Research on Urban Data Visualization Based on Big Data: Transforming Insights into Action

Jibo Guo (guojibo@gdust.edu.cn)
Guangdong University of Science and Technology, Guangdong, China

Zhijing Li* (zhijingbaby@gmail.com), *Corresponding author
School of Artificial Intelligence; National Engineering Laboratory for Big Data System Computing Technology, Shenzhen University, PR China

Qiuying Yu (qiuying_0403@qq.com), Yiping Jiang (jyp2006@163.com)
Guangdong University of Science and Technology, Guangdong, China

Abstract

This paper presents a big data urban visualization platform for Guangdong Province, aimed at enhancing the efficiency and intelligence of urban planning. The platform utilizes Python to collect city-specific data, with data storage implemented using a MySQL database, complemented by NoSQL technologies to support the integration of unstructured data. The Flask backend employs deep learning and data mining algorithms to identify complex relationships among urban data, and we have also integrated graph neural network methods to capture spatial dependencies across different geographic regions within the city. Meanwhile, the ECharts frontend generates dynamic charts to present diverse information. Through a front-end and back-end separation architecture, the system ensures real-time updates, enhancing user experience. This research further emphasizes the need to explore challenges in real-time data integration, expanding data sources, optimizing user interactions, and protecting data privacy, providing important directions for future AI-driven urban planning.

Keywords: Urban Data Visualization; Big Data; AI-driven Urban Planning; Deep Learning

Introduction

Urban planning is a complex but essential field, and big data visualization plays a crucial role in it. By transforming raw data into intuitive visuals, urban planners can effectively process and analyze large datasets. This transformation is vital for identifying trends in urban development, such as population growth and traffic patterns, enabling data-driven decision-making (Bachechi, Po, & Rollo, 2022). For example, visualized traffic data can reveal congestion hotspots, leading to optimized management strategies. Additionally, visualization helps identify inefficiencies in resource distribution and infrastructure usage, allowing planners to target areas for improvement and allocate resources effectively(Kandt & Batty, 2021). Moreover, big data visualization uncovers opportunities for development, highlighting underutilized areas and guiding strategic planning for sustainable growth, such as identifying suitable regions for green spaces(Bibri, 2023).

Traditional methods of data analysis often fall short in handling the volume and variety of data generated in modern urban settings. Big data technologies bridge this gap by providing tools to analyze and visualize extensive datasets, leading to more informed decisions and strategic planning (Son et al., 2023). Moreover, visual representation of data makes it accessible to a broader audience, including policymakers, stakeholders, and the public. It enhances transparency and fosters a collaborative approach to urban development. Citizens can better understand and engage with urban planning initiatives, leading to more community-driven and accepted outcomes (Boeing, 2021). The current urban planning issues in China are as follows:

- Lack of Systematic Data Collection and Analysis Techniques: Many cities in China are at an early stage of data collection and processing, lacking professionals skilled in big data analysis. This results in incomplete data collection, limited analysis capabilities, and challenges in providing quality urban management and decision support.
- Insufficient Selection of Data Visualization Techniques: Effective data visualization is essential for conveying the spatial-temporal relationships of urban data. However, there are deficiencies in selecting appropriate visualization methods, necessitating better use of maps, charts, and tools to present complex data more intuitively for decision-makers and the public.
- Limited Data Analysis and Mining Capabilities: Many cities struggle to perform effective analysis and mining of urban data to identify patterns and insights. The lack of advanced technologies, such as data science and machine learning, restricts the extraction of valuable information from large datasets, hindering the scientific rigor of urban planning.
- Insufficient Consideration of User Needs and Interaction Design: Understanding the needs of diverse user groups (urban management, businesses, citizens) and designing user-friendly interfaces are crucial. Current applications often fall short in this aspect, lacking specificity and ease of use, and failing to meet users' actual needs effectively.

Integrating big data visualization into urban planning is vital for comprehending complex dynamics, optimizing resources, and promoting sustainable development. By converting data into actionable insights, it facilitates efficient urban growth that meets the needs of residents. Our team's work with big data and visualization technologies supports this approach, enhancing the livability and responsiveness of cities. The key innovations of this paper include:

In D. Barner, N.R. Bramley, A. Ruggeri and C.M. Walker (Eds.), *Proceedings of the 47th Annual Conference of the Cognitive Science Society*

- Multidimensional Data Display and Analysis: Visualizing data trends in temporal and spatial dimensions offers an intuitive analysis perspective at both global and local levels.
- Automated Data Updates: Implementing big data technology for automatic scheduled updates ensures data timeliness and accuracy, improving management efficiency.
- Diverse Analysis Scenarios: Analyzing combined data across fields such as population, economy, traffic, housing, and environment uncovers deeper relationships, supporting comprehensive urban management.
- Data-Driven Decision Making: Our approach leverages methodologies such as predictive analytics, machine learning algorithms, and spatial data analysis to conduct a comprehensive analysis of the data.

