REPUBLIC OF TÜRKİYE
YILDIZ TECHNICAL UNIVERSITY
GRADUATE SCHOOL OF SOCIAL SCIENCES
DEPARTMENT OF ECONOMICS

PHD THESIS

TECHNICAL EFFICIENCY AND IRREGULARITIES IN TURKISH MUNICIPALITIES: A DUAL PERSPECTIVE USING DEA AND MACHINE LEARNING

B. BARIŞ EROĞLU
19729002

ADVISOR
PROF. DR. HÜSEYİN TAŞTAN

2025

REPUBLIC OF TÜRKİYE
YILDIZ TECHNICAL UNIVERSITY
GRADUATE SCHOOL OF SOCIAL SCIENCES
DEPARTMENT OF ECONOMICS

PHD THESIS

TECHNICAL EFFICIENCY AND IRREGULARITIES IN TURKISH MUNICIPALITIES: A DUAL PERSPECTIVE USING DEA AND MACHINE LEARNING

B. BARIŞ EROĞLU
19729002
ORCID NO: 0009-0008-4836-4395

ADVISOR
PROF. DR. HÜSEYİN TAŞTAN

SEPTEMBER 2025

B. Barış Eroğlu tarafından hazırlanan “Technical Efficiency and Irregularities in Turkish Municipalities: A Dual Perspective Using DEA and Machine Learning” başlıklı çalışma, **01/09/2025** tarihinde yapılan savunma sınavı sonucunda oybirliği ile başarılı bulunmuş ve jürimiz tarafından İktisat Ana Bilim Dalı İktisat (İngilizce) Programında **DOKTORA** tezi olarak kabul edilmiştir.

Danışman	**İmza**
Prof. Dr. Hüseyin Taştan	……

Jüri Üyeleri	**İmza**
Prof. Dr. Hüseyin Taştan	……
Prof. Dr. Feride Gönel	……
Prof. Dr. Meral Altan	……
Prof. Dr. Emine Tahsin	……
Assoc. Prof. Dinçer Dedeoğlu	……

ABSTRACT

TECHNICAL EFFICIENCY AND IRREGULARITIES IN TURKISH MUNICIPALITIES: A DUAL PERSPECTIVE USING DEA AND MACHINE LEARNING

This dissertation develops an integrated framework that evaluates the efficiency of Turkish municipalities and predicts irregularities from financial accounts, thereby linking service production and governance integrity. The study compiles the largest and most comprehensive dataset to date on Turkish municipalities: a province–year panel spanning 2012–2022 and covering more than 600 municipality–year observations. The dataset integrates unstructured audit report text and structured financial tables extracted algorithmically from Sayıştay reports (in PDF format). Methodologically, the research advances two complementary perspectives: first, Data Envelopment Analysis is applied to measure technical efficiency through service supply and treatment services models, revealing persistent heterogeneity across municipalities and systematic variation by municipality type and service domain; second, novel irregularity indices are derived from Court of Accounts audit reports using text mining techniques such as Latent Dirichlet Allocation (LDA) and are then predicted from financial accounts using Gradient Boosted Machines and Random Forests. The machine learning models yield strong predictive performance, with R^2 values reaching 0.80 for the LDA-based index. The constructed indices together with the machine learning models assign higher weights to irregularities in accounting and asset management, budgeting and financial oversight, and public oversight and social services, highlighting their stronger link to financial statements. On the prediction side, enterprise and ownership revenues, social security expenses, and short-term liabilities emerge as the most influential fiscal predictors of irregularities. SHAP analysis is employed selectively on random samples to illustrate and interpret the model outputs. Econometric panel regressions further extend this analysis by systematically linking efficiency scores with the irregularity indices, thereby offering evidence on how governance failures, particularly in real estate and asset-related domains, translate into lower municipal efficiency. Overall, the dissertation contributes empirically by assembling the richest dataset ever constructed for Turkish municipalities, methodologically by integrating DEA with machine learning on audit-based indices, and substantively by disentangling efficiency from integrity, providing a foundation for evidence-based municipal benchmarking, improving audit performance by providing areas to focus, and policy design.

Keywords: Local Government Efficiency, Data Envelopment Analysis, Audit Reports, Machine Learning

ÖZET

TÜRKİYE BELEDİYELERİNDE TEKNİK ETKİNLİK VE USULSÜZLÜK: VZA VE MAKİNE ÖĞRENMESİ İLE İKİLİ YAKLAŞIM

Bu çalışma, Türkiye il belediyelerinin teknik etkinliğini değerlendiren ve mali tablolardan olası usulsüzlükleri öngören ikili bir yaklaşım geliştirmektedir. Çalışma, 2012–2022 dönemini kapsayan ve 600'den fazla belediye–yıl gözleminden oluşan bir il–yıl paneli ile bugüne kadar Türkiye il belediyeleri üzerine derlenen en büyük ve en kapsamlı veri setini oluşturmuştur. Veri seti, Sayıştay raporlarındaki (PDF formatından) mali tabloların algoritmik yöntemlerle işlenebilir veriye dönüştürülmesi ile denetim raporu bulgularını bütünleştirmektedir. Yöntemsel olarak araştırma iki tamamlayıcı yaklaşım sunmaktadır: Birinci yaklaşımda, bir karar vericinin girdiler ve çıktılar üzerinden teknik etkinliğini ölçen Veri Zarflama Analizi (VZA) uygulanmıştır. Çıktı değişkenleri olarak, temel hizmetler ve arıtma verileri kullanılarak iki ayrı VZA modeli aracılığıyla teknik etkinlik değerleri elde edilerek, belediyeler arasında kalıcı farklılıkları ve belediye tipi ile hizmet alanına göre sistematik değişimleri ortaya koymuştur. İkinci yaklaşımda ise, Sayıştay denetim rapor bulgularını baz alan, Gizil Drichlet Dağılımı (Latent Dirichlet Allocation - LDA) gibi metin analizi teknikleri kullanılarak, usulsüzlük endeksleri türetilmiştir. Bu veri seti mali tablolar ile birleştirilerek, Gradyan Artırmalı Ağaçlar (Gradient Boosted Machines) ve Rasgele Orman (Random Forest) algoritmaları kullanılarak usulsüzlük potansiyellerini öngören modeller oluşturulmuştur. Tahmin sonuçlarında iştirak ve mülkiyet gelirleri, sosyal güvenlik giderleri ve kısa vadeli yükümlülükler, usulsüzlüklerin en etkili mali öngörücüleri olarak öne çıkmaktadır. Bu modelin pratik kullanımını daha net sergilemek için spesifik belediyeler üzerinde uygulama örnekleri Shapley Katkı Açıklamaları (SHapley Additive exPlanations – SHAP) desteğiyle sunulmuştur. Ekonometrik panel regresyonları ise, teknik etkinlik değerlerini usulsüzlük endeksleriyle sistematik biçimde ilişkilendirmiş ve özellikle gayrimenkul ve varlık yönetimi ilgili alanlardaki usulsüzlüklerin belediye etkinliğini düşürdüğüne dair kanıt sunmuştur. Genel anlamda bu tez, ampirik katkı olarak Türkiye belediyeleri için şimdiye kadar inşa edilen en zengin veri setini oluşturmakta; yöntemsel katkı olarak VZA'yı denetim temelli endeksler üzerinde makine öğrenmesi ile bütünleştirmekte; politika katkısı olarak ise ayrı ayrı ele alınan teknik etkinlik ve usulsüzlük analizleri, kanıta dayalı belediye karşılaştırmaları, denetim kurumunu potansiyel usulsüzlüklere odaklayarak denetim veriminin arttırılması ve politika tasarımı için bir temel sağlamaktadır.

Anahtar Kelimeler: Belediye Etkinliği, Veri Zarflama Analizi, Metin Madenciliği, Sayıştay Raporları, Makine Öğrenmesi

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my advisor, Prof. Dr. Hüseyin Taştan, for his professional and rigorous approach throughout this research. His unwavering emphasis on quality and academic excellence served as a guiding force and support at every stage of this work. I am also sincerely thankful to Prof. Dr. Feride Gönel, whose continuous support and mentorship since the beginning of my academic journey at Yıldız have been invaluable, I only wish I had followed her advice more. My special thanks go to Assoc. Prof. Dr. Dinçer Dedeoğlu for kindly accepting to join the committee midway and for his feedback and valuable suggestions.

I am also grateful to lawyers A. İnan Yılmaz and Cemal Polat for their subject-specific support and precious brainstorming sessions regarding the legislative framework relevant to this study.

The blowing wind, the seagulls, the sky and the sea deserve a mention that the toughest months of coding, research and writing practices cannot be possible without the regeneration provided by those, they statistically significantly increased efficiency and decreased irregularities in the thesis.

Above all, my deepest appreciation goes to my beloved partner, Özge Gür, and my precious daughter, Irmak, to whom this thesis is dedicated. Their love, patience, support and unwavering encouragement made this journey possible.

Barış Eroğlu

September, 2025; İstanbul

TABLE OF CONTENTS

ABSTRACT iii
ÖZET iv
ACKNOWLEDGEMENTS v
TABLE OF CONTENTS vi
LIST OF TABLES viii
LIST OF FIGURES x
LIST OF ABBREVIATIONS xi

1. INTRODUCTION 1

2. EVOLUTION OF PUBLIC SPACE AND LOCAL GOVERNMENTS 9
2.1. The Evolution and Discussion of Government Space 9
2.2. Reflections on Local Government Space 11
2.3. Turkish Municipality Framework: Definitions, Resources and Responsibilities 13

3. TECHNICAL EFFICIENCY OF TURKISH MUNICIPALITIES: EVIDENCE FROM DATA ENVELOPMENT ANALYSIS 20
3.1. Literature Review 23
3.2. Data 27
3.2.1. Input Variables 27
3.2.2. Output Variables 30
3.3. Methodology 32
3.4. Results 38
3.4.1. Service Supply Models 39
3.4.2. Treatment Services Models 47
3.5. Discussion 54

4. IRREGULARITY INDICES FROM TURKISH COURT OF ACCOUNTS REPORTS 57
4.1. Literature Review 59
4.2. Methodology and Data 63
4.3. Indices 69
4.3.1. Indicator Count: 69
4.3.2. Weighted Indicator Count: 70
4.3.3. Case Count: 70
4.3.4. Weighted Category-based Case Count: 71
4.3.5. LDA Supported Index: 72
4.4. Comparison and Summary Statistics of Indices 75

5. PREDICTING IRREGULARITIES IN TURKISH MUNICIPALITIES 77
5.1. Introduction 77
5.2. Literature Review 78
5.3. Data 82
5.4. Methodology 82
5.5. Results 89
5.5.1. Case Count Index Results 90
5.5.2. Indicator Count Index Results 91

5.5.3. Weighted Indicator Count Index Results 91
5.5.4. Weighted Category-based Case Count Index Results 92
5.5.5. LDA Supported Index 95
5.5.6. Comparative Assessment of Index Performance and Predictive Modeling 99
5.6. Selected Municipal Case Studies 101

6. BRIDGING EFFICIENCY AND IRREGULARITIES: LINKING DEA SCORES WITH AUDIT-BASED INDICES 107
6.1. Introduction 107
6.2. Literature Review 108
6.3. Data 110
6.4. Methodology 112
6.5. Results 114
6.6. Discussion 122

7. CONCLUSION 125

REFERENCES 131
APPENDIX 138

LIST OF TABLES

Table 1: Municipality Types and Numbers for 2022 14
Table 2: Population Distribution of Municipalities 15
Table 3: Personnel Expenditures of Local Governments of Türkiye by Types of Municipalities 17
Table 4: Staff Decomposition of Municipalities of Türkiye by Municipality Type for 2018 19
Table 5: The Number of Observations (Extracted Balance Sheets) per Year 28
Table 6: Summary Statistics of Variables Used for DEA of Non-Metropolitan Municipalities 29
Table 7: Summary Statistics of Variables Used for DEA of Metropolitan Municipalities 30
Table 8: Yearly Statistics for Output and Input Variables – Non-Metropolitan Municipalities 31
Table 9: Yearly Statistics for Output and Input Variables – Metropolitan Municipalities 32
Table 10: Output Scenarios of DEA Models Matrix 39
Table 11: Slack and Deviation Table for Metropolitan Municipalities, Service Model 43
Table 12: Yearly Slack Analysis for Metropolitan Service Supply Model 44
Table 13: Yearly Slack Analysis for Non-Metropolitan Service Supply Model 44
Table 14: Slack and Deviation Table for Non-Metropolitan Municipalities, Service Model 46
Table 15: Yearly Slack Analysis for Metropolitan Cities, Treatment Model 49
Table 16: Slack and Deviation Table for Metropolitan Municipalities, Treatment Model 50
Table 17: Slack and Deviation Table for Non-Metropolitan Municipalities, Treatment Model 52
Table 18: Yearly Slack Analysis for Non-Metropolitan Municipalities, Treatment Model 53
Table 19: Categories and Category Indicators for Case Determination and Relative Weights Assigned 72
Table 20: LDA Topics 73
Table 21: LDA Topics Summary Statistics 74
Table 22: Summary Statistics of the Indices 75

Table 23: Yearly Summary Statistics for Indices 76
Table 25: Results for Case Count Index 90
Table 26: Metropolitan vs Non-metropolitan Comparison for Case Count Index 91
Table 27: Indicator Count Index Results 91
Table 28: Weighted Indicator Count Index Results 92
Table 29: Weight Assignments of Top 10 Performing Models 94
Table 30: LDA Index Results 95
Table 31: Best Performing Weights for LDA Models 96
Table 32: Ardahan 2021 - SHAP Analysis 102
Table 33: Kayseri 2022 - SHAP Analysis 103
Table 34: Istanbul 2022 - SHAP Analysis 104
Table 35: Izmir 2020 - SHAP Analysis 104
Table 36: Izmir 2020 - LDA Topic Distribution 105
Table 37: Balıkesir 2021 - SHAP Analysis 105
Table 38: Balıkesir 2021 - LDA Topic Distribution 106
Table 39: Correlation Coefficients Between Efficiency and Indices 115
Table 40: Pooled OLS Results – Services Efficiency Model 117
Table 41: Pooled OLS Results – Treatment Efficiency Model 118
Table 42: Fixed Effects - Panel Model 121

LIST OF FIGURES

Figure 1: Türkiye Local Government Expenditures as of % GDP For Years 1978-2017 14
Figure 2: Resource Composition of Local Governments for 2016, OECD Countries 16
Figure 3: Local Government Expenditure Composition for OECD Countries – 2014 17
Figure 4: Yearly Results for Service Supply Model, Metropolitan Municipalities ... 40
Figure 5: Yearly Results for Service Supply Model, Non-Metropolitan Municipalities 41
Figure 6: Efficiency Scores of Treatment Model for Metropolitan Municipalities ... 48
Figure 7: Efficiency Scores of Treatment Model, Non-Metropolitan Subgroup 51
Figure 8: Service Supply and Treatment Services Models, Mean Efficiency Scores 2012-2022 Illustrated on Map 54
Figure 9: Flow Chart for Weighted Category Based Index 65
Figure 10: Flow Chart of LDA Supported Index 68
Figure 11: Word Cloud of Indicator Words for the Sample of 2022 Reports 69
Figure 12: Indicator Term Frequencies for 2022 Reports 70
Figure 13: Indicator Weights 70
Figure 14: Correlation Matrix of Indices Grouped by Municiaplity Type 76
Figure 15: GBM Flow Chart 87
Figure 16: Relative Influences for Case Category Based Index 93
Figure 17: LDA Index Prediction Performance 95
Figure 18: Relative Influences - LDA – GBM 97
Figure 19: Relative Influences LDA - RF Model – MSE 98
Figure 20: Relative Influences LDA - RF Model - Node Purity 98
Figure 21: Correlation Matrix of Node Purity Variables with Fundamentals 99

LIST OF ABBREVIATIONS

BoW:	Bag of Words
DEA:	Data Envelopment Analysis
SFA:	Stochastic Frontier Analysis
ML:	Machine Learning
GBM:	Gradient Boosted Machine
RF:	Random Forest
LDA:	Latent Drichlet Allocation
RMSE:	Root Mean Square Error
SD:	Standard Deviation
TÜİK:	Türkiye İstatistik Kurumu (TURKSTAT - Turkish Statistical Institute)
OECD:	Organisation for Economic Co-operation and Development
PPPs:	Public-Private Partnerships
SHAP:	Shaply Additive Explanations
OLS:	Ordinary Least Squares
FE:	Fixed Effects (Panel Regression)

1. INTRODUCTION

In recent decades, local administrations have experienced a marked expansion in service responsibilities, accompanied by shifts in the division of functions between central and local governments. This transformation has increased the economic importance of local governments as allocators of public resources, providers of essential goods and services, and agents of local economic development. As a result, questions of technical efficiency in resource use, the integrity of fiscal operations, and the accountability of municipal governance have gained greater attention in both academic and policy debates.

This thesis adopts a dual perspective to evaluate municipal performance in Türkiye, combining Data Envelopment Analysis (DEA) to measure technical efficiency with Machine Learning (ML) methods to quantify and predict audit-detected irregularities. The research addresses four central questions:

1. What are the levels, spatial patterns, and temporal trajectories of technical efficiency across provincial municipalities?
2. Can legally grounded, replicable indices of irregularities be constructed from Turkish Court of Accounts (Sayıştay[1]) audit reports, and what do these indices reveal about the evolution and composition of irregularities[2]?
3. To what extent can irregularities be predicted from routine financial statements, and which balance-sheet and operating accounts most strongly drive predicted risk?

[1] Turkish Court of Accounts: Turkish auditory institution responsible for municipalities and other public institutions.

[2] In this research, the term "irregularities" is employed to denote the (judicial) term of "usulsüzlük" in Turkish. This choice is grounded in both institutional practice and linguistic accuracy. The Turkish Court of Accounts (Sayıştay) defines its own audit objectives on the basis of "usulsüzlük" and, in its official English-language publications, systematically translates the term as irregularities. Using this terminology ensures full consistency with the Sayıştay's own usage.

4. Do efficiency scores co-vary with irregularity indices once structural and socio-economic factors are considered, and through which governance domains is any association most pronounced?

These four research questions delineate the analytical framework of the thesis and correspond to distinct yet interrelated dimensions of municipal performance. The first question seeks to uncover the levels, spatial distribution, and temporal evolution of technical efficiency among Turkish provincial municipalities. By applying Data Envelopment Analysis (DEA) across multiple years and jurisdictions, this part of the study aims to identify persistent inefficiencies, (if exists) regional disparities, and trends in service delivery performance, thereby contributing to the literature on public sector productivity in developing and decentralized governance contexts.

The second question addresses the methodological challenge of translating unstructured audit narratives into structured, replicable, and legally grounded indices of irregularities. Drawing upon the official audit reports of the Turkish Court of Accounts (Sayıştay), the study develops text-based measures that capture both the severity and thematic composition of reported governance issues. These indices allow for temporal and cross-sectional comparisons, enabling a systematic investigation of how irregularities evolve over time and across institutional domains.

The third question shifts the focus to predictive modeling and explores whether routine financial and operational statements, such as balance sheet accounts and expenditure categories, contain sufficient informational signals to anticipate irregularities before they are formally detected by audits. By applying machine learning algorithms to financial data, the analysis not only seeks to estimate irregularity risk at the municipal level but also identifies which budgetary or operational variables most consistently contribute to higher predicted risk scores.

Finally, the fourth question integrates the efficiency and irregularity dimensions by examining whether and how the two are empirically related once structural and socio-economic controls are introduced. This part of the study tests the hypothesis that lower technical efficiency may be associated with higher irregularity scores, and further investigates through which governance channels, such as procurement, internal control, or personnel management, this association is most salient. In doing so, the

thesis contributes to a broader understanding of the interplay between performance and accountability in the context of local public administration.

Efficiency benchmarking in local government yields multiple interrelated benefits. It improves economic allocation schemes; resources are used more efficiently which increases public service outputs given fiscal constraints. Provides the tools for budgetary discipline and managerial improvement, guiding optimal resource allocation, cost control, and performance-based budgeting. Enhancing public oversight and trust through accessible performance metrics contributes to transparency and accountability. Improving service quality, coverage, and affordability directly effects citizen welfare.

Increased efficiency frees up resources that can be redirected towards initiatives that attract businesses, stimulate economic growth, and create jobs within the community. This, in turn, fosters a thriving and sustainable local economy, benefiting both residents and businesses alike. Improved service delivery and cost-effectiveness resulting from efficiency efforts can contribute to better social outcomes for citizens. Depending on the framework of responsibility areas of local governments, this may include improved public health, access to quality education, better public transport, and enhanced infrastructure, thereby elevating the overall well-being of the community. Additionally, the efficient use of resources helps minimize waste and environmental impact. This contributes to a more sustainable future for the community, ensuring responsible stewardship of natural resources for generations to come.

From the public officials' perspective, efficiency metrics provide an analytical foundation for data-driven decision-making in budgeting, resource allocation, and strategic planning. Comparative analysis across municipalities enables benchmarking against peers, identification of best practices, and targeted adoption of cost-effective strategies. Demonstrated efficiency gains can improve the credibility of municipal administrations, strengthen fiscal discipline, and support the case for increased fiscal autonomy.

From the citizen's perspective, efficiency assessments translate complex administrative performance into understandable metrics, facilitating informed

participation in elections and public discourse. By identifying and addressing inefficiencies, municipalities can deliver higher-quality services at lower per-unit cost, potentially reducing local tax burdens or expanding service provision without additional resource demands. Enhanced efficiency can also contribute to equitable service access, better public infrastructure, and improved environmental stewardship.

Reliable efficiency analysis requires a comprehensive and consistent dataset, yet in Türkiye, municipality-level studies have historically relied on very small samples. Such limitations have constrained the robustness and generalizability of prior findings. DEA techniques further demand a sufficiently large sample size to produce stable and meaningful efficiency frontiers (Bowlin, 1998).

In the Turkish context, municipality-level research has long been constrained by severe data limitations. As demonstrated in the local literature, most existing studies rely on very small samples, since there is no systematically available dataset that covers all municipalities across time. The only alternative would be to request detailed balance sheets and budgets directly from each municipality, a process that is practically unfeasible given administrative hurdles and inconsistent reporting standards. To overcome these barriers, this study directly addresses these constraints by compiling the largest and most comprehensive dataset to date on Turkish municipalities: a province–year panel spanning 2012–2022 that integrates both unstructured audit report text and structured financial tables extracted directly from Sayıştay Evaluation Reports (that are in PDF file format) using algorithmic data extraction. These financial tables are exceptionally rich, containing not only detailed operational income and expenditure accounts (e.g., personnel, social security, procurement, transfers, tax revenues, property incomes) but also full balance sheets of assets, liabilities, equity, and contingent obligations. This integrity enables the study to jointly analyze fiscal flows and stocks at the municipal level, a level of granularity not present in earlier Turkish literature.

The resulting dataset covers more than 600 municipality–year observations, providing unprecedented temporal and cross-sectional breadth. This expanded coverage forms the basis of input variables for the DEA models. This framework not only enhances the reliability of DEA estimates but also permits the first systematic longitudinal benchmarking of Turkish municipalities at the provincial scale, far exceeding the

observational scope of earlier studies. However, the choice of output variables is necessarily constrained. In Türkiye, systematic data on municipal service delivery are extremely limited, and only a narrow set of indicators is available in official sources. For this reason, the DEA outputs in this study are restricted to the measurable domains of municipal services; namely, basic service provision and treatment services. While a richer output set that includes transportation, road and infrastructure works, nurseries, elder care, cultural services, and environmental protection would substantially enhance the analysis, such data are not consistently available across municipalities or over time. The present design therefore reflects the only feasible output coverage within the Turkish context, enabling robust but necessarily bounded efficiency measurement. Backed by this rich background, two DEA models are estimated as the first empirical analysis:

1. Service supply model: Measuring efficiency in delivering a set of base municipal services.
2. Treatment services model: focusing on treatment services.

These models benchmark municipalities relative to the best-practice frontier, allowing efficiency scores to be interpreted both in cross-sectional and temporal dimensions. Cross-sectionally, they enable comparisons across municipalities of different sizes and institutional capacities by identifying which units transform given fiscal inputs into service outputs most effectively in a given year. Temporally, they trace the trajectory of efficiency for each municipality over the 2012–2022 period, revealing whether local administrations are converging toward or diverging from the efficiency frontier. In this way, the models provide not only a static ranking of municipalities at a point in time but also dynamic insights into how efficiency evolves in response to fiscal pressures, institutional reforms, or changes in service demand.

Besides identifying the inefficiencies emerging from resource allocations and service supply schemes, the irregularities and malpractices of municipalities can be addressed as another source of public resource (in)efficiency. For Turkish municipalities, the Turkish Court of Accounts (Sayıştay) is the supreme audit institution responsible for external auditing. Each year it produces detailed annual audit reports for municipalities, which serve both as compliance checks and performance reviews. These reports are standardized legal documents that contain (i) textual descriptions of

audit findings[3] (cases), including irregularities, procedural breaches, or weaknesses in internal control, and (ii) comprehensive financial statements such as budgets, income–expenditure accounts, and full balance sheets (Representative samples are provided in the Appendix: Figure A2, Figure A3 and Figure A4). Because they are produced through a uniform judicial-administrative process, these reports constitute the most consistent and authoritative source on municipal governance in Türkiye. For the second empirical application of this study, text analysis methods such as Latent Drichlet Allocation are applied to convert unstructured audit text into quantitative indices that can capture the magnitude of the irregularities reported.

The magnitude of irregularities is conceptualized in two dimensions. Integrity of each finding reported by Sayıştay is evaluated by term frequencies that are assigned with a group of specialists, with higher weight given to findings containing high-risk keywords such as “ihale” (procurement), “kamu” (public), “zarar" (loss), “haksız” (unjust), “tahsis” (allocation), among others. The second dimension assigns categories to the findings and assigns weights accordingly for each category. This approach ensures that indices reflect both the frequency and potential fiscal significance of the findings.

These indicators of irregularities, together with the municipality–year financial dataset, are used to train predictive models based on tree-based machine learning methods, specifically Gradient Boosting Machines (GBM) and Random Forests (RF). At the heart of this application lies a fundamental question: can governance irregularities be predicted from routine municipal financial statements? The claim behind this question can be summarized as: irregularities such as mismanagement, fraud, and corruption often leave quantifiable traces in fiscal accounts, even before they are formally detected in audits. GBM and RF are particularly suited to this context because they can capture complex, nonlinear relationships and interactions between fiscal variables without requiring restrictive parametric assumptions. Cross-validated hyperparameter tuning ensures robust performance, while SHAP (SHapley Additive exPlanations) is employed selectively to illustrate how individual financial variables contribute to predicted irregularity scores. In this way, the framework integrates

[3] To improve readability, these two concepts are used interchangeably throughout the text referring the same concept. Sayıştay originally uses “bulgu” referring to findings.

modern statistical learning with governance diagnostics, transforming retrospective audit findings into forward-looking risk assessments.

This methodology has several applied and theoretical contributions. Irregularities such as mismanagement, fraud, and corruption impose significant economic costs by diverting scarce public resources away from productive uses and distorting allocative efficiency. By systematically detecting and predicting these practices, the framework supports better resource stewardship. Similar to the DEA analysis, it will contribute to transparency of public institutions. By uncovering explicit relationships between identified irregularities and specific expenditure patterns, this process translates complex and technical data into policy-relevant evidence, thereby promoting transparency and accountability in local governance.

Moreover, the predictive capability of ML models offers substantial practical utility. Given a municipal balance sheet, the developed predictive algorithm can effectively estimate the likelihood of potential irregularities occurring. With sufficient predictive accuracy, this algorithm serves as an invaluable tool for municipal administrators and auditing institutions, enabling proactive identification and targeted investigation of suspicious financial accounts. Additionally, ML procedures provide insights into the most influential variables contributing to irregularity indices. Identifying these critical variables equips audit institutions, policymakers, and local government officials with targeted analytical insights, significantly enhancing their ability to detect, manage, and mitigate irregularities and financial risks proactively. Ultimately, these insights support evidence-based policymaking and efficient allocation of audit resources, further reinforcing effective municipal governance.

Building on these two strands, last empirical analysis explicitly integrates DEA efficiency scores and irregularity indices through econometric panel regressions. This step provides the first systematic evidence on how governance irregularities and technical efficiency co-vary across Turkish municipalities. By disaggregating the indices into topical categories, the analysis identifies which governance domains, such as procurement, asset management, or real estate, most strongly condition efficiency outcomes. This integrative perspective moves beyond separate treatments of efficiency and irregularity, thereby demonstrating how the dual dimensions of performance interact in shaping municipal outcomes.

The thesis proceeds as follows:

- Chapter 2 examines the evolution of public space and local governance, including the conceptual background, decentralization debates, and the Turkish municipal framework.
- Chapter 3 applies DEA to benchmark municipal technical efficiency, presenting literature, data, methods, and results for service supply and treatment services models.
- Chapter 4 details the construction of audit-based irregularity indices from Sayıştay reports using text analysis and LDA topic modelling.
- Chapter 5 develops ML models to predict irregularities from financial data and interprets their outputs, illustrates practical usage using SHAP analysis.
- Chapter 6 contrasts DEA efficiency scores with irregularity indices, using both correlation and regression analysis (pooled OLS and fixed effects) to explore topic-specific linkages.
- Chapter 7 concludes with key findings, contributions, policy implications, and directions for future research.

2. EVOLUTION OF PUBLIC SPACE AND LOCAL GOVERNMENTS

The evolution of local governments is closely intertwined with broader transformations in the public sphere. The debate regarding government intervention in economic and public domains forms a central axis around which different theoretical perspectives diverge. Indeed, varying attitudes toward the role and extent of public intervention define distinct schools of political and economic thought. Moreover, contemporary political history can often be characterized by varying levels of governmental involvement. To comprehensively understand the evolving role of local governments, it is essential to first examine these broader discussions and the historical transformations of public involvement.

2.1. The Evolution and Discussion of Government Space

Over the past two centuries, the modern state's presence has notably expanded in both political and economic domains. The industrial revolution of the late 19th century significantly elevated public expenditures, driven by burgeoning economic activity and associated societal demands. In the early 20th century, innovations in production techniques, notably Fordist production models, and the rising prominence of finance capitalism introduced new systemic risks and economic crises, further necessitating robust public intervention. The centralized planning and governance structure of the Soviet Union exemplifies how intensive public intervention facilitated rapid industrialization and significant social advancements. Similarly, the Keynesian framework, prevalent post-World War II, advocated for substantial government economic intervention to address systemic market failures. During this era, urbanization remained relatively limited, resulting in constrained roles and responsibilities for local governments compared to their central counterparts (Tanzi & Schuknecht, 2000). However, the economic turmoil of the 1970s, marked by the oil crisis, stagflation, and declining productivity growth, catalyzed a shift toward neoliberal ideology. This school of thought critically questioned extensive government intervention, advocating instead for minimal public presence and greater reliance on

market mechanisms. Questioning government intervention and economic existence and an effort to redefine public space gain popularity both in literature and political space.

Neoliberal proponents argue that excessive public spending undermines economic growth, individual freedoms, and overall efficiency. They posit that government institutions are inherently less efficient than private enterprises, leading to resource misallocation, corruption, principal-agent issues, and rent-seeking behaviors. According to this perspective, markets inherently optimize resource allocation through supply-demand dynamics, purportedly leading to enhanced efficiency and productivity.

Conversely, advocates for robust governmental involvement emphasize inherent market inefficiencies and shortcomings (Bator, 1958; Tobin, 1970). From their viewpoint, unrestricted markets often fail to adequately provide essential services, such as healthcare, education, housing, public infrastructure, and environmental protection. These inadequacies necessitate government intervention to ensure equitable service provision. Additionally, unregulated markets can precipitate economic instabilities, including financial bubbles and market crashes, leading to widespread socio-economic repercussions (Minsky, 1986). Furthermore, market inefficiencies emerging from monopolies and other imperfect market issues, profit-driven behavior within markets may generate unsustainable practices, environmental degradation, and resource depletion, underscoring the need for public regulatory mechanisms to address negative externalities.

Theoretical contributions by Habermas also enrich this discussion by emphasizing the role of public deliberation and communicative rationality. According to Habermas, a well-functioning public sphere, characterized by open discourse and active citizen participation, is crucial for legitimate governance and effective decision-making. The decentralization of authority, in this regard, aligns closely with Habermasian ideals by potentially enhancing democratic participation, transparency, and accountability. Local governments, being closer to the citizens they serve, can foster environments where communicative rationality thrives, facilitating more informed and consensus-oriented policy outcomes (Habermas et al., 1992).

In principle, achieving an efficient model of public intervention could potentially reconcile these competing arguments. Effective and transparent governance structures could negate many neoliberal criticisms regarding public inefficiency. Concurrently, governmental proponents' arguments gain strength if public interventions demonstrably resolve market failures efficiently. Nevertheless, historical evidence consistently demonstrates persistent inefficiencies and potential misuse associated with the management of public resources, irrespective of developmental stages or governance structures. Consequently, continuous efforts toward enhancing public governance efficiency remain indispensable.

2.2. Reflections on Local Government Space

Until the latter part of the 20th century, centralized governance dominated, significantly shaping local government structures. During this period, local authorities typically had limited autonomy, resources, and responsibilities. Primarily funded by central government grants, local governments provided basic public services, with minimal emphasis on citizen participation, transparency, or accountability.

Rapid urbanization in the latter half of the 20th century, combined with evolving political ideologies, redefined the roles and responsibilities of local governments. Numerous countries embraced decentralization, transferring powers from central to local authorities to enhance governance responsiveness, local decision-making, and service delivery efficiency. Decentralization is often advocated for its potential to enhance accountability, empower local communities, foster democratic participation, and stimulate innovation through localized experimentation. Moreover, decentralization can promote political stability by addressing regional disparities, reducing tensions, and facilitating intergovernmental collaboration.

Besides the arguments supporting decentralization, rapid urbanization from the late 20th century into the early 21st century significantly reshaped the scope and responsibilities of local governments. The dramatic growth of urban populations imposed considerable demands on local infrastructure, leading to increased expenditures on public transportation, housing, sanitation, and utilities. Additionally, social welfare responsibilities expanded considerably, with local governments intensifying efforts to address poverty, homelessness, and healthcare accessibility.

Inclusive policies aimed at supporting marginalized communities, including specialized services for individuals with disabilities, emerged as critical components of local governmental agendas.

Decentralization processes have both expanded and diversified the expenditure patterns of local governments. While core areas such as infrastructure, education, and healthcare continue to dominate expenditure portfolios, new priorities like environmental sustainability, cultural initiatives, tourism promotion, and social inclusion programs have also gained prominence. These changes reflect evolving community expectations and demands. Particularly noteworthy is the financial significance of education and healthcare services, which continue to heavily influence local government spending. The extent of local governments' involvement in these domains directly affects the balance of resource allocation and autonomy between local and central authorities. Empirical evidence from various countries indicates that increasing local government responsibility for education and healthcare typically correlates with higher overall local government expenditures.[4]

While decentralization offers substantial benefits, it also introduces potential fiscal risks. Inefficient fiscal management can lead to unsustainable spending levels, resulting in local debt accumulation and consequent fiscal pressures on central governments (Turati et al., 2012). Additionally, international financial institutions advocating decentralization in the 1980s often tied their support to market liberalization and privatization policies, sometimes resulting in inefficiencies and adverse socio-economic outcomes. World Bank and IMF, as the major advocates of decentralization policies, pawed the way for a strong flow of privatization of public services that were usually under local government control. This process resulted in several examples of inefficient private service supplies. Furthermore, to achieve efficiency targets, this transformation resulted in cutting labor costs especially when contracting out is involved, which in turn added macroeconomic and social costs. (International Labour Organisation, 2001).

In contemporary societies, the necessity of local governments is justified through various market failures and public goods provision requirements. Public goods such

[4] For details refer to Figure A1 which shows the expenditure allocation schemes of local governments versus the expenditures as of GDP in comparison in appendix.

as street lighting cannot be efficiently provided by markets due to free-rider problems. Secondly, firm behavior tends to ignore negative externalities, necessitating governmental regulatory interventions through taxation, subsidies, or legislative measures. Monopolistic service provisions and information asymmetries further underscore the importance of local governmental oversight. As the profit focused firms and managements seek minimal costs, there are several examples and complaints regarding dismissed public health considerations. Finally, we can add services that will create positive externalities (potentially in the long term). Libraries, nurseries, nutritional supports can be listed as examples. These also underline the benefits of transferring the authorities and resources from central government to local ones, as local governments can much better diagnose the most important needs of the local community and these needs strongly differ regionally (Adaman & Çarkoğlu, 2000; Batırel, 1990).

