Traffic Speed and Public Transport Use: Data Science Analysis

Motivation

Urban mobility is the biggest challenge facing modern cities. With the increasing population and traffic volume, it is necessary to analyze the interaction of traffic conditions with public transportation demand for better planning. The logic of traffic congestion in this research is that as the average speed decreases, traffic becomes denser, and as the average speed increases, traffic becomes easier. Traffic density is taken as traffic speed, The main question is: Is traffic congestion related to the amount of public transportation use?

Data Sources

Two primary datasets are being used, and both were downloaded from Istanbul Metropolitan Municipality Open Data Portal (data.ibb.gov.tr):

Hourly Public Transportation Data: Contains rows of passenger numberby transportation line, date, and hour.

Hourly Traffic Average Speed Data: Contains vehicle average speeds on corresponding timestamps.

To keep system efficiency, data for only three months were used: October, November, and December 2024. A stratified sample of 1000 rows per month (total 3000 rows) was extracted from each data set to prevent memory overflow in the free environment of Google Colab.

Methodology

After gathering the data and cleaning it, the following was undertaken:

Column Filtering: Only pertinent columns were retained:

For transport: transition_date, transition_hour, number_of_passenger, line_name

For traffic: DATE_TIME, AVERAGE_SPEED

Datetime Alignment:

Public transport data contained a date column and an hour column. These were merged into a date_time column.

Traffic data already contained a DATE_TIME column, which was renamed to date_time.

Data Merging: The two data sets were merged based on the date_time column.

Exploratory Data Analysis (EDA):

Visual inspection using scatter plots.
Missing value inspection (none significant).

Correlation Analysis:

Pearson correlation coefficient was used in order to estimate linear dependency of AVERAGE_SPEED and number_of_passenger.
Heatmap was employed to represent correlation direction and strength.

Analysis & Findings

Integrated dataset consisted of 3000 rows of traffic and public transport hourly-synced timestamps. The variable of interest was AVERAGE_SPEED (average speed in km/h) and number_of_passenger (total passenger boardings).

Scatter Plot

The trend of number of passengers vs. average traffic speed showed no pattern or trend. The points did not seem to be well separated with increasing or decreasing pattern, suggesting there is no correlation.

Correlation Results

Pearson correlation coefficient was -0.0000 and corresponding p-value was 1.0000.

These suggest:

There is no correlation of public transport usage with traffic speed in alinear manner.

The result is not statistically significant; we cannot reject the null hypothesis.

This suggests, based on sample data, traffic congestion (in terms of lower average speeds) does not significantly affect the use of public transport.

Correlation Heatmap

A heatmap for a correlation matrix also confirmed results above. The heat map in colorcoded form showed neutral colors and values close to zero, which, visually, attested to the lack of correlation.

Limitations & Future Work

Sampling Limitation: System constraints allowed analysis only of 3000rows. Processing the entire data could give slightly different results.

Feature Limitation: Only two variables were considered. Control for day-of-week, weather, or line types would be beneficial for fine-tuning.

Modeling Opportunities: Models using regression or classification canbe considered in future research for the prediction of transport use under various traffic and external conditions.

Conclusion

This project is a data analysis project that includes the processes of data collection, integration, cleaning, analysis and interpretation. Contrary to natural assumptions, the analysis did not reveal a strong statistical correlation between public transport usage and average traffic speed in Istanbul in the 4th quarter of 2024. The months when schools were open and public transport alternatives were evaluated instead of walking were taken for the data analysis. As a result of the non-correlation, we can say that similar metro works in areas with heavy traffic will not yield any results. In fact, there may be more traffic during the construction period. This finding emphasizes the importance of evidence-based urban planning and multi-dimensional data analysis in public policy.