**DSA 210 Project**

Through mixing past traffic index data with daily weather data, which is fetched by the Open-Meteo API, that DSA210 project checks how the weather influences traffic in Istanbul. The aim is to study how weather scenarios such as rain, temperature, wind speed, and humidity influence the ratio of traffic and to form a model that can predict that days will have a lot of traffic in the future. The aim of that project is to learn more about how people move around in the real-world scenario and how external parameters influence how people act in traffic. Those consequences can support traffic forecasting, smart city infrastructure planning, and urbanization processes.

Two main websites provide the data I need to handle for that aim. The traffic information which is fetched from a CSV file that is briefly traffic\_index.csv. That file has daily average traffic index values for Istanbul over the years. That knowledge was uploaded manually to Google Colab file section to analyze. The Open-Meteo API was used to fetch the weather situations. That is a free and open-source weather API service that does not require an API key for personal use. By providing the coordinates of Istanbul and the date range that was very crucial , past daily weather factors such as temperature, precipitation, wind speed, and humidity were gotten and saved to an Excel file.

metin, ekran görüntüsü, yazı tipi, sayı, numara içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.

diyagram, öykü gelişim çizgisi; kumpas; grafiğini çıkarma, çizgi içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.

So, in this project I generally used a traffic collection with 3332 amount daily savings from Istanbul city. Each savings illustrates the lowest, highest, and average traffic levels featured for a specific day. A datetime style that takes into consideration time zones is used to preserve the date field. A sample of the data illustrates that the daily average traffic index values are from 23 to 58, that illustrates that the amount of congestion changes and replaces itself within times. The average traffic indexing figure illustrates there is a right skew based distribution, with most days clustering about an average index of 30 and only a few days having very high levels of traffic. Based on that trend, it looks like traffic amounts are regular, but really bad ones do not happen very often.

metin, makbuz, ekran görüntüsü, yazı tipi içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.

The Open-Meteo API was used to collect data about the weather in Istanbul with given coordinates. That API gives away free historical weather data freely. So I got daily averages for highest and lowest temperatures, total precipitation, wind speed, and relative humidity in Istanbul with a specific date range that fits with my traffic data's time range. That instance illustrates standard weather conditions for early August 2015, with highs around from 28 degree to 30 degrees, little to no rain, and mild wind. That API-based data collection method ensured that there was a steady and reliable source for appending to the traffic knowledge and checking how the weather can influence the ratio of traffic in Istanbul.

metin, ekran görüntüsü, sayı, numara, yazı tipi içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.

After that, these two datas are merged with using .merge method.

diyagram, öykü gelişim çizgisi; kumpas; grafiğini çıkarma, çizgi içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.

metin, ekran görüntüsü, çizgi içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.

ekran görüntüsü, metin, çizgi, diyagram içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.

metin, ekran görüntüsü, dikdörtgen, kare içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.

The graphs provide me crucial knowledge about how the weather influences the traffic density in Istanbul city. The average traffic indexing histogram illustrates a right-skew distribution, with middling amounts of traffic happening most days. We can see that there is a very small positive trend in the scatter, which is responsible for comparing the amount of rain to the traffic index. That mainly suggests that traffic can get worse on rainy days, but that range is very big. Moreover, the scatter plot of wind speed vs. traffic index illustrates a small but noticeable upward trend, that is suggesting that higher wind speeds can be an important link that is going to more traffic. Finally, the correlation heatmap illustrates that the minimum, maximum, and average traffic indices are strongly positively correlated with each other. Weather parameters such as rainfall and temperature, on the other hand, are positively correlated with traffic congestion but not in as strong a way. Those initial studies back up the idea that weather has a noticeable, but not major, influence on how people drive every day.

metin, makbuz, yazı tipi, cebir içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.

Two different statistical hypothesis tests I have done to simulate how weather and time of day influenced the traffic density. The first test checked whether more people drive on rainy days. The consequence is statistically significant, with a roughly t-statistic of 7.3605 and a p-value of 0. That means that rain is positively bound to high levels of average traffic index. The second hypothesis checked whether traffic trends are different on weekends when we compare them with weekdays. The Mann-Whitney U test illustrated a U-statistic of 467,841.5 and a p-value of 0; that generally means there is a significant difference in the traffic density on weekends when we compare it with weekdays. Those results support the idea that both the weather and the time of day influence the density of the roads.

metin, ekran görüntüsü, yazı tipi, sayı, numara içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.

metin, ekran görüntüsü, çizgi, sayı, numara içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.

A Random Forest Classifier was used as a main method to form an ML model that represents, from the weather, whether a specific day would have too much traffic. The model was 57% correct, and both its precision and recall scores were roughly around 0.57. That means it did a good job but not a perfect one. With an ROC AUC score of 0.5904, the model looks like it does a little better than selecting at random. The confusion matrix illustrates that the number of correct and incorrect guesses is about the same for the group, which is days with a lot and a few. The most crucial indicators, according to feature importance analysis, were temperature variables in both max and min and speed of wind. Humidity and precipitation came in second and third. That illustrates that temperature and wind parameters are more crucial than rain as alone in specifying how crowded the roads are.