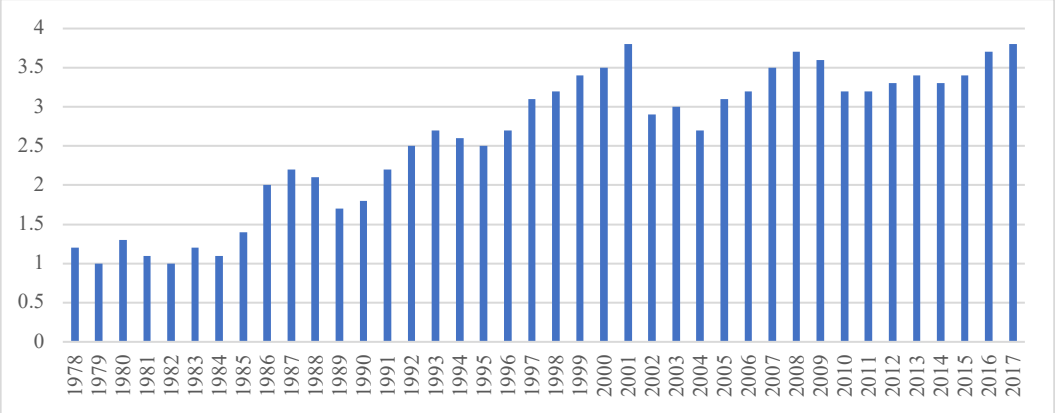
2.3. Turkish Municipality Framework: Definitions, Resources and Responsibilities

Türkiye, despite following global decentralization trends, remains among countries with relatively limited decentralization. The level of local government involvement in education, healthcare, and social protection significantly influences their expenditure and autonomy. For example, Eurozone countries with the highest local government expenditures on education and healthcare allocate, on average, 17.7% of GDP to local governments (OECD, Eurostat, 2013). Conversely, the lowest ten countries allocate approximately 6.1%. Türkiye's average spending on education and healthcare at the local level stands around 6.3% for the years 2008-2022, reaching 7% in 2022, while total local government expenditures form only 3.4% of GDP (TÜİK). Greece provides a comparable case, with local education and health expenditures at 9.9% despite local governments accounting for only 3.6% of GDP.

Other significant indicators of decentralization include local governments' share of overall public spending and employment. In 2021, the European Union, United States, and OECD averages for local government expenditures as a share of GDP were 11.3%, 7.8%, and 12.7%, respectively (Eurostat, FRED, OECD). Türkiye, at 3.4%, ranks among the lowest within the OECD, reflecting its centralized governmental structure,

although there has been an upward trend in recent decades (Figure 1). Local governments' role as public employers further highlights their economic significance. In 2014, OECD local governments accounted for approximately 63% of overall public staff expenditures (OECD, 2016), underscoring personnel costs as a substantial component of local expenditures.

Figure 1: Türkiye Local Government Expenditures as of % GDP For Years 1978-2017



Source: Strateji ve Bütçe Başkanlığı

Municipalities constitute one of the three[5] main categories of local government bodies in the Republic of Türkiye (Local Government Law, Law No. 6360, 2013). There are four types of municipalities[6]: metropolitan municipalities, provincial municipalities, district municipalities, and town municipalities. The definitions and regulations governing municipalities are outlined in the Law of Municipalities (Law No. 5393, 2005). In addition, metropolitan municipalities and their affiliated districts are subject to a separate regulatory framework, the Law of Metropolitan Municipalities (Law No. 5216, 2004). Table 1 presents the overall distribution of municipalities, while Table 2 highlights the distribution of population across these entities. As shown in Table 2, the population distribution is markedly uneven, four of the largest metropolitan cities account for 36.3% of the national population, with the top two alone comprising 26.8%.

Table 1: Municipality Types and Numbers for 2022

Metropolitan	Metropolitan District	Province	District	Town	Total
30	519	51	403	388	1391

Source: YYGM, 2023

[5] The others are special administration of province ("il özel idaresi"), and village.
[6] Respectively: "Büyükşehir Belediyesi", "İl Belediyesi", "İlçe Belediyesi" and "Belde"

Table 2: Population Distribution of Municipalities

	Metropolitan	Province	Metropolitan District	District	Town	Total	Population
0 - 2.000			1	42	93	136	222.919
2.001 - 5.000			13	131	259	403	1.212.464
5.001 - 10.000			41	89	34	164	1.148.235
10.001 - 20.000			81	67	1	149	2.117.352
20.001 - 50.000		5	130	54	1	190	6.044.065
50.001 - 75.000		5	48	11		64	3.915.038
75.001 - 100.000		6	23	5		34	2.968.903
100.001 - 250.000		26	91	4		121	18.669.119
250.001 - 500.000		9	64			73	25.711.770
500.001 - 750.000	1		18			19	11.641.270
750.001 - 1.000.000	5		9			14	12.080.063
1.000.001 - 2.000.000	14					14	18.724.709
2.000.001 - 3.000.000	6					6	13.661.690
3.000.001 - 5.000.000	2					2	7.656.776
5.000.001 -	2					2	21.690.236
Total	30	51	519	403	388	1.391	80.810.660

Source: TÜİK ADNKS, 2023

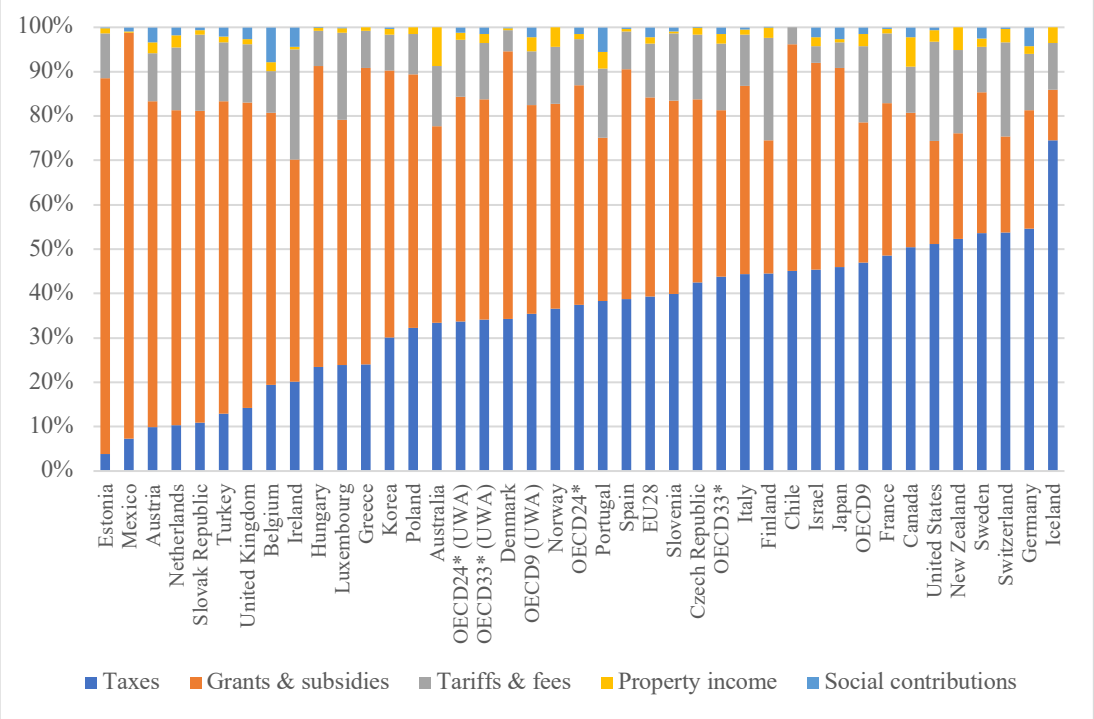
According to the Law of Municipalities, a residential area must have a minimum population of 5,000 to be eligible for municipal status. Additionally, every province and district in Türkiye is required to be governed by a municipality. To establish a metropolitan municipality, the population threshold is set at 750,000. It is noteworthy that, as of 2021, metropolitan municipalities account for approximately 78% of the country's total population.

The financial resources of municipalities consist of allocations from the national tax revenue, internal revenues, transfers from the central government, and borrowing. Internal revenues include fees, income generated by municipal enterprises, grants, and various designated funds.

An analysis of the composition of local government resources in Türkiye reveals that approximately 70% originate from central government transfers. This high proportion raises concerns regarding the fiscal autonomy of local governments and stands in contrast to the global trend toward decentralization. On average, tax revenues

constituted only 10% of municipal resources during the period 2008–2022 (Source: TÜİK). In comparison, OECD countries reported an average composition of 34% grants and 50% tax revenues in 2016 (OECD, 2016). Figure 2 illustrates the resource composition across OECD countries.

Figure 2: Resource Composition of Local Governments for 2016, OECD Countries

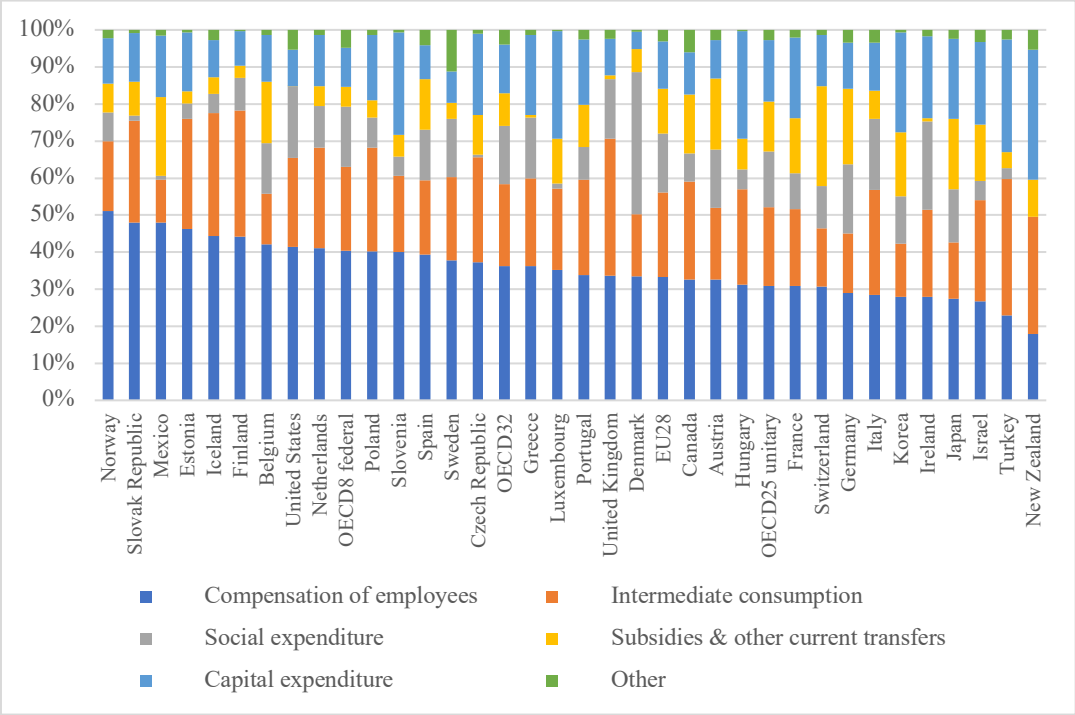


Source: OECD

The competences of local governments vary significantly across countries. In the case of Türkiye, municipalities are primarily responsible for planning and sanitation services, water and clean water supply, transportation, health services (to a limited extent), environmental services, auditing and allocation services, and, to a lesser degree, educational and cultural services. The expenditure structure of local governments is categorized into three main components: recurrent expenditures, investment expenditures, and transfer payments. Recurrent expenditures are predominantly composed of personnel costs, while investment expenditures include spending on construction, machinery, and technical analyses. Transfer payments encompass expropriation costs, debt repayments, economic transfers, and fiscal transfers. Figure 3 presents the composition of local government expenditures across OECD countries, including Türkiye, for the year 2014. The most recent data for

Turkish local governments indicate no substantial change, showing that 59.1% of expenditures are allocated to goods and services and 18.6% to personnel expenses.

Figure 3: Local Government Expenditure Composition for OECD Countries – 2014



Source: OECD

Beyond their share in public expenditures, the role of local governments as employers further underscores their economic significance. In 2016, the average share of public employment expenditures by local governments across the OECD was 36%, whereas Turkish municipalities accounted for only 8.75% of public employment expenditures in 2022. Table 3 and Table 4 provide detailed information on the employment composition and corresponding shares within Turkish municipalities.

Table 3: Personnel Expenditures of Local Governments of Türkiye by Types of Municipalities

	Personnel Expenditures	Total Expenditures	%
Metropolitan Municipalities	11.815.237	198.277.862	5,96
Municipalities	3.169.662	30.442.444	10,41
District Municipalities	19.820.931	169.047.691	11,73

Source: 2022 Yılı Mahalli İdareler Genel Faaliyet Raporu

Table 3 highlights the variation in personnel expenditure shares among different types of municipalities. While metropolitan municipalities allocate only 5.96% of their budgets to personnel expenses, district and provincial municipalities allocate 11.73% and 10.41%, respectively. This disparity suggests that larger municipalities have a higher share of outsourced services and capital investments, while smaller municipalities rely more intensively on human resources for service delivery. Table 4 provides a detailed breakdown of the employment structure across various types of local governments. The majority of staff are concentrated in metropolitan district municipalities (79,747 employees), followed by metropolitan municipalities (52,526), highlighting the importance of these administrative levels in urban service provision. Among all roles, technical staff (27,833) and administrative staff (25,266) dominate the composition, indicating the technical and bureaucratic nature of municipal functions. Notably, critical public service roles such as fire department (8,437) and local security ("Zabıta") staff (15,710) also represent a substantial portion, reflecting the operational functions of municipalities in ensuring urban safety and public order.

The Turkish Law of Municipalities grants municipalities the authority to establish municipal subsidiaries. These subsidiaries function as semi-autonomous enterprises that remain accountable to the municipality while being exempt from the stricter labor regulations governing the public sector. Notably, these entities maintain independent budgets, which are, however, consolidated into the municipal budget on an annual basis (Küçük, 2015). In 2022, municipalities employed a total of 194,354 individuals, representing 4.49% of total public employment. In contrast, municipal subsidiaries accounted for 13.05% of total public employment, with 564,354 employees. The total number of public sector employees in 2022 was 4,328,197 (2022 Yılı Mahalli İdareler Genel Faaliyet Raporu, 2023).

Table 4: Staff Decomposition of Municipalities of Türkiye by Municipality Type for 2018

Type of Local Government	Administrators	Administrative Staff	Technical Staff	Health Staff	Assistive Staff	Local Security ("Zabıta")	Fire Department Staff	Other Staff	Total Officer	Contracted Staff	Workers - Permanent	Workers - Temporary	**Total**
Metropolitan Municipalities	2.322	6.522	12.010	1.090	444	3.585	6.133	473	32.579	6.078	13.605	264	52.526
Provincial Municipalities	940	2.224	2.368	134	147	1.385	1.052	349	8.599	1.188	6.068	1.307	17.162
Metropolitan District Municipalities	4.864	13.705	11.990	1.123	636	9.180	167	1.048	42.713	8.135	25.590	3.309	79.747
District Municipalities	1.005	2.084	1.210	173	276	1.265	891	424	7.328	1.441	7.828	1.164	17.761
Town	240	731	255	18	84	295	194	254	2.071	316	2.517	603	5.507
Total	9.371	25.266	27.833	2.538	1.587	15.710	8.437	2.548	93.290	17.158	55.608	6.647	172.703

Source: 2018 Mahalli İdareler Faaliyet Raporu

3. TECHNICAL EFFICIENCY OF TURKISH MUNICIPALITIES: EVIDENCE FROM DATA ENVELOPMENT ANALYSIS

The notable increase in service demand faced by local governments, accompanied by a reallocation of responsibilities between central and local governments are discussed in the previous part. This transformation has elevated the economic and political significance of local governments, positioning them as key actors in governance and public service delivery. Consequently, issues related to public resource allocation, corruption, and local government efficiency have gained increasing attention. Motivated by these considerations, this chapter examines the technical efficiency of Turkish province-level municipalities using Data Envelopment Analysis (DEA), drawing on a 2012–2022 province–year panel that merges, municipal budget and balance-sheet records, biannual municipal service statistics and official demographic-economic statistics. Inputs are constructed from operational expenditure items, while outputs are specified under two frontiers, a service-provision bundle of core municipal service supply and a treatment services bundle for environmental and utility services.

Before constructing a methodology to measure local government efficiencies, it is necessary to define efficiency properly. Efficiency is a measure of how effectively resources, such as time, money, or labor are utilized to achieve specific objectives. It refers to the extent to which a process, system, or organization produces desired outcomes with minimal waste or effort. Efficiency is commonly expressed as the ratio of output to input or as the percentage of resources used relative to outputs achieved. In economic terms, efficiency frequently describes the optimal allocation of resources to maximize productivity and output. Generally, higher efficiency implies greater production with fewer resources, resulting in lower costs and increased productivity. While output for firms typically equates to profit, the definition becomes more ambiguous for public organizations, such as governments or municipalities.

To perform a technical efficiency analysis, it is essential to clearly define a set of input and output variables. Inputs are typically observable on balance sheets as expenditures

and resource allocations. For output variables, the context is sociological, political, and localized, which precludes defining a universally applicable set of variables. Religious services may be a priority for a highly religious community, whereas elderly care may be significant in communities with a large elderly population. Local geography is another factor influencing service needs. Flood protection measures are critical for municipalities in the Netherlands, whereas earthquake-resistant urban planning and building regulations are significant for Türkiye. Additionally, limitations in available data and difficulties measuring certain variables, particularly qualitative ones, pose further challenges to measuring public services.

Putting aside data availability concerns momentarily, we can discuss an ideal framework of metrics for evaluating local government efficiency. Service delivery emerges as the most critical measure, encompassing not only service volume but also service quality and response speed. Customer satisfaction surveys, request response times, and the proportion of issues resolved upon first contact are key performance indicators. Services should also be weighted according to their community priorities (Adaman & Çarkoğlu, 2000).

Another important efficiency measure is the effective management of budgets and resources, including maintaining balanced budgets, managing debt levels, and ensuring cost-effective program implementation. Transparency and accountability towards constituents can be assessed through public record policies, compliance with open meeting laws, and responsiveness to public information requests.

Community engagement represents another critical efficiency indicator in contemporary societies. Efficiency in this domain can be gauged by the extent to which local governments involve diverse community groups in decision-making processes and the effectiveness of their public outreach efforts.

Staff productivity also significantly contributes to local government efficiency. Productivity can be measured using indicators such as staff-to-population ratios, training and development initiatives, and employee absenteeism levels.

While the ideal framework described above reflects contemporary societal needs, actual studies in the literature often make varied selections of output variables due to data availability constraints and measurement challenges.

Efficiency analysis of local governments offers numerous benefits to various stakeholders. For citizens, enhanced transparency and accountability are primary benefits, as efficiency assessments provide clearer insights into administrative performance, enabling informed decision-making during elections and policy debates, thus fostering greater civic engagement. Improved efficiency can also enhance the quality of public service delivery by identifying inefficiencies, guiding administrators towards targeted improvements. Citizens further benefit from increased cost-effectiveness in public spending, potentially translating into lower tax burdens or more affordable municipal services, such as public transportation. Beyond economic gains, efficiency assessments can enhance transparency and democratic participation, thereby contributing to long-term community welfare (Adaman & Çarkoğlu, 2000).

From public officials' perspective, efficiency analysis supports data-driven decision-making by providing valuable insights into departmental performance. Efficiency metrics facilitate informed decisions on resource allocation, budgeting, and service delivery strategies, promoting an objective, evidence-based governance approach. Benchmarking efficiency metrics across similar municipalities fosters knowledge exchange, enabling officials to identify best practices and continuously improve their strategies. Demonstrating commitment to efficiency and transparency significantly enhances public perception of local governance, fostering trust and citizen support. Audit institutions also benefit from efficiency analyses, as identified inefficiencies can signal potential mismanagement, fraud, or corruption, guiding their auditing efforts.

From a broader community perspective, increased efficiency frees resources that can be redirected toward initiatives promoting economic growth, job creation, and local business attraction, fostering a sustainable local economy. Improved service delivery and cost-effectiveness can also enhance social outcomes, potentially improving public health, education access, transportation, and infrastructure, thus elevating community well-being. Furthermore, efficient resource utilization helps minimize waste and environmental impact, supporting sustainable resource stewardship for future generations.

To achieve these objectives reliably, comprehensive datasets are critical. However, in Türkiye, local government data remains relatively limited, significantly hindering academic research growth in this field. Techniques like DEA, used for measuring technical efficiencies, require large observation numbers for reliable results (Bowlin, 1998). Recognizing this challenge, this study employs a novel data collection strategy, leveraging advancements in text analysis and data extraction methods, resulting in a comprehensive dataset of over 600 observations derived from annual reports published by the Turkish Court of Accounts (Sayıştay). Covering municipality-level data over twelve years, this dataset significantly enhances the reliability of DEA results.

The remainder of this chapter is organized as follows. The next section reviews the relevant literature on the efficiency of local governments, with particular emphasis on the choice of input and output variables in Data Envelopment Analysis (DEA) applications and the evidence from Türkiye. The following section introduces the dataset, detailing the construction of input and output variables, the data extraction process from Sayıştay reports, and the coverage achieved across municipalities and years. The subsequent section outlines the methodological framework, presenting the DEA models applied, their orientation and returns-to-scale assumptions, and the rationale for adopting a non-parametric approach. The empirical results are then reported separately for service supply and treatment services models, complemented by slack analysis to identify sources of inefficiency. The chapter concludes with a discussion that situates the findings in relation to existing literature and highlights their policy implications for local governance in Türkiye.

3.1. Literature Review

The assessment of public sector performance has long been a central concern in both economics and public administration. As governments and public institutions seek to improve service delivery and resource allocation, various analytical tools have been developed to measure efficiency. Among these, Data Envelopment Analysis (DEA) has become one of the most widely used non-parametric methods for evaluating the efficiency of decision-making units. Originally developed for non-profit and public organizations such as hospitals, schools, and military institutions, DEA has since been applied across a wide range of sectors, including banking and private industry (Bowlin,

1998). In evaluating technical efficiency in non-profit contexts, where outputs typically consist of services rather than tangible goods, the choice of output variables becomes particularly important. An extensive review and comparative analysis of output variables used in measuring local government performance is provided by Narbón-Perpiñá & De Witte (2018).

In the efficiency analysis of local governments, one of the most frequently used output variables is the total population served, which is often interpreted as a proxy for the attractiveness or demand level of a given region. However, relying on population as an output requires caution. In the absence of balanced regional economic and industrial development, population figures may reflect migration driven by broader economic conditions rather than the actual performance or efficiency of local governments, a limitation that is particularly relevant in the Turkish context. As a result, many studies incorporate service and infrastructure indicators as more direct output measures. These typically include variables such as street lighting, municipal roads, waste collection, sewage systems, water supply, and electricity. In recent years, additional indicators, such as the provision of parks, sports, cultural amenities, and social services, have gained prominence, reflecting the evolving expectations of citizens from local administrations. Depending on the legal responsibilities assigned to local governments, education and healthcare services may also be considered as output variables. However, in Türkiye, these domains are primarily managed by the central government, limiting their suitability as municipal-level outputs.

One of the pioneering applications of DEA to local governments was conducted by De Borger & Kerstens (1996). To measure the cost efficiency of local governments in Belgium, they employed both parametric and non-parametric methods to compare social, political and economic characteristics of municipalities. Prieto & Zofio (2001) utilized water, sewage, wastewater, roads and lighting services as output variables to measure the efficiencies of Spanish municipalities. The DEA results were used to propose recommendations for resource allocation by the central government to improve these services. Similar research with a broader temporal scope was carried out by Tupper & Resende (2004) for Brazilian municipalities.

A distinct approach was adopted by Woodbury & Dollery (2004) who emphasized the classification of output variables into qualitative and quantitative dimensions. In their

study, the quality and sustainability of water supply were incorporated as quantitative outputs. Another notable DEA application on Brazilian municipalities was conducted by De Sousa et al. (2005), where a strong correlation on efficiency with the size of the municipality is identified. They also noted that natural, political, geographical, demographic, and socio-economic factors contributed to efficiency disparities. In the Portuguese context, Afonso & Fernandes (2008) measured the technical efficiency of 278 municipalities using a composite output indicator and concluded that considerable potential existed for enhancing municipal efficiencies.

Studies examining the efficiency of Turkish municipalities remain relatively scarce in the existing literature. Çağlar (2003) studied the efficiencies of 81 municipalities using DEA. The study relied on data from a specialized government initiative known as the "Local Municipalities Database Project" ("Yerel Yönetimler Bilgi Tabanı Projesi – YERELBİLGİ"), which covered only the year 2001 with an unbalanced data structure. Four different DEA models were constructed using varying combinations of input and output variables. The input variables included current expenditures, investment expenditures, transfer payments, total personnel, total vehicles, sewer network length, zoning staff, number of garbage trucks, number of garbage containers, number of garbage collection staff, tap water capacity, tap water network length, tap water storage capacity, and number of waterworks staff. The output variables included the surface area of municipal jurisdiction, tax revenues, non-tax revenues, grants and funds, the total number of building permits issued, garbage collected, number of households with water coverage, total household water consumption. The findings indicated, consistent with broader literature, that larger municipalities and regions predominantly consisting of large municipalities -such as Marmara region- exhibited higher efficiency scores.

Kaplan, Çelik and Tekeli (2006) conducted a study using data from 16 metropolitan municipalities covering the period from 2002 to 2004. They use a DEA model with the following input variables: personnel expenditures, investment expenditures, social grants, educational and cultural expenditures, and transfer payments. Output variables included population, total recreational area, daily water consumption, and passenger carrying capacity of public transport for peak hours. A study employing a similar analytical framework was conducted by Güneş and Akdoğan (2007), who utilized nearly identical input and output variables with the same set of metropolitan

municipalities. They additionally implemented a model with security and well-being focus where they cover fire trucks and local security personnel ("zabıta").

İlkay & Doğan (2009) choose to focus the research on a narrow area of selected 14 district municipalities of Cappadocia Region for the years 2004 and 2008. They established four different models: garbage services model, zoning services model, financial model, and water services model. In the garbage services model the input variables included the number of garbage collecting personnel, number of garbage vehicles, population coverage while the output variables were municipal jurisdiction area and the total garbage collected. For the zoning services model, the input variables consisted of municipal jurisdiction area, zoning personnel, whereas the output variables were the total zoned area and the total number of building permits issued. In the financial model, the input variables included population coverage, municipal jurisdiction area, tax revenues, non-tax revenues with current expenditures and investment expenditures as output variables. Lastly the water services model employed total water supplied, water network length and water personnel as inputs and total water consumption and total households served as outputs.

Kutlar et al., (2012) analyzed the efficiency of 27 municipalities, including seven metropolitan municipalities, covering the period 2006 to 2008. They applied both DEA and Malmquist index methods to evaluate efficiencies. Personnel expenditures, social security expenditures, goods and services expenditures, current transfer expenditures, capital expenditures, capital transfers and total expenditures are the input variables used. Output variables included total population, proportion of 65+ population, number of pupils, the number of beds in tourism establishments, total number of beds in hospitals and the number of visitors. They used both input-oriented and output-oriented frameworks under constant and variable returns to scale assumptions. Their findings indicated a slight decrease in efficiency scores over the studied period (2006–2008). Contrary to prevailing literature, their results revealed no significant correlation between municipality size and efficiency.

Çelikkaya analyzed efficiency using data from 2015 for 30 metropolitan municipalities (Çelikkaya, 2016). Four different models employed in the research are financial model, infrastructure model, expenditures model and tax model. In the financial model, input variables consisted of personnel expenditures, transfer payments, purchases of goods

and services, capital expenditures whereas output variables included tax revenues, non-tax revenues and grants and funds. For the infrastructure model, the input variables included total number of personnel, number of vehicles, total expenditures and output variables were the length of roads, length of sewer networks, length of water networks and municipal jurisdiction area. For the expenditures model, the input variables consisted of personnel expenditures, social security expenditures, purchases of goods and services, interest payments, current expenditures, capital expenditures with output variables defined as municipality covered population, number of students, municipal jurisdiction area. Lastly for the tax model, the input variables were municipality covered population, municipal jurisdiction area with tax revenue as output variable.

Local literature on municipal efficiency in Türkiye typically use short-span data and often focus on limited samples such as a selection of metropolitan municipalities, regional district municipalities or specific regions. The models developed employed range of output variables, such as water, waste, infrastructure, financials and zoning services. However, possibly due to data limitations, many analyses rely on small sample sizes and unbalanced datasets, frequently sourced from non-recurring projects or surveys. While the studies provide valuable insights over municipal administrations, their comparability and time scope are limited.

3.2. Data

3.2.1. Input Variables

As required by the DEA framework, two sets of data are necessary: input and output variables. The input variables are derived from the expenditure titles from the financial statements of municipalities. However, in the Turkish case, there is a lack of available data covering a broad range of years and municipalities. This data limitation is also evident in prior research within the local literature, where studies typically have limited temporal and geographical coverage. Nevertheless, according to the Municipality Law, all municipalities are required to submit their annual balance sheets to central institutions, including the Ministry of Interior, the Ministry of Environment, Urbanization and Climate Change, and the Turkish Court of Accounts (pursuant to Law No. 5393 on Municipalities and Law No. 5216 on Metropolitan Municipalities).

The Turkish Court of Accounts audits municipal budgets and reports any issues along with the balance sheet tables each year. Although these tables are publicly accessible, they are embedded within report documents rather than available as structured datasets. In this study, we employed text mining techniques to extract the relevant financial data from these embedded tables.

Turkish Court of Accounts selects municipalities for audit based on a set of criteria. These include risk assessment, budget size, recent audit results, public interests and expectations as well as claims and notifications. For instance, in 2022, 42% of district municipalities, 44% of municipalities, and 100% of metropolitan municipalities were audited and reported. As a result, the dataset is incomplete. Additionally, especially for the years before 2016, there are reports that are not suitable[7] for machine reading due to formatting and digitization limitations. The resulting data, number of balance sheets of municipalities extracted for each year is listed below in Table 5.

Table 5: The Number of Observations (Extracted Balance Sheets) per Year

2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
47	65	73	80	64	54	47	67	51	47	44

Four categories of expenditures are used as input variables: personnel expenditures, goods and services expenditures, current transfers and a residual category labeled as other expenditures. Nominal monetary values are deflated to constant prices using the Domestic Producer Price Index (D-PPI) published by TURKSTAT, base 2003=100. Annual deflators are constructed from the 12-month moving-average (December) D-PPI series and applied to convert all nominal figures into real terms. All monetary values are then normalized by the population covered by each municipality, yielding real per-capita expenditure measures. Table 6 and Table 7 shows the summary statistics of the input and output variables for two types of municipalities separately.

[7] Tables reported as pictures, in some cases even unreadable by human eye.

Table 6: Summary Statistics of Variables Used for DEA of Non-Metropolitan Municipalities

	Non-Empty Obs.	Mean	Standard Deviation	Min	Max
Outputs					
Sewage network coverage (%)	144	91.52	10.72	25.00	100.00
Sewage treatment volume	128	67.31	49.10	0.00	223.16
Waste collected	144	95.34	47.80	34.44	338.93
Water distributed	144	116.96	57.86	35.86	464.15
Water treatment volume	103	79.30	57.13	0.00	258.03
Inputs					
Current Transfers	144	12.89	13.81	0.02	92.90
Goods and Services Expenses	144	220.17	79.75	2.88	490.53
Other Expenses	144	117.60	114.98	1.15	891.37
Personnel Expenses	144	145.08	65.25	3.96	375.96

Table 6 shows that sewage network coverage variable exhibits the lowest standard deviation to mean ratio among the output variables, along with a relatively high mean value. This suggests that it is unlikely to emerge as a significant source of inefficiency in the analysis, given its consistent and widespread coverage across non-metropolitan municipalities. In terms of expenditure patterns, goods and services expenditures represent the largest per capita spending category for non-metropolitan municipalities, underlying their central role in resource allocation within this group.

Table 7: Summary Statistics of Variables Used for DEA of Metropolitan Municipalities

	Non-Empty Obs.	Mean	Standard Deviation	Min	Max
Outputs					
Sewage network coverage (%)	133	85.65	10.63	60.00	100.00
Sewage treatment volume	132	49.36	23.74	3.75	169.93
Water distributed	133	50.69	13.17	17.96	100.57
Water treatment volume	118	47.24	26.74	0.00	105.46
Inputs					
Current Transfers	133	11.28	14.63	0.40	76.62
Goods and Services Expenses	133	99.59	51.63	1.07	305.86
Other Expenses	133	92.68	68.12	3.76	399.08
Personnel Expenses	133	65.20	46.84	9.90	396.75

For the metropolitan counterparts, sewage network coverage also exhibits low variation, however with an average coverage of 85.65% of municipal population, there remains room for improvement in achieving full-service access (Table 7). The summary statistics further show that, treatment variables have higher variability compared to service variables. On the expenditure side, goods and services remain the largest spending category. Although the average expenditure levels are relatively close across categories, the within variation is notably higher in metropolitan municipalities.

3.2.2. Output Variables

Data availability on municipality service areas are highly limited for Turkish municipalities. The Turkish Statistical Institute (TÜİK) publishes core service indicators such as water supply, sewerage, waste collection on a biannual basis covering even years. These indicators are often reported in multiple forms, including total service volume (e.g., total water supplied), service coverage (e.g., percentage of the population served), and per capita metrics (e.g., water supplied per person). Coverage variables represent the percentage of the municipal population, as defined by the municipal jurisdiction, that is served by each core service.

In this study, two base models are employed. The first, referred to as the **service supply model**, includes the following output variables: volume of distributed water

(measured in cubic meters per person), sewerage network coverage (percentage of population served), and waste collected (kilograms per person). It is important to note that waste collection and management responsibilities are structured differently in metropolitan municipalities, where district municipalities are responsible for collection and metropolitan municipalities manage subsequent processing. Due to this division of labor and the resulting ambiguity in data attribution, the waste collection variable is excluded from the metropolitan models.

The second model, referred to as the **treatment services model**, focuses on wastewater and water treatment services. This model uses two output variables: sewage treatment volume and water treatment volume, both measured in m^3 per person.

Summary statistics for the output variables in metropolitan and non-metropolitan municipalities are presented in tables Table 6 and Table 7, respectively. A time series overview of these variables is provided in Table 8 and Table 9.

All service output indicators show an improving trend over time with the exception of water treatment volume, which demonstrates fluctuating average values and a high degree of variation among non-metropolitan municipalities within each year, as shown in Table 8. On the expenditure side, all categories peaked in 2016, followed by a consistent decline in spending levels through 2022.

Table 8: Yearly Statistics for Output and Input Variables – Non-Metropolitan Municipalities

Mean and Standard Deviation by Year

	2012		2014		2016		2018		2020		2022	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Outputs												
Sewage network coverage (%)	88.75	11.65	89.11	13.28	88.93	13.95	93.44	5.54	95.82	3.32	95.55	2.94
Sewage treatment volume	73.35	61.01	55.94	42.40	61.21	48.75	60.83	55.13	80.13	42.67	79.42	52.94
Waste collected	118.56	49.98	92.14	37.98	98.64	53.62	101.19	63.91	80.19	38.69	90.34	45.94
Water distributed	108.03	35.80	115.03	73.00	114.16	60.98	123.81	67.42	110.42	41.18	132.17	44.84
Water treatment volume	95.51	63.29	75.84	54.69	86.43	61.53	80.08	58.00	56.97	46.42	87.44	61.08
Inputs												
Current Transfers	16.66	22.68	11.80	9.26	18.65	20.22	10.98	7.98	11.24	6.90	7.20	4.79
Goods and Services Expenses	206.80	50.05	205.97	71.75	245.50	112.14	241.82	104.17	236.95	52.88	187.37	36.66
Other Expenses	90.02	98.77	109.01	90.61	156.83	182.71	161.75	112.63	104.77	77.36	79.75	37.03
Personnel Expenses	164.50	48.80	157.70	68.79	165.26	71.94	167.36	66.48	124.01	49.43	86.51	24.78

Table 9: Yearly Statistics for Output and Input Variables – Metropolitan Municipalities

Mean and Standard Deviation by Year

	2012		2014		2016		2018		2020		2022	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Outputs												
Sewage network coverage (%)	85.56	9.45	84.39	11.83	82.35	11.92	85.20	12.74	85.57	10.62	89.27	7.73
Sewage treatment volume	50.93	51.39	41.47	23.52	46.27	22.09	47.59	21.54	50.55	18.15	57.29	19.39
Water distributed	52.33	21.12	44.11	16.12	49.63	10.91	50.47	10.76	52.23	11.59	55.31	11.29
Water treatment volume	49.66	15.01	45.17	24.38	41.86	28.42	47.41	25.34	51.75	28.52	47.76	29.91
Inputs												
Current Transfers	10.40	13.30	8.25	9.80	10.66	12.92	11.17	13.13	10.29	13.38	16.06	22.58
Goods and Services Expenses	98.42	79.31	69.78	46.82	131.47	71.05	106.98	45.22	95.58	24.42	85.48	29.07
Other Expenses	37.66	37.58	68.76	83.47	102.00	60.90	123.98	81.19	103.40	60.63	70.63	35.26
Personnel Expenses	50.91	20.60	57.40	34.34	66.18	28.32	74.73	54.86	70.44	70.26	57.32	31.91

For the metropolitan counterparts of output variables, 2014 and 2016 represent the periods of lowest service performance, followed by a gradual upward trend extending to 2022, as illustrated in Table 9. When analyzed in conjunction with expenditure data, 2016 emerges as a particularly inefficient year, marked by elevated spending levels alongside weaker service outcomes.

