University of Tehran

Earth Observation and Geomatics Engineering

Online ISSN: 2588-4360

Homepage: https://eoge.ut.ac.ir/

Enhancing Urban Traffic Flow Understanding Through Online Data Analysis and Modeling

Case Study; A western route in Tehran

Zahra Rezaei[1], **Hossein Aghamohammadi** [2✉], and **Mohammad H. Vahidnia**[3], and **Zahra Azizi**[4], and **Saeed Behzadi**[5]

1. Department of Remote Sensing and GIS, Science and Research Branch, Islamic Azad University, Tehran, Iran.
E-mail: Rezaee.ncc@gmail.com
2. Corresponding author, Department of Remote Sensing and GIS, Science and Research Branch, Islamic Azad University, Tehran, Iran.
E-mail: hossein.aghamohammadi@gmail.com
3. Center for Remote Sensing and GIS Research, Faculty of Earth Sciences, Shahid Beheshti University, Tehran, Iran.
E-mail: Vahidnia84@gmail.com
4. Department of Remote Sensing and GIS, Science and Research Branch, Islamic Azad University, Tehran, Iran.
E-mail: zsazizi@yahoo.com
5. Department of Surveying Engineering, Faculty of Civil, Water and Environmental Engineering, Shahid Beheshti University, Tehran, Iran. E-mail: behzadi@sru.ac.ir

Article Info

Article type:
Research Article

Article history:
Received 2025-02-22
Received in revised form 2025-03-25
Accepted 2025-03-27
Available online 13 May 2025

Keywords:
urban resilience,
traffic,
traffic data,
time series,
modelling

ABSTRACT

Effective urban traffic management relies on a thorough understanding of traffic behavior patterns. Traditional methods often struggle to capture the dynamic and complex nature of modern traffic. This research addresses this challenge by utilizing online traffic data from the Mapbox platform to analyze and forecast traffic behavior patterns in Tehran. Autoregressive Integrated Moving Average (ARIMA) and Seasonal Autoregressive Integrated Moving Average (SARIMA) time series models were employed to analyze traffic volume, vehicle speed, and travel time data. The models were evaluated using criteria such as the Akaike Information Criterion (AIC) and Root Mean Square Error (RMSE). The results indicate that these models can accurately simulate both temporal and periodic trends. Specifically, the Moving Average (MA) coefficient (ma.L1) shows a positive and significant impact of the first lag (p-value = 0.000). The analysis reveals that the Autoregressive (AR) coefficient (ar.L1) is -0.0270, indicating a negative impact of the first lag; however, a p-value of 0.588 rejects the significance of this impact. On the other hand, the Moving Average (MA) coefficient (ma.L1) is 0.2028, showing a positive and significant impact of the first lag (p-value = 0.000). Furthermore, the AIC and BIC criteria are -18070.697 and -18036.730, respectively. The study's interpretation of negative values indicates a robust model fit and no extra parameters needed. Additionally, this research analyzed traffic behavior in Tehran by examining modeling results across different hours to capture variations in traffic patterns. The study highlights the effectiveness of using online traffic data and time series modeling to identify factors influencing traffic, improve urban traffic management, and support transportation planning. Overall, it promotes detailed traffic analysis and behavior prediction, contributing to intelligent and sustainable transportation systems, enhanced urban resilience, and improved infrastructure and traffic flow.

Cite this article: Rezaee, Z., Aghamohammadi, H., Vahidnia, M. H., azizi, Z., Behzadi, S. (2025). Enhancing Urban Traffic Flow Understanding Through Online Data Analysis and Modeling Case Study; A western route in Tehran. *Earth Observation and Geomatics Engineering*, Volume 8 (Issue 1), Pages 98-112. http//doi.org/10.22059/EOGE.2025.390975.1169

Publisher: University of Tehran.
DOI: http//doi.org/10.22059/EOGE.2025.390975.1169

1. Introduction

The increasing complexity of urban environments, driven by population growth, environmental crises, and disruptions, highlights the urgent need for resilient urban planning, particularly in transportation systems. Cities face multifaceted challenges such as traffic congestion, which directly impacts functionality, air quality, and quality of life. Addressing these challenges requires a holistic approach that integrates advanced traffic data analysis, behavioral prediction models, and adaptive infrastructure design. Such an approach not only enhances traffic coordination but also contributes to broader goals of sustainability and resilience reducing emissions, improving safety, and ensuring reliable access to urban services. These outcomes address environmental crises by reducing greenhouse gas emissions and improving air quality, while improving quality of life by minimizing delays, reducing stress, and ensuring reliable access to services. Ultimately, urban resilience in transportation planning serves as an essential tool to address contemporary challenges and crises, paving the way for smarter, more sustainable cities (Datola, 2023; Dianat et al., 2022). Urban resilience in the transportation domain means the ability of the transportation system to effectively respond to challenges and unexpected changes, such as accidents or climate change, and even to optimally manage urban traffic (Chen et al., 2022).

Online traffic data is of fundamental importance to the functioning of resilient urban transportation systems. This data, collected from various sources such as sensors, GPS devices, and smartphones, offers a unique opportunity to analyze traffic patterns in real time. By examining this data, cities can make informed decisions to mitigate potential disruptions and respond proactively to crises, including accidents and extreme weather events (Chen et al., 2022). The analysis of this data can provide a deeper understanding of traffic patterns, enabling informed decisions to improve traffic flow, reduce accidents, and enable urban planners to manage traffic effectively during peak hours and special events (Datola, 2023; N. Moghaddam et al., 2021). Researchers have explored various applications of traffic data in their studies, including specific applications in traffic light automation and traffic flow management. Furthermore, a comparative analysis of traffic data before and after the implementation of improvement plans facilitates a precise assessment of their effectiveness (N. Moghaddam et al., 2021), and these data facilitate the optimization of public transportation systems and the identification of areas susceptible to accidents (Chen et al., 2022).

In addition to gathering traffic data, machine learning algorithms can be used to develop traffic prediction models to help engineers predict future traffic and take necessary actions to improve traffic flow (Boukerche and Wang, 2020). However, collecting, processing, and analyzing large datasets demand robust infrastructure and advanced tools, alongside addressing privacy and security concerns. Traffic data studies have evolved significantly over the decades; while initial data was gathered by sensors and analyzed manually, modern data processing algorithms allow for more accurate automatic analyses, particularly with the influx of data from smartphones and GPS systems (Boukerche and Wang, 2020). Recent research has focused on various topics including short-term and long-term traffic prediction, identification of traffic patterns, and developing navigation systems based on traffic data. These analyses have demonstrated substantial contributions to reducing travel times, enhancing safety, and improving public transportation efficiency. Specific studies have investigated optimizing traffic signals, the traffic flow control, and designing transportation networks using these data sources (Karami and Kashef, 2020; Wu et al., 2023).

This study focuses on the application of online traffic data and machine learning algorithms in traffic modelling and using it in urban transportation planning. The significance of this research lies in its potential to improve traffic prediction, and boost urban efficiency, thereby contributing to urban traffic management and intelligent transportation systems. Traffic modeling and time series construction from traffic data are essential for analyzing complex traffic behaviors. Traffic data, including volume, vehicle speed, and travel time, helps in understanding patterns and predicting future conditions. Autoregressive Integrated Moving Average (ARIMA) models are effective for studying traffic behavior due to their capability to analyze and forecast non-stationary time series by incorporating autoregressive, differencing, and moving average components to identify temporal dependencies and periodic patterns. ARIMA models can account for various influencing factors, such as time of day, day of the week, and weather conditions. Additionally, Seasonal ARIMA (SARIMA) models are better suited for periodic data, effectively simulating fluctuations in traffic datasets. The use of these models enhances prediction accuracy and reveals underlying trends and patterns, providing valuable insights for traffic management and urban planning decisions.

This section examines similar studies related to traffic data, highlighting existing gaps, shortcomings in this field, and the need for new research directions. Yang et al. (2021) introduced a network traffic prediction method that integrates Simulated Annealing (SA) optimization with ARIMA models and Backpropagation Neural Networks (BP). Their hybrid framework combines the predictive capabilities of ARIMA with the learning advantages of neural networks, demonstrating significant improvements in prediction accuracy, which aids network managers in resource planning (Yang et al., 2021). Wang et al. (2022) proposed an ARIMA model integrated with multi-instance learning for analyzing and predicting vehicle speed time series. This research emphasizes the combination of traditional statistical models with machine learning techniques to enhance prediction accuracy, providing valuable insights for traffic management and transportation planning (Wang et al., 2022). Additionally, Nassiri et al. (2023) examined temporal trends in road accidents in Iran, utilizing macro-scale traffic flow characteristics to identify patterns influencing road safety. Their findings could inform policymakers and support preventive measures in traffic planning (Nassiri et al., 2023). Deretić et al. (2022) applied a SARIMA modeling approach to predict traffic accidents, analyzing historical data to forecast trends in accident occurrences. Their results are expected to improve traffic management and safety policies, providing a framework for reducing accidents and injuries in sustainable transportation (Deretić et al., 2022). Another study by Wang et al. (2022) introduced a traffic flow prediction method based on periodic features and the SARIMA-NAR model, which enhances prediction accuracy through the analysis of periodic patterns. The findings support traffic managers and urban planners in optimizing transportation systems (Wang et al., 2022). Rabbani et al. (2021) conducted a comparative analysis between the SARIMA model and Exponential Smoothing (ES) for predicting road accidents, highlighting the distinct capabilities of each method in analyzing time series data related to accidents. Their results could aid in developing enhanced roadway safety strategies (Rabbani et al., 2021). Furthermore, Patil (2022) evaluated various time series forecasting methods for urban traffic flow prediction, demonstrating that the choice of forecasting technique significantly influences prediction accuracy, thereby aiding traffic management and reducing congestion in urban settings (Patil, 2022). In another research has also shown a growing emphasis on LSTM, ARIMA, and SARIMA models in Intelligent Transportation Systems (ITS). Verma and Pandey (2024) established ARIMA and SARIMA as conventional statistical forecasting methods, while LSTM captures non-linear dependencies in traffic data. Their comparison using a dataset from the U.S. Government Open Repository highlighted seasonality's role in modeling traffic data accurately, with SARIMA demonstrating superior performance (Verma and Pandey, 2024).