Related Work

Data Visualization Platforms

Data visualization platforms are increasingly vital in urban planning and management, providing comprehensive tools to comprehend and utilize complex urban data effectively (Daly, 2019)(LOS ANGELES, 2008). These platforms serve as essential resources for city officials and the public, offering real-time data and insights into urban dynamics. For instance, Singapore's Smart Nation Initiative has developed a comprehensive smart city platform that integrates various data sources and visualization tools to support informed decision-making in areas such as transportation, energy usage, and public safety (Shamsuzzoha, Nieminen, Piya, & Rutledge, 2021). Similarly, Barcelona's smart city platform utilizes IoT data and advanced visualization techniques to manage urban resources efficiently, contributing to sustainable urban development (Giffinger et al., 2007). The integration of advanced tools like Tableau, D3.js, and ArcGIS further enhances the capabilities of these platforms, allowing for the creation of interactive and dynamic visualizations that can handle large datasets effectively (van Dierendonck, van Tienhoven, & Elid, 2015) (Few & Principal, n.d.).

Advanced Data Analysis and Mining

Advanced data analysis and mining techniques are pivotal in extracting valuable insights from vast urban datasets, enabling informed decision-making and strategic planning. These techniques encompass a range of methods, including predictive analytics and machine learning algorithms, to uncover hidden patterns and trends in urban data. For instance, predictive analytics utilizes historical data to forecast future trends, aiding in traffic management and crime prevention (Patil, 2019) (D'Angelo, Payares, Adelfio, & Mateu, 2024) (Mohler, Short, Brantingham, Schoenberg, & Tita, 2011).

New York City and London exemplify the application of advanced data analysis and mining in urban settings. New York City utilizes machine learning algorithms to predict crime hotspots, enabling proactive law enforcement deployment (Mohler et al., 2011). Meanwhile, London employs data-driven modeling to monitor air quality, identifying pollution hotspots and implementing targeted interventions for public health improvement (Babu Saheer, Bhasy, Maktabdar, & Zarrin, 2022). These case studies demonstrate the efficacy of data-driven approaches in addressing urban challenges and enhancing livability.

Python and R provide extensive libraries for machine learning and data analysis, while Apache Spark enables fast and scalable data processing for large-scale urban datasets (McKinney, 2022) (Zaharia et al., 2016).

User-Centric Design and Interaction

Effective user-centric design begins with comprehensive user research and needs assessment. This involves gathering insights into the preferences, behaviors, and pain points of different user groups. For example, studies such as Shamsuzzoha et al. (Shamsuzzoha et al., 2021) have conducted user experience research to understand how citizens interact with smart city applications, informing the design of user-friendly interfaces and functionalities. User-centric platforms prioritize the development of interactive interfaces that facilitate intuitive navigation and seamless interaction. Features such as customizable dashboards, real-time data updates, and interactive maps enhance user engagement and empower users to explore and analyze urban data effectively. Incorporating feedback mechanisms allows users to provide input and shape the evolution of the platform over time (Hollands, 2015). Additionally, mobile applications leverage native features such as geolocation and push notifications to enhance user experience and engagement (Clark & Mayer, 2023).

Technical Realization

System Structure

The system architecture of this paper aims to achieve the acquisition, processing, analysis, and visualization of relevant data from various cities in Guangdong Province through a series of technologies and components. Figure 1 illustrates our system architecture. Below is a detailed introduction to each component of the system and its functions:

Data Acquisition and Storage. Our system uses Python web scraping to gather data from various cities in Guangdong Province, focusing on metrics like resident population, air quality index, GDP, and per capita consumption. The data undergoes cleaning and processing in Python to ensure quality by removing duplicates, addressing missing values, and standardizing formats. We store structured data in a MySQL database and use NoSQL for unstructured data, creating a dual-database system that enhances reliability, flexibility, and scalability. This integration provides a strong foundation for high-quality data analysis and visualization.

