



## Analysis of Traffic Volume and Population Trends in New York City

Arefa Binti Sumaya -  
23187010

# Introduction



Objective: Integrate traffic and demographic data to improve urban planning and infrastructure.



Main Question: How can population demographics and traffic data guide urban development decisions?

# Data Used

Sources:

- 1. Population Data: StatsAmerica (age, gender, aggregates for NY).
- 2. Traffic Volume Data: NYC Open Data (hourly traffic counts).
- Licenses: Creative Commons Attribution 4.0 (Population) & NYC Open Data Terms of Use (Traffic).

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

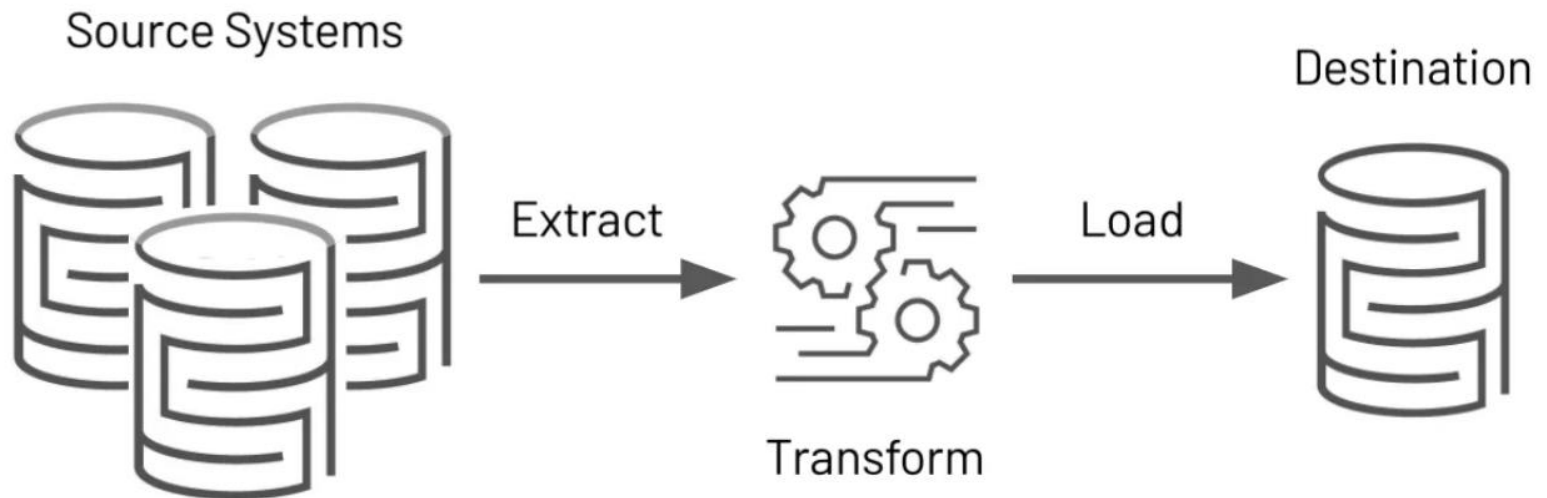
Table 1: Dataset description with key features, value types, and licenses.

## Data Used

Processing:

- Automated ETL pipeline for data readiness.
- Simplification for analysis (e.g., removing unnecessary identifiers).