3.3. Methodology

There are two main approaches used to measure technical efficiencies of local governments: parametric and non-parametric. The most common non-parametric method used is the Data Envelopment Analysis (DEA). It has different sub-versions such as (nonconvex) Free Disposal Hull, super-efficiency DEA, etc. DEA assumes that there is a set of "efficient" decision making units (DMUs – local governments in our case) that represent the best practices through DMUs. The performance of each DMU is evaluated by comparing it to the efficient frontier, which is the set of DMUs that achieve the maximum output for a given set of inputs. DEA evaluates the efficiency of each DMU by constructing a linear programming model that maximizes the weighted sum of outputs relative to the weighted sum of inputs, subject to the constraint that the DMU cannot be more efficient than any of the efficient DMUs. For the interpretation, DEA estimates the relative efficiency of each DMU, which measures how well each DMU uses its inputs to produce outputs relative to the best

practice (i.e., the efficient frontier). DEA can also identify the most efficient DMUs and the sources of inefficiency for each DMU.

Stochastic Frontier Analysis (SFA) is the parametric approach to analyze efficiency in organizations. SFA determines the frontier by using a functional form derived from econometric techniques. SFA assumes that there are two components to the observed output: a stochastic error term and a deterministic component. The error term represents the random variation in the data, while the deterministic component represents the maximum output that could be achieved given the inputs. SFA estimates a production function that relates the observed inputs to the observed outputs, while accounting for the random error term. The efficiency score is then calculated as the ratio of the observed output to the predicted output based on the estimated production function. SFA estimates the technical efficiency of each DMU, which measures how well each DMU uses its inputs to produce outputs relative to the best practice (i.e., the maximum achievable output). SFA can also decompose the efficiency score into two components: pure technical efficiency, which measures how well the DMU uses its inputs to produce outputs, and scale efficiency, which measures how well the DMU uses its inputs to produce outputs relative to its size.

For comprehensive and practical efficiency assessments, DEA models offer greater flexibility and impose fewer assumptions than the parametric models, such as SFA. One of the primary advantages of DEA is its ability to accommodate multiple inputs and multiple outputs without requiring a predefined functional form. This makes DEA particularly suitable for evaluating complex public organizations like municipalities, where diverse services are delivered using heterogeneous resources.

However, DEA is not without limitations. Its deterministic nature means that all deviations from the frontier are attributed to inefficiency, making the model highly sensitive to outliers and measurement errors. This sensitivity can affect the robustness of efficiency scores, particularly in datasets with noise or irregularities.

Additionally, a simpler method occasionally referenced is the ratio index approach, which involves calculating efficiency as the ratio of total outputs to total inputs (Borge et al., 2008). While this method offers ease of implementation and interpretability, it

lacks the ability to account for multiple inputs and outputs or provide relative efficiency scores.

This study adopts DEA as the primary method for evaluating municipal efficiency. First, as public service providers, municipalities do not produce a single, quantifiable output as is often the case for private firms. DEA offers the flexibility to incorporate multiple output variables, enabling the analysis of technical efficiency from various service dimensions. Second, the multi-service, socially oriented and politically influenced nature of municipal production scheme makes it unsuitable to be captured by a specific functional form which is necessary to impose Stochastic Frontier Analysis. Finally, the large number of DMUs included in this study presents additional challenges for SFA, particularly in terms of model complexity and estimation feasibility, further supporting a non-parametric approach.

We evaluate technical efficiency in separate, cross-sectional DEA problems for each calendar year in 2012–2022. Within a given year, let there be n province-level municipalities (DMUs), indexed by $\mathrm{j} = 1, \dots, \mathrm{n}$. For each DMU j, let $\mathrm{x_{kj}}$ denote the quantity of input k for $\mathrm{k} = 1, \dots, \mathrm{K}$, and $\mathrm{y_{rj}}$ denote the quantity of service output r for $\mathrm{r} = 1, \dots, \mathrm{R}$. Because the model is solved year by year, the year index is suppressed in all expressions below. The services specification uses the service-output variables defined in the data subsection; the treatment services specification is obtained by replacing the output vector y with the treatment outputs (same program forms).

CCR/CRS, output-oriented (envelopment form):

For a focal municipality i, the constant-returns, output-oriented problem is (Charnes et al., 1978; Farrell, 1957):

$$\max \varphi$$

subject to

$$\sum_{\mathrm{j}=1}^{\mathrm{n}} \lambda_{\mathrm{j}}\, \mathrm{y_{rj}} \;\geq\; \varphi\, \mathrm{y_{ri}} \quad (\mathrm{r} = 1, \dots, \mathrm{R}),$$

$$\sum_{j=1}^{n} \lambda_j \, x_{kj} \leq x_{ki} \quad (k = 1, \ldots, K), \tag{1}$$

$$\lambda_j \geq 0 \quad (j = 1, \ldots, n).$$

The expansion factor $\varphi \geq 1$ yields output-oriented technical efficiency $TE_i = 1/\varphi \in (0,1]$.

The vector λ selects and weights observed municipalities to form a reference (benchmark) point on the efficient frontier for the focal municipality. Under VRS, the convexity constraint $\sum_j \lambda_j = 1$ ensures the benchmark is a convex combination of observed DMUs. Any $\lambda_j > 0$ identifies municipality *j* as a peer; its magnitude is the share that municipality contributes to the benchmark mix. Geometrically, the DEA projection moves the focal bundle (x_0, y_0) radially inward by φ (and then along slacks which will be discussed) to a point on the convex hull of feasible production points; λ gives the coordinates of that hull point. Because multiple convex combinations can lie on the same supporting facet, λ need not be unique.

In practice, λ serves three purposes: (i) peer identification (which municipalities set the standard), (ii) construction of target inputs/outputs via the weighted sums above (consistent with the φ and slack targets), and (iii) returns-to-scale diagnosis through the chosen convexity restriction (CRS vs. VRS vs. NIRS/NDRS).

BCC/VRS, output-oriented (envelopment form)

Variable returns are imposed by the convexity constraint (Banker et al., 1984; Farrell, 1957):

$$\max \varphi$$

subject to

$$\sum_{j=1}^{n} \lambda_j \, y_{rj} \geq \varphi \, y_{ri} \quad (r = 1, \ldots, R),$$

$$\sum_{j=1}^{n} \lambda_j \, x_{kj} \leq x_{ki} \quad (k = 1, \dots, K), \tag{2}$$

$$\sum_{\mathrm{j}=1}^{\mathrm{n}} \lambda_{\mathrm{j}} = 1, \quad \lambda_{\mathrm{j}} \geq 0 \, (\mathrm{j} = 1, \dots, \mathrm{n}).$$

This program delivers pure technical efficiency under VRS; scale efficiency can be inferred by comparing CCR and BCC scores.

Slack-refined variants

To rule out residual input excesses or output shortfalls after the radial step, we use the standard ε refinement. Let $s_k^- \geq 0$ be input slacks and $s_r^+ \geq 0$ be output slacks; pick a small $\varepsilon > 0 (\mathrm{e.g.}, 10^{-6})$ and we get CCR/CRS with slacks (Cooper et al., 2007):

$$\max \varphi - \varepsilon \left(\sum_{k=1}^{K} s_k^- + \sum_{r=1}^{R} s_r^+ \right) \tag{3}$$

subject to

$$\sum_{\mathrm{j}=1}^{\mathrm{n}} \lambda_{\mathrm{j}} \, \mathrm{x_{kj}} + \mathrm{s_k^-} = \mathrm{x_{ki}} \quad (\mathrm{k} = 1, \dots, \mathrm{K}),$$

$$\sum_{j=1}^{n} \lambda_j y_{rj} - s_r^+ = \theta \, y_{r0}, \quad (r = 1, \dots, R), \tag{4}$$

$$\lambda_{\mathrm{j}} \geq 0, \; \mathrm{s_k^-} \geq 0, \; \mathrm{s_r^+} \geq 0.$$

And BCC/VRS with slacks:

$$\max \varphi - \varepsilon \left(\sum_{k=1}^{K} s_k^- + \sum_{r=1}^{R} s_r^+ \right) \tag{5}$$

subject to the two equalities above and

$$\sum_{j=1}^{n} \lambda_j \; = 1, \quad \lambda_j \geq 0, s_k^- \geq 0, s_r^+ \geq 0 \tag{6}$$

The interpretation outline is as follows:

- Orientation: Output-oriented in the Farrell (1957) sense: φ scales the focal DMU's outputs while holding inputs fixed.
- Returns to scale: CCR assumes CRS; BCC imposes VRS via $\sum_j \lambda_j = 1$. NIRS adoption is via: $\sum_j \lambda_j \leq 1$.
- Efficiencies: Report ($TE_i = 1/\varphi^*$). Comparing CCR vs. BCC identifies scale effects.
- Model scope. Programs are solved separately for each year; all peer weights \lambda reference municipalities in the same year. The treatment services model is obtained by substituting the treatment outputs for y; inputs remain as in the services specification.

By introducing the slack ($s_k^- \geq 0$ and $s_r^+ \geq 0$) variables, equation (3) clarifies two aspects of DEA. Slacks capture the distances from the efficient frontier, quantifying the input excesses and output shortfalls of each DMU relative to the efficient frontier. This enables both an envelopment of the production possibility set and the identification of specific sources of inefficiencies. After estimating the DEA model on municipal data, these inefficiency measures can inform targeted policy recommendations for individual municipalities.

In our analysis, input variables are selected from expenditure titles of municipalities. In theory, municipal efficiency involves two decision layers: resource generation and resource allocation. In a local governance framework where tax revenue forms one of the main sources of revenue for a municipality, we can claim a meaningful involvement of the municipality administration on both resource generating and allocation. However, as previously discussed, Turkish municipalities derive approximately 70% of their funding from central government transfers (2022 figure, consistent with past averages). Given this structural characteristic, applying an input-oriented DEA model, which assumes discretion over input levels, would misrepresent the decision space of Turkish municipalities. Therefore, an output-oriented model is

more appropriate, as it focuses on assessing the capacity of municipalities to maximize service provision with largely predetermined input levels.

As outlined previously, Turkish municipalities are classified into two main subgroups, metropolitan and non-metropolitan municipalities. For the purposes of this study, these two categories differ significantly in two key aspects: the scope of their responsibilities and the scale of their operations in terms of population and budget size. The divergence in service mandates and resource capacities results in substantial structural differences between the subgroups. These imbalances necessitate a disaggregated approach in the DEA analysis to ensure methodological soundness and to allow for meaningful efficiency comparisons within each group.

Furthermore, disaggregation is also methodologically justified by a fundamental data requirement of DEA models: the number of decision-making units (DMUs) must be at least three times the maximum of the number of input or output variables, formally expressed as:

$$n_{\{DMUs\}} \geq 3 \times max\left(n_{\{inputs\}}, n_{\{outputs\}}\right) \quad (7)$$

This condition helps to prevent overfitting and ensures that the efficiency frontier is constructed on a statistically meaningful basis (Banker et al., 1984).

Another important variant within DEA modeling concerns the assumption regarding returns to scale. In the context of local governments, factors such as budgetary limitations, geographical constraints, and operational complexities often restrict the ability to proportionally scale inputs and outputs. These real-world limitations justify the use of non-increasing returns to scale (NIRS) in the efficiency models applied in this study. The appropriateness of this specification is further supported by the statistical test developed by Simar & Wilson (2002), which provides empirical validation for selecting the NIRS assumption over constant returns to scale.

3.4. Results

Based on the selection of the output variables, two primary DEA models are constructed, service supply and treatment services models. To enhance the reliability

of the DEA results, each model was applied separately to metropolitan and non-metropolitan municipalities. This disaggregation is in line with best practices in DEA, which emphasize that greater homogeneity among DMUs improves the robustness of efficiency comparisons. The structure of the model output sets is summarized in Table 10, which presents the matrix of output variable and municipality type combinations.

Table 10: Output Scenarios of DEA Models Matrix
(M: metropolitan, N-M: non-metropolitan)

	Service M	Service N-M	Treatment M	Treatment N-M
Water Supply	✓	✓	-	-
Waste Collected	-	✓	-	-
Sewage Network Coverage	✓	✓	-	-
Water Treatment	-	-	✓	✓
Sewage Treatment Volume	-	-	✓	✓

3.4.1. Service Supply Models

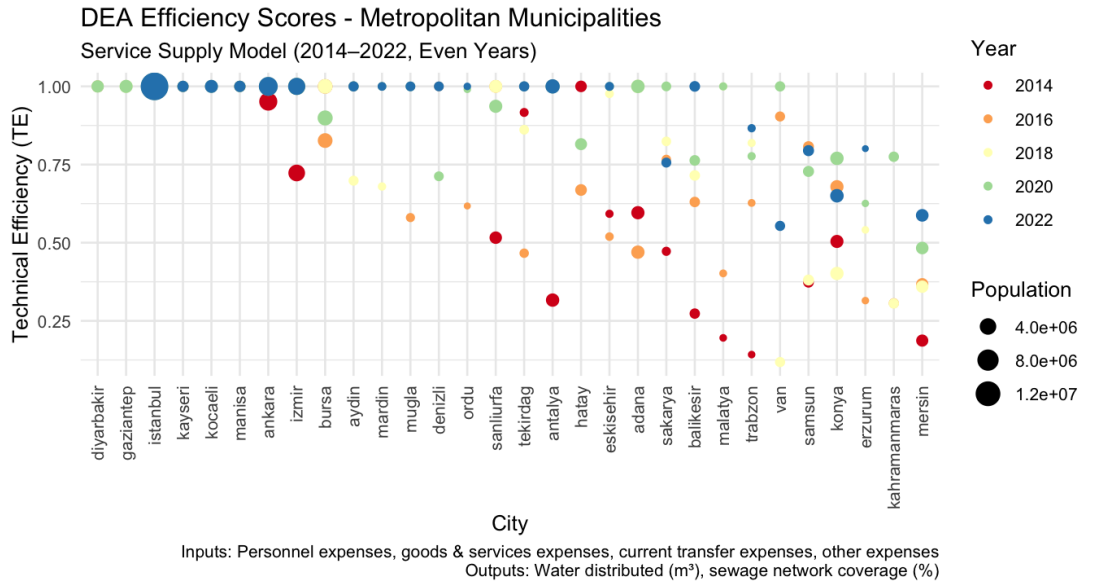
All results of the service supply model along with yearly and city averages are presented in appendix (Table A1 and Table A2). It is important to note that, DEA evaluates relative efficiencies and produces scores ranging from 0 to 1. An efficiency score of 1.00 does not indicate a perfect or absolute performance; rather, it indicates that the corresponding DMU lies on or close to the estimated efficient frontier. The results for the year 2012 are omitted due to the insufficient number of observations that do not meet DEA requirements. The missing observations in the tables are mostly not audited for that year or have non-readable balance sheet reports.

For an overall picture of results, Figure 4 presents the annual efficiency scores alongside population sizes for metropolitan municipalities, ordered by decreasing average efficiency. The yearly average results suggest a gradual improvement in efficiency over time as observed in Figure 4 (the bottom row of metropolitan table of

Table A1 also supports this observation). As the largest city of Türkiye, accounting nearly 20% of the national population, İstanbul appears as a notable outlier in the service supply model, likely due to its scale. Municipalities such as Adana, Malatya and Trabzon exhibit greater variability in efficiency scores across years, while Ankara, İzmir and Bursa show more stable and consistent results.

The municipalities with higher and stable results are benchmark examples, studying their governance structures and policy practices may yield valuable insights for municipalities with persistently low scores. Conversely, the underperforming municipalities have to be deeper analyzed for the root causes of inefficiency. To this end, a more detailed discussion of slack variables is provided in the following sections to support diagnostic interpretation.

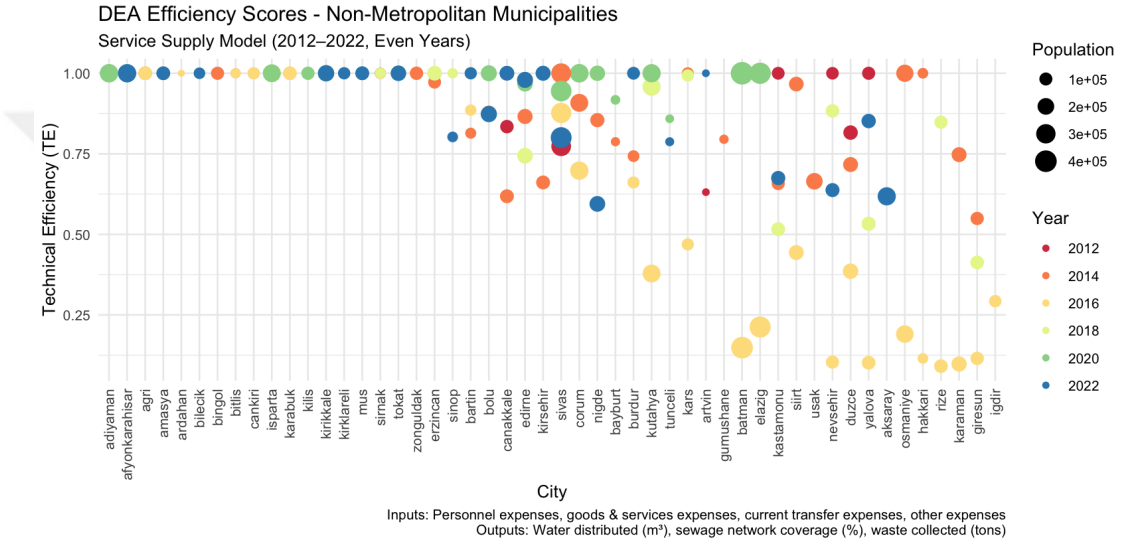
Figure 4: Yearly Results for Service Supply Model, Metropolitan Municipalities



For the non-metropolitan pool of the service supply model, 2012 data have sufficient number of observations. Notably, 36% of the non-metropolitan municipalities in the sample achieve an efficiency score of 1.00, positioning them as benchmark units. Moreover, the overall average efficiency in the non-metropolitan group exceeds that of their metropolitan counterparts. This disparity can be attributed to the greater structural complexity, significantly more service areas and higher volume of service transactions in metropolitan municipalities, which tend to increase operational challenges and contribute to relatively lower efficiency levels.

It is important to emphasize that DEA models were estimated separately for each year. The relatively weak performance of non-metropolitan municipalities in 2016 is evident from the average efficiency scores presented in the bottom row of Table A2 and visualized in Figure 5. Municipalities with consistently low or volatile efficiency scores -such as Karaman, Rize, Giresun- should conduct a detailed analysis of their slack values to identify the sources of inefficiency and establish targeted areas for improvement.

Figure 5: Yearly Results for Service Supply Model, Non-Metropolitan Municipalities



The results discussed above have identified the inefficient DMU's. By utilizing the slack values computed within the DEA framework, we are able to address the specific sources of inefficiency for each of the DMU. Slacks values represent the distance of a specific var of a DMU from the efficient border, thereby allowing for the identification of the input or output components that contribute most significantly to inefficiency.

Table 11 reports the slack values of metropolitan municipalities in the service supply model, focusing on the three lowest performing DMUs in each year. For a cross-check, the table also includes the within group deviations of each variable from its annual mean.

To illustrate, consider 2014 results for Malatya Metropolitan Municipality. The slack analysis reveals that the primary sources of inefficiency are goods and services expenditures and sewage network coverage. The deviation values also support the findings of the DEA with an 37% more than group average of goods and services

expenditures and a 12% lower sewage network coverage from the average coverage value. In simplified terms, this municipality spend higher than average but delivered services lower than average. These results suggest the municipality should review the efficiency and allocation of its goods and services expenditures and prioritize expanding its sewage infrastructure, rather than focusing on water supply, which appears less relevant to its inefficiency profile.

Slack values are calculated for all DMUs, enabling the aggregation of annual average slack scores. These averages provide valuable insights into overall trends in inefficiency, as summarized in Table 12. Moreover, when contrasted with the results presented in Table 13, they allow for a comparative analysis between metropolitan and non-metropolitan municipalities.

Table 11: Slack and Deviation Table for Metropolitan Municipalities, Service Model

Slack & Deviation Matrix

Bottom 3 Inefficient Municipalities Per Year

		Personnel Expenses		Goods and Services Expenses		Current Transfers		Other Expenses		Water distributed		Sewage network coverage (%)	
		% Dev	Slack	% Dev	Slack	% Dev	Slack	% Dev	Slack	% Dev	Slack	% Dev	Slack
2014	malatya	5%	0.0	37%	17.8	−15%	1.1	−74%	0.4	−42%	0.0	−12%	14.5
	mersin	46%	1.7	−47%	0.0	−49%	0.4	−72%	0.1	−25%	0.0	−4%	5.7
	trabzon	76%	2.3	−37%	0.0	−44%	0.2	−37%	3.9	−38%	0.0	−6%	9.1
2016	erzurum	69%	0.0	186%	28.5	−18%	0.0	141%	28.6	7%	0.0	−6%	8.7
	kahramanmaras	86%	0.0	23%	0.0	−60%	0.3	57%	9.8	12%	0.0	−24%	24.7
	mersin	43%	0.0	59%	0.0	−26%	0.5	23%	2.8	−9%	3.1	−4%	0.0
2018	kahramanmaras	8%	0.0	11%	0.7	−30%	0.3	51%	22.5	12%	0.0	−20%	0.0
	mersin	42%	6.4	77%	7.6	−46%	0.0	64%	0.0	−3%	0.0	−12%	0.0
	van	1%	0.0	38%	12.1	220%	1.8	51%	3.5	−32%	12.1	−20%	0.0
2020	denizli	513%	233.6	29%	4.2	−56%	0.0	−5%	0.0	12%	0.0	6%	0.0
	erzurum	35%	13.0	23%	0.0	−64%	0.0	233%	154.4	20%	7.1	−6%	0.0
	mersin	−6%	0.0	40%	0.0	−30%	0.0	117%	54.3	2%	0.0	−11%	4.7
2022	konya	9%	0.0	41%	0.0	41%	0.0	63%	29.7	26%	0.0	7%	0.0
	mersin	−32%	0.0	22%	11.8	−61%	0.0	6%	32.3	3%	0.0	−6%	5.0
	van	−18%	0.0	3%	0.0	−58%	0.0	−31%	3.3	−20%	13.1	0%	0.0

The findings indicate that water service variables are not major contributors to inefficiency in metropolitan municipalities, suggesting a relatively saturated development in this area. In contrast, non-metropolitan municipalities continue to exhibit potential for improvement in water service and waste collection provision (Table 14). Sewage network coverage, on the other hand, appears to be well-established across both municipal types, showing relatively low slack values.

In terms of expenditure related inefficiencies, goods and services expenditures and other expenditures emerge as the most significant contributors. Notably, the pattern of inefficiency related to personnel expenses shows asymmetrical fluctuations over time between the two municipal subgroups, indicating differing dynamics in human resource allocation and efficiency management.

Table 12: Yearly Slack Analysis for Metropolitan Service Supply Model

Average Slack Values by Year

Mean inefficiencies per input/output variable

Year	Personnel Expenses	Goods and Services Expenses	Current Transfers	Other Expenses	Water distributed	Sewage network coverage (%)
2014	6.5	10.1	2.1	18.3	0.6	5.4
2016	1.3	20.0	1.6	9.2	4.2	4.2
2018	3.1	8.1	0.2	11.3	1.4	0.0
2020	11.7	2.9	0.3	14.9	1.4	1.6
2022	0.0	1.9	0.5	6.2	1.1	0.2

Table 13: Yearly Slack Analysis for Non-Metropolitan Service Supply Model

Average Slack Values by Year

Mean inefficiencies per input/output variable

Year	Personnel Expenses	Goods and Services Expenses	Current Transfers	Other Expenses	Water distributed	Sewage network coverage (%)	Waste collected
2012	3.1	3.5	2.5	2.0	7.0	0.5	15.0
2014	7.2	10.0	2.1	13.3	14.8	0.5	2.5
2016	9.7	23.1	1.7	13.8	19.4	5.4	10.3
2018	0.0	34.1	1.2	30.5	7.0	2.7	16.8
2020	6.3	2.7	0.5	10.6	2.4	0.1	2.6
2022	6.8	3.3	1.5	9.1	20.1	0.2	4.5

Table 14 presents the slack values for non-metropolitan municipalities under the services model, highlighting the three lowest-performing municipalities for each year.

The table allows for the identification of specific sources of inefficiency in these poorly performing cases. In contrast to the metropolitan results in Table 11, output slacks are considerably higher among non-metropolitan municipalities. This suggests that basic service provision is more adequately achieved by metropolitan municipalities, whereas non-metropolitan counterparts exhibit greater shortfalls. In other words, there remains significantly more room for improvement in the service delivery performance of non-metropolitan municipalities.

Table 14: Slack and Deviation Table for Non-Metropolitan Municipalities, Service Model

Slack & Deviation Matrix

Bottom 3 Inefficient Municipalities Per Year

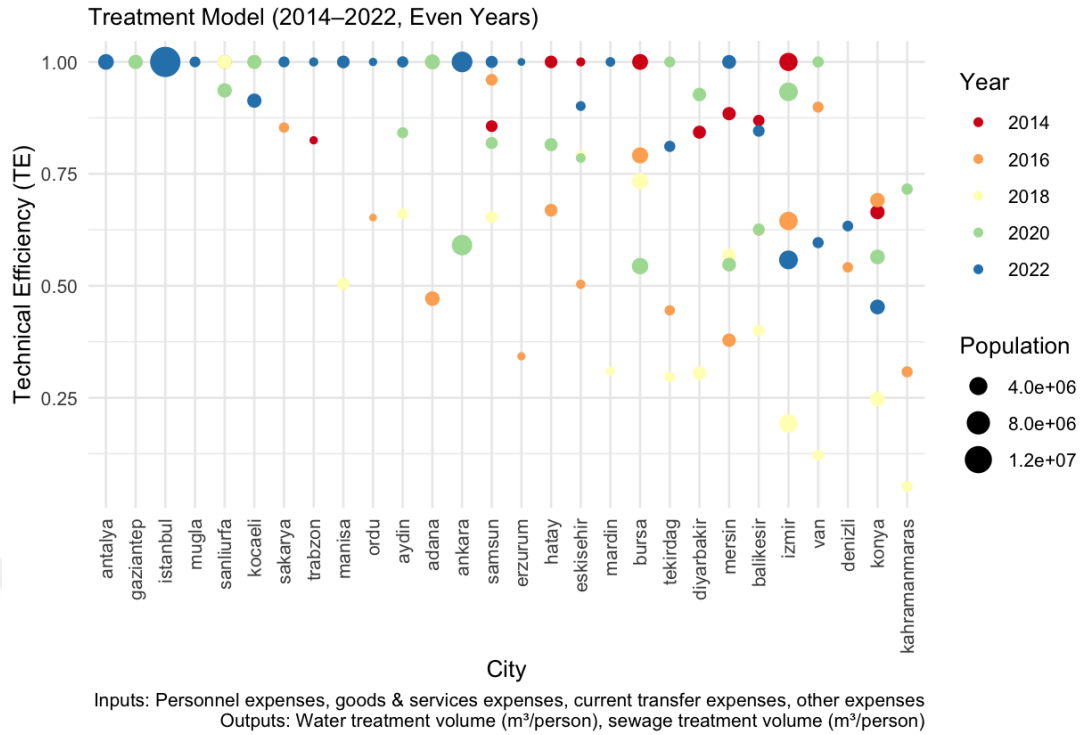
		Personnel Expenses		Goods and Services Expenses		Current Transfers		Other Expenses		Water distributed		Sewage network coverage (%)		Waste collected	
		% Dev	Slack	% Dev	Slack	% Dev	Slack	% Dev	Slack	% Dev	Slack	% Dev	Slack	% Dev	Slack
2012	artvin	60%	7.4	5%	0.0	−26%	1.5	−9%	7.1	5%	0.0	−5%	6.8	31%	60.7
	duzce	−24%	0.0	4%	0.0	14%	2.3	−23%	0.0	−46%	60.4	−1%	0.0	0%	41.4
	sivas	23%	0.0	23%	56.4	−46%	3.1	12%	0.0	−33%	51.5	5%	0.0	−31%	103.1
2014	canakkale	100%	0.0	0%	0.0	34%	0.0	177%	61.5	11%	36.6	3%	0.0	53%	0.0
	giresun	44%	0.0	6%	0.0	246%	18.7	−2%	0.0	−11%	25.2	−13%	0.0	12%	0.0
	kastamonu	40%	0.0	−6%	0.0	5%	0.0	50%	26.1	−12%	38.3	1%	0.0	31%	0.0
2016	karaman	9%	11.5	−16%	15.2	79%	0.0	15%	5.4	−59%	65.8	−3%	7.0	−47%	41.0
	rize	−18%	6.8	94%	36.1	92%	0.0	387%	44.3	−35%	37.4	−11%	15.0	−11%	6.8
	yalova	66%	0.0	73%	13.3	488%	4.8	252%	29.0	13%	0.0	4%	0.0	−16%	14.1
2018	giresun	46%	0.0	71%	11.3	46%	0.1	42%	0.0	−8%	11.2	−9%	16.0	10%	5.7
	kastamonu	69%	0.0	14%	0.0	42%	2.5	55%	0.0	−12%	14.7	3%	4.1	−20%	31.7
	yalova	32%	0.0	39%	22.8	119%	7.4	17%	0.0	13%	0.0	5%	2.2	−16%	43.0
2020	bayburt	22%	62.3	−21%	0.0	39%	7.2	−25%	0.0	−35%	18.1	5%	0.0	−49%	14.9
	sivas	−45%	0.0	22%	59.1	−16%	0.0	32%	43.6	−5%	0.0	6%	0.0	−41%	16.8
	tunceli	47%	75.8	1%	0.0	−42%	4.2	−8%	0.0	25%	0.0	2%	2.0	−25%	26.5
2022	aksaray	−32%	22.1	−29%	0.0	−55%	0.6	−57%	4.5	−38%	106.4	−1%	1.0	−59%	30.3
	nevsehir	−25%	3.0	−16%	0.0	−71%	0.3	−20%	7.3	25%	26.3	3%	0.0	−26%	14.0
	nigde	−52%	2.6	−26%	0.0	−57%	0.3	−31%	21.7	−16%	80.5	−3%	3.0	−36%	9.0

3.4.2. Treatment Services Models

The treatment services model utilizes two variables from TÜİK: water treatment volume and sewage treatment volume. Similar to the service supply model, metropolitan observations for year 2012 were insufficient to meet DEA requirements, thus no model estimated for that year. Figure 6 present the efficiency results of treatment services models for metropolitan municipalities (for all the results: Table A3). The cities of İstanbul, Muğla, and Gaziantep consistently appear as benchmark units, maintaining efficiency scores of 1.00 across all annual models. However, a notable decline in overall efficiency is observed in 2018, warranting further investigation into potential structural or operational disruptions during that period.

A comparison between İzmir and İstanbul reveals divergent efficiency trajectories among the most populous municipalities, indicating that scale alone does not determine treatment services performance. Furthermore, Denizli, Konya, Kahramanmaraş and Van exhibit substantial inefficiency gaps, highlighting the need for targeted improvements in treatment service infrastructure and management.

Figure 6: Efficiency Scores of Treatment Model for Metropolitan Municipalities



The slack values reported in Table 16 offer valuable insights into underlying sources of inefficiency in the treatment services model. For instance, 2018 data for İzmir indicate that the municipality performs relatively well in terms of service provision, as reflected by high treatment volumes. However, its overall efficiency score is weakened due to cost efficiency, suggesting a disproportionate level of expenditure relative to outputs. In contrast, municipalities such as Kahramanmaraş (2016), Bursa (2020), and Van (2018) exhibit low efficiency scores primarily due to insufficient service provision, rather than excessive spending. These cases illustrate the importance of distinguishing between input-driven and output-driven inefficiencies when interpreting DEA results.

The average slack values presented in Table 15 offer a broader perspective on efficiency gaps across all DMUs. In terms of input variables, goods and services expenditures and other expenditures emerge as the primary contributors to inefficiency, indicating potential misallocation or excess spending in these categories. On the output side, water treatment volume consistently appears as the most significant source of inefficiency, suggesting that many municipalities underperform in delivering this service relative to their resource levels. These findings highlight key areas for

targeted efficiency improvements in both expenditure management and service provision.

Table 15: Yearly Slack Analysis for Metropolitan Cities, Treatment Model

Average Slack Values by Year

Mean inefficiencies per input/output variable

Year	Personnel Expenses	Goods and Services Expenses	Current Transfers	Other Expenses	Water treatment volume	Sewage treatment volume
2014	5.7	4.6	0.1	4.9	0.3	2.1
2016	1.5	30.0	3.5	17.4	17.5	5.6
2018	2.0	7.1	0.2	15.4	7.4	0.1
2020	7.1	1.2	1.9	18.3	10.8	3.4
2022	0.1	5.0	2.6	8.3	3.0	2.8

Table 16: Slack and Deviation Table for Metropolitan Municipalities, Treatment Model

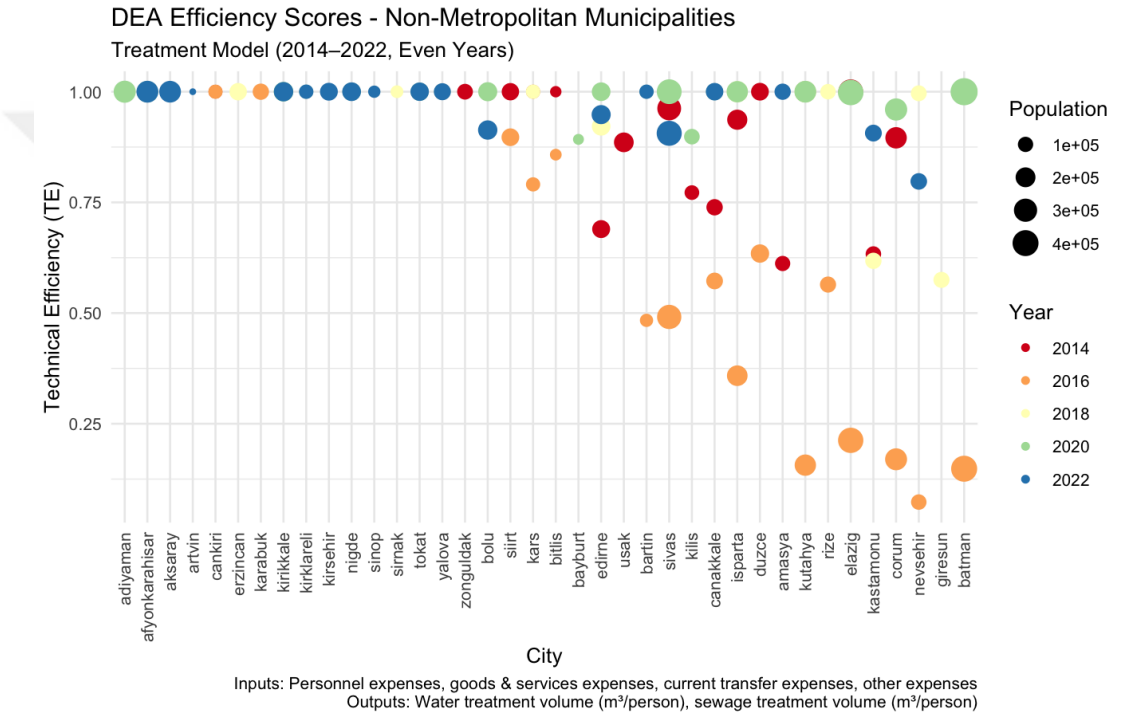
Slack & Deviation Matrix

Bottom 3 Inefficient Municipalities Per Year

		Personnel Expenses		Goods and Services Expenses		Current Transfers		Other Expenses		Water treatment volume		Sewage treatment volume	
		% Dev	Slack	% Dev	Slack	% Dev	Slack	% Dev	Slack	% Dev	Slack	% Dev	Slack
2014	diyarbakir	−37%	0.0	−44%	3.9	−67%	0.0	−74%	0.0	−14%	0.0	−51%	0.9
	konya	−33%	0.0	18%	36.2	−38%	0.0	44%	69.3	−86%	5.5	−33%	0.0
	trabzon	76%	49.6	−37%	0.0	−44%	0.0	−37%	0.0	41%	0.0	−63%	21.5
2016	erzurum	69%	0.0	186%	56.1	−18%	2.2	141%	42.7	62%	0.0	−30%	30.8
	kahramanmaras	86%	0.0	23%	0.0	−60%	0.8	57%	13.8	−98%	73.3	−87%	51.4
	mersin	43%	0.0	59%	0.0	−26%	0.3	23%	1.5	27%	0.0	−5%	1.6
2018	izmir	180%	13.0	10%	0.0	368%	3.8	86%	0.0	20%	0.3	29%	0.0
	kahramanmaras	8%	0.0	11%	0.3	−30%	0.0	51%	3.7	−62%	23.6	−76%	0.0
	van	1%	0.0	38%	12.4	220%	1.9	51%	3.6	−72%	32.8	−43%	0.0
2020	bursa	18%	13.0	13%	0.0	21%	3.1	47%	38.5	−29%	45.1	3%	0.0
	konya	−34%	0.0	4%	1.6	66%	5.6	96%	64.2	8%	40.4	−17%	4.7
	mersin	−6%	0.0	40%	2.7	−30%	0.0	117%	69.5	33%	9.4	19%	0.0
2022	izmir	45%	2.4	−5%	0.0	547%	30.2	21%	0.0	21%	0.0	36%	0.0
	konya	9%	0.0	41%	5.2	41%	0.0	63%	42.2	9%	0.0	−16%	26.4
	van	−18%	0.0	3%	9.9	−58%	0.0	−31%	14.1	−74%	27.2	36%	0.0

In the treatment services model for the non-metropolitan subgroup, 2012 data has insufficient number of observations to meet DEA requirements. As shown in Table A4 and Figure 7, the year 2016 exhibits a noticeably lower average efficiency score compared to other years, which remain relatively stable around 0.95. Among the non-metropolitan municipalities, Afyonkarahisar, Sinop and Yalova demonstrate consistently high efficiency levels across the evaluated period, positioning them as benchmark performers within this subgroup.