Figure 1: System framework diagram.

Data Mining and Analysis. Our system utilizes advanced deep learning and data mining algorithms to analyze complex relationships within urban data, uncovering correlations among diverse datasets. We employ correlation coefficient algorithms to assess associations and integrated graph neural network methods to capture spatial dependencies across the city. These insights allow us to construct analytical graphs that visually represent connections and trends, offering valuable support for informed urban planning.

Backend Architecture and Data Interfaces. Our system employs a front-end and back-end separation architecture. The back-end, built on the Flask framework, provides data interfaces, processes requests, and retrieves data from the MySQL and NoSQL databases in JSON format. It includes functionalities such as querying, filtering, and pagination to support the front-end. This architecture enhances project maintenance, scalability, and development efficiency.

Data Visualization and Frontend Platform. Our system uses ECharts for data visualization, displaying processed information from the backend as charts on the frontend. The frontend retrieves data from backend interfaces, generating various visualizations for metrics like resident population, Air Quality Index (AQI), gross domestic product (GDP), and per capita consumption in Guangdong Province. These charts aid urban planning by helping managers monitor development dynamics.

Built on Flask, the big data backend accesses sources from MySQL and NoSQL databases and employs correlation coefficient algorithms to analyze urban data relationships. This integrated front-end and back-end system, with a separation architecture, forms a real-time updating big data platform.

Front-end Visual Design

ECharts is a JavaScript library for creating intuitive and interactive data visualization charts. The design process for urban data visualization involves four steps: data analysis, chart matching, chart optimization, and testing. First, determine the data to display, which includes metadata and user expectations, requiring backend data processing and interfaces. Next, use ECharts for chart matching and optimization to best represent data characteristics. Finally, conduct testing to ensure accuracy and effectiveness before deploying the frontend and backend on the server to launch the urban data visualization platform.

System Layer Module

The urban data visualization platform mainly consists of the data acquisition layer, data processing layer, data storage layer, data interface layer, and frontend display layer.

Data Acquisition Layer: We use Python web scraping with Requests and LXML libraries to gather data from various sources. This data is stored in MySQL and NoSQL databases before being sent to the data processing layer for refinement.

Data Processing Layer: In the data processing stage, we use Python's Pandas library to clean and transform the acquired data. This includes converting data formats, handling missing values, detecting outliers, and ensuring data quality before sending the processed data to the storage layer.

Data Storage Layer: We store processed data in MySQL and NoSQL databases, organized according to predefined data models for system use.

Data Interface Layer: We use the Flask framework to create data interfaces that retrieve and present data from the storage layer for frontend use. Flask's compatibility and scalability, along with its Jinja2 template engine, enhance frontend code reusability and improve development efficiency.

Frontend Display Layer: The frontend uses Ajax to call data interfaces and retrieve data, utilizing ECharts for visual representations that enhance user understanding and analysis.

Frontend UI Layer: We design the frontend interface with Bootstrap and jQuery for aesthetic appeal, responsiveness, and interactivity, maximizing user experience.

City data visualization platform function display

Province overview

To present data intuitively, we divided the provincial overview page of the urban data visualization platform into ten sections focusing on specific metrics, such as operating vehicles for public transport, passenger volumes, gender ratios, GDP of Guangdong Province, AQI compliance rates, per capita consumption, and population statistics. The page features navigation buttons for the overview, city statistics, and map interfaces, along with a real-time clock. For Guangdong, we use maps and various charts: a pie chart for the male-to-female ratio, a ring chart for AQI rates, a horizontal bar chart for GDP, a line chart for per capita expenditure, a 3D bar chart for annual population, and a multi-bar chart for average housing prices. This is illustrated in Figure 2.