Existing studies highlight the significance of using ARIMA and SARIMA models for modeling time series traffic data due to their effectiveness in enhancing prediction accuracy and managing traffic data volatility. The SARIMA model, in particular, excels at identifying periodic patterns and accommodating seasonal fluctuations, which improves prediction reliability and supports better transportation infrastructure management, potentially reducing congestion and costs (Patil, 2022). The integration of ARIMA and SARIMA models with advanced techniques such as machine learning can yield optimal results, with SARIMA effectively serving as an input for complex models like neural networks (Rabbani et al., 2021; Patil, 2022). Collectively, these studies demonstrate the need for sophisticated modeling approaches to improve traffic prediction and safety, emphasizing the importance of incorporating emerging technologies like IoT and smart sensors for real-time decision-making in traffic management. Overall, employing these models not only enhances prediction accuracy and optimizes traffic flow but also benefits transportation infrastructure and alleviates traffic-related issues (Oladimeji et al., 2023). However, a comprehensive travel time modeling study for Tehran that specifically incorporates key traffic influencing parameters such as weather conditions, accidents, and the impact of historical events on traffic volume (and consequently vehicle speeds and congestion) remains largely unexplored. While the existing literature addresses traffic modeling in various contexts, the unique characteristics of Tehran's traffic patterns influenced by these specific factors require a more tailored approach.

This study addresses existing gaps by utilizing high-resolution, real-time traffic data from the Mapbox platform, incorporating external variables to enhance the modeling process in a specific area of Tehran. The research primarily innovates by employing both ARIMA and SARIMA models to analyze and predict traffic patterns, capitalizing on their strengths in time series analysis to capture autocorrelation and model seasonality. Furthermore, by integrating external factors such as weather conditions, holidays, and accidents alongside real-time data from online traffic services and other sources, the study offers a comprehensive understanding of traffic dynamics in Tehran while utilizing machine learning techniques to assess traffic patterns and predict future conditions, particularly focusing on elements like travel time and evaluating the impact of different conditions on it. This research aims to develop a robust framework for short-term traffic forecasting using ARIMA and SARIMA models to produce accurate travel time prediction models. By providing a detailed understanding of Tehran's unique travel time patterns and comparing the effectiveness of ARIMA and SARIMA models, the study seeks to determine the best modeling approach, ultimately leading to innovative traffic prediction strategies that optimize travel time and reduce overall traffic levels, resulting in reduced traffic. Additionally, while this research focuses on the developing accurate travel time prediction models, its ultimate goal is to provide urban planners and traffic engineers with the tools necessary to make informed decisions that improve traffic flow and reduce congestion.

2. Materials and Methods

This study employs innovative online traffic service sources to collect traffic data in a time series format, enabling a comprehensive analysis of traffic behaviors. The gathered data encompasses critical information such as traffic volume, vehicle speed, and travel time across specific intervals. Following data collection, appropriate storage and preprocessing methods are applied to ensure the quality and accuracy of the data for subsequent analyses. To model the traffic data, statistical models ARIMA and SARIMA are utilized due to their effectiveness in identifying temporal and periodic patterns, particularly in contexts where the data exhibit volatility and non-linear trends. Initially, stationarity tests, including the Dickey-Fuller test, are performed to ascertain whether the data is stationary or requires differencing. Once the stationarity of the data is established, optimal parameters for the models are determined using techniques such as the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF). After fitting the models to the traffic data, the performance of these models is evaluated using various metrics, including Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Akaike Information Criterion (AIC). This systematic evaluation facilitates the selection of the most suitable model for predicting traffic behaviors.

The results derived from the modeling and forecasting processes using ARIMA and SARIMA models are then presented analytically. These analyses include the identification of behavioral traffic patterns, an examination of the impacts of various factors on traffic, and actionable recommendations for improving traffic management in different urban areas. The overarching goal of this research is to provide deeper insights into traffic patterns, thereby assisting urban planners in making more effective decisions in traffic management. By integrating rigorous data collection and preprocessing methods, applying advanced statistical models, conducting thorough stationarity testing, and evaluating model performance, this study significantly contributes to the field of traffic management. It offers valuable insights into traffic behaviors and their influencing factors, ultimately supporting urban planners in developing more effective and sustainable traffic management strategies. Through these efforts, the research aims to enhance decision-making processes, contributing to sustainable urban mobility and improved infrastructure management.

This study makes a significant contribution to the field of traffic management by offering a novel approach to better understanding and forecasting traffic patterns. Ultimately, this approach can result in more effective strategies for optimizing traffic flow in urban environments.

This section presents a flowchart as show in Figure 1, to visually illustrate the sequential steps of the research. It outlines the key phases, starting from the research goal and proceeding through the selection of the study area, the establishment of an online traffic data collection system, data storage and preprocessing, exploration of machine learning methods, and the implementation of ARIMA and SARIMA modeling techniques. The flowchart culminates in the evaluation of results and the discussion of conclusions and recommendations, providing a clearer understanding of the research methodology.

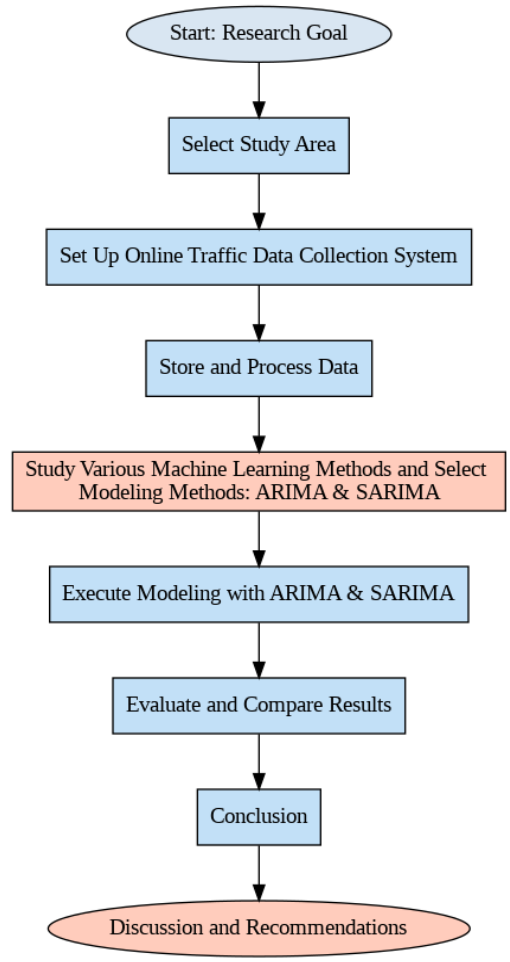

Figure 1- The flowchart of research processes

2.1. Study Area

In this section, a detailed examination of the study area is provided, highlighting its significance as a congested traffic area in Tehran. Tehran, as one of the largest and most populous cities in Iran, exhibits significant spatial variability in traffic patterns and environmental conditions due to its vast geographical expanse and even diverse weather conditions. Given the complexity of Tehran's traffic dynamics, which vary considerably between different districts, this study focuses on a route from District 5 of Tehran to the vicinity of Azadi Square, a significant location in the city, as shown in Figure 2, as a pilot area for detailed analysis. This route, which extends from a point in District 5 near a selected bus station on Hakim Highway to a selected metro station in Azadi Square. The route was selected as a route with a variety of characteristics, including major roads, intersections, highways, and the congested area near the metro stations, etc., to represent typical urban traffic challenges in Tehran, while ensuring the feasibility of real-time data collection and monitoring. By focusing on this route, the study aims to provide a manageable yet representative case for analyzing traffic behavior, allowing for accurate observations and collection of high-quality data. This approach ensures that the results can be used as a basis for broader applications in other parts of the city in future studies.

