

Article

Not peer-reviewed version

Enhancing Urban Resilience: Smart City Data Analyses, Forecasts, and Digital Twin Techniques at the Neighborhood Level

[Andreas F. Gkontzis](#) ^{*}, [Sotiris Kontsiantis](#), [Georgios Feretzakis](#), [Vassilios S. Verykios](#)

Posted Date: 12 January 2024

doi: 10.20944/preprints202401.0967.v1

Keywords: Smart Cities; Digital Twins; Predictive Analytics; Artificial intelligence; Urban Resilience; Sustainable Urban Development; GeoDataFrame; Python; Time Series Analysis; Citizens Reports



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Article

Enhancing Urban Resilience: Smart City Data Analyses, Forecasts, and Digital Twin Techniques at the Neighborhood Level

Andreas F. Gkontzis ^{1,*}, Sotiris Kotsiantis ², Georgios Feretzakis ³ and Vassilios S. Verykios ⁴

¹ Representative of the Innovation Network and Best Practices of Patras Municipality; Municipality of Patras; Patras; Greece; a.gontzis@patras.gr

² Assistant Professor; Department of Mathematics; University of Patras; Patras; Greece; kotsiantis@upatras.gr

³ Adjunct Faculty & Researcher; School of Science and Technology; Hellenic Open University; Patras; Greece; georgios.feretzakis@ac.eap.gr

⁴ Professor; School of Science and Technology; Hellenic Open University; Patras; Greece; verykios@eap.gr

* Correspondence: a.gontzis@patras.gr; Tel.: +306958464166

Abstract: Smart cities, leveraging advanced data analytics, predictive models, and digital twin techniques, offer a transformative model for sustainable urban development. Predictive analytics plays a crucial role in proactive planning, enabling cities to adapt to evolving challenges. Concurrently, digital twin techniques provide a virtual replica of the urban environment, fostering real-time monitoring, simulation, and analysis of urban systems. This research underscores the significance of real-time monitoring, simulation, and analysis of urban systems to support test scenarios that identify bottlenecks and enhance smart city efficiency. The paper delves into the crucial roles of citizen report analytics, prediction, and digital twin technologies at the neighborhood level. The study integrates ETL/ELT processes, AI techniques, and a digital twin methodology to process and interpret urban data streams derived from citizen interactions with the city's coordinate-based problem mapping platform. By employing an interactive GeoDataFrame within the digital twin methodology, dynamic entities facilitate simulations based on various scenarios. This approach enables users to visualize, analyze, and predict the response of the urban system at the neighborhood level. Consequently, antecedent and predictive patterns, trends, and correlations are visualized at the physical level of each city area, leading to improvements in urban functionality, resilience, and resident quality of life.

Keywords: Smart Cities; Digital Twins; Predictive Analytics; Artificial intelligence; Urban Resilience; Sustainable Urban Development; GeoDataFrame; Python; Time Series Analysis; Citizens Reports

1. Introduction

Urban environments are dynamic, intricate ecosystems with challenges that necessitate innovative solutions. In response to the complexities of contemporary city management, the implementation of digital technologies has emerged as a transformative force. The discourse surrounding urban development has evolved significantly, with the emergence of smart cities as a pivotal concept. The objective is clear: to cultivate urban environments that are sustainable, resilient, and centered around the needs of citizens. This transformative paradigm relies on cutting-edge technologies to elevate diverse aspects of urban life, spanning from infrastructure to public services. At the heart of smart city development lies the integration of data analyses, forecasting models, and digital twin techniques, particularly at the neighborhood level. This paper delves into the critical dimensions of these technologies, aspiring to optimize urban functionality, fortify resilience, and enhance the overall quality of life for residents [1,2].

The multifaceted concept of urban resilience encapsulates a city's capacity to adapt, recover, and prosper in the face of myriad challenges—be they demographic shifts, environmental transformations, or unforeseen crises. Smart city initiatives emerge as strategic approaches to fortify urban resilience, capitalizing on the symbiosis of data and technology. These initiatives prioritize the meticulous collection, analysis, and application of extensive urban data to guide decision-making, improve resource allocation, and facilitate proactive planning [2,3]

In this direction, predictive models are keystones for predicting and preparing for future urban scenarios. Leveraging both historical and real-time data, these models predict trends and potential challenges, empowering cities to adopt a proactive stance in addressing issues before they escalate. The integration of predictive analytics into smart city frameworks allows decision-makers to make informed choices regarding infrastructure development, resource distribution, and policy implementation—a critical foresight for cities aiming not only to survive but to thrive in an ever-evolving urban landscape [1,3–5].

This prospect introduces smart cities to digital twin techniques that provide a virtual replica of the urban environment, offering a dynamic and comprehensive simulation of various urban systems such as transportation networks, energy grids, and public services. This digital twin serves as a potent tool for real-time monitoring, scenario testing, and optimization. By crafting a detailed virtual counterpart of the city, decision-makers gain the ability to identify potential bottlenecks, optimize resource utilization, and enhance overall efficiency. This staggered progression through smart city concepts underscores the interconnectedness of these transformative technologies in shaping sustainable urban development [2,4].

This study delves into the realm of urban innovation, focusing on the development and application of a Digital Twin in the city of Patras. Comprising a combination of time series analysis, citizen feedback data, and advanced technologies, this Digital Twin project offers a comprehensive approach to urban management and decision-making. Section 2 navigates through the intricate process of materials and methods, providing a detailed blueprint of the Digital Twin's creation. A custom Python solution, utilizing the Flask web framework, lays the foundation for a simulation and modeling platform. The incorporation of a development server and the utilization of web technologies, exemplified by the interactive map interface, create a robust infrastructure for data handling and communication between the back end and the front end. The next section (section 3) unfolds the Results, presenting the tangible outcomes of the time series analyses and Digital Twin's implementation. Through dynamic predictive visualizations and interactive features, Digital Twin embodies the multifaceted nature of urban issues. The insightful time series analyses, coupled with the forecasted presentation of clusters, markers, and layers, paint a vivid picture of the city's pulse. The application's evolution, as showcased in various updates, culminates in an enriched user experience equipped with layers, filters, and charts for enhanced exploration.

Moving beyond the technical intricacies, section 4 initiates the discussion, engaging in an extended analysis of the results. Drawing key lines from contemporary literature, the study positions the Digital Twin in Patras within the broader context of global urban management initiatives. A comparative discussion with studies analyzing citizen feedback data elucidates the project's resonance with existing trends, emphasizing the project's temporal analysis and machine learning integration. The last section, section 5 encapsulates the essence of this exploration in the Conclusions. Beyond a mere technological endeavor, the Patras Digital Twin emerges as a paradigm shift in urban planning and management. Informed by insights from the literature, the project's potential for optimizing infrastructure, improving public services, and facilitating data-driven decision-making positions it in the core realm of smart city initiatives.

Together, these sections provide a comprehensive narrative of the Digital Twin ecosystem in Patras, unraveling the intricacies of its creation, presenting tangible outcomes, engaging in critical discourse, and ultimately envisioning a future where cities are not just managed but dynamically understood and optimized through the lens of digital novelty.

Related Work

Digital twins serve as virtual copies of physical entities, whether animate or inanimate. The growing appeal of digital twins can be attributed to the tools that facilitate the management of digital twin data, services, modeling, and seamless integration with the physical world. These tools leverage technologies such as machine learning, artificial intelligence, and data analytics to create dynamic digital simulation models, and the models are constantly adapted and synchronized with their real-world counterparts. A digital twin is in a constant state of learning and self-updating, drawing insights from diverse data sources to mirror the physical object. This learning process involves self-reflection, insights gleaned from other similar systems, and knowledge contributed by human experts possessing relevant domain expertise. Furthermore, a digital twin can assimilate knowledge from historical data, incorporating insights from past usage patterns into its digital representation. The digital layer of a smart city plays a crucial role in aggregating the necessary data for simulations within the virtual layer, or digital twin. Subsequently, the outcomes of these simulations are communicated back through the various layers of the city as valuable information. The data sources are diverse and may include inputs from citizens, as well as data from mobile and strategically positioned devices and assets dispersed throughout the urban landscape. Citizens actively contribute to this data ecosystem by utilizing their mobile phones to report information to city authorities. This decentralized approach ensures that the digital layer receives real-time inputs from individuals who can communicate observations, concerns, or relevant data, enhancing the city's overall situational awareness. In essence, the smart city's digital infrastructure serves as a conduit for data flow, facilitating dynamic interactions between the physical environment and its digital representation in the virtual layer.

Towards this end, the paper of [6] introduces an openly accessible digital twin smart cities model focusing on the Docklands area in Dublin, Ireland. The implementation utilizes the commercial software Unity3D (Unity), to load and render the digital twin. Unity3D is employed as the platform for visualizing and interacting with the digital twin, providing a dynamic and immersive experience for users exploring the smart city model. The availability of this digital twin model contributes to the open and transparent development of smart city infrastructure, fostering collaboration and understanding of urban dynamics in the Docklands area of Dublin. The smart city's simulations in the digital twin derive their accuracy from the wealth of data generated within the city itself. Dublin, as an example, contributes extensively to this data pool by making a significant amount of information accessible through the Dublinked open data source. This repository encompasses various data sources related to population, transportation, infrastructure, environmental conditions, and energy consumption. Despite the availability of comprehensive data, a notable gap exists in the fine-grained citizens' interaction data. This absence limits the depth of simulations related to citizens' concerns, potentially resulting in less meaningful and detailed insights. The availability of detailed citizens' interaction data is crucial for a more comprehensive understanding of urban dynamics and can significantly enhance the fidelity of simulations within the digital twin, allowing for more accurate and nuanced representations of the city's social fabric and the concerns of its inhabitants. Efforts to address this data gap can lead to more robust and insightful smart city simulations.

In the study of [7], the researchers aim to develop a digital twin and explore the immersive behaviors of non-expert users in the context of community design and policy decision-making. They highlight the limited research available on citizen participation and integration into a city Information digital twin model, emphasizing its crucial role in realizing a human-centered digital replica of the city. The study focuses on a digital representation of the physical and functional characteristics of the Malvalaan urban neighborhood, consisting of four to five blocks located in the Brabant municipality of Waalre, south of Eindhoven in the Netherlands. This representation was tested in a real-life urban community involving both expert and non-expert stakeholders. The integration of tangible and intangible data, encompassing mobility, road security, and social participation, is thoroughly explained in the paper. This integration empowers local stakeholders to contribute to community design optimization and enhances policy decision-making outcomes. The geo blueprint of the Malvalaan City Information Model (CIM) was generated through aerial LiDAR scanning, utilizing a

photorealistic model of Dutch cities. Commercial software, including Autodesk Revit and SketchUp, was employed to create appropriate content. The study assessed ease of use and usability through a questionnaire, including Likert-scale questions, to comprehensively understand the user experience. Ease of use was evaluated based on users' ability to navigate and make decisions within the virtual environment, while usability was assessed by evaluating users' interaction with various platform elements and their ability to provide geotagged feedback in design scenarios. The detailed analysis of the questionnaire data reveals that, overall, the Malvalaan digital twin successfully realized a user-friendly community design exercise. The platform effectively explored optimal design scenarios, analyzed them, and facilitated users in providing geotagged feedback, demonstrating its potential for enhancing community support, engagement, and decision-making processes.

The findings of the researchers in [8] reveal that the concept of Citizen-Centric Digital Twin (CCDT) has been steadily gaining traction, owing to its predictive, simulative, and visualizing capabilities. The research methodology of this study is based on a systematic literature review of 210 relevant articles based on digital twins and smart cities from 2011 to 2022. The researchers highlight CCDT's potential to address various issues at the city level by actively involving citizens in infrastructure governance. This involvement encompasses issue reporting, providing feedback for city planning and policy decisions, raising concerns about city cleanliness, and identifying environmental hazards. The researchers emphasize that a dynamic and real-time nature is intrinsic to citizen-centric digital platforms, ensuring they provide up-to-date information that encourages citizen participation and promotes quality infrastructure governance. Competent authorities can benefit from both human-centric and sensor-based approaches to validate decisions. However, data derived from voluntary geographic information and social sensing are two prominent methods for centralized digital twin platforms. Most of the projects use data from open-source data platforms, as well as the most accessible data forms for researchers to build their digital twin. While open-source data platforms are commonly used as the foundation for digital twins, social sensing and voluntary geographic information contribute to real-time situational knowledge, enhancing decision-making in infrastructure governance. The article underscores the importance of common data analysis algorithms in citizen-centric digital twins to extract comprehensive information about citizen-centric cities, thereby ensuring effective infrastructure governance. The researchers propose the widespread use of machine learning classifiers for CCDT, capable of performing diverse tasks such as integrating and triangulating sparse data. Utilizing change detection algorithms on remote sensing data, one can identify variations among spatially registered information, such as the dispersion of infrastructure. These algorithms play a crucial role in tasks like analyzing infrastructure dispersion. Furthermore, data derived from geographic information and social sensing, furnish insights into the causes, dispersion patterns, and the severity of the observed phenomena. The study also identifies application processing interfaces for CCDT that facilitate interactive visualization for informed decision-making during citizen participation in infrastructure governance. CesiumJS is highlighted as a significant API, particularly in digital city twin applications in capital cities like Victoria and Sydney. However, it is a commercial platform and the free version is only for personal and not public projects, exploratory development, or educational activities.

The article [9] highlights the growing interest in Digital Twins (DT) in urban and geospatial domains, particularly at the city-scale level. The implementation of Digital Twins for cities faces challenges due to the lack of a common definition, resulting in various frameworks and implementations in practice. The study aims to bridge the gap between conceptual understanding and current realizations of Digital Twins for cities by analyzing literature and survey results. The research collects definitions of Digital Twins for cities and compares them with related concepts used in the literature. A corpus of 26 papers is obtained by refining the results, focusing on computer science, urban planning, and city-level geospatial fields. A key finding underscores the importance of clearly defining concepts and approaches to understand the practical input layer of Digital Twins for cities. Current implementations often confuse the concept with other notions, sometimes treating it as a technological evolution of 3D urban modeling. The study suggests that considering Digital Twins for cities as systems of systems is a prominent approach, with 15 papers defining it as a digital

representation of city elements functioning as an ecosystem of twin systems, without limiting the scope to 3D components. The review analysis also reveals that the City Information Model (CIM) is primarily regarded as the input layer of Digital Twins for cities. Additionally, the establishment of Spatial Data Infrastructure (SDI) is considered crucial for creating a robust geospatial Digital Twin. SDI is seen as the cornerstone for ensuring seamless access to the data hub, along with efficient and standardized data integration.

Through an in-depth examination of 184 pertinent articles carefully selected and analyzed, the paper [10] synthesizes the various disciplinary classifications associated with the functional integration of Geographic Information Systems (GIS) and Building Information Model (BIM). The study distills the significance of data and advocates for an ontology-based data integration approach, specifically emphasizing its application in research on integration within smart cities. While the application cases of smart city platforms have made initial strides in achieving analytical computing through GIS and BIM techniques, ongoing research is dedicated to enhancing interoperability. The researchers stress that the improvement of interoperability in digital twin technology, coupled with the integration of richer data formats, holds the key to developing a more cohesive smart city application system. Consequently, an integrated platform must embody numerous semantic attributes that seamlessly interact with the physical layer of the city. By emphasizing the co-calculation of the importance of time management cycles, a seamless transition and integration can be achieved across various stages, spanning from surveying, planning, and design to construction, operation, and maintenance. This holistic approach not only streamlines processes but also contributes to the reduction of corresponding costs throughout the entire lifecycle of a smart city project.

The researchers in [11] emphasize the evolution of the digital twin concept as a means of extracting value from data and highlight its growing role in the design and management of diverse systems. They discuss the proliferation of proprietary digital twin software solutions in various domains and introduce an open-source software framework for digital twins that stands out as a recently released solution. This framework, operationalized through a browser-based platform using Python and Flask, aims to enhance connectivity between users and data sourced from the physical twin. The paper illustrates how this Digital Twin Operational Platform (DTOP) facilitates the linkage of the physical twin and Internet of Things (IoT) devices to both users and cloud computing services. The choice of Python as the core programming language is justified by its dominance in web-based and scientific applications. Despite competition with commercial solutions, Python's versatility enables the execution of a multitude of algorithms, thanks to widely available libraries for vectorized computing. Within the Python ecosystem, Flask emerges as a pivotal server tool allowing for deployment based on Internet Protocol (IP), facilitating remote hosting and access. Flask's flexibility enables the separation of web interface design from the development of underlying mathematical code, covering tasks such as information dispatching, data recording, and mathematical simulations. The authors underline the vast area for future research and development in open-source digital twin software employing network connectivity. The design of the web interface, accomplished using HTML5, CSS, and JavaScript, ensures a responsive, interactive, and animated browser-based graphical user interface. Flask, in conjunction with Python, enables seamless connectivity among various micro components, spanning local computing, data management, and graphical interfacing. This setup empowers front-end users to interact with the digital twin through a web browser, triggering different computations without requiring specific programming knowledge. Simultaneously, back-end users can download the open-source project, making customized modifications to align with project-specific requirements.

The study [12] has undertaken a scoping review of 162 papers across various urban fields to address the existing gap in the realm of urban sustainability and spatial data analysis. The overview provides insights into the scope of studies utilizing Artificial Intelligence (AI) and Machine Learning (ML) with geospatial data for the analysis of urban areas. The review, conducted using Web of Science to ensure academic rigor and validity, spans the years 2014 to 2021 and reveals a concentration of cases in China and the US, with the UK following. A regional breakdown indicates that 31% of case

studies were in Europe, 29% in Asia, and 27% in North America. The analysis exposes knowledge gaps in ML methods for spatial data science and highlights the need for data specifications to guide future research. While ML methods are gaining popularity, the review identifies a lack of a comprehensive overview across urban domains, hindering researchers from comparing and selecting the most suitable methods for their topics. The existing reviews on remote sensing have demonstrated the effectiveness of Support Vector Machines (SVM), Random Forests (RF), and Boosted Decision Trees (DTs) as powerful methods for the classification of remotely sensed data. However, these studies predominantly focus on the detection and monitoring of the physical surface of the world through remotely sensed images. What remains a gap in the literature is the exploration of the relationship between the physical features of a city and its functions and sustainability. Consequently, this scoping review aims to fill this void by concentrating on studies that primarily leverage geospatial data for the assessment of urban sustainability. The emphasis is on understanding how the spatial characteristics of urban areas are linked to their functionality and long-term viability, providing a more comprehensive perspective on urban studies beyond surface-level classification based on remote sensing imagery. The outcomes underscore the effectiveness of certain algorithms in predicting or classifying data for specific problems, leading to their strong reputation over time. In the selection of supervised machine learning (ML) methods, various factors play a crucial role, including complexity, overfitting properties, parameter requirements, data requirements, and interpretability of results. Complexity refers to the intricacy of the model in capturing relationships within the data. Random Forest (RF) and Support Vector Machines (SVM) are considered more complex compared to simpler models like logistic regression, ordinary least square regressions, or LASSO, primarily because RF and SVM account for non-linear relationships in the data. When considering overfitting properties, a phenomenon where a model performs exceptionally well on training data but poorly on unseen data, RF is often favored. Random Forest is known for its capability to handle overfitting, providing a more robust performance on new, unseen data. This makes RF a popular choice in situations where preventing overfitting is crucial for the model's generalization to new data and real-world scenarios. The study emphasizes the importance of comparing different ML algorithms within specific topics to understand potential differences in results and their implications for urban sustainability. Furthermore, the researchers stress the need for new methodological frameworks that go beyond the application of ML, aiming to explain, translate, and transfer results from ML to urban sciences, practice, and decision policies. While ML and AI gain momentum, the authors emphasize that these technologies should serve to address pressing sustainability issues, such as circularity and resilience. They advocate for standardizing the selection of data, algorithms, and parameters to provide concrete reasoning and clarity for sustainable urban planning decisions and policies. The review also acknowledges limitations in developing spatial ML studies in data-sparse areas, particularly in the Global South.

In this connection, the main goal of the study [13] is to develop a data-driven prediction model for water main failures in the City of Kitchener Ontario Canada; and, hence, reduce field inspections through proactive interventions. Reducing field inspections means limiting field inspections to only those pipes that are more likely to fail in the future. Those pipes can be identified from the developed prediction model. This will contribute to replacing emergency replacements and post-failure repairs with preventive maintenance practices that are less costly. Using such information, decision-makers can cost-effectively update their asset management plans and prioritize replacement plans more effectively. The results showed that more than 72% of issues could have been potentially prevented by monitoring and upgrading only 8% of the network. Six classification algorithms were used in this study, namely: Naïve Bayes (NB), Decision Tree (DT), k-nearest Neighbors (k-NNs), Logistic Regression (LR), Artificial Neural Network (ANN), and Random Forest (RF). From the six machine learning prediction models that were developed, the random forest models outperformed the other machine learning models. This study showed that, when available, the use of city condition indicators can significantly improve problem predictability. Additionally, analyzing factors that contribute to the city's issues provides valuable insights into the root causes of incidents. Identifying the most

strongly correlated factors helps develop accurate predictive models that enable informed decisions about maintenance and repair, improving the reliability, safety, and resilience of city infrastructure.

In the comprehensive analysis of 312 titles and abstracts, along with a detailed examination of 72 full papers, the article [14] explores the intersection of digital twins (DT) and smart cities (SC) in the context of disaster risk management. The review assesses the evolution of DTSCs, explores the intelligent technologies utilized in DTSCs for disaster risk management, and evaluates the technical feasibility of DTSC-driven approaches in this domain. The expertise required to guide the process of utilizing DTSCs for disaster risk reduction, resilience, and sustainable development is also scrutinized. The sustainability of such projects hinges on continuous capacity building and maintenance, underscoring the importance of establishing a robust knowledge infrastructure that considers interactions, synergies, and trade-offs among various stakeholders. The article highlights the need for enhanced techniques to capture the spatial and temporal aspects of disruptions to the built environment and people's adaptive responses. The authors suggest a focus on geoparsing, which involves estimating the locations of users' posts and assessing them against external geographic entity data. They propose the need for the development of a real-time resilience mapping system utilizing geoparsing techniques to extract street-level named entities. Additionally, the article underscores the potential of artificial intelligence (AI) as a prediction system for forecasting natural disasters using high-quality datasets. AI can analyze historical data to identify patterns and anticipate the probabilities of specific disaster occurrences, offering a more accurate and informed approach than conventional methods. AI-based systems leverage learning from historical data, enhancing decision-making options and providing feedback comparable to human judgment. The scalability of AI allows for a deeper understanding of large-scale social phenomena by processing vast amounts of data, selecting essential real data through big data analysis, and simulating scenarios using AI-driven probabilities based on empirical data. The core principle of the Digital Twin for Smart Cities (DTSC) lies in the sensing and monitoring of the status and attributes of the built environment, human systems, and their interconnections. The authors emphasize a significant gap in existing research, particularly the limited focus on identifying crucial infrastructure disruptions and understanding people's responses to such disturbances. This scarcity is attributed to the constraints of geosocial sensing techniques and in gathering information on infrastructure issues. Consequently, a suitable social sensing approach for DTSC should encompass the frequency and impact of city problems, their ramifications on specific city areas and individuals, and the adaptive measures taken in response. The authors highlight the potential contribution of DTSC to strengthening the socio-economic fabric of society and its capacity to enhance resilience by assisting adaptation efforts to meet the challenges of risk management in a dynamically changing environment.

The utilization of Location-Based Data (LBD) for city analysis has garnered significant attention as a promising method for applied research [15]. The paper of [15] delves into the challenges, opportunities, and limitations associated with collecting and using LBD from various platforms for diverse cities in the context of urban phenomena research. The study presents a comprehensive method for requesting and retrieving data from Application Programming Interfaces (APIs), outlining the procedures for obtaining, verifying, filtering, and classifying data. The findings emphasize the need for a meticulous review and, at times, distinct verification steps to preserve the implicit nuances within each dataset derived from geolocated user-generated data, tailoring the approach to each case study. The researchers assert that the use of geodata in city studies is an ongoing trend due to the qualitative insights generated by an exponentially growing community of digitized users. These datasets have the potential to reflect current trends in urban reality more promptly and comprehensively than traditional sources. The study underscores the importance of addressing challenges, particularly the reliance on third-party platforms for geolocation-based data accessibility, which renders the retrieval process susceptible to changes in access conditions. Additionally, the verification of geolocation-based data within large datasets is crucial for achieving more representative sampling and better outcomes in studying urban phenomena. Given the complexity of requesting and retrieving data, the research highlights the importance of properly handling API requirements concerning the shape and size of searches and the number of results per

request. This attention to detail is essential for ensuring the reliability and validity of the collected data for urban research.

The article [16] introduces a methodology that goes beyond conventional constraints, offering a holistic perspective on urban spaces. The approach incorporates both longitudinal and cross-sectional assessments, utilizing geodata processing and semantic segmentation models to derive key metrics such as the green coverage rate and green view index. Osaka City in Japan serves as the exemplar for this study, and the results reveal notable facts, including an uneven distribution of green spaces and variations in green space quality. Consequently, the study advocates for targeted interventions, suggesting improvements in areas with low green visibility and enhancements to existing green spaces. This approach aims to fortify Osaka's urban greening initiatives. The study underscores the limitations of traditional evaluation methods for urban green spaces and emphasizes the necessity for a comprehensive, multidimensional approach rooted in the geographical context and other relevant characteristics. Researchers stress the importance of moving beyond isolated metrics and exploring the interplay and mutual influences of various geolocated indicators. Furthermore, the study highlights the opportunities presented by technological advancements and the availability of street view data to enhance the evaluation and understanding of urban spaces. The integration of these technological tools offers a more nuanced and comprehensive assessment, contributing to a more informed and effective approach to urban planning and greening initiatives.