Statistics by city

We have meticulously structured the city statistics page of our urban data visualization platform into six comprehensive sections, each focusing on key metrics essential for understanding urban dynamics. These sections encompass the following facets: the resident population of each city, the GDP of each city, AQI compliance rates of each city, the top five cities in terms of commodity housing sales, the top five cities in terms of per capita consumption levels, and a national map display. Consistent with the intuitive design principles applied throughout the platform, the navigation bar for city statistics maintains coherence with that of the provincial overview, ensuring a seamless user experience. Within the city statistics graphs, we employ a variety of visualization techniques to effectively convey data insights. For instance, the resident population of each city is presented dynamically through interactive bar charts, enabling users to explore population trends with ease. Similarly, AQI compliance rates are depicted using visually intuitive pie charts, offering a clear representation of air quality standards across cities. The GDP of each city within the province is visualized using comprehensive multi-line graphs, allowing users to discern economic trends over time. To highlight the top performers in commodity housing sales, we utilize pictorial bar charts, offering a visual snapshot of the real estate market landscape. Additionally, funnel charts are employed to illustrate the top five cities for per capita consumption levels, providing users with a comparative view of consumption patterns. Leveraging the capabilities of the Baidu Map API interface, our platform seamlessly integrates city-specific data for visualization, enhancing the depth and breadth of urban insights available to users. This comprehensive approach to city statistics visualization is demonstrated in Figure3, exemplifying our commitment to providing a user-centric and informative urban data visualization experience.

Map display

We have structured the map display page of our urban data visualization platform into two distinct sections. The advantage of structuring the map display page into two sections lies in its user-centric design and interactive features. By dividing the page, users can easily navigate between different types of information, enhancing the platform's usability. In line with the cohesive design approach across the platform, the navigation bar maintains consistency with the provincial overview and city statistics pages. Leveraging the Baidu Map API interface, our platform seamlessly integrates city-specific data for visualization. Users can interact with the map by clicking on red dots, enabling them to access detailed information regarding population demographics, GDP figures, AQI compliance rates, and consumption levels for each city. Moreover, on the right side of the page, users have the option to select specific parameters such as resident population, GDP, and disposable income, allowing them to dynamically scroll through and explore corresponding data trends from 2010 to 2022, as depicted in Figure 4. This interactive and user-centric design empowers users to delve deeper into the nuances of urban data, facilitating comprehensive analysis and informed decision-making.

Discussion

Distinctions from Other Works

Our urban data visualization platform distinguishes itself from other similar works in several key areas:

- Regional Focus: Unlike many urban data visualization projects that provide generic solutions applicable to a wide range of cities, our platform is specifically tailored for the cities in Guangdong Province. This regional focus allows for more accurate and relevant data collection, processing, and analysis, providing insights that are directly applicable to local urban planning and management needs.