Figure 7: Efficiency Scores of Treatment Model, Non-Metropolitan Subgroup



Bayburt (2020) with a slack value 54.5 for personnel expenditure combined with a below average treatment values, indicate substantial inefficiency in personnel management, potentially addressing overstaffing. In contrast, Amasya (2014) exhibits a relatively balanced expenditure profile, yet underperforms in converting these expenditures into treatment service outputs, suggesting inefficiencies in operational effectiveness (Table 17).

Yearly slack analysis presented in Table 18 further highlights 2016 as a notable outlier, with pronounced inefficiencies across multiple variables. If we omit the 2016 as an outlier, the remaining period reveals more balanced slack values and indicates a general improvement trend or at least a convergence among municipalities.

Table 17: Slack and Deviation Table for Non-Metropolitan Municipalities, Treatment Model

Slack & Deviation Matrix

Bottom 3 Inefficient Municipalities Per Year

		Personnel Expenses		Goods and Services Expenses		Current Transfers		Other Expenses		Water treatment volume		Sewage treatment volume	
		% Dev	Slack	% Dev	Slack	% Dev	Slack	% Dev	Slack	% Dev	Slack	% Dev	Slack
2014	amasya	31%	0.0	7%	0.0	−30%	0.0	−9%	1.7	−98%	90.6	−100%	29.9
	edirne	−18%	0.0	53%	38.1	22%	2.1	19%	0.0	23%	0.0	−96%	29.6
	kastamonu	40%	0.0	−6%	0.0	5%	5.1	50%	45.3	−34%	52.8	−84%	8.4
2016	batman	−4%	16.6	−12%	25.8	18%	0.0	131%	32.4	−86%	36.1	−41%	61.3
	kutahya	−8%	17.0	−5%	29.9	12%	0.0	22%	14.6	−64%	18.4	44%	4.4
	nevsehir	70%	0.0	19%	0.0	240%	0.0	658%	54.5	−31%	0.0	22%	13.5
2018	edirne	−19%	0.0	37%	0.0	16%	6.2	62%	36.1	44%	0.0	−64%	0.0
	giresun	46%	0.0	71%	0.0	46%	5.2	42%	0.0	−98%	75.4	−26%	18.2
	kastamonu	69%	0.0	14%	0.0	42%	4.9	55%	0.0	−1%	21.7	−67%	0.0
2020	bayburt	22%	54.5	−21%	0.0	39%	7.8	−25%	8.9	−96%	9.5	−10%	0.0
	corum	−45%	0.0	8%	0.0	3%	0.0	100%	94.5	−23%	0.0	38%	0.0
	kilis	−31%	0.0	82%	102.5	−11%	0.0	−10%	0.0	15%	0.0	−46%	45.1
2022	kastamonu	−29%	0.0	−21%	0.0	−8%	6.9	−38%	0.0	−10%	0.0	38%	0.0
	nevsehir	−25%	0.0	−16%	0.0	−71%	0.4	−20%	0.0	−76%	12.9	−1%	18.8
	sivas	−50%	0.0	2%	0.0	−34%	2.3	−6%	14.7	−2%	0.0	15%	0.0

Table 18: Yearly Slack Analysis for Non-Metropolitan Municipalities, Treatment Model

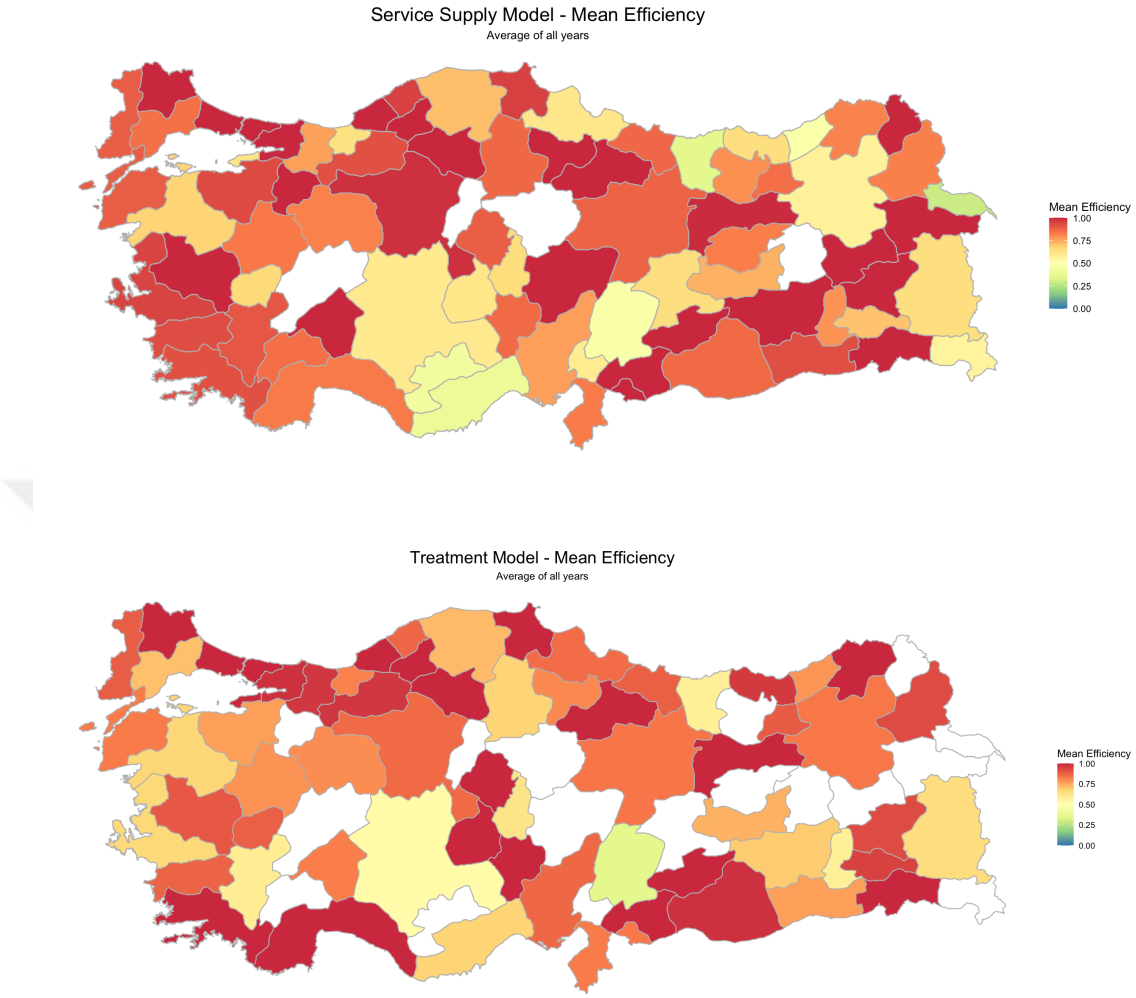
Average Slack Values by Year

Mean inefficiencies per input/output variable

Year	Personnel Expenses	Goods and Services Expenses	Current Transfers	Other Expenses	Water treatment volume	Sewage treatment volume
2014	2.1	5.6	2.6	25.3	21.8	4.3
2016	21.2	48.4	2.7	33.1	8.2	20.1
2018	0.0	0.0	1.2	2.6	8.3	1.3
2020	3.0	9.4	0.4	5.7	0.5	2.8
2022	0.0	1.9	0.7	6.1	0.7	2.0

Figure 8 presents a visual representation of the mean efficiency scores of municipalities over the period 2012-2022, mapped geographically across Türkiye. While Turkish municipalities exhibit considerable heterogeneity in terms of economic activity, population, geography, climate, demography, the spatial distribution of efficiency scores does not reveal a strong pattern of regional dependence. This observation holds for both the service supply and treatment services models, suggesting that efficiency performance is influenced more by local administrative and operational factors than by broad regional characteristics.

Figure 8: Service Supply and Treatment Services Models, Mean Efficiency Scores 2012-2022 Illustrated on Map



3.5. Discussion

The findings of this study both align with and extend the existing literature on municipal efficiency in Türkiye, offering important methodological and empirical advancements. Earlier studies, such as Kutlar et al. (2012) and Çelikkaya (2016), have documented persistent inefficiencies among Turkish municipalities, with a particular emphasis on the gap between spending and service outcomes. Our findings confirm this pattern, particularly through the use of slack analysis, which reveals specific spending categories (e.g., goods and services) that frequently exceed group norms without corresponding service output. This adds a diagnostic layer absent from prior work, which typically reported only composite efficiency scores.

Studies like Kaplan et al. (2006) and Güneş & Akdoğan (2007) suggested that metropolitan municipalities may benefit from economies of scale, though efficiency varied by service type. Our findings provide more nuanced evidence: while some large municipalities such as Istanbul maintain consistently high efficiency, others such as İzmir and Ankara display more volatile performance, particularly in treatment-related services. This supports the conclusion that scale alone does not guarantee efficiency, a claim previously noted but now reinforced by multi-year, multi-model evidence.

İlkay & Doğan (2009) emphasized the importance of modeling efficiency by service domain, a principle that this study adopts through its dual-model structure (Service Supply and Treatment Services). By disaggregating models and treating metropolitan and non-metropolitan municipalities separately, this study provides a more nuanced understanding of performance variation, especially in light of differences in functional responsibilities, resource levels, and data availability.

Moreover, unlike previous studies that largely relied on short time frames (usually 1–3 years) and static datasets, this research utilizes a longitudinal dataset spanning 12 years and obtained from computational text analysis based on Turkish Court of Accounts audit reports. This new data extracting approach overcomes longstanding data limitations in the literature and establishes a scalable, replicable framework for future research.

By transforming technical accounting and financial data into accessible and interpretable efficiency metrics, this methodology contributes meaningfully to transparency and accountability in local governance. As shown in prior literature (e.g. Narbón-Perpiñá & De Witte, 2018), the use of efficiency indicators derived from non-parametric methods such as DEA can be instrumental in evaluating and comparing municipal performance. In well-functioning democracies, such transparency may even support more informed and conscious electoral behavior, empowering citizens to evaluate local administrations beyond political narratives.

The service categories used in the models, such as water distribution, sewage coverage, and treatment volume, remain foundational for small and mid-sized municipalities, and continue to align with standard efficiency measurement frameworks used internationally (e.g., Afonso & Fernandes, 2008; Prieto & Zofio, 2001). However, in

metropolitan municipalities, the role of these basic services is increasingly supplanted by broader social service demands. These include urban transportation, poverty alleviation programs, childcare and elder care services, public health, and recreational infrastructure, which now constitute a growing share of local government responsibilities. Consequently, future analyses would benefit from incorporating a wider array of service indicators to capture the full complexity of urban governance, as also suggested in the extended DEA applications of Woodbury & Dollery (2004) and De Sousa et al. (2005).

Moreover, the rising fiscal importance of local governments relative to central administrations, as observed over the last two decades, underscores the need for greater data transparency and performance monitoring. This trend is mirrored globally in developed and developing economies alike, where municipal-level decision-making increasingly influences public service quality and citizen well-being. With access to more and diversified data on municipal services, the DEA framework applied in this study could be easily expanded to generate more reliable and actionable results.

Importantly, this method provides an antidote to information bias and distortion, which are common in politically polarized environments. Publicly available data, including satisfaction surveys, are often subject to interpretive manipulation, whether intentional or unintentional, by competing political actors. In contrast, the method proposed in this research relies on objectively measurable service outputs and audited financial data, producing results through transparent and replicable algorithms. As such, the framework offers a rare avenue for evidence-based evaluation, independent of political alignment or subjective perception.

4. IRREGULARITY INDICES FROM TURKISH COURT OF ACCOUNTS REPORTS

The purpose of this chapter is to convert unstructured Turkish Court of Accounts (Sayıştay) audit narratives into transparent, replicable, and municipality-year indices of irregularities. It will be explained in detail that exactly how legally grounded audit findings ("bulgu": numbered items within each report) are transformed into quantitative measures that can be compared across places and over time; why alternative representational choices (from simple frequency counts to context-aware topic structures) matter for validity and interpretability; and how the resulting indices should and should not be read as indicators of governance risk. This focus is important because widely used perception-based corruption indicators have limited diagnostic value at the municipal level, whereas audit reports document specific, verifiable findings that can be standardized for systematic analysis, benchmarking, and oversight.

Irregularities, encompassing actions ranging from procedural deviations to explicit corruption, represent significant conflicts between public and private interests. The concept of irregularities, particularly corruption, has been extensively discussed across multidisciplinary fields, including law, philosophy, politics, and economics. In all governance forms, the potential for irregularities exists alongside continuous efforts to prevent them. The opportunity cost of fraudulent behavior, benefiting individual private interests, provides incentives for irregularities, while the threats posed by auditing mechanisms and public oversight counterbalance these incentives.

In this context, irregularities are often characterized by a deliberate asymmetry of information: actors engaging in misconduct strive to conceal their actions, while oversight institutions and the broader public seek to uncover them. Thus, information itself becomes the central arena in the struggle against irregularities and corruption.

Recent advances in computational techniques, particularly text mining and machine learning, have enabled the conversion of large-scale unstructured materials such as

audit reports, balance sheets, and even social media content into structured data, and subsequently, into actionable information. However, the application of these technologies has largely remained within profit-oriented domains. This study aims to demonstrate that such tools can also serve the public interest by enhancing transparency and accountability in public administration.

One of the key challenges addressed in this research is the transformation of more than 600 audit reports from the Turkish Court of Accounts into quantifiable indicators. In addition to the sheer volume of material, the inherently subjective nature of textual interpretation, especially within a politically polarized environment, necessitates a systematic and objective methodology. Manual coding of this magnitude would not only be prohibitively time-consuming and costly but would also risk inconsistency, selective interpretation, and coder bias, undermining replicability. These challenges are compounded by heterogeneity in report formatting and the technical complexity of fiscal language, making purely human classification impractical. To overcome these barriers, the study employs algorithmic data extraction and computational text analysis, aiming to minimize supervision as much as possible while still retaining a degree of structured oversight where necessary. This balance allows the indices to remain transparent, replicable, and scalable, without sacrificing interpretability. In this way, five distinct irregularity indices are constructed, each with its own definitions, methodological underpinnings, advantages, and limitations, which are discussed in detail in the following subsections.

Methodologically, the chapter proceeds step by step on the 2012–2022 audit corpus: (i) building and cleaning the text corpus; (ii) unitizing report content into discrete "audit findings"; (iii) defining a controlled vocabulary of irregularity indicators anchored in audit phrasing; and (iv) constructing five indices that increase in contextual richness. For each index, the chapter reports construction rules, coding and weighting choices, and descriptive properties (levels, dispersion, temporal evolution), alongside subgroup comparisons (metropolitan vs. non-metropolitan) and robustness checks (e.g., alternative thresholds, synonym handling, topic stability). The result is a documented, versioned measurement toolkit that captures both the intensity and the thematic composition of municipal irregularities. Subsequent chapters draw on these indices for downstream empirical analysis, but this chapter's contribution is

measurement: establishing conceptually coherent, reproducible indicators that can support research, auditing practice, and policy design.

4.1. Literature Review

Irregularities, corruption, and fraud represent distinct but closely related concepts in governance research. Irregularities typically refer to breaches of laws or regulations in public financial management, whether or not they involve personal gain. Corruption implies the abuse of public office for private benefit, often operationalized through bribery, embezzlement, or favoritism. Fraud generally denotes intentional deception for financial advantage, often with criminal liability. While analytically separable, these terms overlap in practice: all represent governance failures that distort the allocation of resources, reduce efficiency, and undermine accountability. Accordingly, the literature reviewed here encompasses studies on corruption, fraud, and broader governance irregularities, as they provide complementary methodological insights and measurement strategies.

Measuring such governance failures objectively remains challenging due to the inherent subjectivity of perception-based indices. Typically, corruption indices are derived from perception surveys (CPI by Transparency International), reported corruption experiences (World Bank surveys), or judicial conviction counts. Hessami (2014) analyzed how perceptions of corruption can diverge from actual institutional performance, highlighting biases in survey-based measures. León et al. (2013) used data from Spain and Chile to demonstrate that perception surveys systematically differ depending on the country's development level and regulatory context; they introduced a scale-adjustment algorithm to mitigate cross-country bias. Liu & Mikesell (2014) focused on U.S. municipal governments, showing that perception-based corruption indices are often inconsistent with administrative realities, thereby questioning their reliability for comparative analysis. Neiva de Figueiredo (2013) advanced this line by employing judicial conviction data instead of perceptions, thereby isolating non-compliance rates within communities and providing a more objective benchmark for corruption studies. Finally, Treisman (2007) compared perception-based and experience-based indices across multiple countries, documenting systematic discrepancies between the two and stressing the importance of identifying reliable, context-sensitive measures.

To measure corruption without a perception bias, audit reports or court convictions are more reliable sources. However, there are three challenges in transforming these texts to data manually. Firstly, in most cases it is a huge bulk of text. A large group of readers are necessary, resulting in problems on standardization. It is hard to maintain a standardized grading of the text between different readers. Secondly, this evaluation period will not be feasible because of the time it consumes or the resulting cost. Thirdly, as in the perception bias, reader of the reports has the potential of subjectivity again. Advanced text analysis methods make it possible to generate data based on large amounts of text. With the help of Natural Language Processing (NLP) or text mining applications, it is possible to transform text into data. Moreover, it overcomes all the three problems mentioned above.

For instance, Gentzkow et al. (2016) analyzed U.S. congressional speech texts to measure trends in political partisanship. Their approach began with text mining techniques, converting unstructured speech transcripts into structured data through tokenization and frequency counts, and then applied Natural Language Processing (NLP) methods by training a machine learning classifier capable of predicting a speaker's party affiliation with 73% accuracy from one-minute speech segments. This combination of text mining and NLP demonstrates how political alignment, a highly subjective attribute, can be quantified objectively from large corpora of legislative text.

In a related study, Gentzkow & Shapiro (2010) applied text mining and NLP methods to measure political slant in U.S. newspapers. They first extracted characteristic phrases used by members of Congress to distinguish between Democratic and Republican discourse, converting these linguistic patterns into structured indicators through text mining. They then applied an NLP-based classification approach to map newspaper content against these partisan phrase dictionaries, thereby inferring the ideological orientation of different outlets. This innovative use of text-as-data shows how systematic text analysis can reveal latent political biases in media, transforming qualitative impressions of slant into replicable, quantitative evidence.

Neiva de Figueiredo (2013) addressed the subjectivity in corruption measurement by developing a system model based on judicial convictions of both public officials and the general population. This model contributes significantly by isolating the inherent non-compliance rates in communities, thus clarifying causality in corruption studies.

In public governance, audit strength and transparency reduce irregularities by tightening monitoring and raising detection probability (Olken, 2007; Reinikka & Svensson, 2005; Wu & Zhu, 2011).

In public finance specifically, recent work has leveraged administrative/audit data and machine learning to build municipal risk measures and early-warning systems. Ash et al. (2020), using Brazilian municipal budget-account data as predictors and randomly assigned federal audit outcomes as "gold-standard" labels of fiscal corruption (2003–2010), train tree-based gradient-boosted classifiers to predict corruption and run policy simulations showing that ML-guided targeting could detect substantially more corrupt municipalities than random lottery audits at the same audit rate (or achieve the same detections with fewer audits). This is a canonical example of moving from detection to audit targeting using administrative data.

Using rich municipal micro-data from Brazil, Colonnelli et al. (2019) compare several supervised ML models (e.g., Random Forests, Gradient Boosting) to predict municipality-level corruption in public spending. They also decompose predictive power across feature groups (fiscal structure, procurement, political variables). Their results show that measures of private sector activity, financial development, and human capital are the strongest predictors of corruption, whereas public sector and political features have a secondary influence. The models achieve strong out-of-sample performance, offering actionable risk scores to guide oversight and policy interventions.

Gallego et al. (2021) investigate whether corruption and inefficiency in public procurement can be systematically predicted using machine learning. Leveraging the rapid expansion of e-procurement platforms, they argue that governments and watchdog institutions now possess unprecedented amounts of transaction-level data that can be harnessed for predictive oversight. Their study uses a unique dataset of more than 2 million Colombian public contracts, training ML models to identify which contracts are subsequently investigated for corruption or suffer implementation inefficiencies. Methodologically, they confront the problem of class imbalance—since only a small fraction of contracts are corrupt—through resampling and weighting techniques, while also testing different algorithms. Their models deliver high predictive accuracy and, crucially, illuminate the trade-off between precision and

recall in practical oversight applications. By decomposing feature importance, they highlight which contract characteristics most strongly correlate with malfeasance, offering interpretable guidance for targeting audits. For policy, the study underscores both the promise and challenges of applying ML to procurement oversight: while accuracy and scalability are achievable, decisions about thresholds, false positives, and false negatives remain critical. This work exemplifies how large administrative datasets can be repurposed into governance risk tools, a perspective closely aligned with the methodological contributions of your dissertation.

Public accessibility of information represents another essential factor in mitigating irregularities. Dong & Torgler (2013) analyzed the impact of free information access, highlighting decentralized, primarily internet-based media as particularly effective. Audit structure and strength are equally critical. Wu & Zhu (2011) quantified anti-corruption efforts using the number of audit personnel, audit cases, and uncovered illegal funds, demonstrating these factors significantly reduce corruption, thereby enhancing wealth and reducing income disparities.

Early work in information retrieval and BoW representations established how to translate documents into tokens and counts/weights for quantitative modeling (Manning et al., 2008; Salton & Buckley, 1988). The underlying distributional logic, that meaning can be inferred from usage (Harris, 1954), underpins later developments in text-as-data for the social sciences (Gentzkow et al., 2019; Grimmer & Stewart, 2013). Domain-specific lexicons and weighting schemes have been particularly influential in legal/financial text (Loughran & Mcdonald, 2011).

Within economics and public administration, topic models (notably LDA) provide a principled way to uncover latent thematic structure without imposing categories ex ante (Blei et al., 2003; Griffiths & Steyvers, 2004). Advances in estimation (e.g., collapsed Gibbs sampling) and validation (topic coherence and stability) support robust application in policy settings (Newman et al., 2010; Wallach et al., 2009). These tools are well-suited to legal audit text where findings can span multiple domains (procurement, asset management, budgeting, personnel), and where unsupervised categories improve reproducibility across report vintages.

4.2. Methodology and Data

The objective is to translate Turkish Court of Accounts (Sayıştay) audit reports into comparable, municipality–year measures of irregularities. The unit of analysis is a municipality–year; the corpus spans 2012–2022 and comprises more than 600 reports. Each report is a legal document written in public-law register and organizes identified problems as numbered audit findings ("bulgu", for the reader to better understand the process, sample report texts are introduced in appendix: Figure A3 and Figure A4), each citing legal bases and describing the issue. The pipeline proceeds as follows: (i) build and clean a text corpus of all reports; (ii) unitize the text into finding-level records; (iii) transform text into quantitative representations; and (iv) construct five indices that escalate in contextual richness—Indicator Count, Weighted Indicator Count, Case (Finding) Count, Weighted Category-based Case Count, and an LDA-Supported Index—followed by standardization and robustness checks. Throughout, we favor transparent, reproducible steps grounded in the text-as-data literature in economics and information retrieval (Gentzkow et al., 2019; Manning et al., 2008; Salton & Buckley, 1988).

Corpus construction and preprocessing. PDFs are parsed to plain text; reports are split at the official headings for numbered findings. We normalize to lowercase; remove punctuation, numerals, boilerplate headers/footers; and drop Turkish and legal stop-words. Tokenization yields a document–term representation at the finding level. Let $\mathcal{D}$ denote the set of findings and $\mathcal{V}$ the vocabulary; for finding $d \in \mathcal{D}$ and term $v \in \mathcal{V}$, let x_{dv} be the count (Bag-of-Words/BoW). BoW is a standard, effective baseline for legal/administrative text where cue words carry strong signal, and can be optionally TF-IDF weighted to discount ubiquitous terms.

Index 1 — Indicator Count: Working with legal experts, we build an indicator dictionary $\mathcal{V}_{ind} \subset \mathcal{V}$ of cue terms (and phrases) that signal potential irregularities. For municipality m in year t, let $\mathcal{D}_{mt}$ be its set of findings. The index is the population-standardized sum of indicator hits:

$$I_{mt}^{(1)} = \frac{1}{\text{Pop}_{mt}} \sum_{d \in \mathcal{D}_{mt}} \sum v \in \mathcal{V}_{ind}\, x_{dv} \tag{8}$$

This measure assumes frequency of negative cue words correlates with irregularity intensity.

Index 2 — Weighted Indicator Count: To reflect heterogeneous signaling strength, partition $\mathcal{V}_{ind}$ into tiers $g \in \{1, \dots, G\}$ with weights $w_g > 0$ elicited from legal experts (stronger cues receive higher weights). Define

$$I_{mt}^{(2)} = \frac{1}{\text{Pop}_{\text{mt}}} \sum_{d \in \mathcal{D}_{mt}} \sum g = 1^{G} w_g \sum_{v \in \mathcal{V}^{(g)}_{\text{ind}}} x_{dv} \tag{9}$$

Where 1^G is an indicator function that returns 1 if the (existence) condition is true. Equation 5 yields a severity-weighted frequency proxy.

Index 3 — Case Count: Let $C_{mt} \equiv |\mathcal{D}_{mt}|$ be the number of audit findings in municipality m, year t. The index is a scaled count:

$$I_{mt}^{(3)} = \frac{C_{mt}}{\text{Pop}_{mt}} \tag{10}$$

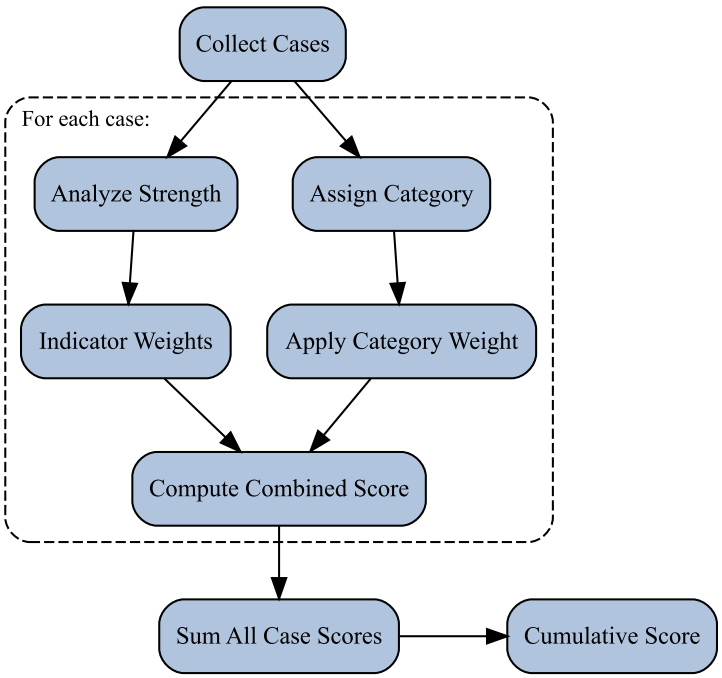
a legally grounded, model-free quantity that tracks the accumulation of distinct issues.

Index 4 — Weighted Category-based Case Count: This index uses supervised category assignments to audit findings. The audit reports were analyzed through a combination of manual review and the application of Large Language Models (e.g., GPT-4) to derive a comprehensive list of case categories that will be introduced in the next section (Table 19). Classification uses a refined keyword mapping with tie-breakers (best-match score), assigning each finding d to a primary category $c(d) \in \mathcal{C}$. Each finding also receives a within-finding severity score s_d based on the weighted indicator algorithm in Index 2. The municipality-year score is:

$$I^{(4)}mt = \sum_{d \in \mathcal{D}_{mt}} \omega_{c(d)} \, s_d, \tag{11}$$

where $\omega_c \geq 0$ are category weights. We calibrate ω via grid search to maximize out-of-sample validity (predictive), with sensitivity checks; weights are kept fixed for comparability across years. Figure 9 presents a flow chart of the algorithm.

Figure 9: Flow Chart for Weighted Category Based Index



Index 5 — LDA-Supported Index (unsupervised topic structure): To reduce subjectivity and capture latent themes, we estimate Latent Dirichlet Allocation (LDA) on the finding-level corpus (Blei et al., 2003). LDA posits: for topics $k = 1, \ldots, K$, draw topic–word distributions $\phi_k \sim \mathrm{Dir}(\eta)$; for each finding d, draw document–topic mixture $\theta_d \sim \mathrm{Dir}(\alpha)$; for each token n in finding d, draw topic $z_{dn} \sim \mathrm{Mult}(\theta_d)$, then word $w_{dn} \sim \mathrm{Mult}(\phi_{z_{dn}})$. The joint distribution is:

$$p(W, Z, \Theta, \Phi | \alpha, \eta) = \left(\prod_{k=1}^{K} p(\varphi_k | \eta) \right) \left(\prod_{d \in \mathcal{D}} p(\theta_d | \alpha) \prod_{n=1}^{N_d} p(z_{dn} | \theta_d) p\left(w_{dn} | \varphi_{z_{dn}}\right) \right) \quad (12)$$

Exact posterior inference is intractable; we use collapsed Gibbs sampling (Griffiths & Steyvers, 2004), integrating out Θ and Φ and sampling token-level z with:

$$p(\, z_{dn} = k \mid z_{-dn}, W, \alpha, \eta \,) \propto \left(n_{dk}^{-dn} + \alpha_k \right) \frac{n_{kw}^{-dn} + \eta_w}{n_{k\cdot}^{-dn} + \sum_{v \in V} \eta_v} \quad (13)$$

where n_{dk}^{-dn} counts tokens in finding d assigned to topic k (excluding token n); n_{kw}^{-dn} counts word type w under topic k; and $n_{k\cdot}^{-dn}$ is the topic-k total. Implementation uses R's topicmodels with method = `"Gibbs"` (Hornik & Grün, 2011). We set $K = 7$ based on interpretability and semantic coherence diagnostics (Blei, 2012), and retain posterior mean topic proportions $\widehat{\theta_d}$.

The choice of the number of topics is a critical step in topic modeling. Too few topics risk conflating distinct irregularity domains into broad categories, while too many lead to fragmentation and unstable themes. In this study, $K = 7$ was selected after testing multiple values using both quantitative diagnostics (perplexity, log-likelihood, and topic coherence scores) and substantive interpretability:

- Quantitative diagnostics: Candidate models between $K = 5$ and $K = 12$ were evaluated. While statistical fit continued to improve marginally with higher K, coherence scores stabilized around $K = 7$, indicating that topics were internally consistent without overfitting noise.
- Semantic interpretability: At $K = 7$, the model produced thematically coherent and policy-relevant clusters that aligned with established governance domains:

    1. Accounting & Asset Management
    2. Budgeting & Financial Oversight
    3. Procurement & Public Contracts
    4. Real Estate & Related Incomes
    5. Personnel & Human Resources
    6. Internal Control & Risk Management
    7. Social Services & Public Oversight

Each cluster corresponded to identifiable areas of irregularities reported by the Court of Accounts, providing face validity and interpretability essential for policy applications. In contrast, smaller K values (e.g., 5) merged unrelated categories (procurement with asset management), while larger values (e.g., 10+) produced fragmented topics with overlapping vocabularies that were harder to interpret consistently.

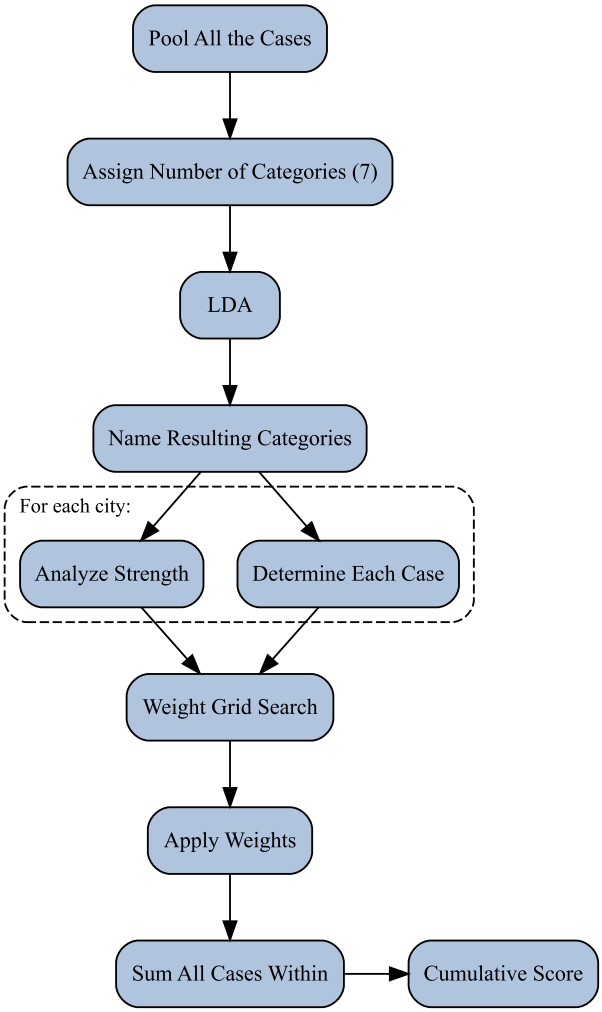
Thus, seven topics offered the best balance between statistical adequacy and substantive clarity, allowing the results to be mapped meaningfully onto fiscal and governance structures.

To convert topics into a municipality-year score, we (i) label topics ex post by their top-probability terms (for interpretability), and (ii) aggregate topic mass across all findings in m, t, applying topic weights λ_k (set to 1 for a neutral variant, or tuned via grid search for criterion validity). A general form is:

$$I_{mt}^{(5)} = \sum_{d \in \mathcal{D}_{mt}} \sum k = 1^{K} \lambda_k \, \hat{\theta}_{dk} \tag{14}$$

This unsupervised index is replicable (given the corpus and hyperparameters) and adapts when the thematic mix of findings shifts over time. Figure 10 presents a simplified flow chart of LDA supported index.

Figure 10: Flow Chart of LDA Supported Index



All indices are population-standardized (per municipal person; results reported per 100,000 where noted) to account for scale. We detect and remove boilerplate segments and conduct spot audits to ensure correct finding segmentation and category assignment.

