**Car Accident Severity**

**Introduction:**

Every year the lives of approximately 1.35 million people are cut short as a result of a road traffic crash. People suffer non-fatal injuries, with many incurring a disability as a result of their injury. Road traffic injuries cause considerable economic losses to individuals, their families. These losses arise from the cost of treatment as well as lost productivity for those killed or disabled by their injuries, and for family members who need to take time off work or school to care for the injured. It would be great to warn a person about the possibility in getting into an accident and how severe it would be, so that the person would drive more carefully or may change his/her travel if it's possible. Here, I have chosen 'car accident severity' dataset of Seattle, which has 'SEVERITYCODE' as it's target variable.

SEVERITYCODE has only two values i.e 1 or 2, where 1 implies 'Property Damage Only Collision' and 2 implies 'Injury Collision'.

Obviously, people will be interested in knowing whether there would be an accident or not. Police would be interested in knowing the probability of an accident to occur so that they can take steps to avoid crashes.

**Data:**

Based on the objective, following factors may influence our decision:

1) Conditions of road, light, weather and driver's condition(is he inattention or under drugs/alcohol.

2) Vehicle's speed, condition.

3) Is area safer or not.

4) Is proper parking area alloted or not.

5) Are pedestrains given a way to walk,etc....

To examine factors that cause an accident I have chosen 'car accident severity' dataset. This dataset includes This includes all types of collisions from 2004 to present in Seattle which was provided within coursera.

**Data Cleaning:**

This is an extensive data set from the Seattle Police Department, with over 190,000 observations collected over the last 15+ years. There are few problems with the data set. Firstly, to accurately build a model to prevent future accidents and/or reduce their severity, we will use the following attributes — ADDRTYPE, WEATHER, ROADCOND, VEHCOUNT, PERSONCOUNT,SPEEDING, SEVERITYCODE, INATTENTIONIND, UNDERINFL, etc….

To reduce redundancy I have dropped some attributes like: x,y,OBJECTID,INCKEY,COLDETKEY,STATUS,LOCATION,ETC..

Secondly, there were lot of missing values for which I filled them with frequency as data consists of categorical attributes more.Though some attributes like SPEEDING, INATTENTIONNID,UNDERINF are very important attributes. I still dropped them because they have lots of missing data.

Thirdly, Data is imbalanced for which I have choosen Oversampling .

**Methodolgy :**

I used Jupyter Notebooks to conduct that analysis and imported all the necessary Python libraries like Pandas, Numpy, Matplotlib, and Seaborn. The data was mostly categorical so I stuck to graphical representation to see correlation between various variables. I started by importing the csv file and to prepare the data by dropping, oversampling, dealt with missing values.

After balancing features, and standardizing , the data has been ready for building machine learning models. I have employed two classification machine learning models:

Logistic Regression

Decision Tree

For logistic regression:

log loss: 0.5890148654572129

f1\_score: 0.6770587000629679

For decision tree:

f1\_score: 0.6396229697458525

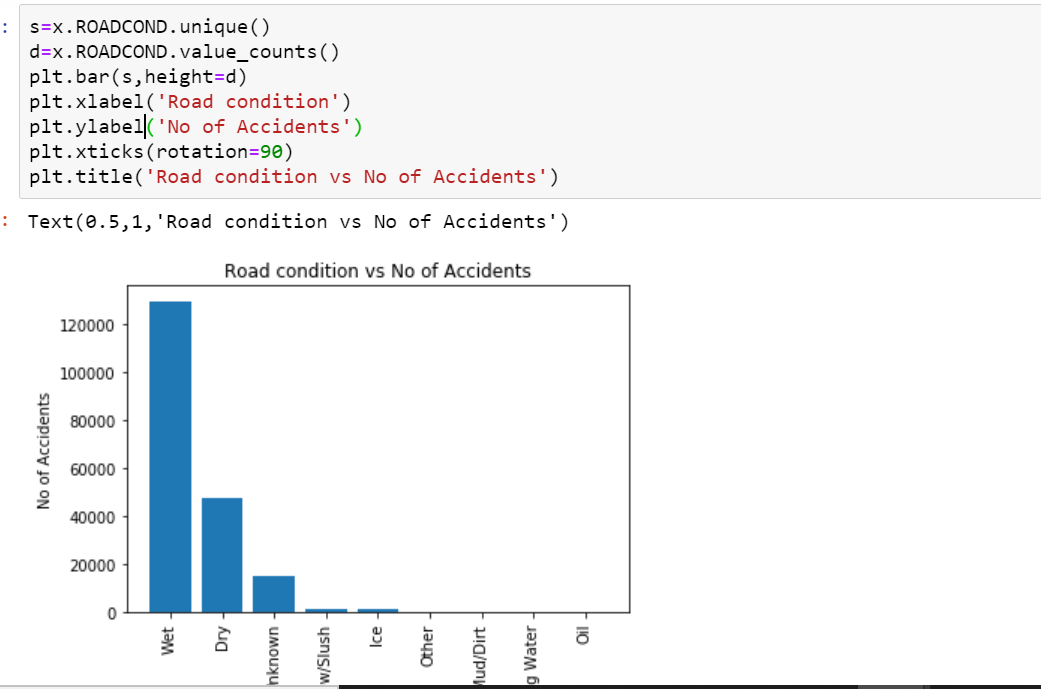
Based on the observation, Logistic regression is the best model to predict car accident sevirity

