ADVANCED LAND USE MIX ANALYSIS IN URBAN AREAS USING POINT-BASED DATA: METHODS AND APPLICATIONS

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF INFORMATICS OF
THE MIDDLE EAST TECHNICAL UNIVERSITY
BY

MEHMET ALI AKYOL

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
IN
THE DEPARTMENT OF INFORMATION SYSTEMS

SEPTEMBER 2024

Advanced Land Use Mix Analysis in Urban Areas Using Point-Based Data: Methods and Applications

submitted by **MEHMET ALI AKYOL** in partial fulfillment of the requirements for the degree of **Doctor of Philosophy in Information Systems Department, Middle East Technical University** by,

Prof. Dr. Banu Günel Kılıç
Dean, **Graduate School of Informatics**

Prof. Dr. Altan Koçyiğit
Head of Department, **Information Systems**

Prof. Dr. Nazife Baykal
Supervisor, **Information Systems, METU**

Prof. Dr. Şebnem Düzgün
Co-supervisor, **Dept. of Mining Engineering, Colorado School of Mines**

Examining Committee Members:

Prof. Dr. Banu Günel Kılıç
Information Systems, METU

Prof. Dr. Nazife Baykal
Information Systems, METU

Prof. Dr. Sinan Kalkan
Dept. of Computer Engineering, METU

Prof. Dr. Ersin Karaman
Dept. of Visual Communication Design, Hacı Bayram Veli University

Assoc. Prof. Dr. Serhat Peker
Management Information Systems, Bakırçay University

Date: 03.09.2024

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Surname: Mehmet Ali Akyol

Signature :

ABSTRACT

ADVANCED LAND USE MIX ANALYSIS IN URBAN AREAS USING POINT-BASED DATA: METHODS AND APPLICATIONS

Akyol, Mehmet Ali

Ph.D., Department of Information Systems

Supervisor: Prof. Dr. Nazife Baykal

Co-Supervisor: Prof. Dr. Şebnem Düzgün

SEPTEMBER 2024, 78 pages

This thesis investigates the role of Land Use Mix (LUM) analysis in urban planning, Geographic Information Systems (GIS) research, and disaster risk assessment, offering novel methodologies and tools to address existing challenges in these areas. Traditional methods for assessing LUM often fall short in terms of scalability, adaptability, and the ability to capture the dynamic nature of urban environments. This research overcomes these limitations by introducing advanced, automated approaches to LUM assessment, leveraging publicly available point-based geospatial data.

One of the key contributions of this thesis is the development of an open-source Python package, landusemix, which facilitates the calculation of LUM using established indices such as the Entropy Index and the Herfindahl-Hirschman Index. This package is designed to be user-friendly, scalable, and adaptable, making it a valuable resource for researchers and practitioners in urban planning and GIS.

Furthermore, the thesis extends the application of LUM analysis to assess temporal variations in urban vulnerability, particularly in the context of earthquake risk. By integrating LUM with temporal population dynamics and land usability assessments, the study provides new insights into how urban vulnerability shifts over time, emphasizing the importance of time-sensitive strategies in urban planning and disaster preparedness.

Overall, this research advances the understanding and application of LUM in urban studies, offering practical tools and methodologies that enhance the ability to analyze, plan, and respond to the complexities of urban environments. The findings contribute to the broader goal of promoting sustainable, resilient, and livable cities.

Keywords: Land Use Mix, Geographic Information Systems, Geospatial Data Mining, Mixed-Use Development, Spatial Analysis

ÖZ

NOKTASAL VERİLER KULLANILARAK KENTSEL ALANLARDA GELİŞMİŞ KARIŞIK ARAZİ KULLANIM ANALİZİ: YÖNTEMLER VE UYGULAMALAR

Akyol, Mehmet Ali
Doktora, Bilişim Sistemleri Bölümü
Tez Yöneticisi: Prof. Dr. Nazife Baykal
Ortak Tez Yöneticisi: Prof. Dr. Şebnem Düzgün

Eylül 2024, 78 sayfa

Bu tez, karışık arazi kullanım analizinin kentsel planlama, coğrafi bilgi sistemleri araştırmaları ve afet risk değerlendirmesindeki rolünü inceleyerek, bu alanlardaki mevcut zorlukları ele alan yenilikçi yöntemler ve araçlar sunmaktadır. Geleneksel arazi kullanım çeşitliliği değerlendirme yöntemleri, genellikle ölçeklenebilirlik, uyarlanabilirlik ve kentsel ortamların dinamik doğasını yakalama konusunda yetersiz kalmaktadır. Bu araştırma, halka açık nokta tabanlı coğrafi verileri kullanarak gelişmiş, otomatikleştirilmiş karışık arazi kullanım değerlendirme yaklaşımları sunarak bu sınırlamaların üstesinden gelmektedir.

Bu tezin bir diğer önemli katkısı, Entropi İndeksi ve Herfindahl-Hirschman İndeksi gibi yaygın indeksleri kullanarak karışık arazi kullanım hesaplamasını kolaylaştıran, açık kaynaklı bir Python paketi olan landusemix'in geliştirilmesidir. Bu paket, kullanıcı dostu, ölçeklenebilir ve uyarlanabilir olacak şekilde tasarlanmıştır ve kentsel planlama ve coğrafi bilgi sistemleri alanındaki araştırmacılar ve uygulayıcılar için değerli bir kaynak oluşturmaktadır.

Ayrıca, karışık arazi kullanım analizinin kentsel kırılganlığın, özellikle deprem riski bağlamında, zamana bağlı etkilerini değerlendirerek bir uygulama alanı sunduğu belirtilmelidir. Karışık arazi kullanımını zamana bağlı nüfus dinamikleri ve arazi kullanılabilirlik değerlendirmeleriyle entegre ederek, kentsel kırılganlığın zamanla nasıl değiştiğine dair yeni içgörüler sunmakta ve zaman duyarlı stratejilerin kentsel planlama ve afet hazırlığındaki önemini vurgulamaktadır.

Genel olarak, bu araştırma, karışık arazi kullanımının kentsel alanlardaki uygulanmasını ilerleterek, kentsel ortamların karmaşıklığını analiz etme, planlama ve bu ortamlara yanıt verme yeteneğini artıran pratik araçlar ve metodolojiler sunmaktadır. Bulgular, sürdürülebilir, dayanıklı ve yaşanabilir şehirlerin teşvik edilmesine yönelik daha geniş hedefe katkıda bulunmaktadır.

Anahtar Kelimeler: Arazi Kullanım Çeşitliliği, Coğrafi Bilgi Sistemleri, Coğrafi Veri Madenciliği, Karma Kullanım Gelişimi, Mekansal Analiz

To my lovely daughter, Defne Akyol...

ACKNOWLEDGMENTS

I would like to express my deepest gratitude to my supervisor, Prof. Dr. Nazife Baykal, for her unwavering support, guidance, and encouragement throughout my journey. Her expertise, patience, and commitment to excellence have been instrumental in shaping this research. Prof. Baykal's insightful feedback and constant belief in my potential have pushed me to reach beyond my limits, and I am profoundly thankful for this.

I am also deeply thankful to my co-supervisor, Prof. Dr. Şebnem Düzgün, for her continuous support and valuable feedback. Her broad knowledge in the field and meticulous attention to detail have greatly enriched my research. The time and effort she dedicated to discussing ideas and refining methodologies have been pivotal in developing this thesis, and I am truly grateful for her mentorship.

I would like to thank Prof. Dr. Tuğba Taşkaya Temizel for her guidance during the initial stages of my research. Her early support provided a strong foundation that has been essential to the progress and completion of this work.

I am also immensely grateful to my jury members, Prof. Dr. Banu Günel Kılıç, Prof. Dr. Sinan Kalkan, Prof. Dr. Ersin Karaman, and Assoc. Prof. Dr. Serhat Peker, for their time and support. I am deeply honored to have had such esteemed scholars as part of my jury.

Special thanks to my colleagues Cemal Kocagöz, Kaan Cındık, and Muhammed Dinçtürk for their friendship and understanding. Working alongside you has been a pleasure, and your support has been greatly appreciated.

I would also like to thank my dear friends, Talat Gökçer Canyurt, Faruk Büyüktekin, Fatih Taş, Esra Emekli, and Emre Emekli, for their constant support and encouragement. Your friendship has been invaluable, providing both intellectual stimulation and emotional support, particularly during the hardest times of this journey. I deeply appreciate your presence and the strength you have given me.

I am profoundly grateful to my family, whose unwavering love and support have been the bedrock of my strength throughout this journey. To my mother, Ummahan, and my father, Hüseyin, thank you for your endless love and sacrifices and for instilling the values of perseverance and hard work in me. To my cousin, Sevda Dalar, thank you for your encouragement and support during tough times; your positivity has been a great source of comfort. To my wife, Şeyma Akyol, thank you for your endless patience, understanding, and encouragement during the most challenging times. Lastly, to my little daughter, Defne Akyol, you have been a source of joy and motivation, reminding me of the importance of this work on a daily basis.

Lastly, I am grateful to everyone who contributed to this work, directly or indirectly, and to those who have supported me in various ways throughout this process. Whether through academic advice, motivational support, or simply lending an ear, your contributions have not gone unnoticed, and I deeply appreciate all your efforts.

TABLE OF CONTENTS

APPENDICES

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

LUM	Land Use Mix
POI	Point of Interest
GIS	Geographic Information Systems
LULC	Land Use / Land Cover
NOAA	National Oceanic and Atmospheric Administration
ESA	European Space Agency
CORINE	Coordination of Information on the Environment
VGI	Volunteered Geographic Information
CNN	Convolutional Neural Networks
SOM	Self-Organizing Maps
CELM	Constrained Extreme Learning Machine
AUC	Area Under the Curve
ROC	Receiver Operating Characteristic
HHI	HHI - Herfindahl-Hirschman Index
KDE	Kernel Density Estimation
CBD	Central Business District
API	Application Programming Interface

CHAPTER 1

INTRODUCTION

Land Use Mix (LUM), which refers to the combination and spatial arrangement of different land use types within a given area, has become a pivotal topic in urban planning, Geographic Information Systems (GIS) research, and disaster risk assessment. The diversity of land uses within urban environments significantly impacts public health, transportation efficiency, economic vitality, and overall quality of life. As cities continue to grow and evolve, understanding and optimizing land use mix is increasingly recognized as essential for fostering sustainable, resilient, and livable urban areas.

Urban areas face unprecedented challenges, including rapid urbanization, environmental sustainability, public health concerns, economic disparities, and susceptibility to natural disasters. Assessing LUM is crucial for understanding and addressing these challenges. LUM analysis helps urban planners and policymakers gain insights into the diversity and distribution of land uses within a city, significantly impacting various aspects of urban life.

The motivation for this research is grounded in the realization that mixed-use areas, characterized by various land uses, bring about numerous benefits. To better understand the overarching significance of LUM analysis, it is essential to consider its multifaceted benefits. Figure 1.1 below illustrates the key motivations for promoting mixed land use within urban environments.

Integrating various land functions such as residential, commercial, recreational, and institutional within close proximity is crucial for enhancing urban utility and promoting community cohesion. A well-balanced land use mix can reduce the need for long commutes, improve public safety, support economic development, and encourage social interactions, thereby creating vibrant, self-sustaining neighborhoods [1]. However, accurately assessing and managing land use mix in complex and dynamic urban settings presents significant challenges.

Traditional methods for identifying and analyzing LUM typically involve manual processes, such as field surveys and site visits, which are labor-intensive, time-consuming, and often limited in their geographic scope [2, 3]. While valuable in certain contexts, these conventional approaches are not easily scalable or adaptable to the rapid changes that characterize modern urban environments. Moreover, the reliance on human intervention in these methods introduces potential biases and inconsistencies, further complicating the assessment of land use mix.

Recent advancements in data science and the proliferation of location-based technologies offer new opportunities to overcome these limitations. With the rise of publicly available point-based geospatial data, such as Point of Interest (POI) data, it is now possible to automate the LUM assessment process, making it more scalable, adaptable, and capable of capturing real-time changes in urban land-

Figure 1.1: Multifaceted benefits of LUM analysis in urban areas.

scapes. These data sources and advanced spatial analysis techniques enable more comprehensive and accurate evaluations of land use patterns, thereby providing valuable insights for urban planners and policymakers.

This thesis aims to address the challenges associated with traditional LUM assessment methods by introducing novel, automated approaches that leverage publicly available geospatial data. Additionally, presents an open-source Python package, which facilitates the calculation of LUM using established indices, such as the Entropy Index and the Herfindahl-Hirschman Index, making the identification of LUM easier for researchers and practitioners in the fields of urban planning and GIS.

In addition to advancing LUM analysis methodologies, this thesis explores the application of LUM in assessing temporal variations in urban vulnerability, particularly in the context of earthquake risk. By integrating LUM analysis with temporal population dynamics and land usability assessments, this research provides new insights into how urban vulnerability evolves over time, highlighting the need for time-sensitive strategies in disaster preparedness and urban planning.

Figure 1.2 presents the three core chapters of this research and their interconnections to provide a clear roadmap of the thesis structure. The chapters are organized hierarchically, each building upon the previous one.

- **Identification of LUM (Foundation):** This chapter lays the groundwork for the study by presenting the core methodology for LUM identification using point-based data. It establishes the practical and theoretical basis for the research.

- **The *landusemix* Python Package (Bridge):** Building upon the foundation, this chapter introduces the *landusemix* Python package, operationalizing the presented methodology. It bridges the methodological framework and practical application, making it accessible and usable for a broader audience.

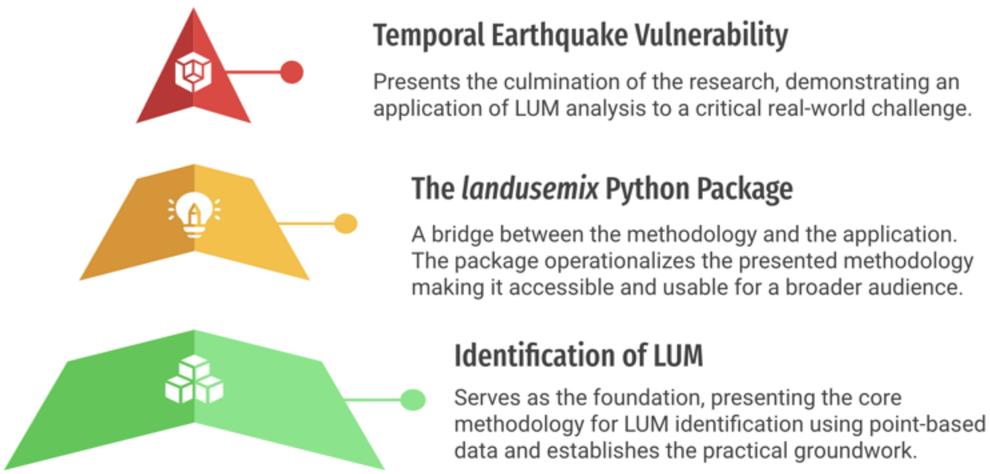


Figure 1.2: Overview of the main chapters of the thesis.

- **Temporal Earthquake Vulnerability (Application):** This chapter represents the culmination of the research, demonstrating an application of LUM analysis to a critical real-world challenge. It showcases how the methodologies developed can be employed to assess temporal earthquake vulnerability in urban areas.

1.1 Study Areas

To validate and demonstrate the proposed methodologies, this research focuses on several different urban areas:

- **Ankara and Kadıköy (Identification of LUM)**:
  - *Ankara*: Serving as the capital of Turkey, Ankara provides a diverse context with various mixed-use scenarios, making it an ideal test bed for developing and validating LUM identification methods.
  - *Kadıköy*: A prominent district in Istanbul known for its vibrant urban life and diverse land use, Kadıköy complements Ankara with its densely populated urban environment, providing a different urban context for comparative analysis.
- **Kadıköy and Avcılar (Temporal Earthquake Vulnerability)**:
  - *Kadıköy*: With its complex mix of residential, commercial, and recreational areas, Kadıköy is used to assess how LUM influences urban vulnerability to earthquakes over time.
  - *Avcılar*: A district in Istanbul characterized by a mix of residential and industrial zones, Avcılar presents a different urban fabric, helping to understand the application's methodology in assessing temporal earthquake vulnerability in varying urban settings.

1.2 Research Questions

This thesis seeks to address several key research questions that are central to advancing the field of LUM analysis and its applications in urban planning, GIS research, and disaster risk assessment:

1. **How can publicly available point-based geospatial data be utilized to automate the identification and classification of LUM in urban areas?**
   - This question explores the feasibility and effectiveness of using POI data and other geospatial datasets to streamline the LUM assessment process, reducing the need for manual intervention and enhancing scalability.
2. **What are the benefits and limitations of integrating Voronoi triangulation with entropy-based indices in modeling LUM, and how does this approach compare to traditional methods?**
   - This question aims to evaluate the novel methodology proposed in this thesis, specifically the combination of Voronoi triangulation and entropy-based calculations, in terms of accuracy, efficiency, and applicability across different urban contexts.
3. **What are the optimal POI density ranges that contribute significantly to the accuracy of LUM classification, and how can these ranges be determined for diverse urban environments?**
   - This question focuses on identifying the POI density thresholds that are most effective for accurate LUM classification, which is critical for applying the proposed methodology in varied urban settings.
4. **How can LUM analysis be integrated into the assessment of temporal urban vulnerability, particularly in the context of earthquake risk, and what insights can this integration provide for disaster preparedness?**
   - This question explores the application of LUM analysis in evaluating how urban vulnerability to natural disasters, such as earthquakes, changes over time. The goal is to understand how LUM can inform more effective, time-sensitive strategies for disaster risk reduction and urban resilience.

1.3 Main Contributions

The contributions of this thesis are summarized as follows:

1. **Development of a Novel Methodology for LUM Identification:**
   - The study introduces a new approach to LUM identification by integrating POI data with Voronoi triangulation and entropy-based calculations. This method reduces the need for extensive fieldwork and human intervention, offering a more efficient and scalable solution for urban planners.

2. **Empirical Validation of the Proposed Methodology:**

   - The thesis provides a thorough empirical evaluation of the proposed methodology using data from real-world urban environments. The study areas, including Ankara and Kadıköy, serve as test cases to demonstrate the method's effectiveness in different urban contexts.

3. **Identification of Optimal POI Density Ranges:**

   - Through detailed analysis, the research identifies optimal POI density ranges that significantly enhance the accuracy of LUM classification. This finding is critical for refining the use of geospatial data in urban analysis.

4. **Application of LUM Analysis to Urban Disaster Vulnerability:**

   - The study extends the application of LUM analysis beyond traditional urban planning to the domain of disaster management. By integrating LUM with temporal vulnerability assessments, the research provides new insights into how mixed land use can influence urban resilience to natural disasters, such as earthquakes.

5. **Contribution to Open Source Tools:**

   - The development and sharing of the *landusemix* Python package as part of this research offers a valuable tool for other researchers and practitioners in the field. This package facilitates the replication and extension of the study's methods, promoting further research and application in urban planning.

6. **Framework for Assessing Temporal Dynamics of Land Use Mix:**

   - The study introduces a framework for incorporating temporal aspects into LUM analysis, which is crucial for understanding how land use patterns evolve over time and their implications for urban resilience. This temporal perspective adds a new dimension to LUM studies, making them more relevant for dynamic urban environments.

1.4 Organization of the Thesis

The outline of this thesis is as follows:

1. **Chapter 2: Related Work**

   - This chapter reviews existing literature on land use mix, geospatial data, and related methodologies. It discusses traditional approaches to LUM identification and highlights the gaps that this study aims to address.

2. **Chapter 3: Identification of Land Use Mix Using Point-Based Geospatial Data in Urban Areas**

   - This chapter details the methodology developed for LUM identification, including the use of POI data, Voronoi triangulation, and entropy-based calculations. It also describes the study areas and the datasets used for the analysis.

3. **Chapter 4: Landusemix: An Open Source Python Package for Calculating Land Use Mix**

    - This chapter introduces the *landusemix* Python package developed as part of this research. It provides an overview of the package's architecture, key features, and illustrative examples of its application.

4. **Chapter 5: Utilizing Land Use Mix to Assess Temporal Earthquake Vulnerability in Urban Areas**

    - This chapter explores the application of LUM analysis in assessing temporal earthquake vulnerability. It presents case studies in the Kadıköy and Avcılar districts, demonstrating how LUM can inform disaster preparedness and urban resilience.

5. **Chapter 6: Conclusion**

    - The final chapter summarizes the findings of the study, discusses the limitations of the research, and suggests directions for future work. It also reflects on the broader implications of the study for urban planning and GIS research.

CHAPTER 2

RELATED WORK

This section presents the background information on the proposed methodology and related studies. Specifically, we review land use/land cover (LULC), LULC classification, and LUM topics in general.

2.1 Land Use/Land Cover (LULC)

LULC refers to how people use the land and interact with the physical land type. Various authorities propose different LULC classes. CORINE (Coordination of Information on the Environment), one of the widely used LULC classes, includes 44 land cover classes in a three-level hierarchy. The main categories are artificial surfaces, agricultural areas, forest and semi-natural areas, wetlands, and water bodies. The National Oceanic and Atmospheric Administration (NOAA) also categorizes land cover such as developed land, agricultural land, grassland, forest land, scrub/scrub land, bushland, palustrine wetlands, estuarine wetlands, and water and submerged land. Additionally, Sentinel, an earth observation program initiated by the European Space Agency (ESA), offers classifications, namely residential, impervious surface, agriculture, bare land, forest, and water [4].

The major limitation of these LULC classes is that they can be too specific or generic. Also, there are significant differences between the types proposed by different sources, making it subjective and problematic in particular cases. However, LULC classes provided by different entities can be useful because they provide a standardized way of categorizing land use and land cover types. Additionally, researchers, policymakers, and other stakeholders can use a common language to compare and contrast land use and land cover across different regions. For instance, the CORINE land cover classification system is widely used in Europe. The European Environmental Agency has adopted it to report on the state and evolution of Europe's environment. The NOAA's land cover classification system is used in the United States for environmental monitoring and reporting purposes.

2.2 LULC Classification

There has been a significant amount of research on LULC classification in the literature. These studies have focused on utilizing various data sources, such as point-based data and satellite imagery, to identify LULC. For instance, the study by [5] combined data from OpenStreetMap road networks, POI data, and satellite imagery to classify urban land use. They used a combination of probabilistic topic models and support vector machines and highlighted the potential for combining remote sensing and social media data for better LULC classification.

Social networks have also been explored for the classification of LULC. [6] explored the relationship between spatial technologies and social media in remote sensing, emphasizing the significance of using social media to monitor various aspects of the urban environment. The study highlighted the importance of social media for monitoring natural disasters and land cover. On the other hand, [7] argued that volunteered geographic information (VGI) from sources like Foursquare could enhance real-time LULC classification. However, the accuracy of VGI information depends on the level of participation, which might be low in some regions. [8] put forth the idea of a “Visitation by Area attraction and Neighborhood” competition to explain the reasons behind people’s check-ins in specific neighborhoods. This study highlights the potential of incorporating information on people’s check-ins in LULC classification models. By understanding this relationship, the authors aim to improve the accuracy of LULC classification. [9] combined social media activity with convolutional neural networks (CNNs) to improve remote sensing for emergency response and rainfall event monitoring. The study demonstrated the effectiveness of using social media and CNNs for remote sensing; however, it is limited by the quality and timeliness of the social media data used.

Other studies also utilized deep learning models for LULC classification. [10] proposed a framework that uses POI data and the Word2Vec model to identify urban land use on the scale of traffic analysis zones, offering a unique approach to incorporating social media data in LULC classification. [11] used geotagged social network images with a Naïve Bayes classifier to explore LULC; however, it is noted that the quality of the geotagged images may limit the performance of this method. Furthermore, [12] used self-organizing maps (SOM) to cluster land uses in urban areas based on patterns of tweet activity, offering a novel approach to utilizing social media data. This study provides a unique perspective in the field of LULC classification by exploring the potential of using social media data in this context. However, it is essential to consider some potential limitations of this approach. For instance, the accuracy of the classification results may be limited by the quality and coverage of the tweet data.