The paper of [17] conducts a comprehensive literature review encompassing four hundred papers focused on mobility-oriented urban studies. The review underscores the dynamic nature of mobility data, emphasizing its pivotal role in enhancing our comprehension of human mobility patterns and the intricate interactions between individuals and urban environments. Emphasizing the significance of the vast volume of big geodata related to human mobility, the paper highlights its potency as a valuable resource, empowering researchers to unravel complex knowledge and foster the development of innovative applications within urban contexts. The wealth of mobility data enables researchers not only to trace individuals' movement trajectories but also to characterize their activities and behaviors in their daily lives. This holistic perspective is pointed out that establishes a foundation for extracting insights into human mobility across diverse urban scenarios. By forging connections between individuals' mobility patterns and their daily activities, researchers can unveil novel perspectives on urban planning, conduct detailed analyses of neighborhood segregation, and recommendation services. Importantly, this approach integrates a cyber-physical-social perspective, underscoring the interconnected nature of human mobility and its far-reaching implications for the dynamics of the urban environment.

The study in [18] pointed out that although the benefits of smart city development in enhancing cities' infrastructure and living environment are well known, few studies have analyzed citizens' perspectives regarding the effects of smart development on their quality of life. As citizens are both beneficiaries and developers in the smart city's process, their interactions are extremely important in determining its success. The researchers highlighted that the study explored a citizen-centric approach using empirical evidence, thereby filling a gap in the literature. Moreover, it is pointed out that the study goes beyond prior studies' focus on citizens' theoretical acceptance and considers citizens' support for the development of a smart city. Therefore, it extends the citizens' involvement in urban development by considering their perceptions not only in the planning stage but also in the assessment process, for example, citizens provide information, feedback, and suggestions. This work adopted a quantitative approach, and a questionnaire survey was used to collect data from 848 Macao citizens about the effects of smart city's quality of life domains on citizens' attitudes, perceptions, and support for citizen-centric smart city development. Factor analysis was used to identify these domains, while multiple linear regression and cluster analysis were used to achieve the research objectives. The researchers conclude that more attention should be given to smart services that enhance citizens' livelihoods and mobility. Furthermore, the lifestyle of citizens and the role that smart city development plays in it should be reconsidered. The city should promote the transition from the traditional way of support to a smarter concept of support by streamlining and enhancing public services through smart technologies. Considering the importance of quality of life in a smart

city identifying current levels of citizen support and citizen concerns can provide a guideline for smart city policymakers to adjust their strategies and promote support. Based on citizens' voices, measures targeting specific city goals can be implemented to better engage them in the development of their smart city.

The researchers in [19] emphasize that the exploration of smart city (SC) user preferences is still in its early stages, with limited empirical insights available. The study employed a standardized questionnaire to assess respondents' digital experience, considering internet usage hours and the frequency of smart service utilization. The majority of the 105 total participants, aged between 25 and 34, were part of a public administration postgraduate program at the Administrative Science Speye of a German university. The sample demonstrated extensive digital experience, with 81% spending more than 10 hours online weekly, and 88% being familiar with using smart services in their daily lives. Additionally, 44% claimed to be frequent users of smart services. The selected sample is deemed a snapshot of the 'creative class' of a city, positioning them as lead users for SC services. These individuals already grapple with the potential benefits and challenges of smart service usage, possessing the motivation, means, and opportunity to adapt to new forms of interaction. The study reveals that respondents prioritize addressing citywide mobility challenges before focusing on improving municipal administration or enhancing existing service structures. The findings suggest a growing public demand for municipal mobile government in the future, indicating that an SC strategy primarily centered on mobile applications could align with users' preferences for mobile public service solutions. However, recognizing the continued importance of other online channels, the study recommends the inclusion of appropriate SC online platforms to complement smart mobile services and meet users' diverse requirements. The survey highlights factors crucial to SC success, including the provision of a comprehensive set of full-featured smart urban services, service responsiveness, and transparency, and the maintenance of data security and privacy. While the study provides interesting insights into the presumed preferences of smart citizens, it acknowledges limitations. The participant selection, based on the assumption that current respondents represent the future SC citizenry and mirror their current preferences, might oversimplify the diversity of future user behaviors. The study focuses on German lead users, and preferences in other countries with significant differences in geographical, demographic, cultural, or socio-economic characteristics may differ.

The researchers extended their study to provide comprehensive insights into the preferences and relevance structure of digital smart city (SC) services from a citizen perspective [20]. A web survey was conducted among the residents of a mid-sized German city between January and February 2021, resulting in 906 valid questionnaires. The majority of respondents were between 35 and 64 years old, used the internet for more than four hours per week, and half of the sample had a college degree. The study found that administrative units related to "Smart Social Services" and "Smart Mobility & Smart Infrastructure" were highly relevant for citizens, outperforming "Smart Resources & Smart Environment" and "Smart Governance & E-Government," which were rated as medium relevant. The study involved 10 public officials from the same city, and they were asked to rate the same set of smart city (SC) services in terms of their relevance within the overall SC context. The research aimed to compare the service preferences of citizens on a broad scale with the estimations of responsible SC executives. The assessments revealed significant differences, particularly in the "Smart Mobility & Smart Infrastructure" administrative unit, which citizens considered important. The study found significant differences in assessments between citizens and public officials regarding the "Smart Mobility & Smart Infrastructure" administrative unit, which citizens considered important. The differences were even more pronounced for the "Smart Governance & E-Government" unit. Interestingly, civil servants tended to perceive their smart city (SC) services as more important than citizens rated them. These disparities in preferences and importance ratings, particularly within dedicated SC service domains, highlight the importance of extensive communication and adaptation efforts. The study underscores the need to better understand citizens' expectations and suggests adjusting the SC service portfolio accordingly to align more closely with citizen concerns.

The authors in [21] conducted a study to explore the relationship between motivation and the intention of public participation in the field of urban planning and management. Using a structural equation model based on a survey in Wuhan, China, the researchers aimed to identify causal relationships between four motivational factors and three intentions to participate in urban planning. The study involved the distribution of an online draft questionnaire among 185 respondents, and based on feedback and suggestions from 107 respondents, the finalized questionnaire was randomly distributed among 625 respondents across all districts in Wuhan, with 502 respondents completing it properly. The characteristics of the sample showed a cluster of 50.5% being female, most respondents being in the 30 to 59 age group and holding an undergraduate degree or lower educational level. The findings revealed significant effects of motivational factors on participation intention in urban planning, specifically in terms of scale, content, and process. Positive motivators, including concern for civil society, personal interest, and social influence, were identified. Conversely, constraints were identified as negative factors hindering citizens' willingness to participate. The study highlighted that compared to citizens' personal interest, social influence, and constraints, the variable of "Civil society" had the strongest influence on participating in urban planning contents, scales, and processes. The researchers inferred that citizens pay more attention to topics and civic culture when participating in urban planning. To improve the engagement environment and motivate citizens, the researchers suggested that decision-makers and policymakers should enhance efforts to encourage civil society participation in urban planning. Strengthening publicity and education on urban planning issues were recommended. Additionally, addressing potential negative effects and mitigating constraints by optimizing methods and tools of participation were emphasized as essential measures.

In the research of [22], the relationship between e-participation, as a form of co-production, and service performance is tested using multiple large longitudinal datasets from Jakarta's smart city mobile platform. Since 2014, the local government in Jakarta has utilized the Qlue MyCity mobile app to improve service performance by engaging citizens in identifying and providing feedback on city problems. The Qlue data, obtained from the app system over a two-month period (January to February 2017), includes various elements such as the category of the report (e.g., waste disposal, potholes), the location (GPS), report time, and the number of users participating in service feedback. The analysis covers 261 sub-districts of Jakarta, with a total of 50,918 unique requests during the specified period. Statistical techniques, including regressions and classical assumption tests, were applied, and analyses included time-lagged measures (e-participation a month lagged) and different service types to assess the impact on clearance rates in subsequent periods. The findings indicate that citizen e-participation via the smart city platform increases the percentage of resolved problems, with a relatively stronger influence on complex issues like damaged roads. These complex problems may involve multiple agencies and require more time for resolution compared to simpler problems like routine waste issues, where the relative influence of the citizen e-participation is lower. The study provides empirical evidence of how e-participation and co-production in smart city projects can improve government service delivery performance. The authors emphasize the need for policymakers to focus on smart city platforms that facilitate citizen e-participation in monitoring government service delivery. Enabling citizen co-production can enhance government transparency and openness, leading to more efficient and responsive use of public resources. However, it is noted that the study's results are based on a single case, limiting generalizability, and the need for further studies in different countries to establish more robust knowledge on the subject is highlighted.

The article [23] explores digital projects aimed at regenerating old urban spaces and implementing smart city systems. Emphasizing the importance of selecting technologies that ensure continuous citizen use, the study also seeks to enhance citizens' satisfaction with facilities, focusing on improving their quality of life within established cities. The research establishes a pre-evaluation system for technologies, facilities, and services using citizens' viewpoints and gathers expert opinions for digital project implementation. The study, conducted in January 2019, engaged 624 residents through a survey, exploring the development of citizen-centric smart cities in three cities in South Korea. The study employs the Hierarchical Regression Model to identify factors influencing citizens' overall preferences for smart city projects, considering demographic, social, and economic features,

and respondents' quality of life. The analysis reveals citizens prefer smart city projects as a means to enhance their quality of life, improve convenience, and upgrade urban infrastructures. Notably, citizens prioritize services closely linked to their daily lives, such as crime prevention and safe pedestrian environments, over novelty services. The findings suggest citizens view smart city projects as solutions to everyday issues and drivers of local economic development, rather than focusing on specific technologies or devices. The study presents implications and suggestions for linking urban regeneration plans with smart city projects, tailoring approaches to the unique characteristics of each city. This citizen-centric perspective highlights the importance of addressing practical problems and aligning smart city initiatives with local economic development goals.

The primary objective of the work of [24] is to assess the current status of smart mobility and explore the factors influencing the proliferation and penetration of modern sharing mobility through ICT applications and social media. The authors conducted a content analysis using a sample of over 1000 reviews from citizens and visitors published on the official digital platforms of the smart city of Milan. The paper also includes a comparison with data obtained from users of public urban mobility services. Through the nodes resulting from the content analysis, the authors applied the Social, Technological, Economic, Environmental, and Political (STEEP) method, a strategic analytical tool, combined with the Strengths, Weaknesses, Opportunities, and Threats (SWOT) dimensions analysis to provide a comprehensive practical contribution to the management of modern municipal smart mobility systems. The results underscore the importance for both smart mobility solution providers and institutions to effectively understand and communicate the implementation of digital services across various user touchpoints and communication channels to enhance the value provided to both citizens' and visitors' concerns.

The study in [25] emphasizes the importance of a structured stepwise framework for Smart City policy transformation, using Helmond, a Dutch city, as an illustrative case. It notes that beyond appropriate steps in a smart city digital framework, feedback loops are crucial for addressing emerging issues. The paper highlights the often underestimated readiness to adopt new smart city policies, emphasizing that the preparation phase is frequently neglected. Furthermore, the study suggests that the process of learning from good practices lacks a systematic approach and comprehensive benchmarking of smart city frameworks. Practical constraints in citizens' daily lives are identified as decisive factors for the success or failure of digital city reconstruction. The researchers stress that the adaptation of smart projects requires a tailor-made approach, with each step adjusted to the local context. The Helmond case, mentioned in the study, underscores the importance of social readiness and citizens' awareness, leading to a need for adjustment and adaptation stages. Concerns are raised about knowledge management and documenting the adjustment process, highlighting the significance of a knowledge management framework. Commitment to meaningful adjustment is identified as a critical success factor in the policy transplantation process. The study concludes by emphasizing the need for a deeper consideration of knowledge management, infrastructure, and commitment in the policy transplantation framework for Smart City initiatives.

2. Materials and Methods

Smart cities represent a paradigm shift in sustainable urban development, tackling the intricacies of dynamic urban environments. At the core of our methodology lies a fusion of advanced data analytics, predictive modeling, and digital twin techniques. Predictive analytics stands as the linchpin, empowering cities to proactively plan for evolving challenges. Simultaneously, digital twin methodologies provide a virtual mirror of the urban landscape, enabling real-time monitoring, simulation, and analysis. Our research emphasizes the criticality of real-time monitoring, simulation, and analysis for supporting test scenarios, revealing bottlenecks, and optimizing smart city efficiency.

This work uses a dataset of 144 text files that include 93,053 citizen reports retrieved through the API from the Sense City platform, a service launched by the Municipality of Patreón, available to the public since 2018. This platform enables citizens to report various issues based on geographic

graphical coordinates, offering a direct channel for community feedback on issues such as infrastructure, services, and various aspects of urban life..

Structured in JSON format, each report includes key details like a unique report identifier (`_id`), a bug identifier (`bug_id`), the current status of the reported issue (`status`), geographic coordinates (`loc`), the reported issue's type (`issue`), a bilingual description in Greek and English (`value_desc`), and the timestamp of the report submission (`reported`). These details provide a comprehensive understanding of reported incidents, encompassing nature, location, and reporting status. Figure 1 shows a sample of a single report.

```

1  "id": "6540dcf9d18942dac7c2b2e2",
2  "bug_id": 146753,
3  "status": "CONFIRMED",
4  "loc": {
5    "type": "Point",
6    "coordinates": [
7      21.7475708,
8      38.2666247
9    ]
10 },
11 "issue": "road-constructor",
12 "value_desc": "\u0391\u03b1\u03b2\u03b3\u03b4\u03b5\u03b6\u03b7 \u03b8\u03b9\u03b1 \u03b2\u03b3\u03b4\u03b5\u03b6\u03b7 \u03b1\u03b2\u03b3\u03b4\u03b5\u03b6\u03b7 \u03b1\u03b2\u03b3\u03b4\u03b5\u03b6\u03b7 \u03b1\u03b2\u03b3\u03b4\u03b5\u03b6\u03b7",
13 "reported": "2023-10-31T10:54:52.113Z"
14 },

```

Figure 1. The sample report of a unique citizen in JSON format.

The report consists of : `_id`: "6540dcf9d18942dac7c2b2e2"(Unique identifier for the report), `bug_id`: 146753 (Bug identifier associated with the report), `status`: "CONFIRMED" (Current status of the reported issue, indicating that it has been confirmed), `loc`: {"type": "Point", "coordinates": [21.7475708, 38.2666247]} (Location information for the reported issue, specifying that it is a point on the Earth's surface with latitude 38.2666247 and longitude 21.7475708), `issue`: "road-constructor"(Type or category of the reported issue, indicating a problem related to road construction), `value_desc`: "Κατάληψη Πεζοδρομίου" (It is encoded using Unicode escape sequences specifically in Greek. When decoded, it represents the text "Κατάληψη Πεζοδρομίου," which translates to "Occupation of the Sidewalk" in English. This is a common method of representing non-ASCII characters in a Unicode string using escape sequences.), `reported`: "2023-10-31T10:54:52.113Z" (Date and time when the issue was reported, which is October 31, 2023, at 10:54:52 UTC).

This report provides details about a confirmed issue related to road construction. The reported problem is the occupation of the sidewalk, and the report includes location coordinates, status, and a timestamp of when it was reported. The `bug_id` and `_id` serve as unique identifiers for tracking and referencing the report.

Geographically, the reports span various locations within Patras, reflecting the broad coverage of citizen concerns across the city. The spatial distribution offers insights into localized challenges and aids in understanding the dynamics of urban issues. The 'issue' field, categorizing reported problems, showcases the diversity of concerns raised by citizens, ranging from road construction to public space-related matters.

In terms of anonymity, the provided JSON seems to be structured in a way that focuses on the details of the reported issue and its location rather than revealing personal information about the citizen making the report. More specifically, the "bug_id" and "_id" fields serve as unique identifiers for the reported issue. These identifiers are important for tracking and managing reports and they don't reveal personal information about the individual making the report. The "loc" field provides coordinates (latitude and longitude) indicating the location of the reported issue. This information points to a specific geographic location and it doesn't directly reveal the identity of the person making the report. The "issue" field specifies the type of problem reported, in the sample, "road-constructor." This field, along with the "value_desc" (description) field, provides information about the nature of the issue but does not reveal personal details. The "reported" field indicates the date and time when the issue was reported. This information can be used for tracking and managing reports, it does not compromise the anonymity of the individual making the report.

In our exploration of the intricate urban fabric, we focus on the neighborhood level, recognizing the pivotal roles of citizen report analytics, prediction, and digital twin technologies. This study integrates Extract, Transform, Load/Extract, Load, Transform (ETL/ELT) processes, artificial

intelligence (AI) techniques, and a digital twin methodology [6–13,26–29]. This integration processes and interprets urban data streams emanating from citizen interactions with the city's coordinate-based problem mapping platform. The synergy of an interactive GeoDataFrame within the digital twin methodology creates dynamic entities, facilitating simulations across diverse scenarios.

Part of the initial stages of the methods in this work is time series analysis with Python serves as a critical methodology for uncovering temporal patterns and trends inherent in sequential data. This analytical approach enables researchers to derive valuable insights, make informed predictions, and understand the dynamics of urban issues. As the work delves into time series analysis for both urban and region-specific issues, the goal is to equip research with a powerful toolbox for interpreting underlying patterns in the data [29–31].

The analysis begins by exploring the daily reported issues that are loaded and processed from a data lake, as discussed in the next methodology steps. The initial time series plot provides a fundamental understanding of overall trends, and then the paper uses a 7-day moving average to smooth out fluctuations, revealing subtler patterns and helping to identify trends. These techniques not only demonstrate the ebb and flow of everyday reported issues but also set the stage for more sophisticated analyses.

Moving beyond the basics, this methodology incorporates seasonal decomposition for a detailed understanding of cyclical patterns within the data. This process breaks down the time series into components—trend, seasonality, and residuals—providing a comprehensive picture of the underlying dynamics. In addition, this paper extends the analysis to urban areas, recognizing that the dynamics of the mentioned issues may differ significantly between different areas. Using similar time series techniques, it gains insight into the temporal dynamics of both urban issues and neighborhoods, providing a holistic picture of the challenges facing the city.

In this neighborhood-centric approach, the article also incorporates prediction techniques such as Triple Exponential Smoothing Holt-Winters. This enables the projection of future values, facilitating proactive decision-making based on expected trends. Through this multifaceted methodology, researchers can not only visualize historical patterns but also anticipate future challenges and opportunities, ultimately contributing to more efficient urban planning and resource allocation. Exploration through time series analysis is part of the initial stage in uncovering the complexity of urban issues and regions. Figure 2 illustrates the time series analysis methodology.

```

1 # Python
2 # Method:Time Series Analysis: Daily Issues
3 df_daily = df.resample('D').sum()
4 # Outcome:Original time series plot of daily issues over the full time span.
5
6 # Method:Time Series Analysis: Daily Issues with 7-day Moving Average
7 df_daily = df.resample('D').sum()
8 df_daily['issue_int_7day_ma'] = df_daily['issue_int'].rolling(window=7, min_periods=1).mean()
9 # Outcome:Plot Of the original time series and a 7-day moving average of daily issues.
10
11 # Method:Seasonal Decomposition for 'issue' with 365-day Period
12 df_daily_issue = df.resample('D').sum()
13 result_issue = seasonal_decompose(df_daily_issue['issue_int'], model='additive', period=365)
14 # Outcome: Four-subplot figure showcasing the original time series, trend, seasonal component, and residuals for 'issue'.
15
16 # Method: Seasonal Decomposition for 'issue_int' every 6 months
17 data = data.set_index('reported_date_time')
18 result_issue = seasonal_decompose(data['issue_int'], period=6)
19 # Outcome: Four-subplot figure with custom colors indicating the observed, trend, seasonal, and residual components for 'issue_int'.
20
21 # Method :Holt-Winters Triple Exponential Smoothing Forecast
22 model = ExponentialSmoothing(df_daily['issue_int'], trend='add', seasonal='add', seasonal_periods=365)
23 fit_model = model.fit()
24 forecast = fit_model.forecast(steps=forecast_steps)
25 # Outcome Original time series plot of daily issues, along with a forecast using Holt-Winters Triple Exponential Smoothing.

```

Figure 2. Time series analysis methods.

Our approach transcends theoretical frameworks, inviting users to actively engage with urban data through an interactive Flask application. This meticulously designed methodology empowers users to navigate urban probabilities, filter narratives, and visualize insights, because of the creation of an interactive map (map.html). This dynamic interface provides a tangible platform for users to explore, analyze, and predict the urban system's response at the neighborhood level. The visualization reveals antecedent and predictive patterns, trends, and correlations, laying the groundwork for tangible enhancements in urban functionality, resilience, and resident quality of life.

As we unfold the specific steps of our methodology, each stage is connected, reflecting a holistic approach to unraveling the complexities of urban dynamics. These steps are not static; rather, they form a continuous loop that can be automated to create a system that perpetually checks and downloads new data from the API, performs necessary data manipulations, updates the machine learning model, and refreshes the digital twin framework. This automation ensures a seamless and real-time experience for users, providing consistently updated maps, charts, scenarios, and probabilities through the Flask app.

The benefits of such automation are manifold. Users experience continuous monitoring, gaining real-time insights into the city's dynamics, and fostering a deeper understanding of urban complexities. Proactive planning becomes a reality, empowering city planners and residents to address emerging challenges with up-to-date predictive capabilities. The user interface evolves dynamically, enhancing engagement and satisfaction. The system proves scalable, accommodating growing data volumes while maintaining responsiveness, and efficiency is optimized through regular automation, minimizing manual intervention.

With these automated processes in place, the system transforms into a powerful tool for urban management. Residents, planners, and decision-makers gain access to a holistic and real-time view of the city's dynamics, enhancing their ability to make informed decisions. The automated digital twin framework now becomes a keystone in the evolution of smart cities, paving the way for a more resilient, responsive, and livable urban environment. Here are the steps of this effort:

Step 1: Data Retrieval (ApiFetch.py)

Initiating the study, the researchers started a data retrieval process from the Sense City API, laying the groundwork for subsequent analyses. Serving as the starting point, this phase ensured the acquisition of high-quality data. Leveraging the Python 'requests' library, they seamlessly interfaced with the API, extracting meaningful statistics regarding confirmed urban issues within a specified timeframe and geographic location.

Step 2: Check Data (.py)

The researchers embarked on a comprehensive examination, integrating the Extract, Transform, Load (ETL)/Extract, Load, Transform (ELT) methodology. Employing the Pandas library, they systematically scrutinized the dataset's structure and intrinsic attributes. This ETL/ELT-driven exploratory analysis yielded valuable insights into data types, forming a basis for subsequent processing and enhancement within the sophisticated data lake infrastructure. ETL involves extracting data from the source system, transforming it into a format that can be used by the digital twin, and then loading it into the digital twin. ELT approach involves extracting data from the source system, loading it into the digital twin, and then transforming it as needed.

Step 3: Convert the 'Reported' Column to Datetime Format (.py)

Recognizing the importance of temporal precision in urban analytics, the researchers transformed the 'reported' column into a datetime format. This not only established a standardized temporal reference but also laid the foundation for sophisticated temporal analyses, enriching the dataset's temporal dimension.

Step 4: Correct a Row of Coordinates (.py)

Addressing the imperative of spatial accuracy in urban studies, the methodical approach to coordinate correction played a pivotal role in ensuring data integrity. This step validated each entry in the 'loc' column for adherence to the expected format, contributing to the spatial reliability of our dataset.

Step 5: Coordinates to Area and New Column (.py)

Delving into the geospatial context, this work executed a precise conversion of coordinates to human-readable area names, further enhancing the spatial granularity of the dataset. OpenLayers, the chosen open-source mapping library, facilitated reverse geocoding, attributing each citizen report to its corresponding urban area, a crucial step for robust spatial analyses within the data lake environment.

Step 6: Weird Characters to Greek (.py)

Addressing encoding intricacies is vital for uniform linguistic representation. The application of 'utf-8-sig' encoding has harmonized character encoding complexities, resulting in a linguistically coherent dataset. This linguistic clarity is integral for diverse analyses within the data lake.

Step 7: Columns Need Intervention for Predictions (.py)

A meticulous assessment of data types revealed nuanced characteristics, necessitating thoughtful consideration. Identification of non-numeric columns, requiring specialized intervention for predictive modeling, set the stage for subsequent machine learning endeavors within our data framework

Step 8: Count the Different Categories Issue_Area (.py)

An enumeration of urban issues and associated areas unfolded, providing a comprehensive understanding of the dataset's categorical composition.

Step 9: Check the Data (.py)

An examination of the dataset unfolded, encompassing essential checks for integrity, completeness, and overall structure. This step, executed within the data lake environment, ensured that subsequent analyses were founded upon a robust and reliable dataset.

Step 10: Converting Non-Numeric Columns (.py)

In preparation for machine learning endeavors, the researchers encoded non-numeric columns. This step involved the precise conversion of categorical variables into a format suitable for predictive modeling, fostering an optimal representation of features within our data lake infrastructure.