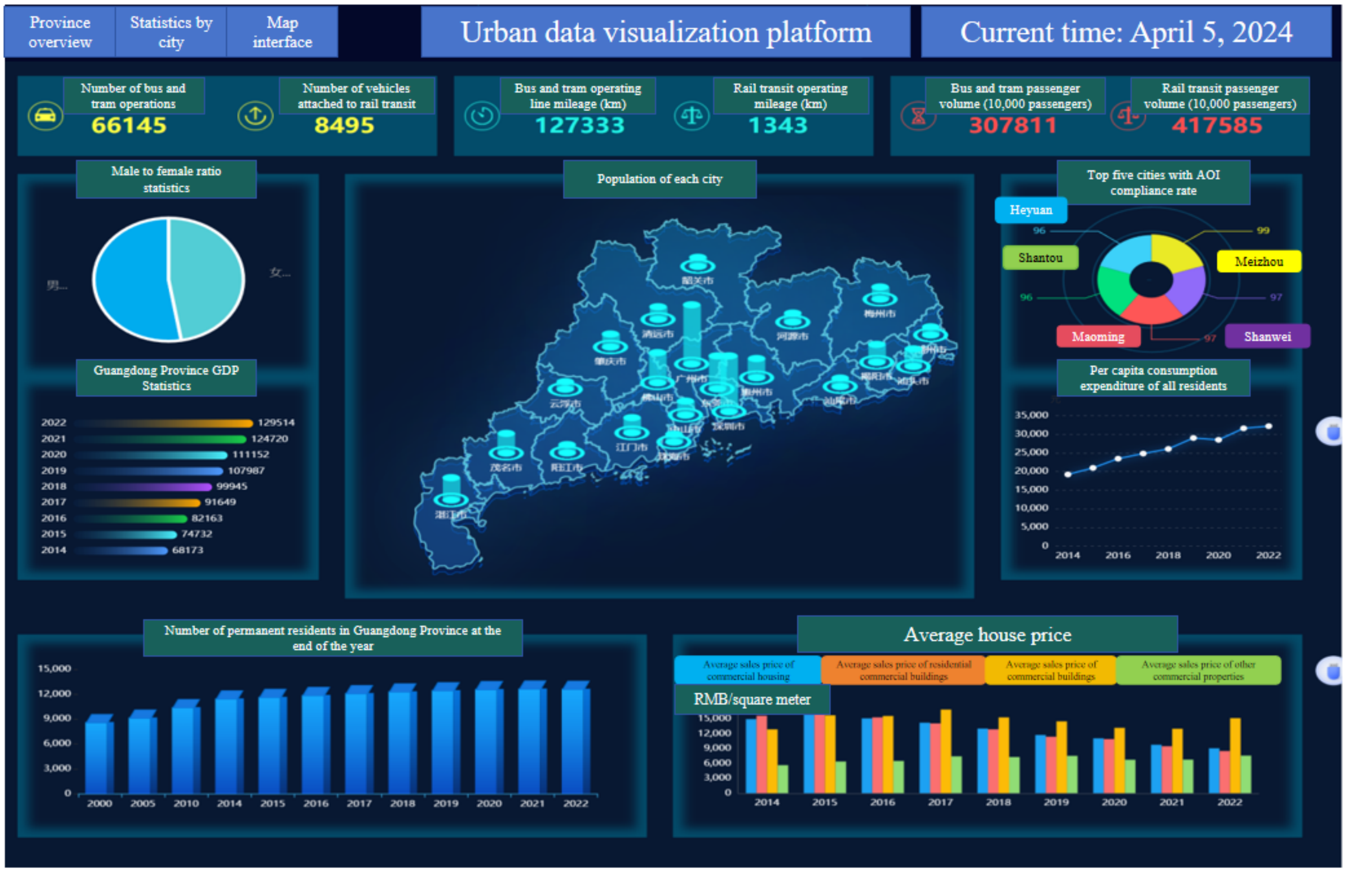

Figure 2: Overview map of Guangdong Province.

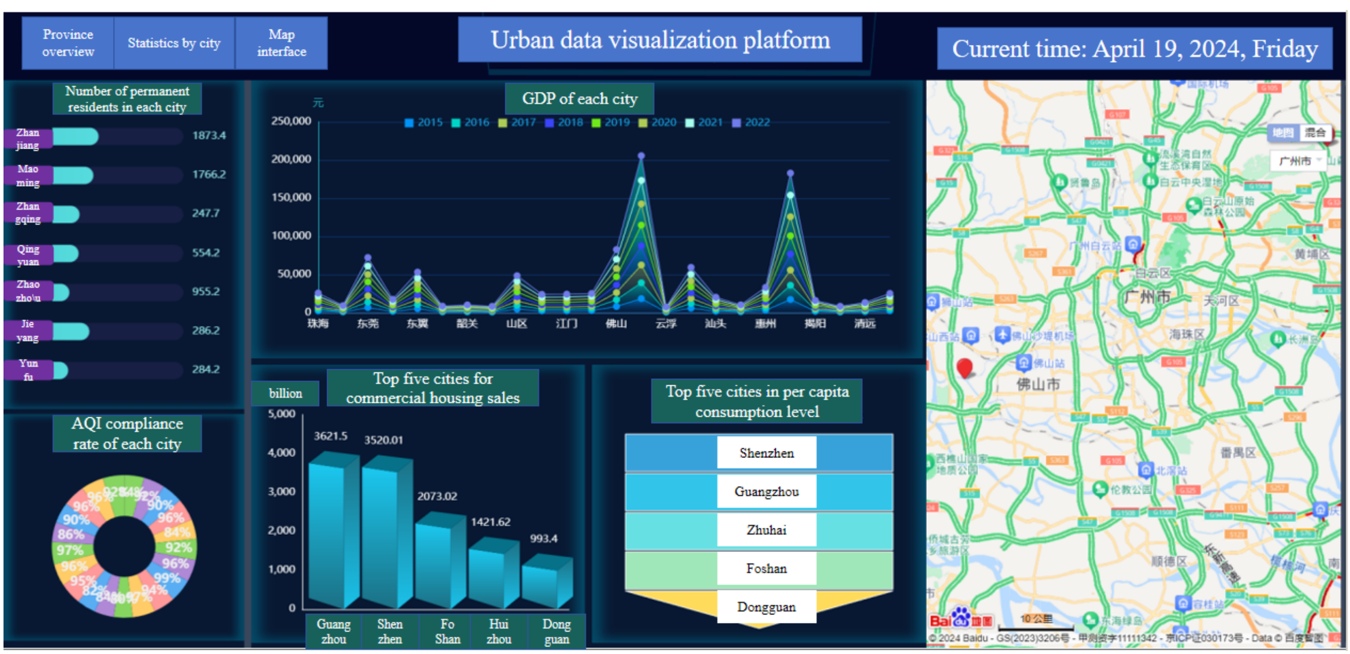

Figure 3: Statistics display chart of cities in Guangdong Province.