Additionally, [13] proposed a deep learning model that utilizes a CNN and a constrained extreme learning machine (CELM), demonstrating promising results in LULC classification. The data used in the study consisted of remote sensing data, including high-resolution satellite images and ground-truth data. The authors tested the proposed model on several land-use types, including residential, commercial, industrial, and open spaces. The authors aimed to improve the accuracy and efficiency of LULC classification by integrating the two models. The study’s results showed that the proposed model outperformed the traditional single model-based approaches, with promising results in classification accuracy. Lastly, [14] presented a system based on a combination of Place2Vec and POI to identify functional urban regions. This study explores the potential of incorporating POI data into deep-learning models for LULC classification. The use of POI data in combination with Place2Vec allows the system to capture the spatial patterns of land use and the underlying functional relationships between urban regions, demonstrating the integration of POI data into the deep learning model, which enhances the ability of the system to capture the functional connections between urban areas.

In conclusion, LULC classification has been a popular research topic in recent years, focusing on utilizing various data sources such as satellite imagery, social media, and POI data to improve accuracy and efficiency. While many studies have shown promising results, the quality and timeliness of social media data, the quality of geotagged images, and the level of participation in VGI information can limit the performance of some methods. The studies reviewed in this section have demonstrated the potential of integrating social media and POI data into deep learning models for LULC classification, focusing on improving accuracy and efficiency. Further research is needed to address these approaches' limitations and continue exploring the potential of incorporating various data sources for improved LULC classification. However, LUM is a different concept than LULC, and the challenge of identifying LUM is much more complex than LULC.

2.3 Land Use Mix

Land use mix has been studied in different fields, from transportation to public health and housing market analysis to urban economics. Fundamentally, LUM refers to the mix of varying land-use types present in an area or a building. Having an appropriate level of LUM in an area has many benefits. It can help lower the commuting times and shorten the commute distance [15]. In urban economics, a place with different land types has more potential to raise land values [16]. Additionally, it can encourage people to travel more within an area to discover recreational land-use types such as community centers or parks. These studies demonstrate the importance of LUM in urban areas and the need for effective methods for measuring and characterizing it.

To measure LUM, a number of metrics have been proposed, such as adjacency, intensity, and proximity, to understand the spatial distribution of land use types in an area [17]. In another study, they used navigational POIs to develop a series of land use mix indicators to characterize neighborhoods and users' mobile phone activity in 24 hours as a proxy of neighborhood vibrancy [18]. Temporal aspects of land use mix and guided land use policies and innovative transportation methods are described in another study [19].

Furthermore, a study presents influence parameters for predicting mixed land use in urban areas. They developed a mathematical model to analyze the relationship between various factors impacting land use patterns, such as zoning regulations, population density, transportation networks, and others. The results showed that this approach could accurately predict the mixture of land uses in a given urban area [20].

Even though there are different methods to model LUM in the literature [21, 22, 23, 17], to the best of our knowledge, identifying LUM based on POI data and Voronoi triangulation has not been investigated before in the literature. Traditional approaches have utilized travel behavior and urban form data to model LUM [21, 22]. In another study, they have also contributed to this field by providing a grid-based approach utilizing a Fishnet model to analyze land use patterns [23].

Recent studies have further enriched the understanding of LUM by exploring its evaluation and spatial patterns in various megacities. For instance, a study on Chengdu utilized weighted optimization of POI data and the introduction of a modified gravity model to measure mixed land use and analyze its spatial patterns, emphasizing the importance of optimizing internal functional structures for effective urban planning [24]. Another study on Shanghai applied entropy and type number indices to measure LUM, revealing high-high spatial agglomeration in central areas and low-low agglomeration in peripheral regions, demonstrating the need for enhanced evaluation methods [25]. Furthermore, research on Beijing's LUM illustrated how higher LUM levels positively impact housing prices, stressing the significance of LUM in urban economics and real estate markets [26].

However, to the best of our knowledge, no study focuses on identifying LUM in urban areas in an efficient and automated way using publicly available data sources. This gap in the literature motivated the current study, which proposes an efficient method utilizing POI-based data and Voronoi triangulation to identify LUM in urban areas.

CHAPTER 3

IDENTIFICATION OF LAND USE MIX USING POINT-BASED GEOSPATIAL DATA IN URBAN AREAS

It is crucial to understand and accurately assess LUM in urban areas to promote sustainable urban development, enhance livability, and foster social interaction. Traditional methods of LUM assessment are inadequate. They rely on limited data sources, require manual data collection, and are not scalable or adaptable across different urban contexts. This chapter addresses these limitations with a novel, automated approach to LUM assessment that leverages publicly available point-based data.

The chapter begins by defining the problem and highlighting the inadequacies of current LUM assessment methods. It then describes the study areas selected for this research—Ankara and Kadıköy—and explains their significance for validating the proposed methodology. The methodology is presented in detail, covering data collection, preprocessing, and the steps involved in calculating LUM using fishnets and Voronoi diagrams. Following this, the datasets used in the study are outlined, with attention to the sources, preprocessing techniques, and challenges encountered during data collection. The chapter then provides the LUM assessment results, including a comparative analysis of the findings from Ankara and Kadıköy. Finally, the chapter discusses the implications of these results for urban planning and policy-making and concludes with a summary of the key contributions.

3.1 Problem Definition

It is crucial to accurately measure and analyze LUM in urban areas. This directly influences urban sustainability, livability, and social interaction. Traditional approaches to LUM assessment are inadequate. They rely on limited or outdated data sources, manual data collection, and methodologies that lack scalability and flexibility. These methods are time-consuming, prone to errors, and not easily adaptable to different urban contexts.

Furthermore, existing tools and techniques for LUM calculation, such as those based on geographic information systems (GIS), require extensive manual input and are not designed to handle large-scale or repetitive analyses. This presents a significant obstacle for researchers and urban planners who must conduct comprehensive LUM studies efficiently across various regions.

Furthermore, land use data is often inconsistent and of poor quality, particularly in rapidly growing urban areas. In many cases, official data sources are either inaccessible or incomplete, so alternative data collection methods are necessary. However, these alternative methods, such as using point-of-interest (POI) data from online platforms, present significant challenges, including data quality issues, inconsistencies, and the need for extensive preprocessing.

These challenges demand a more efficient, scalable, and adaptable methodology for LUM calculation. Such a methodology must leverage readily available data sources, minimize the need for manual intervention, and provide reliable, reproducible results that can be applied across diverse urban environments. This study addresses these gaps head-on by developing and validating a novel approach to LUM assessment using publicly available point-based data and advanced computational techniques.

3.2 Study Area

Our research focuses on two key study areas: Ankara and Kadıköy, chosen for their diverse urban characteristics and varying patterns of land use. The first study area is Ankara, the capital of Turkey, while the second is Kadıköy, a prominent district of Istanbul. Both areas provide a representative context for analyzing mixed land use.

Ankara is located in the center of Turkey and serves as the country's capital (Figure 3.1). It has a population of around 5.6 million and covers an extensive area of 24,521 km^2. With 24 districts, Ankara offers a wide variety of mixed land use scenarios, as many areas consist of intertwined residential and non-residential complexes. This diversity makes Ankara an ideal location for developing and testing our proposed methodology for identifying LUM.

Kadıköy, a district situated on the northern shore of the Sea of Marmara in Istanbul, is the second study area (Figure 3.1). Known for its vibrant urban life, Kadıköy has a population of 467 thousand and spans an area of 25 km^2. The district is recognized for its mixed-use development patterns, with a diverse range of residential and non-residential buildings. Kadıköy complements Ankara by providing insights into LUM in a densely populated urban environment.

We selected Ankara as the primary study area to develop and validate our methodology due to its larger scale and significant number of mixed-use areas. Afterward, Kadıköy was chosen as the second study area to further test and compare the methodology's effectiveness in a different urban context. The results from Kadıköy were used to cross-validate and ensure the robustness of the findings derived from Ankara.

3.3 Methodology

This section outlines the methodological approach used for Land Use Mix (LUM) modeling, based on a structured process that involves data collection, preprocessing, and the generation of fishnets and Voronoi diagrams, LUM calculation, and the evaluation process.

Figure 3.2 illustrates the overall workflow of our methodological approach. It starts with collecting points of interest (POI) data, followed by creating a fishnet grid. Within each grid cell, Voronois are drawn, allowing for the calculation of LUM for each specific area.

After drawing the Voronoi, we calculate areas of residential and non-residential land use sites depending on the Voronoi area associated with each type of POI, which can be either residential or non-

Figure 3.1: Map of the study areas: (1) Ankara and (2) Kadıköy.

residential. As a result, we estimate LUM values per fishnet using an entropy-based LUM formula, as explained in detail in Section 3.3.3. For evaluation, we collect land use registry data for ground truth from the official sources and apply various data pre-processing techniques to clean the data. Next, we apply a point pattern analysis to understand the dispersion of land registry data. Finally, we calculate the ground truth LUM values and evaluate LUM findings per fishnet with respect to the ground truth results.

3.3.1 Fishnet Creation

We generate fishnets equally dividing the urban parts of the city using QGIS (accessed on 30 May 2021) (`https://www.qgis.org/`), an open-source geographic information analysis tool. Each fishnet eye has dimensions of 500 m by 500 m and an area of around 0.2655 km^2. POI data gathered from different sources, such as Google Maps API offering non-residential locations and from a real estate listing website showcasing both residential and workplaces, are mapped to each fishnet. Generated fishnets covering Ankara city and Kadıköy district can be seen in Appendix A.

3.3.2 Voronoi Generation

Voronoi generation is a process that takes a set of points as input and generates a Voronoi mesh. The mesh comprises a set of edges, and each set is shared between two Voronoi cells. The Voronoi cell, for a point, is the region of space that is closer to that point than to any other.

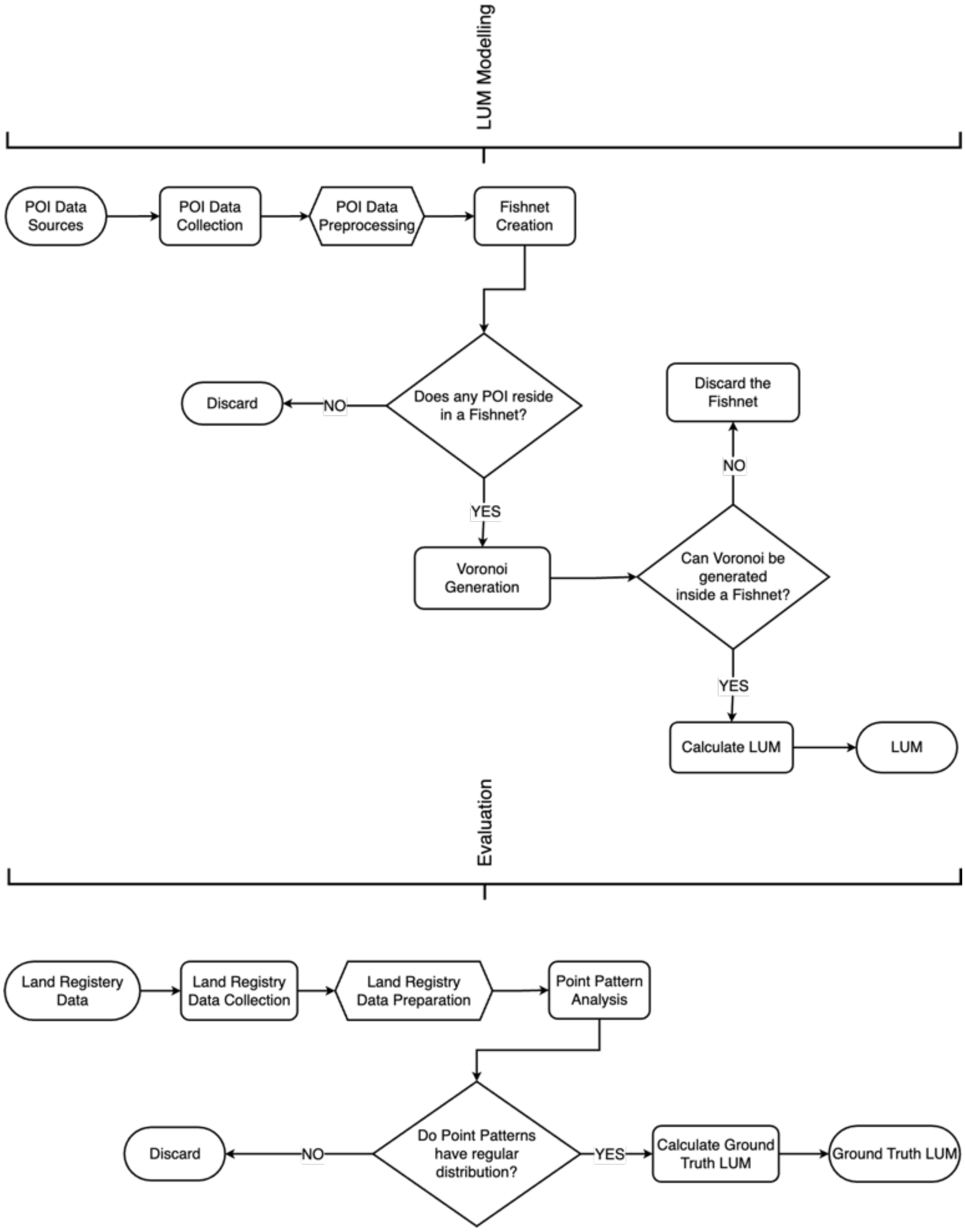


Figure 3.2: Flowchart of LUM modeling and evaluation of results.

This step explains Voronoi generation and depicts a generated Voronoi mesh within a fishnet eye. As no area information was present for each associated POI, we created Voronois to estimate an area for each POI, which we will use next to model LUM based on the POI area.

First, we discard the fishnets with less than four POIs as it is impossible to draw Voronois with less than four points.

Using the Delaunay triangulation [27], Voronois are drawn, which helps calculate the possible associated area for each POI based on the Voronoi coverage. Figure 3.3 shows a generated Voronoi within a fishnet eye with each of the associated POIs.

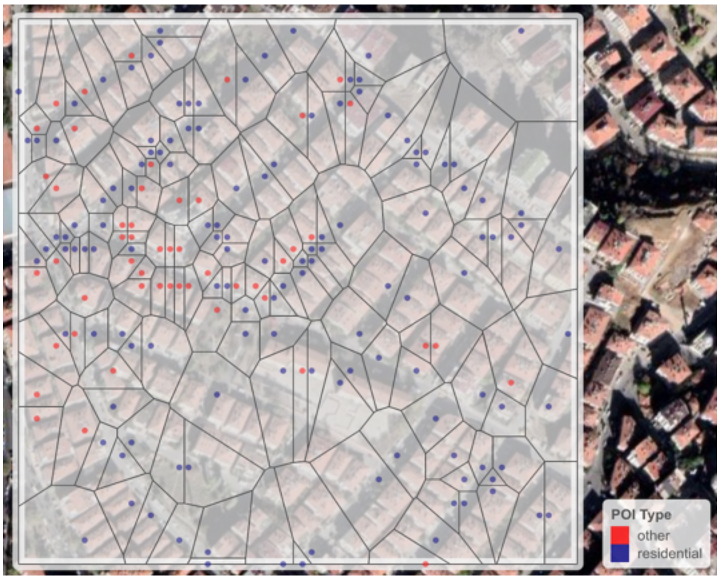


Figure 3.3: Generated Voronoi based on the POI locations and associated POIs are depicted on top of the Voronoi. Blue dots represent residential POIs, and red dots represent non-residential (other) POIs.

3.3.3 Calculate LUM

We calculate how much of the area is used for residential or non-residential purposes using an entropy-based method proposed in [22, 28]. In the original study, they formulated the LUM index based on the two and six land types. In Equation (3.1), we extend it to cover n number of land types where a refers to the area of the specific land-use type, k represents the total area of the particular land use type, and n represents the number of land-use types. Based on the entropy calculation, LUM values range between 0 and 1, where zero indicates no mix, and the higher values indicate a high mix. To calculate LUM, we consider the area of the associated Voronoi of a POI as the area of the particular land use type.

$$LUM = -\sum_{i}[(a_{\mathrm{i}}/k)x\ln(a_{\mathrm{i}}/k)]/\ln(n) \tag{3.1}$$

3.3.4 Evaluation: Point Pattern Analysis

We analyze the dispersion of surface points covering the land registry area using point pattern analysis to measure the quality of the land registry data collection step. We use a well-known technique, the Quadrats method [29] in our study, which divides a region into equal area sub-regions and counts the number of points in each sub-region to measure the distribution of spatial patterns. As a result, the pattern can be asserted as clustered, random, or regular [30, 31, 32] depending on the frequency distribution of the counts compared with the theoretical distribution of spatial randomness [33].

We initially insert regularly distributed synthetic points on each collected land registry area. Our motivation is to understand the adequacy of land registry data via checking the distribution of synthetic points in the area covered by land registry data. If there is a regular distribution of points over the land registry data, it can be concluded as adequate.

An example of surface points can be seen in Figure 3.4a where yellow shapes represent the land registry parcels, and the dots represent the surface points over parcels. The points are inserted with 0.001 latitude and longitude increments.

To determine the optimal grid size, we tried different grid sizes ranging from 2×2 to 14×14 for the Quadrats method. After that, we calculated the number of regular regions with respect to the total number of regions to select the best grid size for our study. Using a method similar to the elbow method in cluster analysis, we selected the grid size (9×9) as the regularity reaches a certain level afterward. Figure 3.5 depicts the percentage of regular places with respect to the grid size.

As a result, we applied the Quadrats method with 9×9 sub-regions. Figure 3.4b demonstrates a sample of quadrants and intersected land registry parcels along with the surface points and quadrant counts.

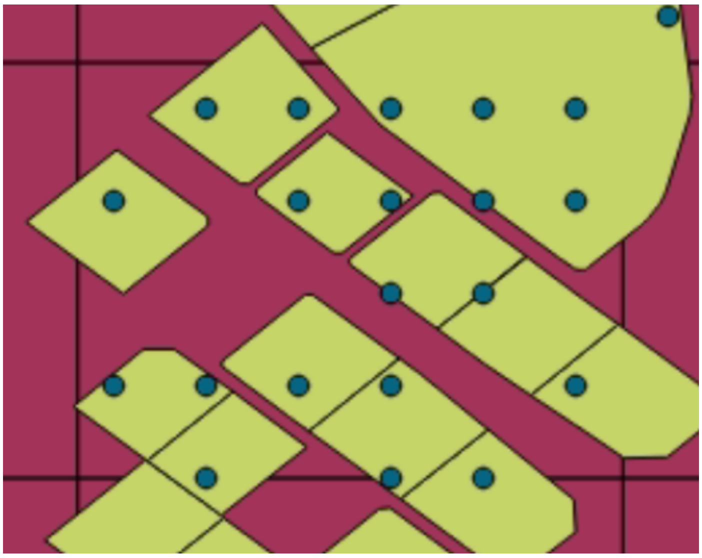
Depending on the quadrat counts, we applied the Quadrat test. As a data preprocessing step, which is detailed in Section 3.4.3, we only selected the fishnets having a regular surface point pattern with a significance threshold of p-value ≤ 0.05.

3.4 Datasets

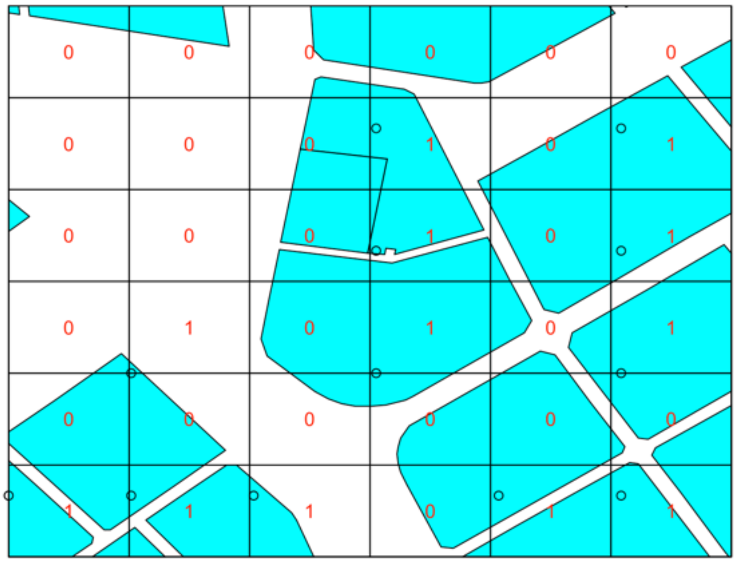
This section will introduce the datasets and preprocessing methods used in our study. The datasets used include publicly available datasets of POI data in addition to land registry data from official sources.

3.4.1 POI Data

We used two primary data sources to acquire POI data for our study. For residential areas, we scraped the data from an ad listing website. For non-residential We used Google Maps API to provide the locations of non-residential areas. A full list of place types offered by Google Maps API at the time of writing can be seen in Appendix B section. For each request to the API, a maximum of 60 results were returned for a particular area, which is limited due to the nature of Google Maps API.



(a)



(b)

Figure 3.4: (**a**) An example of regularly inserted points on each collected land registry data inside a fishnet. (**b**) Another example of a fishnet depicting quadrants, counts, surface points, and parcels. Blue shapes represent the intersecting parcels within the fishnet, and small ovals within blue shapes represent the surface points. The numbers indicate the surface point counts.

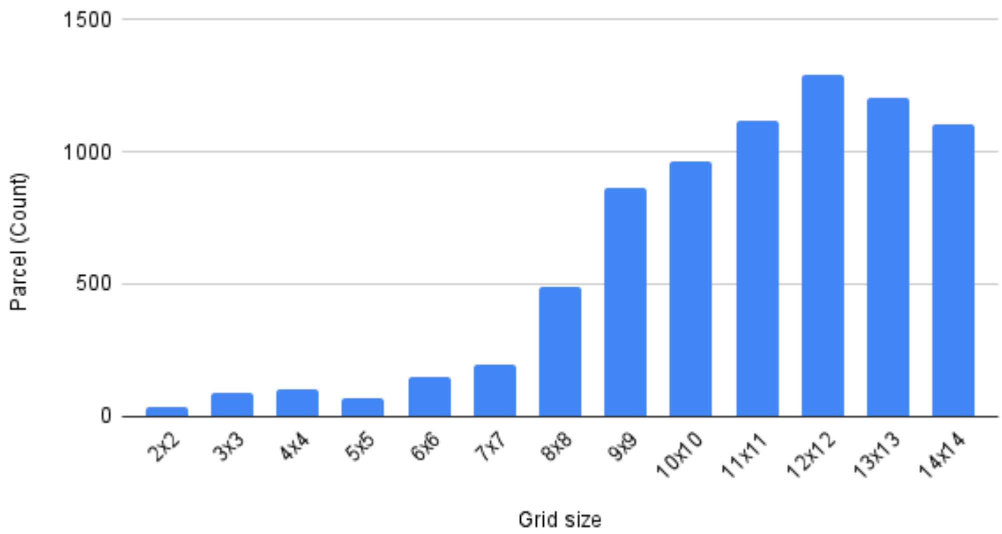


Figure 3.5: Regularity percentage vs grid size.

API calls were continuously made within smaller areas until we identified no additional new POIs. An initial search was performed within an area encompassing a radius of 1500 m (r = 1500) without specifying any particular search query, thereby aiming to capture any available POIs. Following this, we systematically reduced the search radius in decrements until reaching 250 m (r = 250). For each reduced radius, we repeated the API calls to ensure comprehensive coverage within the smaller segments of our study areas.