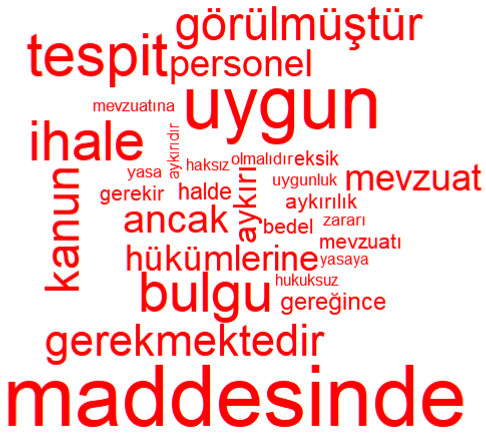
Starting from BoW enables transparent, legally interpretable signals. Weighting by cue strength and by domain folds contextual information into the measure while retaining audit-text traceability. LDA provides an unsupervised cross-check that reduces dependence on hand-crafted categories and captures shifts in the latent structure of findings (Blei, 2012; Blei et al., 2003), with fast, well-understood posterior inference via collapsed Gibbs sampling (Griffiths & Steyvers, 2004). The result is a suite of indices that trade off simplicity and context in a controlled, reproducible way, suitable for descriptive statistics, benchmarking, and downstream empirical work.

4.3. Indices

4.3.1. Indicator Count:

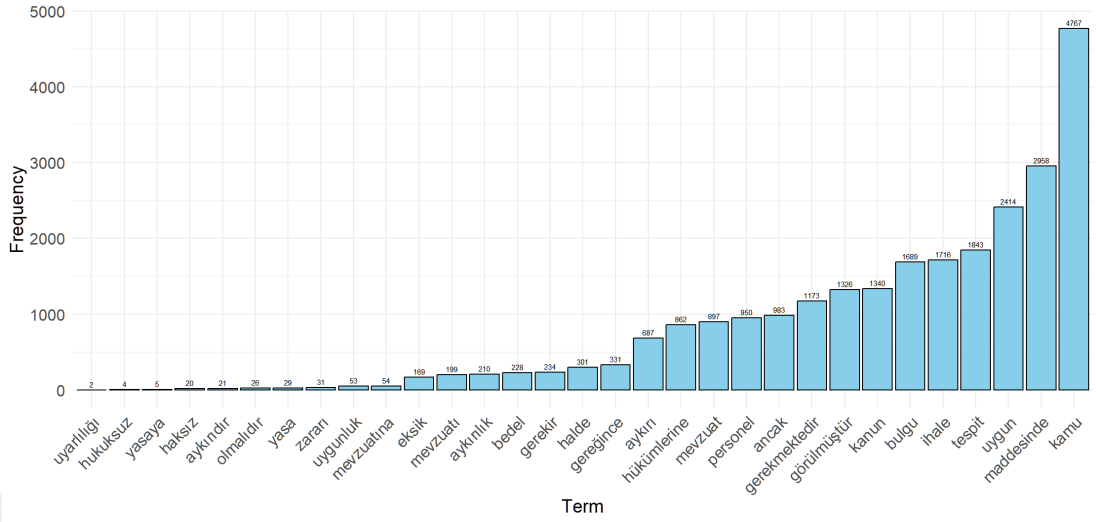
This index can be defined as a bag-of-words (BoW) frequency index that counts the occurrence of a pre-specified lexicon of irregularity indicators[8] within each report and scales by municipal population. Given the specific characteristics of audit reports, namely their reliance on stable legal terminology and exclusive focus on identified problems, a customized word indicator tokenization process can be effectively applied. These two properties make the reports particularly suitable for a basic frequency-based method. Therefore, the identification of irregularity indicator words becomes a critical step. Following the examination of several sample reports and consultations with a group of domain-specific legal experts, a comprehensive list of indicators was compiled. Higher values suggest more frequent appearance of problem-bearing terms across findings. Figure 11 visualizes the indicator lexicon (word cloud) and Figure 12 reports term frequencies for the 2022 sample.

Figure 11: Word Cloud of Indicator Words for the Sample of 2022 Reports



[8] Full list of words is as follows: “ancak", "bulgu", "karşın", "karşin", "aykırı", "aykiri", "aykırıdır", "aykiridir", "uygunluk", "uygun", "halde", "olmalıdır", "olmalidir", "yasaya", "aykırılık", "aykirilik", "gerekir", "gerekmektedir", "hükümlerine", "haksız", "haksiz", "hukuksuz", "tespit", "görülmüştür", "gorulmustur", "rastlanmıştır", "rastlanmistir", "eksik", "bedel", "personel", "kanun", "yasa", "kamu", "zararı", "zarari", "ihale", "mevzuat", "mevzuatına", "mevzuatina", "mevzuatı", "mevzuati", "gereğince", "gereginc e", "uyarlılık", "uyarlilik", "uyarlılığı", "uyarliligi", "uygun", "maddesinde"

Figure 12: Indicator Term Frequencies for 2022 Reports



4.3.2. Weighted Indicator Count:

This index is similar to the first one, but each indicator term is assigned a severity weight. Not all tokens are equally diagnostic in legal jargon. Weighting sharpens the signal by giving higher influence to terms that more directly denote irregularity (e.g., ihale, zarar; “auction”, “loss”). This index still measures intensity, but with severity emphasis embedded. Figure 13 lists the weights and representative terms.

Figure 13: Indicator Weights



4.3.3. Case Count:

This index is basically the total number of numbered findings in each municipal report, scaled per 100,000 population. Sayıştay structures each report around discrete, legally grounded findings. Their count is a direct, case-based measure of irregularity breadth (the number of distinct issues documented), independent of wording frequency.

4.3.4. Weighted Category-based Case Count:

For this index, each finding is assigned to one thematic category and then two weights are applied: (i) a within-finding severity score derived from the weighted indicators (as in **second index**) and (ii) a category weight reflecting the relative governance salience of that domain. The municipality–year score is the population-normalized sum across findings.

By adding thematic context, the index moves beyond raw counts to capture which governance areas drive the burden of irregularities. It therefore reflects both intensity and composition. Categories, indicators, and final weights are reported in Table 19.

To construct a robust algorithm, the audit reports were analyzed through a combination of manual review and the application of Large Language Models (e.g., GPT-4) to derive a comprehensive list of case categories. The subsequent step involved assigning weights to these categories. Weighting was conducted using a grid search approach applied within the context of the best-performing machine learning models. This process enables the identification of case types that are most strongly associated with expenditure patterns, that is, the case types that are most predictive of irregularities.

Table 19: Categories and Category Indicators for Case Determination and Relative Weights Assigned

	Indicators	Weights
Financial Structure	"mali", "finans", "bütçe", "gelir", "gider", "kredi", "borç", "defter", "defterinde", "fatura", "faturalarının", "faturaları", "faturanın", "hesabı", "muhasebeleştirme", "muhasebe", "kayıt", "kayıtları", "kayıtlarının", "amortisman", "hesabına"	0.10
Auction Practices	"ihale", "teklif", "satın alma", "kiralamasında", "doğrudan temin", "pazarlık", "bedel", "proje", "yüklenici", "sözleşme", "rekabet", "rekabete", "saydam", "şeffaf", "satış","ihalede", "işlettirmek", "satışı", "imtiyaz"	0.35
Assets Management	"varlık", "taşınmaz", "taşınmazlar", "taşınmazların", "mülkiyet", "envanter", "amortisman", "arazi", "bina", "kira geliri", "kira", "tahsis", "tahsisler", "tahsislerde", "imar", "plan"	0.25
Internal Control Mechanisms	"kontrol", "denetim", "iç denetim", "yönetim", "hata", "ruhsat", "gayrisıhhi"	0.10
Personnel Issues	"sigorta", "personel", "sgk", "prim", "işçi", "kadro", "kadrosu", "kadrosuna", "ücret", "sözleşmeli", "sözleşme", "personeli", "istihdam", "personele", "657", "sendika", "memur", "çalışan", "çalışanlara"	0.10
Service Purchases	"hizmet alım", "ihale", "kiralama", "işçi", "fiyat"	0.05
Others	Unclassified ones	0.05

4.3.5. LDA Supported Index:

The previous case category enhanced index utilized predefined categories for the evaluation of reports. While this approach allowed for a structured evaluation, it relied on supervised category definitions, which are subjective. A change in category definitions or their indicators could lead to a variation in the index outcomes. To decrease subjectivity even more, an unsupervised topic model (Latent Dirichlet Allocation) is fit to the corpus of findings; each finding receives a topic mixture and topics are characterized by their term distributions. This index captures intensity and latent thematic structure, improving transportability over time and across jurisdictions. Municipality–year scores aggregate finding-level topic-weighted severities, with population normalization.

To summarize, the LDA application has several advantages in this context:

- Objectivity: Categories emerge from data itself.

- Reproducibility: Given the same reports and parameters LDA produces consistent results.
- Interpretability: Once topics are generated, they can be easily named using the most probable terms identified.

In this study, `topicmodels` package in R with collapsed Gibbs sampling is used. The number of topics is set to 7 based on interpretability and coherence based on trials. Following topic modelling, the resulting 7 categories are named as shown in Table 20. Then for each city (report) the cases are evaluated using LDA defined topic specific indicator weights, which are optimized via grid search. Finally, the weighted topic scores for each case are aggregated at the city level to gain a cumulative irregularity score for each city-year combination. The overall summary of the process can be traced in algorithm flow chart presented in Figure 10. This methodological improvement enhances the robustness and generalizability of the index.

Table 20: LDA Topics

LDA Topic Titles and Keywords - 7 Category Case

A summary of topics with their top 10 representative keywords

Topic Title	Top 10 Representative Keywords
Accounting and Asset Management	muhasebe, hesabına, hesabı, kamu, yer, mali, duran, bütçe, alacak, değer
Public Oversight and Social Services	kamu, denetim, sosyal, yıllık, kontrol, iş, iç, izin, personel, yer
Urban Planning and Transportation	toplu, taşıma, ihale, trafik, ruhsat, ulaşım, karar, işyeri, çalışma, denetim
Real Estate and Related Incomes	kira, kamu, ihale, tahsis, kiraya, taşınmaz, hakkı, yer, ait, denetim
Budgeting and Financial Oversight	hesabi, hesabı, faaliyet, giderleri, alacaklar, mali, denetim, gider, vadeli, banka
Taxation and Revenue Collection	tahsil, tahakkuk, vergisi, gelir, vergi, otopark, kamu, denetim, ilçe, ilan
Procurement and Public Contracts	ihale, hizmet, kamu, iş, yapım, yer, denetim, uygun, alımı, mal

The LDA-based topic classification reveals that Accounting and Asset Management consistently dominates the audit case distribution across all years, accounting for nearly one-third of all findings (Table 21). This dominance, however, shows a marked decline over time: from exceptionally high levels in 2012–2014 (above 40%, even exceeding 80% in 2012) to only 21% in 2022. Such a drop suggests either an improvement in internal controls in this domain, a redistribution of audit attention toward other categories, or a combination of both.

Table 21: LDA Topics Summary Statistics

LDA Topic Category Percentages by Year													
Topic Category	Total %	Total	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
Accounting and Asset Management	32.2	2,762	81.1	51.7	40.7	32.8	35.4	35.3	29.1	26.0	34.8	25.4	21.0
Public Oversight and Social Services	11.4	980	1.0	10.3	12.1	12.5	9.3	10.3	8.7	13.3	12.8	12.3	13.7
Urban Planning and Transportation	5.9	503	0.0	1.2	1.0	5.0	4.3	4.9	7.2	6.9	6.7	7.6	8.7
Real Estate and Related Incomes	13.8	1,184	1.0	4.0	11.0	12.3	15.2	14.4	14.5	14.9	13.5	15.2	16.9
Budgeting and Financial Oversight	5.0	431	4.5	15.2	13.5	13.8	8.3	6.0	5.4	1.4	1.0	0.5	0.7
Taxation and Revenue Collection	18.0	1,545	10.4	10.9	14.8	15.8	18.7	18.2	18.0	22.1	18.3	19.0	17.6
Procurement and Public Contracts	13.7	1,179	2.0	6.7	6.9	7.8	8.8	11.0	17.1	15.5	12.9	20.1	21.4

Other prominent categories are Taxation and Revenue Collection, Real Estate and Related Incomes, and Procurement and Public Contracts. Together, these areas make up nearly half of the total findings. Taxation-related irregularities show a relatively stable yet high share (between 10–22% annually), reflecting the persistent fiscal sensitivity of revenue management. Procurement and public contracts gradually gain importance after 2015, rising above 20% by 2022, suggesting heightened audit scrutiny or increased risks in contracting processes. Real estate and related incomes exhibit a steady upward trajectory, climbing from 4% in 2012 to nearly 17% in 2022, signaling growing attention to asset-based governance risks, especially as municipalities expand their involvement in land management and property revenues.

By contrast, Budgeting and Financial Oversight, Urban Planning and Transportation, and Public Oversight and Social Services remain relatively underrepresented throughout the period. The consistently low presence of budgeting-related findings (below 6% in almost all years) may reflect either under-auditing in this category or that formal budgeting rules are adhered to more strictly compared to discretionary domains such as procurement. Urban planning and transportation, although small in share, gain gradual relevance after 2016, perhaps mirroring the rising importance of urban development policies and infrastructure projects in Türkiye's municipal agenda. Social services also grow moderately in share, from 1% to 13%, suggesting an expansion of audit coverage into domains directly affecting citizen welfare.

Overall, the temporal distribution shows a diversification trend: while earlier years were overwhelmingly dominated by accounting irregularities, later years display a more balanced mix across taxation, procurement, real estate, and social service domains. This indicates not only shifts in the substantive risks of municipal

governance but also a potential reorientation of Sayıştay's auditing priorities. Substantively, the results highlight that while "hard" fiscal domains (accounting, taxation, procurement) remain central, issues tied to assets, contracts, and social accountability are increasingly shaping the irregularity landscape.

4.4. Comparison and Summary Statistics of Indices

Table 22 gives the summary statistics of all the indices. Remind that numbers in the tables represent the values per 100.000 municipal population. The indices collectively indicate a shift from concentrated irregularities in the early 2010s toward a more widespread, complex, and varied audit environment by the early 2020s. This is evidenced by increasing means and standard deviations in both supervised (e.g., Case Category Weighted) and unsupervised (LDA Weighted) measures, supporting the robustness and multidimensionality of the composite irregularity framework.

Table 22: Summary Statistics of the Indices

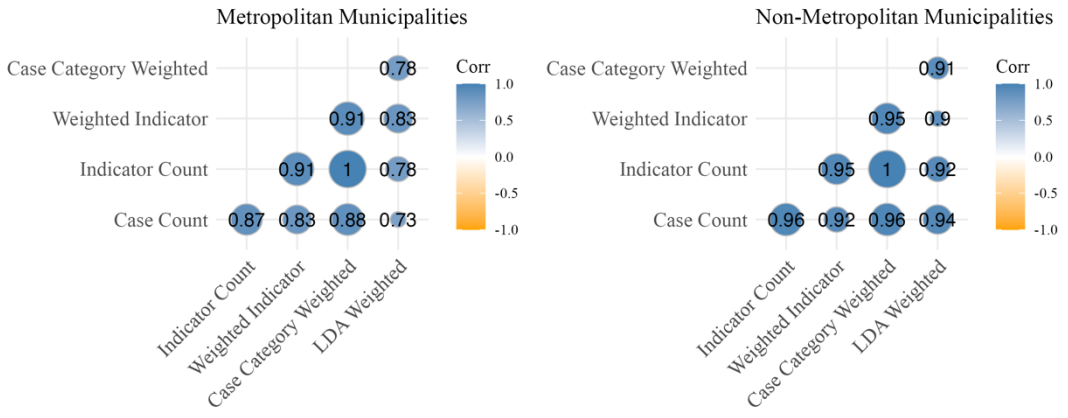
	Count	Mean	SD	Min	Max
Case Count	543	7.29	13.99	0.03	143.94
Indicator Count	543	150.60	249.49	1.12	2,355.30
Weighted Indicator	543	13.12	25.72	0.02	211.02
Case Category Weighted	543	259.07	428.59	2.08	3,921.14
LDA Weighted	543	74.18	157.90	0.00	1,627.95

Table 23: Yearly Summary Statistics for Indices

	Case Count		Indicator Count		Weighted Indicator		Case Category Weighted		LDA Weighted	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
2012	3.9	4.2	49.9	50.1	1.7	2.0	94.7	102.2	35.8	46.8
2013	3.0	3.7	57.8	63.0	4.0	7.0	92.3	109.2	22.4	41.5
2014	4.6	5.6	103.8	97.9	5.4	6.6	172.0	164.1	30.9	53.8
2015	4.8	6.1	120.5	117.9	6.1	8.4	201.2	192.4	35.0	60.7
2016	8.9	15.0	156.1	245.8	16.7	27.5	265.9	410.2	78.0	157.8
2017	10.0	20.5	187.8	355.2	22.0	41.6	311.0	587.9	124.0	266.8
2018	5.7	9.1	173.3	209.2	13.6	18.4	303.8	363.8	69.3	111.5
2019	10.2	15.3	203.6	291.2	18.8	30.0	360.8	494.0	106.0	170.9
2020	5.9	9.7	116.8	174.3	9.8	15.1	203.8	309.7	75.2	127.5
2021	11.7	26.5	225.4	458.4	21.7	43.1	399.9	807.8	124.9	272.2
2022	10.7	15.4	223.8	278.2	21.9	29.3	389.1	480.4	117.2	161.7

As demonstrated in Figure 14, the LDA-based scoring method appears to contribute more independent explanatory value in metropolitan municipalities, where irregularities possibly be more varied and multidimensional. In contrast, the extremely high correlations in non-metropolitan municipalities suggest parallel results between different scoring methods, potentially due to less complex structures.

Figure 14: Correlation Matrix of Indices Grouped by Municiaplity Type



5. PREDICTING IRREGULARITIES IN TURKISH MUNICIPALITIES

5.1. Introduction

The purpose of this chapter is to test whether municipal irregularity risk can be predicted directly from routine financial statements and, if so, to identify which balance-sheet and budget lines carry the strongest predictive signal. Using the audit-based irregularity indices constructed in the previous chapter as targets, the analysis asks a simple but policy-salient question: given a municipality's balance sheet and operating accounts, what is the expected level and composition of audit-relevant irregularities? The significance of such a mapping is twofold. First, it enables a scalable early-warning tool that prioritizes scarce audit capacity before findings materialize in ex-post reports. Second, it provides actionable diagnostics for municipal managers by pinpointing the financial levers most associated with predicted risk.

Concretely, the chapter addresses four research questions:

1. Predictability: To what extent do standard financial accounts predict out-of-sample variation in audit-based irregularity indices?
2. Feature salience: Which accounts (e.g., personnel/social security expenses, enterprise and ownership revenues, short-term liabilities, tax revenues) most strongly drive predicted risk, and how stable are these effects across indices and municipality types?
3. Representation quality: How does predictive performance vary across target constructions, from simple counts to context-aware (topic-model) indices, holding the learner class fixed?
4. Interpretability and use: Can predictions be decomposed into transparent, case-selectable explanations that support oversight decisions without conflating correlation with causation?

To answer these questions, the study trains Gradient Boosted Machines (GBM) and Random Forests (RF) on a province–year panel of municipal financials (2012–2022), with five irregularity indices as continuous targets. Inputs comprise core balance-sheet and operating items; outputs are predicted index values. Model development follows an 80/20 train–test split with five-fold cross-validation and grid search for hyperparameters. Performance is evaluated with R^2, RMSE, and RMSE standardized by the target's standard deviation (RMSE/SD) to allow comparability across indices. To maintain decision transparency, SHAP (SHapley Additive exPlanations) is used to attribute each prediction to specific accounts, yielding municipality- and year-specific "why" diagnostics alongside the point forecasts. The chapter concludes with brief case illustrations showing how a single year's balance sheet translates into a predicted irregularity score and a SHAP breakdown that can guide targeted audit or managerial action.

The chapter is structured as follows. The second section presents a brief review of the relevant literature on the application of machine learning in public sector oversight and irregularity detection. The third section introduces the dataset used in the analysis, including both balance sheet variables and the five irregularity indices developed earlier. The fourth section outlines the methodological framework, including ML discussion, model training, validation procedures, and performance evaluation criteria. The fifth section reports the prediction results across all five indices. Finally, the sixth section presents selected municipal case studies to illustrate the practical applicability of the proposed models in identifying high-risk areas and guiding oversight efforts.

5.2. Literature Review

The availability of the structured and unstructured public data has led to growing interest in the use of ML techniques for public sector applications, particularly in the domains of fraud detection, audit support, and financial oversight. Unlike traditional econometric models, which require strong assumptions about functional forms and variable distributions, ML algorithms are designed to uncover complex, non-linear patterns in high dimensional datasets, even in unbalanced scenarios. This flexibility has made them especially valuable for identifying irregularities that may manifest subtly or vary significantly across contexts.

The evolution of machine learning methodologies, from early decision tree algorithms to advanced ensemble techniques such as Random Forests (Breiman, 2001) and Gradient Boosted Machines (Friedman, 2001), has significantly improved both the accuracy and interpretability of predictive analytics in public administration. These models have been increasingly applied in domains such as tax evasion detection, procurement fraud, and municipal risk assessment. In particular, ensemble methods offer robustness to overfitting and the capacity to rank variable importance, a feature particularly relevant in evaluating financial indicators that may be associated with irregularities.

In parallel with advances in prediction, interpretability has become a central concern in the deployment of ML models in sensitive domains. Tools such as SHAP (SHapley Additive exPlanations) have emerged to bridge the gap between predictive power and transparency, enabling researchers and practitioners to attribute model outputs to specific features in a theoretically grounded manner (S. Lundberg & Lee, 2017). This is especially critical in public sector applications, where accountability, transparency, and decision justification are paramount.

Recent studies have demonstrated the efficacy of ML in auditing contexts. For example, Jans et al. (2010) applied classification algorithms to detect internal control weaknesses in accounting systems, while Baesens et al. (2015) provided a comprehensive overview of data mining applications in fraud detection. In the field of public finance, ML has been employed to assess fiscal sustainability, forecast revenue and expenditure patterns, and flag anomalous reporting behaviors (Li et al., 2012; Sun et al., 2016). Despite these advances, the integration of ML into municipal-level audit analytics remains relatively underexplored, particularly in developing and middle-income country contexts.

This study contributes to the literature by applying GBM and Random Forest models to predict audit-based irregularity scores using financial and operational data from Turkish municipalities. In addition to advancing methodological practice, the inclusion of SHAP-based interpretability ensures that model outputs are accessible and meaningful to policymakers and oversight institutions.

Numerous studies have explored the primary mechanisms countering irregularities, namely electoral accountability and auditing processes. Ferraz & Finan (2009)

demonstrated empirically, using Brazilian data, that electoral accountability significantly reduces corrupt behavior. They highlight information asymmetry as a critical factor influencing electoral incentives and auditing effectiveness. Publicly available audit reports can substantially affect election outcomes; Ferraz & Finan, (2008) found that each additional reported irregularity reduced re-election probabilities by approximately 20%, a relationship amplified by local media presence, such as radio stations. Winters & Weitz-Shapiro (2013) examined voter support for politicians engaged in irregular practices, concluding that voters, when provided sufficient information, show zero tolerance toward corrupt officials. Complementary findings by Bobonis et al. (2016) indicate that disseminating audit reports prior to elections temporarily reduces corrupt practices. Moreover, Clausen et al. (2011) associate higher corruption levels with reduced confidence in public institutions and diminished political engagement, including increased emigration.

A second key research area involves examining the manifestations of irregularities within public budgets. Transparency International defines corruption specifically as "the abuse of entrusted power for private gain," typically occurring through public resource misallocation. Hence, budget composition emerges as a critical indicator of irregularities. Daniele & Giommoni (2021) analyzed the effect of budget restrictions in small Italian municipalities, demonstrating decreased irregularities specifically during pre-election periods due to constraints on capital expenditures, without negatively impacting overall welfare. Hessami (2014), employing OECD data from 1996 to 2009 and using Transparency International's Corruption Perceptions Index, found significant shifts in public expenditure allocation correlated with perceived corruption levels. Specifically, higher corruption was associated with increased spending on health and environmental protection. However, Hessami's reliance on perception-based indices limited insights into the actual magnitude of irregularities. Liu & Mikesell (2014) addressed this limitation by analyzing corruption convictions across U.S. states, concluding that increased irregularities correlated positively with excessive per capita expenditures in specific budget categories, notably capital projects and construction, and negatively with education and healthcare funding. Similarly, Mauro (1998) highlighted the education sector as particularly susceptible to budgetary distortions stemming from irregularities.

Another significant strand of literature addresses irregularity prediction. López-López-Iturriaga & Sanz (2018) developed an early-warning corruption detection model using neural networks, focusing on economic and institutional variables within Spanish provinces. Their findings indicate that real estate taxation, economic growth rates, and local governance tenure significantly predict corruption occurrences. Gallego et al. (2021) adopted a predictive approach centered on public contract data from Colombia, identifying key variables linked to faulty transactions. Similarly, Colonnelli et al. (2019), applying machine learning techniques to Brazilian municipalities, identified local economic variables, financial development, demographics, and human capital as significant predictors of irregularities. Nevertheless, causality between economic prosperity and increased irregularities remains contested.

An emerging literature emphasizes applying machine learning techniques for irregularity detection and providing policy guidance to audit institutions. Colonnelli et al. (2019) trained multiple machine learning models, Random Forest, Gradient Boosting, neural networks, and LASSO, on Brazilian data, subsequently developing an ensemble model to address individual model weaknesses. Ash et al. (2020) developed a corruption detection model using budgetary account data from Brazilian municipalities, contributing to the literature by creating extensive datasets. Their predictive index, covering data from Brazil's 26 states and approximately 5,500 municipalities between 2001-2012, included 797 fiscal categories obtained from the Finance Ministry's online database. Government auditors identified irregularities through a randomized selection process, generating reports utilized as corruption indicators. Utilizing a Gradient Boosting classifier (XGBoost), Ash et al. employed an 80-20 training-test split and five-fold cross-validation for hyperparameter tuning. Their model substantially outperformed Ordinary Least Squares (OLS) regression, achieving an AUC-ROC score of approximately 0.8 and an F1 score around 0.67, compared to OLS scores of 0.56 and 0.41, respectively.

Local literature on corruption in Türkiye, although extensive, notably lacks empirical research, particularly concerning municipal budgets. Kıvratma (2016) highlights the limited availability of municipal budget data, examining municipalities exceeding 100,000 population in 2013 for diversity analysis. Akdoğan (2008) explored relationships between financial and service performance of municipalities, while

Karpuz (2008) conducted a descriptive analysis of municipal debt structures from 1996 to 2006. Thus, this study significantly advances empirical research by applying robust machine learning methodologies to Turkish municipal financial data.

5.3. Data

The data that will be used in this chapter are constructed and used in the previous chapters, a brief summary and corresponding framework will be introduced for this chapter. For machine learning (ML) algorithms, the dataset must be structured to clearly define both predictors and a target variable. Predictors include the financial variables previously introduced in the DEA section, complemented by additional balance sheet variables. Specifically, the annual service supply costs categorized under operational expenditures represent input variables for the DEA model, consistent with the existing literature. However, irregularities identified through audits typically encompass a wider scope than just operational expenditures, extending to all balance sheet and operational income titles. Therefore, to adequately capture these irregularities, balance sheet values and operational income variables have been integrated into the initial input set derived from the DEA model, thus forming the comprehensive predictor set for the ML analysis.

5.4. Methodology

The primary motivation of this research is to trace municipal irregularities through their reflection in balance sheet data. The underlying hypothesis suggests that irregularities in municipal transactions should manifest in balance sheet records, which can be detected and analyzed effectively through Machine Learning (ML) methods. Thus, ML models enable proactive policy recommendations for municipalities and auditing institutions by signaling potential irregularities in advance. Additionally, these models facilitate the identification of balance sheet variables most significantly associated with irregularities.

At the foundation of many supervised machine learning algorithms lies the decision tree, a flowchart-like structure where internal nodes represent decision rules based on input variables, branches represent outcomes of those decisions, and leaf nodes represent predicted outputs. Decision trees are valued for their simplicity,

interpretability, and capacity to handle both categorical and numerical data. However, single decision trees are prone to overfitting, particularly when the data is noisy or complex (Quinlan, 1986).

To overcome these limitations, ensemble learning approaches such as Random Forests (RF) and Gradient Boosted Machines (GBM) have been developed. These methods aggregate multiple decision trees to improve predictive performance and generalizability.

Two ensemble methods are selected for their strong performance in structured data tasks and their interpretability: Random Forests (Breiman, 2001) and Gradient Boosted Machines (Friedman, 2001). Compared to distance-based models such as Support Vector Machines or k-Nearest Neighbors, both RF and GBM are more robust to outliers, variable scaling, and non-linear interactions. Their decision-tree foundations allow for natural handling of categorical variables and missing values.

RF is particularly noted for its stability and resistance to overfitting, while GBM excels in capturing subtle patterns through iterative model refinement. Both models provide feature importance scores, enabling transparency and interpretability, an essential feature for public sector applications where explainability is paramount.

The Random Forest (RF) algorithm is employed to enhance robustness and provide complementary insights. RF combines multiple decision trees built independently using bootstrapped samples of the dataset, ensuring each tree has reduced correlation through random subset selection of predictors. Unlike GBM's sequential error minimization, RF averages predictions across trees, mitigating overfitting and enhancing prediction stability. Additionally, RF efficiently manages high-dimensional data and variable interactions, provides clear importance rankings, and requires minimal parameter tuning.

The Random Forest algorithm constructs an ensemble of decision trees using bootstrap aggregation (bagging). Let (X_i, y_i) be the input-output pairs in the training dataset of size n. For each tree t, a bootstrap sample S_t is drawn from the dataset with replacement (Breiman, 2001).

Each tree is trained independently using S_t, and at each node, a random subset of predictors of size m is selected to determine the best split, adding decorrelation among trees. The final prediction $\hat{y}$ is the average (regression) or majority vote (classification) across all trees:

$$\hat{y}_i = \frac{1}{T}\sum t = 1^T h_t(X_i) \quad (15)$$

Where h_t is the prediction from tree t and T is the total number of trees.

RF provides variable importance by aggregating how much each variable contributes to reducing node impurity across all trees, often measured via Gini importance or mean decrease in accuracy (Breiman, 1996, 2001; Hastie et al., 2009).

GBM is particularly chosen for its capability in handling non-linear relationships and complex interactions via its iterative boosting process (Friedman, 2001; Hastie et al., 2009). This approach is robust against diverse data types, requiring minimal preprocessing. Its decision-tree-based structure effectively manages categorical variables and missing data. Unlike distance-based methods such as Support Vector Machines or K-Nearest Neighbors, GBM is unaffected by variable scales, and its threshold-based splits minimize the influence of outliers.

GBM functions by sequentially generating multiple weak decision trees, each tree aimed at reducing the errors of its predecessor. This sequential boosting process contrasts with RF's parallel bagging approach, where trees are built independently. GBM utilizes gradient descent optimization, guided by the gradient of the loss function with respect to predictions, iteratively minimizing errors. Initially, GBM makes preliminary predictions for observations, calculates residuals, and fits subsequent trees to these residuals, scaled by a learning rate (shrinkage parameter). The iteration continues until a specified number of iterations is reached or until further improvements become negligible.

Unlike RF, which builds trees in parallel, GBM constructs trees sequentially, each one correcting the residuals of the previous. Starting from an initial prediction $f_0(x)$, GBM updates the model iteratively (Friedman, 2001):

$$f_m(x) = f_{m-1}(x) + \eta \cdot h_m(x) \tag{16}$$

where $h_m(x)$ is a weak learner (typically a shallow tree) trained on the negative gradient of the loss function $L\big(y, f(x)\big)$, and η is the learning rate (shrinkage parameter).

At iteration m, the residuals are:

$$r_{im} = -\left[\frac{\partial L\big(y_i, f(x_i)\big)}{\partial f(x_i)}\right] f(x) = f_{m-1}(x) \tag{17}$$

where:

- y_i: Actual value for observation i
- $f(x_i)$: Model's prediction for observation i
- $L\big(y_i, f(x_i)\big)$: Loss function – e.g., squared error, log-loss
- $f_{m-1}(x)$: The prediction function up to iteration $m-1$
- r_{im}: The residual (or pseudo-residual) for observation i at iteration m

Note that this residual is not the usual difference $y_i - \widehat{y_i}$, but rather the negative gradient of the loss function at the current prediction. These residuals serve as the target outputs for the next decision tree in the boosting sequence, enabling the model to incrementally reduce the prediction error by learning from the most recent mistakes.

Both GBM and RF methods are advantageous as they yield feature importance scores, clearly identifying the most influential balance sheet items indicative of irregularities. These results provide targeted insights beneficial for auditing institutions and municipal administrators.

To train and evaluate both GBM and RF models, the dataset comprising between 200 to 600 observations is divided into training (80%) and testing (20%) subsets. The training subset further undergoes a 5-fold cross-validation process to reduce overfitting and ensure model reliability (Hastie et al., 2009; Kohavi, 1995). Each fold utilizes 80% of the training data for model construction and the remaining 20% for

validation, systematically rotating this process to ensure each observation contributes to both model training and evaluation.

Hyperparameter tuning for GBM is conducted through grid search, optimizing the following parameters:

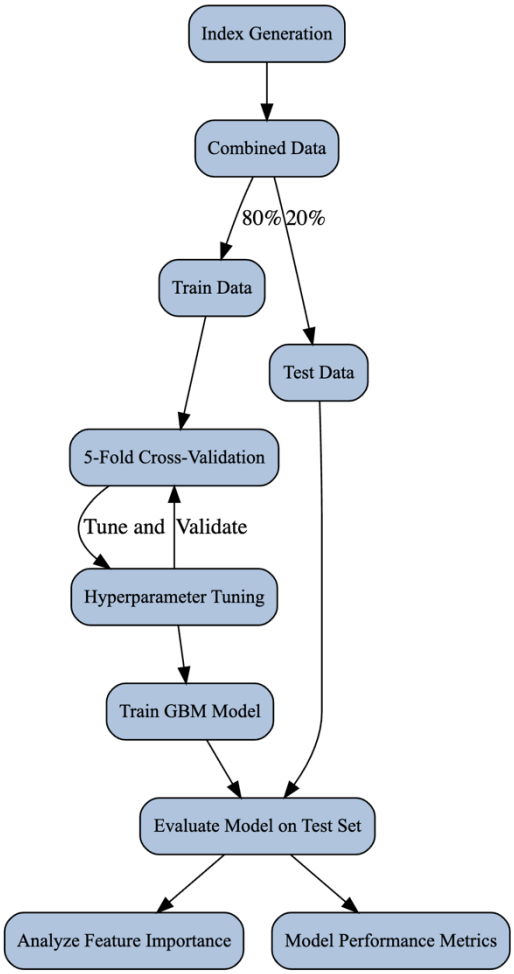
- Number of trees: number of boosting iterations (100, 500, 1000, 2000).
- Learning rate: step size for updating models (0.001, 0.01, 0.05, 0.1).
- Interaction depth: complexity of interactions in trees (2, 5, 8).
- Minimum observations in nodes: minimum required observations in terminal nodes (5, 10, 15).

For RF, hyperparameter tuning is also conducted via grid search, optimizing parameters including:

- Number of trees: number of trees in the forest (100, 500, 1000).
- Maximum features: number of variables randomly selected at each split ($\sqrt{p}$, $\log_2 p$, $p/3$, where p is the number of predictors).
- Minimum observations per leaf node: controlling node size (1, 5, 10).

Figure 15 illustrates a simplified structure of the GBM algorithm. In this simplified context presented in the figure, RF has the same flow.

Figure 15: GBM Flow Chart



The key difference between the GBM and RF models lies in the training algorithm. In the GBM procedure, trees are built sequentially (each one correcting the previous), with boosting, whereas RF trees are built independently, in parallel, using different data subsets (bagging). We should also underline here that RF necessitates complete data whereas GBM allows for unbalanced observations. As the data employed is unbalanced, RF loses a significant amount of data in this research.

Model performance is measured by predictions based on test data portion. To evaluate and compare indices, we employ two primary metrics: R^2, which assesses the proportion of variance in the irregularity index explained by the model, and root mean square error (RMSE), which measures prediction accuracy. RMSE values are particularly useful for comparing the performances of different models within the same index. Different indices have different distributions, including varying means and standard deviations. To compare the indices performances, we additionally report

RMSE values divided by standard deviation values of the indices. This can be interpreted as the distance of model's mean prediction errors measured in terms of standard deviation. Three metrics can be summarized as follows:

- R^2: indicates the proportion of variance in indices explained by the model.
- RMSE: measures the prediction error magnitude.
- RMSE/SD: normalizes the RMSE by dividing it by the standard deviation of the respective indices, providing standardized error metrics across indices.

Both GBM and RF provide feature importance scores, highlighting variables that significantly contribute to irregularity predictions. These critical variables are presented and analyzed for models demonstrating adequate predictive performance.

