DSA210 Final Report
Ece Akkozak 31010

This study analyzes how weather conditions affect traffic patterns in Istanbul, with hourly data of January 2024. As the case in the large cities, Istanbul is known for its heavy traffic, and unfavorable weather conditions often make things worse. The goal was to find the effect of rain, temperature, wind, and visibility parameters on traffic flow rate. Initially, I wanted to combine both traffic and public transport datasets, but unfortunately, there was a lot of information missing from the public transport dataset. Therefore, I focused only on the weather and traffic data for a month. I collected traffic data from IBB, and weather information from Visual Crossing. After cleaning the data and merging the datasets, I had more than 957,000 records. This was enough to analyze traffic speed patterns and test various hypotheses about traffic speed. All analysis and visualizations were run in Jupyter Notebook (.ipynb), and all the graphs included in this report were created there.

Exploratory Data Analysis (EDA)

My first step was to look at the data with graphs and measurable values. Box plots and histograms allowed me to analyze how various parameters, such as rainfall and temperature, fluctuated over the period. I also checked traffic volume for both hourly and daily observables.

Features

To make better use of the data, I defined new attributes such as time of day, day of week, and temperature and precipitation bins. For example, I created an attribute called is_weekend that distinguishes between weekday and weekend. I also converted the temperature and precipitation values into labeled bins such as "low," "medium," and "high" to clarify the visualization of these attributes.

Hypothesis Testing

I used correlation and p-value tests to determine which variables had a tangible effect on traffic. Weather factors such as rain, visibility, and wind had weak but significant effects. Traffic-related factors such as minimum and maximum speeds had strong effects on average speeds.

Machine Learning

Based on the data provided, I constructed a linear regression model for traffic speed prediction where data from January 1 to 24 was used to predict traffic speeds from January 25 to 31. Given that the data relationships appeared linear, this model was preferred due to its simple interpretation.

Outcome:

Effectiveness Of The Model

- R^2 Score: 0.8257
- RMSE: 0.66 km/h
- MAE: 0.61 km/h

The built-in prediction of traffic speeds proved to be accurate, as the model fitted the daily value changes strongly and behaved optimally during predicted speed reductions on busy days. Most of the daily predicted values were within 1 km/h of the actual figures.

Outputs

• The speed decrease is minimal when it rains or when more vehicles are added than normal conditions.

• When the minimum and maximum speeds are set, the average speed is reasonably estimated.

• Traffic congestion occurs on all routes during rush hour.

• Speed is slightly affected by weather conditions.

This project taught me how to clean and combine large datasets to address real-world problems. Many seemingly insignificant patterns can help with predictions. The weather effect, while not the strongest component, was strong enough to improve the model. There is also some validation of the results; rain and traffic volumes lead to slower movements. If I had more time, I would have included map data to identify areas with the most traffic. More feature interactions and nonlinear models would also benefit from elaboration. Such as:

• Apply time-series models and maps for field-based analysis.

• Incorporate more accessible machine learning algorithms to track detailed patterns.

• Examine results from other months or years to analyze trends over time and assess the impact of holidays or other important dates.

• With a full year of data, it is easier for the model to learn broader trends, which increases accuracy, especially with most machine learning models. However, this creates a very large dataset that will require more sophisticated methods to manage and process big data efficiently.

This project showed that weather information can contribute marginally to predicting traffic speed. The model performed well and the results could be important for managing traffic in Istanbul. I also learned how to use data science tools to clean, analyze, and model real-world data. The feedback I received helped me improve my previous project. As a result, this technique constructively blended learning with real-life problem solving.