Final Report Ece Akkozak 31010

This project looks at how weather affects traffic in Istanbul, using hourly data from January 2024. Traffic in Istanbul is often heavy, and bad weather can make it worse. The goal of the project was to find out if conditions like rain, temperature, wind, and visibility have an effect on how fast traffic moves. At first, I tried to use both traffic and public transport data, but the public transport data was incomplete. So I focused only on traffic and weather data, which covered the entire month. I used traffic data from IBB and weather data from Visual Crossing. After cleaning and merging the datasets, I ended up with over 957,000 records. This gave me enough information to study patterns and test predictions about traffic speed. All analysis and visualizations were conducted in a Jupyter Notebook (.ipynb), and plots used in this report were exported from there.

Exploratory Data Analysis (EDA)

I started by exploring the data using graphs and summaries. Boxplots and histograms helped me see how variables like rain and temperature changed over time. I also looked at hourly and daily patterns in traffic.

Feature Engineering

To make the data more useful, I created new features such as hour of day, day of week, and bins for temperature and precipitation. For example, I added a binary feature called `is_weekend` to capture differences between weekdays and weekends. I also converted temperature and precipitation values into labeled bins (e.g., "low", "medium", "high") to help visualize their effects more clearly.

Hypothesis Testing

I used correlation and p-value tests to see which variables had a real effect on traffic. Weather factors like rain, visibility, and wind showed weak but statistically significant effects. Traffic factors like minimum and maximum speed had strong effects on average speed.

Modeling

I built a linear regression model using data from January 1 to 24 to predict traffic speed for January 25 to 31. I chose this model because the relationships in the data seemed mostly linear and it was easy to interpret.

Results:

Model Performance

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

- MAE: 0.61 km/h

The model predicted traffic speeds with good accuracy. Most daily predictions were within 1 km/h of the real values. It followed trends well, such as speed drops on busy days.

Outputs

- Traffic slows down slightly when it rains or when there are more vehicles.
- Minimum and maximum speeds are strong predictors of average speed.
- Rush hours clearly affect traffic flow.
- Weather has a small but noticeable effect on speed.

Several of these trends are shown in the visualizations. For example, the boxplot of average speed by precipitation (see Figure: `avg_speed_by_precip_bin.png`) shows a clear drop in speed during heavy rainfall.

This project helped me better understand how to clean and combine large datasets and use them to answer real-world questions. I learned that even small patterns can be useful for making predictions.

Although weather effects were not very strong, they still helped improve the model. The results also matched real-life experience—rain and traffic volume do slow things down.

If I had more time, I would try more advanced models like Random Forest and also include map data to see where traffic is affected the most. I would also explore feature interactions and non-linear models that could capture more subtle effects.

- Use time-series models.
- Try more machine learning models to capture complex patterns.
- Add spatial data for area-based analysis.
- Compare results with other months and years to better understand seasonal patterns and the effects of special dates like holidays.
- Since the current dataset covers only one month, it does not fully reflect long-term or seasonal variations in traffic behavior. For example, weather and traffic patterns can change significantly during public holidays, school breaks, or summer months.
- Using a full year of data would allow the model to adapt to broader trends and improve prediction accuracy, especially for machine learning models. However, this would also lead to a very large dataset, requiring more advanced techniques for handling and processing big data efficiently.

This project showed that weather data can help predict traffic speed, even if the impact is small. The model performed well, and the insights could be useful for planning traffic systems in Istanbul. I also learned how to use data science tools to clean, analyze, and model real data. The feedback I received helped me build a better version of my earlier project. Overall, this was a useful and meaningful learning experience that allowed me to combine technical skills with real-world problem solving.