To illustrate the results, a selection of sample predictions is presented in the final part of the results section. While the predicted irregularity scores alone offer limited interpretive value, actionable insights can only be derived by examining the contributions of individual balance sheet titles to these scores. To this end, the Shapley Additive Explanations (SHAP) method is employed to quantify the marginal contribution of each feature to the model's prediction.

SHAP is a game-theoretic approach for interpreting the output of predictive machine learning models. It attributes the contribution of each feature to a given prediction by computing Shapley values. Originating from cooperative game theory, Shapley values address the question: "How much does each feature contribute to pushing the prediction away from the average prediction?" Specifically, the Shapley value of a feature represents its average marginal contribution across all possible combinations of feature subsets. This makes SHAP a theoretically grounded and consistent method for model explanation.

For a specific observation, SHAP decomposes the model prediction into the sum of a base value and individual feature contributions. This can be formally expressed as:

$$\hat{y}_i = \phi_0 + \sum_{j=1}^{p} \phi_j y_j \tag{18}$$

where $\hat{y}_i$ denotes the predicted value for municipality i, p is the total number of features, and $\phi_j \in \mathbb{R}$ represents SHAP value (feature attribution) for feature j. The binary variable $y_j \in \{0,1\}^p$ indicates whether a feature is present (1) or withheld (0) during the computation. The term ϕ_0 is the base value of the model, typically corresponding to the mean prediction across all training instances (S. M. Lundberg et al., 2018).

These contributions allow us to identify which financial indicators are most strongly associated with irregularity risk in each prediction.

5.5. Results

The input variables used in the machine learning models consist of balance-sheet items, major revenue heads (e.g., tax revenues, enterprise/ownership revenues, intergovernmental transfers), expenditure lines (e.g., personnel and social security, goods and services, capital outlays), and position accounts (e.g., short-term liabilities, fixed assets, cash/bank balances), together with operational expenditure components. For each of the five irregularity indices introduced earlier, RF and GBM are applied to identify patterns of irregularity that can be traced through these input variables. Model performance is evaluated using two principal metrics: the coefficient of determination (R^2), RMSE and the normalized RMSE/SD, which allows for comparability across indices with varying scales.

Prior to model training, the dataset is partitioned to enable robust performance evaluation. Given that the number of observations in this study constitutes a moderately sized dataset by machine learning standards, an 80/20 split between training and test sets is adopted. This allocation ensures that the models are trained on a sufficient volume of data while retaining an adequate portion for out-of-sample validation. Furthermore, a 5-fold cross-validation procedure is implemented on the training set to fine-tune hyperparameters and mitigate the risk of overfitting. This methodology strikes a balance between learning capacity and generalizability, thereby enhancing the reliability of the resulting performance metrics.

Machine learning models are inherently suited for pattern recognition and can uncover latent clusters within the data. As dataset size increases, these models tend to yield

more stable and accurate pattern detection. Consequently, the models are applied to the full dataset without segmenting municipalities into metropolitan and non-metropolitan categories. This unified approach also facilitates the identification of potential structural differences between municipality types, should such distinctions emerge from the model outputs.

The results of the machine learning models are presented in the subsequent subsections, following the order in which the indices were introduced. A comparative evaluation of model performance across indices is provided at the end of the section.

5.5.1. Case Count Index Results

Case count index is the simplest approach of indices, with only covering the basic count of cases handled in audit reports. Since the cases cover several issues with divergent significance and complexity levels, we can expect a low performance on predictive power. Results justify this expectation with poor R^2 performances below or around 0.5 as shown in Table 24. Given this weak results, further analysis such as prediction performance graphs, variable relevance analysis and sample predictions based on these models are skipped.

Table 24: Results for Case Count Index

Model	Variable Set	R^2	RMSE/SD
GBM	All Vars	0.38	0.54
	Top 5 Vars	0.53	0.38
RF	All Vars	0.52	0.38
	Top 5 Vars	0.28	0.62

It is expected for this index to better perform in a non-complex subgroup of observations. (To give insights about the predictive power of the model) Only aiming to justify this claim, GBM model is run separately over metropolitan and non-metropolitan subgroups. A weaker performance should be observed for metropolitan municipalities where both the inputs and irregularities have complex patterns. Results presented in Table 25 confirm this framework and support the simplified character of the index.

Table 25: Metropolitan vs Non-metropolitan Comparison for Case Count Index

R^2 Values for Case Count by Municipality Type		
Target Variable	Metropolitan	Non-Metropolitan
Case Count	0.31	0.71

These results direct the research towards more complex index trials. Case count as a basic index just covers the simple structure and is not able to capture the patterns between budget values and irregularities.

5.5.2. Indicator Count Index Results

Second index is based on equal weighted accumulation of predefined irregularity indicator key words. Our claim here is as the appearance of the signals increases, severity of the irregularities follows. Table 26 gives the results for both of the ML models. Case count index does not give satisfying predictive power, too. A slight improvement is achieved yet far from enough.

Table 26: Indicator Count Index Results

Model	Variable Set	R^2	RMSE/SD
GBM	All Vars	0.52	0.44
	Top 5 Vars	0.52	0.46
RF	All Vars	0.57	0.41
	Top 5 Vars	0.35	0.58

By focusing on indicators this index captures the irregularities quantitatively. Case count and indicator count are based on cumulative approaches which in turn together lack the qualitative refinement. More precisely, they are blind to strengths of the cases addressed in the reports. This guides the study to an index that can distinguish the strengths of the cases reported.

5.5.3. Weighted Indicator Count Index Results

The core idea of this index, which is an updated version of the indicator count, is the signaling power of the indicators should be different. Thus, indicator terms should be weighted accordingly for a better predictive power. The results of the ML models

present an improvement with R^2 values reaching to 0.66. Also, R^2 values are more consistent along the two models.

Table 27: Weighted Indicator Count Index Results

Model	Variable Set	R^2	RMSE/SD
GBM	All Vars	0.61	0.36
	Top 5 Vars	0.66	0.34
RF	All Vars	0.59	0.39
	Top 5 Vars	0.64	0.40

Besides the improvement, these R^2 values are still can be improved and moreover, this index misses the context of the cases reported. A further refinement can be context capturing algorithms.

5.5.4. Weighted Category-based Case Count Index Results

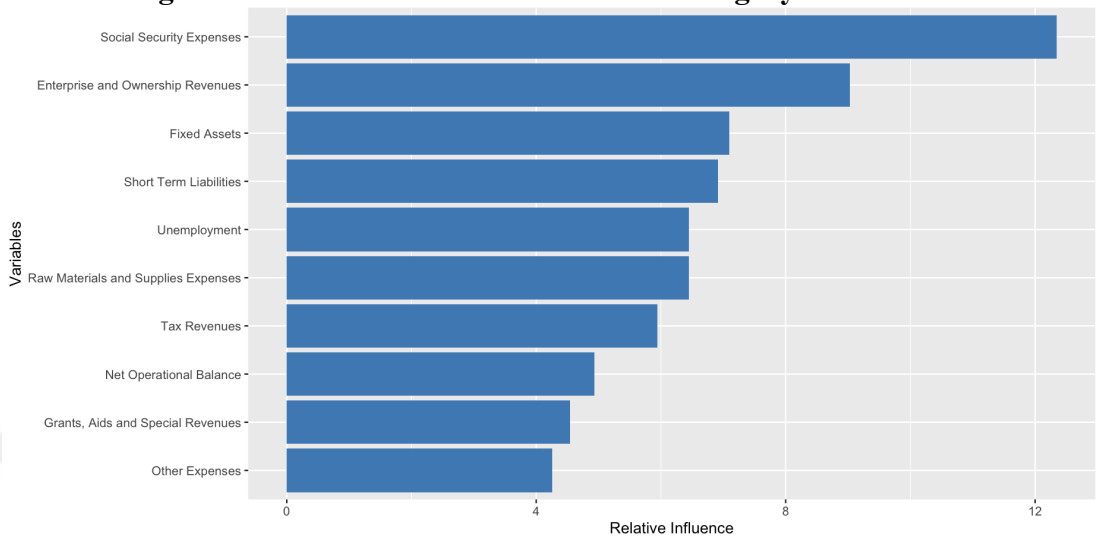
As mentioned earlier, Sayıştay reports the detected irregularities of the municipalities under ordered titles of cases. Each case reported is a detected specific irregularity which also can be categorized under titles. More specifically, cases can be categorized under titles such as: Financial structure, auction practices, goods and services etc. These titles are explained and reported in detail in the data part.

Categorization of the cases helps improve the index by adding context. Case categories can be claimed to have different impacts on expenditure schemes and balance sheet values. For the first approach regarding this classification of the cases, predefined categories are introduced, and a weight grid search is applied to capture most contributing titles.

This index returns an R^2 value of 0.74 and an RMSE/SD ratio of 0.41. The RMSE/SD ratio for case weighted index is 0.41, indicating that the model's predictive error is less than half of the natural variability of the index. This, together with an R^2 of 0.74, demonstrates that the model provides highly accurate predictions and effectively explains the underlying variation in the case category based weighted scores. Since the model returns robust results, detailed analysis and interpretations based on this index can be applied.

ML models report the relative influences of the predictor variables. GBM model returns the relative influences order as presented in Figure 16.

Figure 16: Relative Influences for Case Category Based Index



The large and complex nature of social security outlays makes them especially vulnerable to misreporting, misallocation, or procedural errors. This complexity increases the risk of audit findings, especially in municipalities with limited administrative capacity. Besides these difficulties, social security expenditures are politically salient and may be subject to manipulation or exceptional spending in response to local electoral dynamics or social pressure.

Enterprise and ownership revenues refer to the income municipalities generate from their business activities, municipal subsidiaries, and the management of public assets. In the Turkish context, these revenue streams have grown in importance as local governments have expanded their economic activities and diversified their financial bases. These revenues often provide municipalities with greater fiscal autonomy, but also introduce additional layers of complexity and managerial challenge that can increase audit risk. Municipal subsidiaries and asset management are frequently cited in audit reports for issues related to transparency, valuation, and compliance with procurement and public finance laws. The pursuit of additional revenue through municipal enterprises may create incentives or opportunities for off-budget transactions, creative accounting, or conflicts of interest.

The relative influence analysis reveals that unemployment is a non-negligible predictor of audit irregularity risk in Turkish municipalities. This finding aligns with

the economic rationale on public sector risk, where economic distress is associated with increased pressure on local administrations to deliver services or generate employment, sometimes through non-compliant or irregular practices. In the Turkish context, persistent regional unemployment disparities may amplify these pressures, suggesting that economic development and job creation are aligned with local governance quality and transparency.

For the weight assignment process, a grid search is applied. Table 28 presents the mean weights assigned to each audit case category by the ten top-performing GBM models, reflecting the relative contribution of each category to predictive model accuracy. These weights offer insight into which types of irregularities are most predictive of overall risk or irregularity scores. The weight distribution confirms that core financial management, asset oversight, and procurement processes are the dominant sources of predictive information in audit-based risk modeling. This suggests that targeted improvements in these areas are likely to yield the greatest reduction in municipal audit irregularities.

Table 28: Weight Assignments of Top 10 Performing Models

Case Category	Mean Weight
Financial Structure	0.30
Auction Practices	0.16
Assets Management	0.23
Internal Control Mechanisms	0.08
Personnel Issues	0.12
Service Purchases	0.05
Others	0.05

This index with introducing context to the text conversion significantly improves the model performances. However, the major drawback of this index is the degree of supervision. The categories are predefined based on human reading samples. Next index is generated through the exact same algorithms except category extraction procedures, which are also transferred to computation thus gains reproducibility and robustness.

5.5.5. LDA Supported Index

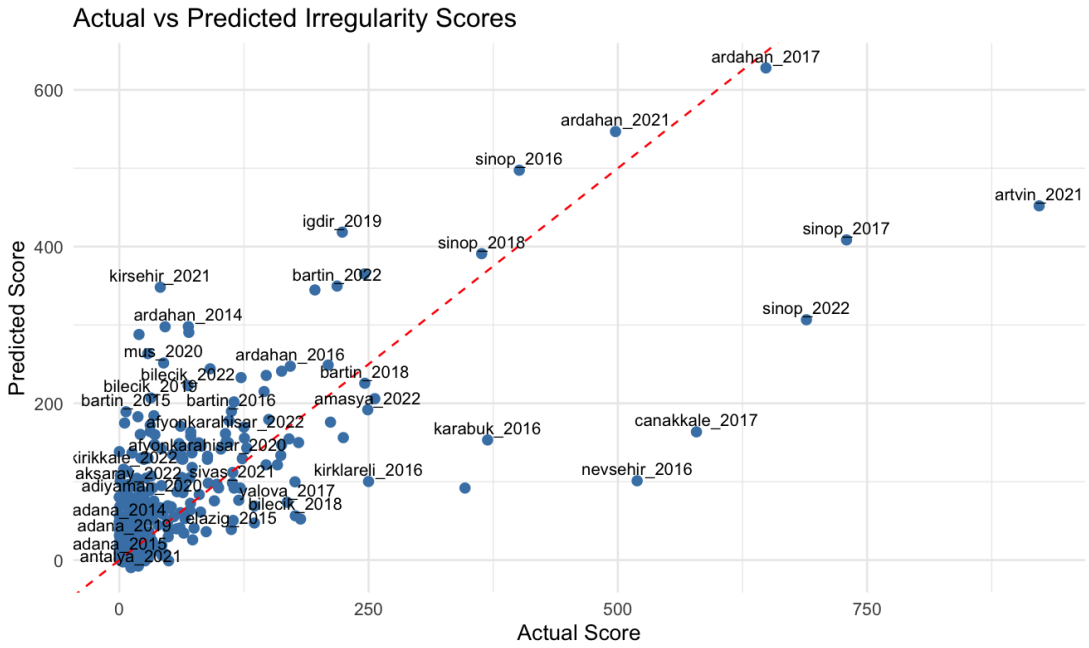
As a final refinement towards an unsupervised index, LDA is employed. Instead of predefined case categories, we let LDA determine categories. Details of the index can be found on data section. LDA supported index improves ML results with R^2 values 0.80 for both RF and GBM models (Table 29 and Figure 17). GBM model exhibits a lower RMSE/SD ratio indicating its predictions are closer to the observed values relative to the variability of the index.

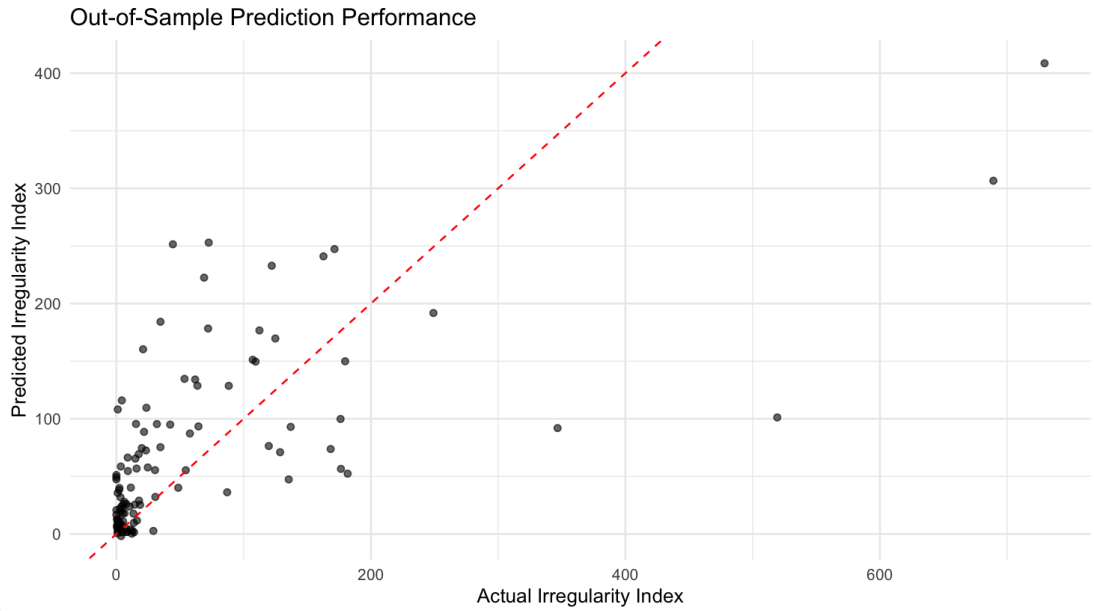
Table 29: LDA Index Results

7 Category Setup

Metric	GBM	RF
Mean Index	53.34	84.75
Standard Deviation of Index	100.97	165.39
R-squared	0.80	0.80
RMSE	48.87	147.95
RMSE/SD	0.58	0.88

Figure 17: LDA Index Prediction Performance





Presented in Table 30, analysis of weight distributions across categories reveals that "Accounting and Asset Management" and "Budgeting and Financial Oversight" are major titles for both of the models with slight differences. "Real Estate and Related Incomes" and "Taxation and Revenue Collection" are the titles that are inconsistent among models with RF giving more weights to these titles.

Table 30: Best Performing Weights for LDA Models

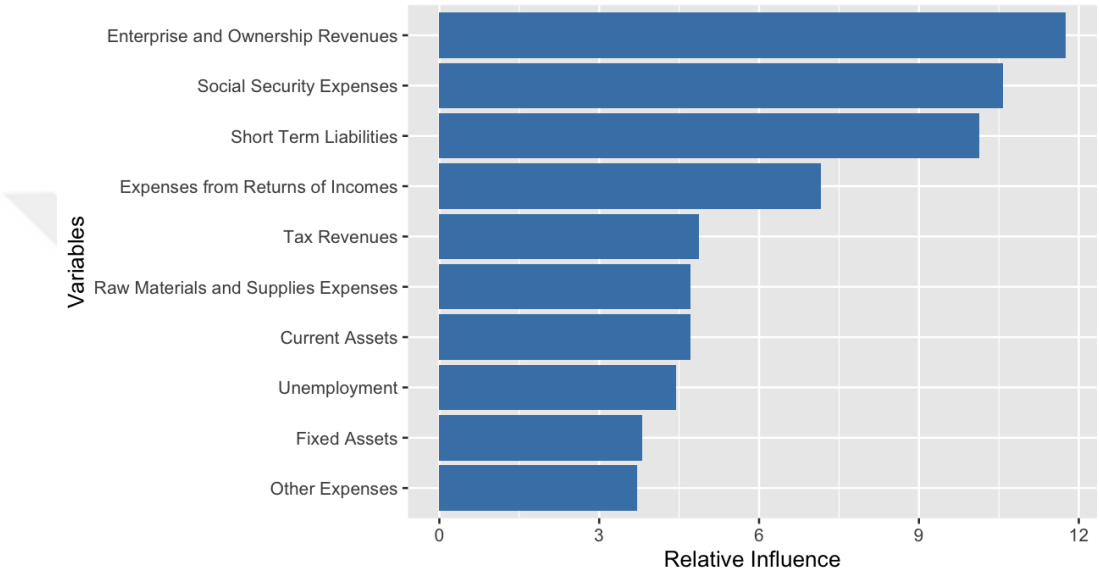
Mean Weights Comparison for LDA Models
GBM vs Random Forest

Category Name	GBM	RF
Accounting and Asset Management	0.37	0.27
Public Oversight and Social Services	0.13	0.08
Urban Planning and Transportation	0.04	0.04
Real Estate and Related Incomes	0.07	0.19
Budgeting and Financial Oversight	0.25	0.15
Taxation and Revenue Collection	0.04	0.20
Procurement and Public Contracts	0.10	0.07

Regarding explanatory powers of variables, GBM model reports "Enterprise and Ownership Revenues", "Social Security Expenses" and "Short Term Liabilities" as the

leading determinants of the model as illustrated on Figure 18. Enterprises can be considered as non-opaque and complex operations of municipalities, increasing the potential of irregularities as the model captures. Ownership revenues and personnel regimes (social security expenses) are among the leading titles of corruption and nepotism practices. "Unemployment" as a significant social and economic indicator as discussed before, is again among the top influencers.

Figure 18: Relative Influences - LDA – GBM



There are two types of relative influence analysis in RF models. Relative influence by node purity measures the total reduction in impurity. It reflects the cumulative usefulness of a variable in partitioning the data for more accurate prediction. In contrast, relative influence by mean increase in mean square error (MSE) measures the average rise in prediction error when a given variable is muted. This metric more directly measures the marginal contribution of each variable to overall accuracy.

"Tax Revenues" is the leading determinant for both of the relative influence approaches in RF model (Figure 19 and Figure 20). "Metropolitan" dummy also plays a key role on predicting irregularities based on this index.

Figure 19: Relative Influences LDA - RF Model – MSE

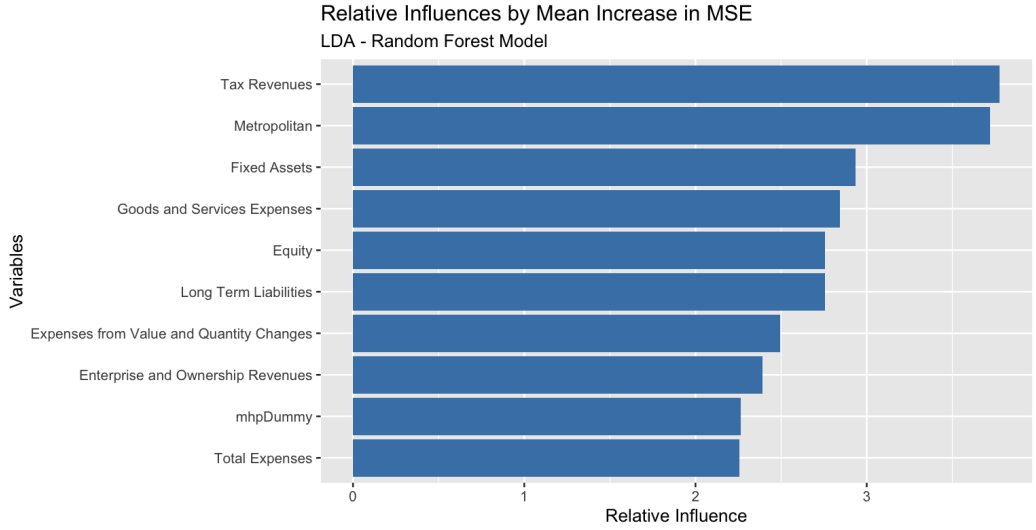
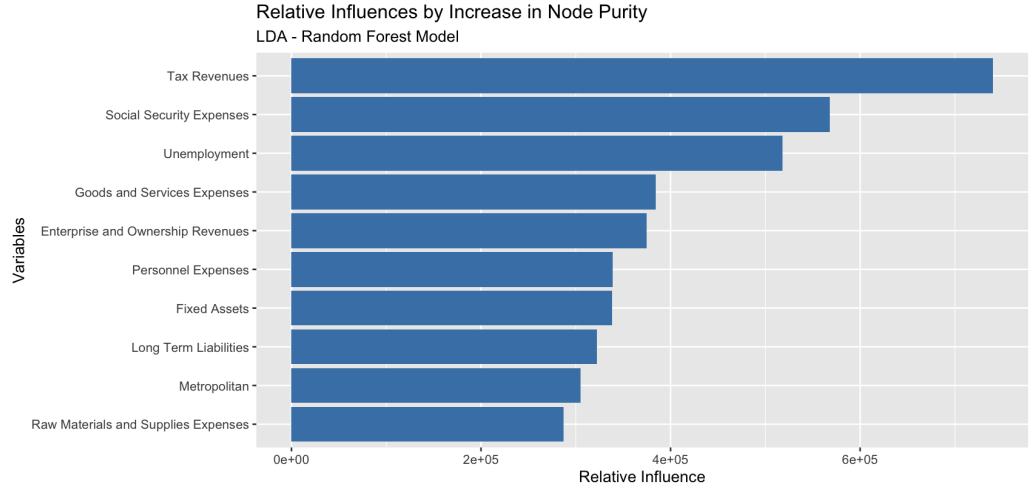


Figure 20: Relative Influences LDA - RF Model - Node Purity



Regarding node purity results, all the leading three variables are linked with socio-economic conditions and scale of the municipality. The influence order reveals that, these three variables are the key clusters for subbranches. Metropolitan status and "Fixed Assets" are more highly ranked in MSE metric, likely because their absence substantially degrades model performance whereas they are not as much useful in splitting. Conversely, "Social Security Expenses" and "Unemployment" feature more dominantly in node purity than MSE metric. These differences highlight the strengths of the two metrics: one reflects predictive power, the other signals structural relevance.

To get further insights on relative influence results, a correlation matrix is presented in Figure 21. Along with the top three variables, fundamentals such as municipal population, municipal type and GDP per capita are added. Remind that, "Tax

Revenues" and "Social Security Expenses" are per person values. The significant negative correlation between tax revenues and metropolitan status and municipal population may partly explain the high predictive power of both variables in the RF model, implying structural divergences can be traced via tax revenues. Similarly, the correlation between unemployment and GDP per capita further supports structural predictive power of unemployment.

Figure 21: Correlation Matrix of Node Purity Variables with Fundamentals



When compared with the GBM model, there is substantial consistency among the variables "Tax Revenues", "Enterprise and Ownership Revenues", and "Social Security Expenses" between the two models. However, RF model tends to assign higher importance to municipality type and broader expenditure categories, while the GBM model underlines more specific revenue streams and liabilities. This divergence likely reflects methodological differences: GBM, with its sequential boosting approach, hones in on the most powerful predictors, whereas Random Forest captures a broader array of variables whose effects may be more distributed. These results reinforce the value of using multiple interpretability metrics and modeling strategies to obtain a comprehensive understanding of risk factors in municipal audit outcomes.

5.5.6. Comparative Assessment of Index Performance and Predictive Modeling

A comparative analysis of the five irregularity indices, ranging from the simplest case count to the most sophisticated LDA-supported index, demonstrates a clear trajectory

of methodological refinement and predictive gain. The case count index, rooted in the plain count of cases reported, yields limited predictive power, as reflected by R^2 values consistently below 0.5 for both GBM and RF models. The indicator count index, which incorporates key-word-based signals, offers only a marginal improvement, highlighting the limitations of purely quantitative accumulation in capturing the complexity of municipal audit irregularities.

Advancing this framework, the weighted indicator count index introduces a differentiated approach by assigning variable importance to indicator terms. This strategy delivers a moderate increase in predictive performance, with R^2 values around 0.66, suggesting that recognizing heterogeneity in irregularity signals enhances model accuracy but remains insufficient for robust prediction.

Substantial improvements arise with the case category weighted index, which leverages predefined case categories and optimized weights via grid search. The inclusion of qualitative context and category-specific weighting elevates the models' explanatory power, yielding R^2 values of 0.74 and RMSE/SD ratios as low as 0.41. This indicates that the models capture a significant proportion of the true variance in irregularity scores. Notably, the most influential predictors in this framework pertain to financial management, asset oversight, and procurement practices, affirming the centrality of fiscal governance in audit risk.

The highest level of predictive performance is achieved by the LDA supported index, which replaces subjective category definitions with data-driven topic modeling. Both GBM and RF models attain R^2 values of 0.80, and the GBM exhibits a particularly low normalized error, confirming the value of unsupervised, replicable category assignment. The LDA supported index not only enhances model accuracy but also enriches interpretability by revealing the latent thematic structure of audit findings. Variable importance analyses across both modeling approaches consistently highlight the roles of tax revenues, enterprise and ownership revenues, social security expenses, and indicators of municipal scale and economic context.

In summary, the progression from basic counts to unsupervised, context-aware indices is mirrored in a monotonic improvement in predictive accuracy. The findings suggest that robust, contextually nuanced, and data-driven representations of irregularities,

combined with state-of-the-art machine learning, can substantially enhance the capacity to monitor and predict municipal audit risk. Methodologically, the results highlight the importance of both variable selection and the structural design of indices, advocating for the use of advanced text analysis and ensemble modeling to inform public sector risk analytics.

5.6. Selected Municipal Case Studies

To better illustrate the practical contributions of this research and demonstrate the operational applicability of the proposed methodology, a couple of municipalities were randomly selected for in-depth analysis.

Ardahan. 2021 serves as the first example. Despite its relatively small population and economic size, the city had a total of 33 irregularities reported in the 2021 audit. Its LDA-based irregularity index is 497.67, while the GBM model predicted a higher value of 546.72. Consistent with the SHAP analysis in Table 31, 9 of 33 cases reported are directly related with the Enterprise and Ownership Revenues. Additionally, Social Security Expenses and Tax Revenues are among the top contributors, aligning with the themes observed in the actual audit report.[9].

[9] https://www.sayistay.gov.tr/reports/download/wPoAKXLQey-ardahan-belediyesi

Table 31: Ardahan 2021 - SHAP Analysis

Top Positive SHAP Contributors for Ardahan 2021 Predicted Index: 546.72 | Actual Index: 497.67 | Difference: 49.06	
Balance Sheet Title	SHAP Contribution
Enterprise and Ownership Revenues	92.39
Social Security Expenses	67.24
Tax Revenues	66.78
Short Term Liabilities	37.74
Expenses from Returns of Incomes	36.04
Current Assets	28.95
Personnel Expenses	27.94
Goods and Services Expenses	25.68
GDP per Capita	19.54
Grants, Aids and Special Revenues	19.36
Raw Materials and Supplies Expenses	19.07
Interest Expenses	15.74
Metropolitan	15.48
Other Expenses	13.25

The second example, Kayseri 2022, is presented in Table 32. Predicted and actual indices in this particular case are 24.4 and 17.27 respectively. According to the SHAP results, Enterprise and Ownership Revenues again emerges as the most significant contributor. A closer inspection of the audit report confirms this finding: nearly half of the reported irregularities pertain to ownership-related practices, affirming the model's predictive capacity.[10]

[10] https://www.sayistay.gov.tr/reports/download/X6QLAZqYyp-kayseri-buyuksehir-belediyesi

Table 32: Kayseri 2022 - SHAP Analysis

Top Positive SHAP Contributors for Kayseri 2022

Predicted Index: 24.4 | Actual Index: 17.27 | Difference: 7.12

Balance Sheet Title	SHAP Contribution
Enterprise and Ownership Revenues	37.09
Grants, Aids and Special Revenues	6.91
Short Term Liabilities	4.80
Other Revenues	3.78
Personnel Expenses	2.83
Equity	2.11
Total Expenses	1.70
Capital Transfers	1.64
Net Operational Balance	0.69
Net Migration	0.37
Depreciation Expenses	0.35
Expenses from Value and Quantity Changes	0.10
Off-balance Sheet Accounts	0.02

Further verification is provided in Table 34 through comparison with the corresponding audit report. The primary issues reported[11] for Istanbul in 2022 align with three of the top four categories identified by the SHAP algorithm, excluding Short-Term Liabilities.

[11] https://www.sayistay.gov.tr/reports/download/DJYy3lpgwp-istanbul-buyuksehir-belediyesi

Table 33: Istanbul 2022 - SHAP Analysis

Top Positive SHAP Contributors for Istanbul 2022 Predicted Index: 20.14 | Actual Index: 16.41 | Difference: 3.72	
Balance Sheet Title	SHAP Contribution
Enterprise and Ownership Revenues	23.18
Short Term Liabilities	11.06
Goods and Services Expenses	7.44
Personnel Expenses	7.07
Current Assets	5.89
Other Revenues	3.91
Capital Transfers	1.93
Equity	1.62
Net Operational Balance	1.38
Net Migration	0.92
Combined Personnel Expenses	0.84
Total Expenses	0.47
Expenses from Value and Quantity Changes	0.36
Depreciation Expenses	0.25

To further illustrate the model's effectiveness, the following examples incorporate LDA-derived categories identified in the audit reports. The case of İzmir (2020) demonstrates a strong alignment between the leading SHAP contributors from the machine learning model and the dominant LDA topics. Notably, the model successfully identified issues related to the personnel regime as the primary area of concern (Table 34 and Table 35).

Table 34: Izmir 2020 - SHAP Analysis

Top Positive SHAP Contributors for Izmir 2020 Predicted Index: 21.9 | Actual Index: 13.97 | Difference: 7.93	
Balance Sheet Title	SHAP Contribution
Personnel Expenses	15.39
Off-balance Sheet Accounts	7.57
Combined Personnel Expenses	3.77
Other Revenues	1.36
Equity	0.78
Net Operational Balance	0.38
Short Term Liabilities	0.19
Total Expenses	0.12
Depreciation Expenses	0.10
Net Migration	0.07

Table 35: Izmir 2020 - LDA Topic Distribution

LDA-Derived Topic Distribution for Izmir 2020	
LDA Topic Title	Relative Share
Procurement and Public Contracts	30.8%
Real Estate and Related Incomes	26.9%
Taxation and Revenue Collection	19.2%
Public Oversight and Social Services	11.5%
Urban Planning and Transportation	7.7%
Accounting and Asset Management	3.8%

In the case of Balıkesir (2021), the SHAP analysis reveals a balanced distribution among the predictor variables, as presented in Table 36. Similarly, the LDA topic distribution in Table 37 aligns with the model's results, also reflecting a balanced representation across the identified categories.

Table 36: Balıkesir 2021 - SHAP Analysis

Top Positive SHAP Contributors for Balikesir 2021 Predicted Index: 13.61 | Actual Index: 10.22 | Difference: 3.39	
Balance Sheet Title	SHAP Contribution
Grants, Aids and Special Revenues	5.86
Personnel Expenses	5.53
Other Revenues	3.78
Depreciation Expenses	1.69
Equity	0.90
Net Migration	0.22
Combined Personnel Expenses	0.13
Net Operational Balance	0.10
Expenses from Value and Quantity Changes	0.09
Total Revenues	0.05
Current Transfers	0.05

Table 37: Balıkesir 2021 - LDA Topic Distribution

LDA-Derived Topic Distribution for Balikesir 2021	
LDA Topic Title	Relative Share
Procurement and Public Contracts	29.4%
Urban Planning and Transportation	23.5%
Accounting and Asset Management	17.6%
Real Estate and Related Incomes	11.8%
Taxation and Revenue Collection	11.8%
Public Oversight and Social Services	5.9%

These randomly selected examples help clarify the practical applicability of the methodology developed in this research. Based on SHAP results, audit institutions can predefine high-risk areas for more targeted examination. Similarly, municipal administrations can proactively identify and address embedded issues using available balance sheet data, even in the absence of a formal audit.

To sum up, up to a high level of accuracy, the models developed in this study demonstrate that irregularities can indeed be predicted from municipal financial statements. Nevertheless, it is important to emphasize that these results are not definitive, and the algorithms employed are not strict decision-making tools. Rather, they serve as supportive mechanisms that highlight potential problem areas, albeit with a margin of error. As such, they are not intended to replace traditional audit functions but to enhance their efficiency in two key dimensions: cost-effectiveness and improved targeting. These models can facilitate the development of cross-check mechanisms, enabling auditors to more precisely identify areas that warrant further scrutiny. This, in turn, reduces the time and resources required for comprehensive evaluations of public institutions. Moreover, by directing attention toward high-risk areas, this predefined focus can increase the likelihood of detecting irregularities.

6. BRIDGING EFFICIENCY AND IRREGULARITIES: LINKING DEA SCORES WITH AUDIT-BASED INDICES

6.1. Introduction

This section bridges two core perspectives in the evaluation of Turkish municipal governance: production-side performance, captured by DEA efficiency scores, and governance-risk diagnostics, captured by audit-based irregularity indices. Theoretically, several mechanisms suggest a negative association between irregularities and technical efficiency. In a principal–agent framework, weak monitoring and low accountability encourage slack behavior and rent-seeking, diverting effort away from productive service provision (Alesina & Tabellini, 2008; Rose-Ackerman, 1999). These tendencies often manifest in the expenditure mix, where resources are steered toward budget categories with higher discretion, such as capital projects or personnel spending, rather than toward core service delivery. Such shifts crowd out productive uses of funds and weaken the efficiency of municipal outputs. The presence of soft budget constraints reinforces this dynamic: expectations of bailouts or lenient fiscal oversight reduce the pressure to minimize costs, while opportunistic procurement practices and weak asset management undermine scale economies, raising unit costs for given service bundles. At the same time, the relationship is not uniform across all municipalities. Larger metropolitan administrations, with their broader mandates, larger budgets, and more complex service portfolios, face more opportunities for irregular practices and higher managerial challenges. Population size and structural complexity may thus mediate the link between irregularities and efficiency, shaping how these dynamics unfold in practice.

From this logic follow three testable hypotheses:

- Municipalities with higher irregularity burdens tend to exhibit lower technical efficiency;
- The association is topic-specific and strongest where fiscal discretion is concentrated (budgeting and financial oversight; procurement and public contracts; accounting and asset management);
- Scale and complexity condition the relationship, with metropolitan status and population potentially mediating both irregularities and efficiency.