This approach enabled us to progressively gather POIs that might have been missed in broader searches. Consequently, by July 2021, we had successfully collected a total of 19,943 non-residential POI locations for Ankara. The full list of the number of ads listings per district is presented in Appendix B.2. Also, Appendix section B presents the distribution of collected ad listings for the largest districts, such as Çankaya, Keçiören, Yenimahalle, and Etimesgut, having disproportional ad listing counts in a bar chart.

Similarly, we replicated this methodology for the Kadıköy region in November 2023. By employing the same systematic approach of reducing the search radius and making numerous API calls, we obtained 15,527 non-residential data points for Kadıköy.

For the ad listings data, we undertook a comprehensive web scraping exercise to gather all online ad listings for Ankara from Hepsiemlak (accessed on 30 July 2019) (`https://www.hepsiemlak.com/`), one of Turkey's largest property ad listing websites, encompassing residential and non-residential properties. Hepsiemlak, initially established in 2006, has become a leading platform in the Turkish real estate market. We have conducted data collection from Hepsiemlak between June 2019 and July 2019. During this period, we managed to accumulate a total of 84,580 ad listings. Following a rigorous data cleansing process (the details are explained extensively in Section 3.4.3) we refined this dataset to 60,031 unique ads. The data cleansing involved procedures to remove duplicates, correct inconsistencies, and ensure the accuracy and reliability of the ad listings.

In addition to the data from Hepsiemlak, in November 2023, we extended our data collection efforts to include Kadıköy by using the Sahibinden (accessed on 29 November 2023) (https://www.sahibinden.com/). Sahibinden, established in 2000, is another prominent platform in Turkey that offers comprehensive listings for residential and commercial properties and various other categories and services. Like Hepsiemlak, Sahibinden has a significant user base and provides detailed listings.

Using our established methodology, we collected 5942 ad listings from Sahibinden for Kadıköy. After applying the same rigorous data cleansing techniques, this dataset was narrowed down to 4847 unique ads. This additional dataset reinforces the robustness of our study by providing a broader geographical scope and adding temporal depth to our data.

3.4.2 Land Registry Data

The difficulty of reaching the official government data is one of the main reasons we aim to acquire LUM solely by using openly and widely available point-based data. Nevertheless, we need ground truth data from the official sources to evaluate the results and compare the predicted LUM values.

Hence, we legally and ethically gathered 650,000 official parcel data in the form of shapefiles from the Turkish land registry website (accessed on 15 May 2020) (https://parselsorgu.tkgm.gov.tr) via various web automation and web scraping techniques on the cloud in a fair usage manner. We used Python's Selenium package to implement the scraper and deployed it on the Google Cloud. We utilized virtual machines with 2 GBs of RAM and 20 GBs of storage on the cloud and managed 100 servers for two months to make the data acquisition faster. We also used Google Cloud Storage to consolidate all the collected data.

The crawled data were packaged as compressed shapefiles containing various information about the land, including but not limited to the borders, area, and land type. These shapefiles were later used to calculate the ground truth results.

One of the vital information included in the official land registry data is the land type labels. After scrutinizing the labels, we found that the following labels can be used to infer the residential areas: home, apartment, housing, domicile, and villa.

Therefore, we labeled the areas associated with the above tags as residential and others as non-residential.

3.4.3 Data Pre-Processing

We encountered data quality problems with our geospatial data. Many data points appear to be mislocated or mislabeled in our collected dataset.

The descriptive statistics revealed that ads are being published in very close locations. For instance, in an area of approximately 100 m^2 located tens of real estates, indicating an issue of disguised missing data, which is defined as intentionally and systematically coding missing data as valid data values. In [34], a police station's coordinates were entered as the locations of some traffic accidents in a region, which appeared as a hot spot in the analysis. Similar to this problem, the same coordinates might be systematically used for different real estates.

To solve this problem, we first rounded the latitude and longitude decimals to a certain accuracy after experimenting with a different number of decimal places and considering the GPS signal sensitivity. We have rounded them to up to four decimal places, resulting in, on average, a 3.78 m change between the actual locations and the rounded latitude and longitude decimals. Table 3.1 shows the highest number of POI counts per latitude and longitude pairs rounded up to four decimal places.

Table 3.1: Number of real estate locations per rounded latitude and longitude.

Latitude	**Longitude**	**Real Estate Count**
39.9836	32.8218	201
39.9838	32.8217	153
39.9755	32.8281	112
39.9838	32.8218	107
40.0430	32.8998	97

As a result, we ended up with only 61,096 unique locations for Ankara and 4921 unique points for Kadıköy. However, we still observed that even rounded coordinates of the locations included several ads, which is questionable as it is very unlikely to see so many ads in a very limited area. Therefore, we manually checked these locations. We noticed that several real estate ads are located very close to a real estate agency, indicating that the coordinates of the ads belong to the real estate agency but not the advertised address in the text. For instance, Figure 3.6 presents a building where tens of real estate locations are advertised as the same building and having a real estate agency in the same building according to the Google Maps data.

To solve the issue of the disguised missing data problem, we finally used the Google Maps API to check whether there is any real estate agency around each rounded location. We discovered that at least one or more real estate properties are available in places with many real estate advertisements. We kept looking at the vicinity of places with many real estate accumulations until we did not discover any real estate agency nearby. In the end, we found that places with less than three real estate do not have any close real estate agencies within a radius of 250 m. That is why we assumed all the rounded locations having more than two real estates as disguised and discarded. After removing them, we ended up with 60,031 unique ad listings for Ankara and 4847 unique ad listings for Kadıköy.

We noticed that out of the 650,000 shapefiles in the land registry data, only 45,000 of them were unique. The reason why we have seen lots of duplicate shapefiles result from the small step size we defined in order not to miss any shapefiles. After eliminating the duplicates, we examined data quality issues. We noticed that some of the parcels that we downloaded have issues of conflicting with other parcels. That is, the parcels intersect with others.

Usually, one does not expect to see any intersection with parcels as each parcel represents someone's registered land, according to Turkey's land registry office. Therefore, after carefully examining some of the cases, we noticed that intersections occur in the case of a more general land type associated with the parcel whose areas are much larger than the other intersected parcel. Hence, we discarded the one

Figure 3.6: Dots represent the real estate located in the same building, and a small image located on the top left presents the real estate agency located in the same spot.

with the larger area and more general land type. As a result, we ended up with 43,000 shapefiles to continue our ground truth LUM calculation further.

Additionally, to ensure that we have enough data, as mentioned in the previous chapter, we applied the quadrats method and quadrat tests to eliminate fishnets without sufficient land registry data for ground truth LUM calculation. As a result, out of 1862 fishnets, 158 of them were discarded from the evaluation.

We encountered data limitations as well. First of all, we could only calculate LUM for some fishnet eyes. We could not calculate LUM because there is not a sufficient number of POIs in the fishnet to draw a Voronoi. There must be at least four points to draw a Voronoi. After examining the places with less than four POIs, we saw that green areas, wetlands, or agricultural areas mostly cover those areas.

In addition to that, the other reason could be the lack of point-based data, which we collected from an ad listing website for the sale/rent of houses and offices and Google Maps API for various shops, workplaces, and business locations with our best effort.

Lastly, we also saw that some areas do not have an official record of the registry on the land registry website. In Figure 3.7a, we depict all the available land registry data, but on the left of the figure, it is clear that there is either an apartment or a villa, but the official sources do not confirm.

Although it is an official data source and each parcel indicates registered land according to the land registry office, we discovered issues of collisions between parcels. In Figure 3.7b, it can be seen that one big shapefile overlaps with other small shapefiles. In general, we saw that the bigger shapefile

(a)

(b)

Figure 3.7: (**a**) An urban area covered by all the available official land registry shapefiles. (**b**) An example of shapefiles having overlaps with each other, indicating a problem with the official data source, overlapping area zoomed within the blue rectangle for better clarity.

usually has a generic label such as land or field. That is why we decided to discard the shapefiles with a bigger area in the case of an intersection.

3.5 Results

In this section, we present our findings and provide the evaluation results concerning different aspects, such as Local Moran's I and the POI intensity per fishnet. Additionally, we present and compare POI-based LUM identification outcomes from these two distinct urban settings.

Using the point-based data from Google Maps API and ad listings data to predict LUM, we present the LUM results on a map to further understand their distribution across Ankara city.

Figure 3.8 shows the areas in Ankara according to their LUM values, in which gray areas show not applicable results. The colored squares indicate the LUM values of the associated places.

According to the map, darker colors correspond to higher LUM values. Areas far from the city center typically have lower LUM values, while areas with higher LUM values often include both residential and non-residential activities. Gray areas represent fishnets with insufficient POI data.

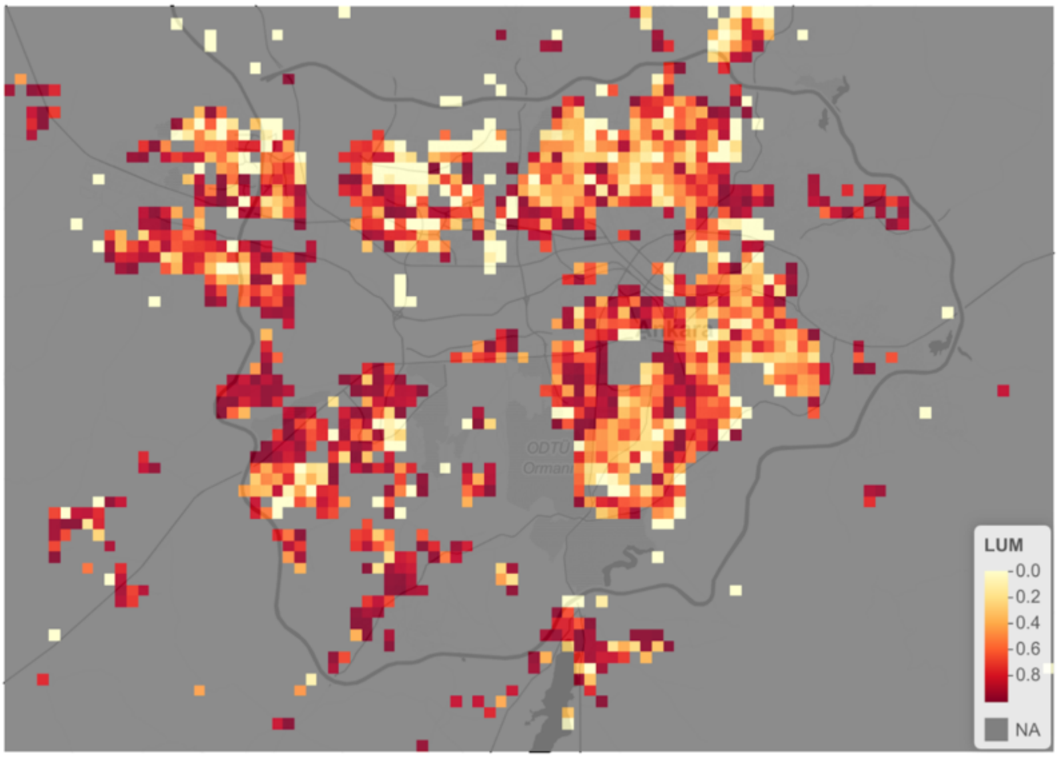


Figure 3.8: Predicted LUM values are presented on a map, where darker colors indicate higher LUM values.

To evaluate our LUM findings, we first observed the overall results with respect to the ground truth. Rather than directly comparing the predicted and ground truth LUM values as continuous numbers, we classified the areas into two categories:

- **Mixed Land Use (LUM $\geq$ 0.5)**: Areas with a high degree of land use diversity.
- **No Mixed Land Use (LUM < 0.5)**: Areas lacking significant land use diversity.

To determine the cutoff point, we used stratified sampling and Receiver Operating Characteristic (ROC) analysis on a subset of the regions, ultimately setting the cutoff at 0.5, which yielded an Area Under the Curve (AUC) of 0.57. To determine the cutoff point, we first set 20% of the regions we managed to identify LUM values aside based on a stratified sampling where we obtain samples from places having [0–0.1), [0.1–0.2), [0.2–0.3), ..., [0.9–1] LUM values. We then tested different thresholds on that portion via drawing ROCs for each of the cutoff points, which are depicted in Figure 3.9. Afterward, we evaluated the classification accuracy, precision, recall, and F1-score based on this classification.

Additionally, we also set two baselines to compare the results—one for all the fishnets with no mix representing homogeneous single land use and one for all the fishnets having mixed land use.

Our overall classification results are presented in Table 3.2. Specifically, our approach achieved an accuracy of 51%, a precision of 48%, a recall of 69%, and an F1-score of 57%. These metrics reflect the performance of our binary classification model.

Additionally, we observed the classification metrics based on the local spatial autocorrelation measures, which brings about landscape fragmentation in comparison to landscape metrics [35]. It finds spatial dependency of features in space to imply landscape heterogeneity. In this study, we used one of the local spatial autocorrelation methods, Local Moran's I (LMI), [36] to identify the spatial pattern of different fishnet LUM values.

Table 3.2: Overall accuracy, precision, recall, and F1-score.

	Accuracy	**Precision**	**Recall**	**F1**
Overall	0.51	0.48	0.69	0.57
Baseline all 0 s	0.57	1.00	0.57	0.72
Baseline all 1 s	0.43	0.00	NA	NA

Local Moran's I fundamentally has two interpretations. On the one hand, it lets you find hot spots. On the other hand, it identifies outliers. Thanks to LMI, we detected local clusters and local spatial outliers to analyze our LUM classification findings further, depending on the Local Moran's I index of each fishnet's LUM value.

As Tobler's first law of geography states, spatial distance affects the relationship among things. Hence, ignoring spatial distance among fishnets will make the results biased. Therefore, we calculate the LMI for each fishnet based on the assigned LUM values.

In Table 3.3, we presented the classification accuracy, precision, recall, and F1-score for different significance ranges of the LMI analysis. The main motivation was to investigate whether the scores increase proportionally to the significance level, and it is confirmed that when the significance of LMI increases, our classification scores also become better. For instance, for p-values less than and equal to 0.01, we saw 71% classification accuracy, 63% precision, 92% recall, and 75% F1-score, with the baseline of all 0 s being 58%.

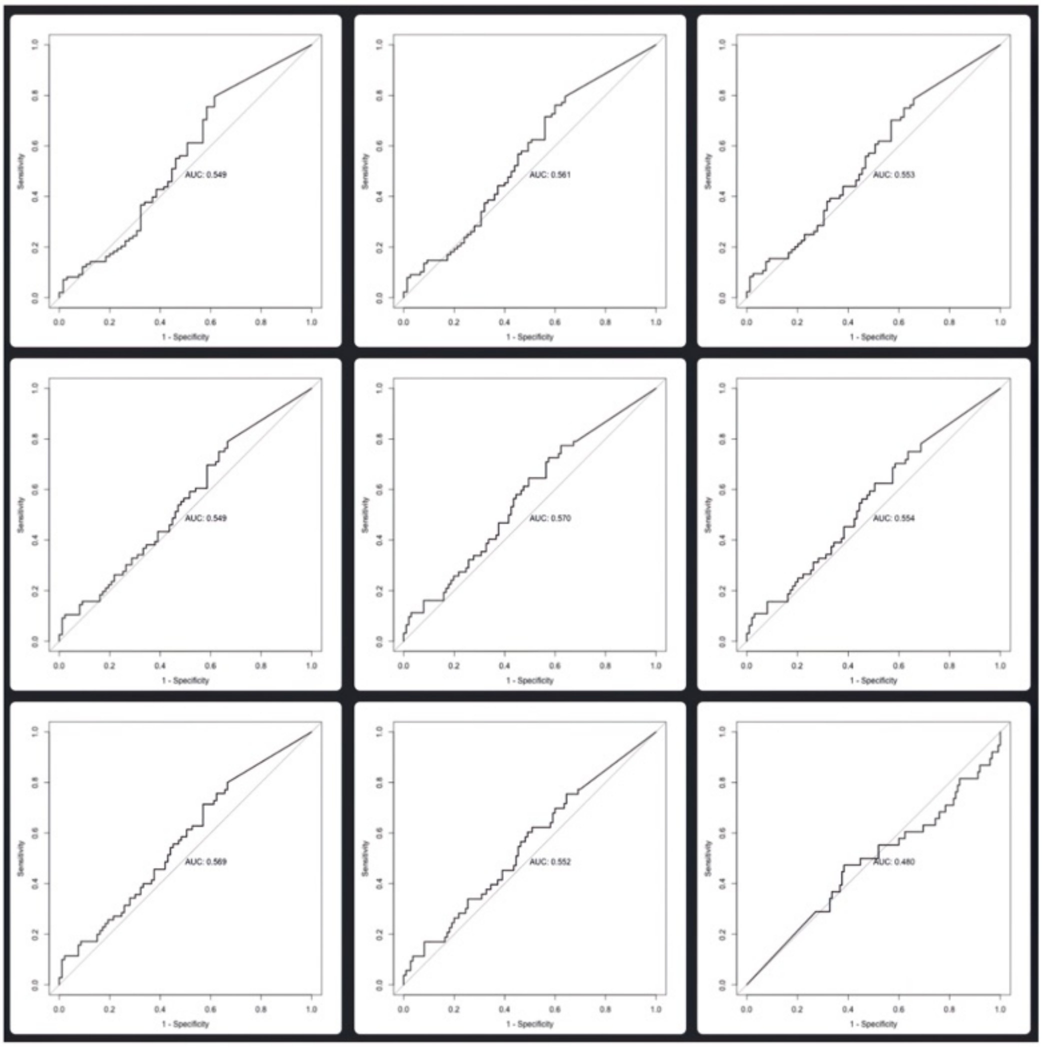


Figure 3.9: AUC curves for different cutoff points. The top-left AUC plot presents cutoff = 1 and the bottom-right plot presents cutoff = 0.9.

Table 3.3: Accuracy, precision, recall, and F1-score for different significance ranges of the Local Moran's I findings along with the number of places residing in each significance level and all 0 s baseline accuracy results for comparison.

Sig. Level	Accuracy	Precision	Recall	F1	# of Places	Baseline Acc.
$p <= 0.01$	0.71	0.63	0.92	0.75	28	0.58
$p <= 0.02$	0.61	0.46	0.92	0.62	38	0.58
$p <= 0.03$	0.51	0.34	0.92	0.50	49	0.61
$p <= 0.04$	0.49	0.31	0.92	0.46	55	0.61
$p <= 0.05$	0.48	0.27	0.92	0.41	65	0.59
$p <= 0.10$	0.45	0.19	0.92	0.31	96	0.57
$p <= 0.20$	0.41	0.13	0.92	0.23	135	0.57
$p <= 0.30$	0.39	0.15	0.70	0.25	161	0.56

Additionally, we look at the classification accuracy, precision, recall, and F1-score concerning the different numbers of POIs residing on fishnets. In order to specify the bins, we first started with equal-depth binning and then tweaked the ranges to make sure that distributions of LUM values in each bin are different from the others. We used a non-parametric Mann–Whitney U test with a p-value ≤ 0.05 to make sure that they were not coming from the same distributions. As a result, we created five bins depending on the POI counts with LUM values coming from different distributions.

In Table 3.4, the results are presented for different bins such as [4–5], [6–11], [12,17], [18–28], and [29–60]; as a result, we found that the accuracy of finding out whether a fishnet has mix land use or no-mix land use increases up to 65% followed by the 67% precision, 89% recall, and 77% F1-score when the POI count is between 29 and 60. In other words, considering that we are working with fishnets with an area of 0.2655 km^2, if the POI density ranges between 109 POI/km^2 and 226 POI/km^2, we can reach the highest mixed land use classification accuracy of 77% and F1-score of 80%.

Table 3.4: Accuracy, precision, recall, and F1-score for different POI bins along with the number of places residing in each bin and all 0 s baseline accuracy results for comparison.

Bins	**Accuracy**	**Precision**	**Recall**	**F1**	**# of Places**	**Baseline Acc.**
[4–5]	0.41	0.34	0.55	0.42	113	0.66
[6–11]	0.42	0.26	0.51	0.35	156	0.54
[12–17]	0.51	0.53	0.55	0.54	107	0.65
[18–28]	0.57	0.56	0.78	0.65	175	0.48
[29–60]	0.77	0.74	0.88	0.80	111	0.56

Finally, we compared the results from Ankara to Kadıköy with the same assumption of having between 29 and 60 points within a fishnet eye, we find the following results presented in Table 3.5.

Table 3.5: Accuracy, precision, recall, and F1-score for Ankara and Kadıköy, fishnets having [29–60] POIs, along with the number of places and all 0 s baseline accuracy results for comparison.

Bins	**Accuracy**	**Precision**	**Recall**	**F1**	**# of Places**	**Baseline Acc.**
Ankara [29–60]	0.77	0.74	0.88	0.80	111	0.56
Kadıköy [29–60]	0.78	0.77	0.91	0.83	18	0.66

Comparing the results from Ankara to those from Kadıköy, we observe that the latter demonstrates a slightly better accuracy and F1-score. This is likely due to Kadıköy's higher density of mixed-use developments which aligned well with our POI-based methodology.

3.6 Discussion

Dynamically identifying LUM using POI data is a non-trivial and challenging task. This study demonstrates that point-based data, combined with Voronoi triangulation, can effectively classify land use into mixed and non-mixed categories. As shown in Table 3.2, our methodology achieved classification metrics of 51% accuracy, 48% precision, 69% recall, and an F1-score of 57%.

Considering local spatial autocorrelation, we observed an increase in classification accuracy for fishnets with LUM values similar to their neighboring areas. For fishnets displaying high-high or low-low LUM value clusters, the classification accuracy reached 71%, with 63% precision, 92% recall, and a 75% F1-score, according to Local Moran's I with a significance level of p-value less than or equal to 0.01.

Additionally, analyzing POI intensity ranges revealed that the number of POIs within a fishnet is critical for accurate LUM classification. When the number of POIs ranged between 29 and 60, the accuracy improved to 78%, with 77% precision, 91% recall, and an F1-score of 83%. This confirms that higher POI densities lead to better classification performance.

Comparing results from Ankara to Kadıköy, we noted that the latter had slightly better accuracy and F1-score due to its higher density of mixed-use developments. This finding underscores our methodology's capability to classify land use effectively across different urban contexts.

These findings emphasize the robustness and applicability of our POI-based approach for LUM classification in diverse urban settings. The results from Kadıköy, in particular, underscored the effectiveness of our methodology in a densely populated urban environment with a high density of mixed-use developments. This further validates that a minimum POI density of approximately 109 POIs per km^2 is optimal for predicting an area's LUM classification accurately.

Moreover, the method's adaptability allows for real-time updates and dynamic tracking of urban land use changes, essential for responsive urban planning. The high accuracy observed in densely populated areas like Kadıköy underscores the model's potential for application in similar urban settings.

Future research could expand on this by integrating additional data sources, such as social media check-ins or transportation data, to enhance the model's robustness and applicability. Additionally, exploring the method's applicability in different urban contexts can provide insights into its scalability and generalizability.

3.7 Conclusions

This study presents an effective method for identifying LUM and classifying land use into mixed and non-mixed categories in urban areas using point-based geospatial data. By employing Voronoi triangulation and an entropy-based LUM formula, we demonstrated that it is possible to dynamically determine LUM without requiring extensive fieldwork or human intervention. The methodology was validated in two distinct urban settings: Ankara and Kadıköy. The results underscore the robustness and adaptability of our approach, achieving up to 78% accuracy and an F1-score of 83% in Kadıköy, where the density of mixed-use developments is higher.

Crucial findings from this study indicate that the number and density of POI within a given area significantly influence the accuracy of LUM classification. Higher accuracy was observed in fishnets with a POI density ranging between 29 and 60, emphasizing the importance of sufficient and representative data for reliable LUM classification.

The comparative analysis revealed that different urban environments could yield varying performance levels. Kadıköy's slightly higher accuracy rates than Ankara highlight the method's effectiveness in densely populated areas with complex urban mixes.