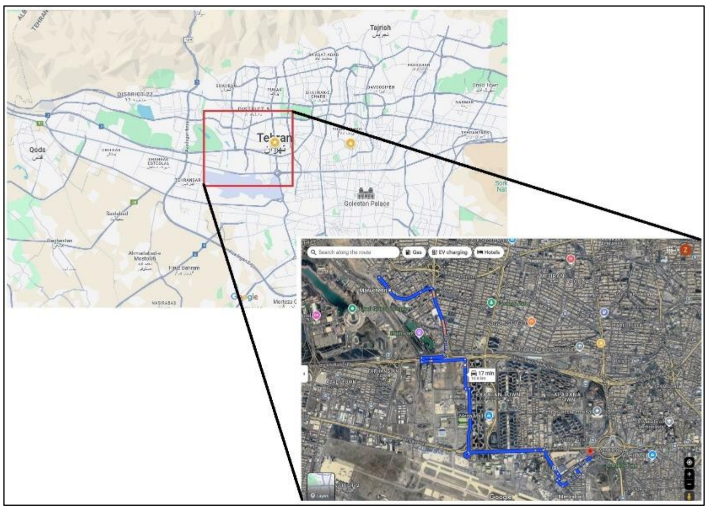

Figure 2 – The Study Area

This specific route was chosen to analyze traffic patterns and conditions that are representative of urban transportation challenges in Tehran. The required data was gathered through the API of this platform over the period from April 2, 2024, to June 4, 2024, resulting in the collection of a time series dataset comprising approximately 6,500 records. While the two-month timeframe may seem limited, it was chosen to capture high-resolution data (e.g., every 5–10 minutes) sufficient for training and validating ARIMA and SARIMA models. The dataset's granularity and size (6,500 records) provide a strong basis for short-term traffic forecasting, the study's primary focus. However, longer-term data collection would enhance the capture of broader seasonal trends and improve model generalizability.

3. Data Sources and Collection Methods

To collect traffic data, this study utilized online mapping platforms such as Mapbox. The developed system allows for real-time monitoring and recording of online traffic data at any moment. Figures 3 illustrate two examples of the system's functionality in collecting and recording real-time data at various dates and times. This platform offers access to online and real-time traffic data, encompassing information on vehicle speeds, traffic conditions, and road incidents.

The study acknowledges that the two-month period falls within a single season (02-04-2024) to (04-0٦-2024), (in Persian date; since 1403/01/14 to 1403/03/14) limiting observable seasonal variations. While the SARIMA model accounts for daily cycles, the analysis focuses on short-term periodic traffic behavior, such as daily and weekly fluctuations, crucial for real-time traffic management. Future work could extend data collection to multiple seasons to better explore seasonal impacts on traffic patterns. The system was developed using JavaScript, which facilitated the integration with Mapbox to retrieve online traffic data seamlessly. This approach allowed for efficient data handling and real-time updates, ensuring that the traffic information collected was both current and relevant for analysis.

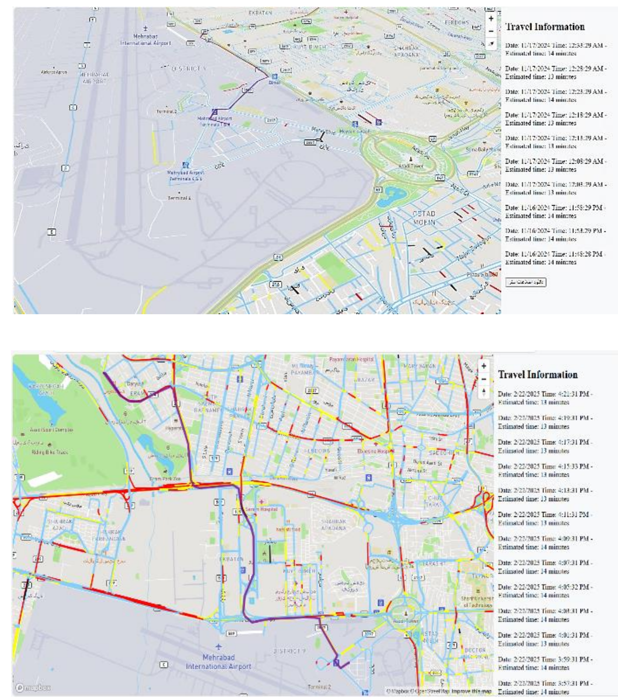

Figure 3 – the environment of the Developed System with Mapbox online Services with functionality in various time

To facilitate the retrieval of traffic data, an account was created on Mapbox, and an API key was obtained to access its online traffic services. Specific points on the road network were selected, focusing on main roads and key traffic intersections based on traffic patterns and the significance of these locations. To automate data collection, scripts were developed using Python, along with the Requests library, which periodically sent requests to the Mapbox API at ten-minute intervals. This methodology allowed for efficient acquisition of traffic information, which was then stored as a time series in either a local or cloud-based SQL database. The collected data includes several parameters that influence traffic, such as timestamps, geographic locations, traffic volume status (Traffic Condition Factor), and records of road incidents (e.g., accidents). Table 1 illustrates a representative sample of this traffic data, showcasing records from various time intervals and days. This diversity in recorded data is crucial for accurately representing how traffic information is distributed within the dataset. The structure of this data included several parameters that influence traffic, such as time, geographic locations, and traffic volume status (Traffic Condition Factor), etc.

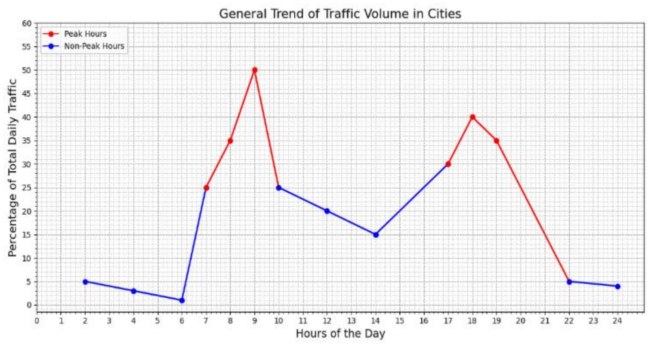

Figure 4 – The general graph of Traffic Volume in Tehran and big cities

Within the "time" parameter, various dates and types of days (such as holidays, weekdays, and special events) were considered, as these can significantly influence traffic patterns in Tehran. An example of a traffic behavioral algorithm developed for Tehran is the “Traffic Peak Factor Algorithm”. This algorithm analyzes historical traffic data to identify distinct temporal patterns prevalent under normal traffic conditions. For instance, and as it shown in Figure 4, the algorithm may determine that traffic congestion tends to peak on weekdays during the morning rush hour (7:00 AM to 9:00 AM) and again in the evening (5:00 PM to 7:00 PM). It takes into account external factors such as the occurrence of local events (e.g., concerts, sports events) that may exacerbate traffic volume on specific days. The traffic peak factor is calculated by analyzing data points during these times and establishing a ratio that represents traffic flow relative to off-peak hours. By incorporating this factor into each data record, the algorithm can more precisely monitor and predict traffic flow. For example, if historical data shows that traffic volume increases by 50% on weekday mornings compared to weekends, this insight can be utilized to adjust traffic signal timings or inform commuters about expected delays during those peak times. Thus, the Traffic Peak Factor Algorithm enables effective traffic management strategies by allowing authorities to anticipate congestion and implement proactive measures that facilitate smoother traffic flow throughout the city.

The raw traffic data was programmatically collected from Mapbox's Traffic API v1 using Python 3.9 and the Requests library (v2.28.1). Authenticated HTTPS requests were used to query API endpoints targeting specific geocoordinates along the study route. The raw JSON responses were parsed into structured fields and organized by temporal parameters (date, time), weather conditions, and traffic variables (Traffic Condition Factor, duration) in the database schema. This structured storage approach ensures seamless integration with downstream analyses using Python libraries such as Pandas and NumPy, as well as ARIMA and SARIMA modeling, while maintaining data fidelity for reproducibility and enabling efficient querying for modeling purposes.