At the same time, countervailing considerations caution against expecting a uniform, strong correlation. First, measurement scope differs: DEA efficiency in this study is constructed on a focused set of service outputs, whereas the irregularity indices aggregate findings across all audited functions; some irregularities may be orthogonal to the production frontier under evaluation. Second, reporting intensity and audit salience may vary with size or fiscal capacity, creating spurious associations (e.g., richer or larger municipalities both produce more services and attract deeper audits). Third, reverse causality is possible if inefficient service delivery invites more intense scrutiny. Accordingly, the empirical strategy proceeds in tiers. We begin with descriptive Pearson correlations between efficiency scores and each irregularity index. We then estimate pooled OLS regressions that relate efficiency to overall and topic-level irregularity measures, controlling for log population, GDP per capita, and unemployment to absorb observable scale and socio-economic factors. Finally, we exploit the panel structure by estimating municipality fixed-effects models with year dummies and clustered standard errors, thereby purging time-invariant heterogeneity and common shocks while focusing on within-municipality co-movements over time. This design allows a direct assessment of hypothesis and, by disaggregating irregularities into LDA-derived topics, isolates the governance domains most plausibly connected to technical efficiency.

6.2. Literature Review

Early theoretical work established why weaknesses in institutions and incentives can translate into both allocative inefficiency and irregular practices. Classic principal–agent and rent-seeking formulations emphasize how discretionary authority, weak monitoring, and information asymmetries distort input choices and effort provision

(Alesina & Tabellini, 2008; Rose-Ackerman, 1999). Complementary frameworks argue that bribery and influence reallocate resources toward more "extractable" or opaque activities, raising the effective tax on production and depressing output (Klitgaard, 1988; Shleifer & Vishny, 1993). A large cross-country literature then documented macroeconomic consequences: higher corruption is associated with lower investment and growth (Mauro, 1995) and with a shift in the composition of public spending toward capital-intensive, rent-rich items and away from operations and maintenance (Tanzi & Davoodi, 1997). These channels provide clear theoretical routes through which irregularities can degrade technical efficiency at the production frontier.

Beginning in the 2000s, micro-evidence clarified mechanisms linking oversight, leakage, and performance. Reinikka & Svensson (2004) showed how greater transparency about intergovernmental transfers sharply reduced capture in Uganda's education spending. Olken (2007) exploited randomized audit intensity to demonstrate that external monitoring significantly curtailed missing expenditures in Indonesian road projects. Related work distinguished "active" (corruption) from "passive" (organizational) waste in procurement, showing both can be large and policy-relevant (Bandiera et al., 2009). In the U.S. context, Glaeser & Saks (2006) documented substantial subnational variation in corruption and linked it to lower income growth. Synthesizing this evidence, (Olken & Pande (2012) reviewed the causal literature and highlighted monitoring, media, and institutional design as central levers for reducing corruption without sacrificing service delivery.

A subsequent strand focused on audits and procurement design as operational tools to improve performance. Duflo et al. (2013) showed how contract design for third-party auditors affects truth-telling and compliance in environmental regulation. Avis et al. (2018) found that public release of audit results reduced corruption and improved electoral accountability in Brazilian municipalities. On the procurement side, Lewis-Faupel et al. (2016) provided evidence that e-procurement can speed delivery and improve match quality in infrastructure projects. A comprehensive overview by (Bosio et al. (2022) links procurement rules to competition and value-for-money, detailing how institutional choices shape efficiency risks.

Turning to local-government efficiency, empirical work connects institutional quality to performance at the subnational level. Del Monte & Papagni (2001) show that

corruption lowers long-run economic performance via resource misallocation. Geys et al. (2010) examine Flemish municipalities and associate inefficiency with lower transparency and accountability. Using DEA for Romanian counties, Bostan et al. (2018) report a negative correlation between efficiency scores and audit-reported irregularities, suggesting that managerial slack and misuse often co-occur. For Brazil, M. da C. S. De Sousa et al. (2005) document that weaker transparency and budgetary oversight—conditions conducive to irregularities—are associated with lower local efficiency. Relatedly, Finocchiaro Castro et al. (2025) (using SFA for Italian waste services) find that weaker control of corruption is linked to higher technical inefficiency, underscoring the governance–production connection at the service-line level. Taken together with the audit and procurement literature above, these contributions imply that the efficiency–irregularity nexus should be most visible in domains with high discretion (budget execution, accounting/asset management, procurement), and potentially mediated by scale and complexity.

This cumulative evidence motivates the empirical design in this chapter: test whether municipal DEA efficiency co-varies with audit-based irregularity measures overall, and whether topic-specific irregularities, especially in budgeting/financial oversight, procurement/public contracts, and accounting/asset management, exhibit stronger associations after accounting for size, socio-economic conditions, and time-invariant heterogeneity.

6.3. Data

Given that the LDA-derived topics capture distinct thematic areas of irregularities, it is reasonable to expect that not all topics will be equally associated with technical inefficiency. To further explore this relationship, the irregularity scores are disaggregated by LDA topics. As a reminder, the LDA analysis identified seven thematic titles in the audit reports, and each finding in a given report was assigned to one of these titles. Consequently, for each municipality-year observation, seven separate irregularity scores are obtained, one for each LDA topic.

To quantify irregularities by governance area, we construct four topic-specific indices for each municipality–year, using the seven LDA-derived topics ($\text{Topic}_1, \ldots, \text{Topic}_7$). Each index yields a seven-element vector per observation; topic–years with no observed finding are coded as zero.

1) LDA Dummies (binary presence by topic).

An indicator equal to 1 if at least one audit finding in year t for municipality i is assigned to topic k, and 0 otherwise:

$$D_{itk} \in \{0,1\} \tag{19}$$

This captures the existence of irregularities within each topic.

2) LDA Counts (case counts by topic).

The number of distinct audit findings assigned to topic k:

$$\text{C}_{\text{itk}} = \#\{\text{findings in topic } k\} \tag{20}$$

This measures the frequency/breadth of irregularities.

3) Plain Counts by Topic (unweighted indicator totals).

For each topic, the within-topic sum of unweighted indicators attached to findings:

$$\text{P}_{\text{itk}} = \sum_{\text{f}\in(\text{i,t,k})} \text{indicator}_{\text{f}} \tag{21}$$

This aggregates intensity signals without applying severity weights.

4) Weighted Counts by Topic (severity-weighted totals).

For each topic, the within-topic sum of indicators after applying pre-specified severity weights:

$$W_{itk} = \sum_{f \in (i,t,k)} weight_f \times indicator_f \quad (22)$$

This reflects severity-adjusted intensity of irregularities.

We should note that, for scaling purposes, indices are adjusted to per 10,000 people and efficiency scores are adjusted to per million people.

6.4. Methodology

To investigate the association between municipal efficiency and irregularity patterns identified in audit reports, we employed two complementary regression strategies: pooled Ordinary Least Squares (OLS) and two-way Fixed Effects (FE) panel models. The dependent variable in all regressions is the technical efficiency scores from the models of the Data Envelopment Analysis (DEA).

The pooled OLS specification serves as a baseline, exploiting both the cross-sectional and temporal variation in the dataset without controlling for unobserved heterogeneity across municipalities. The FE model extends this approach by accounting for time-invariant unobserved heterogeneity at the municipality level, thereby controlling for structural characteristics such as geographic location, governance traditions, or long-standing infrastructural differences that might otherwise bias coefficient estimates.

Linear Panel Regression Model

Let i index municipalities and t years. For each outcome $y_{it} \in \{\text{Efficiency Services}, \text{Efficiency Treatment}\}$, we estimate

$$y_{it} = \beta_0 + \sum_{k=1}^{7} \beta_k \, T_{itk} + \beta_8 \ln(\text{pop}_{it}) + \beta_9 \, \text{GDPpc}_{it} + \beta_{10} \, \text{Metro}_i + \varepsilon_{it} \quad (23)$$

where:

- T_{itk} is the intensity of Topic k in municipality i and year t (the seven topics are: Accounting and Asset Management; Budgeting and Financial Oversight; Procurement and Public Contracts; Real Estate and Related

Incomes; Taxation and Revenue Collection; Urban Planning and Transportation; Public Oversight and Social Services). This intensity is calculated in four different algorithms that are explained in the data section (LDA dummies, LDA counts, plain counts, weighted counts).

- $\ln(\text{pop}_{it})$ is the log population.
- GDPpc_{it} is GDP per capita measured in thousands (we divide raw GDP per capita by 1,000 to stabilize magnitudes).
- Metro_i is a time-invariant metropolitan dummy (1 if metropolitan).
- ε_{it} is an idiosyncratic error.

Note that both y_{it} and T_{itk} are normalized by population (e.g., per 100,000 residents); coefficients then represent effects in those units.

Estimation is by OLS on the pooled panel, no fixed effects are included in this model. Standard errors are reported in the tables.

Each β_k measures the marginal association between topic k and the outcome, holding the other six topics and controls constant. β_k is the change in the outcome per unit change in topic intensity per 100,000 population.

While pooled OLS provides a useful baseline, it does not control for unobserved heterogeneity across municipalities that may bias the estimated relationships. For instance, structural characteristics such as geographic location, long-standing governance practices, or socio-economic composition may simultaneously influence both efficiency and irregularity levels. To address this issue, we employ a Fixed Effects (FE) panel data specification. The FE estimator controls for all time-invariant, municipality-specific factors by focusing exclusively on within-municipality variation over time.

Fixed Effects Panel Model

To absorb time-invariant municipal characteristics and common shocks, we estimate a two-way FE model:

$$y_{it} = \alpha_i + \lambda_t + \sum_{k=1}^{7} \gamma_k \, T_{itk} + \delta_1 \log(\text{pop}_{it}) + \delta_2 \, \text{GDPpc}_{it} + u_{it} \tag{24}$$

with:

- α_i is the municipality fixed effect,
- λ_t is the year fixed effect,
- u_{it} is the idiosyncratic error term.

Because Metro_i is time-invariant, it is absorbed by α_i and not estimated separately in (24). We report heteroskedasticity-robust standard errors clustered at the municipality level.

Each γ_k is a within-municipality estimate: it captures how changes over time in topic k for a given municipality relate to changes in the outcome, net of the other six topics, controls, and year dummies.

6.5. Results

We begin by examining the Pearson correlations between municipality-level DEA efficiency scores (Services and Treatment models) and the set of irregularity indices. Table 38 reports coefficients for the full sample and for metropolitan vs. non-metropolitan subsamples (asterisk denotes $p < 0.10$). Across both efficiency measures, correlations are small in magnitude ($|r| \leq 0.22$) and predominantly non-significant. Two weak, positive associations appear for the Services model, between the LDA-supported index and efficiency in the overall sample ($r = 0.14$) and within metropolitan municipalities ($r = 0.22$). For the Treatment model, the overall correlations with the indicator-count index ($r = 0.15$) and the category-weighted index ($r = 0.16$) are similarly modest; metropolitan subsample correlations are near zero, and estimates for non-metropolitan municipalities are positive but not statistically distinguishable from zero.

Table 38: Correlation Coefficients Between Efficiency and Indices

Index	Overall	Metropolitan	NonMetropolitan
	Correlations with Services Model Efficiency		
Case Count Index	0.09	0.04	0.09
Indicator Count Index	0.07	-0.01	0.06
Weighted Indicator Count	0.04	0.06	0.01
Case Category Based Weighted Index	0.08	-0.01	0.06
LDA Supported Index	0.14*	0.22*	0.08
	Correlations with Treatment Model Efficiency		
Case Count Index	0.14	-0.02	0.13
Indicator Count Index	0.15*	-0.12	0.15
Weighted Indicator Count	0.11	-0.12	0.09
Case Category Based Weighted Index	0.16*	-0.12	0.16
LDA Supported Index	0.02	-0.02	0.17

Taken together, these results indicate that efficiency and audit-based irregularity indices move largely independently, suggesting they capture distinct dimensions of municipal performance. Beyond multidimensionality, a second reason for the weak correlations is the scope mismatch between outputs used in the DEA and the coverage of the irregularity indices. The two efficiency measures are constructed from a narrow set of outputs, basic services and treatment services, representing only a subset of municipal activities. By contrast, the irregularity indices aggregate findings across all audited functions. It is therefore plausible that movements in overall irregularities do not align closely with efficiency measured on this narrower service bundle.

These results also highlight the limited explanatory power of cumulative indices, which aggregate irregularities across diverse topics. As such, they motivate a shift toward topic-specific modeling, which enables a deeper grasp of irregularities. To clear potential ambiguities, lets remind the different types of indices again:

- Overall indices, which are cumulative irregularity scores used in the ML models and correlation analysis (e.g., case count, indicator count, and their weighted versions). For this category a single index per report is assigned.

- Topic-based indices, which disaggregate irregularities into seven thematic categories (by LDA) and are jointly entered into the regressions for deeper interpretability. For this category, a 7 dimensioned vector is assigned per report.

We next estimate pooled OLS regressions of topic-specific irregularity measures on efficiency and controls (log population, GDP per capita, unemployment), reported separately for the Services and Treatment efficiency models and for all/metro/non-metro samples (Table 39 and Table 40). This time we are searching for a topic specific association between efficiencies and irregularity topics.

Table 39: Pooled OLS Results – Services Efficiency Model

	Dependent variable:			
	Efficiency Services			
	LDA Dummies	LDA Counts	Plain Counts	Weighted Counts
	(1)	(2)	(3)	(4)
Accounting & Asset Management	76.153***	6.924***	0.286***	0.087***
	(3.700)	(0.874)	(0.057)	(0.020)
Budgeting & Financial Oversight	10.043**	4.581**	0.242*	0.094
	(4.324)	(2.098)	(0.137)	(0.059)
Procurement & Public Contracts	-13.222**	-1.513	-0.061	-0.041
	(5.389)	(5.815)	(0.125)	(0.048)
Real Estate & Related Incomes	-0.619	-1.478	-0.144**	-0.046**
	(4.558)	(2.643)	(0.059)	(0.018)
Taxation & Revenue Collection	18.358***	13.138***	0.691***	0.237***
	(3.574)	(3.444)	(0.215)	(0.079)
Urban Planning & Transportation	-10.083*	-4.971***	-0.036	0.013
	(5.316)	(1.859)	(0.081)	(0.029)
Public Oversight & Social Services	1.427	-3.899	-0.183*	-0.067
	(5.104)	(2.851)	(0.102)	(0.041)
Log Population	-11.396***	-39.723***	-46.578***	-48.083***
	(3.320)	(4.566)	(4.825)	(4.862)
GDP per Capita (× 1,000)	1.148**	2.710***	3.127***	3.152***
	(0.559)	(0.859)	(0.927)	(0.941)
Metropolitan Dummy	17.882**	44.809***	53.817***	55.472***
	(7.539)	(11.533)	(12.404)	(12.571)
Constant	134.703***	501.691***	587.232***	607.098***
	(40.001)	(53.769)	(56.690)	(57.039)
Observations	253	253	253	253
R^2	0.891	0.737	0.692	0.683
Adjusted R^2	0.887	0.727	0.680	0.670
Residual Std. Error (df = 242)	25.824	40.160	43.477	44.102
F Statistic (df = 10; 242)	198.748***	67.989***	54.458***	52.245***

Note: * p<0.10, ** p<0.05, *** p<0.01

Table 40: Pooled OLS Results – Treatment Efficiency Model

	Dependent variable:			
	Efficiency Treatment			
	LDA Dummies	LDA Counts	Plain Counts	Weighted Counts
	(1)	(2)	(3)	(4)
Accounting & Asset Management	66.778***	5.316***	0.176***	0.054***
	(6.068)	(0.740)	(0.043)	(0.016)
Budgeting & Financial Oversight	11.185***	7.271***	0.505***	0.213***
	(3.969)	(2.322)	(0.109)	(0.045)
Procurement & Public Contracts	-1.089	-6.453*	-0.155*	-0.075**
	(4.449)	(3.895)	(0.083)	(0.032)
Real Estate & Related Incomes	-3.712	-1.322	-0.091**	-0.027*
	(4.418)	(2.005)	(0.045)	(0.014)
Taxation & Revenue Collection	10.415***	5.574*	0.403**	0.145**
	(3.897)	(2.844)	(0.180)	(0.067)
Urban Planning & Transportation	4.078	0.645	0.148**	0.068**
	(6.426)	(1.507)	(0.073)	(0.027)
Public Oversight & Social Services	3.386	-2.621	-0.134*	-0.041
	(4.334)	(2.166)	(0.070)	(0.028)
Log Population	-5.874**	-22.077***	-24.814***	-25.236***
	(2.344)	(3.276)	(3.481)	(3.526)
GDP per Capita (× 1,000)	0.712*	1.801***	1.928***	1.919***
	(0.388)	(0.584)	(0.627)	(0.638)
Metropolitan Dummy	5.717	11.061	14.798*	15.131*
	(5.020)	(7.711)	(8.216)	(8.366)
Constant	71.988**	292.176***	326.372***	332.191***
	(28.702)	(39.220)	(41.608)	(42.082)
Observations	194	194	194	194
R^2	0.916	0.802	0.773	0.765
Adjusted R^2	0.911	0.791	0.760	0.753
Residual Std. Error (df = 183)	15.909	24.451	26.167	26.590
F Statistic (df = 10; 183)	199.587***	73.943***	62.244***	59.698***
Note:				* p<0.10, ** p<0.05, *** p<0.01

Four specifications are reported for each efficiency outcome (Services and Treatment), employing LDA dummies, LDA counts, plain counts, and weighted counts.

- Accounting & Asset Management consistently shows a positive and highly significant association with both efficiency outcomes, regardless of the index type. For instance, the coefficient is 0.087 ($p<0.01$) in the weighted counts specification for Services, and 0.054 ($p<0.01$) for Treatment.
- Budgeting & Financial Oversight is also significantly and positively associated, especially in the LDA and plain counts versions.
- Taxation & Revenue Collection exhibits strong positive effects across all specifications.
- Procurement & Public Contracts and Real Estate & Related Incomes show negative and statistically significant associations in some cases, suggesting these irregularities may be systematically linked with inefficiencies.
- Urban Planning & Transportation is weakly associated and only sporadically significant.
- Public Oversight & Social Services generally shows no significant impact.

Among the index types, the LDA Dummies specification tends to produce inflated and less reliable coefficients. Due to its binary and coarse structure, it is best interpreted as a robustness check rather than a preferred model.

The pooled OLS regressions for both efficiency models reveal consistent and meaningful associations between specific irregularity topics and municipal performance. Notably, Accounting & Asset Management and Taxation & Revenue Collection show strong positive associations across all model types and efficiency outcomes, suggesting that municipalities with intensive irregularities in these domains may also exhibit signs of operational focus or resource mobilization that correlate with higher measured efficiency. In contrast, topics such as Real Estate & Related Incomes and Procurement & Public Contracts consistently yield negative coefficients, implying that irregularities in these areas may undermine efficient public service provision. The case of Urban Planning & Transportation presents a more mixed picture, with positive associations in treatment services models and weaker or negative effects in services, underscoring potential differences in how spatial planning affects various dimensions

of municipal functioning. These differentiated results highlight the interpretive value of disaggregated, topic-specific indices over cumulative irregularity measures.

While pooled OLS offer an initial understanding of cross-sectional associations, they do not account for unobserved heterogeneity across municipalities that may bias coefficient estimates. Factors such as governance quality, geographic context, or institutional capacity, often constant over time, may confound the observed relationships. To address this, we estimate fixed effects (FE) panel models, which control for all time-invariant municipal characteristics and allow for a more robust assessment of within-municipality variation over time.

To account for unobserved time-invariant heterogeneity, FE panel regressions were estimated for Efficiency Services. As shown in Table 41, the significant topics corroborate some of the OLS findings.

- Accounting & Asset Management and Taxation & Revenue Collection maintain positive and significant associations, particularly in the weighted and topic counts models.
- Real Estate & Related Incomes again shows a consistently negative and significant relationship, further validating its adverse connection with service efficiency.
- Other topics remain insignificant in the FE context.

Table 41: Fixed Effects - Panel Model

	Dependent variable:		
	Efficiency Services		
	Plain Counts	Weighted Counts	Topic Counts
	(1)	(2)	(3)
Accounting and Asset Management	5.383**	1.961**	53.414
	(2.567)	(0.891)	(46.054)
Budgeting and Financial Oversight	5.742	2.688	66.006
	(5.690)	(2.433)	(104.110)
Procurement and Public Contracts	-4.976	-2.449	-109.669
	(5.502)	(2.089)	(254.715)
Real Estate and Related Incomes	-9.653***	-3.051***	-594.507***
	(2.826)	(0.852)	(140.638)
Taxation and Revenue Collection	-16.428	-5.869	-390.192**
	(11.682)	(4.374)	(187.030)
Urban Planning and Transportation	-1.186	-0.631	130.516
	(3.023)	(1.077)	(81.524)
Public Oversight and Social Services	-1.464	-0.463	-68.540
	(3.768)	(1.483)	(115.223)
Log Population	-1,188.462***	-1,191.903***	-1,090.837***
	(287.592)	(285.932)	(291.089)
GDP per Capita (× 1,000)	-15.827	-16.496	-12.244
	(16.348)	(16.345)	(16.197)
Observations	253	253	253
R^2	0.284	0.288	0.295
F Statistic (df = 9; 160)	7.058***	7.182***	7.425***
Note:			* p<0.10, ** p<0.05, *** p<0.01

Notably, the R^2 values are lower in FE models, reflecting the restricted within-unit variation, yet the significance of certain topics like Accounting & Asset Management is robust to this stricter specification.

In all panel specifications, the regressions control for (log) population size and GDP per capita, both of which are widely used to account for structural and socioeconomic differences across municipalities. Among these controls, the population variable consistently shows a negative and statistically significant association with service efficiency, suggesting that larger municipalities may face scale-related coordination challenges not captured by per capita financial metrics. On the other hand, GDP per capita does not show a significant relationship in most models, indicating that

economic development alone may not drive variations in efficiency when irregularity dimensions are taken into account. These controls, combined with city and year fixed effects, strengthen the internal validity of the FE estimations by isolating the role of topic-specific irregularity scores in predicting municipal efficiency.

6.6. Discussion

Taken together, the evidence points to a weak and highly topic-specific link between technical efficiency and audit-based irregularities. Simple Pearson correlations are small in magnitude and mostly insignificant across both efficiency definitions and across metropolitan splits. This is consistent with a scope mismatch: the DEA scores capture efficiency for a narrow bundle of service outputs, whereas the irregularity indices summarize findings across the municipality's full activity set.

The pooled OLS results revealed significant and economically meaningful associations between certain topic-specific irregularities and efficiency scores. For both municipal service provision and infrastructure treatment, irregularities in Accounting & Asset Management, Procurement & Public Contracts, and Budgeting & Financial Oversight were consistently associated with lower efficiency. These three domains align with those identified in the literature as having high discretionary scope and susceptibility to capture (Alesina & Tabellini, 2008; Olken & Pande, 2012; Tanzi & Davoodi, 1997).

Conversely, irregularities in topics such as Taxation & Revenue Collection or Urban Planning & Transportation often had insignificant or even positive associations with efficiency scores. While these patterns require cautious interpretation, they may reflect cases where efficiency is achieved at the expense of compliance or transparency, a notion supported by Bandiera et al. (2009) and Glaeser & Saks (2006), who discuss the trade-offs between rule adherence and output metrics.

In both service supply and treatment services models, control variables behaved as expected. Higher population size (log scale) and GDP per capita generally predicted higher efficiency, although effect sizes varied across specifications. These variables help account for scale economies and socioeconomic capacity in public service provision.

The fixed effects panel regressions refine the previous findings by controlling for time-invariant municipality characteristics that may bias pooled estimates. This approach allows us to better isolate the within-unit variation in efficiency outcomes attributable to irregularity patterns.

Among the topic-specific irregularities, Accounting & Asset Management maintains a positive and statistically significant relationship with efficiency across the first two specifications. This initially counterintuitive finding may reflect administrative diligence correlated with both efficiency and irregularity detection, or possibly improved detection or reporting capacity in better-performing municipalities.

By contrast, Real Estate & Related Incomes exhibits a consistently negative and highly significant association with municipal efficiency. This supports the hypothesis that irregularities in asset valuation, association, or land-based revenue practices may introduce allocative distortions and managerial inefficiencies.

Irregularities in Taxation & Revenue Collection are also significantly and negatively associated with efficiency in the topic count model. This aligns with the literature emphasizing the critical role of reliable local revenue systems for sustained service delivery and efficient budgeting (e.g. Glaeser & Saks, 2006; Klitgaard, 1988).

In contrast, irregularities in Procurement & Public Contracts and Budgeting & Financial Oversight do not reach statistical significance under fixed effects, diverging from pooled OLS results. This suggests that cross-sectional variation, rather than within-municipality changes, may be driving their earlier significance, potentially due to structural or institutional differences across municipalities.

Regarding control variables, population remains strongly and negatively associated with efficiency, suggesting that scale introduces complexity, coordination problems, or managerial slack that impairs performance. GDP per capita, while consistently negative, does not reach significance in any specification, implying that local wealth levels do not systematically translate into higher technical efficiency once irregularity patterns are considered.

Taken together, the regression results are strongly consistent with the theoretical literature linking discretion, weak oversight, and inefficient allocation (Rose-

Ackerman, 1999; Shleifer & Vishny, 1993). The finding that procurement and budgeting-related irregularities are especially harmful is well-aligned with micro-level audit studies (Avis et al., 2018; Olken, 2007) and subnational efficiency research (De Sousa et al., 2005; Geys et al., 2010). These domains are particularly susceptible to resource leakage, inflated costs, and non-meritocratic allocation, all of which depress technical efficiency on the production frontier.

The evidence thus suggests that irregularities in budgeting and procurement are not merely symptoms of governance failure but active drivers of inefficiency. Moreover, by quantifying irregularity exposure at the topic level, this study moves beyond binary corruption indicators and offers a more actionable framework for diagnosis and policy design.

This study used two different efficiency scores based on basic services and treatment services. These outputs are a focused small portion of the services supplied by the municipalities; the models reflect efficiencies subject to these narrow outputs. On the other hand, covering all the actions taken by municipality, irregularities reflect overall performance of the municipality. A wider scope of efficiency outputs can significantly improve the analysis.

7. CONCLUSION

This thesis addresses two dimensions of local governance in Türkiye by examining technical efficiency and irregularities in provincial municipalities through Data Envelopment Analysis (DEA) and Machine Learning (ML) methodologies. The study contributes substantially to the analysis of municipal efficiency and irregularities, providing both methodological innovations and empirical insights.

The introductory and theoretical framework chapters establish the foundation by situating the research within broader debates on public intervention, decentralization, and the evolving responsibilities of local governments. The discussion highlights how global ideological shifts, from Keynesian interventionism to neoliberal market-oriented reforms, have shaped the scope of public authority and, consequently, the functions of local administrations. In this context, the Turkish case is characterized by a relatively centralized governance model, limited fiscal autonomy, and a modest share of local government expenditure in GDP compared to OECD averages.

The historical overview of decentralization underscores both its potential benefits, enhanced accountability, responsiveness, and service delivery, and its fiscal and managerial risks, including inefficiencies and increased reliance on outsourcing. The Turkish municipality framework is presented in detail, covering types, population distribution, resource composition, and expenditure patterns. This structural understanding is essential to interpreting the empirical analyses that follow, as it clarifies the fiscal, institutional, and demographic constraints within which municipalities operate.

By combining these insights with a dual methodological approach, DEA for efficiency assessment and ML for irregularity prediction, the groundwork is laid for a comprehensive evaluation of municipal performance. This dual perspective acknowledges that improving local governance requires not only optimizing resource

allocation and service provision but also mitigating irregularities and strengthening fiscal integrity. The analyses thus bridge theoretical debates with the empirical strategy, ensuring that results are both contextually grounded and policy-relevant.

The DEA analysis assessed the technical efficiency of Turkish provincial municipalities using two complementary models: a service provision model capturing the breadth of municipal service outputs, and a treatment services model focusing on infrastructure and environmental services. Drawing on a uniquely compiled longitudinal dataset covering 2012–2022 from audit reports, municipal budget data, and official statistics, the study addressed the data scarcity challenge often noted in Turkish local government research. The results revealed substantial heterogeneity in efficiency scores across time and space. Some municipalities, in both metropolitan and non-metropolitan categories, consistently achieved high efficiency, illustrating the feasibility of aligning inputs and outputs effectively within existing constraints. Others persistently underperformed, echoing earlier literature on mismatches between expenditure priorities and service needs.

Structural determinants also emerged. Metropolitan municipalities benefited from economies of scale in some areas, but these advantages were offset by complexity in service delivery. Smaller municipalities displayed greater variability: some leveraged leaner structures to good effect, while others struggled with resource limitations. Over time, trajectories differed markedly, with some municipalities improving steadily and others stagnating or declining. With the help of slack analysis, these patterns offer policymakers, auditors, and municipal managers diagnostic benchmarks to target interventions.

From a policy and practical standpoint, DEA-based benchmarking supports better budget prioritization, targeted fiscal transfers, and performance-oriented oversight. Nevertheless, limitations remain, particularly the restricted availability of detailed output data, time inconsistent audit coverage, and the exclusion of district-level municipalities, all of which suggest avenues for future research, including broader service categories and extended institutional coverage.

The second major component of the thesis developed five distinct indices to quantify municipal-level irregularities using Turkish Court of Accounts audit reports, ranging from simple frequency counts to an LDA-supported semantic index. The approach responds to limitations of perception-based measures by leveraging standardized, case-based legal findings, ensuring replicability and reducing bias.

The indices were linked to financial and socio-economic data via gradient boosting (GBM) and random forest (RF) models. Predictive accuracy improved systematically with richer index construction, peaking at ~0.80 R^2 for the LDA-based index. The constructed indices placed greater emphasis on irregularities in accounting and asset management, budgeting and financial oversight, and public oversight and social services, underscoring their stronger ties to municipal financial statements. On the prediction side, enterprise and ownership revenues, social security expenses, and short-term liabilities emerged as the most influential fiscal predictors of irregularities. SHAP value analysis, applied selectively to random samples for illustrative purposes, addresses municipality specific areas of improvement for irregularities.

By transforming unstructured legal audit texts into structured, interpretable indices and linking them to routine financial accounts, the thesis develops a scalable governance risk assessment framework. Beyond methodological novelty, the predictive capability of the machine learning models demonstrates practical utility: given a municipal balance sheet, the algorithm can estimate the likelihood of irregularities, equipping auditors and administrators with actionable early-warning signals. Combined with SHAP-based insights into the most influential fiscal variables, the framework both translates complex legal findings into policy-relevant evidence and provides oversight bodies with targeted tools for prioritization. This dual contribution ensures that the approach is simultaneously replicable across public-sector contexts and directly usable for proactive, risk-based governance.

The final analytical step linked DEA efficiency scores with the irregularity indices to explore potential connections between service-production efficiency and governance risk. Pearson correlations between the two sets of measures were uniformly small ($|r| \leq 0.22$) and largely insignificant across the full sample and metropolitan/non-metropolitan splits, indicating that efficiency and irregularities are largely independent dimensions of performance. This reflects a scope mismatch: DEA models capture

efficiency in a narrow bundle of outputs, while irregularity indices cover the municipality's entire activity set.

To explore domain-specific relationships, the analysis used topic-specific irregularity measures. Pooled OLS regressions identified consistent patterns:

- Accounting & Asset Management and Taxation & Revenue Collection were positively and significantly associated with both efficiency measures in most index types.
- Real Estate & Related Incomes and Procurement & Public Contracts showed negative associations in several models, consistent with inefficiency risks from asset mismanagement and procurement discretion.
- Other topics exhibited mixed or insignificant effects.

For the control variables, population size was often negatively associated with efficiency, while GDP per capita showed a weaker and less consistent role.

Fixed-effects panel regressions refined these findings by controlling for time-invariant municipal characteristics. Real Estate & Related Incomes retained a robust negative association with efficiency, and Accounting & Asset Management remained positively significant in some specifications. The earlier pooled OLS significance of procurement and budgeting irregularities diminished under FE, pointing to cross-sectional rather than within-municipality drivers.

Overall, these results indicate a modest and highly topic-specific efficiency–irregularity link, with meaningful overlap only in high-discretion, asset-related governance domains.

The thesis makes three principal contributions:

1. Empirical – Assembles one of the most comprehensive longitudinal datasets for Turkish municipalities, integrating service, financial, and audit-derived governance measures.
2. Methodological – Introduces a dual DEA–ML framework with topic-level irregularity modelling and a novel LDA-based index.

3. Substantive – Shows that efficiency and governance integrity are complementary but largely orthogonal, with overlaps concentrated in specific, high-discretion domains.

For municipal managers, DEA benchmarking can guide operational improvements; for audit institutions, topic-specific indices support risk-based oversight; for central government, combined efficiency–irregularity diagnostics can inform targeted fiscal and managerial interventions. A joint efficiency–irregularity monitoring system could provide a practical dashboard for aligning resource allocation, service performance, and governance safeguards.

Future work could enhance both components of the framework developed in this thesis. On the DEA side, efficiency measurement could be strengthened by incorporating a richer and more diverse set of outputs that reflect the full scope of municipal responsibilities, for example, transportation services, road and infrastructure works, nurseries, elder care, cultural facilities, and environmental protection. This would narrow the scope mismatch between efficiency frontiers and the broader service portfolio of municipalities, enabling a more comprehensive assessment of performance.

On the irregularity measurement side, indices could be further refined using transformer-based natural language processing (NLP) models such as BERT, RoBERTa, or Longformer. These architectures can capture deeper semantic relationships in audit texts, improving both the thematic precision of topic extraction and the predictive quality of the resulting indices. Combined with rigorous bias and sensitivity analysis, such methods could yield more nuanced and portable governance risk measures.

Another easily adoptable future work can be extension of indices and ML analysis to municipal subsidiaries and district municipalities. Introducing a political dummy set, indicators such as ruling party, central government match status, trustee status can give political insights on both efficiencies, irregularities and unbiasedness of the audit.

Taken together, the thesis advances an integrated governance-analytics framework for Turkish municipalities that bridges efficiency analysis and governance risk

assessment. First, it demonstrates that large-scale, longitudinal efficiency benchmarking is both technically and institutionally feasible using administrative datasets, enabling evidence-based decisions in resource allocation, intergovernmental transfer design, and performance-based budgeting. Second, it shows that unstructured audit texts can be algorithmically transformed into valid, replicable irregularity indices, which can then be linked to routine fiscal accounts to produce interpretable risk scores. Such metrics provide an operational basis for targeted oversight and risk-adjusted fiscal management.

The joint deployment of these tools allows for a clearer conceptual and empirical distinction between production efficiency (maximizing outputs from given inputs) and governance integrity (minimizing misallocation and compliance risks), while also identifying cases where these pillars intersect. This separation, and potential reconciliation, addresses a common limitation in public sector performance debates, where efficiency and integrity are often conflated, leading to misaligned incentives and suboptimal policy design. With investments in data quality, integration, and analytical capacity, alongside ethical safeguards for model use, the approach can support a sustainable shift toward accountable, efficient, and transparent local governance, both nationally and internationally.

REFERENCES

2022 Yılı Mahalli İdareler Genel Faaliyet Raporu. (2023). [Resmi Rapor]. T.C. Çevre, Şehircilik ve İklim Değişikliği Bakanlığı, Yerel Yönetimler Genel Müdürlüğü. https://yerelyonetimler.csb.gov.tr/faaliyet-raporlari-i-88463

Adaman, F., & Çarkoğlu, A. (2000). *Türkiye'de Yerel ve Merkezi Yönetimlerde Hizmetlerden Tatmin, Patronaj İlişkileri ve Reform*. TESEV Yayınları.

Afonso, A., & Fernandes, S. (2008). Assessing and explaining the relative efficiency of local government. *Journal of Behavioral and Experimental Economics (Formerly The Journal of Socio-Economics)*, *37*(5), 1946–1979.

Akdoğan, İ. (2008). Belediyelerin Mali Performansı ve Hizmet Performansı Arasındaki İlişki. *Unpublished Phd Thesis*, 110.

Alesina, A., & Tabellini, G. (2008). Bureaucrats or politicians? Part II: Multiple policy tasks. *Journal of Public Economics*, *92*(3–4), 426–447. https://doi.org/10.1016/j.jpubeco.2007.06.004

Ash, E., Galletta, S., & Giommoni, T. (2020). A Machine Learning Approach to Analyzing Corruption in Local Public Finances. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3589545

Avis, E., Ferraz, C., & Finan, F. (2018). Do Government Audits Reduce Corruption? Estimating the Impacts of Exposing Corrupt Politicians. *Journal of Political Economy*, *126*(5), 1912–1964. https://doi.org/10.1086/699209

Baesens, B., Van Vlasselaer, V., & Verbeke, W. (2015). *Fraud analytics: Strategies and methods for detection and prevention*. John Wiley & Sons.