Step 11: Train RandomForest and Save the Model (.py)

In the pursuit of predictive modeling, we employed AI techniques, specifically machine learning, the RandomForest Classifier, as the vehicle for understanding complex patterns within the data. Trained with precision, this model serves as an analytical instrument, capable of discerning intricate relationships among various features. The selection of the classifier was a result of a thorough performance comparison among various models tailored to the specific dataset. Specifically, we approached the urban issues in the city as a multiclass classification problem, given the presence of eight categories of issues (garbage, lighting, road constructor, green, protection-policy, environment, plumbing, and parking). Recognizing the intricacies of this multiclass classification task, we initially assessed the data distribution to address the imbalance, scrutinizing the class distribution within the target variable. Figure 3 shows the distribution of the issues classes.

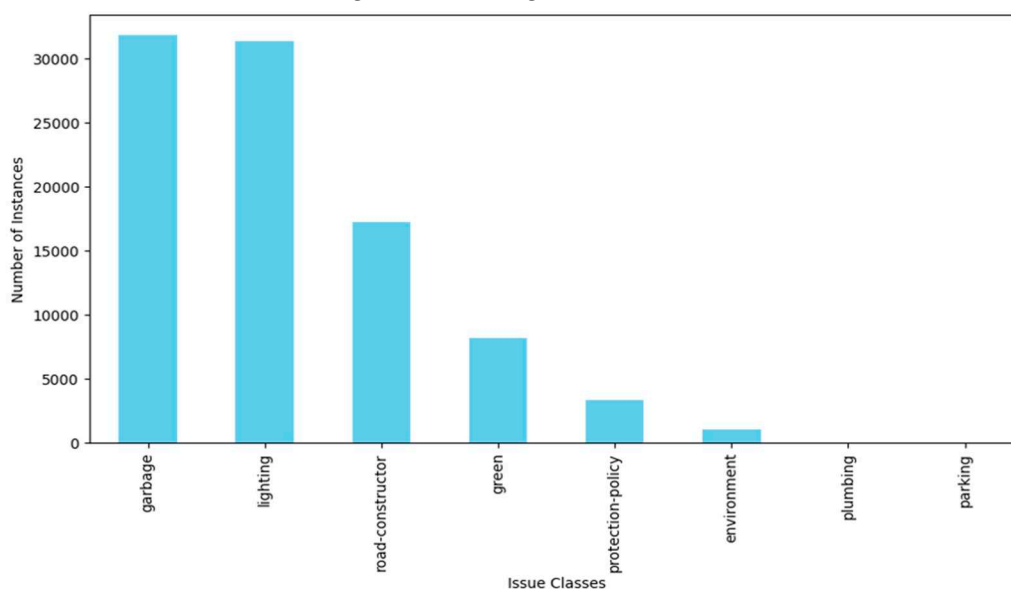


Figure 3. The distribution of the classes.

Confronted with data imbalance, the researchers proceeded to employ Resampling Techniques, combining oversampling and undersampling to effectively manage class imbalance, and trained several classifiers capable of handling unbalanced data, including Random Forest, Gradient Boosting,

KNN, SVM, and Neural Networks, on the designated training set [28]. The Random Forest classifier emerged as the top performer based on accuracy, achieving a very good score of 79.04%. This accuracy metric denotes the proportion of correctly predicted cases within the test set, showcasing the model's effectiveness in discerning and categorizing urban issues. Figure 4 illustrates the accuracy of the trained classifiers. Regarding ROC -AUC is a commonly used metric for binary classification problems.

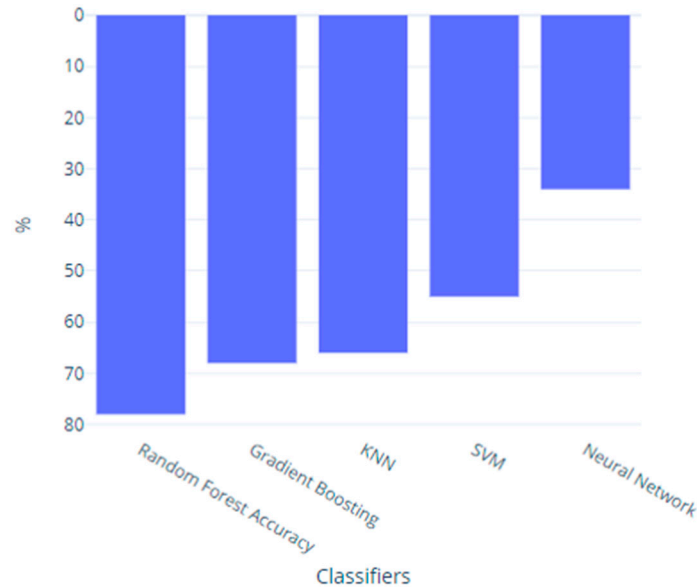


Figure 4. The performance of the trained classifiers.

Step 12: Pretrained Model and Predict 6 Months Later (.py)

With the trained model at the disposal of this work, the researchers projected the analyses into the future, specifically a six-month horizon. This step not only showcases the predictive capabilities of the model but also places the research in a long-term time frame, laying the groundwork for future urban insights.

Step 13: Patras Flask Laptop Filters Probability (.py)

The deployment of a Flask application marked an interactive phase, allowing end-users to navigate through intricate probabilities and filter urban narratives with ease. This immersive approach fosters user engagement, turning abstract data into tangible urban narratives through an intuitive and visually appealing interface.

The Flask application [11], in combination with Leaflet and Chart.js, leverages digital twin concepts and technologies to model and visualize urban data. Digital twins, in the context of smart cities, refer to virtual replicas of physical objects, processes, or systems. The framework facilitates the integration of various data sources and provides tools for visualization, prediction analysis, and interaction. Here's why the provided approach aligns with a digital city framework:

Data Integration and Processing: The Flask application loads urban data from the output of the data lake, representing a digital twin of the city. This data includes information about reported issues, areas, years, and issue probabilities. The data is processed to enhance its quality and to provide additional insights. For example, a 'year' column is added based on the 'reported_date_time' field.

Visualization: The framework uses Leaflet, a popular JavaScript library for interactive maps, to visualize spatial data. The map displays markers and clusters representing different issues and their locations in the city. Chart.js is utilized to create visualizations such as charts representing issue counts, area counts, and average issue probabilities. These visualizations enhance the understanding of urban data trends.

Interactivity and User Engagement: The application provides an interactive user interface with filters for issues, years, areas, and issue probabilities. Users can dynamically explore and analyze the digital twin data based on their preferences. Users can choose specific filters to update the displayed

data on the map and charts, allowing for a more personalized and insightful exploration of the city's digital twin.

Real-Time Updates and Monitoring: The framework can be extended to support real-time updates from various sensors and IoT devices in the city. This would enable monitoring and analysis of the city's state in near real-time.

Scalability and Extensibility: The architecture of the Flask application allows for scalability and extensibility. Additional features, data sources, or visualization components can be integrated to enhance the overall framework.

Below is a more general structure for Flask code that separates the backend (Flask) from the front end (HTML and JavaScript) and serves as a web page that includes a map, graphs, and filters. Further implementation of map preparation, graphs, manipulation filters, and data updates in JavaScript will depend on each specific requirement.

```
# Simplified Flask Code with Data Lake Processing and Output
from flask import Flask, render_template, jsonify
import pandas as pd
import json
app = Flask(__name__)
# Function to load processed data from an output CSV
def load_data_from_processed_csv(csv_path):
    try:
        # Load data from the CSV file
        data = pd.read_csv(csv_path)
        return data
    except Exception as e:
        print(f"Error loading data from CSV: {str(e)}")
        return pd.DataFrame()
# Extract unique values for filters
def get_unique_values(data):
    unique_years = data['year'].unique().tolist()
    unique_areas = data['Areas'].unique().tolist()
    unique_issues = data['issue'].unique().tolist()
    unique_issue_Probability = data['issue_Probability'].unique().tolist()
    return unique_years, unique_areas, unique_issues, unique_issue_Probability
# Route to serve the main page
@app.route('/')
def index():
    try:
        # Set the path to the processed CSV file from the data lake
        csv_path = r"G:\DataLake\processed_data_output.csv"
        # Load data from the processed CSV file
        digital_twin_data = load_data_from_processed_csv(csv_path)
        # Check if data is empty
        if digital_twin_data.empty:
            return render_template('index.html', issues='[]', unique_issue_Probability='[]')
        # Get unique values for filters
        unique_years, unique_areas, unique_issues, unique_issue_Probability = get_unique_values(digital_twin_data)
        # Convert DataFrame to a list of dictionaries and handle NaN values
```

```

        data_list = digital_twin_data.replace({pd.NaT:
None}).to_dict(orient='records')
        # Pass data to the HTML template
        return render_template('index.html',
issues=json.dumps(data_list), unique_issue_Probability=list(map(int,
unique_issue_Probability)))
    except Exception as e:
        print(f"Error processing data: {str(e)}")
        # Pass empty arrays if an error occurs
        return render_template('index.html', issues='[]',
unique_issue_Probability='[]')
# Run the Flask app
if __name__ == '__main__':
    app.run(debug=True)

```

Key points about the general structure of the Flask code:

Load Data from data lake output: The `load_data_from_processed_csv` function is responsible for loading processed data from the output csv file specified by the `csv_path` variable. This function handles exceptions and returns an empty DataFrame if there is an error loading the data.

Error Handling: The code includes try-except blocks for loading data from the csv file. If an error occurs during data loading, it prints an error message to the console, and an empty DataFrame is returned.

Data Processing: The code assumes that the processed data is stored in the specified csv file. The `csv_path` variable appends to the correct path of the csv file of the processed data.

Unique Values Extraction: The `get_unique_values` function extracts unique values for years, areas, issues, and issue probability according to the data of the current work. This information is likely used for setting up filters on the frontend.

HTML Template Rendering: The code renders the HTML template ('index.html') and passes the processed data to it. If the data is empty, it passes empty arrays to ensure that the frontend gracefully handles the absence of data.

Run the Flask App: The script runs the Flask app when executed directly. The `if __name__ == '__main__':` block ensures that the app starts only when the script is executed directly, not when it's imported as a module.

The following code is a general template of the `templates/index.html` structure based on the placeholders used in the previous Flask code and can be replaced with the appropriate HTML and JavaScript code according to the needs of each project.

```

<!DOCTYPE html>
<html lang="en">
<head>
    <!-- Head content including title, CSS, and scripts -->
    <title>The Title</title>
    <!-- Include CSS styles -->
    <!-- Include Leaflet and Chart.js libraries -->
    <script
src="https://unpkg.com/leaflet@1.7.1/dist/leaflet.js"></script>
    <script
src="https://unpkg.com/leaflet.markercluster@1.4.1/dist/leaflet.markerclu
ster.js"></script>
    <script src="https://cdn.jsdelivr.net/npm/chart.js"></script>
</head>
<body>
    <div id="map"></div>
    <div id="chartContainer1">
        <!-- Chart container for Issue Count -->

```

```

</div>
<div id="chartContainer2">
  <!-- Chart container for Area Count -->
</div>
<div id="filters">
  <!-- Filters container including Issue, Year, Area, and Issue
Probability filters -->
</div>
<div id="chartContainer3">
  <!-- Chart container for Average Issue Probability -->
</div>
<!-- Include the custom JavaScript code for initializing the map
and charts, handling filters, and updating data -->
<script>
  // JavaScript code for initializing the map and charts, handling
filters, and updating data
  var issues = {{ issues|safe }};
  var uniqueIssueProbability = {{ unique_issue_Probability|safe
}};

  // Use the 'issues' and 'uniqueIssueProbability'
variables to initialize your charts and map
  // Example: Initialize a chart using Chart.js
  var
      ctx
      =
document.getElementById('chartContainer1').getContext('2d');
  var myChart = new Chart(ctx, {
    type: 'bar',
    data: {
      labels: ['Label1', 'Label2', 'Label3'],
      datasets: [{
        label: 'Issues Count',
        data: [10, 20, 30],
        backgroundColor: 'rgba(75, 192, 192, 0.2)',
        borderColor: 'rgba(75, 192, 192, 1)',
        borderWidth: 1
      }]
    }
  });
  // Example: Initialize a map using Leaflet
  var map = L.map('map').setView([51.505, -0.09], 13);

L.tileLayer('https://{s}.tile.openstreetmap.org/{z}/{x}/{y}.png', {
  attribution: '© OpenStreetMap contributors'
}).addTo(map);
// Add markers to the map based on 'issues' data
// The additional JavaScript code for handling filters and
updating data
</script>
</body>
</html>

```

Key Points about the general structure of templates/index.html code:

Dynamic Data: The Flask variables `{{ issues|safe }}` and `{{ unique_issue_Probability|safe }}` are used to dynamically inject data into the HTML template. The structure of these variables must align with the expected data format in the project's JavaScript code.

Leaflet and Chart.js: The HTML file includes the necessary scripts for Leaflet and Chart.js. and using the provided placeholders, the initialization of the map and charts can be customized.

Custom JavaScript: The `<script>` tag includes a placeholder for your custom JavaScript code. Replace this with the actual code needed to initialize the map and charts, handle filters, and update data based on your application's requirements.

CSS Styling: You may want to include CSS styles for styling your HTML elements. Add a `<link>` tag in the `<head>` section to link your CSS file if needed.

The provision of the Flask and HTML code frameworks serves the purpose of offering readers an overview of the project's development structure without delving into the extensive details inherent in the Python programming language code. This framework allows readers to grasp the project's rationale and development flow efficiently. Regarding the specific elements used in Flask and the HTML code of the work, it is necessary to note the following aspects:

The highlighted aspects of this work's detailed Flask application code

Efficient Data Loading: The code efficiently loads data from the data lake output in csv format, demonstrating a seamless pandas integration for data manipulation.

Data Processing for Enhanced Analysis: The addition of the 'year' column based on the 'reported_date_time' field demonstrates thoughtful data processing, enabling users to analyze data over time.

Dynamic Filtering Options: The collection of unique years, areas, issues, and issue probabilities allows users to dynamically filter and explore the dataset. This feature enhances the user's ability to derive insights based on specific criteria.

Error Handling and Graceful Degradation: The code includes error-handling mechanisms that print informative messages. In case of an error, the application gracefully provides default values for filters, ensuring a smooth user experience even in unexpected situations.

Interactive HTML Rendering: The `render_template` function dynamically renders the 'map.html' template, passing data to the front end for display. This approach supports an interactive and responsive user interface.

Conversion to JSON for Frontend Integration: The conversion of the DataFrame to a list of dictionaries and subsequent conversion to JSON ensures seamless integration with the frontend. This is a key aspect of rendering dynamic content.

User-Friendly Display of Data: The application provides a user-friendly display, passing empty arrays if no data is available. This thoughtful consideration contributes to a positive user experience.

Debugging Information: The use of print statements for debugging purposes during data loading ensures that developers can quickly identify and address any issues.

Conditional Execution: The `if __name__ == '__main__':` block ensures that the Flask app is only run when the script is executed directly, promoting modular and reusable code.

Support for Future Enhancements: The modular structure and integration with a frontend template suggest a codebase that is extensible and open to future enhancements. This is crucial for the longevity and adaptability of the application.

The highlighted aspects of this work's detailed maps.html code

Integration of Maps and Charts: The seamless integration of Leaflet for interactive maps and Chart.js for dynamic charts creates an engaging user experience. This combination allows users to visually explore spatial and statistical aspects of the data.

User-Friendly Filters: The inclusion of user-friendly filters for issues, years, areas, and issue probabilities empowers users to tailor their data exploration. This level of interactivity is crucial for a meaningful and personalized user experience.

Clustered Marker Representation: The use of clustered markers on the map, each representing the most frequent issue in that cluster, adds a layer of sophistication. It condenses information while providing insights into prevalent issues in specific geographical areas.

Dynamic Data Loading: The code demonstrates flexibility by allowing data to be loaded from the data lake output. This adaptability showcases a robust system capable of handling different data sources, increasing its applicability in various scenarios.

Real-Time Data Updates: The real-time update functionality, triggered by filter changes, ensures that users receive instant feedback. This feature facilitates a dynamic and responsive data exploration experience.

Appealing Visual Design: The use of custom styles, including vibrant colors, clear typography, and strategic layout, contributes to an aesthetically pleasing design. A visually appealing interface can captivate users and make the exploration process more enjoyable.

Chart Variety: The inclusion of different chart types (line and bar charts) adds diversity to data representation. This variety not only caters to different learning styles but also provides a comprehensive view of the dataset.

Random Color Generation: The inclusion of a function to generate random colors for chart elements adds a playful and dynamic element to the visual representation. It enhances the overall visual appeal and contributes to a lively user interface.

Responsive Design: The implementation of responsive design principles ensures that the application remains accessible and functional across various devices and screen sizes. This adaptability reflects a commitment to user convenience.

Potential for Further Customization: The code structure and modular design suggest a foundation that can be easily extended or customized. This feature encourages further development, allowing users to adapt the application to their specific needs or integrate additional functionalities.

Before the next section, the paper illustrates the data distribution and statistical analysis so that the reader can derive additional information and seamlessly delve into the body of results. Figure 5 presents the percentage of issues by area. The visualization is divided into three horizontal bar graphs, each representing 54 different regions out of the 162 unique regions in total. Areas are sorted by number and bar graphs use different colors for a visually appealing representation.

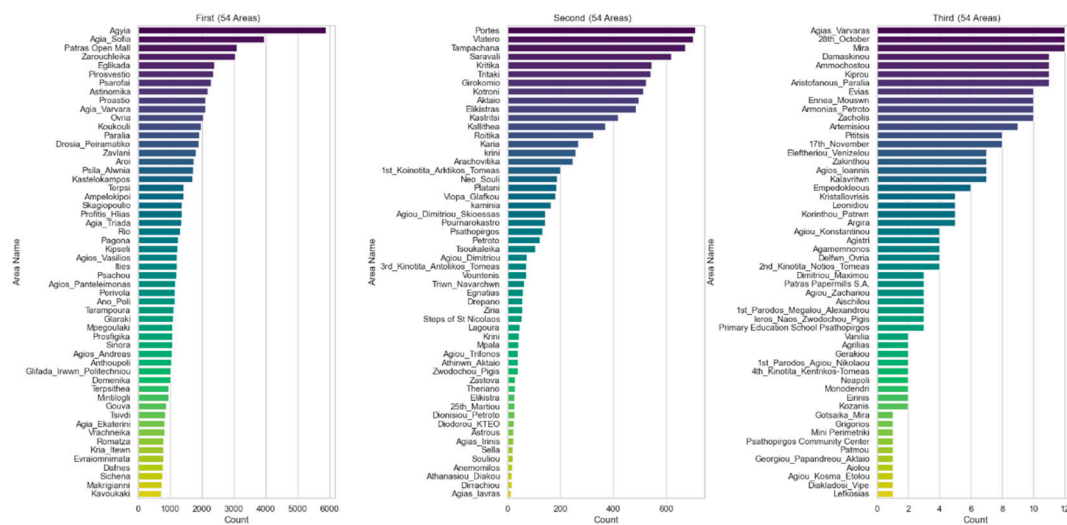


Figure 5. The count of areas.

Figure 6 is the output of the Python code that creates a stacked bar chart for three different parts (First Half, Second Half, and Third Half) based on the data lake output. The image attempts to present the information in a clear and visually appealing way, taking into account the fact that there are 162 unique areas in total, and the population of regions increases the difficulty of a valuable visualization. Each segment consists of a subplot with regions represented by stacked bars, where the height of each bar segment corresponds to the percentage of a particular issue within the region, with region names appearing on each line. The legend is included in the upper right corner of the first subplot, providing information about the issues represented by different colors in the graph. The image attempts to present the information in a clear and visually appealing way, taking into account the fact that there are 162 unique areas in total, and the population of regions increases the difficulty of a valuable visualization.

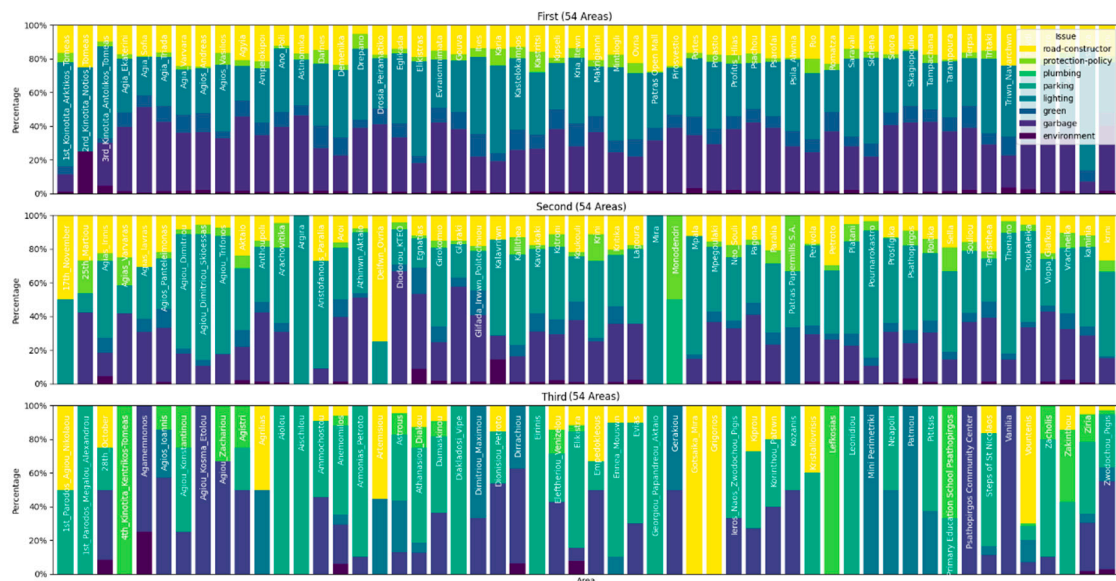


Figure 6. The percentage of issues by region.

Figure 7 is an attempt to show the overall distribution of subjects in each region over time. The color dimensions categorize regions based on the total number of releases from 2018 to the date of this article's writing, and a sample of the region color groupings is shown on the right side of the chart.

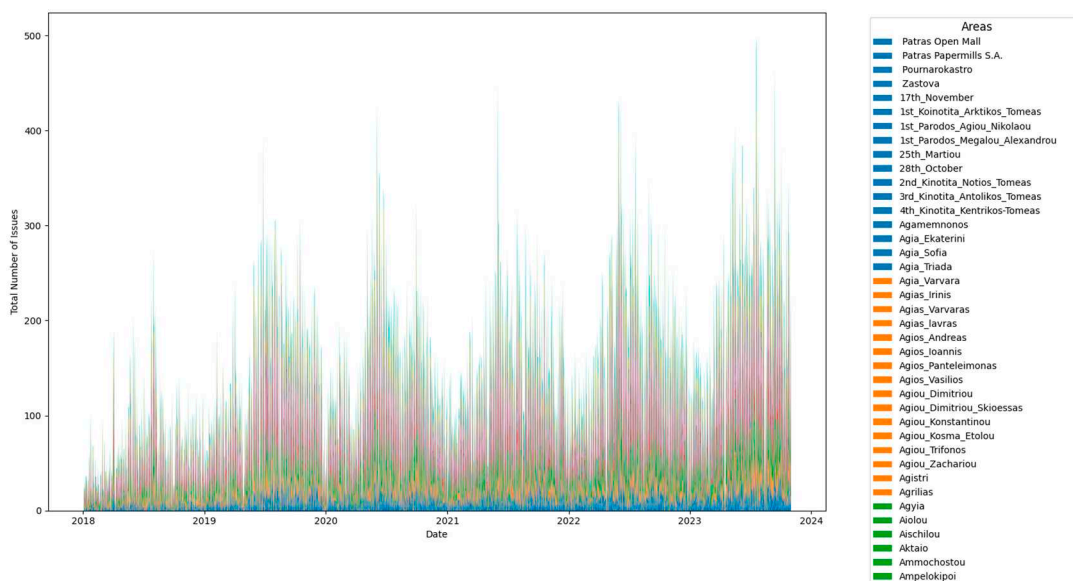


Figure 7. The total number of issues in each area over time.

Table 1 is the statistical summary of the lake data output that provides a comprehensive overview of the entities as a whole, including distribution, central tendency, and variability of the numerical column. The terminal output provides a statistical summary of the data generated and here is a detailed explanation of each segment

Count: Latitude, Longitude, Areas, Areas_int, issue, issue_int, reported_date_time, year, month, day, hour, minute, issue_Probability: These columns represent the number of non-null entries in the dataset. For instance, there are 93053 entries for each of these columns, indicating the total number of records in the dataset.

Unique: Latitude, Longitude, Areas, Areas_int, issue, issue_int, reported_date_time, year, month, day, hour, minute, issue_Probability: This shows the number of unique values in each

column. For example, there are 162 unique areas, 8 unique issue categories, and 89003 unique reported date and time entries.

Top: Areas, issue, reported_date_time: Indicates the most frequently occurring value in each column. For instance, "Agyia" is the most common area, "garbage" is the most common issue, and "2019-10-14 07:32:00" is the most common reported date and time.

Freq: Areas, issue, reported_date_time: Represents the frequency of the top value. For example, "Agyia" appears 5881 times in the "Areas" column.

Mean: Latitude, Longitude, Areas_int, issue_int, year, month, day, hour, minute, issue_Probability: Represents the mean (average) value for each numeric column.

Std: Latitude, Longitude, Areas_int, issue_int, year, month, day, hour, minute, issue_Probability: Represents the standard deviation, a measure of the amount of variation or dispersion in each numeric column.

Min: Latitude, Longitude, Areas_int, issue_int, year, month, day, hour, minute, issue_Probability: Represents the minimum value in each numeric column.