Our study offers valuable insights for urban planners and policymakers, supporting efficient resource allocation and the planning of sustainable, mixed-use urban environments. Although the focus was on residential and non-residential land use types, future research could expand this approach to include additional land use categories such as recreational, commercial, and transportation. Additionally, integrating social media data could provide further insights into human interactions with urban spaces, enhancing the understanding and prediction of mixed land use dynamics.

Overall, our study's primary contribution is the development of an efficient, automated method for identifying LUM in urban areas using publicly available point-based data, eliminating the need for fieldwork and human intervention. This methodology represents a significant advancement in urban land use studies, providing a scalable and efficient method for modern city planning.

CHAPTER 4

LANDUSEMIX: AN OPEN SOURCE PYTHON PACKAGE FOR CALCULATING LAND USE MIX

LUM assessment is a fundamental aspect of urban planning, influencing key factors such as sustainability, livability, and social interaction within cities. Despite the critical role LUM plays in shaping urban environments, traditional tools and methods for its calculation often lack the flexibility, scalability, and user-friendliness required for modern, large-scale analyses. To address these challenges, we introduce landusemix, an open-source Python package designed to calculate LUM using two widely recognized indices: the Entropy Index and the Herfindahl-Hirschman Index (HHI).

This chapter provides a comprehensive overview of the landusemix package, outlining its motivation, architecture, key features, and contributions to urban planning and Geographic Information Systems (GIS) research. By integrating robust computational tools, landusemix offers an efficient and scalable solution for analyzing land use diversity and concentration, enabling researchers, urban planners, and policymakers to make data-driven decisions more accurately and easily. The chapter is structured to guide the reader through the rationale behind the package's development, its technical architecture, key features and functionalities, illustrative examples of its application, omissions, limitations, and broader implications for urban studies.

4.1 Motivation and significance

Urban planning and development play critical roles in shaping the sustainability and livability of cities. One significant aspect of urban planning is integrating diverse land uses within a geographic area. A well-balanced land use mix (LUM) can enhance urban diversity, improve service accessibility, reduce transportation needs, and foster social interaction [37]. Consequently, accurately measuring LUM is essential for evaluating urban diversity and sustainability.

Urban areas are characterized by various land uses, including residential, commercial, industrial, and recreational areas. The integration and distribution of these different land uses significantly impact urban sustainability, quality of life, and overall urban functionality. Measuring LUM provides valuable insights into urban diversity and informs urban planning and policy-making [38]. Traditional methods of calculating LUM often involve manual processes or lack flexibility and scalability, presenting a need for more advanced tools.

The concept of LUM has been explored extensively in the literature, with several indices proposed to quantify it [39, 40]. The Entropy Index and the Herfindahl-Hirschman Index (HHI) are the most commonly used [41]. They provide alternative methods to calculate LUM, capturing different aspects of land use concentration. The Entropy Index measures diversity by indicating how evenly land uses are distributed, while the HHI focuses on concentration, highlighting dominance or lack thereof. These indices are widely recognized and frequently used in land use studies, offering both complementary and comprehensive analysis. Previous studies have utilized these indices to address various urban planning challenges, such as evaluating land use diversity [42, 43], monitoring the balance of commercial and residential areas [41, 44], and assessing the impacts of urban development policies [45, 46].

In recent years, integrating geospatial technologies with Python programming has significantly advanced the field, offering powerful tools for land use classification, suitability analysis, and change modeling. Previous tools and methods for calculating these indices have limitations in terms of usability and scalability. For instance, some GIS software includes functions for land use analysis, but these often require extensive manual input and are not explicitly tailored for LUM calculations. Additionally, the `landscapemetrics` is a R package offers comprehensive landscape metrics for spatial analysis [47]. Similarly, `pylandstats` provides various diversity indices for ecological studies [48]. Also, [49] offers a comprehensive approach to analyzing urban sprawl and morphology metrics using OpenStreetMap data. The package employs Kernel Density Estimation (KDE) to estimate the probability density function of different land uses over a spatial grid, utilizing weighted and logarithmic transformations to enhance the density estimations. However, the `landusemix` Python package focuses on LUM identification and gathers indices that help in this specific analysis. By integrating LUM identification indices, `landusemix` simplifies the calculation of LUM metrics, making it an essential tool for urban planners and GIS researchers seeking practical and targeted solutions.

To address this gap, we introduce `landusemix` [50, 51], a Python package specifically designed to calculate LUM using the Entropy Index and the HHI. Offering both indices in the `landusemix` package allows researchers to choose the measure that best fits their specific analysis needs. This tool aims to provide GIS researchers, urban planners, and policymakers with a researcher-friendly, efficient, and flexible means for measuring the diversity and concentration of land use. It provides easy-to-understand documentation and integrates with familiar libraries like `pandas` [52, 53], `geopandas` [54], and `rasterio` [55]. By leveraging these libraries, the package is designed to efficiently handle large datasets and perform bulk calculations without requiring heavy GIS applications, which are often less flexible for large-scale or repetitive LUM analysis. Additionally, it supports multiple data formats (CSV, Shapefiles, GeoJSON, raster) and offers a modular design for customization. The package streamlines the calculation of LUM indices, reducing the need for advanced GIS tools and enabling large-scale or repetitive analysis with less manual intervention, making it ideal for integration into broader research workflows. Accurately measuring LUM is critical for evaluating urban diversity, which influences various socio-economic and environmental factors [56]. The package facilitates the systematic analysis of land use patterns, providing precise and reproducible metrics that can be incorporated into broader urban studies. The package is designed to be installed and used straightforwardly, supporting input in land use area data and returning calculated indices for further analysis or visualization.

Researchers and developers can easily integrate `landusemix` into their existing workflows. An example usage scenario involves loading land use data, instantiating the LandUseMixIndices class with this data, and then calling the methods to calculate the Entropy Index and HHI. Detailed documentation and examples are provided to help users get started quickly.

The remainder of this chapter is structured as follows: Section 2 provides a detailed software description of the `landusemix` package, including its architecture and functionalities. Section 3 discusses the omissions and limitations. Section 4 presents illustrative examples to demonstrate the practical usage of the package. Section 5 discusses the impact of the package on urban planning and GIS research. Finally, Section 6 concludes with a summary of findings and proposes future research directions.

4.2 Software Description

4.2.1 Overview

`landusemix` is a Python package designed to calculate LUM using the Entropy Index and the HHI. This package provides GIS researchers, urban planners, and policymakers tools to efficiently measure and analyze land use diversity and concentration. By leveraging Python's powerful data processing libraries, `landusemix` provides a practical and researcher-friendly solution for LUM calculations, allowing for efficient and scalable analysis without requiring extensive manual setup common in traditional GIS software.

4.2.2 Architecture

The `landusemix` package is organized into several key components. Fig. 4.1 below is an image illustrating the software architecture: Each component is designed to perform specific functions related to LUM calculations:

- **Core Module (`indices`)**: This module contains the main class and functions for calculating the Entropy Index and HHI. Implementation of this module is shared in Appendix C for reference.
- **Utilities Module (`utils`)**: Provides additional utility functions for data handling and preprocessing.
- **Data Module (`data`)**: Includes sample land use data for testing and experimentation.
- **Tests Module (`tests`)**: A suite of unit tests to ensure the correctness and reliability of the package.

4.2.3 Key Features and Functionalities

The `landusemix` package offers the following key features:

- **Entropy Index Calculation**: Computes the Entropy Index, a measure of land use diversity [57], using the formula:

$$\text{ENT} = -\frac{\left[\sum_{i=1}^{k} P_i \ln\left(P_i\right)\right]}{\ln(k)} \tag{4.1}$$

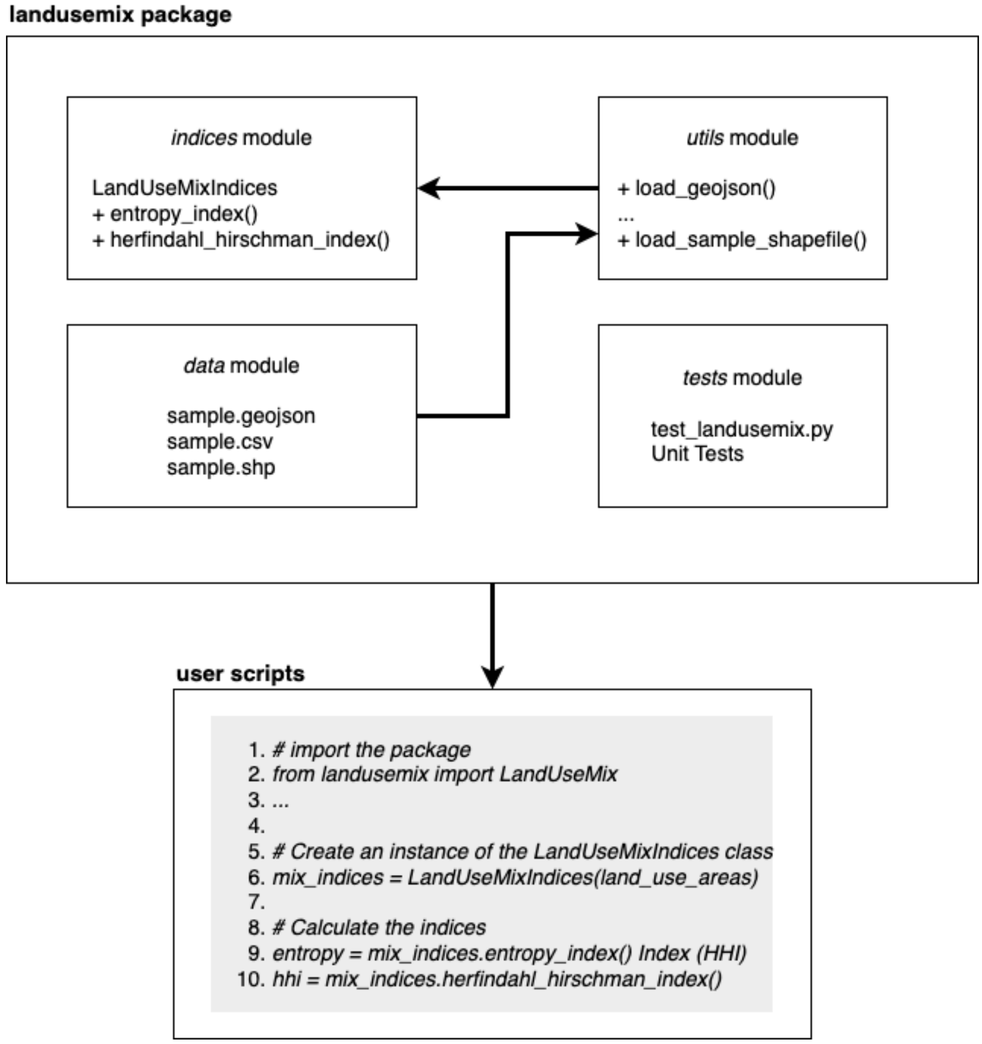


Figure 4.1: Architecture of the landusemix Python package

where P_i represents the proportion of the total area occupied by the i-th land use type, and k is the total number of different land use types. The Entropy Index ranges from 0 (no diversity) to 1 (maximum diversity), indicating the diversity of land use types [16].

- **HHI Calculation**: Computes the HHI, a measure of land use concentration, using the formula:

$$\text{HHI} = \sum_{i=1}^{k} (100 \times P_i)^2 \tag{4.2}$$

where P_i represents the proportion of the total area occupied by the i-th land use type. The HHI ranges from 0 (many small equally-sized areas) to 10,000 (one single area), indicating the concentration of land use types [41].

- **Utility Functions**: Includes functions for data preprocessing, such as loading of different types of data formats such as CSV, Shapefiles, GeoJSON, and raster data.

- **Documentation and Examples**: Provides detailed documentation and practical usage examples to assist users in getting started quickly.

4.3 Omissions and Limitations

While the `landusemix` package offers robust methods for calculating LUM using the Entropy Index and HHI, there are several limitations to consider:

- **Index Scope**: The package is limited to two indices (Entropy Index and HHI). Although these indices are widely used, other indices, such as the Dissimilarity Index [58], the Gini Coefficient, the Atkinson Index [59], or the Clustering Index [60], could provide additional insights into the LUM. These indices often require specific assumptions about spatial arrangement and social factors, which may not apply universally to all study areas or contexts. Therefore, they have been omitted from this package to maintain a broader applicability.
- **Spatial Context**: The current version of the package does not consider the spatial arrangement of land use types. The indices are based solely on the proportion of the land use areas without considering their spatial distribution. Considering the spatial distribution could provide valuable information for urban planning contexts. However, some other indices, such as the Atkinson Index, Clustering Index, and Dissimilarity Index, do consider the spatial context. Still, they have been omitted in this version to keep it generic enough for researchers.
- **Visualization Tools**: While the package provides accurate calculations, it lacks built-in visualization tools for presenting LUM results. Users need to rely on external libraries or software to visualize their findings.

Future versions of `landusemix` could incorporate several enhancements to address the current limitations by including additional indices and integrated visualizations.

4.4 Illustrative Examples

4.4.1 Installation

You can use `pip` to install the `landusemix` package. The following command installs the package and its dependencies:

```
pip install landusemix
```

4.4.2 Example Usage

Here, we provide an example of using the `landusemix` package to calculate the Entropy Index and HHI for a given set of land use areas.

Importing the Package

First, we import the necessary modules from the package:

```
from landusemix import LandUseMixIndices
from landusemix.utils import *
```

Loading Data

We can load the land use data from a dictionary or other compatible format. For this example, we define a dictionary representing different land use areas in square meters:

```
land_use_areas = {
    'residential': 5000,
    'commercial': 3000,
    'industrial': 2000,
}
```

You can also load data via the package's Utils functions. Figure 4.2 depicts a shapefile with three different land use types.

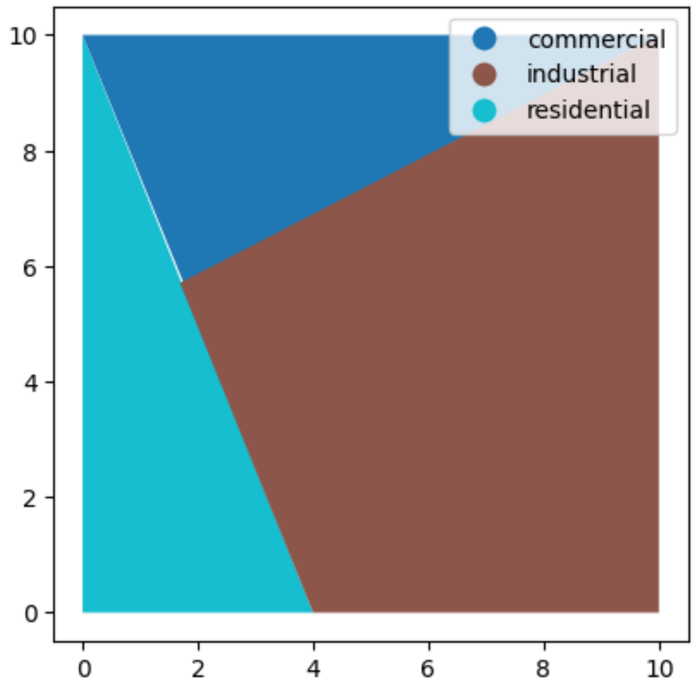


Figure 4.2: An example of a shapefile having three different land use types.

We can load the shapefile and calculate areas as such:

```python
shp_gdf = load_shapefile('landusemix/data/shapefiles/multiple.shp')
shp_gdf['area'] = shp_gdf.geometry.area
land_use_areas = shp_gdf.groupby('use')['area'].sum().to_dict()
print(land_use_areas)
# {'commercial': 21.46, 'industrial': 58.67, 'residential': 20.0}
```

Calculating Indices

Next, we create an instance of the `LandUseMixIndices` class and use its methods to calculate the Entropy Index and HHI:

```python
# Create an instance of the LandUseMixIndices class
mix_indices = LandUseMixIndices(land_use_areas)

# Calculate the Entropy Index
entropy = mix_indices.entropy_index()
print(f"Entropy Index: {entropy}") # Entropy Index: 0.88

# Calculate the HHI
hhi = mix_indices.herfindahl_hirschman_index()
print(f"Herfindahl-Hirschman Index: {hhi}")
# Herfindahl-Hirschman Index: 4291.41
```

4.4.3 Results Interpretation

The output of the above code provides the calculated Entropy Index and HHI for the given land use areas. The higher the Entropy Index, the more diverse the LUM; conversely, a higher HHI indicates a higher concentration of specific land use types.

The calculated values for the Entropy Index and HHI provide meaningful insights into land use diversity and concentration regarding LUM. For instance, an Entropy Index of 0.88 indicates a high diversity of land use types, suggesting an even distribution. Conversely, an HHI value of 500 indicates moderate concentration, with some land use types potentially dominating the area. Values closer to 0 for either index indicate lower diversity or higher concentration, respectively.

4.5 Impact

The `landusemix` package advances the tools available for urban planning and GIS research. This package facilitates a deeper understanding of urban diversity and its implications for sustainability and livability by providing a researcher-friendly and efficient means to calculate LUM indices.

4.5.1 Contributions to Urban Planning

Accurate and efficient measurement of LUM is essential for several reasons in urban planning:

- **Informing Policy Decisions**: The insights gained from analyzing the LUM can guide policy decisions to promote balanced urban development and enhance accessibility to essential services. These indices can be applied within specific boundaries or zones, using additional data, such as shapefiles or GeoJSON files, to generate localized insights. This approach has proven effective in real-world applications, as demonstrated in our recent work on identifying land use mix within fishnet grids using point-based geospatial data [61]. The HHI can identify areas with a significant concentration of specific land uses, which may indicate potential issues with service accessibility. For instance, high HHI values in residential areas might highlight a lack of nearby commercial or recreational facilities. Researchers can use these insights to make data-driven recommendations for zoning changes or new service allocations to improve accessibility.
- **Evaluating Urban Diversity**: By quantifying the diversity and concentration of land use, planners can identify areas requiring interventions to improve functionality and livability. The Entropy Index can help assess urban diversity by indicating how evenly different land uses are distributed within a geographic area. High entropy values suggest diverse land use, often associated with sustainable urban environments. Planners can use these insights to promote mixed-use development, enhancing overall urban sustainability.
- **Supporting Sustainable Development**: Understanding land use patterns and their impact on urban sprawl, transportation needs, and resource allocation helps promote sustainable urban growth. Both indices can provide insights into land use patterns that foster or hinder social interactions. Urban planners can design strategies to encourage community well-being by analyzing areas with diverse or concentrated land uses. For instance, lower HHI and higher entropy values could support the creation of public spaces that facilitate social interactions and community cohesion.

The `landusemix` package empowers urban planners with reliable metrics that can be easily integrated into various planning processes, from initial assessments to long-term strategic planning.

4.5.2 Contributions to GIS Research

The role of GIS in urban studies is critical, as it allows for the spatial analysis and visualization of land use patterns. The `landusemix` package enhances GIS research in the following ways:

- **Enhanced Data Analysis**: The package provides practical tools for quantitatively analyzing land use data, enabling large-scale and flexible analysis that can be easily integrated into further data mining and machine learning methods, offering advantages over traditional GIS tools.
- **Reproducible Research**: The package promotes reproducible research by offering standardized methods for calculating LUM, ensuring consistency and accuracy in urban studies.

GIS researchers can leverage the capabilities of `landusemix` to perform advanced analysis of land use diversity, contributing to a broader understanding of urban dynamics.

4.6 Conclusions

In this chapter, we introduced `landusemix`, a Python package developed to calculate LUM using the Entropy Index and the HHI. Accurate measurement of LUM is essential for evaluating urban diversity and sustainability, and the `landusemix` package provides a reliable, researcher-friendly, and efficient tool for this purpose.

We discussed the motivation and significance of the package, highlighting its importance in urban planning and GIS research. By simplifying the calculation of key LUM indices, `landusemix` reduces the potential for human error and lessens the dependence on heavy GIS applications such as ArcGIS and QGIS. The package is designed to streamline and scale analysis, making it possible to efficiently handle larger datasets and perform repetitive calculations more easily, thanks to its integration with efficient libraries such as `pandas`, `geopandas`, and `rasterio` and allowing the integration the results for further analysis within the Python ecosystem.

We also discussed the broader impact of the package on urban planning and GIS research, emphasizing its contributions to policy decision-making, sustainable development, and comprehensive urban studies.

While `landusemix` offers robust features, there are still areas for improvement, such as incorporating additional indices, accounting for spatial context, and integrating visualization tools. These enhancements could provide a more comprehensive assessment of the LUM and cater to more advanced user needs.

In conclusion, the `landusemix` package is a valuable resource for researchers and practitioners in urban planning and GIS, providing essential tools for analyzing and understanding land use diversity and its implications.

CHAPTER 5

UTILIZING LAND USE MIX TO ASSESS TEMPORAL EARTHQUAKE VULNERABILITY IN URBAN AREAS

Urban areas are dynamic environments where land use and population density can fluctuate significantly depending on the time of day, day of the week, or specific events. These temporal variations in how space is used directly influence the vulnerability of urban populations to natural disasters, particularly earthquakes. Understanding the relationship between land use mix (LUM) and these temporal patterns of population distribution is crucial for developing effective disaster preparedness and response strategies.

This chapter explores the application of LUM analysis to assess urban vulnerability, focusing on how vulnerability changes at different times of the day and week. By leveraging the patterns associated with different land use types—such as residential, commercial, recreational, and institutional—we aim to identify urban areas that may be more or less vulnerable to earthquake damage depending on when an event occurs. This analysis is critical for urban planners and policymakers tasked with mitigating disaster risk and enhancing the resilience of urban populations.

5.1 Motivation

Traditional vulnerability assessments are flawed because they treat population distribution as static, overlooking the reality that urban environments change over time. Recent research definitively challenges this approach by incorporating dynamic population patterns. [62] developed a framework integrating human mobility data to assess community vulnerability, and their findings revealed significant temporal changes in vulnerable population distributions.

Similarly, [63] created vulnerability maps for different time periods, considering both working and non-working hours. [64] proposed a method to estimate population distribution by the time of day, week, and season, accounting for various population groups, including shoppers and tourists, who are often neglected. Their study demonstrated how the population affected by a hazard can vary significantly depending on timing. [65] examined how land use changes impact social vulnerability, finding that urban expansion generally increases vulnerability due to higher population density in residential areas.

These studies highlight the importance of considering temporal dynamics in vulnerability assessments. Based on these, we assume that different land use types exhibit distinct patterns of use depending on the time of day and whether it’s a weekday or weekend.

For instance:

1. Residential Areas: Typically experience low activity during weekday daytime hours as residents are often at work or school but see increased activity in the evenings and on weekends [66, 67].
2. Commercial Areas: See high activity during weekday daytime hours, particularly in central business districts (CBDs) and retail zones, with varying levels of activity in the evenings and weekends depending on the type of commercial activity.
3. Recreational Areas: Attract moderate to high activity during lunch breaks, after work hours, and especially on weekends, when parks, sports facilities, and cultural venues are most visited [68].
4. Institutional Areas: Such as schools and offices, experience high activity during weekday daytime hours but are generally quiet in the evenings and on weekends, except during special events [69].

Understanding these temporal dynamics is essential for accurately assessing earthquake vulnerability. By analyzing LUM and making inferences about where people are likely to be located at different times, we can better predict the potential impact of an earthquake, depending on when it occurs. This approach provides a more detailed understanding of urban vulnerability, allowing for more targeted and effective disaster preparedness measures.

5.2 Problem Definition

The main problem addressed in this chapter is identifying urban areas that could be particularly vulnerable to earthquakes at different times due to the temporal variations in land use and population distribution. Traditional vulnerability assessment methods often fail to account for these temporal fluctuations, potentially overlooking critical periods of heightened risk.

