Analysis of Traffic Volume and Population Trends in New York City

Arefa Binti Sumaya - 23187010

Introduction

Planning urban mobility and infrastructure requires understanding the connection between traffic volume and population demographics. These factors play a vital role in designing effective transportation system and tackling urban development challenges. By Integrating traffic and demographic data, policymakers can identify the trends, potential issues and make better strategies that improve urban planning and improve the quality life for residents. Thus, this report aims to offer practical insights into the question, “How can we combine and evaluate population demographics and traffic data to guide urban development decisions?”. To answer this question, we performed advance data analysis to reveal relationships and provide actionable insights for New York City.

Used Data

The analysis was conducted using an SQLite database comprising two primary tables: Traffic Volume Data and Population Data by age and sex. The data was processed through an automated ETL pipeline to make it ready for analysis. This datasets provides important information about population and traffic trends so we organized to simplify our analysis.

The Population Data by age and sex comes from StatsAmerica and provides demographic details for U.S., including states and countries. As shown in Table 1, It contains geographic identifier like state and country codes, as well as descriptive names such as “New York”. The data also provides different population segments, including groups (e.g., ”Population 0-4,” ”Population 18-54”) and gender categories (”Male Population,” ”Female Population”). After processing, the output data focus only on New York State, also removing unnecessary identifiers like numeric codes. This simplified format makes it easier to analysis trends by age and gender. The Traffic Volume Data is sourced from the NYC Open Data portal

Dataset	**Key Features**	**Value Types**	**License**
Population Data	Geographic identifiers, year, age groups, gender, and population aggregates (e.g., ”Population Under 18”).	Numeric, Text	Creative Commons Attribution 4.0 (CC BY)
Traffic Volume Data	Road identifiers, date, hourly traffic counts, and aggregated time intervals (e.g., ”12:00-4:00 AM”).	Numeric, Text, Date	NYC Open Data Terms of Use

Table 1: Dataset description with key features, value types, and licenses.

and contains hourly vehicle counts for various road segments in the city. It includes information such as road names, recording dates, and hourly traffic volumes (e.g., ”12:00-1:00 AM”). The processed output data, we aggregated into larger time intervals of four hours, like ”12:00-4:00 AM” and ”4:00-8:00 AM”. This allows us to more efficient analysis of traffic pattern and trends. Both datasets have specific licenses that guide their use. The Population Data by Age and Sex follows the Creative Commons Attribution 4.0 International (CC BY 4.0) license, which allows sharing and adapting the data with proper credit. The Traffic Volume Data follows the NYC Open Data Terms of Use, which requires proper attribution and does not allow commercial use.

To comply with these licenses, the sources will be credited in all reports, presentations, or publications using this data. Any adapted versions of the datasets will also include the same licenses to ensure transparency and responsible usage.

Analysis

A. Correlation Analysis:

Correlation analysis is a statistical method used to measure the strength and direction of the relationship between two variables. It helps to understand whether the variables increase or decrease together and how strong the connection is. This analysis is useful and make us understand how population demographics, such as various age groups, impact traffic patterns at specific times, providing important insights for urban planing and transportation system [1].

i. Method:

To study the relationship between population demographics and traffic volumes, we followed several steps. First, we prepared the data by selecting population data for New York and also focusing on age groups (e.g., 0-4, 25-44). We focused on the two busiest time intervals from the traffic data: 8:00 AM – 12:00 PM in the morning and 8:00 PM – 12:00 AM in the late evening. Then we used the Spearman correlation coefficient to measure the relationships between population groups and traffic volumns [2]. Also, data was normalized using MinMax scaling to compare variables. Finally, we used scatter plot and heatmap for visualization that create to better understand the results.

ii. Results:

For this result analysing we seen a clear patterns in traffic and population relationships. The scatter plot (see Figure 1) showed a strong positive relationships between the working age (5-17, 18-24, 25-44) and traffic volumes during the 8:00 AM–12:00 PM time interval. In contrast, weaker relationship were observed for children (0-4) and elderly (65+) that indicates that lower traveling frequencies these groups during peak hours.