25%, 50%, 75%: Latitude, Longitude, Areas_int, issue_int, year, month, day, hour, minute, issue_Probability: These values represent the quartiles, indicating the distribution of the data. For example, 25% of the reported dates and times fall on or before the 57th minute.

Max: Latitude, Longitude, Areas_int, issue_int, year, month, day, hour, minute, issue_Probability: Represents the maximum value in each numeric column.

Table 1. The statistical summary of the generated data.

	stat_report[numeric_cols] = stat_report[numeric_cols].applymap(lambda x: f'{int(x) if pd.notna(x) else ""}')												
	Latitude	Longitude	Areas	Areas_int	issue	issue_int	reported_date_time	year	month	day	hour	minute	issue_Probability
count	9305 3	93053	93043	93053	9305 3	9305 3	9305 3	9305 3	9305 3	9305 3	9305 3	9305 3	9305 3
unique	162	8	89003										
top	Agyia	garbage	14/10/2019	7:32:00									
freq	5881	31836	4										
mean	38	21	NaN	90	NaN	2	nan	2020	6	15	10	29	69
std	0	0	NaN	53	NaN	1	nan	1	3	8	4	17	17
min	38	21	NaN	1	NaN	1	nan	2018	1	1	0	0	23
25%	38	21	NaN	41	NaN	1	nan	2019	4	8	7	15	57
50%	38	21	NaN	89	NaN	2	nan	2021	7	16	9	30	72
75%	38	21	NaN	144	NaN	4	nan	2022	9	23	12	45	84
max	38	21	NaN	171	NaN	8	nan	2023	12	31	23	59	100

From this summary, somebody can identify patterns such as the common areas and issues, and can also observe the distribution and variability of numeric columns. However, identifying outliers

might require additional visualization techniques, like box plots or scatter plots. The summary provides a good overview, but further exploration through data visualization and more advanced statistical techniques may be needed for a deeper understanding.

Figure 8 provides boxplots for the "Latitude", "Longitude", "Areas_int" and "issue_int" columns. According to the boxplots which show the distribution of the number of citizen reports received on each day are very useful in highlighting some observations: The boxes are relatively wide, which indicates that there is a lot of variation in the data. The medians of the boxes are all above the first quartile, which indicates that the data is positively skewed. There are a few outliers in the data, as indicated by the whiskers of the boxplots. The overall trend of the data is increasing, as indicated by the upward slope of the center lines of the boxes.

Based on these observations, somebody can conclude that the data is likely to represent a large number of citizen reports, with a significant amount of variation in the number of reports received each day. The data is also positively skewed, which means that there are more days with a high number of reports than days with a low number of reports. There are a few outliers in the data, which could be due to unusual events or data entry errors. Finally, the overall trend of the data is increasing, which suggests that the number of citizen reports is increasing over time.

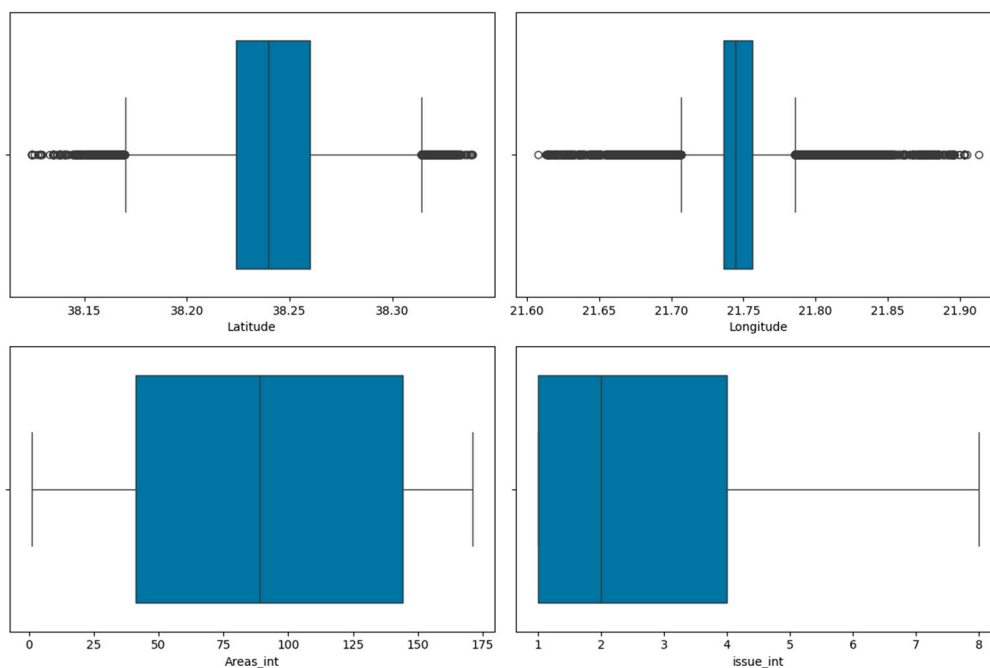


Figure 8. Boxplots for the "Latitude", "Longitude", "Areas_int" and "issue_int" columns.

Figure 9 shows a scatterplot for "Latitude" versus "Longitude" with "Areas_int" as hue and "issue_int" as size. The points are concentrated in the city of Patras, with a few outlying suburbs scattered around the city. The largest concentrations of reports are in the city center and in the areas to the north and south of the city. The size of the points in the scatter plot suggests that there is a significant variation in the number of reports received at different locations in the city. Some areas, such as the city center and the north and south areas, receive a high number of reports, while other areas receive a relatively low number of reports. This variation in the number of reports could be due to a number of factors, such as:

Population density: Areas with a higher population density are likely to receive more citizen reports.

Land use: Areas with a mix of land uses, such as residential, commercial, and industrial land uses, are likely to receive more citizen reports than areas with a single land use.

Socioeconomic status: Areas with a lower socioeconomic status are more likely to receive more citizen reports.

Crime rates: Areas with higher crime rates are more likely to receive more citizen reports.

Based on this information, it is highlighted that the citizen reports are distributed throughout the city of Patras, with a few outliers in the surrounding areas. The number of reports received on each day varies widely, with more reports being received on some days than others.

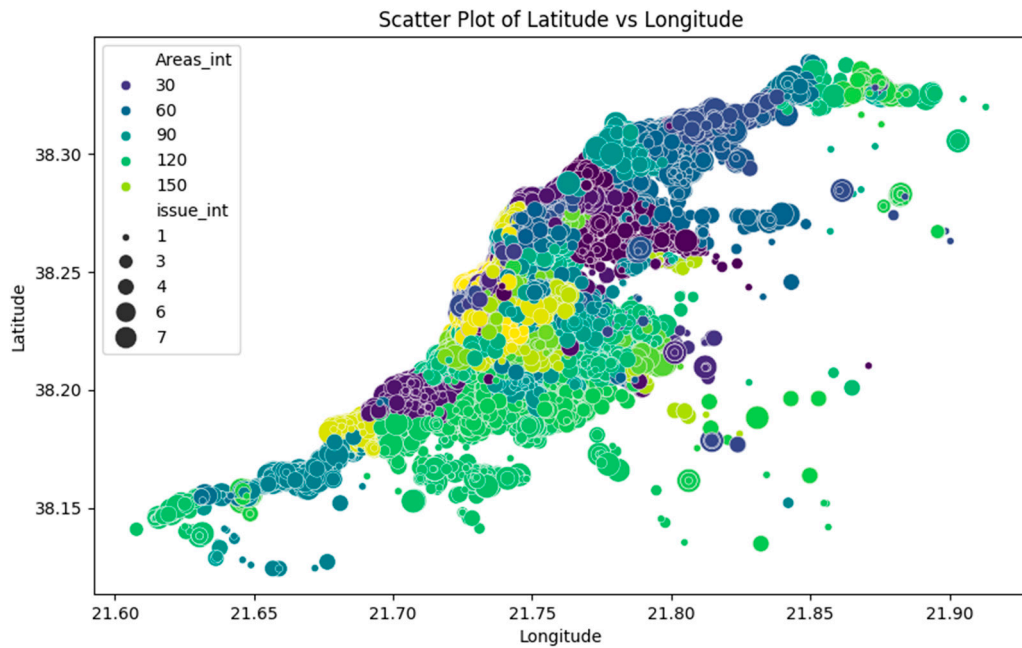


Figure 9. This scatterplot for "Latitude" versus "Longitude" with "Areas_int" and "issue_int".

As the last figure in this section, the paper also lists the analysis of each data histogram in the figure: According to the visualizations of the Figure 10 data histograms, an analysis follows for each of them.

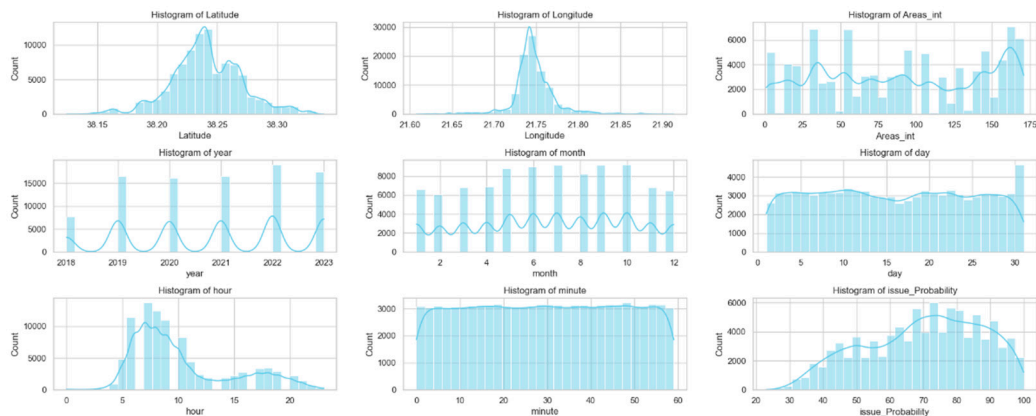


Figure 10. Histograms of data.

Histogram 1: Latitude. The histogram of latitude shows that the majority of citizen reports are located in the central part of Patras. There is also a smaller concentration of reports in the western and northern parts of the city. The histogram is relatively symmetrical, with a slightly longer tail on the left side.

Histogram 2: Longitude. The histogram of longitude shows that the majority of citizen reports are located in the eastern part of Patras. There is also a smaller concentration of reports in the central and southern parts of the city. The histogram is relatively symmetrical, with a slightly longer tail on the right side.

Histogram 3: Areas_int. The histogram of Areas_int, which is the numerical representation of the areas, shows that the majority of citizen reports come from areas with a medium to high population density. There are also a smaller number of reports from areas with a low population density. The histogram is positively skewed, with a longer tail on the right side. This suggests that there are more citizen reports from areas with a high population density than from areas with a low population density.

Histogram 4: Year. This histogram shows that the number of citizen reports has increased steadily over time. The histogram is positively skewed, with a longer tail on the right side. This suggests that there have been more citizen reports in recent years than in previous years.

Histogram 5: Month. This histogram shows that the number of citizen reports is highest in the summer months and lowest in the winter months. The histogram is slightly skewed to the right, with a longer tail on the right side. This suggests that there are more citizen reports in the summer months than in the winter months.

Histogram 6: Day. This histogram shows that the number of citizen reports is highest on weekdays and lowest on weekends. The histogram is slightly skewed to the right, with a longer tail on the right side. This suggests that there are more citizen reports on weekdays than on weekends.

Histogram 7: Hour. This histogram shows that the number of citizen reports is highest during the day and lowest at night. The histogram is slightly skewed to the right, with a longer tail on the right side. This suggests that there are more citizen reports during the day than at night.

Histogram 8: Minute. This histogram is relatively uniform, with a slight peak at the beginning of each hour. This suggests that citizen reports are distributed evenly throughout the hour.

Histogram 9: issue Probability. This histogram shows that the majority of citizen reports have a low to medium probability of being resolved. There are also a smaller number of reports with a high probability of being resolved. The histogram is slightly skewed to the left, with a longer tail on the left side. This suggests that there are more citizen reports with a low probability of being resolved than with a high probability of being resolved. Overall, the histograms of citizen reports show that the majority of reports are located in the central part of Patras, come from areas with a medium to high population density, and are highest in the summer months and on weekdays. The histograms also show that the number of citizen reports has increased steadily over time.

In essence, this section has examined the methodological roadmap employed in this study to create a digital twin application. It encompasses various components, including features, user interactions, data interactions, rules, and predictions related to neighborhood issues based on citizen reports of problems across different city areas. A comprehensive presentation of the data's nature and distribution has been provided, facilitating the reader's exploration as they further explore the subsequent section to understand the outcomes.

3. Results

The section presents the results of the previous methodologies to analyze data derived from citizen interaction with the coordinate-based problem mapping platform of the Municipality of Patras. The results are divided into two subsections. The first subsection is the presentation of the time series analysis that comes with the application, rendering the temporal analysis of the data attributed to the digital twin, and the second subsection is the presentation of the use of the digital twin.

3.1. Time Series Analysis

Time series analysis is a powerful statistical technique used to analyze sequential data points collected over time. This method provides meaningful insights, identifies patterns, and makes predictions based on the temporal order of observations. One key aspect of time series data is seasonality, a recurring pattern that occurs at regular intervals. Seasonality refers to the periodic fluctuations or patterns that repeat over a specific time frame, often influenced by external factors such as seasons, holidays, or business cycles. Understanding seasonality is crucial for various fields, including finance, economics, meteorology, and, in this case, urban data analysis.

In this analysis, the work delves into the time series data related to citizens' issues reported over time. The paper explores the concept of seasonality, investigating whether there are recurring patterns or trends within specific time intervals. The goal is to gain insights into the temporal dynamics of reported urban issues, helping us identify regularities, anomalies, and potential factors contributing to the observed patterns.

To accomplish this, we will employ the seasonal decomposition technique, a method that breaks down a time series into its components—trend, seasonality, and residuals. By dissecting the data in this manner, we aim to uncover underlying patterns in reported city issues and city areas that may be obscured in the raw time series.

Furthermore, we extend our analysis to include forecasting, providing a comprehensive view of how urban issues are anticipated to evolve over time. By examining the temporal nature of the data, this exploration of time series analysis with a focus on seasonality equips us to make informed decisions and predictions regarding future urban challenges. This analysis will provide a holistic view of how urban issues vary over time, enabling city stakeholders to make informed decisions and predictions based on the temporal nature of the data.

Let's begin this exploration of time series analysis with a focus on seasonality and forecasting to unravel the dynamics of urban issues over distinct time intervals.

Figure 11 generates a time series plot of the sum of areas by day, for the entire time span of the data. The plot shows that there has been a general increase in the sum of areas over time, with some seasonal fluctuations. The seasonal fluctuations are likely due to the fact that some types of issues are more common during certain times of year (e.g., some issues are more common during the summer months). The plot also shows that there have been some spikes in the sum of areas at certain points in time. These spikes could be due to a number of factors, such as: Major weather events (e.g., storms, floods, etc.), Construction projects, Changes in government policy or regulations, Increased awareness of the platform and its features.

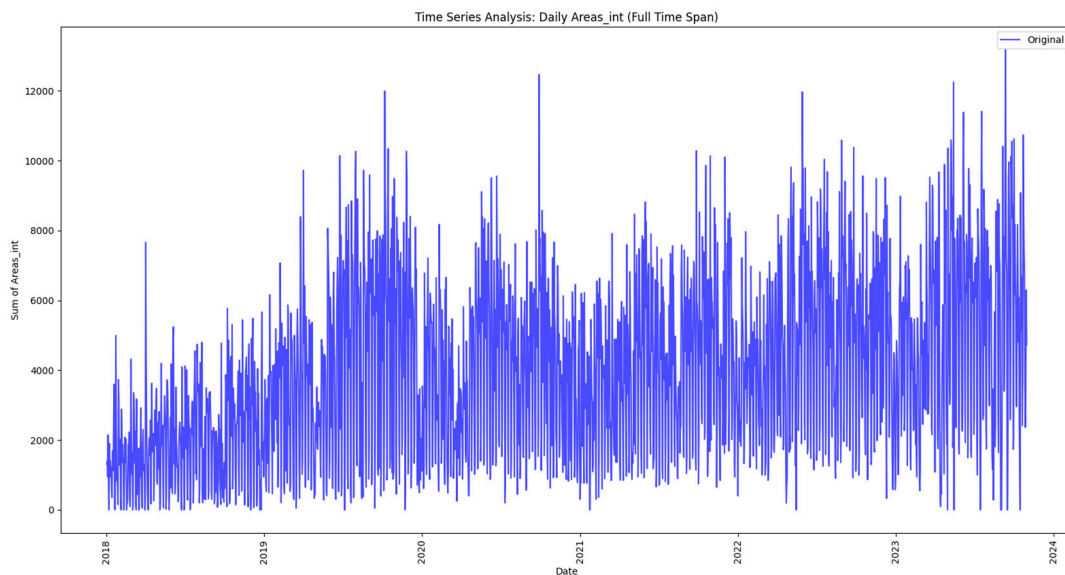


Figure 11. Time analysis of daily areas.

The simple time analysis of daily areas shows a general upward trend in the sum of areas over time. This indicates that the total number of issues being reported to the SenseCity platform is increasing. However, the simple time analysis does not provide any insights into the underlying causes of this trend or the seasonal fluctuations in the data. The time series plot of daily areas with a 7-day moving average reveals a more nuanced picture of the data (Figure 12). The general upward trend is confirmed, but the 7-day moving average helps to smooth out the seasonal fluctuations. This makes it easier to identify spikes in the data, which can be helpful for identifying potential contributing factors.

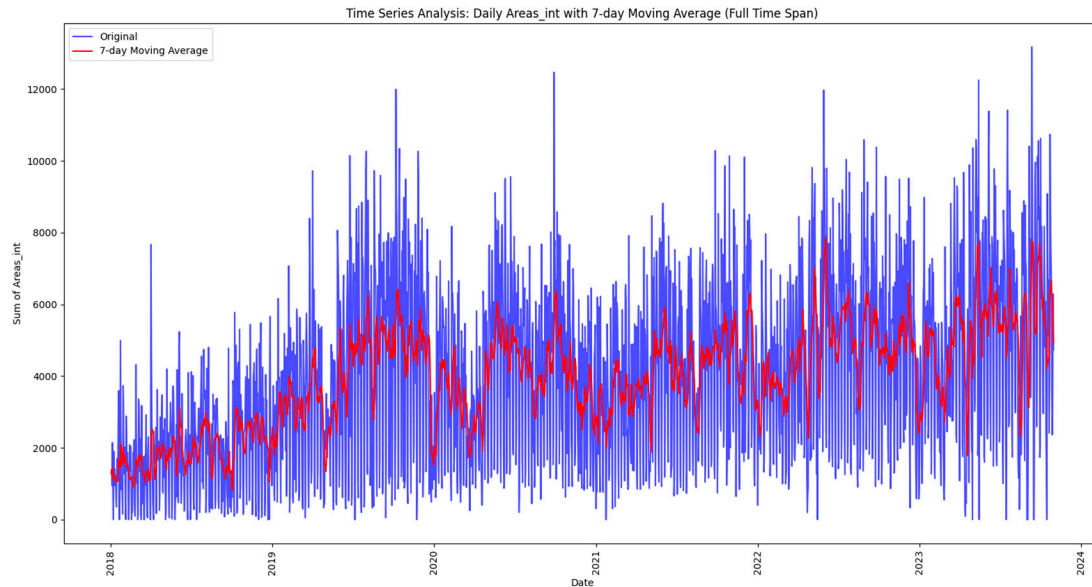


Figure 12. Time analysis of daily areas with a 7-day moving average.

The key difference between the simple time analysis and the time series plot with a 7-day moving average lies in their ability to reveal the underlying trend and seasonal fluctuations in the data. The simple time analysis only shows the raw values of the data, which can make it difficult to see the overall trend or the impact of seasonal factors. The time series plot with a 7-day moving average, on the other hand, effectively smooths out the seasonal fluctuations and highlights the spikes in the data, making it easier to identify potential contributing factors.

The time series plot of daily areas with a 7-day moving average shows a clearer upward trend in the sum of areas over time compared to the time series plot of daily areas. This is because the 7-day moving average smooths out the seasonal fluctuations in the data, allowing us to focus on the overall trend. In addition, the time series plot of daily areas with a 7-day moving average highlights the spikes in the data. These spikes can be helpful for identifying potential contributing factors, such as weather events, construction projects, or public awareness campaigns. Here is a more detailed analysis of the spikes in the time series plots:

January 2021: This spike could be attributed to a combination of factors, such as Increased awareness of the SenseCity platform: The SenseCity platform was launched in early 2021, and this period could have coincided with an initial surge in user registrations and issue reporting. Weather events: The winter of 2021 was particularly harsh in Greece, with heavy snowfall and icy conditions. This could have led to an increase in road and infrastructure issues that were reported through the SenseCity platform.

September 2021: This spike could be associated with Construction projects: Patras is a city undergoing significant infrastructure development, and construction projects could have contributed to the increase in reported issues. Government policy changes: The Greek government implemented new environmental regulations in 2021, which may have led to increased reporting of environmental issues. These regulations are expected to have a significant impact on the environment in Greece in the coming years. For example: a) Regulation (EU) 2020/852 on the establishment of a framework for the sustainable use of natural resources (also known as the Circular Economy Package) aims to reduce the environmental impact of the manufacturing and consumption of goods. The regulation sets target for recycling and reuse of materials and establishes a new system for managing waste electrical and electronic equipment (WEEE). b) Law 4818/2021 on the Prevention and Control of Air Pollution aims to improve air quality in Greece. The law sets stricter limits on air pollutants and establishes a new system for monitoring and reporting air pollution. c) Law 4936/2022 on Environmental Impact Assessment strengthens the environmental impact assessment process for

major infrastructure projects. The law requires more detailed assessments and public participation in the process.

July 2022: This spike could be attributed to Tourist season: Patras is a popular tourist destination, and the summer months see a significant influx of visitors. This could have resulted in an increase in issues related to tourism, such as littering and public safety concerns. Extreme weather events: The summer of 2022 was characterized by extreme weather events, including heat waves and wildfires. These events could have led to an increase in issues related to infrastructure damage and environmental pollution.

November 2022: This spike could be associated with Public awareness campaigns: The city of Patras may have implemented public awareness campaigns to encourage residents to report issues through the city platform. This could have led to a temporary surge in reported issues. Improved access to the platform: The city platform may have been made more accessible to residents, such as through increased public outreach or mobile app development. This could have led to a larger number of people reporting issues.

The time series plot of Figure 13 shows the number of issues reported to the platform in Patras, Greece, from 2018 to 2023. The absence of the 7-day moving average in the analysis of daily city issues returns the general trend of the number of issues reported to the city platform time and is a useful way to see the overall trend of issue reporting, but it can be difficult to see the smaller fluctuations in the data. The trend is generally upward, with some seasonal fluctuations and spikes. This suggests that the number of issues being reported to the platform is increasing over time. The presence of the 7-day moving average analysis, Figure 14, is a useful way to confirm the overall trend of the issue report and the fluctuations in the data. This means that it calculates the average of the issue for each day, taking the previous 7 days into account. In summary, it plots the number of issues reported each day and applies a smoothing technique to visualize trends. The blue line represents the original time series, and the red line represents the 7-day moving average. The plot shows that there has been a general increase in the number of issues reported over time. There are also some seasonal fluctuations, with more issues being reported during the summer months. The 7-day moving average helps to smooth out these fluctuations and reveal the underlying trend. The following are some key observations from the plot: the number of issues reported has been increasing steadily over time, there are some seasonal fluctuations, with more issues being reported during the summer months, the 7-day moving average shows a more gradual increase in the number of issues reported, there were some peaks in the number of issues reported in early 2021 and mid-2022.

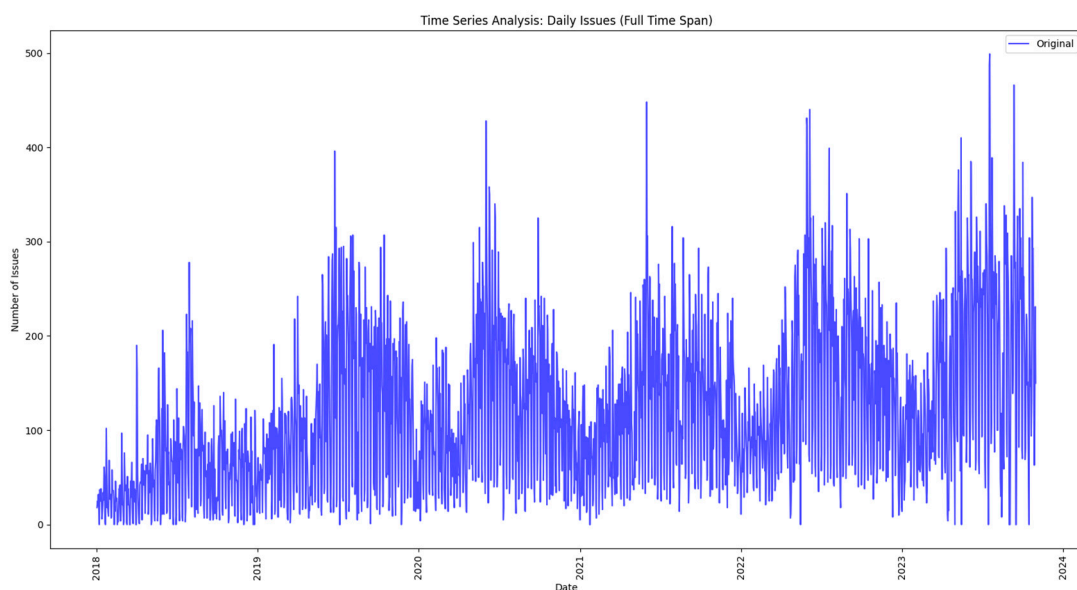


Figure 13. Time analysis of daily city issues.