To address this problem, we use LUM analysis to infer the likely temporal distribution of people across different land use types. By calculating LUM and identifying dominant land use types, we can make educated assumptions about when and where people are most likely to be present in an area. This analysis is further informed by the Land Usability Map provided by the municipality, which offers additional context for understanding how land use patterns might influence vulnerability. This chapter aims to provide a more comprehensive assessment of urban earthquake vulnerability by focusing on these temporal aspects.

5.3 Study Area

In this chapter, we aim to evaluate the identification of urban areas that could be particularly vulnerable to earthquakes at different times due to temporal variations, and we picked the Kadıköy and Avcılar districts of Istanbul, Türkiye as our study area, which is depicted in Figure 5.1. Istanbul has a rich cultural and historical heritage, located at the intersection of Europe and Asia. However, its location also puts it in a seismically active region, making it vulnerable to earthquakes. Understanding and reducing earthquake risks in urban areas like Kadıköy and Avcılar is vital to increasing the resilience of the population and infrastructure.

5.3.1 Kadıköy

Kadıköy, located on the Anatolian side of Istanbul, is a densely populated district with a diverse mix of residential, commercial, and cultural spaces. Its proximity to the Marmara Sea and the Bosphorus Strait makes it subject to geological vulnerabilities, and understanding the seismic characteristics of this area is essential for developing effective risk mitigation strategies.

5.3.2 Avcılar

Avcılar, situated on the European side of Istanbul, is characterized by a mix of residential and industrial zones. As part of the greater metropolitan area, Avcılar's susceptibility to earthquake hazards demands careful examination. The geological features and urban infrastructure in Avcılar contribute to the complexity of earthquake risk assessment in this district.

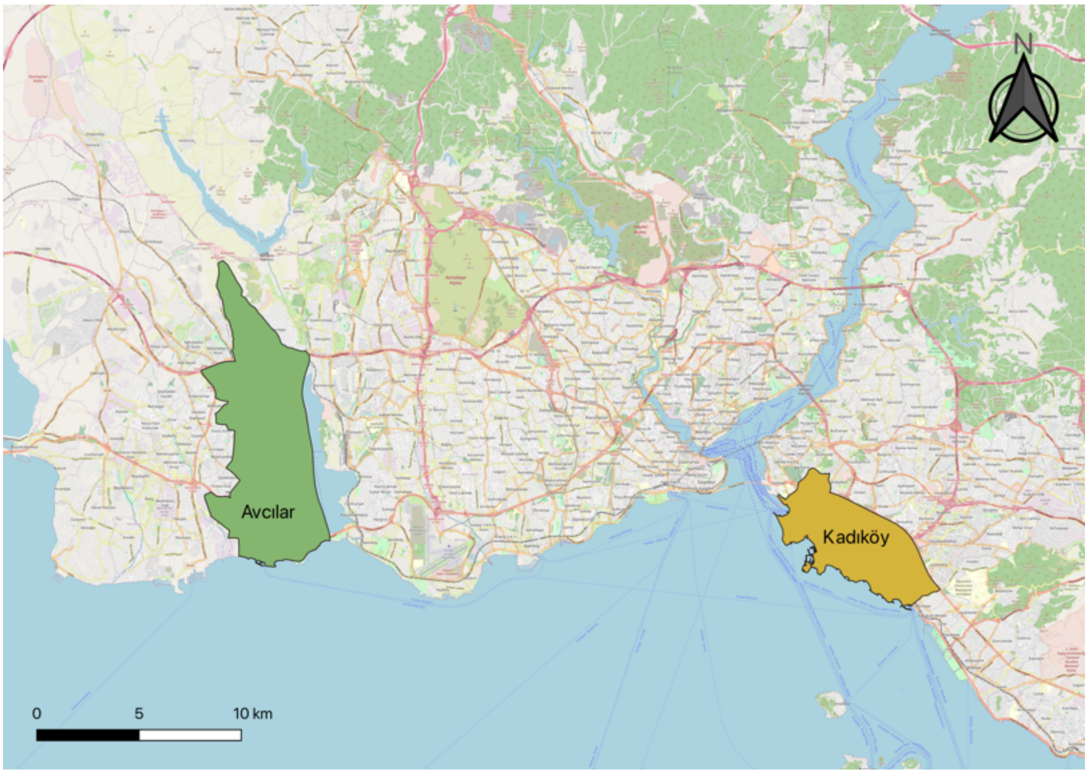


Figure 5.1: Area of interest. The green area on the left is Avcılar, and the yellow area on the right is Kadıköy.

5.4 Methodology

This section presents the methodology for assessing temporal earthquake vulnerability using LUM analysis. The process depicted in Figure 5.2 begins with data collection and preparation, followed by calculating LUM and identifying dominant land use types using the chi-square test. Based on the dominant land use type and the mixed-use type interpretation, the results are integrated with the Land Usability Map to assess urban vulnerability comprehensively.

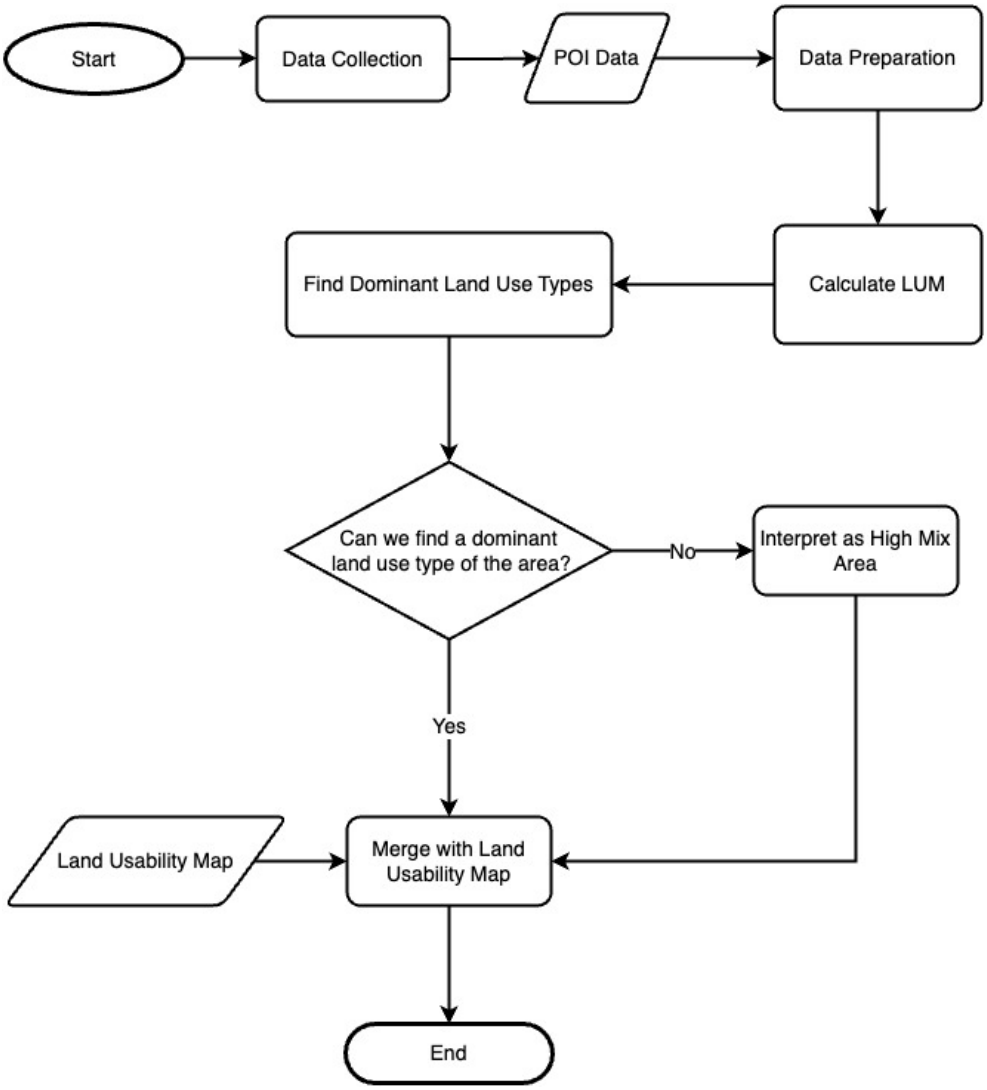


Figure 5.2: Flowchart of the methodology for assessing temporal earthquake vulnerability using LUM analysis. The process includes data collection, LUM calculation, identification of dominant land use types using the chi-square test, and integration with the Land Usability Map to assess urban vulnerability comprehensively.

5.4.1 Calculate LUM

The LUM calculation in this section extends the methodology outlined in Chapter 3 by incorporating a more detailed classification of land use types. Whereas the previous analysis focused on classifying land uses into binary categories of mixed land use (LUM $\geq$ 0.5) versus no mixed land use (LUM $<$ 0.5), this section calculates LUM based on four distinct types: residential, commercial, recreational, and institutional.

As described in Section 3.3, the general process of LUM calculation involves dividing the study area into fishnet cells, aggregating POI data within each cell, and drawing Voronois based on the POI locations. Fishnets created for two study areas are depicted in Appendix A. However, the following modifications were made to accommodate the more detailed land use classification in this chapter:

1. **Expanded Land Use Categories:** In this analysis, POIs were categorized into four land use types:
   - **Residential:** Areas primarily used for housing.
   - **Commercial:** Areas used for business activities, including retail and office spaces.
   - **Recreational:** Areas designated for leisure and entertainment, such as parks, sports facilities, and cultural venues.
   - **Institutional:** Areas occupied by institutions such as schools, government buildings, and hospitals.
2. **LUM Calculation:** The LUM for each grid cell was calculated using the entropy index formula, quantifying the diversity of the four land use types. The formula is as follows:

$$\text{ENT} = -\frac{\left[\sum_{i=1}^{k} P_i \ln\left(P_i\right)\right]}{\ln(k)} \tag{5.1}$$

   Where:

   - P_i is the area of POIs of land use type i in the grid cell,
   - T is the total area of POIs in the grid cell,
   - n is the total number of land use types (which is 4 in this case).

   The resulting LUM value ranges from 0 to 1, where 0 indicates no mix (a grid cell dominated by a single land use type), and 1 indicates a high mix (a grid cell with a balanced representation of all four land use types).

The calculated LUM values are crucial for the subsequent analysis, as they provide the basis for inferring population distributions at different times of the day and week. By understanding the mix of residential, commercial, recreational, and institutional land uses within each grid cell, we can better predict how population density might fluctuate, thereby assessing potential temporal vulnerability to earthquakes.

5.4.2 Finding Dominant Land Use Types

After calculating the LUM for each grid cell, the next step involves identifying each cell's dominant land use type. This identification is critical for inferring the likely population distribution at different times of the day and week, informing the assessment of earthquake vulnerability.

To determine the dominant land use type, we applied the chi-square test to compare the observed distribution of land use types within each grid cell to an expected distribution. The chi-square test is a statistical method used to assess whether there is a significant difference between categorical data's observed and expected frequencies [70].

Null and Alternative Hypotheses:

- **Null Hypothesis ($\mathbf{H_0}$):** Land use is uniformly distributed across the grid cell, meaning there is no dominant land use type, indicating a highly mixed area.
- **Alternative Hypothesis ($\mathbf{H_1}$):** Land use is not uniformly distributed, suggesting the presence of a dominant land use type within the grid cell.

Procedure:

1. **Observed Frequencies:** The observed frequencies were derived from the actual distribution of POIs across the four land use categories (residential, commercial, recreational, and institutional) within each grid cell.
2. **Expected Frequencies:** The expected frequencies were calculated based on the assumption of a uniform distribution of land use types across the entire study area. This step establishes a baseline for comparison, assuming no particular land use type dominates any specific grid cell.
3. **Chi-square Calculation:** The chi-square statistic was calculated using the following formula:

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} \tag{5.2}$$

   Where:

   - O_i represents the observed frequency of land use type i in the grid cell,
   - E_i represents the expected frequency of land use type i,
   - n is the number of land use types (which is 4 in this case).
4. **Determining Dominance:** A p-value threshold of *0.05* was used to determine statistical significance. If the chi-square test indicated a significant difference between observed and expected frequencies (i.e., rejecting the null hypothesis), the land use type with the highest observed frequency was considered dominant in that grid cell.

Identifying the dominant land use type in each grid cell allows us to infer where populations will likely be concentrated at different times. For example, cells dominated by residential land use are likely to have higher population densities during evenings and weekends, whereas commercial-dominated cells are more active during weekday daytime hours. These inferences are crucial for assessing the temporal aspects of earthquake vulnerability.

5.4.3 Merge with Land Usability Map

The methodology's final step involves integrating the LUM calculations results and the chi-square test with the Land Usability Maps derived from the microzoning studies in Istanbul. These maps were developed as part of a comprehensive assessment of earthquake hazards, and they categorize areas based on their suitability for settlement and the associated seismic risks [71, 72].

Integration and Visualization Process:

1. **Side-by-Side Visualization:** To effectively analyze earthquake vulnerability, we visualized the Land Usability Maps alongside the LUM maps for different times of the day and week. This side-by-side comparison allows us to assess where high population density (inferred from the dominant land use type) coincides with zones of varying land usability, including areas marked as suitable (Uygun Alanlar, UA), precautionary (Önlemli Alanlar, ÖA), and unsuitable (Uygun Olmayan Alanlar, UOA) for settlement.

2. **Identifying Temporal Vulnerability:** By comparing these maps, we identified particularly vulnerable areas at specific times. For instance, a region with high LUM during weekday daytime hours, situated in a precautionary or unsuitable zone, might be identified as a high-risk area during that time period.

3. **Mapping Vulnerability Hotspots:** The side-by-side visualization facilitated the identification of "hotspots"—areas that are both heavily populated at certain times and located in zones of low land usability. These hotspots are crucial for understanding the temporal dynamics of earthquake vulnerability, especially in regions with high seismic risk, as identified in the microzoning studies.

4. **Temporal Dynamics of Vulnerability:** The integration of these visualizations enabled the analysis of how vulnerability changes over time. For instance, a commercial area might be highly vulnerable during working hours but less so during nights and weekends, while residential areas might show the opposite pattern. This analysis is grounded in the findings of the microzoning studies, which consider factors such as soil liquefaction potential, proximity to fault lines, and other geotechnical characteristics specific to Istanbul.

Using side-by-side visualizations of the Land Usability Maps and LUM maps provides urban planners with a clear understanding of when and where populations are most at risk. The Land Usability Maps, which are based on detailed seismic risk assessments, highlight areas that require careful consideration in urban planning and disaster preparedness. By identifying hotspots, planners can allocate resources more effectively and prioritize interventions in areas where the risk is highest during specific times of the day or week. This approach is vital for enhancing the effectiveness of earthquake preparedness and response strategies in Istanbul.

5.5 Datasets

This section introduces our study's datasets and preparation methods, focusing on data collected for the Kadıköy and Avcılar districts. The datasets include publicly available POI data and Land Usability Maps obtained from official sources.

5.5.1 POI Data

The POI data for Kadıköy and Avcılar were collected from two primary sources: Sahibinden and Google Maps API. The data collection followed the methodology outlined in Chapter 3, with some modifications to accommodate the specific land use types relevant to this study.

Sources and Labeling:

- **Residential:** POIs were identified as residential if they were ad listings marked as residences on Sahibinden or labeled as `lodging` in Google Maps.
- **Commercial:** POIs labeled as `shopping_mall`, `store`, `market`, `food`, `moving_company`, or `storage` in Google Maps were categorized as commercial.
- **Institutional:** POIs were labeled as institutional if they were identified as `school`, `university`, `bank`, `dentist`, `doctor`, or `mosque` in Google Maps.
- **Recreational:** POIs such as `mosque`, `museum`, `church`, and `park` were categorized as recreational.

The data was collected using Python scripts, with the Selenium library employed for automated data scraping. After the data cleansing, a total of 4847 POIs were collected for Kadıköy and 3276 for Avcılar in November 2023. The POIs depicted in each region with fishnets are presented in Figure 5.3.

5.5.2 Land Usability Map

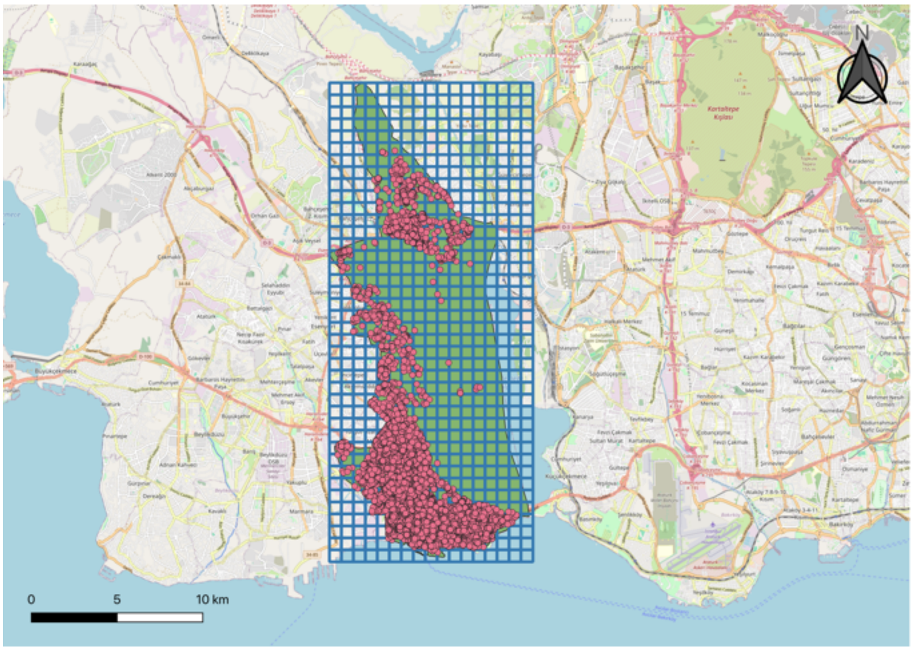
The Land Usability Maps used in this study were obtained from the comprehensive microzoning studies conducted by the Istanbul Metropolitan Municipality [71, 72]. These maps categorize areas into suitable, precautionary, and unsuitable for settlement based on seismic risk factors such as soil stability, liquefaction potential, and proximity to fault lines.

Integration with POI Data: The Land Usability Maps were digitized using QGIS to facilitate spatial analysis. The maps served as the basis for assessing the vulnerability of different land use areas, particularly in regions identified as precautionary or unsuitable.

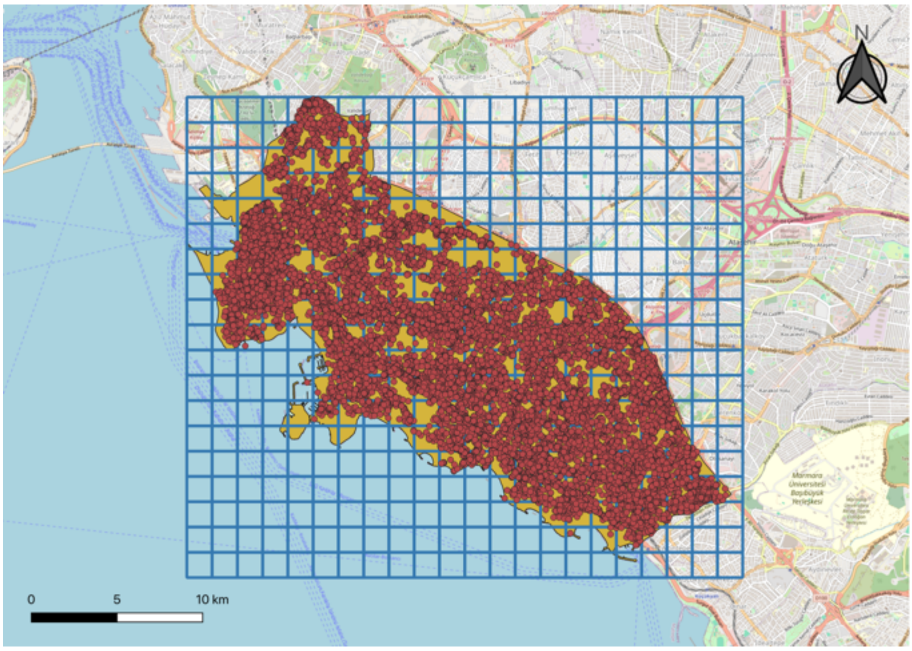
5.5.3 Data Preparation

The data preparation process followed the general methodology described in Chapter 3, with additional steps tailored to the specific needs of this study.

POI Labeling: Based on the above criteria, POIs were categorized into residential, commercial, institutional, and recreational land use types. This labeling was critical for the LUM calculations and the chi-square test used to identify dominant land use types. The number of POIs with categories are listed in Table 5.1.



(a)



(b)

Figure 5.3: Visualization of POIs and fishnet grids for the Avcılar (a) and Kadıköy (b) districts. The POIs, representing various land use types, are overlaid on each district's fishnet grids.

Table 5.1: Categorization of POIs Collected for Kadıköy and Avcılar

Land Use Type	Kadıköy (n = 4847)	Avcılar (n = 3276)
Residential	1650	1100
Commercial	1400	925
Institutional	1020	775
Recreational	627	371
Others	150	105

Land Usability Map Digitization: The Land Usability Maps were digitized using QGIS to create a spatial dataset that could be overlaid with the POI data. This step involved careful georeferencing to ensure accurate spatial alignment between the POI data and the Land Usability Maps.

Tools Used: The data preparation was conducted using a combination of Python and QGIS. Python, with libraries such as Pandas and Geopandas, was used for data processing and analysis, while QGIS was employed for spatial data manipulation and visualization.

The integration of these datasets provided a robust foundation for the subsequent analysis of temporal earthquake vulnerability in Kadıköy and Avcılar.

5.6 Results and Interpretation

This section introduces the results and interpretation of our findings for Kadıköy and Avcılar by presenting the LUM maps, dominant land use types, and the comparison with land usability maps for both places.

5.6.1 LUM Map Overview in Kadıköy

The LUM map for Kadıköy is presented in Figure 5.4, where the district is divided into a grid, and each cell is analyzed to determine the LUM index. The color-coded map ranges from blue (low LUM, indicating less diverse land use) to red (high LUM, indicating more diverse land use). This visualization highlights the varying levels of land use diversity across Kadıköy.

The LUM analysis reveals distinct patterns of land use distribution within Kadıköy. Areas with high LUM values (red) indicate regions where there is a high diversity of land use types, while areas with low LUM values (blue) are more homogeneous, typically dominated by a single land use type.

In Kadıköy, high LUM values are observed in the central parts of the district, which are characterized by a mix of residential, commercial, and recreational land uses. These areas are likely to be vibrant urban spaces with varied activities throughout the day and evening. On the other hand, peripheral areas exhibit lower LUM values, suggesting that these regions are more specialized and less diverse in their land use composition.