Table (1) The collected Data

Date-number	Weath er	Typ e date	record 's time1	Traff ic Peak	Traffic Conditi on Factor	Accide nt	Durati on
April 2, 2024	0	0.8	7:30	1	Red	0	28
April 2, 2024	0	0.8	8:30	1	Red	0	25
April 2, 2024	0	0.8	16:00	0.6	Red	0	23
April 2, 2024	0	0.8	21:20	0.5	Yellow	0	20
April 20, 2024	0	1	7:40	1	Red	0	30
April 20, 2024	1	1	8:20	1	Red	1	27
April 20, 2024	0	1	15:50	0.5	Yellow	0	24
April 20, 2024	0	1	20:50	0.9	Yellow	0	26
April 20, 2024	0	1	22:00	0.1	Blue	0	17

By structuring the data in an organized database, quick and effective access was ensured for the modeling phase. In addition to these parameters, traffic flow speed can be computed from the traffic color codes provided in the Mapbox interface, indicating traffic conditions as shown in the referenced image. By analyzing the color-coded congestion levels and accounting for the posted speed limits of each roadway segment, the traffic flow speed can be estimated using the following formula - an approach empirically validated by Aljanahi et al. (1999) who demonstrated that vehicle speeds under free-flow conditions closely adhere to speed limits, while progressively deviating under congested states due to traffic density effects and safety considerations:

$$\textit{Traffic Flow Speed} = \textit{Speed Limit} \times \textit{Traffic Condition Factor}$$

The Traffic Condition Factor (ranging from 0 for complete stoppage to 1 for free flow) operationalizes this inverse correlation between congestion intensity and achievable speeds. It is derived from color-coded congestion levels:

- Blue (free flow, factor ≈1)
- Yellow (moderate congestion, ~0.5)
- Red (heavy congestion, approaching 0)

The "Traffic Condition Factor" serves as a critical indicator of congestion levels throughout different times of the day, where higher congestion ("Red") correlates with increased traffic volumes and reduced speeds, particularly during peak hours (evidenced by elevated "Traffic Peak" values in Table 1). Conversely, "Blue" conditions reflect lower volumes and higher speeds. Furthermore, the Traffic Flow Speed formula utilized herein demonstrates that vehicle speed is directly influenced by the Traffic Condition Factor and the corresponding speed limit; during peak traffic hours, as indicated by elevated values of "Traffic Peak," vehicle speeds tend to diminish due to congestion. Additionally, the recorded duration of travel in Table 1 provides empirical insights into travel times, which are significantly impacted by both traffic volume and vehicle speed. By correlating this duration with varying traffic conditions, one can gain a comprehensive understanding of how these factors interrelate and ultimately influence overall travel times. Thus, the criteria in Table 1 offer invaluable contextual data, enabling a nuanced exploration of their relationships with traffic flow dynamics, thereby enhancing predictive modeling capabilities and informing effective traffic management strategies. In the modeling phase, the gathered traffic data was utilized to identify and forecast traffic patterns and trends. Analyzing these data allows for a better understanding of the factors influencing traffic, leading to improved decision-making in traffic management (Sattarzadeh et al., 2023). By systematically analyzing:

- Speed-flow relationships (via the Traffic Condition Factor)
- Temporal congestion patterns (peaks, durations)
- Ancillary factors (accidents, weather)

The approach enhances real-time traffic monitoring and management. The criteria in Table 1 thus provide a multidimensional foundation for predictive modeling, linking micro-level speed/volume dynamics to macro-level traffic management strategies in urban planning.

3.1. The Structure of the Traffic Data Database

The data storage system established for this research comprises a comprehensive dataset that facilitates detailed analysis and model training essential for understanding traffic conditions. An SQL database was specifically designed as shown in Figure 5, to accommodate the diverse parameters collected during the study. The schema of this database includes several critical columns: the Date (recorded as Date) captures the exact date of data collection; Weather (represented as INT) denotes the prevailing weather conditions, where a value of 0 signifies clear weather and a value of 1 indicates rainy conditions; Type Date (noted as FLOAT) categorizes the day type, with values reflecting weekdays or weekends; Record's Time (stored as TIME) specifies the precise time when each data record was obtained; Traffic Peak (also recorded as FLOAT) indicates whether the time falls within peak traffic hours, with a value of 1 for peak times and 0 for non-peak periods; the Traffic Condition Factor (FLOAT) conveys the level of congestion, derived from associated traffic color codes; the Accident column (INT) indicates the occurrence of accidents, with 1 representing an accident and 0 indicating no incident; and finally, Duration (INT) records the duration of the traffic condition in minutes.

To populate the SQL database, data insertion was performed utilizing SQL INSERT statements executed through Python scripts. These scripts automated the retrieval of real-time data from Mapbox, effectively storing the information in the database at predetermined intervals. In terms of data retrieval, this structured database allows for efficient querying via SQL commands, promoting flexible analysis of various traffic trends and patterns. The structured format not only enhances accessibility for modeling and forecasting traffic behaviors but also ensures a coherent framework for ongoing data analysis.

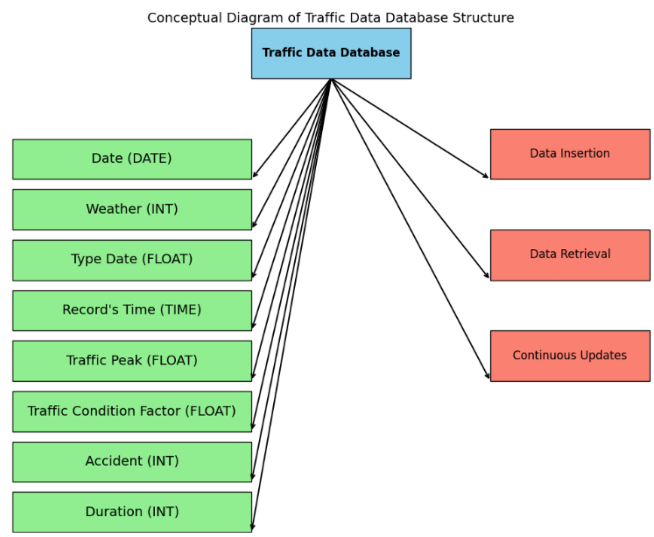

Figure 5 – The conceptual graph of Traffic Database

Furthermore, the design of the data collection and storage processes accommodates continuous updates. As new data is procured from the API, it is seamlessly appended to the existing database, thereby maintaining a current and relevant repository for further analysis. By establishing a well-structured database with comprehensive traffic data, the study robustly supports advanced data analysis and modeling, ultimately contributing to improved traffic management strategies.

4. System Architecture

This research utilized a variety of tools and software, each specifically designed to enhance the data collection, processing, and analysis processes. Python was chosen as the primary programming language because of its powerful and flexible capabilities, which facilitated the writing of scripts for data collection and initial data processing from the Mapbox platform. Well-established and widely used libraries such as Pandas and NumPy were used for in-depth data processing and numerical analysis after the data was retrieved from the SQL database. In addition, the Matplotlib library was used to visualize the results obtained from the

ARIMA/SARIMA modeling, presenting the data in a visually appealing way. Finally, to efficiently store and organize the collected data, an SQL database was implemented to optimize data access, management, and retrieval. Figure 6 illustrates the system architecture, showing the flow of data from Mapbox, through the various processing stages, to the final visualization of the traffic forecasts. This architecture highlights the relationship between data collection, data storage, data processing, modeling, and the presentation of results. Specifically, it shows how data collected from Mapbox using Java scripts is stored in an SQL database in the number and text format, then retrieved and processed using the strong Python's libraries like Pandas and NumPy before being used for ARIMA/SARIMA modeling. The results of the modeling are then visualized using Matplotlib. This coherent and efficient structure allows for systematic data analysis, ensuring that the results obtained are of a high quality. This architecture underscores the importance of a well-integrated system where each element interacts with the others to facilitate the processes of data collection, processing, analysis, and ultimately, the generation of actionable insights in traffic studies.

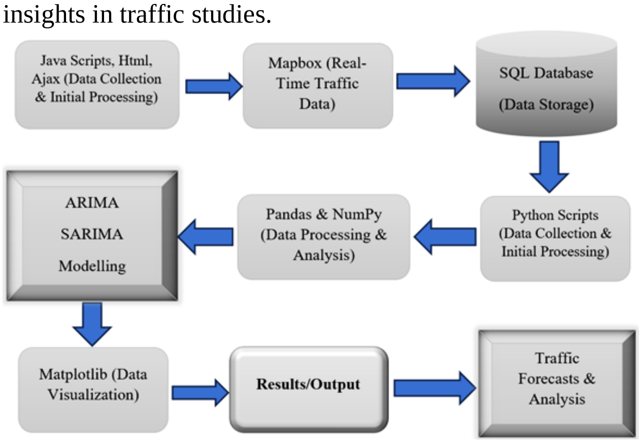

Figure 6 - System architecture and the relationship between its components

5. Time Series Modeling

As described above, online traffic data was collected from the Mapbox platform. Once the data was collected, several stages of processing and analysis were implemented. First, data cleaning was performed to identify and remove any incomplete or anomalous records in order to maintain the overall quality of the dataset. Next, descriptive analyses were performed using data analysis tools such as Pandas and Matplotlib to identify traffic patterns and fluctuations over time. In the next phase, various time series models, including ARIMA and SARIMA, were used to predict traffic patterns. The collected data was divided into training and test sets. The time series modeling used two commonly applied statistical techniques: ARIMA and SARIMA. These methods were chosen for their effectiveness in modeling and forecasting non-stationary and periodic time-dependent data, especially in the context of predicting traffic variables.