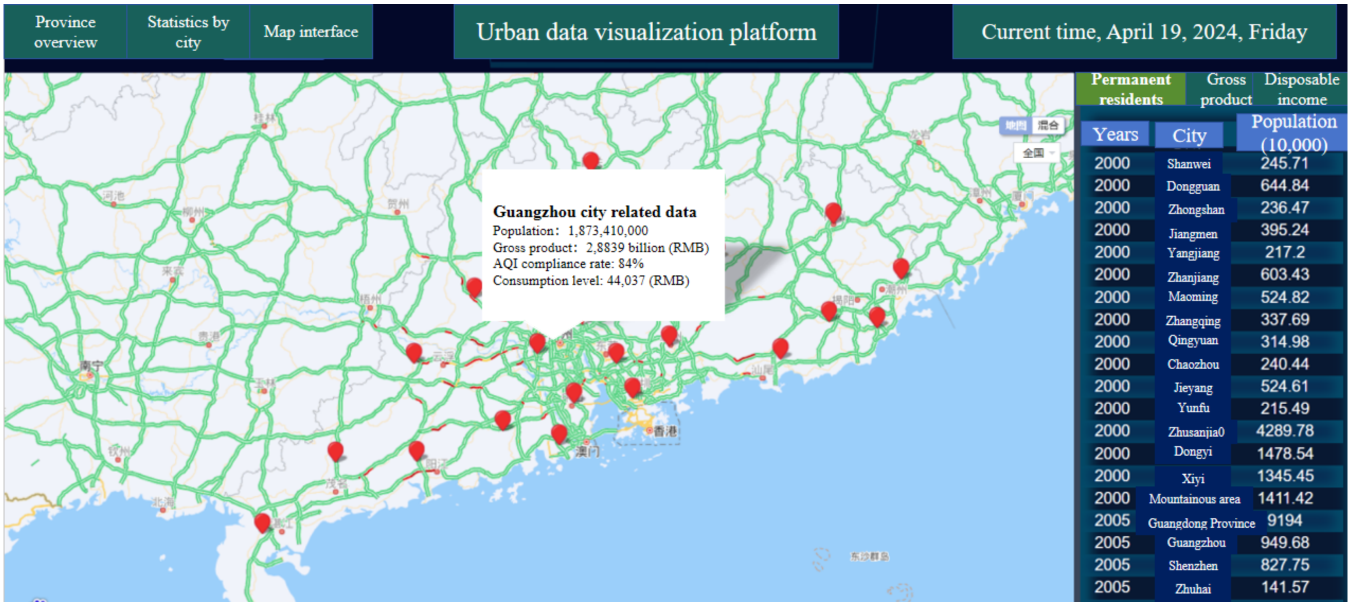

Figure 4: Map interface display.

- Comprehensive Data Integration: Our platform integrates a wide variety of urban data metrics, including population statistics, GDP, AQI compliance rates, housing prices, and consumption levels. This comprehensive approach enables a multi-dimensional analysis of urban data, offering a more holistic view of the city's dynamics compared to other platforms that may focus on a narrower set of data points.

- User-Centric Design: The platform places a strong emphasis on user experience, with a user-friendly interface and interactive features that cater to the needs of various stakeholders, including city planners, policymakers, and the general public. By employing techniques such as front-end and back-end separation and utilizing frameworks like Flask and Bootstrap, we ensure that the platform is both easy to use and highly responsive.

- Scalability and Flexibility: The architecture of our platform is designed to be scalable and flexible, accommodating future expansions and updates. This includes the ability to integrate additional data sources and analytical tools as they become available.

These distinctions highlight our commitment to providing a specialized, comprehensive, and user-friendly urban data visualization solution for Guangdong Province. However, several challenges remain, including the lack of real-time data integration due to scheduled updates, which can delay access to current information. Future efforts should prioritize incorporating real-time data streams.

While key metrics are in place, adding data on social behavior, transport efficiency, and environmental impact would deepen insights into urban dynamics. Enhancing the platform with predictive analytics, machine learning, and interactive features—like drag-and-drop, customizable dashboards, and smart filtering—would improve usability. Robust data privacy measures, including encryption and secure access, are also essential.

