**Introduction to Course era capstone Project**

In this project we will discuss about the Car Severity Accident Prediction of a city in USA.

We will divide this into two parts.

1. Business Understanding.
2. Data Understanding.

**Business Understanding** :-

Car accidents are one of the most common issues found across the globe to be severe. Accidents might sometimes be due to our negligence or due to natural reasons or anything. Sometimes, we might be too lazy or negligent to drive costing our lives as well as the others. Whereas sometimes, due to heavy rain or heavy gales etc. we might unknowingly droop into an accident with the other car. Whatever the reason maybe, car accidents not only lead to property damage but cause injuries and sometimes even leading to people's death.

Road Traffic Accidents cause millions of deaths and injuries every year across the world.

It causes significant economic losses and -3% GDP growth for most of the countries.

To overcome this cause of the accident we need to do some analysis on the existing accident data so that we can identify what are the factors be the reason of a car accident and how we can get rid of it and take a necessary precaution to avoid it as much as we can.

In this case we are trying to predict a classification model on the basis of **Severity Code**.

Insights on factors leading to high accident injury risk provide valuable guidance for public traffic polices.

Also benefits public individuals for safer road trips.

**Data Understanding** :-

Example dataset was used with goal to predict accident severity (code1-property, code 2-injury).

Unmatched data and duplicated columns were dropped.

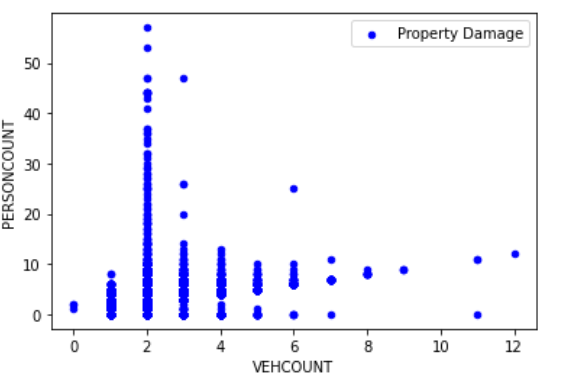
Data rows with missing entries were dropped.

10 Features have been selected to build the classification model.

1. Weather
2. Collision type
3. Address type
4. Road Condition
5. Light Condition
6. UnderInfl
7. PersonCount
8. Vehcount
9. Pedcount
10. Pedcylcount

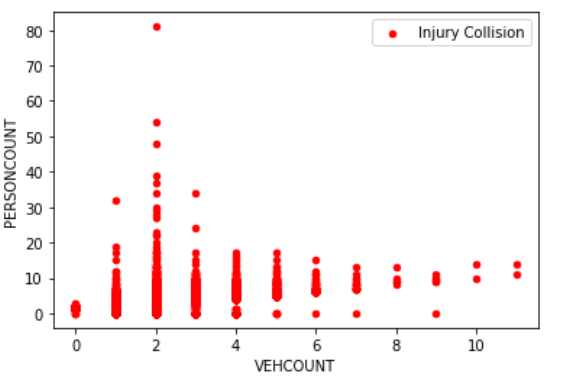
But first of all I needed a lot of analysis to choose the features.

I started my analysis with the vehcount and personcount for the severitycode=’1’(i.e Property Damage).I derived that most of the belongs to vehcount=’2’ i.e Most property Damage happened when Vehcount=’2’

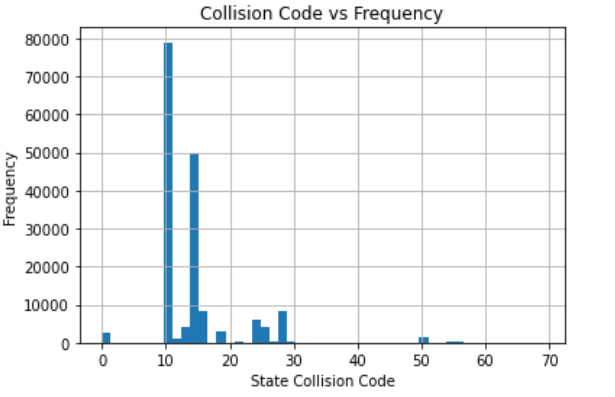


When it comes to severitycode=’2’(i.e Major Injury) we can see that in that case the Major Injury happened when accident occurred between 2 vehicles.