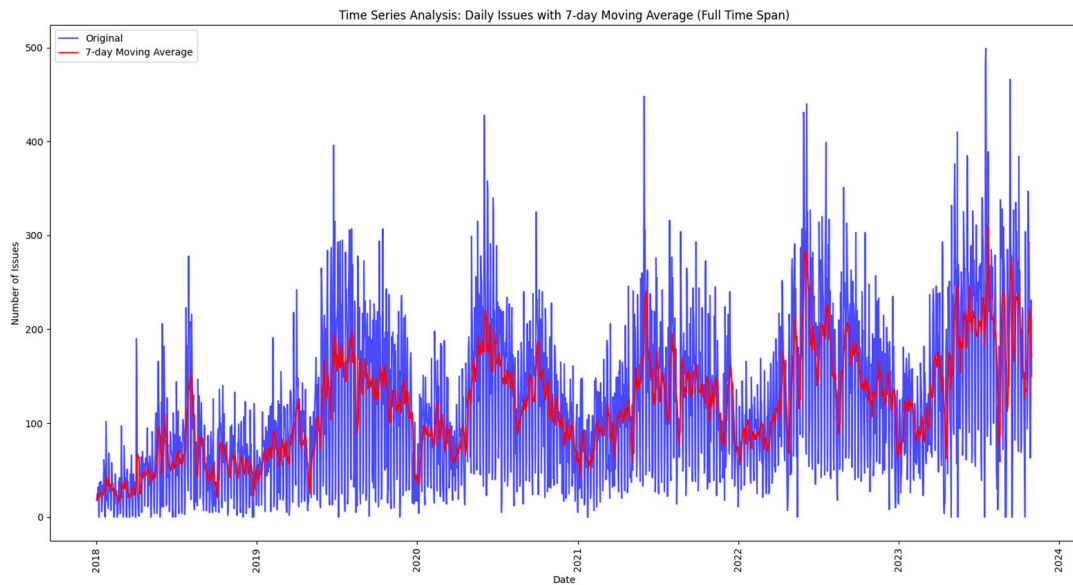


Figure 14. Time series analysis: daily issues with a 7-day moving average.

The seasonal fluctuations are likely due to a variety of factors, such as the weather, tourism (especially domestic tourism), and the availability of outdoor activities. For example, there are more issues reported during the summer months, when there are more tourists in Patras and people are more likely to be spending time outdoors. The spikes in the data could be due to a number of factors, such as construction projects, public awareness campaigns, and changes in government policy or regulation. For example, there was a spike in the data in early 2021, which could be due to the increased awareness of the platform following its updates in late 2020. Overall, the time series plot suggests that the number of issues being reported to the city platform in Patras is increasing. This is a positive sign, as it indicates that the platform is being used more regularly by residents and that the city is becoming more aware of the issues that need to be addressed. The following is a detailed analysis of the spikes in the time series analysis of the number of issues.

Spike 1: January 2021. This spike could be due to a number of factors, including increased awareness of the SenseCity platform following its update in late 2020, increased use of the platform during the winter months when people are more likely to be staying indoors and using technology, and the winter weather conditions in Greece, which could have led to an increase in the number of reported issues related to infrastructure, such as potholes and blocked drains.

Spike 2: September 2021: This rise may be due to the return of tourists to Patras (as Patras is the port gateway to Italy) as well as citizens returning to normality after the COVID-19 pandemic, increased construction projects in the city, and the implementation of new government policies or regulations related to urban development (as mentioned in the time series analyses for areas).

Spike 3: July 2022. This increase is probably due to the summer season when people move more often in the city and discover things, but also to the extreme weather phenomena that occurred in Greece in the summer. Greece experienced a number of extreme weather events during the summer of 2022, including floods, storms, and fires. These events can damage infrastructure, create dangerous conditions, and force people to evacuate their homes. This can lead to an increase in the number of reported problems such as damaged roads, fallen trees, and power outages leading to lighting problems. In addition to the direct impact of weather events, the consequences of these events can also contribute to the increase in reports. For example, residents may be more likely to report problems if they are concerned about the safety or well-being of themselves or others. They may also be more likely to report issues if they are frustrated by a lack of response from the authorities. Also, as a tourist season, tourists often contribute to an increase in reported issues such as lighting, littering, and vandalism, especially when tourism is domestic. Extreme weather events such as floods, storms, and fires can also lead to an increase in the number of reported problems.

Spike 4: November 2022. This spike could be due to public awareness campaigns and/or improved access to the city platform. The city of Patras may have implemented public awareness campaigns to encourage residents to report issues or the platform may have been made more accessible to residents, such as through new features or improved user interfaces.

Time series analysis of the number of issues reported on the city platform shows a general upward trend, with some seasonal fluctuations and peaks. The time series analysis of areas, on the other hand, shows a stronger upward trend, with fewer seasonal fluctuations and peaks. This suggests that the number of issues reported on the city platform is growing at a faster rate than the number of areas with high concentrations of reported issues. The time series analysis of areas shows that the number of areas with high concentrations of reported issues is increasing at a slower rate than the number of issues being reported to the city platform. This suggests that the issues in the city are becoming more widespread, with more areas experiencing issues. This could be due to a number of factors, such as Increased urbanization, as the city becomes more densely populated, it is more likely for issues to arise. Changes in demographics, the city's demographics are changing, with an increasing proportion of elderly residents and young families. These groups may be more likely to experience certain types of issues, such as accessibility. Changes in the built environment, the city's built environment is changing, with new development projects and infrastructure upgrades. These changes may also lead to new issues.

The following analysis improves understanding of the underlying patterns of city areas and issues as reported by citizens, including any trends, seasonality, and residual patterns. The 'seasonal_decompose' function from 'statsmodels.tsa.seasonal' decomposes the time series into these components, making it easier to analyze and interpret the individual contributions to the overall behavior. Figure 15 presents a seasonal decomposition time series analysis for the city areas and highlights that there is a clear seasonal pattern in the area data. The number of areas with high concentrations of reported issues is typically higher during the summer months and lower during the winter months. This is likely due to a number of factors, such as increased external and internal tourism, increased construction activity, and changes in weather patterns. The trend component of the time series analysis shows that there is a general upward trend in the number of areas with high concentrations of reported issues. This suggests that the problem of areas with high concentrations of reported issues is becoming more widespread. The seasonal component of the time series analysis shows that the seasonal pattern in the data is relatively stable over time. The residual component of the time series analysis shows that there is some variability in the data that is not explained by the trend or seasonal components. This variability could be due to a number of factors, such as random events.

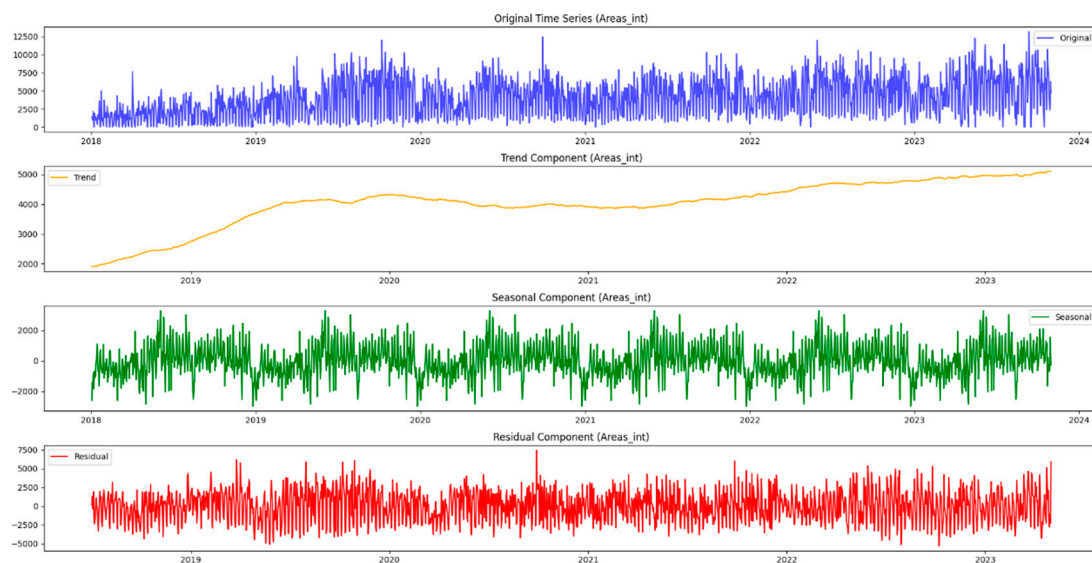


Figure 15. A seasonally decomposed time series analysis for areas.

Figure 16 illustrates the time series analysis of the issues and highlights that there is a clear seasonal pattern in the data and, in general, is in alignment with the seasonally decomposed time series analysis of areas. The number of issues reported to the platform is typically higher during the summer months and lower during the winter months. The trend component shows that there is a general upward trend in the number of issues. This suggests that the problem of issues in Patras is becoming more widespread over time. The seasonal component of the time series analysis shows that the seasonal pattern in the data is relatively stable over time. The residual component of the time series analysis shows that there is some variability in the data that is not explained by trend or seasonal components and may be due to random events. Figure 17 and Figure 18 present the analysis of the time interval of 3 and 6-months for the city issues.

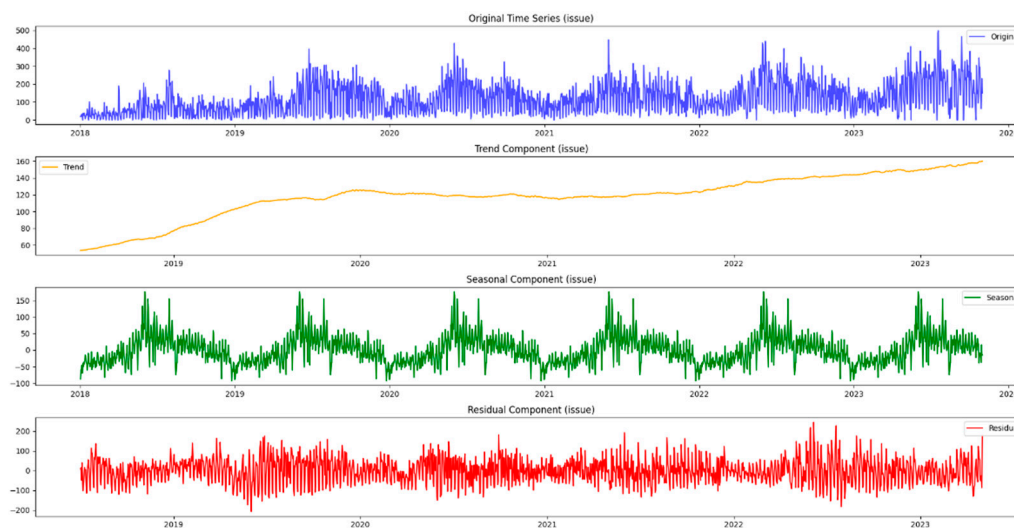


Figure 16. A seasonally decomposed time series analysis for issues.

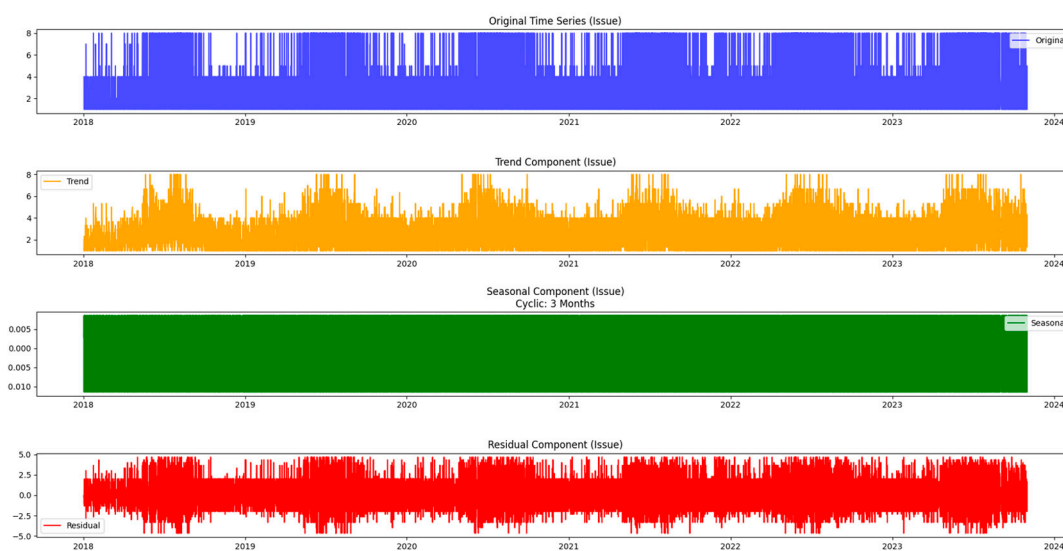


Figure 17. The seasonal component indicates a 3-month cyclical behavior for issues.

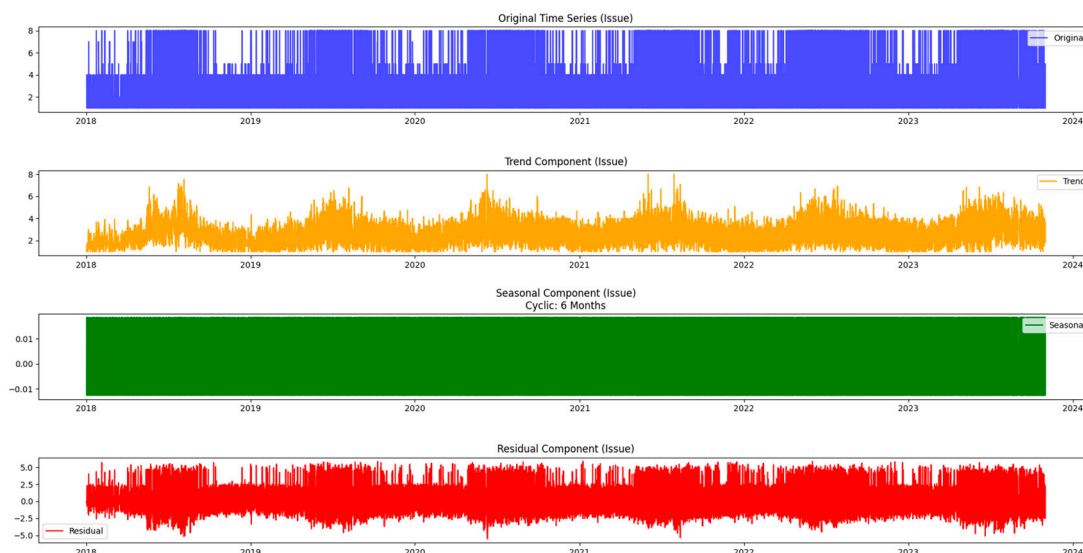


Figure 18. The seasonal component indicates a 6-month cyclical behavior for issues.

The main difference between the two images is that the cyclical component of the data in the 6-month image is a little more pronounced than the one in the 3-month image. This could suggest that cyclical fluctuations in the data occur over a longer period of time than previously thought. Another difference between the two images is that the trend component of the data in the 6-month image is more upward-sloping than the trend component of the data in the 3-month image. This likely suggests that the number of issues reported on the city's platform is increasing at a faster rate than previously thought.

Figure 19 and Figure 20 present the analysis of the time interval of 3 and 6-months for the city issues. Both graphs show that the number of areas reported is steadily increasing and provides strong seasonal but more stable trends in area-specific reports compared to issues. The residual component of the time series analysis in images shows that the variability in the data is due to random events such as various annotations of instructions or other markings made by citizens.

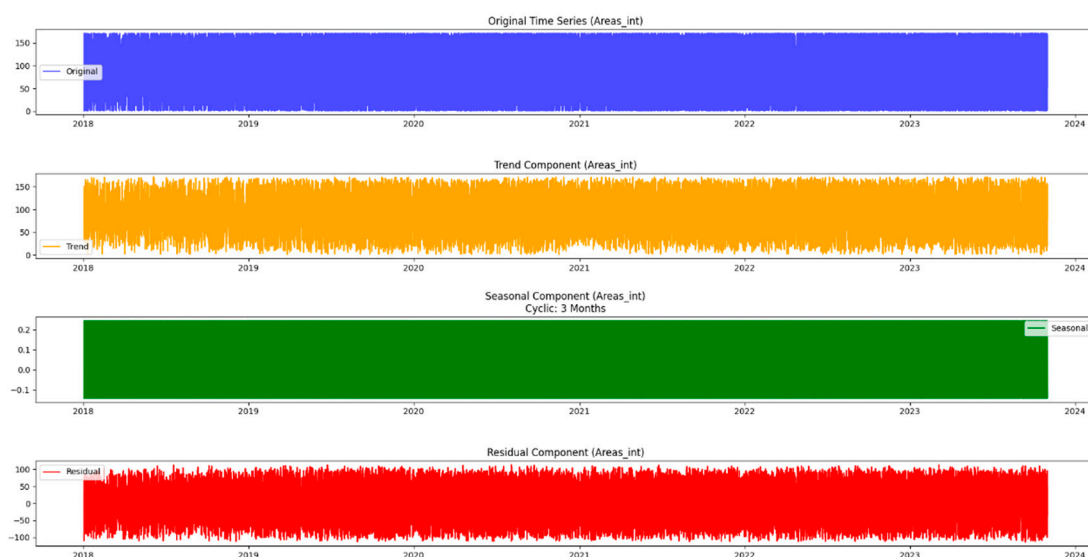


Figure 19. The seasonal component indicates a 3-month cyclical behavior for Areas.

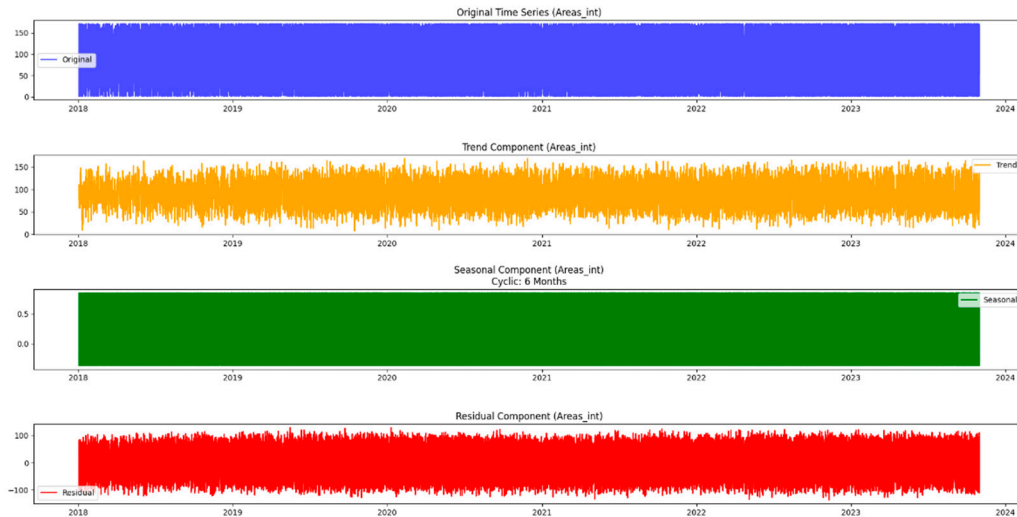


Figure 20. The seasonal component indicates a 6-month cyclical behavior for Areas.

Because of the trends depicted in the data set, researchers extend the time series analysis by applying the Holt-Winters method, which is particularly useful when dealing with trending time series data. More specifically, this work uses Holt-Winters Triple Exponential Smoothing, a time series forecasting technique that extends the Holt-Winters method to handle trend and seasonality as well. The result provides a visual representation of the original time series and the forecasted values, allowing somebody to capture future trends and seasonality in the data. Figure 21 and Figure 22 highlight the projected values of city and town issues over the coming years.

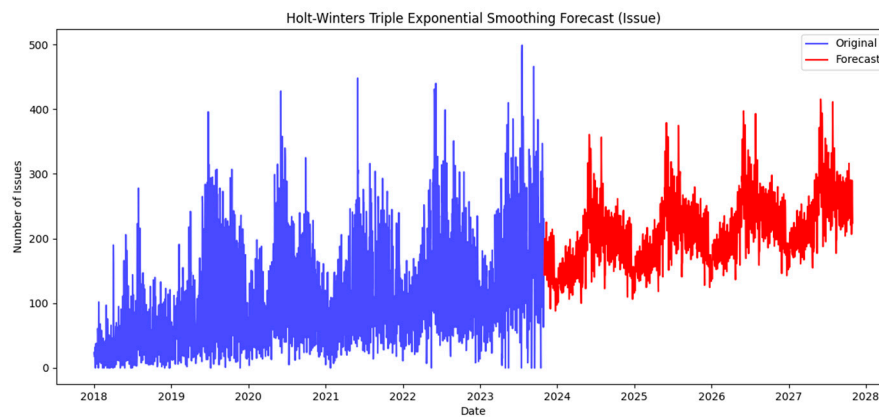


Figure 21. The Holt-Winters Triple Exponential Smoothing Forecast for city issues.

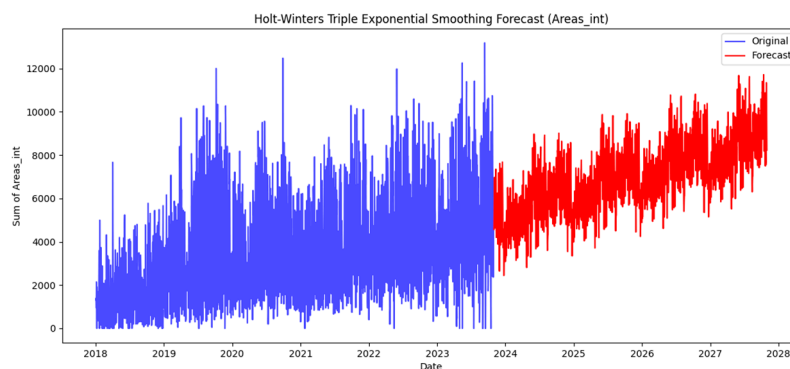


Figure 22. The Holt-Winters Triple Exponential Smoothing Forecast for city areas.

An initial observation reveals a consistent upward trend in both the overall issues reported in the city and the issues reported in specific areas over the next few years. For the city-wide problems, the seasonal variations exhibit a constant pattern, marked by pronounced peaks during the summer months. This suggests a recurring pattern of increased reporting during these warmer seasons. In contrast, when examining issues within specific regions, the seasonality is intricately woven into the overall trend. The trend reflects the dynamic nature of areas in reporting problems, highlighting fluctuations that align with the unique characteristics of each neighborhood. This nuanced approach captures the interplay between the overarching trend and the seasonal dynamics of reported issues in different areas.

3.2. Digital Twin Application

Within the realm of urban innovation, our digital twin application, powered by Flask, not only presents a dynamic snapshot of the city's current state but also delves into the future through predictive modeling. At the heart of this endeavor is a web of AI techniques, predominantly machine learning, employing the RandomForest Classifier as a vehicle for unraveling the complexities embedded in our urban data.

Trained with accuracy, this machine learning model stands as an analytical technique, adept at deciphering intricate relationships among diverse features within the city's landscape. What sets the digital twin apart is its ability to project these analyses into the future, providing a futuristic gaze into the urban landscape—a gaze extending six months ahead. These services enrich the predictive capability of the city stakeholders and position them within a long-term time frame, serving as a roadmap for future urban insights.

The synergy between the Flask application and the machine learning model offers a new possibility to the leadership of the city of Patras, which before this project did not exist. In addition to visualizing the current state of the city, this application now extends its capabilities to display predictions regarding the likelihood of each reported issue occurring in specific neighborhoods. This forward-looking feature enhances the utility of the digital twin, transforming it into a valuable tool for not just understanding the present but also anticipating and preparing for the challenges that lie ahead.

As the research progresses in the following paragraphs of the application presentation, it becomes clear that the digital twin, in addition to being a dynamic observer of the urban core, is also a proactive assistant in the resilience of the city. The finalization of the application unfolded through a gradual and evolutionary process, a narrative worth sharing as it underscores the evident requirement for a holistic digital twin of the city's physical layer. Preceding the development of the updated application version, incremental efforts were undertaken to address the needs stemming from the reports. These requirements center on gaining complete control over historical data and enhancing predictive capabilities for requests and areas within the city.

To provide readers with a quick insight into the spatial distribution of citizen requests related to city issues based on specific coordinates, Figure 23 delineates the city's dimensions on the map, while Figure 24 elucidates the distribution of issues. It is observed that a multitude of issues have arisen from these reports, spanning a wide range of city areas where authorities are called upon to promptly manage, prioritize, and intervene to resolve the problems. For the swift resolution of these issues, a key element is the clustering of the reported issues, as illustrated by the clusters in Figure 25. The user is informed about the number of problems per area through the respective clusters. Additionally, the interactive map is enhanced by the development of a pop-up window at the final points, displaying information about the issue, the area, and the reporting date. Simultaneously, the map's functionalities are enriched through the implementation of a filter linked to the frequency of citizen-reported issues and the utilization of color-coded clusters for a more straightforward identification of each particular issue within an area (Figure 26). Users receive ongoing updates on the number of problems per region prior to the full expansion of the respective cluster.



