**Introduction/ Business Problem**

In this project, we are given a dataset about accidents happened in a Seattle. Information such as number of vehicles involved, number of people involved, and severity of the accident are provided. From this dataset, we are interested to know how weather, light condition, road condition and other criteria affect the severity of accident. Based on the insights generated, Seattle government could take action to reduce accidents happened in Seattle. In the end of this project, we will build a classification model to predict the severity of accident based on the information.

**Data**

This dataset contains 38 column and 194673 rows. We only select 11 columns which we think contribute more to determine the severity of accidents. The 11 columns are SEVERITYCODE, LOCATION, SEVERITYDESC, COLLISIONTYPE, PERSONCOUNT, VEHCOUNT, JUNCTIONTYPE, WEATHER, ROADCOND, LIGHTCOND and SPEEDING.

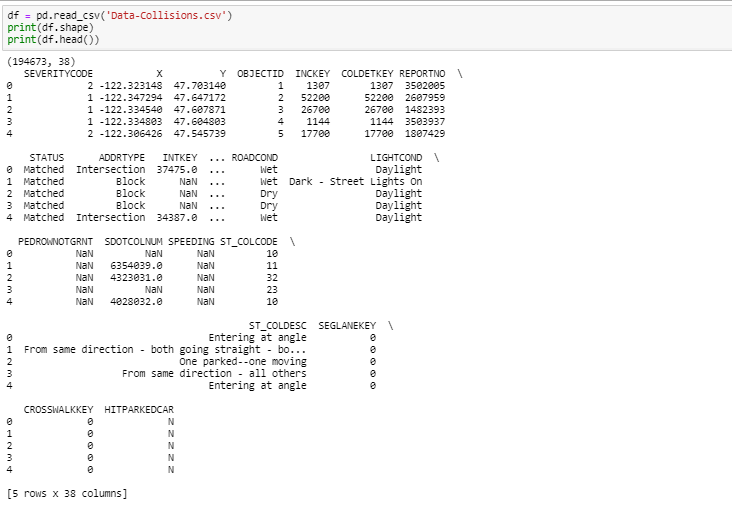


Figure 1. Full Dataset

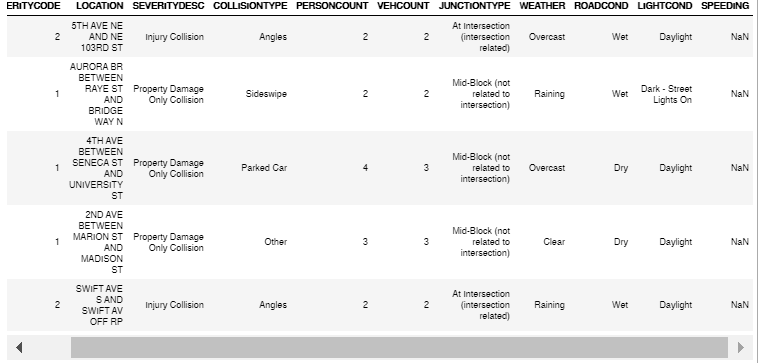
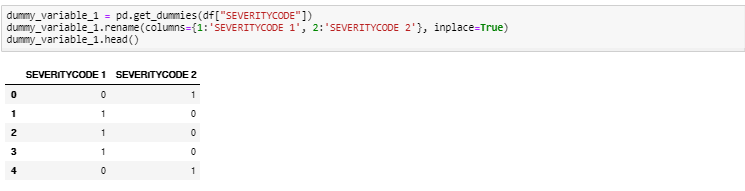


Figure 2. Dataset with selected columns

We split the SEVERITYCODE column into 2 by getting dummy variables for future usage.



In the next section, we will discuss the top 20 locations of most accidents happened, and how weather, light condition and road condition affect severity of accident by generating pie charts respectively.

**Methodology**

1. **Top 20 locations of most accidents happened in Seattle**

We are interested in column LOCATION, SEVERITY CODE 1 and SEVERITY CODE 2 from the dataset. By counting number of accidents of severity code 1 and 2 happened in each location, we plot a bar chart and get the following result.