## ETL Process



# Analysis Methods

## 1. Correlation Analysis:

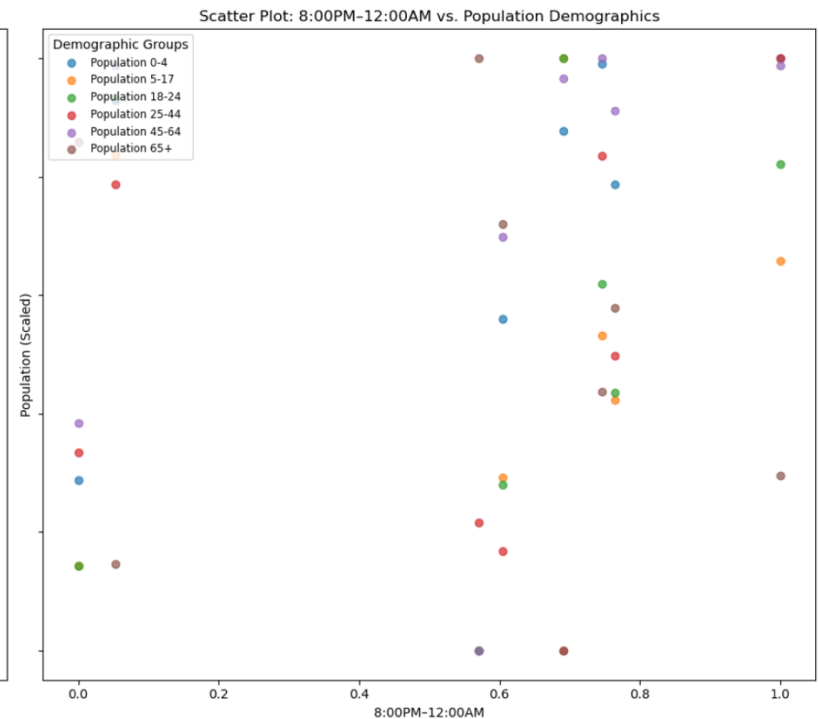
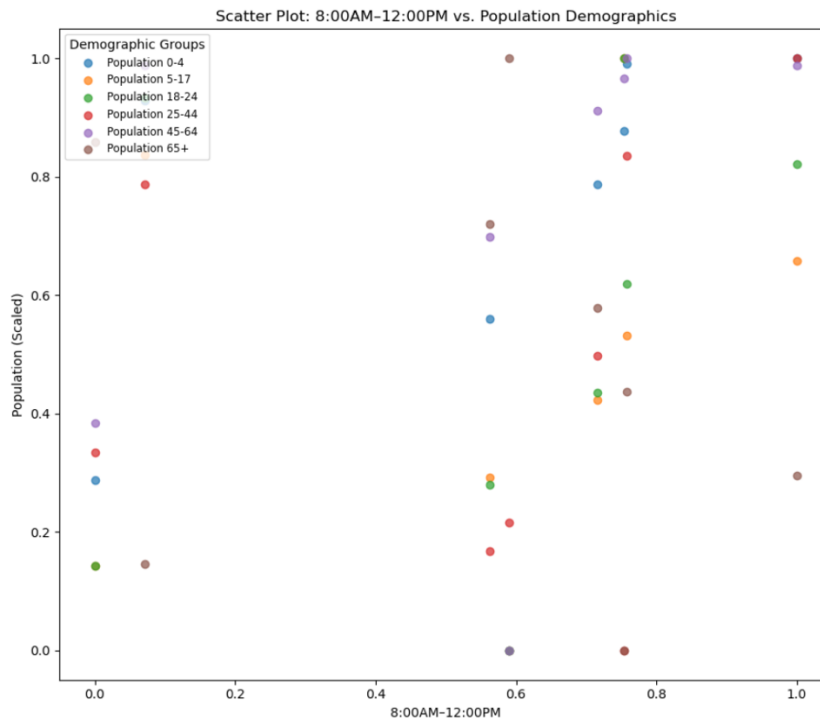
- Measures relationships between age groups and traffic volumes.
- Focus: Peak intervals (8:00 AM-12:00 PM & 8:00 PM-12:00 AM).
- Spearman's Correlation & MinMax Scaling.
- Visuals: Scatter plots & heatmaps.

## 2. Time Series Decomposition:

- Breakdown: Trend, Seasonality, Residuals.

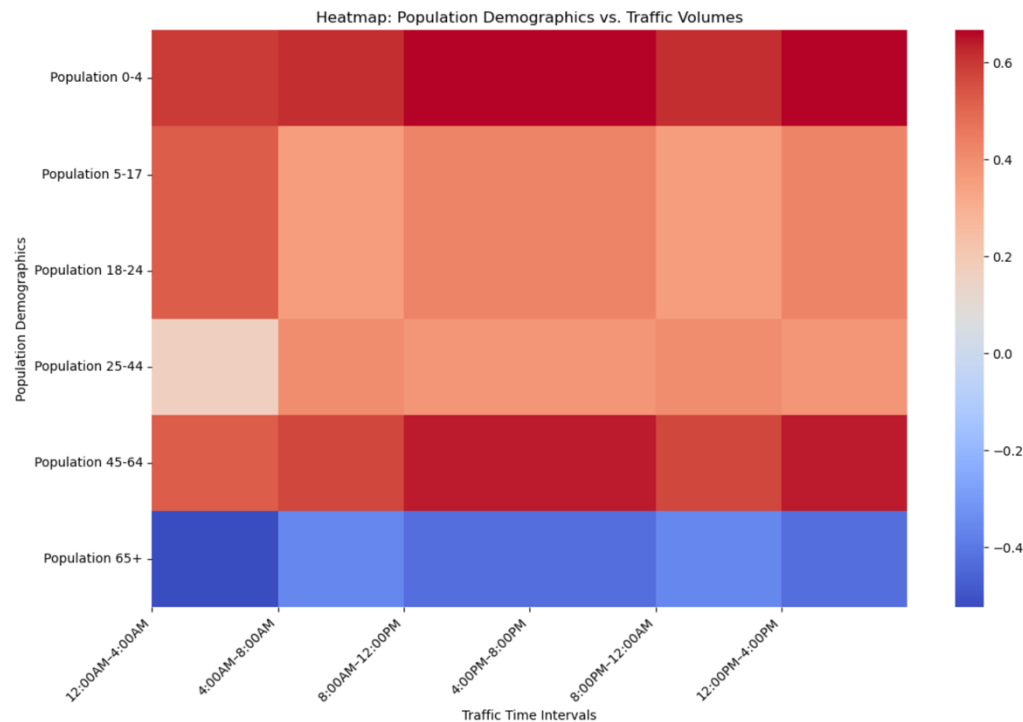
# Correlation Analysis: Scatter Plot Results