Conclusion

The urban data visualization platform we developed demonstrates the power of advanced technologies for urban data analysis. It efficiently handles large datasets and presents them in an accessible format. Our platform tackles key challenges in urban data visualization, including high-quality data acquisition, effective processing and storage, a scalable backend architecture, and engaging frontend visualizations. By organizing the platform into sections—provincial overview, city statistics, and map display—we ensure easy navigation and valuable insights. Looking ahead, the platform's real-time update capability and scalability will adapt to the evolving data needs of urban environments. By addressing current limitations and exploring new research directions, we aim to continually enhance the platform's functionality and impact.

Acknowledgments

The study is supported by an internal project of the National Engineering Laboratory for Big Data System Computing Technology, Shenzhen University, Shenzhen 518060, PR China (No. SZU-BDSC-IF2024-XX); Supported by the Natural Science project of Guangdong University of Science and Techmology "Comprehensive knowledge fusion and dynamic decision-making: multi-source data collaboration project" (Grant no: GKY-2023BSQD-43); Supported by Natural Science Project of Guangdong University of Science and Technology (Grant No. GKY-2023BSQD-42).

References

Babu Saheer, L., Bhasy, A., Maktabdar, M., & Zarrin, J. (2022). Data-driven framework for understanding and predicting air quality in urban areas. *Frontiers in big Data*, *5*, 822573.

Bachechi, C., Po, L., & Rollo, F. (2022). Big data analytics and visualization in traffic monitoring. *Big Data Research*, *27*, 100292.

Bibri, S. E. (2023). Data-driven smart eco-cities of the future: an empirically informed integrated model for strategic sustainable urban development. *World Futures*, *79*(7-8), 703–746.

Boeing, G. (2021). Spatial information and the legibility of urban form: Big data in urban morphology. *International Journal of Information Management*, *56*, 102013.

Clark, R. C., & Mayer, R. E. (2023). *E-learning and the science of instruction: Proven guidelines for consumers and designers of multimedia learning*. john Wiley & sons.

Daly, R. (2019). *Boston*. Weigl Publishers.

D'Angelo, N., Payares, D., Adelfio, G., & Mateu, J. (2024). Self-exciting point process modelling of crimes on linear networks. *Statistical Modelling*, *24*(2), 139–168.

Few, S., & Principal, P. E. (n.d.). Designing tables and graphs to enlighten.

Giffinger, R., Fertner, C., Kramar, H., Kalasek, R., Pichler-Milanovic, N., & Meijers, E. J. (2007). Smart cities. ranking of european medium-sized cities. final report.

Hollands, R. G. (2015). Critical interventions into the corporate smart city. *Cambridge journal of regions, economy and society*, *8*(1), 61–77.

Kandt, J., & Batty, M. (2021). Smart cities, big data and urban policy: Towards urban analytics for the long run. *Cities*, *109*, 102992.

LOS ANGELES, C. O. (2008). City of los angeles. *Policy*, *14*, 28.

McKinney, W. (2022). *Python for data analysis*. " O'Reilly Media, Inc.".

Mohler, G. O., Short, M. B., Brantingham, P. J., Schoenberg, F. P., & Tita, G. E. (2011). Self-exciting point process modeling of crime. *Journal of the american statistical association*, *106*(493), 100–108.

Patil, P. (2019). Machine learning for traffic management in large-scale urban networks: A review. *Sage Science Review of Applied Machine Learning*, *2*(2), 24–36.

Shamsuzzoha, A., Nieminen, J., Piya, S., & Rutledge, K. (2021). Smart city for sustainable environment: A comparison of participatory strategies from helsinki, singapore and london. *Cities*, *114*, 103194.

Son, T. H., Weedon, Z., Yigitcanlar, T., Sanchez, T., Corchado, J. M., & Mehmood, R. (2023). Algorithmic urban planning for smart and sustainable development: Systematic review of the literature. *Sustainable Cities and Society*, 104562.

van Dierendonck, R., van Tienhoven, S., & Elid, T. (2015). D3. js: Data-driven documents. *JS: Data-Driven Documents*.

Zaharia, M., Xin, R. S., Wendell, P., Das, T., Armbrust, M., Dave, A., ... others (2016). Apache spark: a unified engine for big data processing. *Communications of the ACM*, *59*(11), 56–65.