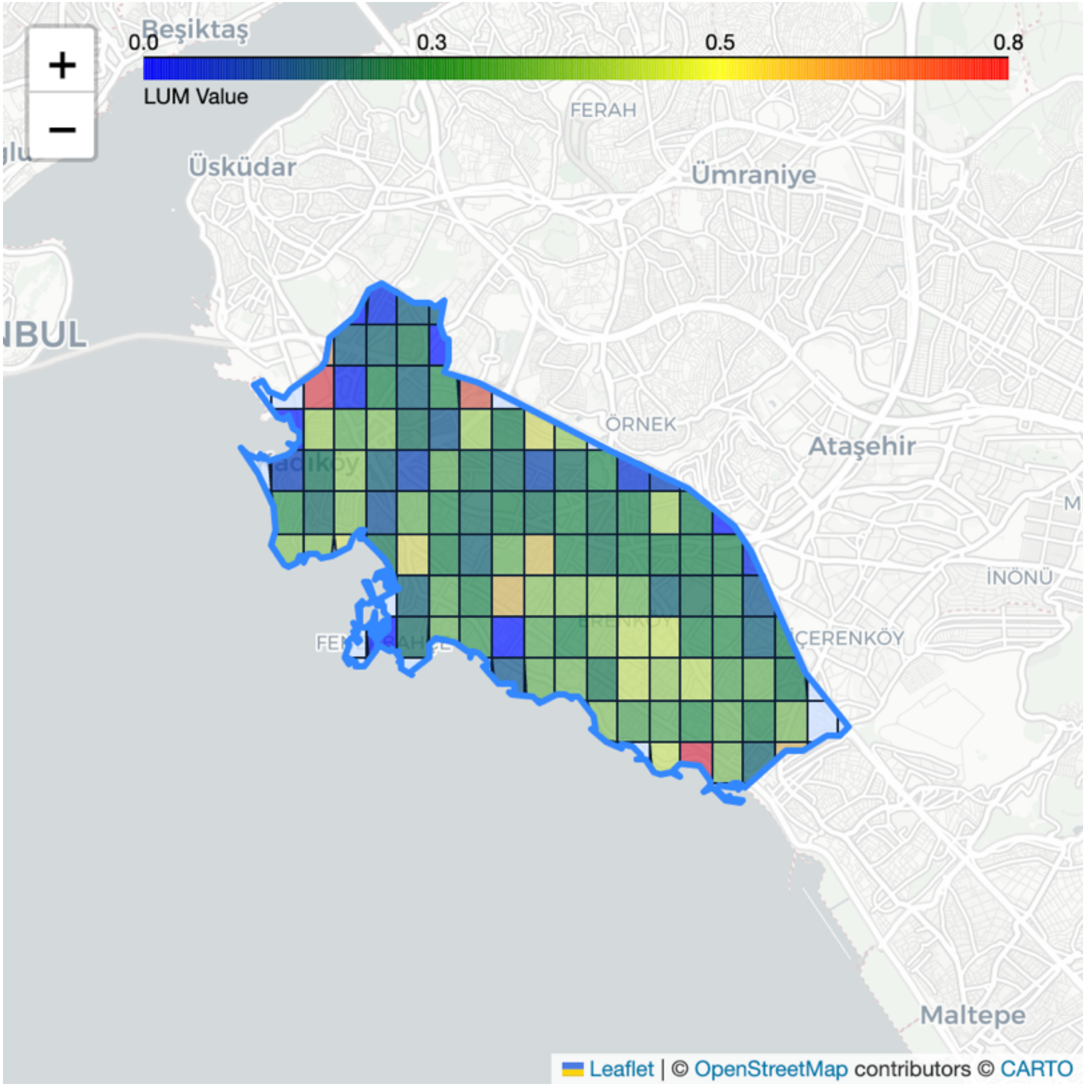


Figure 5.4: LUM Map of Kadıköy District. This map visualizes the LUM values across Kadıköy, with each grid cell representing a specific area’s LUM index. The color scale ranges from blue (low LUM, indicating less diverse land use) to red (high LUM, indicating more diverse land use).

5.6.2 LUM Map Overview in Avcılar

Similarly, the LUM map for Avcılar is shown in Figure 5.5. The grid cells are color-coded to represent the LUM index, with the same color scale as used for Kadıköy. This map provides insights into the land use diversity within Avcılar, allowing for a comparative analysis between the two districts.

The LUM analysis in Avcılar similarly identifies areas with varying degrees of land use diversity. The central and northern parts of Avcılar exhibit higher LUM values, indicating a mix of different land use types. The southern parts, closer to the coast, show lower LUM values, suggesting a more uniform land use distribution, likely dominated by residential areas.

5.6.3 Kadıköy Dominant Land Use

Figure 5.6 displays the dominant land use type map for Kadıköy, where each grid cell is color-coded based on the predominant land use type: Residential (red), Commercial (green), Recreational (blue), Institutional (yellow), or Mixed Land (black). This map highlights the spatial distribution of different land uses within the district, identifying areas where certain land uses are more prevalent.

The analysis shows that residential land use (red) dominates much of Kadıköy, particularly in the northern and central areas. Commercial land use (green) is primarily concentrated along major roadways and near the central business district. Recreational land use (blue) and institutional land use (yellow) are scattered throughout the district, with notable clusters near parks and educational institutions. Mixed Land (black) areas, which indicate a balance of different land use types, are found in the central zones, contributing to the area's vibrancy and complexity.

5.6.4 Avcılar Dominant Land Use

The dominant land use type map for Avcılar is shown in Figure 5.7. The map uses the same color coding as the Kadıköy map, allowing for easy comparison of dominant land use types between the two districts.

In Avcılar, residential land use (red) is predominant in the southern and western parts of the district, while commercial land use (green) is more common in the northern regions, particularly near major commercial hubs. The presence of recreational and institutional land uses is less pronounced compared to Kadıköy, indicating that Avcılar may be more residential in nature with fewer mixed-use developments.

5.6.5 Overlay Analysis in Kadıköy

To assess the alignment of dominant land use types with land usability, the dominant land use map is overlaid with the Land Usability Map for Kadıköy, as shown in Figure 5.8. The overlay helps identify areas where high land use diversity intersects with zones marked as Suitable (blue), Precautionary-A (yellow), Precautionary-B (orange), or Unsuitable (red) for urban development.

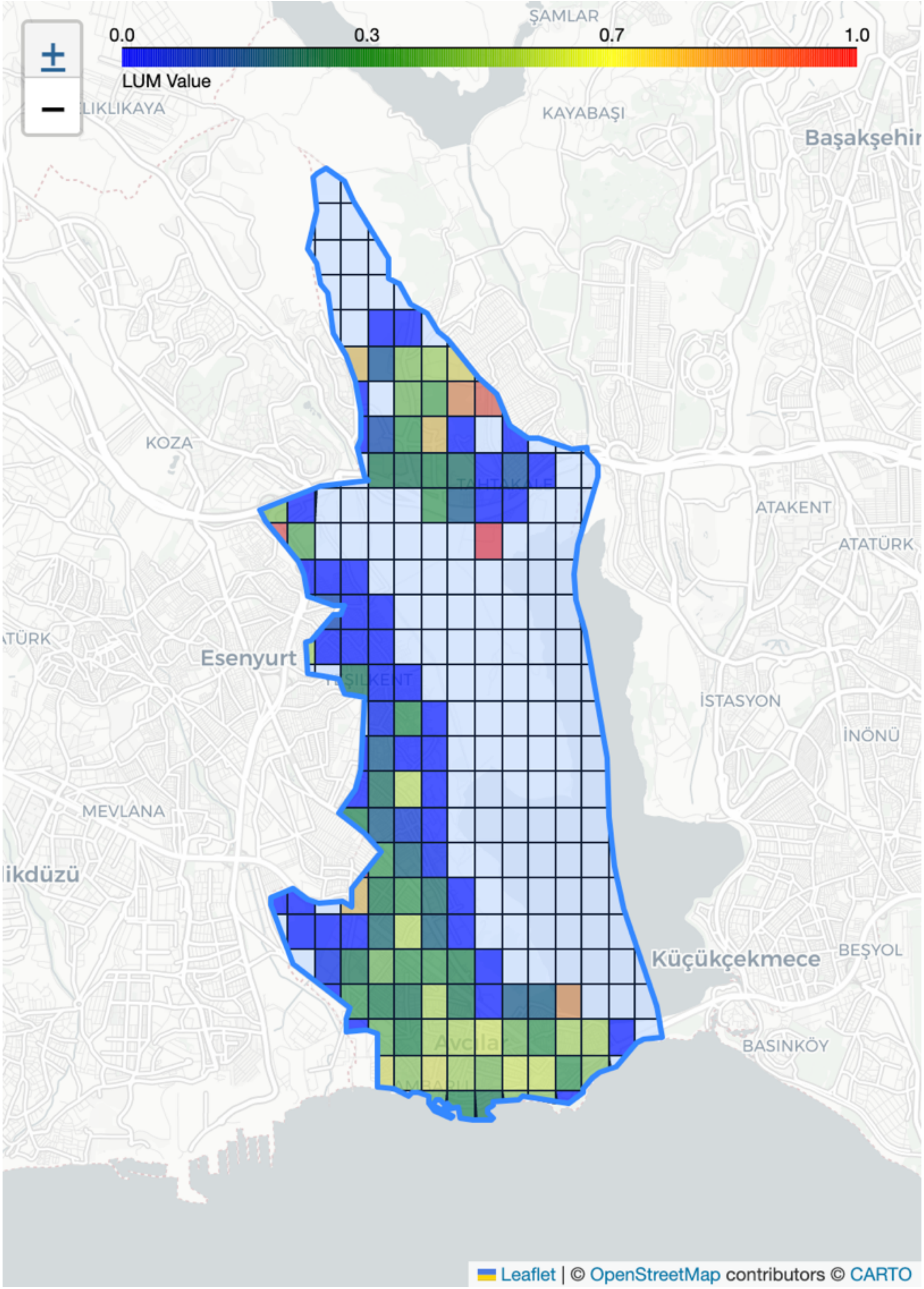


Figure 5.5: LUM Map of Avcılar District. This map visualizes the LUM values across Avcılar, with each grid cell representing a specific area's LUM index. The color scale ranges from blue (low LUM, indicating less diverse land use) to red (high LUM, indicating more diverse land use).

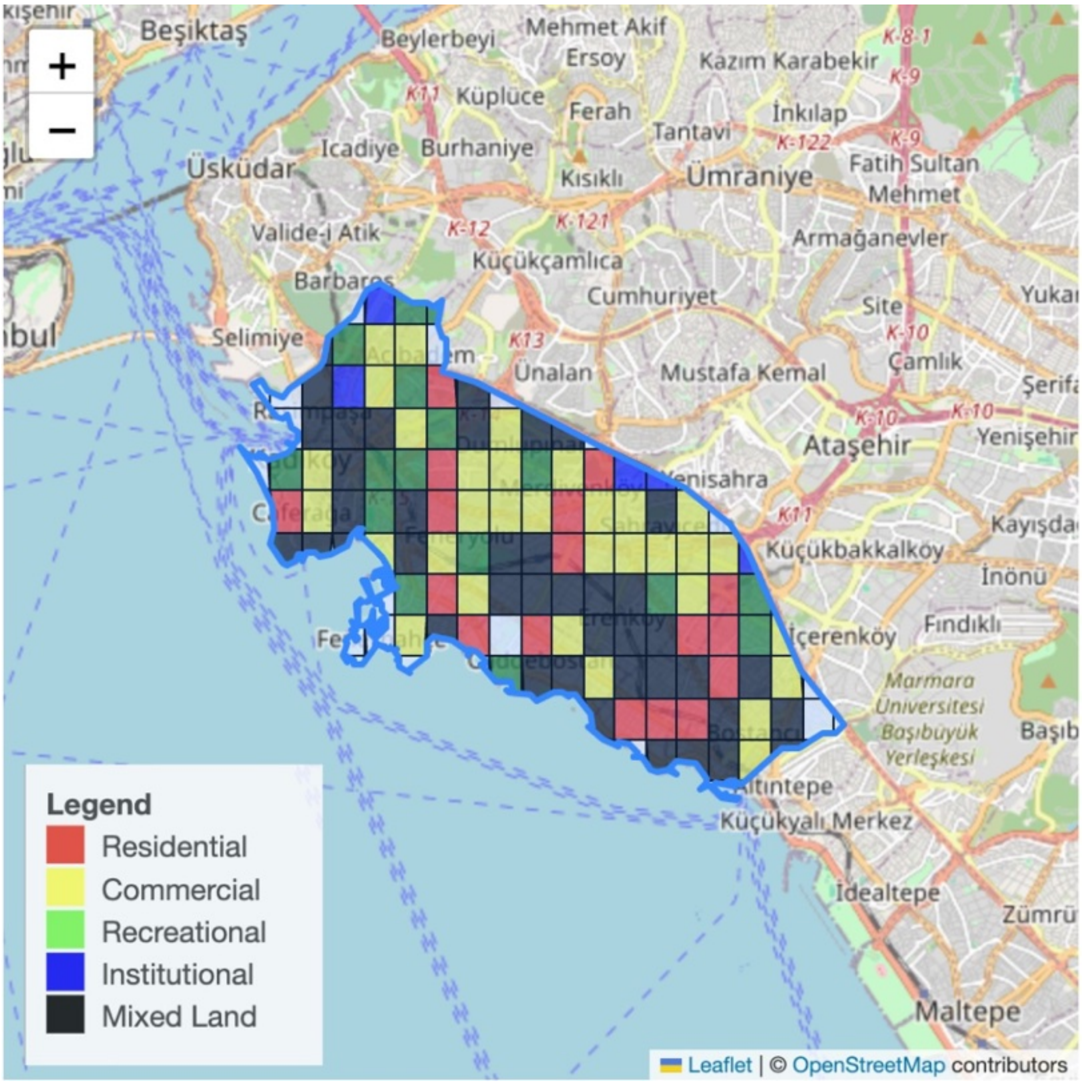


Figure 5.6: Dominant Land Use Type Map of Kadıköy District. Each cell is color-coded based on the dominant land use type: Residential (red), Commercial (green), Recreational (blue), Institutional (yellow), or Mixed Land (black).

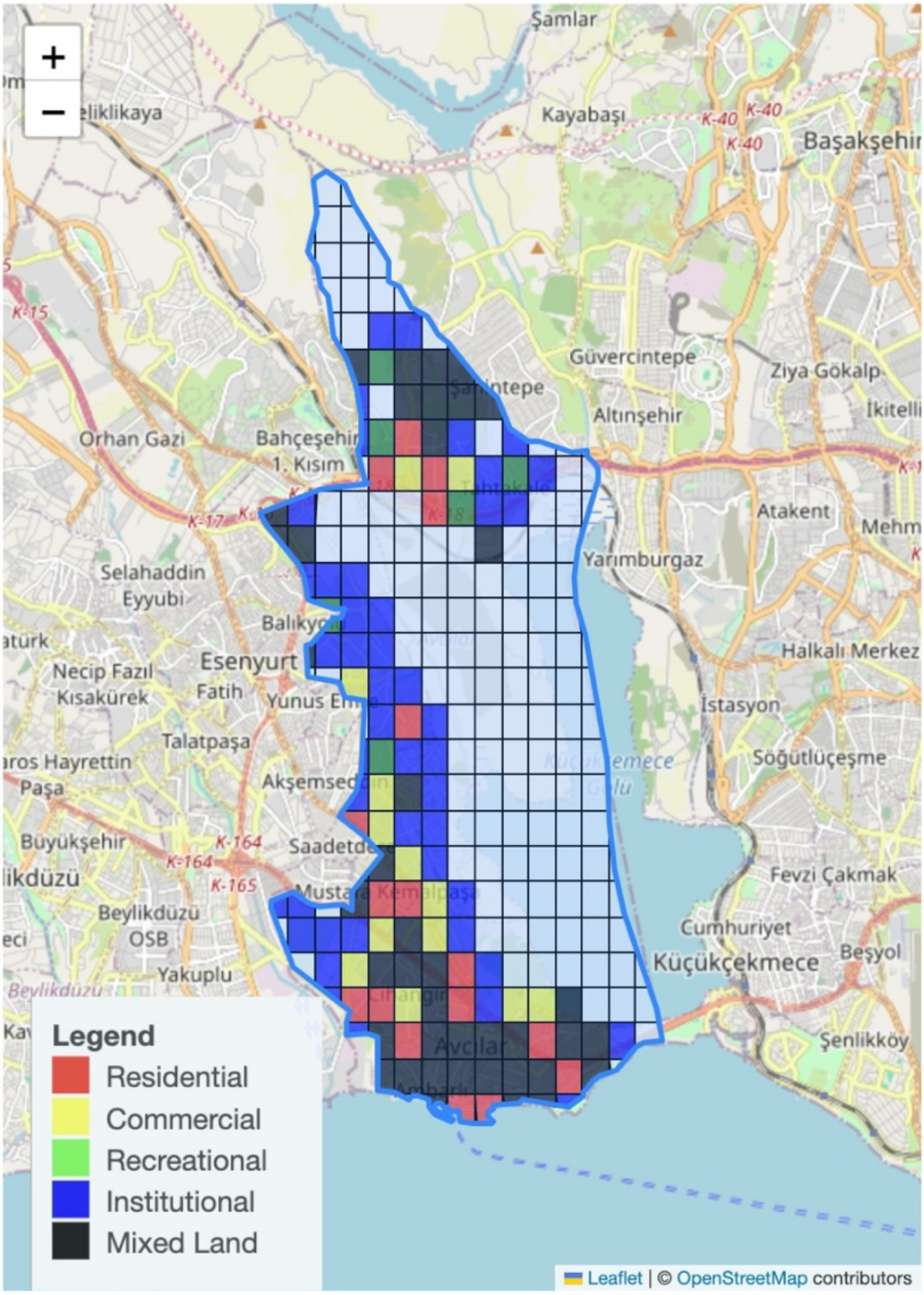


Figure 5.7: Dominant Land Use Type Map of Avcılar District. Each cell is color-coded based on the dominant land use type: Residential (red), Commercial (green), Recreational (blue), Institutional (yellow), or Mixed Land (black).

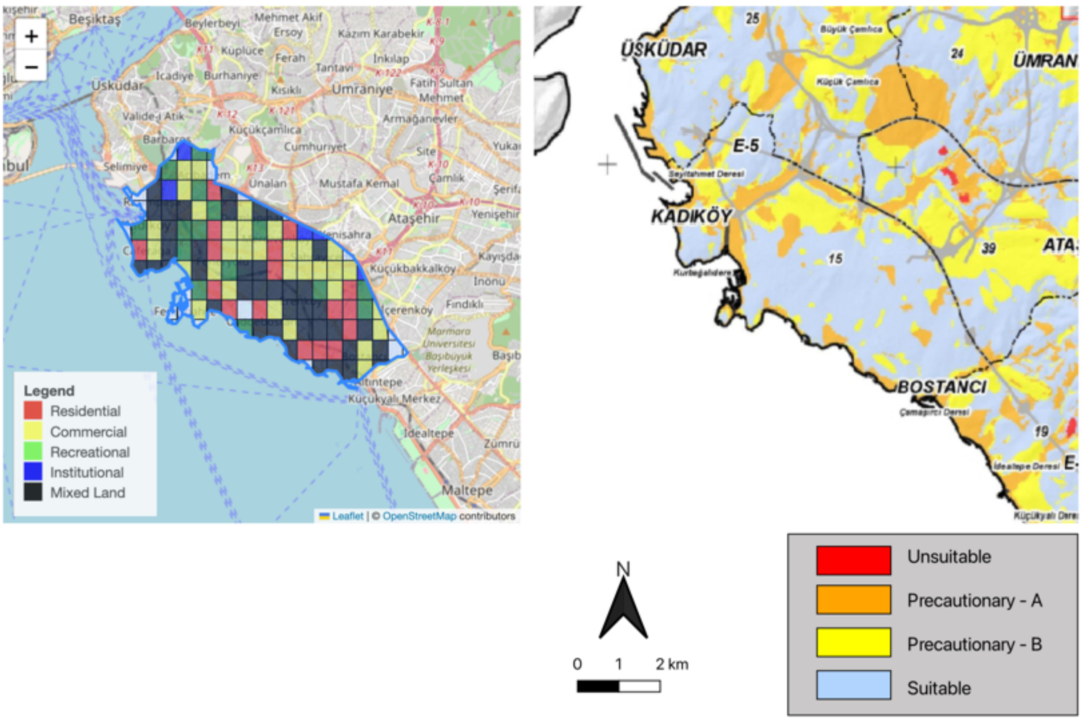


Figure 5.8: Overlay of Dominant Land Use Type and Land Usability Maps for Kadıköy District. The left panel displays the dominant land use map for Kadıköy, where the district is divided into a grid, and each cell is color-coded based on the dominant land use type: Residential (red), Commercial (green), Recreational (blue), Institutional (yellow), or Mixed Land (black). The right panel shows the Land Usability Map, indicating areas categorized as Suitable (blue), Precautionary-A (yellow), Precautionary-B (orange), and Unsuitable (red) for urban development.

The overlay analysis in Kadıköy reveals that areas with high land use diversity (Mixed Land, black) are often located within zones marked as Suitable (blue). However, there are instances where diverse land use areas intersect with Precautionary or Unsuitable zones, particularly in the northern parts of the district. These areas may require special attention from urban planners to mitigate potential risks associated with land suitability constraints.

5.6.6 Overlay Analysis in Avcılar

Figure 5.9 presents the overlay of the dominant land use map with the Land Usability Map for Avcılar. This analysis follows the same methodology as for Kadıköy, facilitating a comparative understanding of land use patterns and their suitability in both districts.

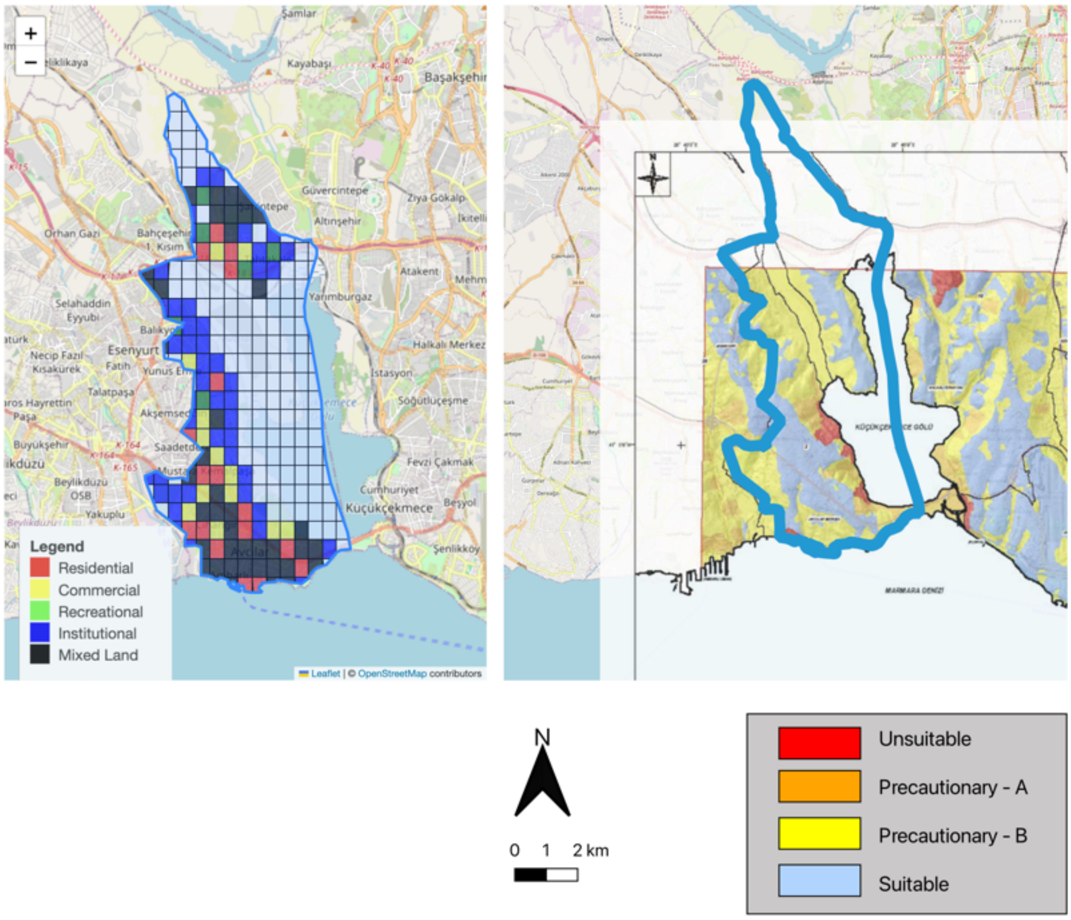


Figure 5.9: Overlay of Dominant Land Use Type and Land Usability Maps for Avcılar District. The left panel displays the dominant land use map for Avcılar, where the district is divided into a grid, and each cell is color-coded based on the dominant land use type: Residential (red), Commercial (green), Recreational (blue), Institutional (yellow), or Mixed Land (black). The right panel shows the Land Usability Map, indicating areas categorized as Suitable (blue), Precautionary-A (yellow), Precautionary-B (orange), and Unsuitable (red) for urban development.