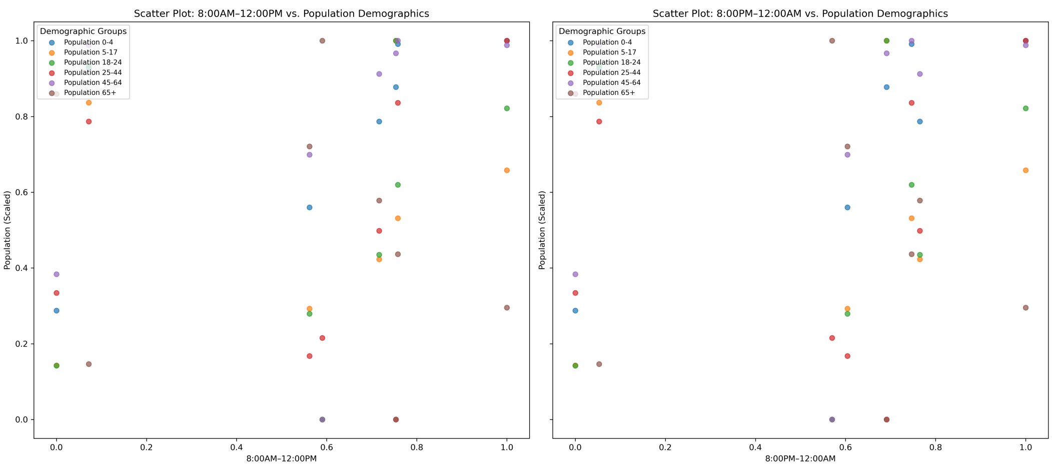

Figure 1: Scatter plots showing the relationship between traffic volumes and population demographics during 8:00 AM–12:00 PM (left) and 8:00 PM–12:00 AM (right).

The heat map (see Figure 2) showed that the working-age population(18-24, 25-44, 45-64) had the highest correlations with peak-hour traffic volume (Spearman $\rho \approx 0.7$–0.8). The elderly population (65+) had negative correlations during peak hour traffic but moderate connection in mid day travaling (12:00 PM–4:00 PM).

iii. Interpretation:

The working age population creates most of the traffic during morning (8:00 AM–12:00 PM) and midday (8:00 PM–12:00 AM). In contrast, elderly (65+) and children (0-4) are less travel during that hour. For the urban planning, public transportation and road systems should be improved to better serve working age commuters during busy morning hours. In midday transit services should be adjusted with flexible schedules to support elderly people.

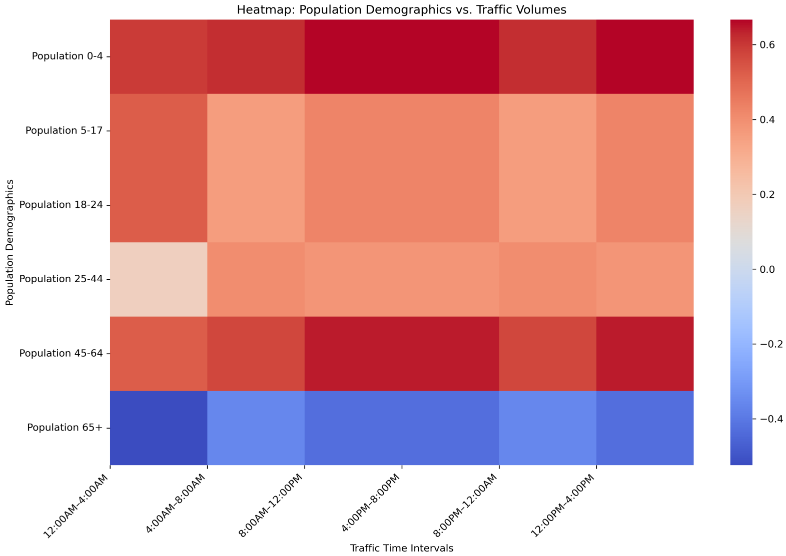

Figure 2: Heatmap showing the correlation between population demographics and traffic volumes across different time intervals.

B. Time Series Decomposition Analysis:

In our analysis, We used time series decomposition analysis to break down traffic volume and population data into trend, seasonality, and residuals [3]. It helped us understand how these factors influence urban development needs, such as transit planning and infrastructure optimization.

i. Method:

In TSDA method, Firstly, the data separates the data into three parts: trends, which shows the overall direction; seasonality, showing repeating patterns at regular intervals; and residuals, capturing the random changes. Secondly, traffic data from 2012 to 2019 was analyzed to to detect long time trends, seasonal patterns, and irregular changes. On the other hand, population data was decomposed to find stable trends and seasonal shifts. Lastly, we chosen this approach because it allow us separate and examine each value independently.

ii. Results:

For traffic volumes (see Figure 3), the trend was relatively flat, so there is no significant increase or decrease between 2012 and 2019. The seasonality showed same patterns, likely tied to holidays, weather or specific times of the year. The residuals observed unusual changes or irregular variations its causes because of construction or special occasions.