We can say, road, and light conditions pointing to certain classes, we can conclude that particular conditions have a somewhat impact on whether or not travel could result in property damage (class 1) or injury (class 2).

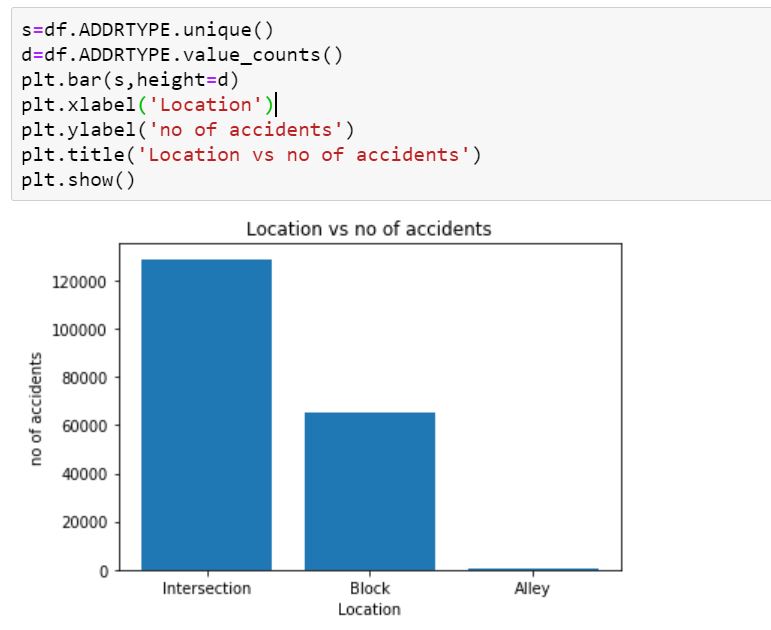
**Results:**

Since most are categorical variables I have done graphical representation to know the corelation .

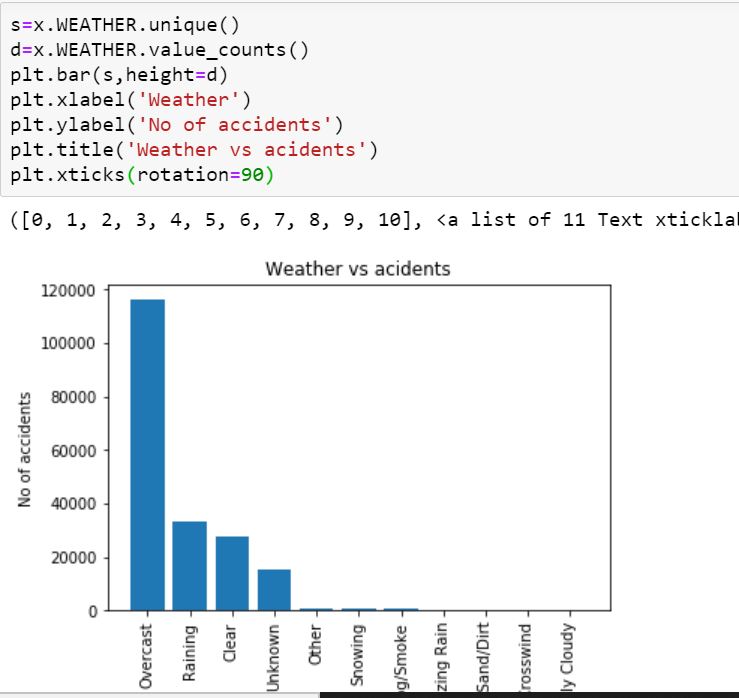
The below figure shows that wet roads contribute the most for accident on compared to other types of roads.