The overlay analysis in Avcılar shows that much of the residential land use is located in areas marked as Suitable (blue), which is favorable for continued urban development. However, there are significant regions, particularly in the northern parts, where residential and commercial areas intersect with Precautionary-A and Precautionary-B zones (yellow and orange). These findings suggest that these areas may be at a higher risk of vulnerability, requiring careful consideration in future urban planning efforts.

5.6.7 Identification of Temporal Vulnerability Hotspots

Given the diverse land use in both Kadıköy and Avcılar, it is crucial to consider how population dynamics might vary by time of day and day of the week. Residential areas are likely to have higher population densities during the evenings and weekends, while commercial and institutional areas are busier during weekdays.

In Kadıköy, areas identified with high LUM values (red on the LUM map) are likely to experience significant temporal shifts in population density. For example, the central areas with a mix of residential, commercial, and recreational land uses might have a dense population during both the day (due to commercial and institutional activities) and the night (due to residential occupancy). This dynamic poses challenges for emergency response planning, as these areas could be densely populated at any time.

Similarly, in Avcılar, high LUM areas near major commercial hubs could see large crowds during the day, while the more residential southern regions would be more densely populated at night. Understanding these temporal shifts is essential for accurately assessing the vulnerability of these areas to hazards such as earthquakes, where the risk to human life can vary dramatically depending on the time of day.

Potential Risk Areas by Time

By integrating the LUM findings with temporal activity patterns, we can identify potential vulnerability hotspots. The key findings are as follows:

- **Weekday Daytime Risk:**
  - **Commercial and Institutional Areas:** On weekdays, these areas are particularly vulnerable during the daytime when they experience high foot traffic and dense occupancy. In Kadıköy, this risk is prominent in central commercial zones and areas near schools and universities. In Avcılar, similar risks are observed near major shopping centers and business districts.
  - **Residential Areas:** Even though these areas are less populated during the day on weekdays, they may still pose a risk if they are located in Precautionary or Unsuitable zones, especially in areas where residential land use is mixed with commercial or institutional uses.
- **Weekday Nighttime Risk:**
  - **Residential Areas:** At night, residential areas become more densely populated as people return home from work. In Kadıköy, this is particularly true for the northern and central

residential zones that overlap with Precautionary-A and B areas. In Avcılar, the southern residential zones may face increased risks.

  - **Commercial Areas:** Commercial zones may experience lower risks at night as activity levels drop, but mixed-use areas where residential and commercial land uses overlap may still pose a concern.

- **Weekend Daytime Risk:**
  - **Recreational and Commercial Areas:** On weekends, recreational and commercial areas become more active, with high foot traffic in parks, malls, and entertainment venues. This increases the vulnerability of these areas, particularly those located in Precautionary or Unsuitable zones. For example, the coastal areas of Kadıköy, which attract visitors for leisure activities, may face heightened risks during weekends.
  - **Institutional Areas:** Generally quieter during weekends, institutional areas might not pose a significant risk unless they are near recreational or commercial zones.

- **Weekend Nighttime Risk:**
  - **Residential Areas:** Similar to weekday nights, residential areas remain the primary concern during weekends. However, some areas might see even higher densities as people tend to stay home or host gatherings. This is especially concerning in mixed-use zones where residential land use dominates but is interspersed with commercial and recreational activities.
  - **Recreational Areas:** Some recreational areas, especially those with nightlife, could also see increased activity during weekend nights, necessitating focused risk mitigation strategies.

By considering both the time of day and the day of the week, the analysis highlights the importance of tailored urban planning and emergency preparedness strategies. Different areas may require varied approaches depending on their land use mix, the day of the week, and the time of day, ensuring that all potential vulnerabilities are adequately addressed.

The results of the LUM analysis, combined with the Land Usability Maps and consideration of temporal dynamics, provide a comprehensive understanding of the urban landscape in Kadıköy and Avcılar. Key findings include:

- LUM Distribution: Both districts show significant variation in LUM, with Kadıköy exhibiting higher land use diversity in its central areas compared to the more residentially dominated Avcılar.

- Dominant Land Use Types: Residential land use is predominant in both districts, but Kadıköy has a more mixed land use in its central areas, while Avcılar is more uniformly residential, especially in the south.

- Alignment with Land Usability: In both districts, there are significant overlaps between high LUM areas and zones marked as Precautionary or Unsuitable in the Land Usability Maps, indicating potential areas of concern.

5.7 Conclusions

This chapter demonstrates the importance of LUM analysis as a key tool for assessing and understanding urban vulnerability. When used in conjunction with Land Usability Maps and temporal considerations, LUM analysis is an indispensable tool for urban planners and disaster managers. This research provides valuable insights directly applicable to urban planning and disaster management by exploring the relationship between land use diversity, spatial suitability, and temporal population dynamics.

The case studies of Kadıköy and Avcılar districts irrefutably demonstrate that urban vulnerability is a multi-faceted issue, influenced not only by the spatial distribution of land uses but also by the temporal patterns of population density. The LUM maps clearly showed that there were varying degrees of land use diversity across different parts of these districts. When overlaid with land usability maps, they unquestionably highlighted specific areas where urban development is particularly risky.

This study makes a significant contribution by focusing on the temporal aspects of vulnerability. It acknowledges that the risk to human life and property in urban areas fluctuates depending on the time of day and day of the week. This dynamic perspective challenges traditional, static vulnerability assessments and demands more sophisticated, time-sensitive urban planning and emergency response strategies.

These findings have significant and far-reaching practical implications. Urban planners must integrate temporal vulnerability considerations into their decision-making processes. They must tailor infrastructure, resource allocation, and public policies to address the specific risks identified in this study. Furthermore, it is imperative that the public be educated about these risks, as this will significantly enhance community preparedness and resilience.

Further studies should build on this research, refining and expanding the methodologies used here. Incorporating real-time data, such as human mobility patterns captured via mobile devices, will provide more precise assessments of urban vulnerability. Furthermore, applying these findings in other urban contexts will help to generalize the approach and solidify its utility in diverse settings.

CHAPTER 6

CONCLUSION

The thesis has explored the role of LUM analysis in urban planning, GIS research, and disaster risk assessment. Introducing novel methodologies and tools has addressed the limitations of traditional LUM assessment methods, which often struggle with scalability, adaptability, and capturing the dynamic nature of urban environments. Through integrating POI data with advanced spatial analysis techniques like Voronoi triangulation and entropy-based calculations, this research offers a more efficient and scalable approach to LUM identification, reducing the need for extensive fieldwork and human intervention. Furthermore, we identify optimal POI density ranges essential for effective LUM classification and apply our methodology to assess earthquake vulnerability in urban settings. The empirical validation of these methods in real-world urban contexts, such as Ankara and Kadıköy, demonstrates their effectiveness and robustness.

We also developed a Python package, an open-source tool designed to facilitate LUM analysis using established indices like the Entropy Index and the Herfindahl-Hirschman Index. This user-friendly, scalable, and adaptable package makes it a valuable resource for researchers and practitioners in urban planning and GIS.

Additionally, applying LUM analysis to assess temporal variations in urban vulnerability, particularly in the context of earthquake risk, has provided new insights into how urban vulnerability evolves over time, emphasizing the importance of time-sensitive strategies in urban planning and disaster preparedness, which is also demonstrated in Kadıköy and Avcılar.

Our methodology offers a robust tool to identify and manage mixed land use zones, thus mitigating potential negative urban consequences such as congestion and pollution. By incorporating our methods into urban planning frameworks, planners can ensure that mixed-use areas are optimally balanced, fostering sustainable and livable urban environments. For instance, identifying LUM provides urban planners with specific guidelines for better urban design and resource allocation, enhancing both functionality and aesthetics in urban spaces.

By utilizing public geospatial data for LUM analysis, we demonstrated a method that proved to be accurate in classifying mixed and no-mixed areas using publicly available data. This approach not only makes LUM analysis more accessible and cost-effective but also democratizes the data, enabling wider use and adaptation in different urban contexts without the need for expensive proprietary data sources.

We significantly enhanced LUM patterns' spatial detail and reliability by integrating Voronoi triangulation and entropy indices within our research. This integration allows for a more detailed understanding of urban spaces, capturing the complexity and diversity of land use in a manner that traditional methods might overlook. These improvements are crucial for urban planners aiming to create more detailed and effective urban development plans based on point-based data.

Our analysis identified specific POI density ranges that optimize mixed-used classification accuracy. This finding delivers practical insights for urban data analysis, allowing planners and researchers to effectively collect and utilize geospatial data for LUM classification. The defined density ranges help ensure that the data used is neither too sparse to be meaningful nor too dense to process efficiently, striking a balance that facilitates accurate and actionable urban planning insights.

Applying our LUM methodology to earthquake-prone zones revealed new insights into urban vulnerability, contributing to disaster risk management. Understanding land use patterns in relation to earthquake risk zones can aid in urban development and emergency preparedness planning. This approach highlights the practical applicability of LUM analysis in enhancing the resilience of urban environments to natural disasters by informing zoning laws, building codes, and emergency response strategies.

The significant contributions of our methodology have been validated through empirical analysis, extending its utility to urban planners and disaster risk managers. Our solution's scalability and practical application for LUM identification and risk assessment are evident, signifying a substantial advancement in urban planning methodologies. Our novel methodology, combined with the open-source `landusemix` Python package, represents a leap forward for urban planners and researchers. It streamlines LUM identification and presents a scalable solution adaptable to various urban environments, promoting more resilient and sustainable urban development.

Integrating LUM analysis with earthquake risk assessment introduces a new approach to disaster risk management. This methodology has the potential to lead to safer and more resilient urban environments, providing a crucial tool for urban planners to address and mitigate the impacts of natural disasters.

This work's empirical validation and practical applications underscore its relevance and utility in addressing modern urban challenges. The contributions from this thesis lay a robust foundation for future urban studies and enhancements, paving the way for resilient, sustainable, and more livable cities.

6.1 Limitations and Future Work

Despite the notable contributions of the proposed methods and applications, the research has certain limitations. The accuracy of LUM identification is contingent on the quality and completeness of the POI data used, which may be lacking in certain areas. For instance, lower-income or less digitally mapped regions might have insufficient POI data, leading to less accurate LUM analysis. This dependence on POI data underscores the need for continuous updates and possible integration with additional data sources to ensure comprehensive and accurate LUM classifications.

While the proposed methodology is more scalable than traditional approaches, it still demands significant computational resources, especially when applied to large and complex urban areas. High-resolution LUM analysis over extensive regions can produce a substantial computational load, requiring efficient data processing and algorithm optimization techniques. Moreover, the *landusemix* package could benefit from further refinement to address specific edge cases and improve its adaptability across different urban contexts. Specific urban configurations and anomalies in the data can pose challenges that require tailored adjustments in the methodology.

As future work, enhancing the *landusemix* package by expanding its features and integrating it with other urban planning tools could significantly increase its utility. Adding functionalities such as automated data cleaning, improved visualization tools, and enhanced compatibility with GIS platforms could help streamline end-user's workflow. Applying the developed methodology to a broader range of urban environments would also help to validate its adaptability and robustness. Testing the approach in diverse city scapes, from megacities to smaller towns, would provide insights into its generalizability and performance under varying conditions.

Incorporating real-time data sources, such as mobile phone data or social media activity, could further enhance the dynamic capabilities of LUM analysis, providing more immediate insights for urban planners. Real-time data can capture transient land use changes and human activity patterns, offering a more dynamic and responsive analysis tool. This would be particularly beneficial for applications that require up-to-date information, such as disaster response or event management. Furthermore, leveraging additional data sources such as geotagged tweets and satellite imagery could enrich the dataset, providing more comprehensive LUM assessments. Geotagged social media data can offer insights into human behavior and activity patterns. At the same time, high-resolution satellite imagery can provide detailed, up-to-date information on changes in physical land use and urban morphology.

Finally, exploring the relationship between LUM and other critical urban challenges, such as climate change resilience, transportation efficiency, and social equity, could broaden the impact of LUM analysis in urban studies. For instance, understanding how mixed land use contributes to or mitigates urban heat islands can inform climate adaptation strategies. Investigating the interplay between LUM and public transportation networks can aid in designing more efficient and accessible transit systems. Furthermore, examining how land use diversity affects social equity can lead to more inclusive and equitable urban development policies.

In conclusion, this thesis has significantly contributed to understanding and applying LUM analysis in urban environments. By providing practical tools and methodologies, it supports the broader goal of promoting sustainable, resilient, and livable cities, equipping urban planners and researchers with the means to better analyze, plan, and respond to the complexities of modern urban landscapes. Moving forward, addressing the identified limitations and exploring the outlined future work avenues will further enhance the relevance and applicability of LUM analysis, fostering the development of smarter, more adaptive, and equitable urban areas.

REFERENCES

[1] J. Jacobs, *The death and life of great American cities*. Random House, New York, 1961.

[2] S. Swamy and D. Baindur, "Managing urban freight transport in an expanding city—case study of ahmedabad," *Research in Transportation Business & Management*, vol. 11, pp. 5–14, 2014.

[3] P. Jia, X. Pan, F. Liu, P. He, W. Zhang, L. Liu, Y. Zou, and L. Chen, "Land use mix in the neighbourhood and childhood obesity," *Obesity reviews*, vol. 22, p. e13098, 2021.

[4] P. Thanh Noi and M. Kappas, "Comparison of random forest, k-nearest neighbor, and support vector machine classifiers for land cover classification using sentinel-2 imagery," *Sensors*, vol. 18, no. 1, p. 18, 2018.

[5] X. Liu, J. He, Y. Yao, J. Zhang, H. Liang, H. Wang, and Y. Hong, "Classifying urban land use by integrating remote sensing and social media data," *International Journal of Geographical Information Science*, vol. 31, no. 8, pp. 1675–1696, 2017.

[6] J. Li, J. A. Benediktsson, B. Zhang, T. Yang, and A. Plaza, "Spatial technology and social media in remote sensing: A survey," *Proceedings of the IEEE*, vol. 105, no. 10, pp. 1855–1864, 2017.

[7] C. Aubrecht, J. Ungar, D. O. Aubrecht, S. Freire, and K. Steinnocher, "Mapping land use dynamics using the collective power of the crowd," in *Earth Observation Open Science and Innovation*, pp. 247–253, Springer, Cham, 2018.

[8] T.-N. Doan and E.-P. Lim, "Modeling location-based social network data with area attraction and neighborhood competition," *Data Mining and Knowledge Discovery*, vol. 33, no. 1, pp. 58–95, 2019.

[9] J. Li, Z. He, J. Plaza, S. Li, J. Chen, H. Wu, Y. Wang, and Y. Liu, "Social media: New perspectives to improve remote sensing for emergency response," *Proceedings of the IEEE*, vol. 105, no. 10, pp. 1900–1912, 2017.

[10] Y. Yao, X. Li, X. Liu, P. Liu, Z. Liang, J. Zhang, and K. Mai, "Sensing spatial distribution of urban land use by integrating points-of-interest and google word2vec model," *International Journal of Geographical Information Science*, vol. 31, no. 4, pp. 825–848, 2017.

[11] A. Sitthi, M. Nagai, M. Dailey, and S. Ninsawat, "Exploring land use and land cover of geotagged social-sensing images using naive bayes classifier," *Sustainability*, vol. 8, no. 9, p. 921, 2016.

[12] V. Frias-Martinez and E. Frias-Martinez, "Spectral clustering for sensing urban land use using twitter activity," *Engineering Applications of Artificial Intelligence*, vol. 35, pp. 237–245, 2014.

[13] Q. Weng, Z. Mao, J. Lin, and X. Liao, "Land-use scene classification based on a cnn using a constrained extreme learning machine," *International Journal of Remote Sensing*, vol. 39, no. 19, pp. 6281–6299, 2018.

[14] W. Zhai, X. Bai, Y. Shi, Y. Han, Z.-R. Peng, and C. Gu, "Beyond word2vec: An approach for urban functional region extraction and identification by combining place2vec and pois," *Computers, Environment and Urban Systems*, vol. 74, pp. 1–12, 2019.

[15] J. S. Miller, "Impact of jobs–housing balance on average jurisdiction commuting times: Virginia macroscopic analysis," *Transportation Research Record*, vol. 2244, no. 1, pp. 18–26, 2011.

[16] Y. Song and G.-J. Knaap, "Measuring the effects of mixed land uses on housing values," *Regional Science and Urban Economics*, vol. 34, no. 6, pp. 663–680, 2004.

[17] S. Eom, T. Suzuki, and M.-H. Lee, "A land-use mix allocation model considering adjacency, intensity, and proximity," *International Journal of Geographical Information Science*, vol. 34, pp. 899 – 923, 2020.

[18] Y. Yue, Y. Zhuang, A. G. Yeh, J.-Y. Xie, C.-L. Ma, and Q.-Q. Li, "Measurements of poi-based mixed use and their relationships with neighbourhood vibrancy," *International Journal of Geographical Information Science*, vol. 31, no. 4, p. 658–675, 2017.

[19] S. R. Gehrke and K. J. Clifton, "Toward a spatial-temporal measure of land-use mix," *Journal of Transport and Land Use*, vol. 9, no. 1, p. 171–186, 2016.

[20] P. A. Ghosh and P. M. Raval, "Modelling urban mixed land-use prediction using influence parameters," *GeoScape*, vol. 15, no. 1, 2021.

[21] R. Cervero and K. Kockelman, "Travel demand and the 3ds: Density, diversity, and design," *Transportation research. Part D, Transport and environment*, vol. 2, no. 3, pp. 199–219, 1997.

[22] L. D. Frank, T. L. Schmid, J. F. Sallis, J. Chapman, and B. E. Saelens, "Linking objectively measured physical activity with objectively measured urban form: findings from smartraq," *American journal of preventive medicine*, vol. 28, no. 2, pp. 117–125, 2005.

[23] Y. Xu, L. Wang, C. Fu, and T. Kosmyna, "A fishnet-constrained land use mix index derived from remotely sensed data," *Annals of GIS*, vol. 23, no. 4, pp. 303–313, 2017.

[24] P. Deng, Q. Huang, F. Liu, Z. Wang, and M. Liu, "The research on identification and spatial pattern of urban mixed land-use: A case study of chengdu," *Ecological Indicators*, vol. 154, p. 110891, 2023.

[25] H. Zheng, Y. Zhuo, Z. Xu, C. Wu, J. Huang, and Q. Fu, "Measuring and characterizing land use mix patterns of china's megacities: A case study of shanghai," *Growth and Change*, vol. 52, no. 4, pp. 2509–2539, 2021.

[26] H. Yang, M. Fu, L. Wang, and F. Tang, "Mixed land use evaluation and its impact on housing prices in beijing based on multi-source big data," *Land*, vol. 10, no. 10, p. 1103, 2021.

[27] D. A. Field, "Laplacian smoothing and delaunay triangulations," *Communications in Applied Numerical Methods*, vol. 4, no. 6, p. 709–712, 1988.

[28] L. D. Frank, J. F. Sallis, T. L. Conway, J. E. Chapman, B. E. Saelens, and W. Bachman, "Many pathways from land use to health: associations between neighborhood walkability and active transportation, body mass index, and air quality," *Journal of the American Planning Association*, vol. 72, no. 1, p. 75–87, 2006.

[29] T. C. Bailey and A. C. Gatrell, *Interactive Spatial Data Analysis*, vol. 413. Longman Scientific & Technical Essex, 1995.

[30] E. Pielou, "The effect of quadrat size on the estimation of the parameters of neyman's and thomas's distributions," *The Journal of Ecology*, p. 31–47, 1957.

[31] A. Getis, "Temporal land-use pattern analysis with the use of nearest neighbor and quadrat methods," *Annals of the Association of American Geographers*, vol. 54, no. 3, p. 391–399, 1964.

[32] A. Rogers, "Quadrat analysis of urban dispersion: 1. theoretical techniques," *Environment and Planning A*, vol. 1, no. 1, p. 47–80, 1969.

[33] H. Shu, T. Pei, C. Song, T. Ma, Y. Du, Z. Fan, and S. Guo, "Quantifying the spatial heterogeneity of points," *International Journal of Geographical Information Science*, vol. 33, no. 7, p. 1355–1376, 2019.

[34] R. Belen, T. T. Temizel, and O. Kaygısız, "A data quality case study for turkish highway accident data sets," in *Road Safety on Four Continents: 15th International Conference, Abu Dhabi, United Arab Emirates, 28-30 March 2010. Paper*, 2010.

[35] C. Fan and S. Myint, "A comparison of spatial autocorrelation indices and landscape metrics in measuring urban landscape fragmentation," *Landscape and Urban Planning*, vol. 121, p. 117–128, 2014.

[36] L. Anselin, "Local indicators of spatial association—lisa," *Geographical Analysis*, vol. 27, no. 2, p. 93–115, 1995.

[37] A. Iannillo and I. Fasolino, "Land-use mix and urban sustainability: Benefits and indicators analysis," *Sustainability*, vol. 13, no. 23, p. 13460, 2021.

[38] J. Jiao, J. Rollo, and B. Fu, "The hidden characteristics of land-use mix indices: An overview and validity analysis based on the land use in Melbourne, Australia," *Sustainability*, vol. 13, no. 4, p. 1898, 2021.

[39] K. Manaugh and T. Kreider, "What is mixed use? Presenting an interaction method for measuring land use mix," *Journal of Transport and Land use*, vol. 6, no. 1, pp. 63–72, 2013.

[40] Y. Zhuo, X. Jing, X. Wang, G. Li, Z. Xu, Y. Chen, and X. Wang, "The Rise and Fall of Land Use Mix: Review and Prospects," *Land*, vol. 11, no. 12, p. 2198, 2022.

[41] Y. Song, L. Merlin, and D. Rodriguez, "Comparing measures of urban land use mix," *Computers, Environment and Urban Systems*, vol. 42, pp. 1–13, 2013.

[42] H. N. Im and C. G. Choi, "The hidden side of the entropy-based land-use mix index: Clarifying the relationship between pedestrian volume and land-use mix," *Urban Studies*, vol. 56, pp. 1865 – 1881, 2019.

[43] J. Jiao, J. Rollo, and B. Fu, "The Hidden Characteristics of Land-Use Mix Indices: An Overview and Validity Analysis Based on the Land Use in Melbourne, Australia," *Sustainability*, 2021.

[44] H. N. Im and C. G. Choi, "Measuring pedestrian volume by land use mix: Presenting a new entropy-based index by weighting walking generation units," *Environment and Planning B: Urban Analytics and City Science*, vol. 47, pp. 1219 – 1236, 2020.

[45] D. Zhou, J. Xu, and Z. Lin, "Conflict or coordination? Assessing land use multi-functionalization using production-living-ecology analysis.," *The Science of the Total Environment*, vol. 577, pp. 136–147, 2017.

[46] M. Jehling and R. Hecht, "Do land policies make a difference? A data-driven approach to trace effects on urban form in France and Germany," *Environment and Planning B: Urban Analytics and City Science*, vol. 49, pp. 114 – 130, 2021.

[47] M. H. K. Hesselbarth, M. Sciaini, K. A. With, K. Wiegand, and J. Nowosad, "landscapemetrics: an open-source R tool to calculate landscape metrics," *Ecography*, 2019.

[48] M. Bosch, "PyLandStats: An open-source Pythonic library to compute landscape metrics," *PLoS ONE*, vol. 14, 2019.

[49] L. Gervasoni, "Urbansprawl: Open framework for calculating spatial urban sprawl indices and performing disaggregated population estimates using open data," 2018–.