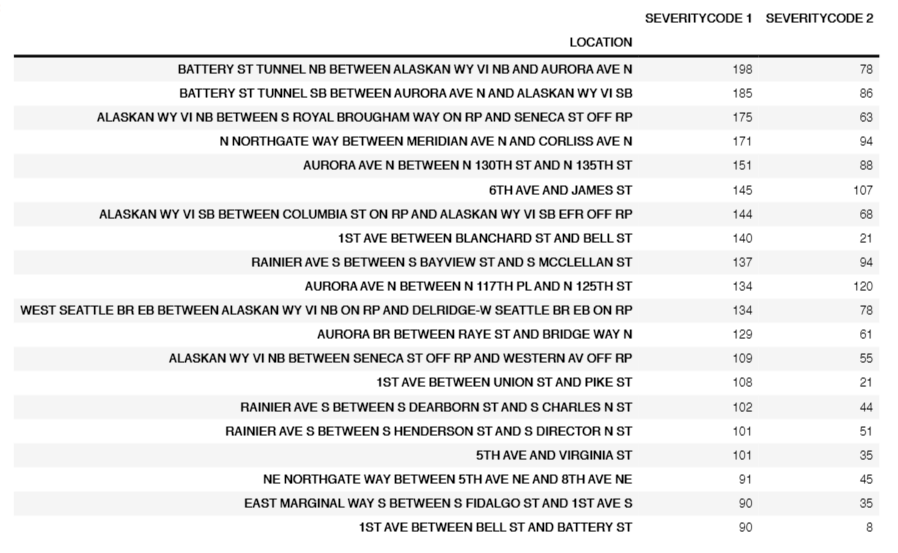


Table 1. Top 20 locations of most accidents happened in Seattle

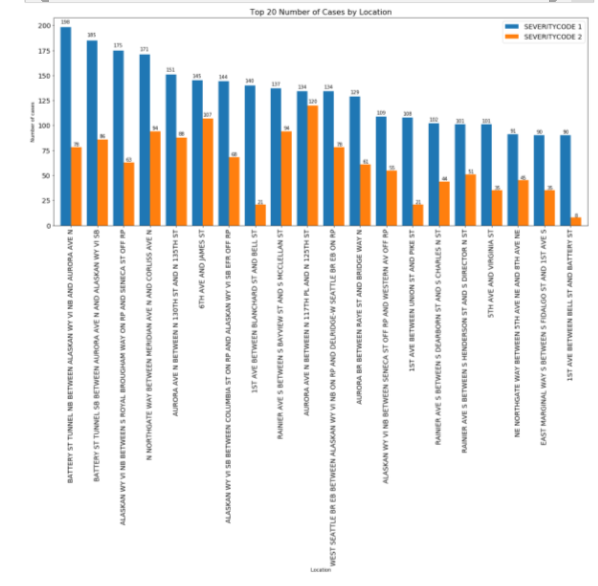


Figure 3. Top 20 locations of most accidents happened in Seattle in bar chart

Based on the result, we can conclude that Battery St Tunnel has the most cases happened. Seattle government should raise awareness of people passing by that area by setting up sign board and speed limit.

1. **Number of accidents under different weather condition**

Before analysing the dataset, we predict that more accidents happened under bad weather condition than good weather condition. Surprisingly, most of the accidents happened under clear weather, followed by overcast, raining and unknown weather condition.

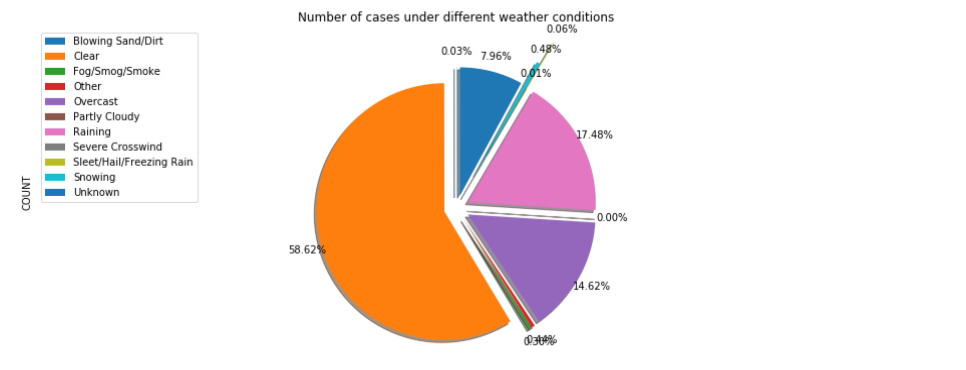


Figure 4. Number of accidents under different weather condition pie chart

1. **Number of accidents under different light condition**

The light condition in this dataset includes Dark - No Street Lights, Dark - Street Lights Off, Dark - Street Lights On, Dark - Unknown Lighting , Dawn , Daylight, Dusk, Other and Unknown. Based on the following pie chart, we could conclude that most of the accidents (61.29%) happened during daytime, 25.60% happened during nighttime with street lights on and 7.11% of unknown light condition.