- Scatter plots showing the relationship between traffic volumes and population demographics during peak intervals:
- 8:00 AM–12:00 PM: Strong positive relationships for working-age groups (5-17, 18-24, 25-44)
- 8:00 PM–12:00 AM: Moderate correlations.



# Correlation Analysis: Heat Map Results

- Heatmap results showing Spearman correlations between population demographics and traffic volumes across different time intervals:
- High correlations for working-age groups (18-24, 25-44, 45-64) ( $\rho \approx 0.7-0.8$ ) during peak hours.
- Negative correlations for elderly (65+) during peak hours but moderate midday correlations.



# Time Series Decomposition Findings

## Traffic Volume Trends:

- Stable trends over time.
- Seasonal spikes linked to holidays/weather.
- Residuals show irregular disruptions (e.g., construction).

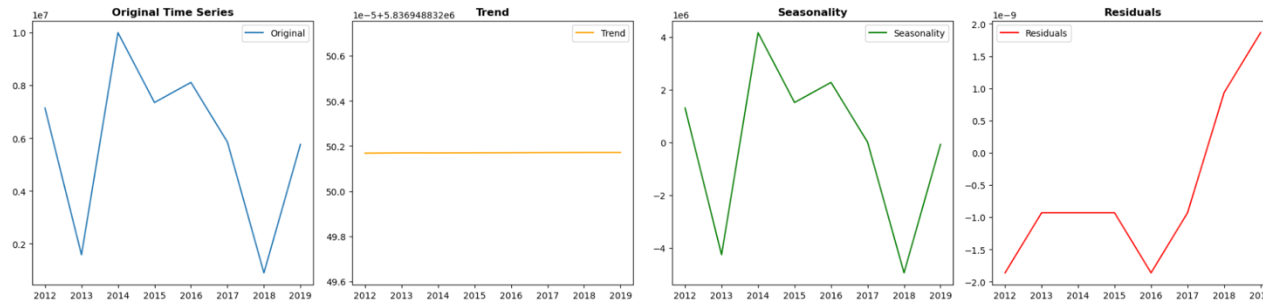
## Population Trends:

- Stable overall trends.
- Seasonal shifts due to migration or temporary factors.

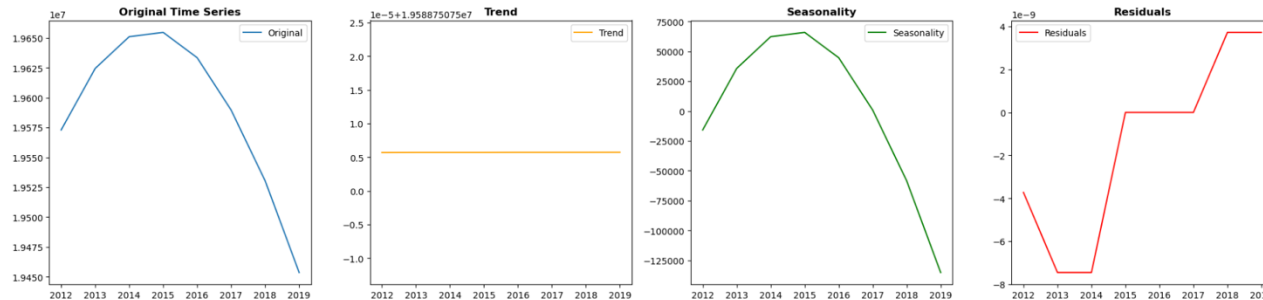
## Implications:

- Seasonal spikes demand adaptive planning.
- Current infrastructure suffices for long-term trends.

Traffic Volume Decomposition:



Population Decomposition:





# Limitations and Future Work



General data scope (regional specificity lacking).



Excludes weather, economic, and future trends.



Regional and external factor integration.



Predictive analysis for future trends.