Bandiera, O., Prat, A., & Valletti, T. (2009). Active and Passive Waste in Government Spending: Evidence from a Policy Experiment. *American Economic Review*, *99*(4), 1278–1308. https://doi.org/10.1257/aer.99.4.1278

Banker, R. D., Charnes, A., & Cooper, W. W. (1984). Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis. *Management Science*, *30*(9), 1078–1092. https://doi.org/10.1287/mnsc.30.9.1078

Batırel, Ö. F. (1990). *Kamu Maliyesi Teorisine Giriş*. Filiz Kitabevi.

Bator, F. M. (1958). The Anatomy of Market Failure. *The Quarterly Journal of Economics*, *72*(3), 351–379. https://doi.org/10.2307/1882231

Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM*, *55*(4), 77–84. https://doi.org/10.1145/2133806.2133826

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, *3*(Jan), 993–1022.

Bobonis, G. J., Cámara Fuertes, L. R., & Schwabe, R. (2016). Monitoring Corruptible Politicians. *American Economic Review*, *106*(8), Article 8. https://doi.org/10.1257/aer.20130874

Borge, L.-E., Falch, T., & Tovmo, P. (2008). Public sector efficiency: The roles of political and budgetary institutions, fiscal capacity, and democratic participation. *Public Choice*, *136*(3–4), 475–495. https://doi.org/10.1007/s11127-008-9309-7

Bosio, E., Djankov, S., Glaeser, E. L., & Shleifer, A. (2022). Public Procurement in Law and Practice. *Journal of Economic Perspectives*, *36*(4), 29–56. https://doi.org/10.1257/jep.36.4.29

Bostan, I., Onofrei, M., Popescu, C., Lupu, D., & Firtescu, B. (2018). Efficiency and Corruption in Local Counties: Evidence from Romania. *Lex Localis - Journal of Local Self-Government*, *16*(2), 215–229. https://doi.org/10.4335/16.2.215-229(2018)

Bowlin, W. F. (1998). Measuring Performance: An Introduction to Data Envelopment Analysis (DEA). *The Journal of Cost Analysis*, *15*(2), Article 2. https://doi.org/10.1080/08823871.1998.10462318

Breiman, L. (1996). Bagging Predictors. *Machine Learning*, *24*(2), 123–140. https://doi.org/10.1023/A:1018054314350

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32.

Çağlar, A. (2003). *Veri Zarflama Analizi ile Belediyelerin Etkinlik Ölçümü* [PhD Thesis, Hacettepe University]. https://acikbilim.yok.gov.tr/handle/20.500.12812/482214

Çelikkaya, F. (2016). *Türkiye'deki Büyükşehir Belediyelerinin Etkinlik Analizi* [PhD Thesis]. Gaziosmanpaşa Üniversitesi.

Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, *2*(6), Article 6. https://doi.org/10.1016/0377-2217(78)90138-8

Clausen, B., Kraay, A., & Nyiri, Z. (2011). Corruption and Confidence in Public Institutions: Evidence from a Global Survey. *The World Bank Economic Review*, *25*(2), Article 2. https://doi.org/10.1093/wber/lhr018

Colonnelli, E., Gallego, J. A., & Prem, M. (2019). *What predicts corruption?* (Documentos de Trabajo No. 017144; Issue 017144). Universidad del Rosario. https://ideas.repec.org/p/col/000092/017144.html

Cooper, W. W., Seiford, L. M., & Tone, K. (2007). *Data Envelopment Analysis: A Comprehensive Text with Models, Applications, References and DEA-Solver Software*. Springer US. https://doi.org/10.1007/978-0-387-45283-8

Daniele, G., & Giommoni, T. (2021). *Corruption under Austerity* (CESifo Working Paper Series No. 8909; Issue 8909). CESifo. https://ideas.repec.org/p/ces/ceswps/_8909.html

De Borger, B., & Kerstens, K. (1996). Cost efficiency of Belgian local governments: A comparative analysis of FDH, DEA, and econometric approaches. *Regional Science and Urban Economics*, *26*(2), Article 2. https://doi.org/10.1016/0166-0462(95)02127-2

De Sousa, M. da C. S., Cribari-Neto, F., & Stosic, B. D. (2005). Explaining DEA Technical Efficiency Scores in an Outlier Corrected Environment: The Case of Public Services in Brazilian Municipalities. *Brazilian Review of Econometrics*, *25*(2). https://EconPapers.repec.org/RePEc:sbe:breart:v:25:y:2005:i:2:a:2507

Del Monte, A., & Papagni, E. (2001). Public expenditure, corruption, and economic growth: The case of Italy. *European Journal of Political Economy*, *17*(1), 1–16. https://doi.org/10.1016/S0176-2680(00)00025-2

Dong, B., & Torgler, B. (2013). Causes of corruption: Evidence from China. *China Economic Review*, *26*, 152–169. https://doi.org/10.1016/j.chieco.2012.09.005

Duflo, E., Greenstone, M., Pande, R., & Ryan, N. (2013). Truth-telling by Third-party Auditors and the Response of Polluting Firms: Experimental Evidence from India*. *The Quarterly Journal of Economics*, *128*(4), 1499–1545. https://doi.org/10.1093/qje/qjt024

Farrell, M. J. (1957). The Measurement of Productive Efficiency. *Journal of the Royal Statistical Society. Series A (General)*, *120*(3), 253–290.

Ferraz, C., & Finan, F. (2008). Exposing Corrupt Politicians: The Effects of Brazil's Publicly Released Audits on Electoral Outcomes *. *Quarterly Journal of Economics*, *123*(2), Article 2. https://doi.org/10.1162/qjec.2008.123.2.703

Ferraz, C., & Finan, F. (2009). *Electoral Accountability and Corruption: Evidence from the Audits of Local Governments* (NBER Working Papers No. 14937; Issue 14937). National Bureau of Economic Research, Inc. https://EconPapers.repec.org/RePEc:nbr:nberwo:14937

Finocchiaro Castro, M., Guccio, C., Romeo, D., & Vidoli, F. (2025). How does institutional quality affect the efficiency of local government? An assessment of Italian municipalities. *Economia Politica*, *42*(2), 569–597. https://doi.org/10.1007/s40888-025-00359-0

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 1189–1232.

Gallego, J., Rivero, G., & Martínez, J. (2021). Preventing rather than punishing: An early warning model of malfeasance in public procurement. *International Journal of Forecasting*, *37*(1), Article 1. https://doi.org/10.1016/j.ijforecast.2020.06.006

Gentzkow, M., Kelly, B., & Taddy, M. (2019). Text as Data. *Journal of Economic Literature*, *57*(3), 535–574. https://doi.org/10.1257/jel.20181020

Gentzkow, M., & Shapiro, J. M. (2010). What Drives Media Slant? Evidence From U.S. Daily Newspapers. *Econometrica*, *78*(1), Article 1. https://doi.org/10.3982/ECTA7195

Gentzkow, M., Shapiro, J., & Taddy, M. (2016). *Measuring Group Differences in High-Dimensional Choices: Method and Application to Congressional Speech* (No. w22423; Issue w22423, p. w22423). National Bureau of Economic Research. https://doi.org/10.3386/w22423

Geys, B., Heinemann, F., & Kalb, A. (2010). Voter involvement, fiscal autonomy and public sector efficiency: Evidence from German municipalities. *European Journal of Political Economy*, *26*(2), 265–278. https://doi.org/10.1016/j.ejpoleco.2009.11.002

Glaeser, E. L., & Saks, R. E. (2006). Corruption in America. *Journal of Public Economics*, *90*(6–7), 1053–1072. https://doi.org/10.1016/j.jpubeco.2005.08.007

Griffiths, T. L., & Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National Academy of Sciences*, *101*(suppl_1), 5228–5235. https://doi.org/10.1073/pnas.0307752101

Grimmer, J., & Stewart, B. M. (2013). Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts. *Political Analysis*, *21*(3), 267–297. https://doi.org/10.1093/pan/mps028

Güneş, İ., & Akdoğan, M. (2007). Büyükşehir Belediye Hizmetlerinin Göreli Etkinlik Analizi. *Çagdaş Yerel Yönetimler Dergisi*.

Habermas, J., Burger, T., & Lawrence, F. (1992). *The structural transformation of the public sphere: An inquiry into a category of bourgeois society*. The MIT press.

Harris, Z. S. (1954). Distributional Structure. WORD, *10*(2–3), 146–162. https://doi.org/10.1080/00437956.1954.11659520

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning*. Springer New York. https://doi.org/10.1007/978-0-387-84858-7

Hessami, Z. (2014). Political corruption, public procurement, and budget composition: Theory and evidence from OECD countries. *European Journal of Political Economy*, *34*, 372–389. https://doi.org/10.1016/j.ejpoleco.2014.02.005

İlkay, M. S., & Doğan, N. Ö. (2009). Veri zarflama analizi ile Kapadokya Bölgesindeki Belediyelerin etkinlik ölçümü: 2004 ve 2008 yıllarına ilişkin bir karşılaştırma. *Erciyes Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, *0*(32), Article 32. http://search/yayin/detay/96083

International Labour Organisation (Ed.). (2001). *The impact of decentralization and privatization on municipal services: Report for discussion at the Joint Meeting on the Impact of Decentralization and Privatization on Municipal Services*. Joint Meeting on the Impact of Decentralization and Privatization on Municipal Services, Geneva. International Labour Office.

Jans, M., Lybaert, N., & Vanhoof, K. (2010). Internal fraud risk reduction: Results of a data mining case study. *International Journal of Accounting Information Systems*, *11*(1), 17–41. https://doi.org/10.1016/j.accinf.2009.12.004

Kaplan, M., Çelik, T., & Tekeli, R. (2006). *Türkiye'de Büyükşehir Belediyelerinin Performanslarının Ampirik Analizi, 2002-2004*. Bilgi, Ekonomi ve Yönetim Kongresi.

Karpuz, N. B. (2008). *Belediyelerin Finansman Kaynağı Olarak Borçlanma Ve Tekirdağ Belediyesı (1996-2006)*. 118.

Kıvratma, Ş. (2016). *Belediyelerin Ekonomik Sınıflandırma Bütçelerinin Yapısal Analizi*. 101.

Klitgaard, R. (1988). *Controlling Corruption*. University of California Press. http://www.jstor.org/stable/10.1525/j.ctt1pnj3b

Kohavi, R. (1995). A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. In C. S. Mellish (Ed.), *Proceedings of the 14th International Joint Conference on Artificial Intelligence (IJCAI-95)* (Vol. 2, pp. 1137–1143). Morgan Kaufmann.

Küçük, H. (2015). Türkiye'de Belediye Şirketlerinin Denetimi Üzerine Bir Değerlendirme. *Journal of International Management*, 14.

Kutlar, A., Bakirci, F., & Yüksel, F. (2012). An analysis on the economic effectiveness of municipalities in Turkey. *African Journal of Marketing Management*, *4*, 80–98.

León, C. J., Araña, J. E., & de León, J. (2013). Correcting for Scale Perception Bias in Measuring Corruption: An Application to Chile and Spain. *Social Indicators Research*, *114*(3), Article 3. https://doi.org/10.1007/s11205-012-0185-7

Lewis-Faupel, S., Neggers, Y., Olken, B. A., & Pande, R. (2016). Can Electronic Procurement Improve Infrastructure Provision? Evidence from Public Works in India and Indonesia. *American Economic Journal: Applied Economics*, *8*(3), 43–72.

Li, S.-H., Yen, D. C., Lu, W.-H., & Wang, C. (2012). Identifying the signs of fraudulent accounts using data mining techniques. *Computers in Human Behavior*, *28*(3), 1002–1013. https://doi.org/10.1016/j.chb.2012.01.002

Liu, C., & Mikesell, J. L. (2014). The Impact of Public Officials' Corruption on the Size and Allocation of U.S. State Spending. *Public Administration Review*, *74*(3), Article 3. https://doi.org/10.1111/puar.12212

Local Government Law, No. 6360, Official Gazette of Türkiye, No. 28489 (2013). https://www.mevzuat.gov.tr/mevzuat?MevzuatNo=6360&MevzuatTur=1&MevzuatTertip=5

López-Iturriaga, F. J., & Sanz, I. P. (2018). Predicting Public Corruption with Neural Networks: An Analysis of Spanish Provinces. *Social Indicators Research*, *140*(3), Article 3. https://doi.org/10.1007/s11205-017-1802-2

Loughran, T., & Mcdonald, B. (2011). When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks. *The Journal of Finance*, *66*(1), Article 1. https://doi.org/10.1111/j.1540-6261.2010.01625.x

Lundberg, S., & Lee, S.-I. (2017). *A Unified Approach to Interpreting Model Predictions* (Version 2). arXiv. https://doi.org/10.48550/ARXIV.1705.07874

Lundberg, S. M., Erion, G. G., & Lee, S.-I. (2018). *Consistent Individualized Feature Attribution for Tree Ensembles* (Version 3). arXiv. https://doi.org/10.48550/ARXIV.1802.03888

Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to Information Retrieval* (1st edn). Cambridge University Press. https://doi.org/10.1017/CBO9780511809071

Mauro, P. (1995). Corruption and Growth. *The Quarterly Journal of Economics*, *110*(3), 681–712. https://doi.org/10.2307/2946696

Mauro, P. (1998). Corruption and the composition of government expenditure. *Journal of Public Economics*, *69*(2), Article 2. https://doi.org/10.1016/S0047-2727(98)00025-5

Minsky, H. P. (1986). *Stabilizing an Unstable Economy*. Yale University Press. https://books.google.com.tr/books?id=XTl9QgAACAAJ

Narbón-Perpiñá, I., & De Witte, K. (2018). Local governments' efficiency: A systematic literature review—part I. *International Transactions in Operational Research*, *25*(2), 431–468. https://doi.org/10.1111/itor.12364

Neiva de Figueiredo, J. (2013). Are corruption levels accurately identified? The case of U.S. states. *Journal of Policy Modeling*, *35*(1), Article 1. https://doi.org/10.1016/j.jpolmod.2012.01.006

Newman, D., Lau, J. H., Grieser, K., & Baldwin, T. (2010). Automatic Evaluation of Topic Coherence. In R. Kaplan, J. Burstein, M. Harper, & G. Penn (Eds), *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics* (pp. 100–108). Association for Computational Linguistics. https://aclanthology.org/N10-1012/

OECD. (2016). *OECD Regions at a Glance 2016*. OECD. https://doi.org/10.1787/reg_glance-2016-en

Olken, B. A. (2007). Monitoring Corruption: Evidence from a Field Experiment in Indonesia. *Journal of Political Economy*, *115*(2), 200–249. https://doi.org/10.1086/517935

Olken, B. A., & Pande, R. (2012). Corruption in Developing Countries. *Annual Review of Economics*, *4*(1), 479–509. https://doi.org/10.1146/annurev-economics-080511-110917

Prieto, A. M., & Zofio, J. (2001). Evaluating Effectiveness in Public Provision of Infrastructure and Equipment: The Case of Spanish Municipalities. *Journal of Productivity Analysis*, *15*(1), 41–58. https://doi.org/10.1023/A:1026595807015

Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning*, *1*(1), 81–106. https://doi.org/10.1007/BF00116251

Reinikka, R., & Svensson, J. (2004). Local Capture: Evidence from a Central Government Transfer Program in Uganda. *The Quarterly Journal of Economics*, *119*(2), 679–705. https://doi.org/10.1162/0033553041382120

Reinikka, R., & Svensson, J. (2005). Fighting Corruption to Improve Schooling: Evidence from a Newspaper Campaign in Uganda. *Journal of the European Economic Association*, *3*(2/3), 259–267.

Rose-Ackerman, S. (1999). *Corruption and Government: Causes, Consequences, and Reform* (1st edn). Cambridge University Press. https://doi.org/10.1017/CBO9781139175098

Salton, G., & Buckley, C. (1988). Term-weighting approaches in automatic text retrieval. *Information Processing & Management*, *24*(5), 513–523. https://doi.org/10.1016/0306-4573(88)90021-0

Shleifer, A., & Vishny, R. W. (1993). Corruption. *The Quarterly Journal of Economics*, *108*(3), 599–617. https://doi.org/10.2307/2118402

Simar, L., & Wilson, P. W. (2002). Non-parametric tests of returns to scale. *European Journal of Operational Research*, *139*(1), 115–132. https://doi.org/10.1016/S0377-2217(01)00167-9

Sun, A., Lachanski, M., & Fabozzi, F. J. (2016). Trade the tweet: Social media text mining and sparse matrix factorization for stock market prediction. *International*

Review of Financial Analysis, *48*, 272–281. https://doi.org/10.1016/j.irfa.2016.10.009

Tanzi, V., & Davoodi, H. R. (1997). *Corruption, Public Investment, and Growth* (IMF Working Paper No. WP/97/139). International Monetary Fund. https://doi.org/10.5089/9781451929515.001

Tanzi, V., & Schuknecht, L. (2000). *Public Spending in the 20th Century: A Global Perspective* (1st edn). Cambridge University Press. https://doi.org/10.1017/CBO9780511625800

Tobin, J. (1970). On Limiting the Domain of Inequality. In M. J. Boskin (Ed.), *The Public Economics of Government Redistribution* (pp. 99–121). Harvard University Press.

Treisman, D. (2007). What Have We Learned About the Causes of Corruption from Ten Years of Cross-National Empirical Research? *Annual Review of Political Science*, *10*(1), Article 1. https://doi.org/10.1146/annurev.polisci.10.081205.095418

Tupper, H. C., & Resende, M. (2004). Efficiency and regulatory issues in the Brazilian water and sewage sector: An empirical study. *Utilities Policy*, *12*(1), 29–40. https://doi.org/10.1016/j.jup.2003.11.001

Turati, G., Boetti, L., & Piacenza, M. (2012). Decentralization and Local Governments' Performance: How Does Fiscal Autonomy Affect Spending Efficiency? *FinanzArchiv*, *68*(3), 269. https://doi.org/10.1628/001522112X653840

Türkiye, R. of. (2004). *Law on Metropolitan Municipalities (Law No. 5216)*. https://www.mevzuat.gov.tr/mevzuat?MevzuatNo=5216&MevzuatTur=1&MevzuatTertip=5

Türkiye, R. of. (2005). *Law of Municipalities (Law No. 5393)*. https://www.mevzuat.gov.tr/mevzuat?MevzuatNo=5393&MevzuatTur=1&MevzuatTertip=5

Wallach, H., Mimno, D., & McCallum, A. (2009). Rethinking LDA: Why Priors Matter. In Y. Bengio, D. Schuurmans, J. Lafferty, C. Williams, & A. Culotta (Eds), *Advances in Neural Information Processing Systems* (Vol. 22). Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2009/file/0d0871f0806eae32d30983b62252da50-Paper.pdf

Winters, M. S., & Weitz-Shapiro, R. (2013). Lacking Information or Condoning Corruption: When Do Voters Support Corrupt Politicians? *Comparative Politics*, *45*(4), Article 4.

Woodbury, K., & Dollery, B. (2004). Efficiency Measurement in Australian Local Government: The Case of New South Wales Municipal Water Services. *Review of Policy Research*, *21*(5), Article 5. https://doi.org/10.1111/j.1541-1338.2004.00098.x

Wu, Y., & Zhu, J. (2011). Corruption, anti-corruption, and inter-county income disparity in China. *The Social Science Journal*, *48*(3), Article 3. https://doi.org/10.1016/j.soscij.2011.05.001

APPENDIX

A Supplementary Graphs and Tables

Figure A1: Health and Education Expenditure Shares for Local Governments vs Public Expenditure Share of Local Governments

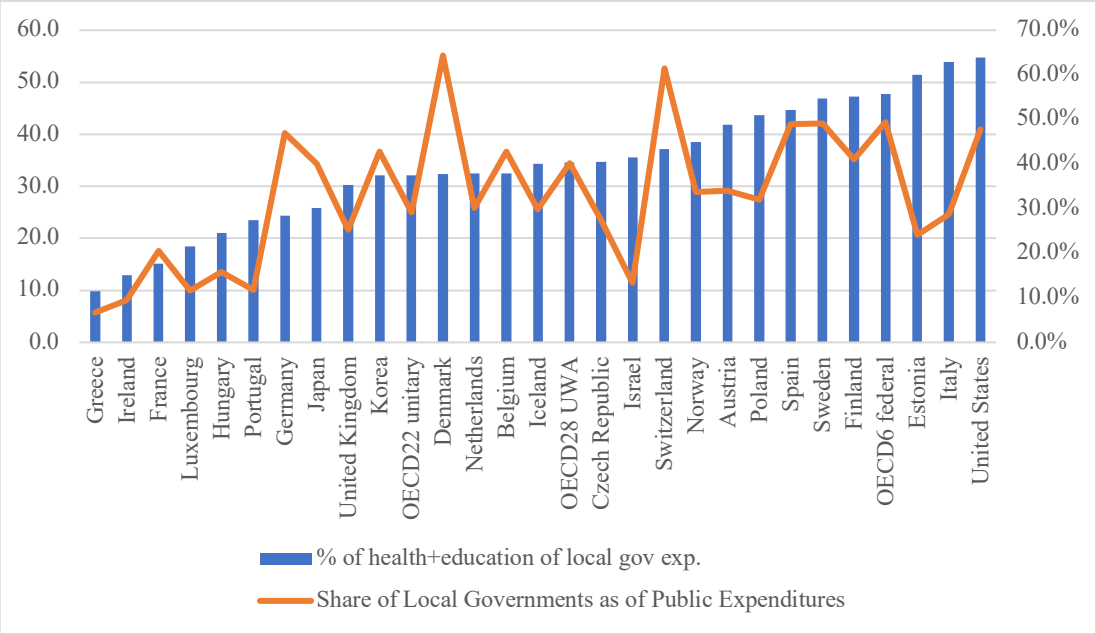


Table A1: Services Model Efficiency Results – All Metropolitan Municipalities

City	2014	2016	2018	2020	2022	Average
diyarbakir	1.00		1.00	1.00		1.00
istanbul	1.00	1.00	1.00	1.00	1.00	1.00
kayseri	1.00	1.00	1.00	1.00	1.00	1.00
kocaeli	1.00	1.00	1.00	1.00	1.00	1.00
manisa	1.00	1.00	1.00	1.00	1.00	1.00
gaziantep			1.00	1.00		1.00

City	2014	2016	2018	2020	2022	Average
ankara	0.95			1.00	1.00	0.98
izmir	0.72	1.00	1.00	1.00	1.00	0.94
bursa	1.00	0.83	1.00	0.90		0.93
aydin		1.00	0.70	1.00	1.00	0.92
mardin	1.00		0.68	1.00	1.00	0.92
mugla	1.00	0.58	1.00	1.00	1.00	0.92
denizli		1.00		0.71	1.00	0.90
ordu		0.62		0.99	1.00	0.87
sanliurfa	0.52	1.00	1.00	0.94		0.86
tekirdag	0.92	0.47	0.86	1.00	1.00	0.85
antalya	0.32	1.00		1.00	1.00	0.83
hatay	1.00	0.67		0.82		0.83
eskisehir	0.59	0.52	0.98	1.00	1.00	0.82
adana	0.60	0.47	1.00	1.00		0.77
sakarya	0.47	0.77	0.82	1.00	0.76	0.76
balikesir	0.27	0.63	0.72	0.76	1.00	0.68
malatya	0.20	0.40	1.00	1.00		0.65
trabzon	0.14	0.63	0.82	0.78	0.87	0.65
van		0.90	0.12	1.00	0.55	0.64
samsun	0.37	0.81	0.38	0.73	0.79	0.62
konya	0.50	0.68	0.40	0.77	0.65	0.60
erzurum		0.31	0.54	0.63	0.80	0.57
kahramanmaras		0.31	0.31	0.77		0.46

City	2014	2016	2018	2020	2022	Average
mersin	0.19	0.37	0.36	0.48	0.59	0.40
Average	**0.69**	**0.73**	**0.79**	**0.91**	**0.91**	**0.81**

Table A2: Services Model Efficiency Results – All Non-Metropolitan Municipalities

City	2012	2014	2016	2018	2020	2022	Average
afyonkarahisar	1.00	1.00	1.00		1.00	1.00	1.00
ardahan	1.00	1.00	1.00				1.00
kirklareli	1.00				1.00	1.00	1.00
mus	1.00	1.00			1.00	1.00	1.00
sirnak	1.00	1.00	1.00	1.00			1.00
tokat	1.00				1.00	1.00	1.00
agri		1.00	1.00				1.00
amasya		1.00			1.00	1.00	1.00
bilecik		1.00	1.00	1.00		1.00	1.00
bingol		1.00					1.00
bitlis		1.00	1.00				1.00
cankiri		1.00	1.00				1.00
isparta		1.00	1.00	1.00	1.00		1.00
kilis		1.00			1.00		1.00
zonguldak		1.00					1.00
karabuk			1.00				1.00
adiyaman					1.00		1.00
kirikkale					1.00	1.00	1.00

City	2012	2014	2016	2018	2020	2022	Average
erzincan	1.00	0.97	1.00	1.00			0.99
sinop		1.00	1.00	1.00		0.80	0.95
bartin	1.00	0.81	0.89	1.00	1.00	1.00	0.95
bolu		0.88			1.00	0.87	0.92
canakkale	0.83	0.62	1.00		1.00	1.00	0.89
edirne		0.87		0.74	0.97	0.98	0.89
kirsehir		0.66		1.00		1.00	0.89
sivas	0.77	1.00	0.88		0.94	0.80	0.88
corum		0.91	0.70		1.00		0.87
nigde		0.85		1.00	1.00	0.59	0.86
bayburt		0.79			0.92		0.85
burdur		0.74	0.66		1.00	1.00	0.85
kutahya		1.00	0.38	0.96	1.00		0.83
tunceli					0.86	0.79	0.82
kars		1.00	0.47	0.99			0.82
artvin	0.63					1.00	0.82
gumushane		0.80					0.80
batman	1.00	1.00	0.15		1.00		0.79
elazig		1.00	0.21		1.00		0.74
kastamonu	1.00	0.66		0.52		0.67	0.71
siirt		0.97	0.44				0.71
usak		0.66					0.66
nevsehir	1.00		0.10	0.88		0.64	0.66

City	2012	2014	2016	2018	2020	2022	Average
duzce	0.82	0.72	0.39				0.64
yalova	1.00		0.10	0.53		0.85	0.62
aksaray						0.62	0.62
osmaniye		1.00	0.19				0.60
hakkari		1.00	0.11				0.56
rize			0.09	0.85			0.47
karaman		0.75	0.10				0.42
giresun		0.55	0.12	0.41			0.36
igdir			0.29				0.29
Average	**0.94**	**0.90**	**0.61**	**0.87**	**0.99**	**0.89**	**0.83**

Table A3: Treatment Services Model Efficiency Results – All Metropolitan Municipalities

City	2014	2016	2018	2020	2022	Average
istanbul	1.00	1.00	1.00	1.00	1.00	1.00
mugla	1.00	1.00	1.00	1.00	1.00	1.00
gaziantep			1.00	1.00		1.00
antalya					1.00	1.00
sanliurfa	1.00	1.00	1.00	0.94		0.98
kocaeli	1.00	1.00	1.00	1.00	0.91	0.98
sakarya	1.00	0.85	1.00	1.00	1.00	0.97
trabzon	0.83	1.00	1.00	1.00	1.00	0.97
manisa	1.00	1.00	0.50	1.00	1.00	0.90

City	2014	2016	2018	2020	2022	Average
ordu		0.65		1.00	1.00	0.88
aydin		1.00	0.66	0.84	1.00	0.88
adana	1.00	0.47	1.00	1.00		0.87
ankara	1.00			0.59	1.00	0.86
samsun	0.86	0.96	0.65	0.82	1.00	0.86
erzurum		0.34	1.00	1.00	1.00	0.84
hatay	1.00	0.67		0.82		0.83
eskisehir	1.00	0.50	0.79	0.79	0.90	0.80
mardin			0.31	1.00	1.00	0.77
bursa	1.00	0.79	0.73	0.54		0.77
tekirdag	1.00	0.45	0.30	1.00	0.81	0.71
diyarbakir	0.84		0.31	0.93		0.69
mersin	0.88	0.38	0.57	0.55	1.00	0.68
balikesir	0.87	0.62	0.40	0.63	0.85	0.67
izmir	1.00	0.64	0.19	0.93	0.56	0.67
van		0.90	0.12	1.00	0.60	0.65
denizli		0.54			0.63	0.59
konya	0.66	0.69	0.25	0.56	0.45	0.52
kahramanmaras		0.31	0.05	0.72		0.36
Average	**0.94**	**0.73**	**0.65**	**0.87**	**0.89**	**0.81**

Table A4: Treatment Services Model Efficiency Results – All Non-Metropolitan Municipalities

City	2014	2016	2018	2020	2022	Average
afyonkarahisar	1.00	1.00		1.00	1.00	1.00
cankiri	1.00	1.00				1.00
sinop	1.00	1.00	1.00		1.00	1.00
zonguldak	1.00					1.00
karabuk		1.00				1.00
yalova		1.00	1.00		1.00	1.00
erzincan			1.00			1.00
nigde			1.00	1.00	1.00	1.00
sirnak			1.00			1.00
adiyaman				1.00		1.00
kirikkale				1.00	1.00	1.00
kirklareli				1.00	1.00	1.00
tokat				1.00	1.00	1.00
aksaray					1.00	1.00
artvin					1.00	1.00
kirsehir					1.00	1.00
bolu	1.00			1.00	0.91	0.97
siirt	1.00	0.90				0.95
kars	1.00	0.79	1.00			0.93
bitlis	1.00	0.86				0.93
bayburt				0.89		0.89
edirne	0.69		0.92	1.00	0.95	0.89

City	2014	2016	2018	2020	2022	Average
usak	0.89					0.89
bartin		0.48	1.00	1.00	1.00	0.87
sivas	0.96	0.49		1.00	0.91	0.84
kilis	0.77			0.90		0.84
canakkale	0.74	0.57		1.00	1.00	0.83
isparta	0.94	0.36	1.00	1.00		0.82
duzce	1.00	0.63				0.82
amasya	0.61				1.00	0.81
kutahya	1.00	0.16	1.00	1.00		0.79
rize		0.56	1.00			0.78
elazig	1.00	0.21		1.00		0.74
kastamonu	0.63		0.62		0.91	0.72
corum	0.90	0.17		0.96		0.68
nevsehir		0.07	1.00		0.80	0.62
giresun			0.57			0.57
batman		0.15		1.00		0.57
Average	**0.91**	**0.60**	**0.94**	**0.99**	**0.97**	**0.89**

Figure A2: Part of a Sample Financial Table as Included in the Sayıştay Report



BİLANÇO

Kurum Kodu : 46.34.01.00 **İSTANBUL BÜYÜKŞEHİR BELEDİYESİ** **Yıl : 2021**

	AKTİF HESAPLAR	Cari Yıl Tutar TL		PASİF HESAPLAR	Cari Yıl Tutar TL
1	**DÖNEN VARLIKLAR**	**19.945.593.179,60**	**3**	**KISA VADELİ YABANCI KAYNAKLAR**	**17.860.200.332,73**
10	**Hazır Değerler**	**7.513.886.489,47**	**30**	**Kısa Vadeli İç Mali Borçlar**	**1.044.159.929,88**
101	ALINAN ÇEKLER HESABI	0,00	300	BANKA KREDİLERİ HESABI	1.044.159.929,88
102	BANKA HESABI	3.435.943.149,45	303	KAMU İDARELERİNE MALİ BORÇLAR HESABI	0,00
103	VERİLEN ÇEKLER VE GÖNDERME EMİRLERİ HESABI (-)	-1.057.895.319,57	309	KISA VADELİ DİĞER İÇ MALİ BORÇLAR HESABI	0,00
			31	**Kısa Vadeli Dış Mali Borçlar**	**4.956.176.731,87**
104	PROJE ÖZEL HESABI	5.182.055.127,92	310	CARİ YILDA ÖDENECEK DIŞ MALİ BORÇLAR HESABI	4.956.176.731,87
105	DÖVİZ HESABI	31.737.216,22	**32**	**Faaliyet Borçları**	**6.284.892.839,13**
106	DÖVİZ GÖNDERME EMİRLERİ HESABI (-)	-96.489.237,59	320	BÜTÇE EMANETLERİ HESABI	6.284.892.839,13
108	DİĞER HAZIR DEĞERLER HESABI	1.669.250,00	329	DİĞER ÇEŞİTLİ BORÇLAR HESABI	0,00
109	BANKA KREDİ KARTLARINDAN ALACAKLAR HESABI	16.866.303,04	**33**	**Emanet Yabancı Kaynaklar**	**4.574.379.593,50**
			330	ALINAN DEPOZİTO VE TEMİNATLAR HESABI	851.950.151,50
11	**Menkul Kıymet ve Varlıklar**	**1.405.403.250,00**	333	EMANETLER HESABI	3.722.429.442,00
112	KAMU KESİMİ TAHVİL SENET VE BONOLARI HESABI	1.405.403.250,00	**36**	**Ödenecek Diğer Yükümlülükler**	**678.522.309,72**
12	**Faaliyet Alacakları**	**2.805.705.372,69**	360	ÖDENECEK VERGİ VE FONLAR HESABI	100.609.389,46
120	GELİRLERDEN ALACAKLAR HESABI	908.777.938,54	361	ÖDENECEK SOSYAL GÜVENLİK KESİNTİLERİ HESABI	37.851.162,37
121	GELİRLERDEN TAKİPLİ ALACAKLAR HESABI	1.300.134.317,43			
122	GELİRLERDEN TECİLLİ VE TEHİRLİ ALACAKLAR	73.672.199,01	362	FONLAR VEYA DİĞER KAMU İDARELERİ ADINA	18.612.305,33

Figure A3: An Example of Findings Listed in Sayıştay Reports

T.C. Sayıştay Başkanlığı

7. DENETİM BULGULARI

Raporda yer alan bulgular, denetimler sonucunda tespit edilen hususlara kamu idaresi tarafından verilen cevapların değerlendirilmesi suretiyle düzenlenmiştir.

A. DENETİM GÖRÜŞÜNÜN DAYANAĞI BULGULAR

Bu bölümde, mali rapor ve tablolara verilen görüş ile ilgili olan bulgular yer almaktadır.

BULGU 1: İcra Daireleri Tarafından İdarenin Banka Hesaplarından Doğrudan Kesilen Tutarların Hatalı Muhasebeleştirilmesi

İcra dairelerince İdarenin banka hesaplarından doğrudan yapılan icra kesintilerinin hatalı muhasebe hesaplarında izlendiği görülmüştür.

Genel Yönetim Muhasebe Yönetmeliği'nin "Nazım hesaplara ilişkin ilkeler" başlıklı 12'nci maddesinde düzenlendiği üzere; kamu idarelerinin varlık, kaynak, gelir ve gider hesaplarının dışında, muhasebenin bilgi verme ve izleme görevi yüklediği işlemleri, muhasebe

Figure A4: Another Example of Findings Listed in Sayıştay Reports

T.C. Sayıştay Başkanlığı

Sonuç olarak, meri mevzuat ve yapılan değerlendirmeler neticesinde; İBB Meclisi tarafından onaylanan tarife ile İdare uygulamasının birbirine aykırılık teşkil etmeyecek şekilde gerçekleştirilmesi gerekmektedir.

BULGU 66: Herhangi Bir Sözleşmeye Dayanmadan Belediye Şirket Personelinin Kurumda İstihdam Edilmesi

Belediyenin iştirak şirketlerinde istihdam edilen 195 personelin herhangi bir sözleşme, protokol, ihale vs. olmadan ve Belediye bütçesinden herhangi bir gider yapılmadan İdarenin çeşitli birimlerinde çalıştırıldığı görülmüştür.

Kamu kurum ve kuruluşlarında görev yapan personelin, kurumlar arasında geçiş yapabilmesine veya geçici görevlendirilmesine ilişkin yasal mevzuat bulunmasına karşın, Türk Ticaret Kanunu'na tabi ve kurumlar vergisi mükellefi olan, bu kapsamda gelir ve giderini tam olarak beyan etme yükümlülüğü bulunan ve kar amacı güden belediye şirket personelinin başka kurum, kuruluş ve şirketlerde bedelsiz çalışabileceğine ilişkin herhangi bir düzenleme bulunmamaktadır.

Belediye şirketleri, 6102 sayılı Türk Ticaret Kanunu'na tabi şirketlerdir. Türk Ticaret Kanunu'nun 16 ve 18'inci maddesinde belirtildiği üzere, belediye tarafından ticari şekilde işletilmek üzere kurulan kuruluşlar da tacirdir ve her tacirin, ticaretine ait bütün faaliyetlerinde basiretli bir iş adamı gibi hareket etmesi gerekmektedir.