

Seattle car accident severity analysis

Introduction:

In United States around 23 percent of the total accidents that are caused every year is weather related. Wet roadways, having ice or water in the pavement, and the rainfall are the major cause of these accidents.

Drivers can reduce the risk of getting into accident if they are alarmed upfront about the possible road accidents on their chosen route. In this way they can decide either to drive carefully or change the travel plans if possible.

In Seattle City there are 11,083 total crashes in 2019. We opted to predict the possibility of car accidents due to the weather and road conditions in Seattle. To predict the possibility of getting into a road accident in Seattle city based on the weather and road condition. We will also estimate the possible severity of the road accident. The business goal is to reduce the road accidents due to the weather and road conditions in Seattle city using our timely predictions.

Data Section:

Data Description:

The data for this Capstone project is obtained from the Seattle Department of Transportation. It has the data starting from 2004 to present on the dates of occurred accidents. The given dataset contains 166,689 records of data starting from 2004 till the current date. This dataset has many attributes like longitude, latitude, location severity, collision type, affected number of persons, date and time of the accident, weather condition, road condition, and light condition.

Data Exploration:

Given data is cleaned initially by removing the unnecessary attributes like OBJECTID, INCKEY, COLDETKEY, REPORTNO, STATUS, EXCEPTSNCODE, SEVERITYCODE, SDOT_COLCODE, and INTKEY. Two new attributes named as WEEKEND and HOD are created by using the date and time of an accident. WEEKEND attribute has the values of True or False based on the fact if the accident happened over the weekend or during the weekday. HOD attribute represents the hour of the day that accident occurred. HOD is calculated by rounding the time of the accident to the lower floor value like 6 AM for the data point of 6:30 AM. By using these two variables we can also take the effect of rush hour traffic along with weather and road condition into our model estimation. The final selected attributes to use in the model are SEVERITY CODE, LONGITUDE, LATITUDE, WEATHER, ROADCOND, LIGHTCOND, WEEKEND, and HOD.

After cleaning the data, we have the 144481 data points in total. We opted to remove the rows based on the null values. In our first step of exploratory data analysis we found the following facts

Exploratory analysis of the final attributes

WEATHER:

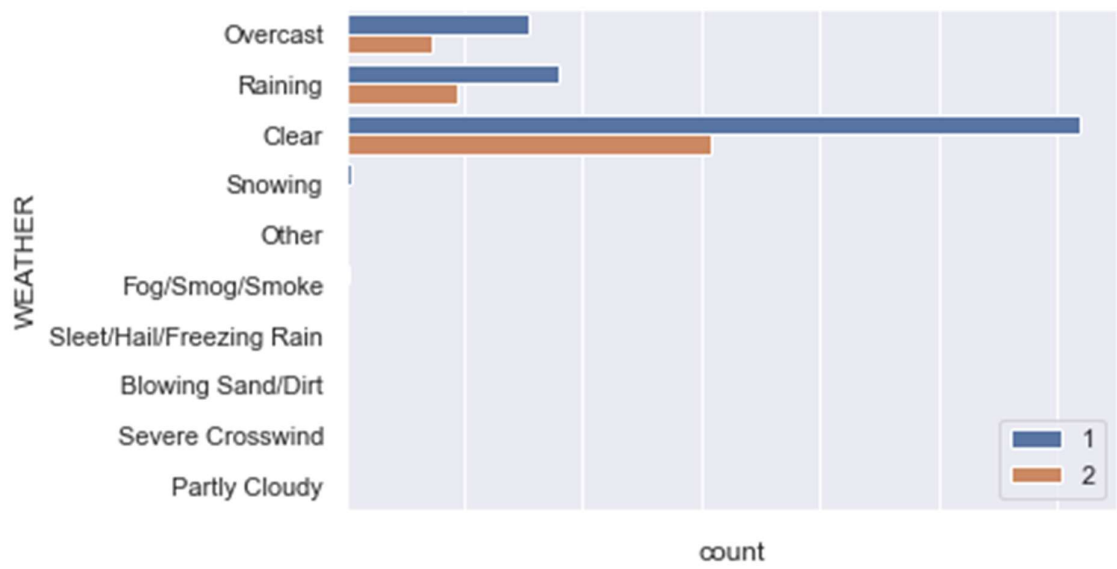


Figure 1 Weather condition Vs Accident Severity

Above figure clearly shows that most road accidents happened in the clear weather condition followed by rain. Severity number 1 reflects the property damage only accidents and the severity number 2 represent the accidents that involve injury collisions.

