Weather Impact on Traffic Accidents in Istanbul in 2024

Nazlı Uçar 32042

Introduction

The goal of this project is looking into the relationship between weather conditions and daily traffic accidents in Istanbul in 2024, whether it has a quantitative effect on the accidents. As one of the million drivers in Istanbul, traffic safety is one of my biggest daily concerns. I always try to maximize my safety, but sometimes some factors are not determined by the drivers. In this project, I tried to figure out whether weather conditions are one of the factors or not.

Data Cleaning

I found my traffic accidents data in Istanbul in 2024 from İBB's website, and I found the daily weather conditions from Visual Crossing website. I prepared the raw data of daily traffic accidents and weather conditions in Istanbul to make my analysis clearer. At first, the traffic accident data was not in a daily format but was ordered by announcement with lots of unnecessary details such as the number of closed lanes, the exact location, etc. Additionally, it was in a csv format. First, I changed the csv format into an excel file, then I changed the format of the data into daily and deleted the unnecessary information. I did similar steps to the weather data. I simplified the weather conditions into 3 categories: temperature, wind gust and humidity percentage. Additionally, I transformed the txt files into excel files. I did all of these with python codes.

Exploratory Data Analysis

With Jupiter Notebook, I created descriptive statistics table, correlation matrix, line plot, histograms of numeric features, bar charts with y axis having the accidents rates and the x axis having the weather feature, boxplots, and scatterplots. I observed that the average of the accidents happening in Istanbul were 38 and it ranged between 20-60 values. A variety of weather features happened in Istanbul in 2024, so we can say that many feature changes have been observed which help us to determine the features which influences accidents. Correlation analysis indicated that there is no strong relationship between the features and the accidents; however, wind gust and humidity had a weak correlation with accident rates. Line plots confirmed this due to showing no trend. Histograms show that the accident counts look

normally distributed centered around 40, temperature was bimodal with peaks of approximately 43 and 53, wind gust being left skewed and humidity looking like normally distributed which centered around 55. Bar charts indicated that accident counts remained relatively stable across temperature ranges but slightly decreased as wind and humidity levels increased which explains the correlation matrix. This conclusion was also confirmed by boxplots, where higher wind gust and humidity bins showed lower medians and tighter interquartile ranges, indicating fewer and more consistent accident counts in those conditions. Finally, scatterplots confirmed this conclusion once again by highlighting the weak negative correlation between accidents and temperature and wind gust while showing no relationship between accidents and temperature. To conclude, correlation matrix, bar charts, boxplots and scatterplots all agree that there is no correlation between temperature and the accident rates in Istanbul 2024 and weak negative correlation between the accidents and wind gust and humidity.

Hypothesis Testing

To understand and calculate the relationship between each of the weather conditions (wind gust, temperature, humidity) I created two opposite hypotheses per feature with Jupiter Notebook. The null hypotheses of the features were defending the zero correlation between the indicated feature and the accident rates. The alternative hypotheses were defending that there were some correlations. To assess the statistical significance of these hypotheses, I conducted a two-sided p test for each feature with the significance level of 0.05. For temperature, the p value was 0.75, so we could not reject H0. For humidity, p value was 0.0116, and for wind gust, p value was 0.0001. Both were smaller than 0.05, so H0 was rejected. These results match with the EDA results: Wind gust and humidity affected accident rates, but temperature did not.

Machine Learning Techniques

MLT was used to predict the daily number of traffic accidents by using weather conditions data (temperature, wind gust and humidity) in Istanbul 2024. The target variable was continuous, so I applied regression approaches like linear regression, random forest and k-nearest neighborhood. The linear regression performed the best with explaining 20% of the variation. K nearest was the poorest, it performed worse than just taking the average of the rates. Linear regression and random forest had predictions in the similar range of the difference of the actual results by 7-10. On the other hand, K-nearest had a much wider range.

This result concludes that weather has a minor effect on the accident rates, but still there are many factors in the amount of accident rates in a day, more information must be used to predict the daily accident rates accurately.

Findings

My studies conclude that in Istanbul, traffic accidents were not related to temperature as we could not reject the null hypothesis of temperature, but slightly negatively related to humidity and wind gust as it was indicated in EDA and hypothesis testing. As wind gusts and humidity increased, the accidents decreased slightly. We can interpret that the increase of wind gust and humidity can indicate rain. If so, we can conclude that drivers are more careful on rainy days due to harsh weather conditions. Still, these features are not enough to explain the number of daily accidents happened in Istanbul 2024 as we found out in regression models. The results shocked me because I always thought that accidents were related to weather conditions. The harsh conditions make car travel more dangerous due to slippery roads, less clear vision, etc. This study shows that these issues are not the number one cause of accidents.

Limitations & Future Work

The R square value we obtained at the MLT indicates that this data set lacks some information to predict the ratio accurately. To fix this issue, more features can be added such as holidays, whether it is rush hour or not, etc. Furthermore, specifying the target variable might be useful as well such as predicting a specific level of accident rates (high, low). In the future, more work like these additions can be done on this model. I will advance my variables and explore the issue in a different perspective other than weather.

To conclude, this project changed my perspective on traffic safety. Many of us, drivers in Istanbul, tend to blame the weather for our mistakes: The vision is limited, the roads are slippery, etc. This project says that these excuses are not valid, our more cautious behaviors decrease traffic accidents. Additionally, it indicates that Istanbul drivers are more conscious in harsh conditions which makes me feel safer in the traffic. Data driven analysis in urban safety is very important because it can change myths and prejudice on important topics such as this. In the future, more work must be done in urban safety to change the perspectives on significant topics and find key relationships between dangerous situations and their possible factors. By doing so, these situations can be minimized, and public safety can be maximized.