5.1. ARIMA Model

The ARIMA model is a widely used approach in time series analysis, designed to model time series data characterized by specific trends and cycles. This model has three basic components: first, the autoregressive (AR) component, which refers to the dependence of current observations on their previous values, where the parameter (p) indicates the number of lags. Second, the Integrated (I) component refers to the differencing process applied to the data to achieve stationarity, where the parameter (d) denotes the number of differencing steps required to achieve stationarity. Finally, the Moving Average (MA) component takes into account the effect of random noise on the data, with the parameter (q) indicating the number of moving averages considered (Kontopoulou et al., 2023). The general equation for the ARIMA model is shown in equation (1) below:

$$Y_t=\phi_1 Y_{t-1}+\phi_2 Y_{t-2}+\ldots+\phi_p Y_{t-p}+\theta_1\epsilon_{t-1}+\theta_2\epsilon_{t-2}+\ldots+\theta_q\epsilon_{t-q}+\epsilon_t \quad (1)$$

In this equation, (Y_t) is the predicted value at time (t). Also, (ϕ) and (θ) represent the autoregression (AR) and moving average (MA) coefficients, respectively. In addition, (ϵ_t) is considered as random noise or model error.

5.2. SARIMA Model

The SARIMA model, as an extension of the ARIMA model, is designed for time series data with seasons and periodic patterns. This model contains the same components as ARIMA, but with the addition of periodic parameters that can represent seasons, periods, and time cycles. In general, the SARIMA model is given by equation (2)

$$SARIMA(p,d,q)(P,D,Q)s \quad (2)$$

where p, d, and q represent the order of non-seasonal autoregressive (AR), integrated (I), and moving average (MA) components, respectively. P and Q represent the order of periodic AR and MA components, respectively. D specifies the order of periodic differencing, and s determines the periodic period; for example, for hourly data, the value of s will be 24 (Kumar and Hariharan, 2022). The general equation of the SARIMA model is as follows (Kumar and Hariharan, 2022; Box et al., 2015):

$$\phi(B)\Phi(B^s)(1-B)^d(1-B^s)^D Y_t=\theta(B)\Theta(B^s)\epsilon_t \quad (3)$$

In the SARIMA model formulation, denoted as equation (3), several key components interact to capture both non-periodic and periodic patterns within the time series data. These parameters are explained as follows;

- Y_t: The time series value at time t.

- B: The backshift operator (i.e., $B\,Y_t = Y_{t-1}$). Therefore, $B^s Y_t = Y_{t-s}$.
- $\phi(B)$: The non-periodic autoregressive (AR) polynomial of order p
- $\Phi(B^s)$: The periodic autoregressive (AR) polynomial of order P
- $(1-B)^d$: The non-periodic differencing operator of order d. This is used to make the time series stationary.
- $(1-B^s)^D$: The periodic differencing operator of order D. This removes periodic non-stationarity.
- $\theta(B)$: The non-periodic moving average (MA) polynomial of order q
- $\Theta(B^s)$: The periodic moving average (MA) polynomial of order Q
- ϵ_t: The error term (also known as white noise), assumed to have a mean of zero and constant variance.

6. Implementation

To effectively implement ARIMA and SARIMA models for time series forecasting, a systematic approach is essential. The first step involves testing the time series data for stationarity, which can be accomplished using statistical tests such as the Dickey-Fuller test or the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test. These tests help ascertain whether the time series exhibits a constant mean and variance over time, which is a prerequisite for applying ARIMA models. If the data is found to be non-stationary, appropriate differentiation steps must be taken to stabilize the mean of the time series. This can be achieved through first-order differencing, where the difference between consecutive observations is calculated, or higher-order differencing if necessary. Additionally, transformations such as logarithmic or periodic differencing may be applied to address any periodic patterns present in the data.

Once the time series data is rendered stationary, the next phase involves fitting the ARIMA and SARIMA models by determining the optimal values for their respective parameters: (p; the number of lag observations), (d; the degree of differencing), and (q; the size of the moving average window) for ARIMA, and (P), (D), and (Q) for SARIMA, which account for periodic effects. This can be systematically approached using the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots to identify suitable parameter values. The ACF helps determine the (q) parameter by indicating the number of lagged observations that significantly correlate with the current observation, while the PACF assists in identifying the (p) parameter.

In this study, several key criteria were established to evaluate the performance of the ARIMA and SARIMA models for predicting travel times in Tehran. Model evaluation criteria such as the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) were used to assess model fit, with lower values indicating a better fit. Additionally, Root Mean Squared Error (RMSE) was calculated to measure the accuracy of the travel time predictions produced by both models. After fitting the models, the Augmented Dickey-Fuller (ADF) test was conducted to ensure the stationarity of the time series data, which is critical for reliable forecasting.

Furthermore, the Ljung-Box test assessed the autocorrelation of the residuals to confirm that the models adequately captured the dynamics of the data. The significance of the AR and MA coefficients was also evaluated, revealing the impact of these parameters on traffic behavior predictions. Lastly, tests for heteroscedasticity and normality of residuals were performed to identify any potential issues that could affect model performance.

After evaluating the models, cross-validation techniques can be implemented to assess predictive performance on unseen data, ensuring robustness. Once the best model is selected, it can be utilized to predict travel times. This involves applying the fitted model to new data and generating forecasts, which can be further refined through ensemble methods or by incorporating additional predictors, such as weather conditions or special events, to enhance accuracy. Overall, this structured methodology provides a comprehensive framework for researchers and practitioners to effectively implement ARIMA and SARIMA models for analyzing and predicting traffic patterns, leading to more informed decision-making in urban traffic management.

6.1. Experimental Results

In this section, we present a comprehensive analysis of the results obtained from the implementation of ARIMA and SARIMA models on the traffic data. The models were trained on historical travel time data, and their performance was evaluated using statistical metrics such as the AIC, BIC, and RMSE. The primary objective of this analysis is to evaluate the effectiveness of these models in forecasting traffic behavior, which refers to the patterns and dynamics of traffic-related variables such as travel time, traffic volume, and vehicle speed. These variables collectively describe how traffic conditions evolve over time and are influenced by factors such as time of day, day of the week, weather conditions, and special events. The forecast graphs for the ARIMA and SARIMA models are presented in Figures 7 and 8, respectively. These visual representations illustrate the predicted values against the actual observed traffic data, allowing for a clear comparison of the models' performance over the specified time period. Additionally, the numerical results obtained from these models are summarized in Tables 2 and 3. These tables provide a detailed overview of the model parameters, evaluation metrics, and the corresponding forecasts, facilitating a deeper understanding of the models' efficacy in predicting traffic patterns.

Table (2) ARIMAX Results

No. Observations		6592	
Log Likelihood		9038.397	
ADF Statistic		-15.079762	
Critical Values t1% t5% t10%		Values -3.431 -2.862 -2.567	
Covariance Type		opg	
Ljung-Box (L1) (Q)	0.00	Jarque-Bera (JB)	34187.68
Prob(Q)	0.98	Prob(JB)	0.00
Heteroskedasticity (H)	0.85	Skew	0.83
Prob(H) (two-sided)	0.00	Kurtosis	14.03

Parameters	coef	std err	z	P>|z|	[0.025	0.975]
ar.L1	-0.0270	0.050	-0.542	0.588	-0.125	0.071
ma.L1	0.2028	0.049	4.155	0.000	0.107	0.298
sigma2	0.0038	2.58e-05	146.247	0.000	0. 004	0.004

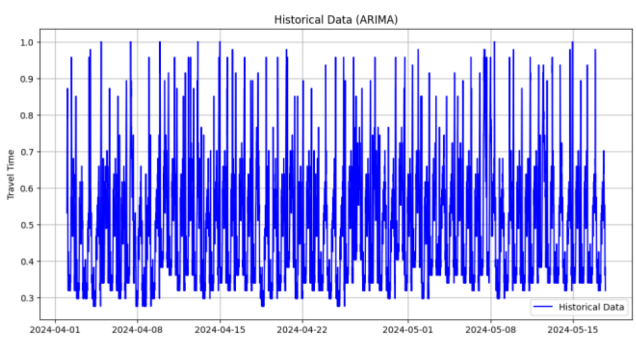

Figure 7 - ARIMA model for Historical Data

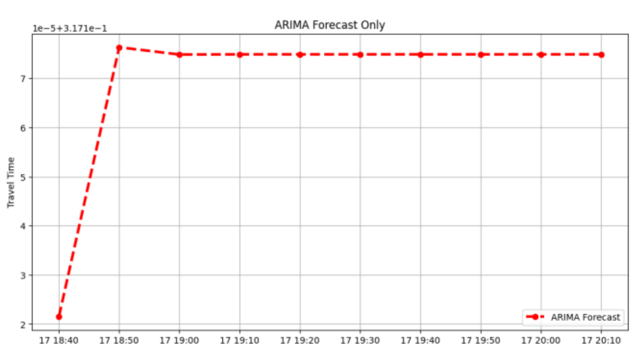

Figure 8 - ARIMA Forecast line

Table (3) SARIMAX Results

No. Observations		6592	
Log Likelihood		9040.348	
ADF Statistic		-15.079762	
Critical Values t1% t5% t10%		Values -3.431 -2.862 -2.567	
Covariance Type		opg	
Ljung-Box (L1) (Q)	0.00	Jarque-Bera (JB)	34264.88
Prob(Q)	0.96	Prob(JB)	0.00
Heteroskedasticity (H)	0.85	Skew	0.83
Prob(H) (two-sided)	0.00	Kurtosis	14.05