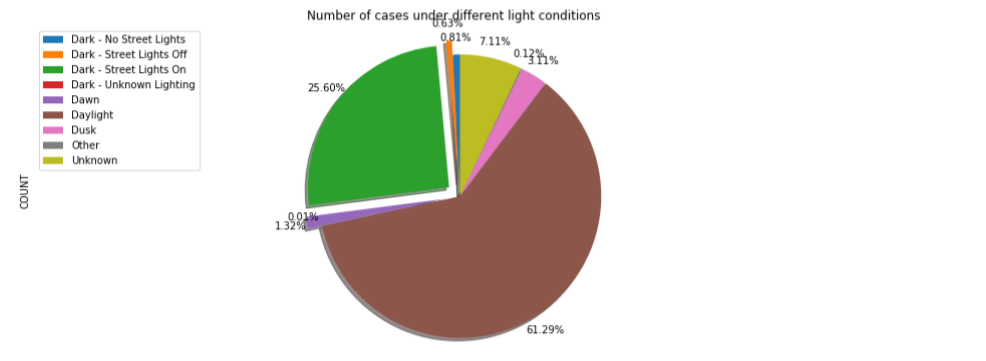


Figure 5. Number of accidents under different light condition pie chart

1. **Number of accidents under different road condition**

There are in total 9 road conditions stated in this dataset, which are dry, ice, oil, sand/mud/dirt, snow/slush, standing water, wet and other conditions. Majority of car accidents happened on dry road (65.65%), 25.03% happened on wet road and 7.95% of unknown road condition.

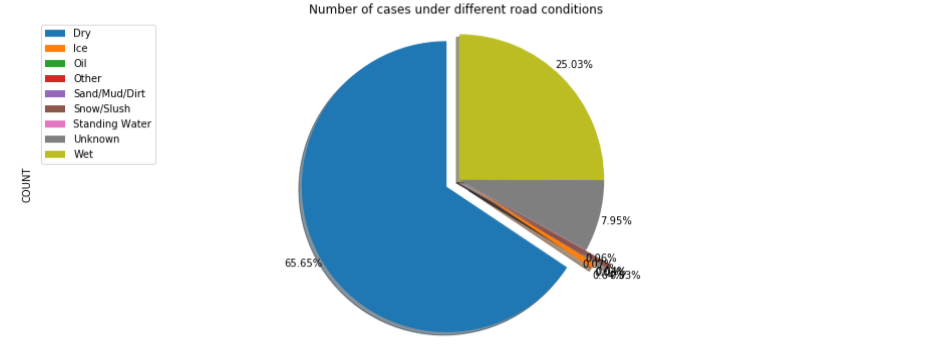


Figure 5. Number of accidents under different light condition pie chart

The result shows that most of the accidents happened during clear daytime on dry road, and Battery St Tunnel is the most dangerous place to drive in Seattle.

1. **Classification Model on severity code (1 or 2)**

Based on information given in the dataset, we decided to build a decision tree to determine the severity code of accident. Columns used as input are COLLISIONTYPE, PERSONCOUNT, VEHCOUNT, JUNCTIONTYPE, WEATHER, ROADCOND, LIGHTCOND, and SPEEDING.

**Result**

Finally, the following graph shows the decision tree with F1 score 0.71 and Jaccard Similarity Score 0.74.A picture containing clock

Description automatically generated

**Discussion**

As the F1 score and Jaccard Similarity Score are quite satisfying, we can conclude that the information stated will affect the severity of accidents.

**Conclusion**

In order to decrease the number of accidents happened, Seattle government could take actions such as setting up signboards on junctions, setting speed limit at dangerous areas and taking care of road conditions.