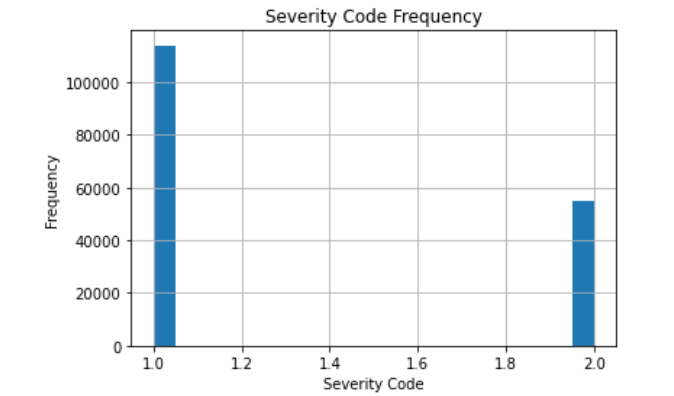
So we can conclude that the overall accident severity project the vehicle count plays an important feature to build the classification model.



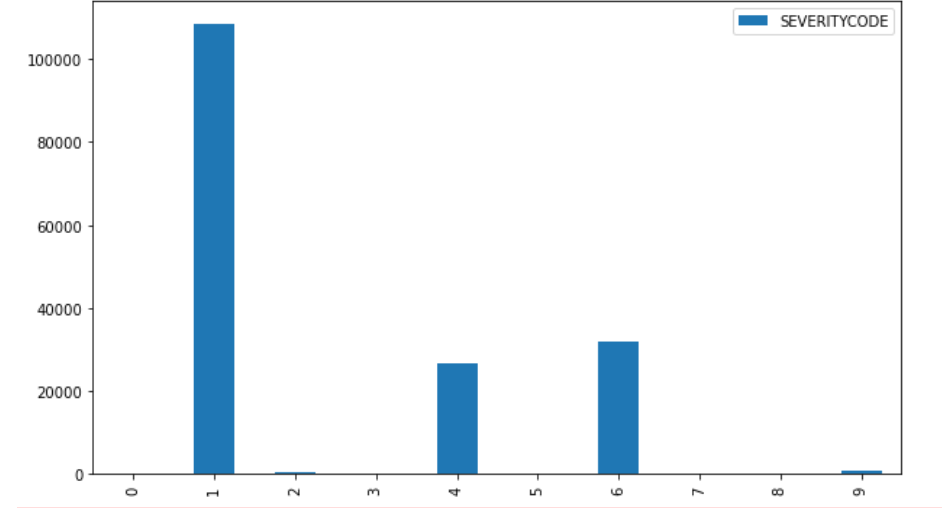
After that I did a comparative analysis between collision Code and frequency and found that the most accident happened on collision\_code=’10’.



After that I wanted to see that what is the count of each accident severity code and found that the major effect on property damage.

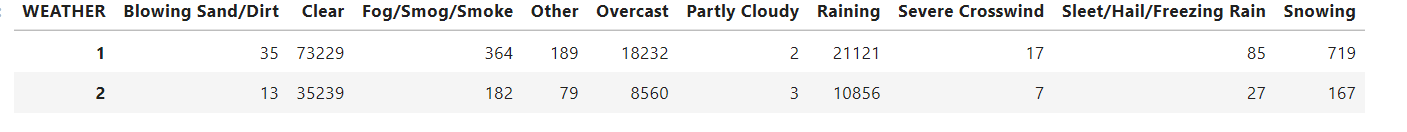


After that I checked with Weather Condition and found that the weather is a major feature which can have serious effect on accident.



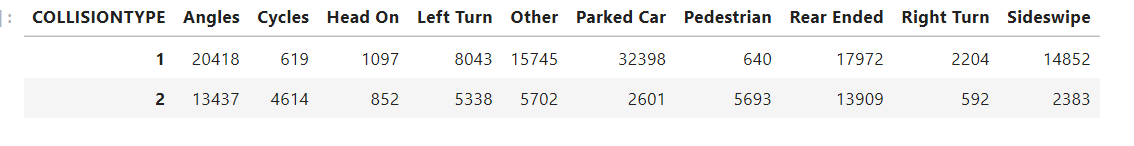
I use **CROSSTAB** Functionality of python to check the individual Features Contribution of an accident.