Parameters	coef	std err	z	P>|z|	[0.025	0.975]
ar.L1	-0.0268	0.050	0.538	0.591	-0.124	0.071
ma.L1	0.2033	0.049	4.179	0.000	0.108	0.299
ar.S.L24	0.1473	0.602	0.245	0.807	-1.033	1.328
ma.S.L24	-0.1234	0.604	-0.204	0.838	-1.308	1.061
sigma2	0.0038	2.58e-05	146.247	0.000	0. 004	0.004

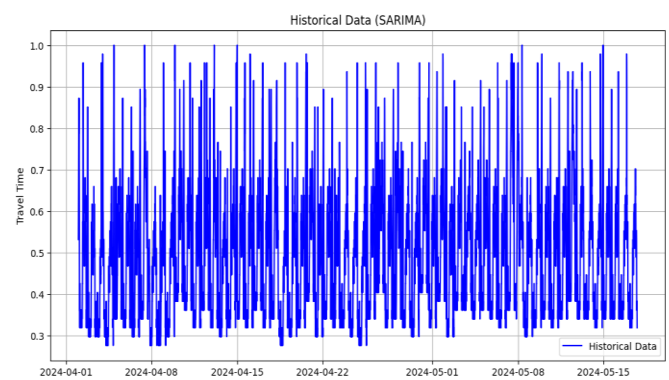

Figure 9 - SARIMA model for Historical Data

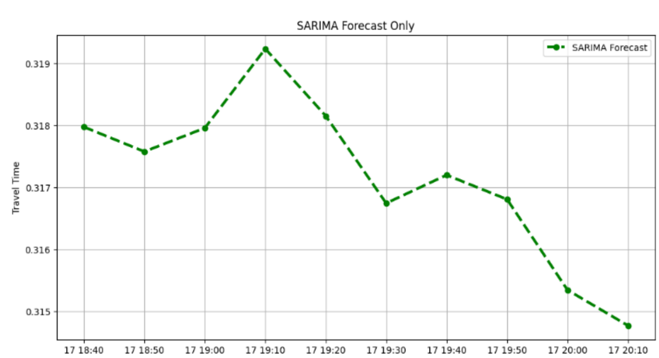

Figure 10 - SARIMA Forecast line

According to the values shown in Tables (2 and 3), the results obtained from the ARIMA and SARIMA models are analyzed.

The independent forecasts generated by the ARIMA and SARIMA models provide valuable insights into the expected future behavior of travel time.

The ARIMA model, which utilizes the parameters (p=1, d=1, q=1), produced a forecast characterized by a red dashed line in its corresponding plot. This model's predictive performance is grounded in its ability to capture the underlying temporal dynamics of the data, as evidenced by the ADF test results, which indicate a strong rejection of the null hypothesis of non-stationarity (ADF statistic = -15.08, p-value < 0.001).

In addition to the ARIMA and SARIMA forecasts, the visual representation of the predictions in separate plots further elucidates the distinct characteristics of each model. In Figures (8 and 10). In Figure 8, plot displays the ARIMA forecast for travel time, represented by a red dashed line with markers indicating the forecasted values over the next 10-time intervals (10-minute increments). The x-axis represents the date and time, while the y-axis indicates the predicted travel time.

So, if the red forecast line shows a consistent pattern, this indicates that the ARIMA model effectively captures the underlying trend in the travel time data. For example, if the forecast suggests a gradual increase or decrease in travel time, it may reflect anticipated changes due to factors like traffic patterns or upcoming events. However, if the forecast deviates significantly from historical patterns, it could indicate that the ARIMA model is not fully capturing the dynamics of the data, possibly due to periodic variations or other external factors not included in the model. In contrast, the SARIMA model, incorporating periodic components (1, 1, 1)x(1, 0, 1, 24), is represented by the green dashed line in its respective plot as Figure 10. This plot shows the SARIMA forecast for travel time, represented by a green dashed line with markers for the predicted values over the same 10 future time intervals. Like the ARIMA plot, the x-axis indicates date and time, while the y-axis displays the predicted travel time.

The SARIMA model extends the ARIMA model by incorporating periodic components, making it more suitable for datasets with periodic fluctuations. The periodic order (1, 0, 1, 24) indicates that the model considers one periodic autoregressive term, no periodic differencing, one periodic moving average term, and a periodic period of 24 (which could represent daily cycles in this context). So, if the green forecast line closely tracks the historical travel times, it suggests that the SARIMA model is effectively capturing both the trend and periodic variations in the data. This is particularly important in travel time forecasting, where factors such as time of day, day of the week, or special events can significantly impact travel patterns. Any significant discrepancies between the SARIMA forecast and historical data may indicate that the model needs adjustments, such as refining periodic parameters or incorporating additional explanatory variables.

Totally, the two forecasting plots illustrate the ARIMA forecast, showcasing its ability to predict future values without considering periodic effects, and the SARIMA forecast, highlighting its sensitivity to periodic trends. The third and fourth plots showed in Figures (7 and 9) could be utilized to compare the residuals of both models, providing insights into their predictive accuracy and potential areas for improvement. The SARIMA model's forecasts account for periodic variations, enhancing its predictive capability in contexts where periodic fluctuations are present. Notably, the SARIMA model yielded a marginally better log-likelihood and lower AIC values compared to the ARIMA model, suggesting a more robust fit to the data.

Overall, these independent forecasts facilitate a comprehensive comparative analysis, allowing for an informed selection of the most appropriate model based on performance metrics and the nature of the underlying data. Such analyses are crucial for making data-driven decisions in time series forecasting, particularly in fields where understanding temporal trends is essential. Altogether, these findings underscore the effectiveness of the SARIMA model in capturing the dynamics of traffic volume and provide a solid foundation for utilizing these models in practical traffic management applications. The insights gained from this analysis can inform strategies for optimizing traffic flow and enhancing urban transportation systems.

In summary, this section presented the findings from implementing ARIMA and SARIMA models to forecast traffic behavior, particularly travel time, using a dataset of approximately 6,500 records collected from the Mapbox platform over two months for a major route in Tehran. The ARIMA model effectively captured temporal trends, while the SARIMA model demonstrated superior performance by incorporating daily periodic fluctuations, achieving a marginally better log-likelihood and lower AIC values. The MA coefficient showed a significant positive impact (p-value = 0.000), whereas the AR coefficient was non-significant (p-value = 0.588). Residual analysis confirmed the absence of autocorrelation (Ljung-Box Q = 0.00, p-value = 0.98), but non-normality and heteroscedasticity were detected, indicating areas for refinement.

6.2. Evaluation of the model parameters

The analysis of the SARIMA model with parameters (1, 1, 1) reveals a log likelihood of 9038.397, accompanied by AIC and BIC values of -18070.794 and -18050.414, respectively, indicating an adequate fit to the data. A closer examination of the model’s standard errors and z-values shows that the MA parameter has a significant coefficient (p-value = 0.000), reflecting its meaningful impact on the dependent variable, traffic volume. In contrast, the AR parameter is not significant (p-value = 0.588), suggesting it does not substantially affect the dependent variable. Specifically, the AR coefficient (ar.L1) is -0.0270, suggesting a negative influence from the first lag, though its effect is not significant due to the high p-value. The MA coefficient (ma.L1) is 0.2028, indicating a positive and statistically significant effect with a p-value of 0.000. The error variance (sigma²) is 0.0038, illustrating the dispersion of errors in the forecasts.

Additionally, the SARIMAX model with parameters (1, 1, 1) x (1, 0, 1, 24) was assessed, yielding a log likelihood of 9040.348 and AIC and BIC values of -18070.697 and -18036.730, respectively. In this model, the MA parameter continues to demonstrate a significant positive effect (p-value = 0.000). However, the periodic AR and MA parameters (ar.S.L24 and ma.S.L24) are not statistically significant (p-values above 0.05), indicating they do not influence traffic forecasts over the 24-hour period.

Additional tests were performed to evaluate the residual characteristics of the model. The Ljung Box test yielded a Q statistic of 0.00 and a p-value of 0.98, confirming the absence of autocorrelation in the residuals (with a p-value close to 1), indicating the appropriateness of the selected models. Furthermore, the results of the Jarque-Bera test indicate that the distribution of the residuals is not normal (p-value = 0.00), suggesting a possible need for further investigation into the normality of the residuals. In addition, the presence of heteroscedasticity was indicated by a p-value of 0.00, which suggests a likelihood of instability in the error variance.

The ARIMA and SARIMA plots provide valuable insights into the expected future behavior of travel time. The ARIMA forecast focuses on capturing the trend based solely on historical data, while the SARIMA forecast enhances this by accounting for periodic variations, leading to potentially more accurate predictions. By comparing both forecasts, stakeholders can evaluate which model better captures the dynamics of travel time and make informed decisions based on the predicted trends.

The overall evaluation of the models, as illustrated in Figures 11 and 12, which depict the residuals and their distribution, indicates that the SARIMA model has been well fitted to the data. However, the results of the Jarque-Bera test and indications of heteroscedasticity necessitate further investigation. Additionally, the non-significance of the AR coefficient may suggest that there is no need to include this component in the model.