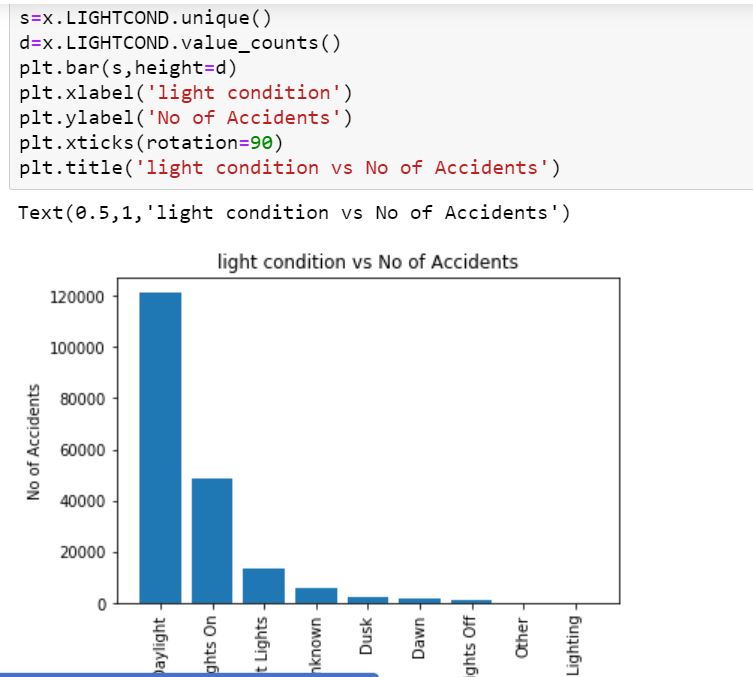
It is also important to find out where most accidents take place. Upon analyzing the data, it turned out that intersections are the most common accident zones. This could be because drivers don’t heed the stop sign, or maybe some intersections can use more stop signs, or maybe there need to be more pedestrian crossings. In any case, this should be an area to look into more in-depth.



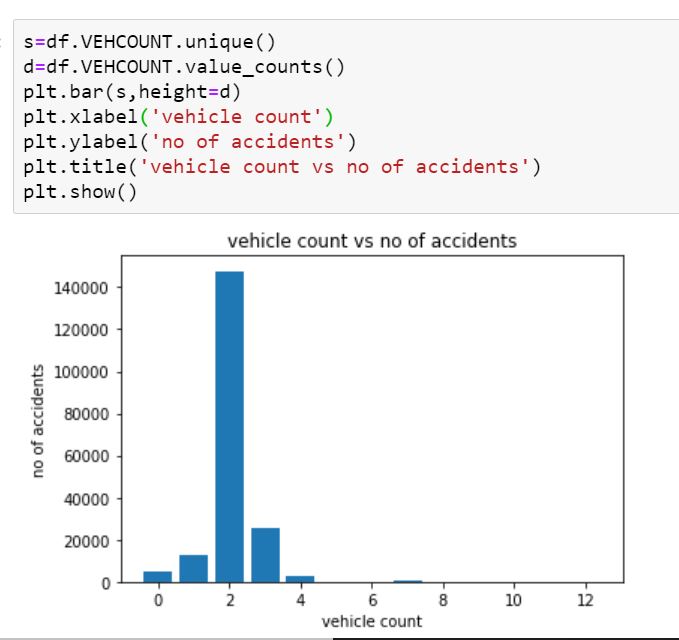
Next in my analysis, I wanted to check how the weather conditions affect the accidents. To my surprise, overcast conditions cause most accidents, rather than rainy or snowy conditions.



Next I examined effect of light on accidents. But, to my surprise again in daylight conditions more accidents occurred.



 I checked the number of vehicles involved in most accidents. I found out that most accidents included 2–3 vehicles at once.



# **Discussion:**

At the start of our analysis, I was trying to figure out the severity and frequency of road accidents based on weather conditions, road conditions, and other factors. Even though our data was a good size, there were a number of missing elements and we needed to clean the data in order to get a good result. We had to drop ‘SPEED’ because there were too many missing elements but I think that is an important factor that should be considered. From the analysis, it is clear that most accidents involve solo drivers, on wet roads, bad weather, at intersections, and are minor in nature. This could be helpful to the police department in understanding where to install more stop signs, or maybe adding cameras to intersections to compel people to slow down. We also live in a technologically friendly world so maybe we can develop some inbuilt technology in our cars that warn us when the road and weather conditions are bad, or the car is approaching a stop sign.

# **Conclusion:**

Although this analysis has given us some good insight, there needs to be a closer inspection of certain other variables. It seems like a lot of these accidents are minor and avoidable. Having said that there is still a considerable amount of loss of property and these findings can be helpful to the Seattle PD in enforcing some new measures to prevent future accidents.