[50] M. A. Akyol, "Land Use Mix (landusemix) Python Package Code," 5 2024. Software available at `https://figshare.com/articles/software/Land_Use_Mix_landusemix_Python_Package_Code/25864297`.

[51] M. A. Akyol, S. Duzgun, and N. Baykal, "landusemix: A python package for calculating land use mix," *SoftwareX*, vol. 27, p. 101861, 2024.

[52] T. pandas development team, "pandas-dev/pandas: Pandas," Feb. 2020.

[53] Wes McKinney, "Data Structures for Statistical Computing in Python," in *Proceedings of the 9th Python in Science Conference* (Stéfan van der Walt and Jarrod Millman, eds.), pp. 56 – 61, 2010.

[54] J. V. den Bossche, K. Jordahl, M. Fleischmann, M. Richards, J. McBride, J. Wasserman, A. G. Badaracco, A. D. Snow, B. Ward, J. Tratner, J. Gerard, M. Perry, cjqf, G. A. Hjelle, M. Taves, E. ter Hoeven, M. Cochran, R. Bell, rraymondgh, M. Bartos, P. Roggemans, L. Culbertson, G. Caria, N. Y. Tan, N. Eubank, sangarshanan, J. Flavin, S. Rey, and J. Gardiner, "geopandas/-geopandas: v1.0.1," July 2024.

[55] S. Gillies *et al.*, "Rasterio: geospatial raster I/O for Python programmers," 2013–.

[56] H. E. Christian, F. C. Bull, N. J. Middleton, M. W. Knuiman, M. L. Divitini, P. Hooper, A. Amarasinghe, and B. Giles-Corti, "How important is the land use mix measure in understanding walking behaviour? Results from the RESIDE study," *International Journal of Behavioral Nutrition and Physical Activity*, vol. 8, pp. 1–12, 2011.

[57] M. G. Turner, R. H. Gardner, R. V. O'neill, and R. V. O'Neill, *Landscape ecology in theory and practice*, vol. 401. Springer, 2001.

[58] O. D. Duncan and B. Duncan, "A methodological analysis of segregation indexes," *American Sociological Review*, vol. 20, no. 2, pp. 210–217, 1955.

[59] A. B. Atkinson *et al.*, "On the measurement of inequality," *Journal of Economic Theory*, vol. 2, no. 3, pp. 244–263, 1970.

[60] D. S. Massey and N. A. Denton, "The dimensions of residential segregation," *Social Forces*, vol. 67, no. 2, pp. 281–315, 1988.

[61] M. A. Akyol, T. T. Temizel, S. Duzgun, and N. Baykal, “Identification of Land Use Mix Using Point-Based Geospatial Data in Urban Areas,” *Applied Sciences*, vol. 14, no. 16, 2024.

[62] C. Xia, Y. Hu, G. Chi, and J. Chen, “Assessing dynamics of human vulnerability at community level – using mobility data,” *International Journal of Disaster Risk Reduction*, 2023.

[63] R. Debnath, “An assessment of spatio-temporal pattern of urban earthquake vulnerability using gis: a study on dhaka city,” *Annals of GIS*, vol. 19, pp. 63 – 78, 2013.

[64] R. Bian and C. Wilmot, “Spatiotemporal population distribution method for emergency evacuation,” *Transportation Research Record*, vol. 2532, pp. 106 – 99, 2015.

[65] C. Botezan, A. Radovici, and I. Ajtai, “The challenge of social vulnerability assessment in the context of land use changes for sustainable urban planning—case studies: Developing cities in romania,” *Land*, 2021.

[66] X. Zhang, Y. Chai, Z. Chen, and Y. Tan, “Analysis of spatial and temporal patterns of daily activities of suburban residents based on gps data: A case study of the shangdi-qinghe area of beijing,” *International Review for Spatial Planning and Sustainable Development*, vol. 4, pp. 4–16, 2016.

[67] J. M. Fields, *The Timing of Noise-Sensitive Activities in Residential Areas*. Bionetics Corporation, 1985, 1985.

[68] J. Veitch, A. Carver, G. Abbott, B. Giles-Corti, A. Timperio, and J. Salmon, “How active are people in metropolitan parks? an observational study of park visitation in australia,” *BMC Public Health*, vol. 15, 2015.

[69] H. Yu, S. C. Mulhearn, H. van der Mars, J. M. Griffo, and P. H. Kulinna, “Off-hours facility use in public high schools: Physical activity potential vs. reality,” *Research Quarterly for Exercise and Sport*, vol. 93, pp. 341 – 351, 2020.

[70] S. Williams, “Chi-square test for goodness of fit,” in *Encyclopedia of Measurement and Statistics*, 2007.

[71] Istanbul Metropolitan Municipality, “Avrupa yakası güneyi i. etap mikrobölgeleme Çalışması: Yönetici Özeti,” 2007. Accessed: 2024-08-10.

[72] Istanbul Metropolitan Municipality, “Anadolu yakası mikrobölgeleme Çalışması: Yönetici Özeti,” 2009. Accessed: 2024-08-10.

APPENDIX A

GENERATED FISHNETS

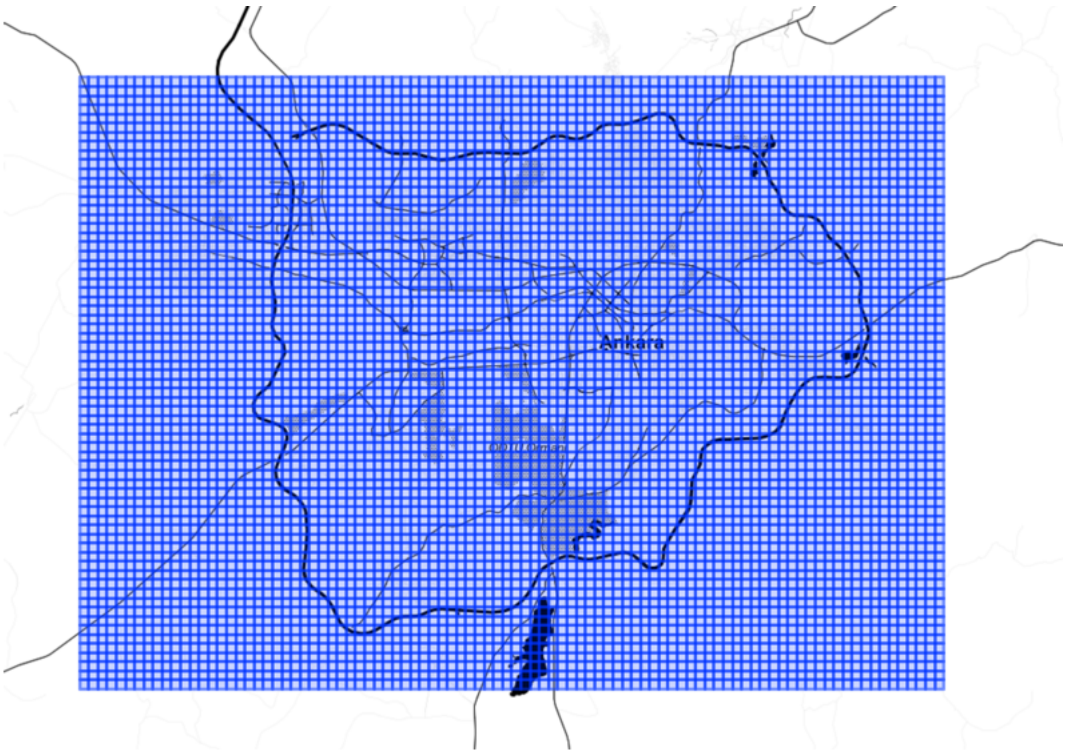


Figure A.1: Fishnet covering the city of Ankara.

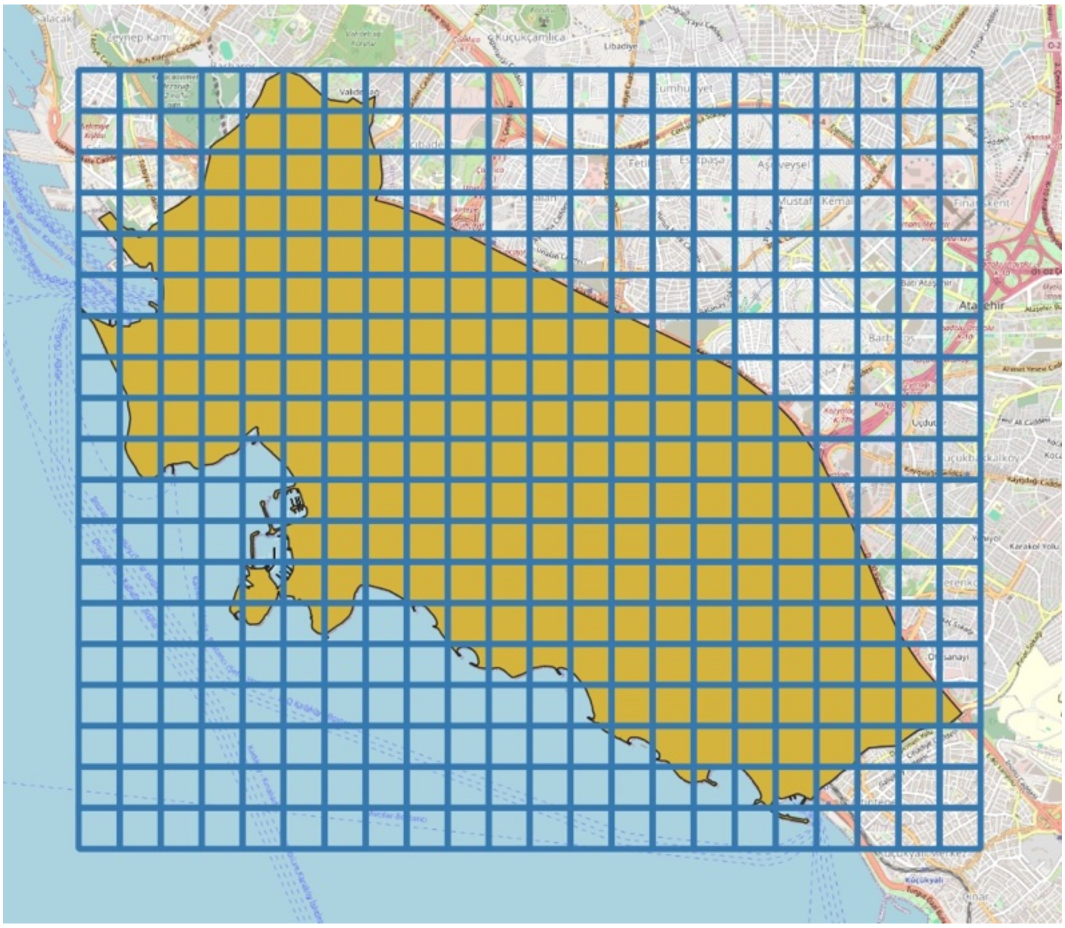


Figure A.2: Fishnet covering the Kadıköy district.

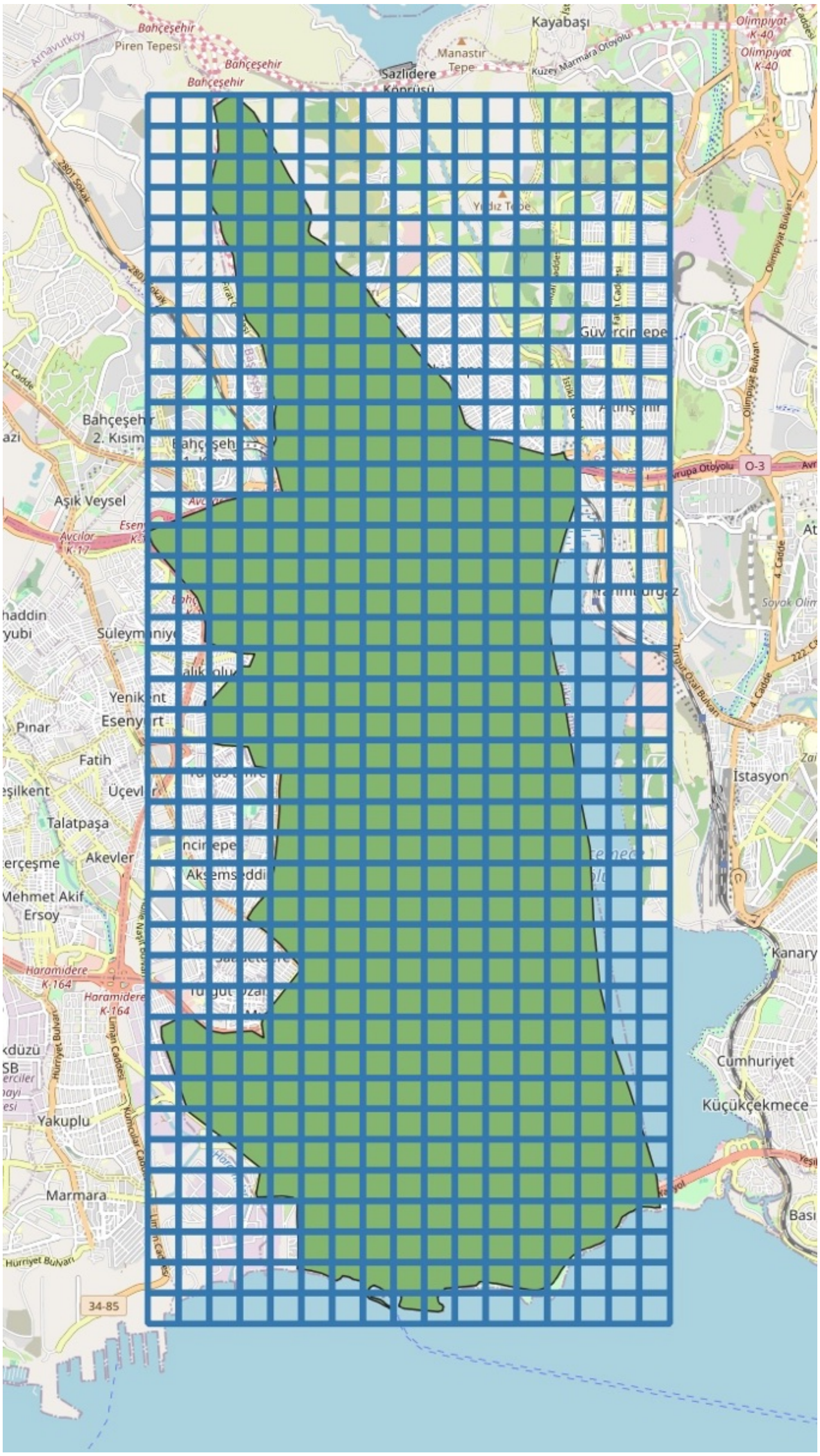


Figure A.3: Fishnet covering the Avcılar district.

APPENDIX B

POI DATA DETAILS

Table B.1: All the available labels in the Google Maps API.

accounting	lawyer	park
airport	library	parking
amusement_park	light_rail_station	pet_store
aquarium	liquor_store	pharmacy
art_gallery	local_government_office	physiotherapist
atm	locksmith	plumber
bakery	lodging	police
bank	meal_delivery	post_office
bar	meal_takeaway	primary_school
beauty_salon	mosque	real_estate_agency
bicycle_store	movie_rental	restaurant
book_store	movie_theater	roofing_contractor
bowling_alley	moving_company	rv_park
bus_station	museum	school
cafe	night_club	secondary_school
campground	painter	shoe_store
car_dealer	electronics_store	shopping_mall
car_rental	embassy	spa
car_repair	fire_station	stadium
car_wash	florist	storage
casino	funeral_home	store
cemetery	furniture_store	subway_station
church	gas_station	supermarket
city_hall	grocery_or_supermarket	synagogue
clothing_store	gym	taxi_stand
convenience_store	hair_care	tourist_attraction
courthouse	hardware_store	train_station
dentist	hindu_temple	transit_station
department_store	home_goods_store	travel_agency
doctor	hospital	university
drugstore	insurance_agency	veterinary_care
electrician	laundry	zoo

Table B.2: All of the districts in Ankara with the number of ads collected per district.

District	Ad Count
Çankaya	44083
Keçiören	9040
Yenimahalle	8306
Etimesgut	7981
Altındağ	3444
Pursaklar	2565
Mamak	2110
Sincan	1984
Gölbaşı	1663
Polatlı	721
Kazan	173
Çubuk	127
Akyurt	91
Elmadağ	65
Ayaş	17
Beypazarı	17
Çamlıdere	16
Bala	15
Kızılcahamam	7
Haymana	5
Şereflikoçhisar	5
Güdül	4
Nallıhan	3
Kalecik	1

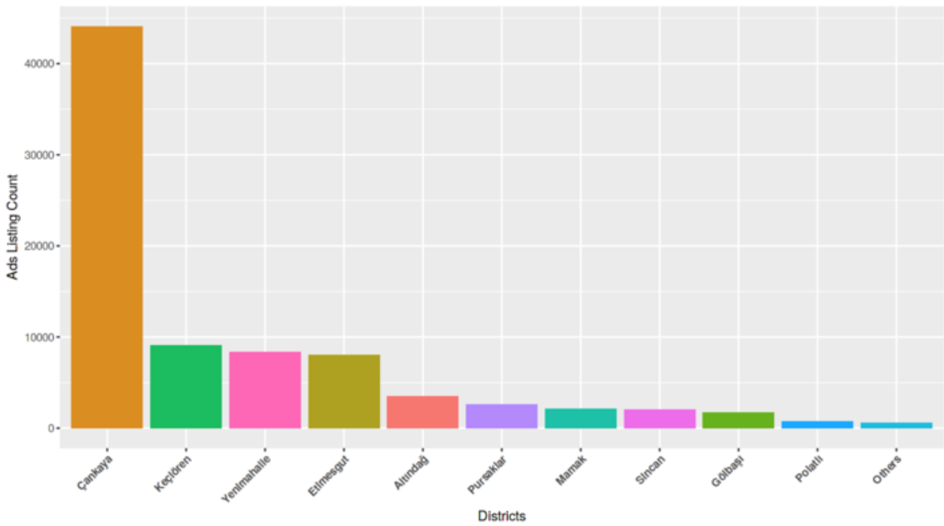


Figure B.1: Map showing the scraped ads listings by top districts, the y-axis represents the number of ads collected in that specific area.

APPENDIX C

PYTHON MODULE CODE FOR LANDUSEMIX PACKAGE

```python
"""
This module contains a class that calculates land use mix indices.
"""
import math

class LandUseMixIndices:
    """
    A class that calculates land use mix indices.

    Attributes:
        land_use_areas (dict): A dictionary containing land use areas.

    Methods:
        entropy_index: Calculates the entropy index.
        herfindahl_hirschman_index: Calculates the Herfindahl-Hirschman.
    """

    def __init__(self, land_use_areas):
        """
        Initializes a LandUseMixIndices object.

        Args:
            land_use_areas (dict): A dictionary containing land use areas.
        """
        self.land_use_areas = land_use_areas

    def entropy_index(self):
        """
        Calculates the entropy index.

        Returns:
            float: The entropy index value.
        """
        total_area = sum(self.land_use_areas.values())
```

```
        k = len(self.land_use_areas)
        if k == 1:
            return 0
        entropy = -sum((area / total_area) * math.log(area / total_area)
                       for area in self.land_use_areas.values()) / math.log(k)
        return round(entropy, 2)

    def herfindahl_hirschman_index(self):
        """
        Calculates the Herfindahl-Hirschman index.

        Returns:
            float: The Herfindahl-Hirschman index value.
        """
        total_area = sum(self.land_use_areas.values())
        hhi = sum((100 * (area / total_area)) **
                  2 for area in self.land_use_areas.values())
        return round(hhi, 2)
```

CURRICULUM VITAE

Personal Information

Name: Mehmet Ali Akyol
Place of Birth: Muğla, Türkiye

Education

- Ph.D. in Information Systems, METU, ongoing
- M.S. in Information Systems, METU, 2017
- B.A. in Business Administration, Anadolu University, 2016
- B.S. in Electrical and Electronics Engineering, Bilkent University, 2013
- Muğla Science High School, 2008

Professional Experience

- R&D Engineer, Beko, 2021-ongoing
- Founder/Software Developer, Startup Buffer, 2016-2021
- Research Assistant, METU, 2014-2020
- Systems Design Engineer, Aselsan, 2013-2014

Selected Publications

1. **Akyol, Mehmet Ali**; Düzgün, Şebnem; Baykal, Nazife; "landusemix: A Python package for calculating land use mix", SoftwareX, vol. 27, p. 101861, 2024, Elsevier.
2. **Akyol, Mehmet Ali**; Temizel, Tuğba Taşkaya; Düzgün, Şebnem; Baykal, Nazife; "Identification of Land Use Mix Using Point-Based Geospatial Data in Urban Areas," Applied Sciences, vol. 14, no. 16, pp. 6871, 2024, MDPI.
3. Salah, Albert Ali; Altuncu, M Tarık; Balcisoy, Selim; Frydenlund, Erika; Mamei, Marco; **Akyol, Mehmet Ali**; Arslanlı, Kerem Yavuz; Bensason, Ivon; Boshuijzen-van Burken, Christine; Bosetti, Paolo; "Policy implications of the D4R challenge," Guide to mobile data analytics

in refugee scenarios: the 'Data for Refugees Challenge' study, pp. 477-495, 2019, Springer International Publishing.

4. Kılıç, Özgün Ozan; **Akyol, Mehmet Ali**; Işık, Oğuz; Kılıç, Banu Günel; Aydınoğlu, Arsev Umur; Surer, Elif; Düzgün, Hafize Şebnem; Kalaycıoğlu, Sibel; Temizel, Tuğba Taşkaya; "Data Analytics without Borders: Multi-Layered Insights for Syrian Refugee Crisis," Proceedings of the Data for Refugees Challenge Workshop, 2019.

5. **Akyol, Mehmet Ali**; Kılıç, Özgün Ozan; Taşkaya Temizel, Tuğba; Aydınoğlu, Arsev Umur; Sürer, Elif; "Sosyal Fayda için Büyük Veri Analizi: Suriyeli Mülteciler için bir Durum Çalışması," 2019.

6. Gökalp, Mert Onuralp; Kayabay, Kerem; **Akyol, Mehmet Ali**; Koçyiğit, Altan; Eren, P Erhan; "Big data in mHealth," Current and Emerging mHealth Technologies: Adoption, Implementation, and Use, pp. 241-256, 2018, Springer International Publishing.

7. **Akyol, Mehmet Ali**; Gökalp, Mert Onuralp; Kayabay, Kerem; Eren, P Erhan; Koçyiğit, Altan; "A context aware notification architecture based on distributed focused crawling in the big data era," Information Systems: 14th European, Mediterranean, and Middle Eastern Conference, EMCIS 2017, Coimbra, Portugal, September 7-8, 2017, Proceedings 14, pp. 29-39, 2017, Springer International Publishing.

8. Kayabay, Kerem; **Akyol, Mehmet Ali**; Gökalp, Mert Onuralp; Koçyiğit, Altan; Eren, Pekin Erhan; "Current Research Topics in Industry 4.0 and an Analysis of Prominent Architectural Frameworks,", 4th International Management Information Systems Conference"Industry 4.0", 2017.

9. Gokalp, Mert Onuralp; Kayabay, Kerem; **Akyol, Mehmet Ali**; Eren, P Erhan; Koçyiğit, Altan; "Big data for industry 4.0: A conceptual framework," 2016 International Conference on Computational Science and Computational Intelligence (CSCI), pp. 431-434, 2016, IEEE.

10. Kayabay, Kerem; Gökalp, Mert Onuralp; **Akyol, Mehmet Ali**; Koçyiğit, Altan; Eren, P. Erhan; "Big Data for Future Enterprises: Current State and Trends," 3rd International Management Information Systems Conference, İzmir, Turkey, pp. 298-307, 2016.