ROADCOND (Road Condition):

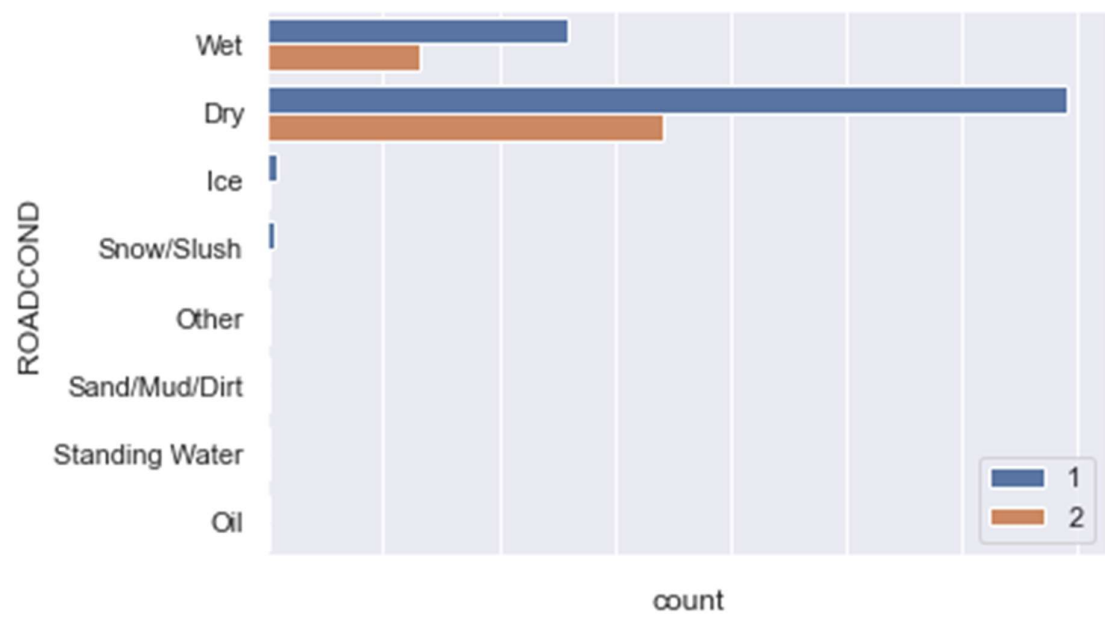


Figure 2 Road Condition Vs Accident Severity

In figure 2 it shows that higher number of road accidents happened in the dry road condition followed by wet road condition.

LIGHTCOND (Light Condition):

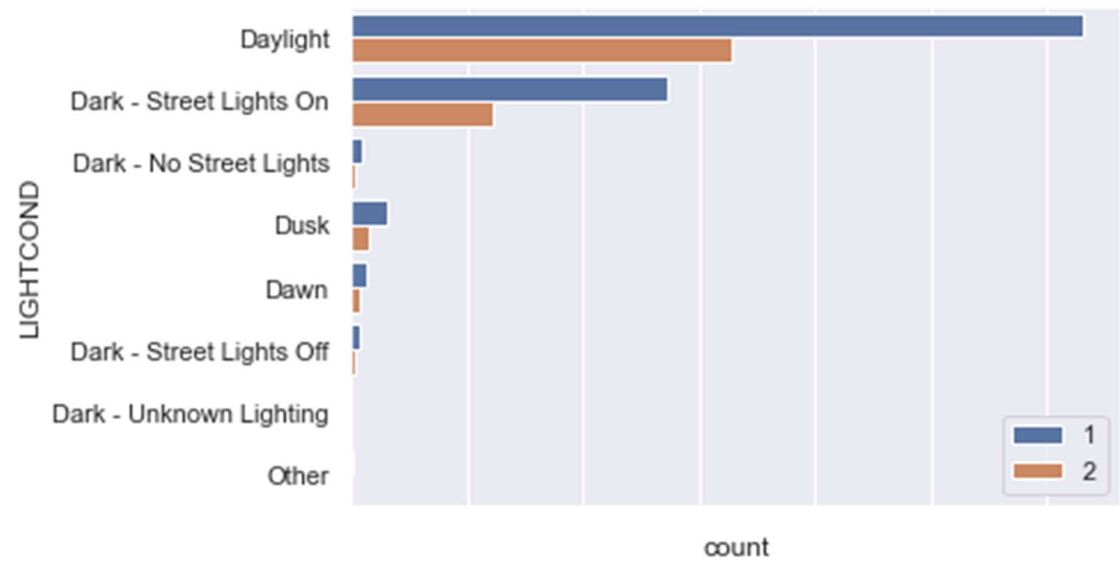


Figure 3 Light Condition V Accident Severity

From the Figure 3 it shows that most of the road accidents happened during the day light and where the property damage only collisions are at higher rate compared to the injury collisions.