To enhance the model, it is recommended to adjust the parameters or to implement alternative methodologies, such as generalized autoregressive conditional heteroskedasticity (GARCH) models, to address the issue of heteroscedasticity. These findings highlight the model's ability to forecast time series data and underscore the need for further research aimed at improving prediction accuracy. Overall, the results demonstrate the strong performance of ARIMA and SARIMA models in forecasting traffic behaviors, although improvements are warranted regarding the normality of the residuals. Such insights can assist urban planners in making informed decisions for better traffic management.

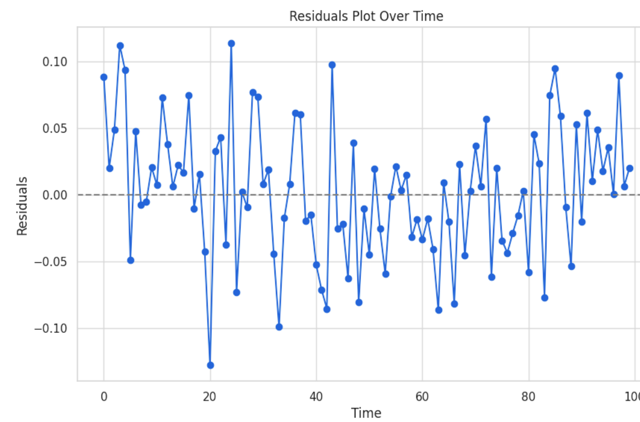

Figure 11 - Plot of randomly generated residuals and their line plots to check Normality and Heteroscedasticity

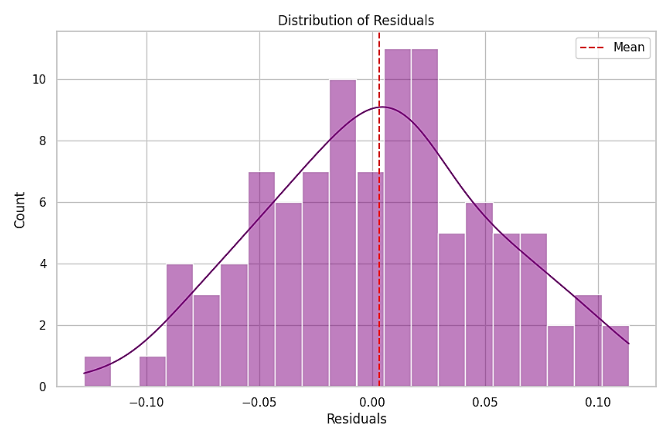

Figure 12 - Histogram of residual distribution

While the primary focus of this study is on travel time prediction, the ARIMA and SARIMA models implemented in this research are versatile and capable of analyzing other traffic-related variables, such as traffic volume and vehicle speed. The methodology described in the "Implementation" section is applicable to any time series data, and the models can be adapted to predict different traffic metrics depending on the specific objectives of the analysis. In this study, travel time was selected as the key output due to its direct relevance to urban traffic management and its significant impact on commuter experience and transportation efficiency. However, the same approach can be extended to other variables, such as traffic volume or vehicle speed, by adjusting the input data and model parameters accordingly.

7. Results' Evaluation

The implementation of the ARIMA and SARIMA models in Tehran's study area has yielded significant insights into traffic patterns at various times of the day. The results can be summarized as follows:

1. Peak Hour Analysis: Morning Rush Hour (7 AM - 9 AM): The models indicated a substantial increase in traffic volume, with predicted congestion levels rising by approximately 40% compared to off-peak hours. The average travel time during this period extended by an estimated 20 minutes due to heightened vehicle density, particularly on major routes.
2. Midday Conditions (12 PM - 2 PM): During lunchtime, the traffic volume showed a moderate increase, approximately 15% higher than pre-peak hours. The models forecasted a slight rise in average travel times, suggesting that local business activities contribute to increased congestion.
3. Evening Rush Hour (5 PM - 7 PM): Similar to the morning rush, the evening peak presented significant congestion, with estimates indicating a 35% increase in traffic volume. The models predicted average travel times to escalate by nearly 25% as commuters returned home, leading to traffic bottlenecks in residential areas.
4. Late Night and Early Morning (11 PM - 6 AM): In contrast, during late-night hours, traffic volumes decreased dramatically by around 60%. The models indicated minimal congestion, with average travel times reflecting a notable improvement, averaging only 15 minutes on arterial roads.
5. Impact of Weather Conditions: The models also revealed that adverse weather conditions (e.g., heavy rain) could lead to a further 15-20% increase in average travel times across all hours. For instance, traffic patterns during heavy rainfall not only saw a rise in congestion but also demonstrated a tendency for more pronounced delays, particularly for those traveling during peak hours.
6. Holiday Traffic Patterns: Results showed that on national holidays, traffic volumes can fluctuate significantly. For example, during holidays, traffic levels may be 30% lower compared to regular weekdays, presenting unique opportunities for urban planners to optimize road usage and public transportation systems.

Overall, the successful implementation of these forecasting models provides valuable insights into Tehran's traffic dynamics, allowing authorities to anticipate congestion, optimize traffic flow, and facilitate better urban planning during various hours of the day and under different conditions.

7.1. Comparison of SARIMA Travel Time Modeling with Google Maps

This study investigates the potential of ARIMA and SARIMA models as statistically-driven alternatives for real-time travel time prediction, contrasting these models with services like Google Maps, which utilize advanced hybrid approaches combining model-based and data-driven techniques. The findings indicate that with meticulous parameter selection, ARIMA and SARIMA models can yield accuracy comparable to Google Maps, particularly in short-term forecasting scenarios. This means that for relatively brief prediction windows, the travel times estimated by SARIMA models can closely reflect those provided by Google Maps. However, it is important to acknowledge the inherent limitations of ARIMA and SARIMA models in fully capturing the complexities and variability of real-world traffic dynamics. Unlike Google Maps, which benefits from a vast array of real-time data points and sophisticated algorithms, the traditional time series models may struggle to adapt dynamically to sudden changes, such as accidents or unexpected road closures.

A comprehensive comparison between these methods proves challenging due to their fundamental differences. Google Maps uses a robust algorithm that integrates historical data, current traffic conditions, and user-reported incidents to provide real-time updates, enabling it to adjust predictions on the fly. In contrast, ARIMA and SARIMA are primarily focused on historical data analysis and do not inherently account for real-time fluctuations unless supplemented with additional information. To bridge these gaps, the study suggests several future research directions. One promising avenue involves integrating neural network models with traditional time series approaches to better address parameter uncertainty and enhance predictive accuracy. For instance, ARIMA or SARIMA models can be employed to establish a baseline prediction, which can then be refined by incorporating insights from neural networks that utilize external data sources, such as weather conditions or social media reports. This hybrid methodology aims to leverage the strengths of both traditional statistical models and modern machine learning techniques, potentially leading to more accurate and responsive travel time predictions. In summary, while ARIMA and SARIMA models present a viable statistical alternative for travel time forecasting, their limitations necessitate further exploration of hybrid approaches that combine these models' strengths with the real-time adaptability of platforms like Google Maps.

8. Conclusion

Urban growth intensifies traffic problems. Analyzing traffic data (volume, speed, travel times) informs better traffic management and infrastructure decisions. Integrating advanced statistical methods with real-time traffic data can reduce travel times, enhance safety, and improve public transport efficiency. Online data analysis empowers urban planners to manage traffic, coordinate signals, and optimize networks. Key challenges include data handling, infrastructure, and privacy. Ultimately, this leads to increased urban resilience through effective traffic forecasting and management. This research explored the importance of using online traffic data in the analysis and application of ARIMA and SARIMA time series models to online traffic data using the Mapbox platform for urban traffic analysis and forecasting. The data collected from the

Mapbox platform and associated analyses not only revealed traffic patterns, but also facilitated the provision of optimized recommendations for traffic management and flow improvement. Our results corroborate previous findings on the effectiveness of these models in predicting traffic trends and identifying periodic variations, and demonstrate their utility for analyzing real-world traffic data collected from a contemporary online platform. Specifically, we found a significant positive impact of the first-lag moving average (MA) coefficient (0.2028, $p < 0.000$), indicating the importance of recent traffic conditions on current predictions. While the first-lag autoregressive (AR) coefficient (-0.0270) showed a negative impact, its non-significance ($p = 0.588$) suggests that its impact is less pronounced in our dataset.

Nevertheless, the distinctive contribution of this study is its emphasis on travel time modeling in Tehran, a city where research on traffic data series and the incorporation of real-time traffic data into the modeling process has been restricted. Additionally, it integrates various pivotal parameters that have been frequently disregarded in prior research. Specifically, our model incorporates the influence of dates and times (accounting for holidays and peak hours), weather conditions, and the occurrence of accidents. By incorporating these factors, which have been shown to have a significant impact on traffic volume in Tehran, we aim to provide a more realistic and accurate representation of travel time dynamics. This methodological advancement enables us to transcend the limitations of conventional historical trends and instead capture the intricate interplay of factors that collectively contribute to Tehran's pervasive traffic congestion. The insights derived from this approach have the potential to inform efficacious solutions that alleviate this persistent problem.