Figure 23. The city of Patras in Greece.

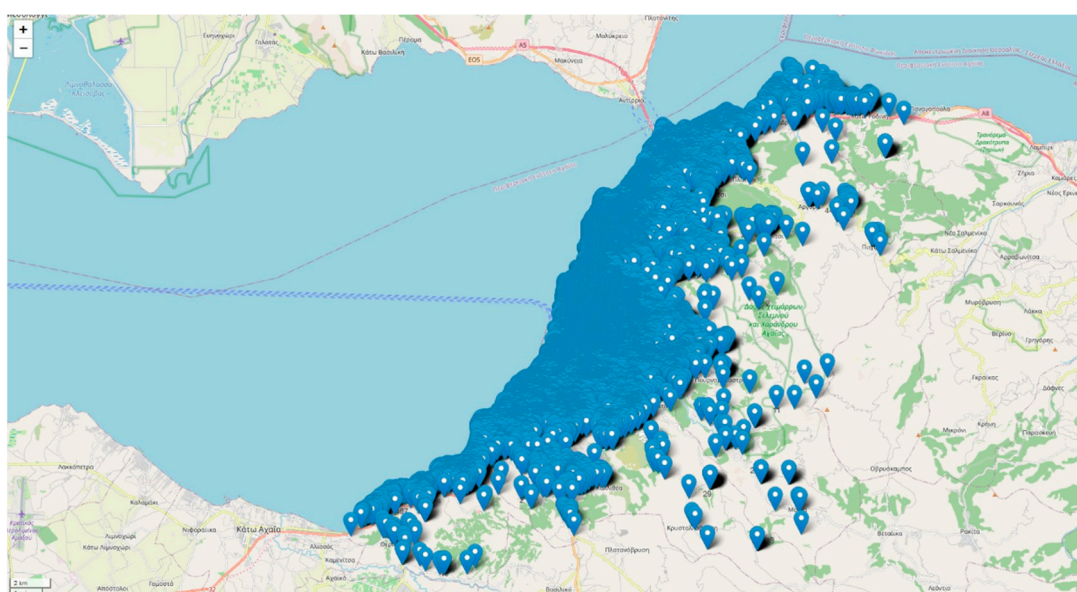


Figure 24. The spread of the reported issues in the city in the last 5 years.

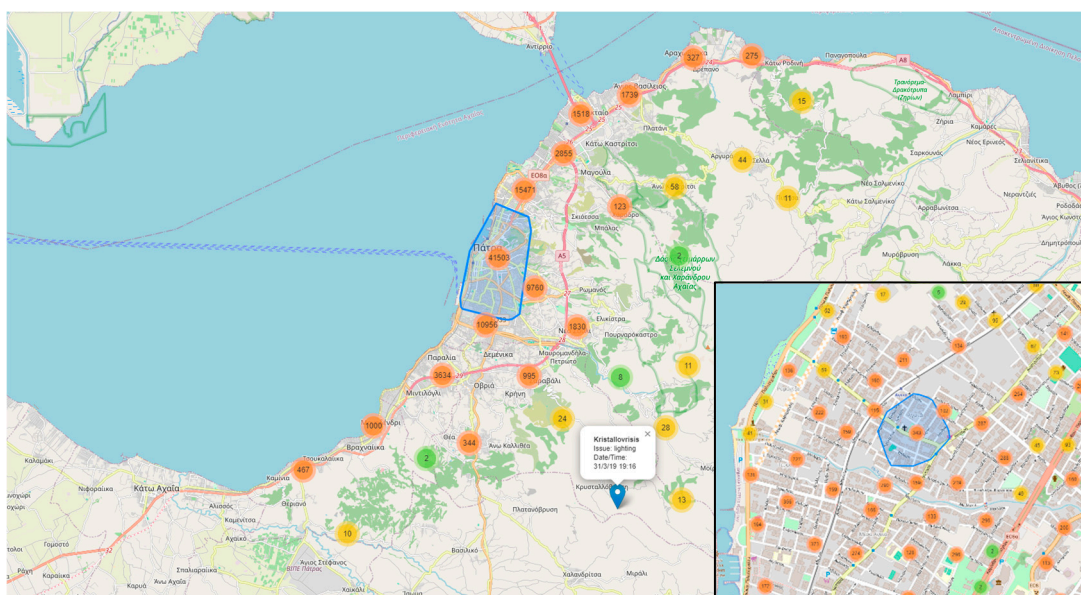


Figure 25. Issue clustering and information pop-up tag.

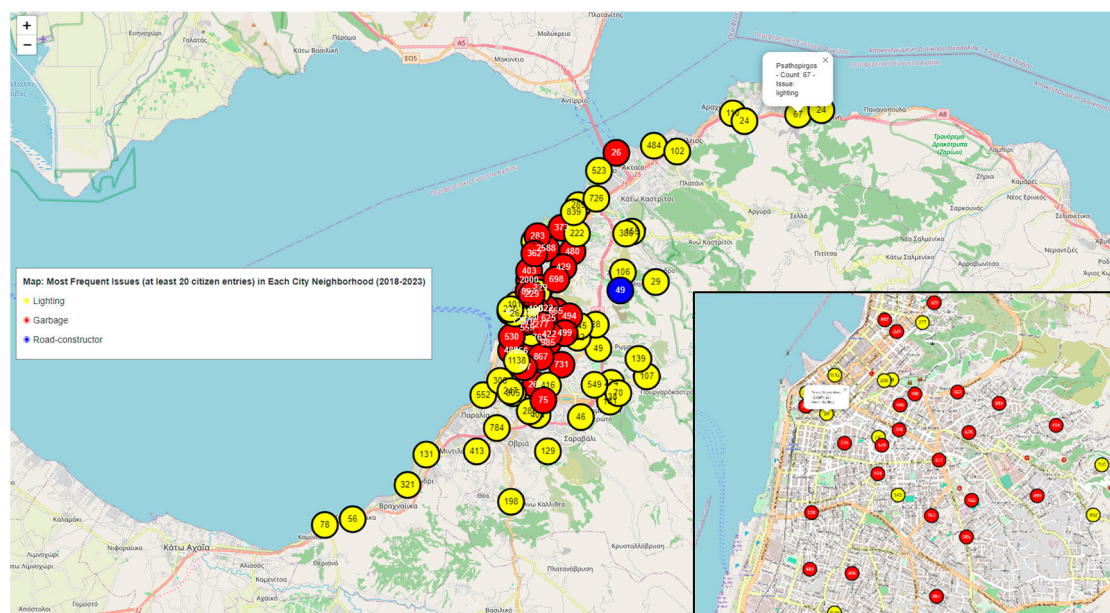


Figure 26. Filtering the frequency of issues and color clustering.

Subsequently, the trained machine learning model was employed to predict the likelihood of each issue reoccurring in every neighborhood for the next six months. Figure 27 illustrates the probabilities of issue-by-area occurrences on the interactive map for the upcoming three months. Using the mean values in this context is a way to aggregate or summarize the data. By calculating the mean probability for each combination of 'Areas' and 'issue', you are representing a central tendency for the probabilities associated with each issue in each area. This summary provides a high-level view of the data, allowing you to identify areas where certain issues have a higher or lower mean probability. The choice to use mean values might be based on the assumption that a single value can represent the overall probability for a specific issue in a particular area, smoothing out potential variations or outliers in the data. This approach can simplify the visualization and highlight general patterns in the dataset. In this iteration, users can now conveniently stay informed about the predominant issue in each area through the color-coded clustering on the map. Furthermore, they have the ability to assess the likelihood percentage of the most prevalent issue reoccurring in the upcoming 6 months. The map has undergone enhancements in terms of the information presented in each pop-up window, providing insights into the probability of the most common issue resurfacing in the respective area.

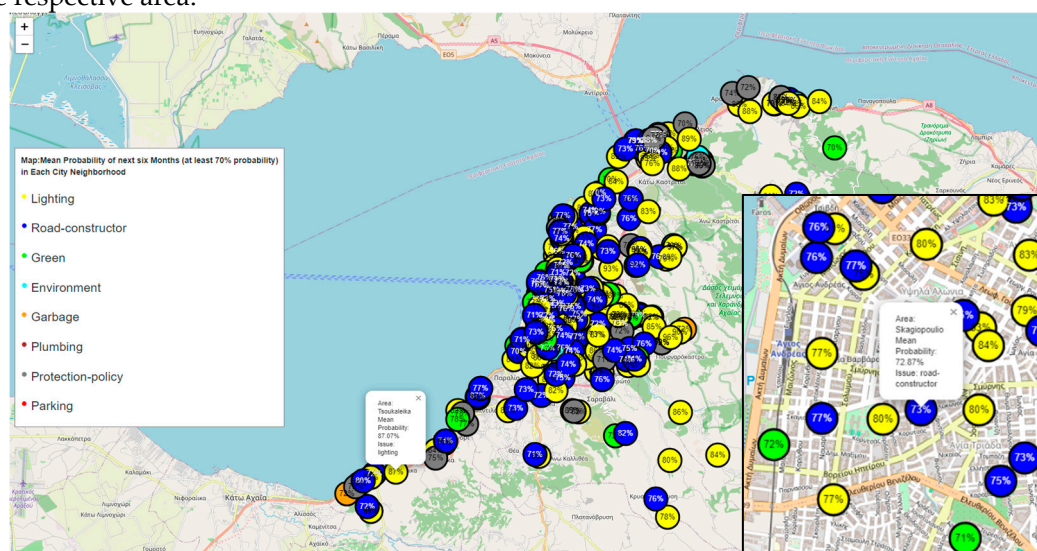


Figure 27. Issue probability for the next six months in each city neighborhood.

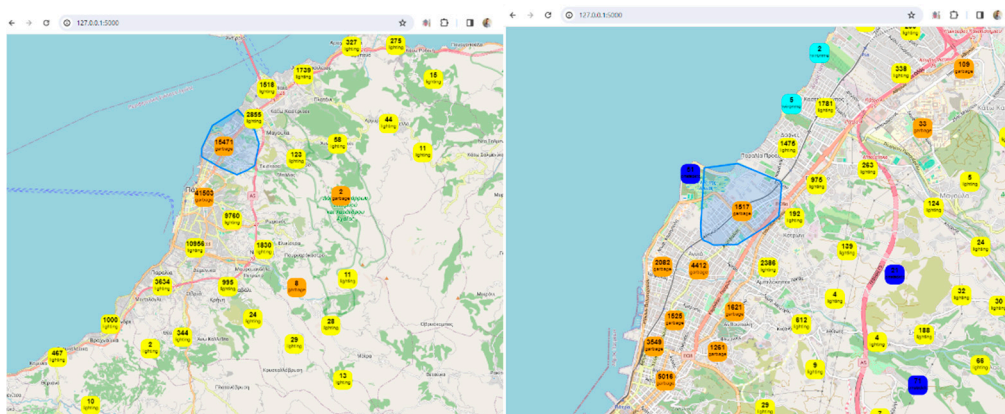


Figure 29. The digital twin app, an interactive service.

This approach to visualizing and interacting with data related to a physical or virtual environment is common in Digital Twin projects. In the case of this current work, the map serves as a representation of different areas, and the markers highlight issues reported in these areas. Users can explore the map, zoom in for more details, and click on markers to obtain specific information about each issue.

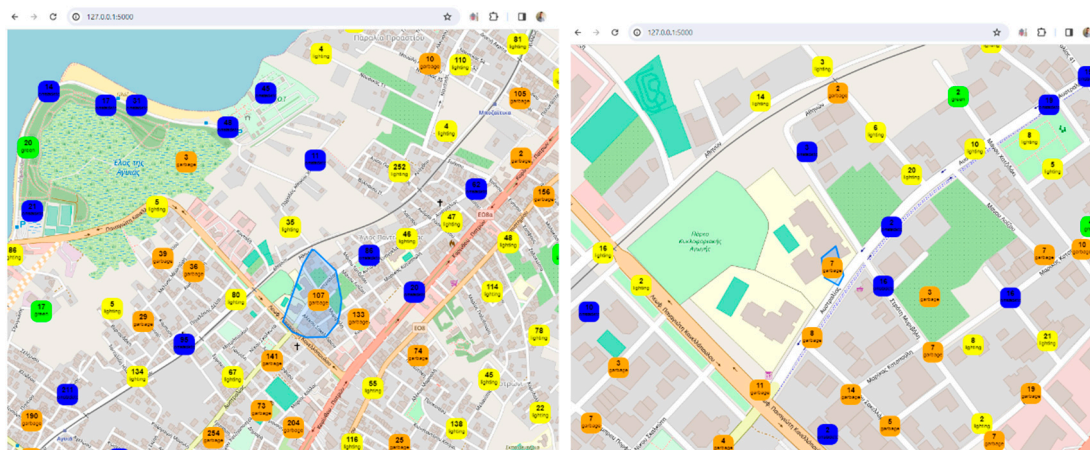


Figure 30. The digital twin app, zooms to the lowest level of the neighborhood.

All the aforementioned steps led to a stable yet dynamic understanding of the needs of the digital twin operator. These transformations contributed to the development of the platform, with a focus on capturing the unique characteristics of the physical layer in a digital twin environment. The continuous vigilance of the work through ongoing additions and improvements for a more representative presentation of the city's pulse in its digital twin version led to the further enrichment of the web application with new features (Figure 31).

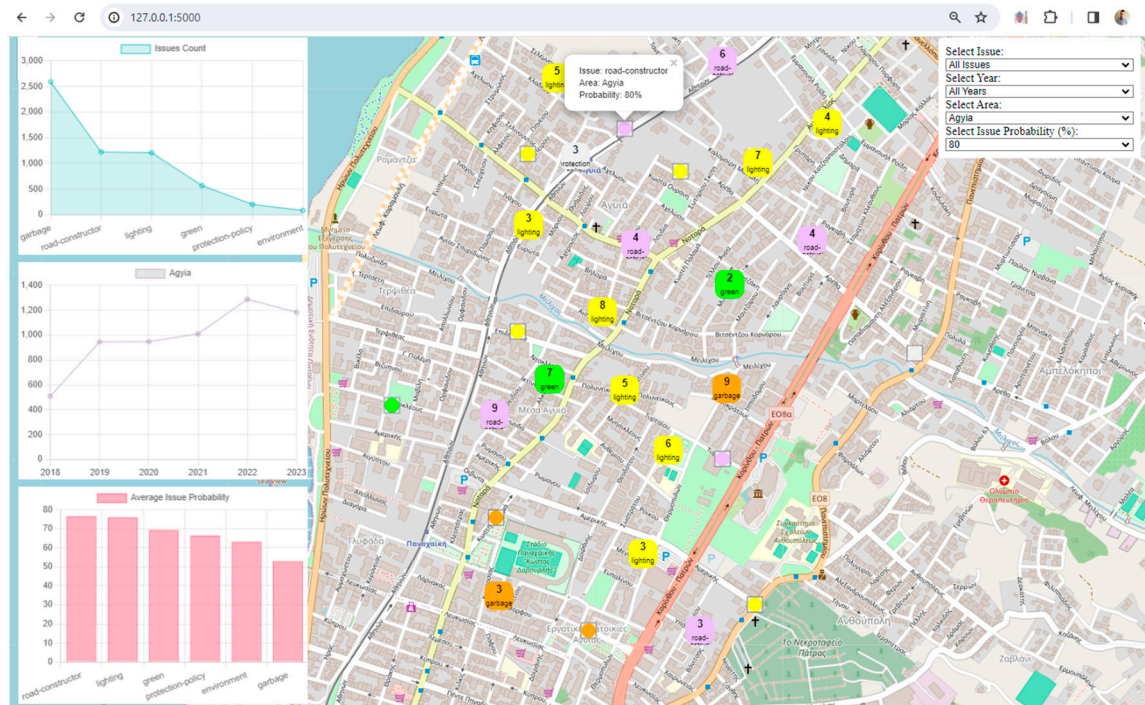


Figure 32. The digital twin app can display forecasts, issues and area counts for each reported issue by area.

4. Discussion

The Digital Twin project in the city of Patras, as outlined in the previous sections, introduces a sophisticated approach to urban management as is able to learn from historical data and real-time data to make informed decisions. The significance of this research lies in its potential to provide city planners, policymakers, and administrators with a valuable tool for understanding and addressing the diverse challenges that characterize Smart Cities. By harnessing the power of multiclass classification, cities can streamline their decision-making processes, allocate resources judiciously, and work towards creating more resilient and adaptive urban spaces. Building upon temporal analysis and digital twin applications, this paper explores the synergies between the developed multiclass classification model and the digital twin. By integrating the insights garnered from temporal analysis with the real-time capabilities of the digital twin, the work demonstrates a holistic approach to Smart City issue resolution. This integrated framework holds the potential to enhance the accuracy and efficiency of the multiclass classification model by incorporating real-time data and simulations.

The time series analysis and the utilization of a custom Python solution with open-source technologies, including Flask as the web framework, produce a commitment to tailoring the technology to the specific needs of the city. The literature review provides a foundation for interpreting the Digital Twin results [6–14,30,31]. These studies underline the potential of citizen feedback data to inform urban planning and decision-making. The Digital Twin in Patras aligns with this literature, utilizing similar methodologies to analyze and visualize citizen-reported issues. Here is a more extensive analysis of the various elements of this work.

Time Series Analysis: The time series analysis that comes with the application, rendering the temporal analysis of the data attributed to the digital twin application, introduces a valuable dimension to urban management. The temporal analysis allows for a nuanced understanding of the evolving nature of reported issues. By tracking trends over time, city officials can identify patterns, anticipate recurring challenges, and make informed decisions. Drawing insights from similar studies, the Digital Twin's time series analysis can be contextualized. Patterns observed in the reports, similar to those in the literature, may indicate specific issues prevalent in the urban landscape. The positive correlation between the number of reports and the presence of certain urban elements, as found in

the literature, can provide a basis for understanding the significance of reported issues in the Digital Twin [1,2,10–16,30,31].

Custom Python Solution: The decision to employ a custom Python solution for the Digital Twin project is strategic, providing flexibility and customization. Python's versatility, coupled with the Flask micro web framework, allows for the seamless integration of back-end functionalities. The implementation of a local intranet development server, as detailed in section materials and methods, ensures an efficient and localized environment, contributing to the creation of a robust Digital Twin platform. The paper on Flask-based digital twin applications contributes to understanding the technical architecture of the Digital Twin in Patras. The proposed frameworks in these papers highlight the versatility of Flask for developing smart city digital twins [6–17]. The integration of machine learning models in these frameworks mirrors the approach taken in the Patras Digital Twin, emphasizing the role of Flask in building scalable and intelligent urban management platforms.

Interactive Visualization: The heart of the Digital Twin project lies in its ability to present complex urban data through interactive visualization. The incorporation of Leaflet and Chart.js in the web application enhances the user experience, offering dynamic maps and charts. The dynamic filters and charts, as discussed in the results, empower users to interact with the data, tailoring their exploration to specific issues, areas, and timeframes. This not only facilitates a deeper understanding of the urban landscape but also ensures that stakeholders can derive meaningful insights from the vast dataset.

Predictive Modeling and Machine Learning: The integration of machine learning models, specifically the Random Forest Classifier, adds a predictive dimension to the Digital Twin. The models, trained on historical data, contribute to forecasting potential issues in different neighborhoods over the next six months. This forward-looking capability positions the Digital Twin as a proactive tool for urban planning, allowing city officials to anticipate and mitigate challenges before they escalate. The studies on machine learning applications in urban settings, offer a parallel to the Digital Twin's use of machine learning models. The application of machine learning algorithms allows for a more sophisticated analysis of the vast dataset, enabling the identification of patterns that might not be immediately apparent. This resonates with the Digital Twin's use of a Random Forest Classifier for predictive modeling. The literature showcases a combination of digital twin technology and machine learning for enhancing public services [5,8,12–14,22]. In a similar vein, the Digital Twin in Patras leverages machine learning within a Flask-based framework to optimize urban planning. This integration signifies a synergy between the dynamic representation of the city (Digital Twin) and advanced analytical tools (machine learning) to improve public services.

Spatial Clustering: The spatial clustering of reported issues on the interactive map is a pivotal feature. Clusters serve as a visual aid, quickly highlighting areas with a higher concentration of problems. This functionality streamlines decision-making processes, enabling city officials to prioritize interventions and allocate resources efficiently.

Continuous Development and Adaptability: The commitment to continuous development, as evident in the frequent updates and additional features, underscores the adaptability of the Digital Twin. The inclusion of new filters and charts, as showcased in Figure 31, ensures that the platform remains responsive to the evolving needs of its operators. This adaptability positions the Digital Twin as a scalable solution capable of incorporating additional functionalities and datasets in the future.

User Empowerment: The emphasis on user interaction and empowerment through the application's features is noteworthy. By providing users with the ability to filter issues, years, areas, forecasts and visualize data based on their preferences, the Digital Twin becomes a versatile tool for diverse stakeholders. This user-centric approach enhances engagement and ensures that the platform is accessible and beneficial to a wide range of users, from city officials to residents. The literature on the potential of digital twins for urban planning aligns with the broader implications of the Digital Twin project in Patras. The ability to simulate different urban planning scenarios corresponds to the Digital Twin's capability to predict and simulate future issues based on historical data. This reinforces the transformative potential of digital twins in revolutionizing urban planning methodologies.

Potential for Real-Time Integration: The architecture's extensibility, particularly the incorporation of Flask, opens doors to future enhancements. The suggestion of integrating real-time data from various sensors and IoT devices positions the Digital Twin as a platform that can evolve to incorporate emerging technologies. This potential for real-time data integration holds promise for creating a more dynamic and responsive urban management system.

5. Conclusions

In conclusion, the Digital Twin project in Patras stands as a modern and comprehensive urban management solution, leveraging time series analysis, citizen feedback data, and advanced technologies. The incorporation of time series analysis adds a crucial temporal dimension to issue tracking, enabling city officials to discern evolving patterns and make informed decisions. This aligns seamlessly with findings from studies on citizen feedback data, emphasizing the importance of temporal trends in understanding urban challenges. The comparison with the literature, particularly studies analyzing citizen feedback data from various cities worldwide, provides a robust context for interpreting the results. The Digital Twin's capacity to identify and visualize urban issues resonates with similar findings in the literature, reinforcing the efficacy of citizen-reported data in shaping urban policies. The integration of machine learning within the Digital Twin aligns with the broader trends identified in the literature. Studies showcasing the power of machine learning algorithms to analyze vast datasets and improve urban planning find practical applications in the Digital Twin's predictive modeling and issue forecasting. The Flask-based framework, as demonstrated in the literature, proves to be a versatile and scalable solution for developing intelligent urban management platforms.

Furthermore, the Digital Twin's combination of machine learning and a dynamic representation of the city echoes successful models discussed in the literature. This synergy, as seen in studies on improving public transportation and assessing the potential of digital twins for urban planning, positions the Patras Digital Twin at the core of smart city initiatives. In essence, the Digital Twin in Patras represents a technological advancement and embodies a paradigm shift in urban planning and management. By drawing on insights from literature and implementing cutting-edge technologies, the project addresses current urban challenges and strengthens the foundation for future advancements in smart city initiatives. The potential for informed decision-making, infrastructure optimization, and enhanced public services positions the Digital Twin in Patras as an additional transformative force in transforming the cities of tomorrow.

In summary, this paper offers a comprehensive investigation into the amalgamation of data analytics, predictive forecasting, and digital twin technologies. Through an in-depth exploration of these elements, the paper seeks to unravel the transformative potential of these innovations in sculpting the urban landscapes of tomorrow, where cities not only respond intelligently to current challenges but also proactively shape a sustainable and resilient future for their residents.

The study identifies a limitation related to the absence of standardized protocols and formats for digital twins, posing challenges to interoperability. Different systems and platforms may struggle to communicate and share data seamlessly. Integrating various digital twin models and datasets from different domains, such as transportation, energy, and healthcare is complex.

Another noteworthy limitation highlighted by the study pertains to machine learning algorithms, which heavily rely on historical data that may inherently contain biases. If not addressed, this can result in unfair or discriminatory outcomes, especially in areas such as public services and law enforcement. Understanding and explaining the decisions made by machine learning models is crucial for accountability and building trust.

Author Contributions: Andreas F. Gkontzis, and Vassilios S. Verykios; methodology, Andreas F. Gkontzis.; software, Andreas F. Gkontzis; validation, Andreas F. Gkontzis, Sotiris Kontsiantis, Georgios Feretzakis, and Vassilios S. Verykios; formal analysis, Andreas F. Gkontzis.; investigation, Andreas F. Gkontzis, and Sotiris Kontsiantis; resources, Andreas F. Gkontzis; data curation, Andreas F. Gkontzis, Sotiris Kontsiantis, Georgios Feretzakis and Vassilios S. Verykios; writing—original draft preparation, Andreas F. Gkontzis, and Vassilios S.