WEEKEND:

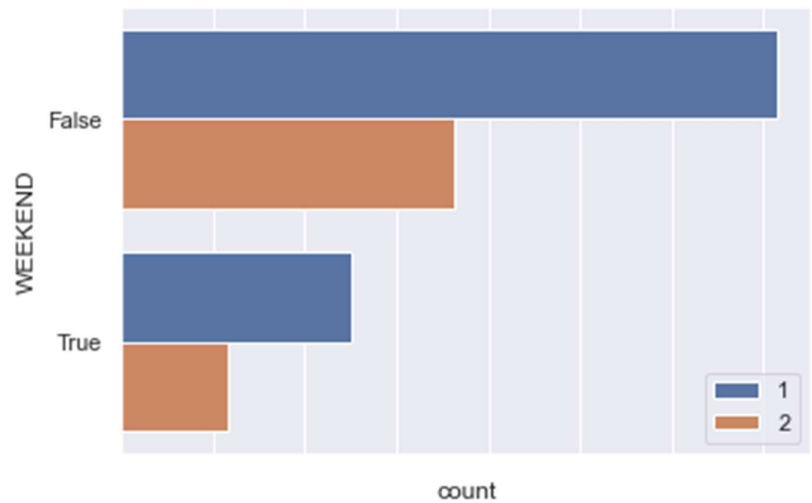


Figure 4 Weekend Vs Accident Severity

From Figure 4 it is cleat that on an average the number of accidents over the weekends are lower compared to the weekdays.

HOD (Hour of the Day)

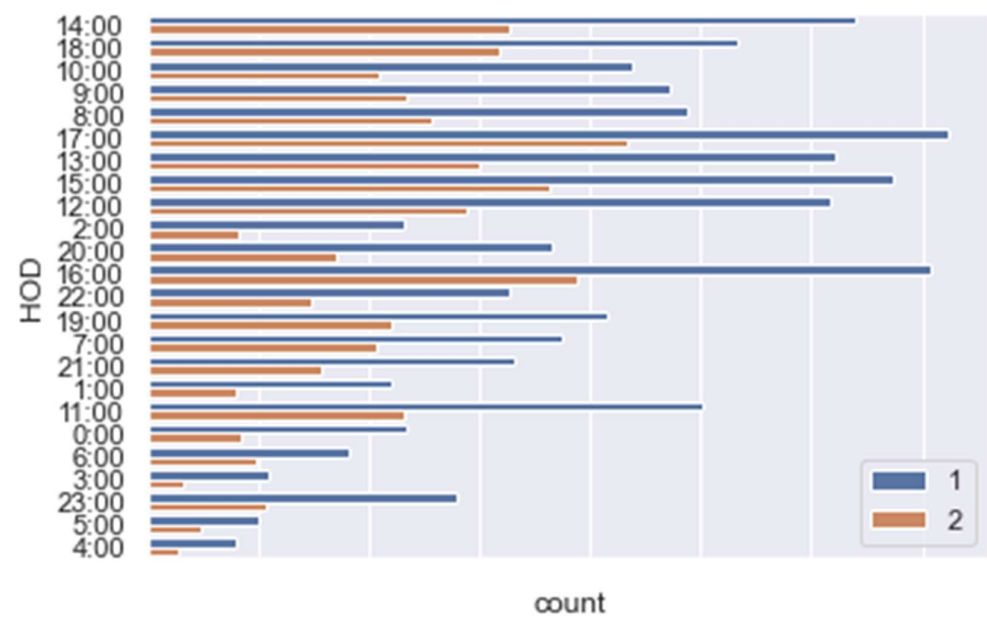


Figure 5 Hour of the Day Vs Accident Severity

In the Figure 5 it is shown that increased number of road accidents happened between the day times of 12 – 18 hours.

From our initial exploratory analysis, its is clearly evident that combination of weather conditions and the time of the day can be used to predict the possibility of an accident.