In general, ARIMA and SARIMA models are effective for analyzing periodic and non-stationary traffic data. In the future, urban traffic management will become more intelligent through AI, IoT, and big data. This study demonstrates the use of online traffic data for urban planning and infrastructure improvement, and provides a model for future research to improve traffic resilience and efficiency amidst urban challenges.

The findings of our analysis suggest that traffic patterns on weekdays are highly predictable, and the occurrence of accidents and inclement weather can result in increased travel times. This information can be used to optimize traffic signal timing during peak hours, thereby improving traffic flow and reducing delays.

9. Discussion and Recommendations

Our research highlights the effectiveness of ARIMA and SARIMA models in simulating and forecasting traffic patterns in Tehran. By incorporating critical parameters such as specific dates/times (holidays and peak hours), weather conditions, and accident occurrences, we have captured the unique dynamics of traffic congestion. The models generated outputs detailing predicted traffic volumes, average travel times, and congestion levels, revealing that during peak hours, congestion levels markedly increase compared to regular periods. Furthermore, variations in weather conditions, such as rain, indicated a correlation with longer travel times.

These findings align with previous studies that suggest environmental factors and specific temporal elements significantly impact traffic flow. For instance, existing literature has consistently shown that adverse weather conditions correlate with increased travel delays (insert reference). Additionally, the effectiveness of time series modeling in traffic prediction has been supported by various studies, highlighting a robust methodology for addressing urban transportation issues (insert references). The outcomes of our research emphasize the importance of leveraging online traffic data for effective traffic management. By utilizing the predictive capabilities of ARIMA and SARIMA models, urban planners and traffic management authorities can gain insights into congestion patterns and make data-driven decisions. For instance, recognizing peak hours and their relationship with holiday traffic can inform logistical planning and resource allocation.

To address traffic congestion in Tehran, several key actions are recommended. First, investing in robust ICT infrastructure is crucial for managing large volumes of traffic data, enabling real-time monitoring and improved analysis. Integrating machine learning techniques can enhance traffic forecasting by revealing specific driver behavior patterns and improving travel demand predictions. Collaboration among stakeholders, including the Municipality of Tehran and law enforcement, is essential to develop comprehensive traffic management strategies. Training local experts in data analysis will build the capacity needed to tackle traffic dynamics effectively. Prioritizing data privacy and security is vital to maintain public trust as traffic data is collected. Lastly, long-term research focusing on urban resilience and intelligent transportation systems will promote sustainable innovation in Tehran's transport planning. By implementing these strategies, the city can improve its transportation system's resilience and efficiency.

References

Datola G. Implementing urban resilience in urban planning: A comprehensive framework for urban resilience evaluation. Sustainable Cities and Society. 2023 Nov 1;98:104821.
https://doi.org/10.1016/j.scs.2023.104821

Dianat H, Wilkinson S, Williams P, Khatibi H. Choosing a holistic urban resilience assessment tool. International Journal of Disaster Risk Reduction. 2022 Mar 1;71:102789.
https://doi.org/10.1016/j.ijdrr.2022.102789

Chen H, Zhou R, Chen H, Lau A. A resilience-oriented evaluation and identification of critical thresholds for traffic congestion diffusion. Physica A: Statistical Mechanics and its Applications. 2022 Aug

15;600:127592.
https://doi.org/10.1016/j.physa.2022.127592

Najafi Moghaddam Gilani V, Hosseinian SM, Ghasedi M, Nikookar M. Data-Driven Urban Traffic Accident Analysis and Prediction Using Logit and Machine Learning-Based Pattern Recognition Models. Mathematical problems in engineering. 2021;2021(1):9974219.
https://doi.org/10.1155/2021/9974219

Boukerche A, Wang J. Machine learning-based traffic prediction models for intelligent transportation systems. Computer Networks. 2020 Nov 9;181:107530.
https://doi.org/10.1016/j.comnet.2020.107530

Karami Z, Kashef R. Smart transportation planning: Data, models, and algorithms. Transportation Engineering. 2020 Dec 1;2:100013.
https://doi.org/10.1016/j.treng.2020.100013

Wu P, Chen T, Wong YD, Meng X, Wang X, Liu W. Exploring key spatio-temporal features of crash risk hot spots on urban road network: A machine learning approach. Transportation research part A: policy and practice. 2023 Jul 1;173:103717.
https://doi.org/10.1016/j.tra.2023.103717

Yang H, Li X, Qiang W, Zhao Y, Zhang W, Tang C. A network traffic forecasting method based on SA optimized ARIMA–BP neural network. Computer Networks. 2021 Jul 5;193:108102.
https://doi.org/10.1016/j.comnet.2021.108102

Wang Y, Yu C, Hou J, Chu S, Zhang Y, Zhu Y. ARIMA Model and Few-Shot Learning for Vehicle Speed Time Series Analysis and Prediction. Computational Intelligence and Neuroscience. 2022;2022(1):2526821.
https://doi.org/10.1155/2022/2526821

Nassiri H, Mohammadpour SI, Dahaghin M. Forecasting time trend of road traffic crashes in Iran using the macro-scale traffic flow characteristics. Heliyon. 2023 Mar 1;9(3).
https://doi.org/10.1016/j.heliyon.2023.e14481

Deretić N, Stanimirović D, Awadh MA, Vujanović N, Djukić A. SARIMA modelling approach for forecasting of traffic accidents. Sustainability. 2022 Apr 7;14(8):4403
https://doi.org/10.3390/su14084403

Wang Y, Jia R, Dai F, Ye Y. Traffic flow prediction method based on seasonal characteristics and SARIMA-NAR model. Applied Sciences. 2022 Feb 19;12(4):2190.
https://doi.org/10.3390/app12042190

Rabbani MB, Musarat MA, Alaloul WS, Rabbani MS, Maqsoom A, Ayub S, Bukhari H, Altaf M. A comparison between seasonal autoregressive integrated moving average (SARIMA) and exponential smoothing (ES) based on time series model for forecasting road accidents. Arabian Journal for Science and Engineering. 2021 Nov;46(11):11113-38.
https://doi.org/10.1007/s13369-021-05650-3

Patil P. A Comparative Study of Different Time Series Forecasting Methods for Predicting Traffic Flow and Congestion Levels in Urban Networks. International Journal of Information and Cybersecurity. 2022 Jan 9;6(1):1-20.
https://doi.org/10.3390/fi14100294

Verma R, Pandey S, Rup S. Traffic Flow Prediction Using LSTM, ARIMA & SARIMA Model for Intelligent Transportation System. In2024 International Conference on Intelligent Computing and Sustainable Innovations in Technology (IC-SIT) 2024 Nov 21 (pp. 1-6). IEEE.
https://doi.org/10.1109/IC-SIT63503.2024.10862828

Oladimeji D, Gupta K, Kose NA, Gundogan K, Ge L, Liang F. Smart transportation: an overview of technologies and applications. Sensors. 2023 Apr 11;23(8):3880.
https://doi.org/10.3390/s23083880

Medina-Salgado B, Sánchez-DelaCruz E, Pozos-Parra P, Sierra JE. Urban traffic flow prediction techniques: A review. Sustainable Computing: Informatics and Systems. 2022 Sep 1;35:100739.
https://doi.org/10.1016/j.suscom.2022.100739

Yuan H, Li G. A survey of traffic prediction: from spatio-temporal data to intelligent transportation. Data Science and Engineering. 2021 Mar;6(1):63-85.
https://doi.org/10.1007/s41019-020-00151-z

Aljanahi AA, Rhodes AH, Metcalfe AV. Speed, speed limits and road traffic accidents under free flow conditions. Accident Analysis & Prevention. 1999 Jan 1;31(1-2):161-8.
https://doi.org/10.1016/S0001-4575(98)00058-X

Sattarzadeh AR, Kutadinata RJ, Pathirana PN, Huynh VT. A novel hybrid deep learning model with ARIMA Conv-LSTM networks and shuffle attention layer for short-term traffic flow prediction. Transportmetrica A: Transport Science. 2023 Jul 27:1-23.
https://doi.org/10.1080/23249935.2023.2236724

Kontopoulou VI, Panagopoulos AD, Kakkos I, Matsopoulos GK. A review of ARIMA vs. machine learning approaches for time series forecasting in data driven networks. Future Internet. 2023 Jul 30;15(8):255.
https://doi.org/10.3390/fi15080255

Kumar PB, Hariharan K. Time series traffic flow prediction with hyper-parameter optimized ARIMA models for intelligent transportation system. Journal of Scientific & Industrial Research. 2022 May 4;81(04):408-15.
http://op.niscair.res.in/index.php/JSIR/article/view/50791

Box GE, Jenkins GM, Reinsel GC, Ljung GM. Time series analysis: forecasting and control. John Wiley & Sons; 2015 May 29.
ISBN:978-0-13-060774-4

Mapbox. (n.d.). Real-time traffic data. Retrieved Since April 1, 2024 to Present, from https://www.mapbox.com