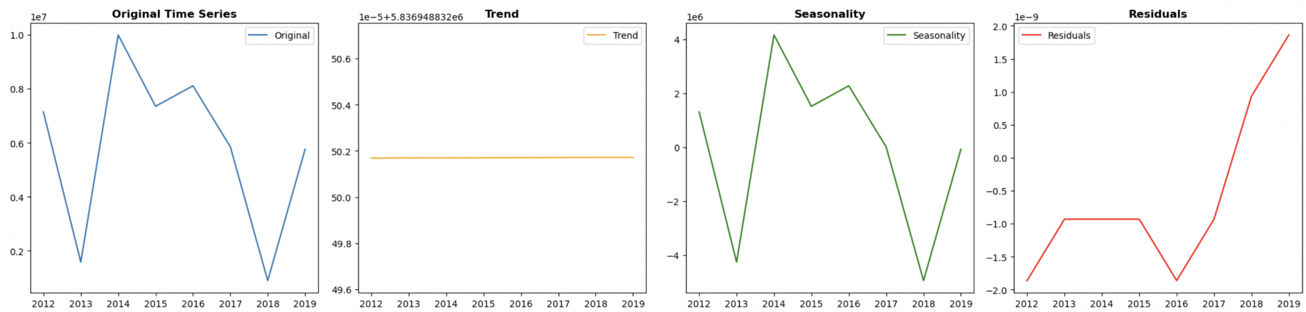

Figure 3: Traffic Volume Decomposition

For population data (see Figure 4), the trend was stable so there is no change over time. The seasonality showed regular changes, possibly caused by migration, jobs, or temporary population shifts. The residuals showed unusual changes or one-time events in population dynamics.

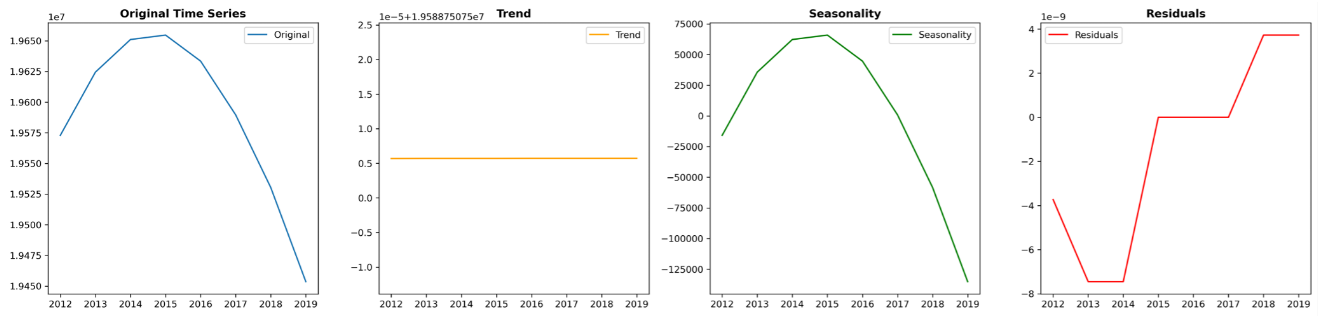

Figure 4: Population Decomposition

iii. Interpretation:

Traffic volumes stayed stable over the years, so it suggest that infrastructure demand has remained same. Although, seasonal pattern showed spikes because of holiday or tourism. Population trends stayed steady there is minimal long term growth and existing infrastructure is sufficient. Seasonal population changes could affect local traffic because of migration or temporary stays. For urban planning, seasonal traffic patterns can help improve public transit schedules and infrastructure during rush hours [4]. While stable population trends showed that current infrastructure is enough, but we need to give more attention to seasonal spikes and local changes. Additionally, resudual analysis highlights the need to plan for unexpected events like road closures or demographic shifts [5]. Overall, this decomposition analysis gives useful insights to better transportation systems with population trends and seasonal patterns.

Conclusion

Through this analysis, we aimed to understand how population and traffic data can combined to guide urban development decisions. We used correlation analysis and time series decomposition to study the relationship between demographics and traffic patterns. The correlation analysis showed strong relationships between working-age populations and peak-hour traffic volumes to describe the importance of commuters in demand. Time series decomposition showed long term trends, seasonal patterns, and irregular changes in both traffic and population data, giving insights into recurring demands and unexpected disruptions. However, there are some limitations. The analysis based on general data, but more specific regional data could provide more understanding. It did not include factor like weather, economic changes which may affect traffic volume. It only focused on past trend (2012-2019), so the future predictions uncertain. In spite of these limitations, the analysis result provide useful guideline to improve public transit and infrastructure for demographic needs and seasonal patterns.

References

[1] K. Pandey and R. Sharma, “Correlation and time series analysis in urban planning,” *Urban Studies Journal*, vol. 58, no. 7, pp. 1–15, 2021. DOI: `10.1177/0042098021993234`.

[2] C. Spearman, “The proof and measurement of association between two things,” *American Journal of Psychology*, vol. 15, no. 1, pp. 72–101, 1904. DOI: `10.2307/1412159`.

[3] R. J. Hyndman and G. Athanasopoulos, *Forecasting: Principles and practice*, 2nd. OTexts, 2018, Retrieved from `https://otexts.com/fpp2/`. [Online]. Available: `https://otexts.com/fpp2/`.

[4] National Highway Traffic Safety Administration (NHTSA), *Traffic safety facts annual report*, `https://www.nhtsa.gov/`, Retrieved from `https://www.nhtsa.gov/`, 2019.

[5] G. E. P. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time Series Analysis: Forecasting and Control*, 5th. Wiley, 2015.