1. **Weather** :-



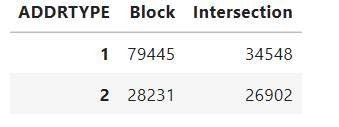
The left side indicates the accident severity code.We can directly see that the maximum number of accident occurred when the weather was ‘**CLEAR**’ .

1. **Collision type** :-



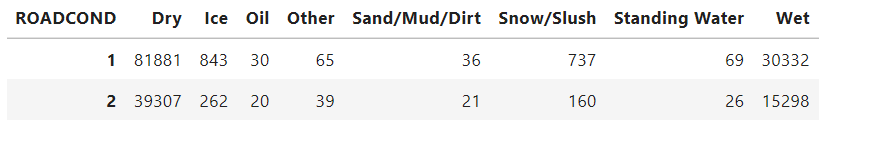
We can see that the property damage happened mostly due to ‘**Parked car’** but major injury happened due to ‘**Angles**’ and overall responsible is ‘**ParKed Car’**.

1. **Address Type** :-



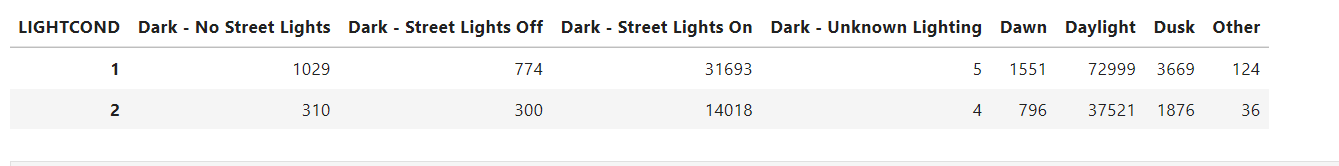
The Property Damage as well as Major Injury happened mainly on ‘**Block**’ .

1. **Road Condition** :-



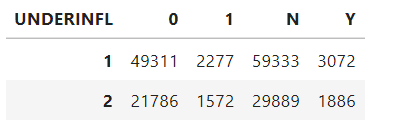
The most accident occurred when the road was in ;**Dry**’ Condition.

1. **Light Condition** :-



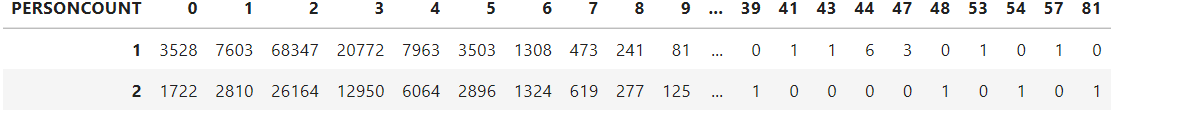
The maximum number of accident occurred in the **daylight**.

1. **Underinfl** :-



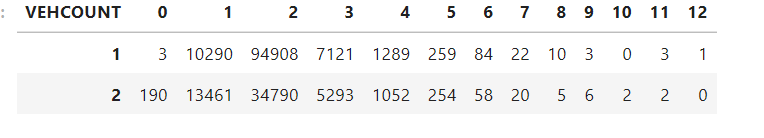
We can see that the maximum number of accident happened when **underinfl=’N’**.

1. **PersonCount** :-



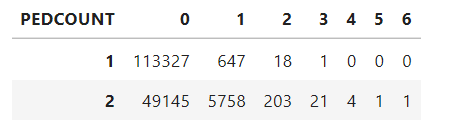
In the all accident the major 2 persons category affected.

1. **Vehcount:-**



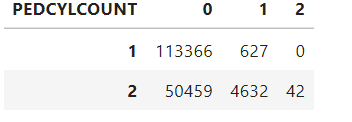
**We c**an see that the majority accident occurs when there is 2 vehicles involved.

1. **PedCount**:-



The major accident happens when pedcount=’0’.

1. **PedcylCount** :-



The major accident happens when pedcylcount=’0’.

I calculated the correlation between each feature with severity code but I did not get any good correlation ☹.

**Model Building** :-

Algorithm I have used here is

1. K Nearest Neighbour
2. Decision Tree