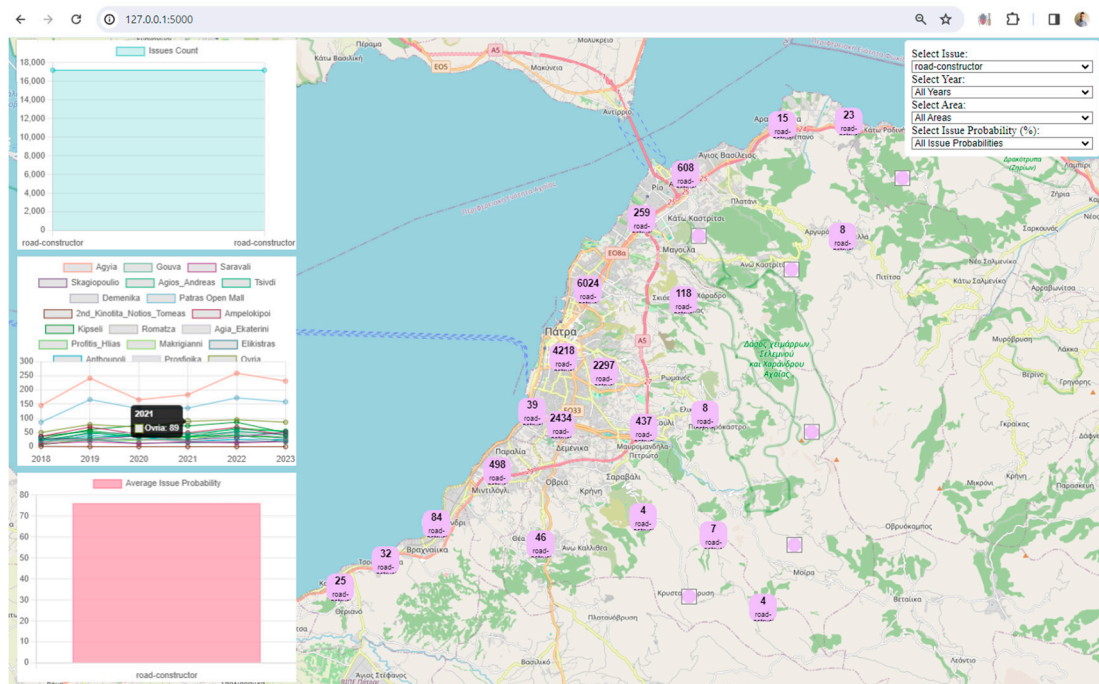


Figure A3. Digital twin app, updates charts, issue clusters, and areas (filter issue: road-constructor).

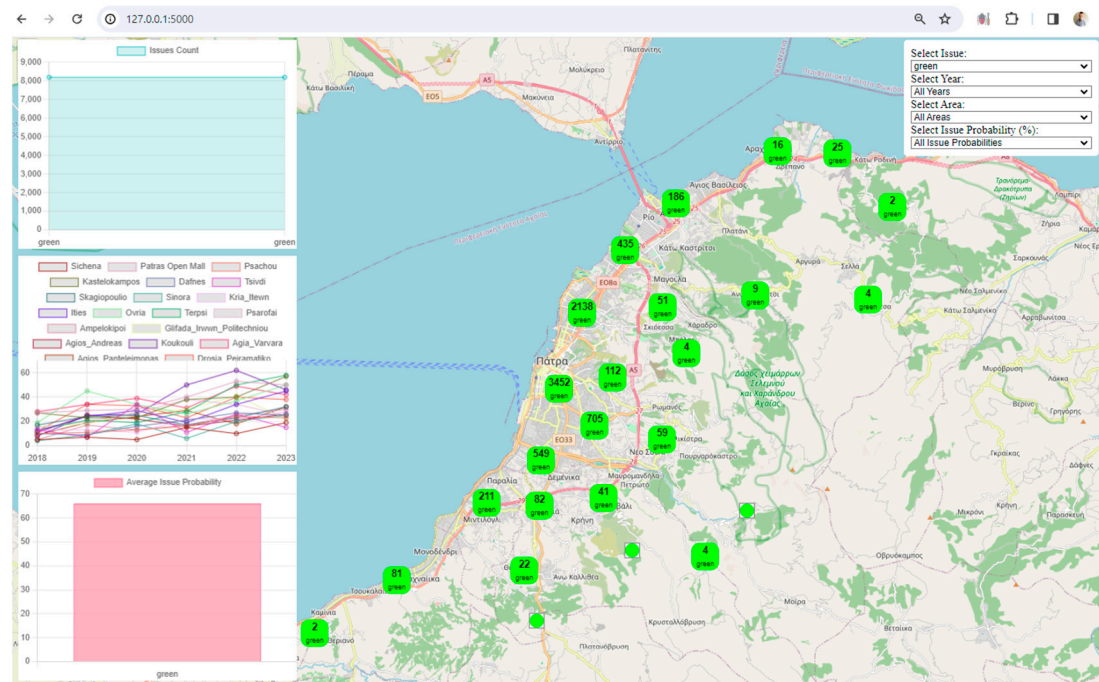


Figure A4. The digital twin app, updates charts, issue clusters, and areas (filter issue: green).

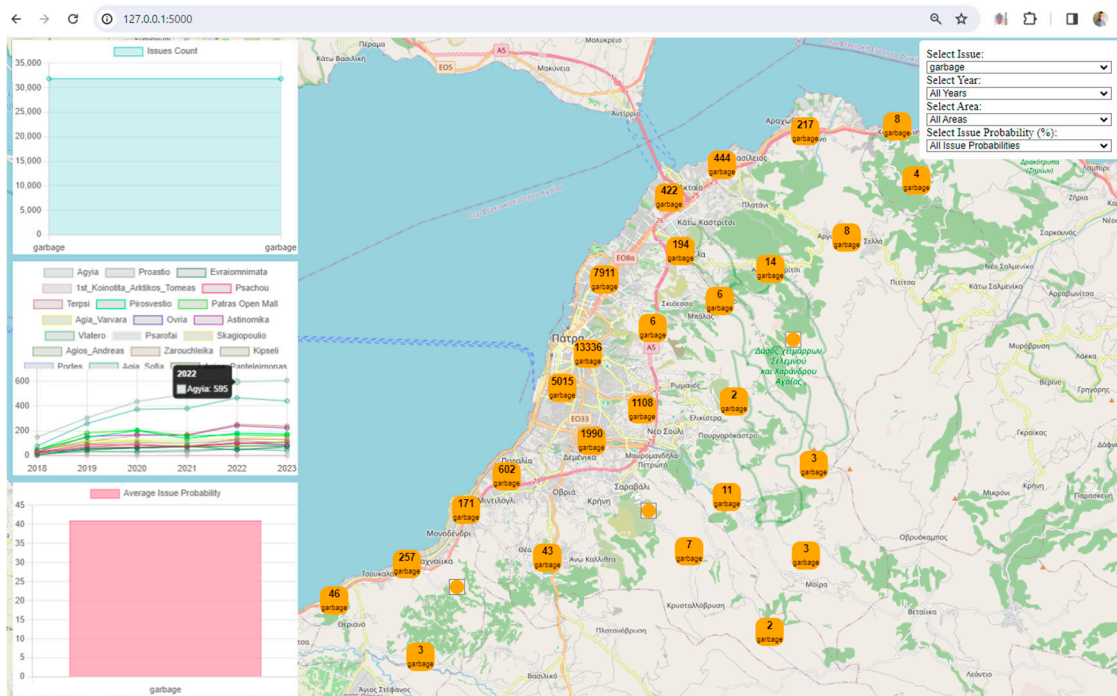


Figure A5. Digital twin app, updates charts, issue clusters, and areas (filter issue: garbage).

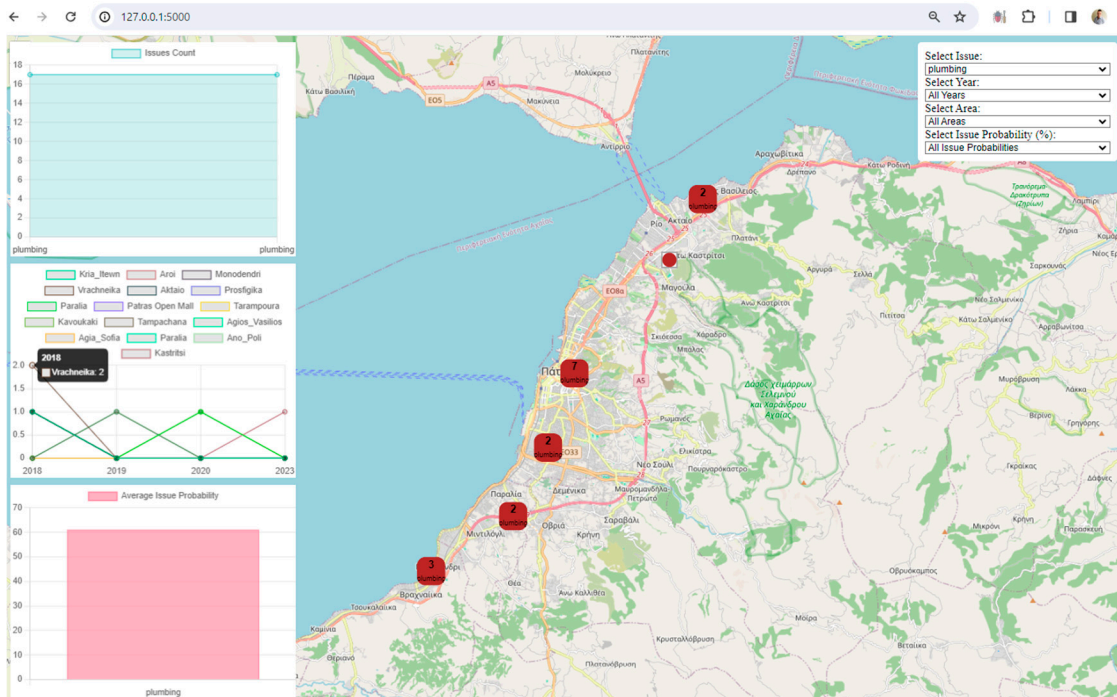


Figure A6. The digital twin app, updates charts, issue clusters, and areas (filter issue: plumbing).

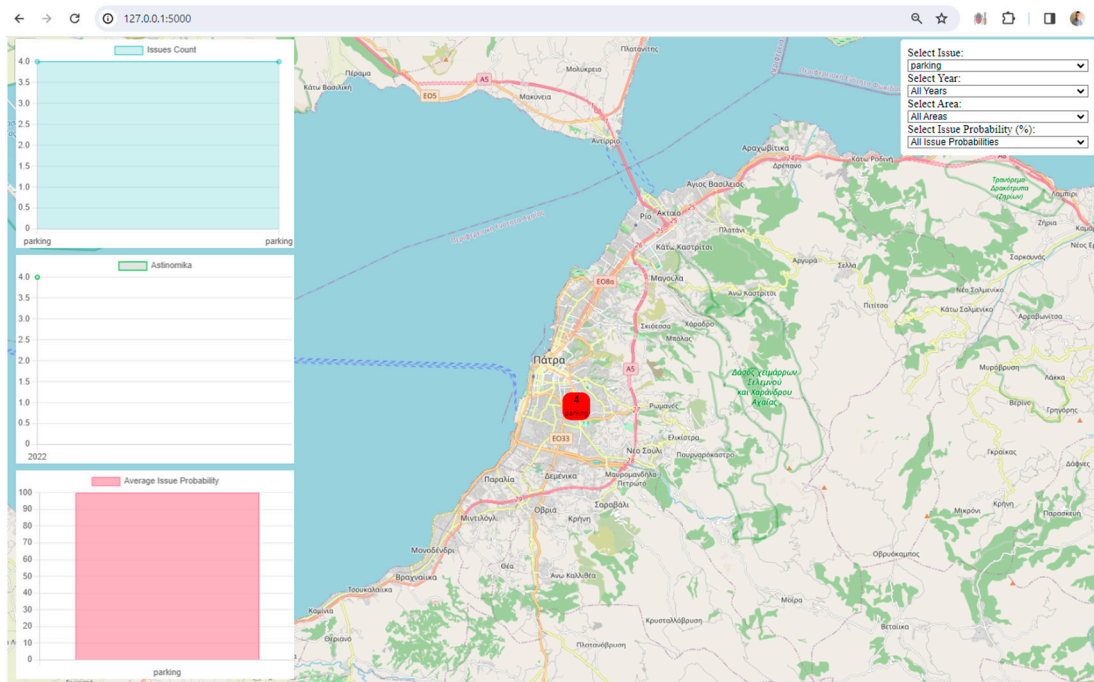


Figure A9. Digital twin app, updates charts, issue clusters, and areas (filter issue: parking).

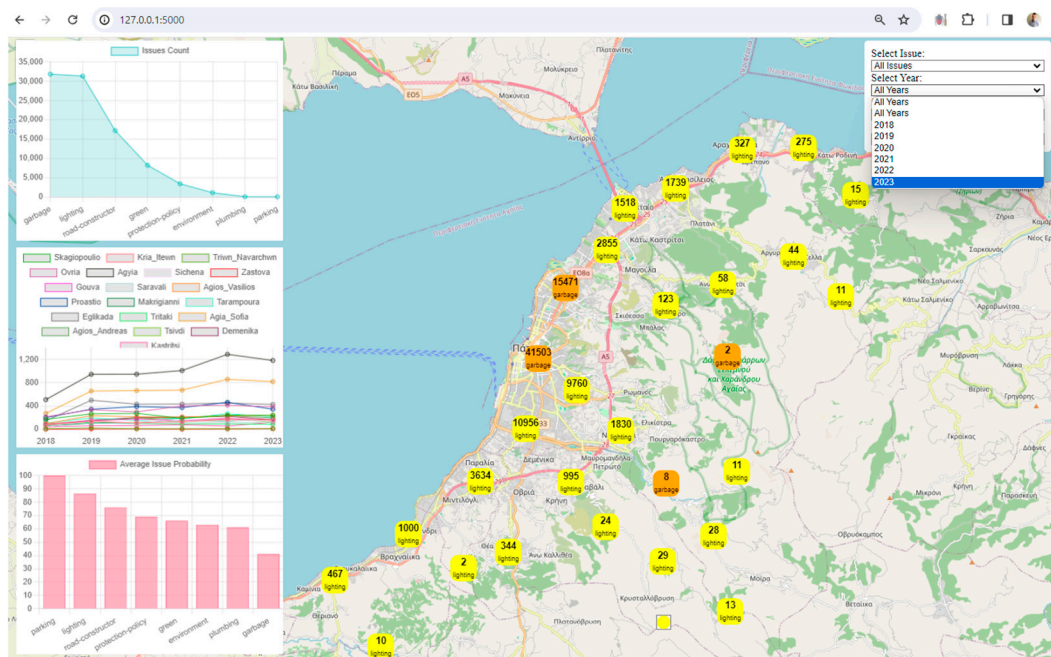


Figure A10. Digital twin app, updates charts, issue clusters, and areas (filter: years).

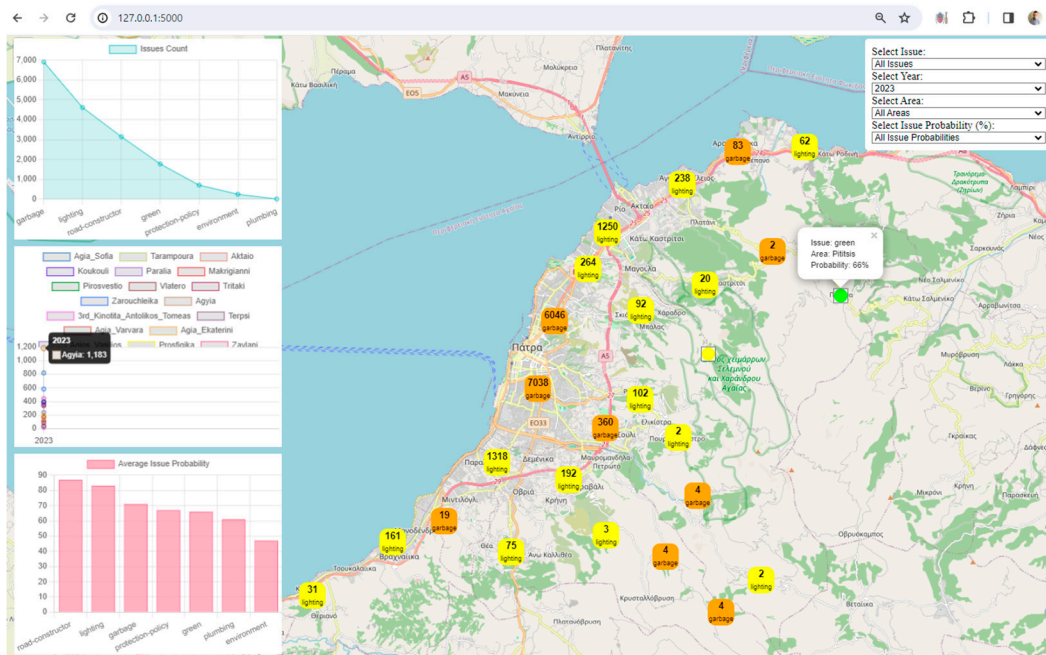


Figure A11. Digital twin app, updates charts, issue clusters, and areas (filter years: 2023).

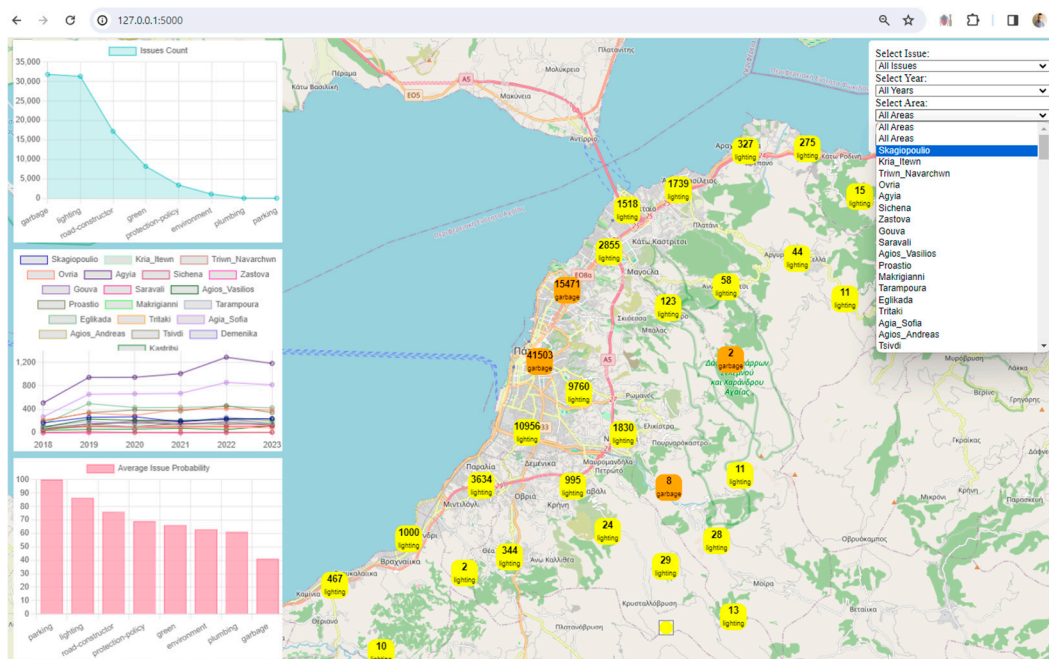


Figure A12. The digital twin app, updates charts, issue clusters, and areas (filter: areas).

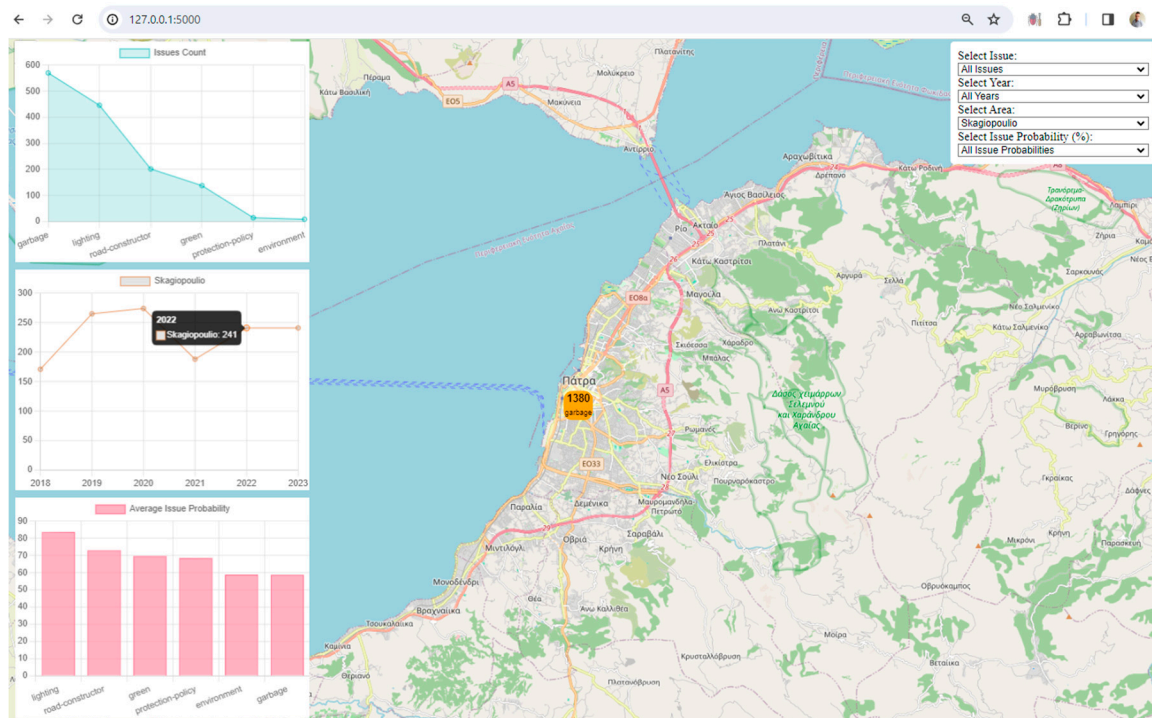


Figure A13. Digital twin app, updates charts, issue clusters, and areas (filter areas: skagiopoulio).

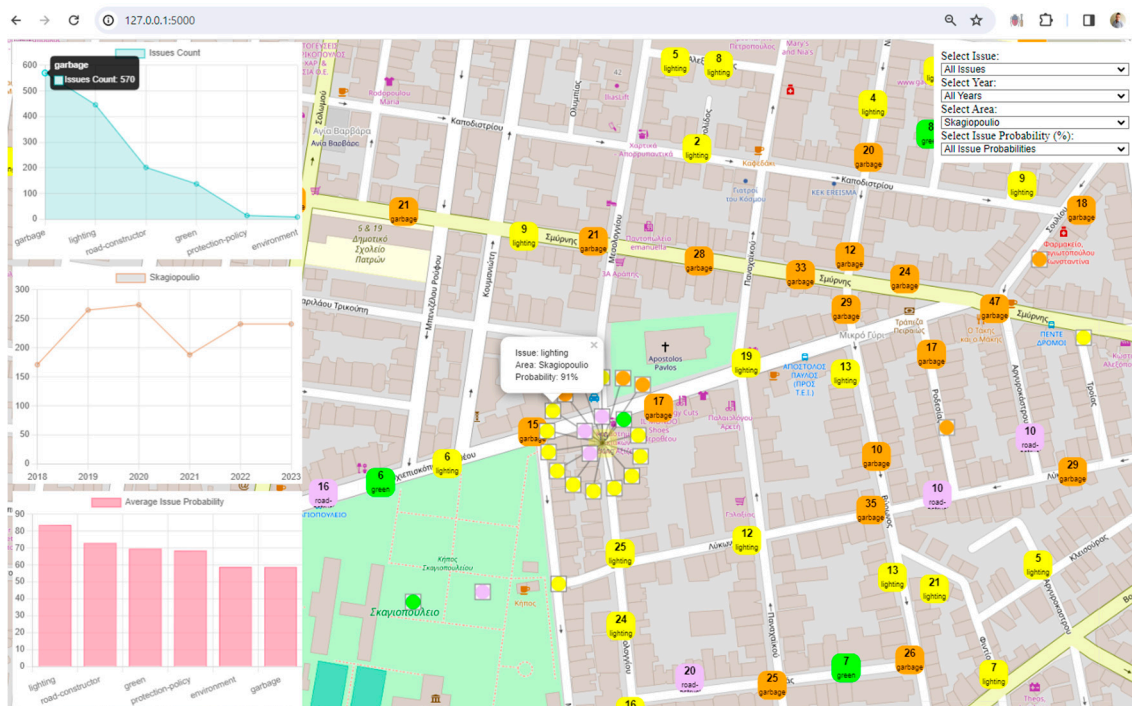


Figure A14. Digital twin app, updates charts, issue clusters, and areas (filter areas: skagiopoulio).

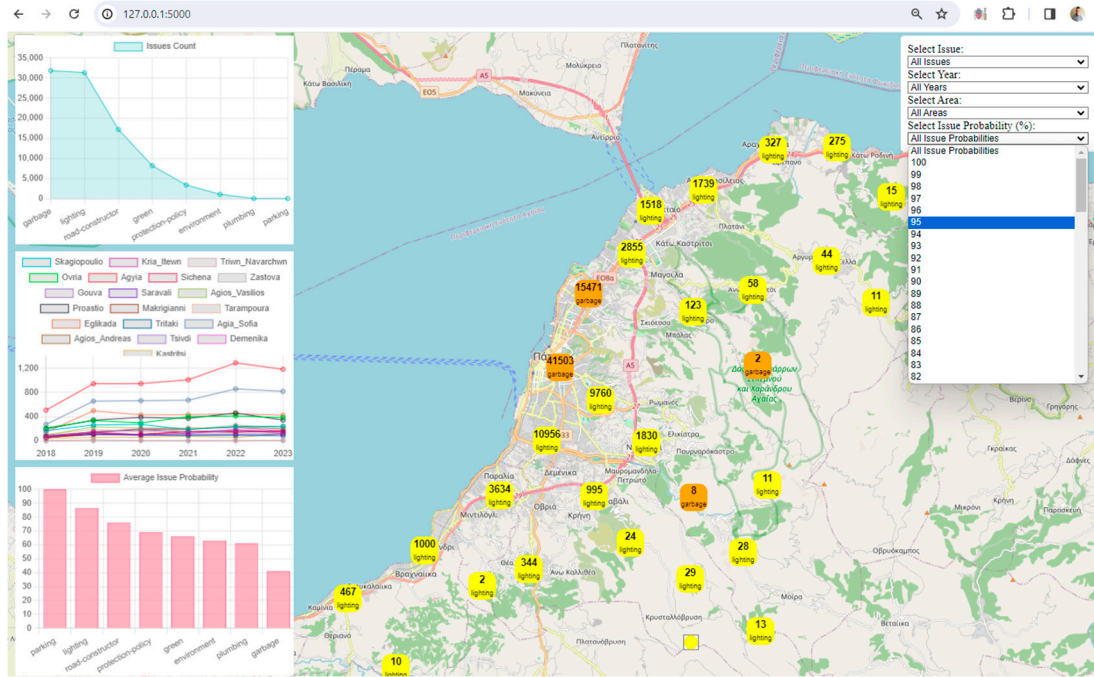


Figure A15. The digital twin app, updates charts, issue clusters, and areas (filter: probabilities).

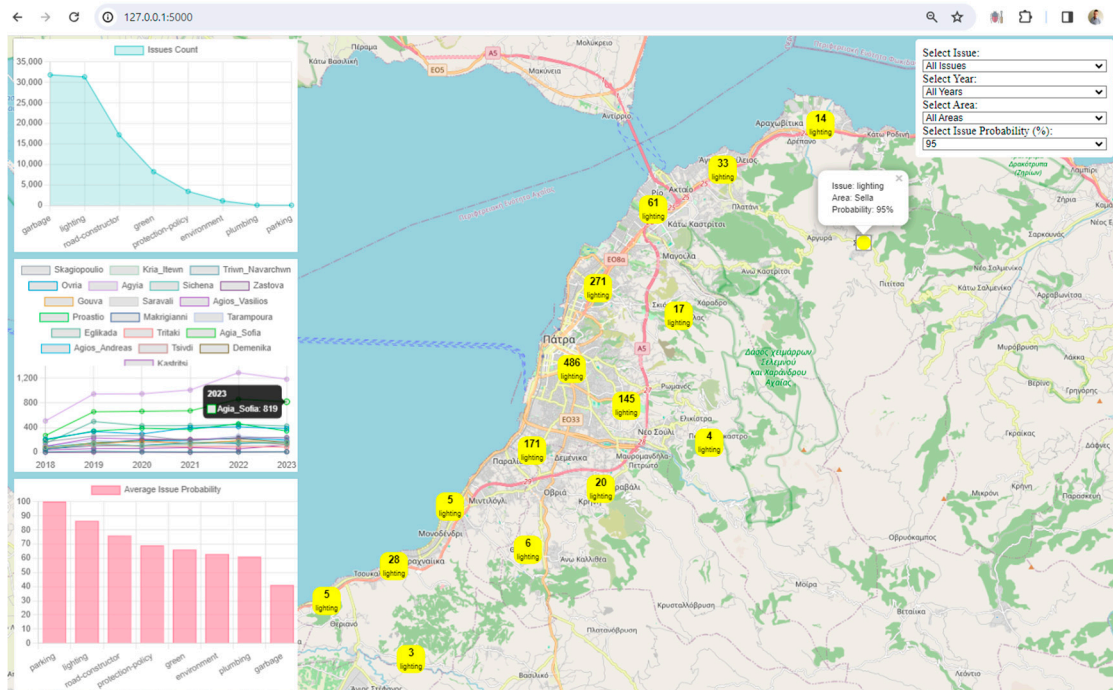


Figure A16. Digital twin app, updates charts, issue clusters, and areas (filter probabilities: 95%).

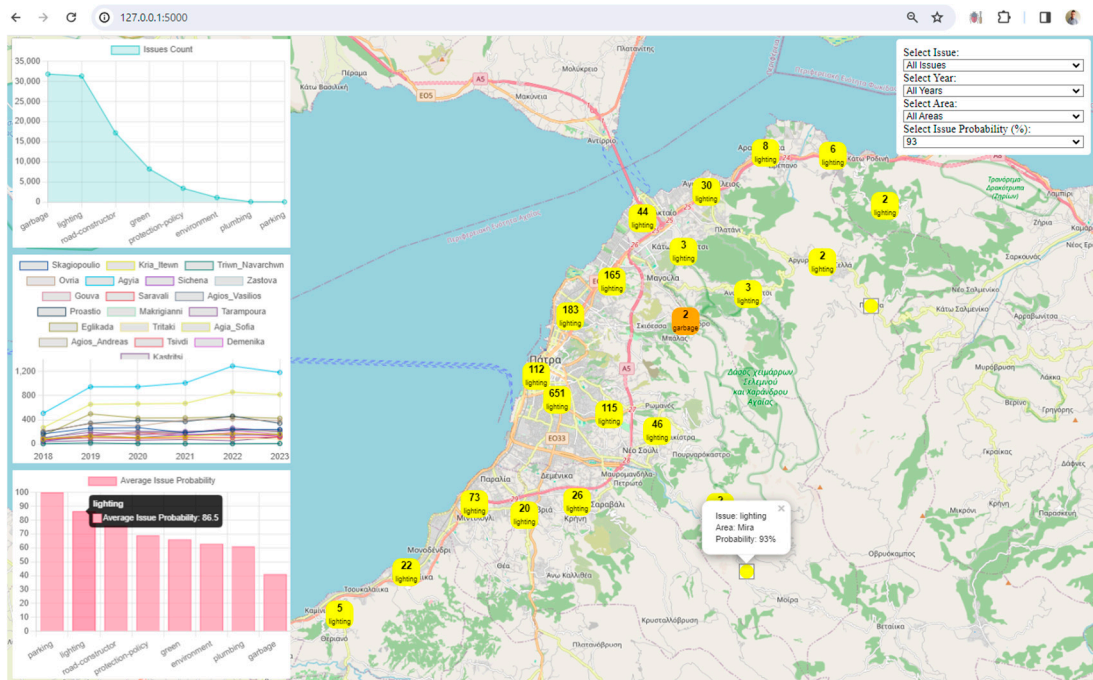


Figure A17. Digital twin app, updates charts, issue clusters, and areas (filter probabilities: 93%).

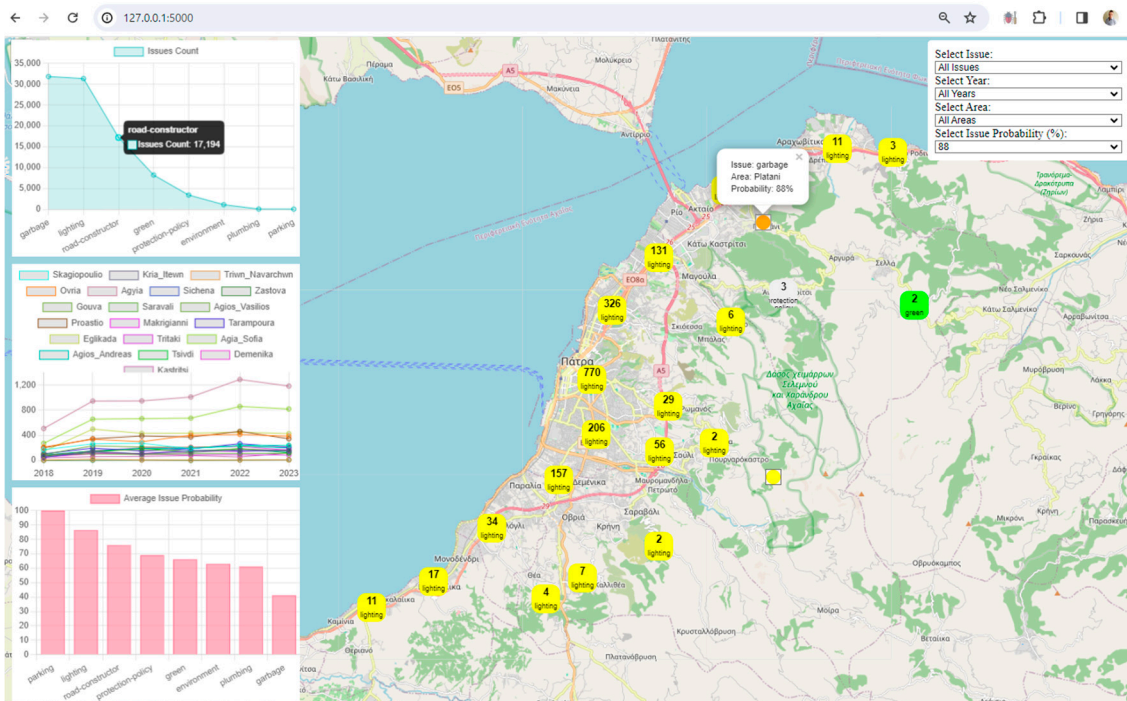


Figure A18. Digital twin app, updates charts, issue clusters, and areas (filter probabilities: 88%).

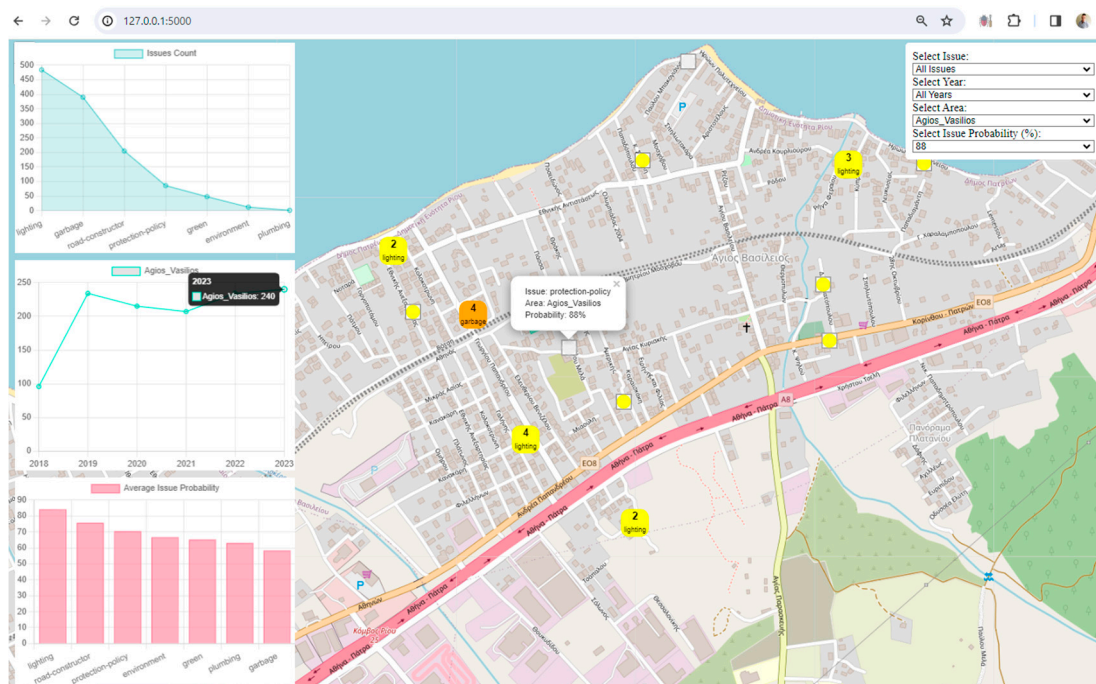


Figure A19. Digital Twin app, updates charts, issue clusters, and areas (two filters: area, probabilities).

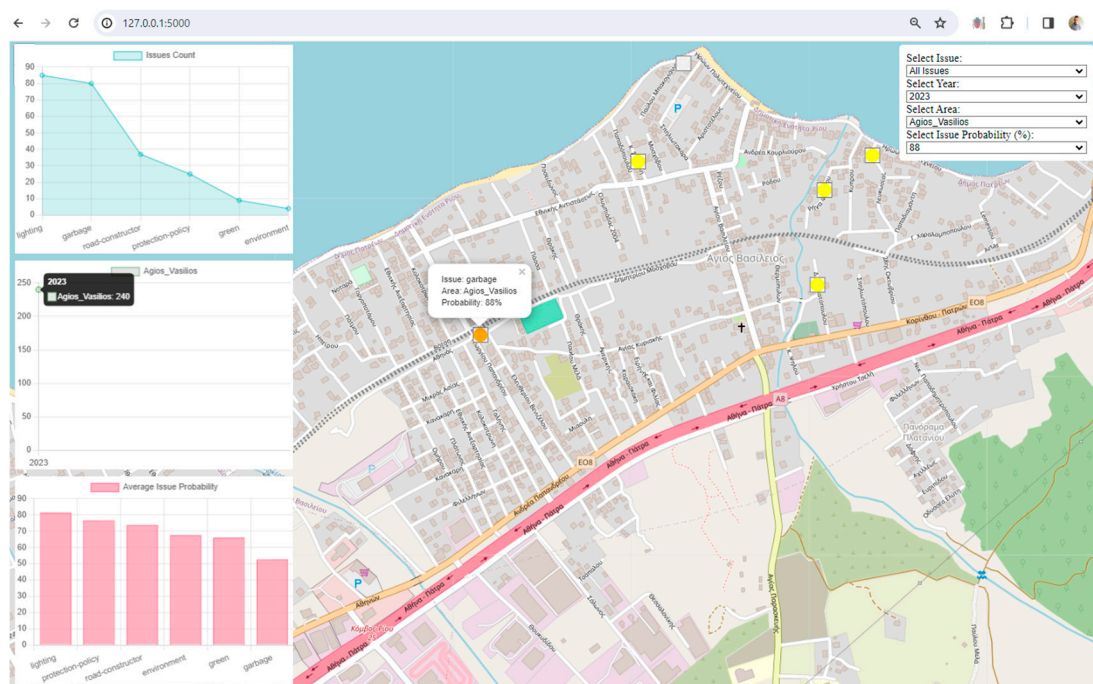


Figure A20. The digital twin app, updates charts, issue clusters, and areas (three filters: year, area, and probabilities).

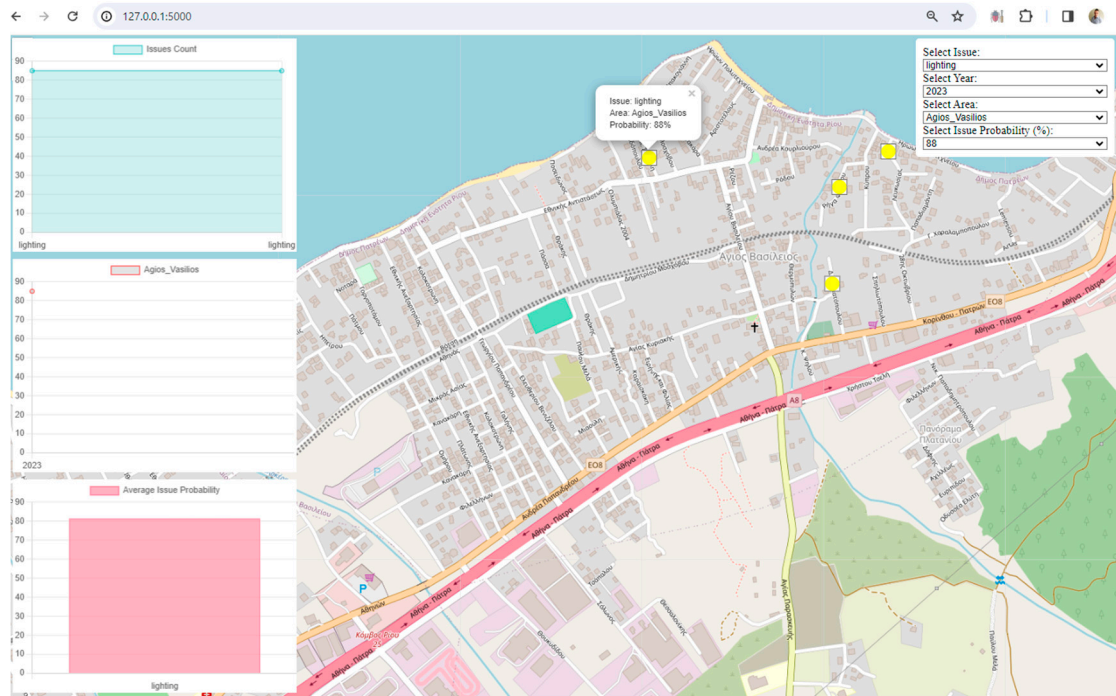


Figure A21. The digital twin app, updates charts, issue clusters, and areas (four filters: issue, year area, probabilities).

References

1. Weil, C.; Bibri, S. E.; Longchamp, R.; Golay, F.; Alahi, A. Urban Digital Twin Challenges: A Systematic Review and Perspectives for Sustainable Smart Cities. *Sustainable Cities and Society* 2023, 99, 104862. <https://doi.org/10.1016/j.scs.2023.104862>.
2. Deng, T.; Zhang, K.; Shen, Z.-J. (Max). A systematic review of a digital twin city: A new pattern of urban governance toward smart cities. *Journal of Management Science and Engineering* 2021, 6(2), 125-134. <https://doi.org/10.1016/j.jmse.2021.03.003>
3. Wolf, K.; Dawson, R. J.; Mills, J. P.; Blythe, P.; Morley, J. Towards a digital twin for supporting multi-agency incident management in a smart city. *Scientific Reports* 2022, 12(1). <https://doi.org/10.1038/s41598-022-20178-8>
4. Gkontzias, A.F.; Kalles, D.; Paxinou, E.; Tsoni, R.; Verykios, V.S. A Big Data Analytics Conceptual Framework for a Smart City: A Case Study. In: Fitsilis, P. (eds) *Building on Smart Cities Skills and Competences. Internet of Things*. Springer, Cham, 2022. https://doi.org/10.1007/978-3-030-97818-1_16
5. Salamanis, A.I.; Gravvanis, G.A.; Kotsiantis, S.B.; Vrahatis, M.N. Novel Sparse Feature Regression Method for Traffic Forecasting. *International Journal on Artificial Intelligence Tools* 2023, 32(01). <https://doi.org/10.1142/S0218213023500082>
6. White, G.; Zink, A.; Codecá, L.; Clarke, S. A digital twin smart city for citizen feedback. *Cities* 2021, 110, 103064. <https://doi.org/10.1016/j.cities.2020.103064>
7. Najafi, P.; Mohammadi, M.; Wesemael, P.; Le Blanc, P.M. A user-centred virtual city information model for inclusive community design: State-of-art. *Cities* 2023, 134, 104203. <https://doi.org/10.1016/j.cities.2023.104203>
8. Abdeen, F.N.; Shirowzhan, S.; Sepasgozar, S.M.E. Citizen-centric digital twin development with machine learning and interfaces for maintaining urban infrastructure. *Telematics and Informatics* 2023, 84, 102032. <https://doi.org/10.1016/j.tele.2023.102032>
9. Jeddoub, I.; Nys, G.-A.; Hajji, R.; Billen, R. Digital Twins for cities: Analyzing the gap between concepts and current implementations with a specific focus on data integration. *International Journal of Applied Earth Observation and Geoinformation* 2023, 122, 103440. <https://doi.org/10.1016/j.jag.2023.103440>
10. Xia, H.; Liu, Z.; Efrechokina, M.; Liu, X.; Lin, C. Study on city digital twin technologies for sustainable smart city design: A review and bibliometric analysis of geographic information system and building

- information modeling integration. *Sustainable Cities and Society* 2022, 84, 104009. <https://doi.org/10.1016/j.scs.2022.104009>
11. Bonney, M.S.; de Angelis, M.; Dal Borgo, M.; et al. Development of a digital twin operational platform using Python Flask. *Data-Centric Engineering* 2022, 3, e1. <https://doi.org/10.1017/dce.2022.1>
 12. Casali, Y.; Aydin, N.Y.; Comes, T. Machine learning for spatial analyses in urban areas: a scoping review. *Sustainable Cities and Society* 2022, 85, 104050. <https://doi.org/10.1016/j.scs.2022.104050>
 13. Omar, A.; Delnaz, A.; Nik-Bakht, M. Comparative analysis of machine learning techniques for predicting water main failures in the City of Kitchener. *Journal of Infrastructure Intelligence and Resilience* 2023, 2(3), 100044. <https://doi.org/10.1016/j.iintel.2023.100044>
 14. Ariyachandra, M.F.; Wedawatta, G. Digital Twin Smart Cities for Disaster Risk Management: A Review of Evolving Concepts. *Sustainability* 2023, 15, 11910. <https://doi.org/10.3390/su151511910>
 15. Martí, P.; Serrano-Estrada, L.; Nolasco-Cirugeda, A. Social Media data: Challenges, opportunities and limitations in urban studies. *Computers, Environment and Urban Systems* 2019, 74, 161-177. <https://doi.org/10.1016/j.compenvurbsys.2018.11.001>
 16. Hu, A.; Yabuki, N.; Fukuda, T.; Kaga, H.; Takeda, S.; Matsuo, K. Harnessing multiple data sources and emerging technologies for comprehensive urban green space evaluation. *Cities* 2023, 143. <https://doi.org/10.1016/j.cities.2023.104562>
 17. Wang, R.; Zhang, X.; Li, N. Zooming into mobility to understand cities: A review of mobility-driven urban studies. *Cities* 2022, 130, 103939. <https://doi.org/10.1016/j.cities.2022.103939>
 18. Chen, Z.; Chan, I.C.C. Smart cities and quality of life: a quantitative analysis of citizens' support for smart city development. *Information Technology & People* 2023, 36(1), 263-285. <https://doi.org/10.1108/ITP-07-2021-0577>
 19. Wirtz, B.W.; Müller, W.M.; Schmidt, F.W. Digital Public Services in Smart Cities – an Empirical Analysis of Lead User Preferences. *Public Organiz Rev* 2021, 21, 299–315. <https://doi.org/10.1007/s11115-020-00492-3>
 20. Wirtz, B.W.; Becker, M.; Schmidt, F.W. Smart city services: an empirical analysis of citizen preferences. *Public Organiz Rev* 2022, 22, 1063–1080. <https://doi.org/10.1007/s11115-021-00562-0>
 21. Li, W.; Feng, T.; Timmermans, H.J.P.; Li, Z.; Zhang, M.; Li, B. Analysis of citizens' motivation and participation intention in urban planning. *Cities* 2020, 106, 102921. <https://doi.org/10.1016/j.cities.2020.102921>
 22. Allen, B.; Tamindael, L.E.; Bickerton, S.H.; Cho, W. Does citizen coproduction lead to better urban services in smart cities projects? An empirical study on e-participation in a mobile big data platform. *Government Information Quarterly* 2020, 37(1), 101412. <https://doi.org/10.1016/j.giq.2019.101412>
 23. Oh, J. Smart City as a Tool of Citizen-Oriented Urban Regeneration: Framework of Preliminary Evaluation and Its Application. *Sustainability* 2020, 12, 6874. <https://doi.org/10.3390/su12176874>
 24. Savastano, M.; Suciuc, M.-Ch.; Gorelova, I.; Stăvîță, Gh.A. How smart is mobility in smart cities? An analysis of citizens' value perceptions through ICT applications. *Cities* 2023, 132, 104071. <https://doi.org/10.1016/j.cities.2022.104071>
 25. Noori, N.; Hoppe, T.; Jong, M. D.; Stamhuis, E. Transplanting good practices in Smart City development: A step-wise approach. *Government Information Quarterly* 2023, 40(2), 101802. <https://doi.org/10.1016/j.giq.2023.101802>
 26. Bala, M.; Boussaid, O.; Alimazighi, Z. A Fine-Grained Distribution Approach for ETL Processes in Big Data Environments. *Data & Knowledge Engineering* 2017, 111, 114-136. <https://doi.org/10.1016/j.datak.2017.08.003>
 27. Alexandropoulos, S.-A.N.; Kotsiantis, S.B.; Vrahatis, M.N. Data preprocessing in predictive data mining. "The Knowledge Engineering Review" 2019, 34, e1. <https://doi.org/10.1017/S026988891800036X>
 28. Gkontzis, A.F.; Kotsiantis, S.; Kalles, D.; Panagiotakopoulos, C.T.; Verykios, V.S. Polarity, Emotions, and Online Activity of Students and Tutors as Features in Predicting Grades. *International Journal of Intelligent Decision Technologies* 2020, 14(3), 409-436. <https://doi.org/10.1080/10494820.2019.1709209>
 29. Arafet, K., & Berlanga, R. Digital Twins in Solar Farms: An Approach through Time Series and Deep Learning. *Algorithms* 2021, 14, 156. <https://doi.org/10.3390/a14050156>
 30. Hu, W., He, Y., Liu, Z., Tan, J., Yang, M., & Chen, J. Toward a Digital Twin: Time Series Prediction Based on a Hybrid Ensemble Empirical Mode Decomposition and BO-LSTM Neural Networks. *J. Mech. Des.* 2020, 143. <https://doi.org/10.1115/1.4048414>

31. Zhang, T., Liu, X., Luo, Z., Dong, F., & Jiang, Y. Time series behavior modeling with digital twin for Internet of Vehicles. EURASIP J. Wirel. Commun. Netw. 2019. <https://doi.org/10.1186/s13638-019-1589-8>

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.