**Back Ground**

We have a data of various accidents in particular area since 2004 to 2020. In this data the time and date of accidents are also mentioned. The weather, road conditions and light conditions are also mentioned during these accidents. Using location and attitude one can easily predict and get any proper conclusion. The kind of property damage and collision type is also mentioned. It is important to downsize the amount of accidents so we can make a resident friendly zone.

Here I have put more focus on SDOT\_COLCODE(reason for many accidents) value and hours of accident(for this I also split the date and time for measuring)

**Problem**

By using this data, we need to decrease the road accidents. We can also predict some times and zones where the probability of accidents are higher. In this project, we need to suggest the tasks to government in order to decrease accidents and make the region more secure.

**Interest**

By using nearly 2 lakh records in 35 different parameters, we can easily draw a conclusion in order to decrease the accidents. Government can also take appropriate steps in order to make it more safe.

Parked Car is reason for ¼ accidents of SDOT\_COLCODE 14. So I mostly focused on it and tried to solve it with recomedations.

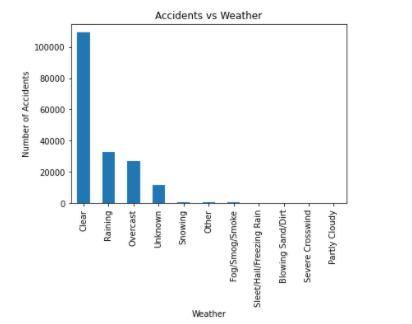
**Data Acquisition and cleaning**

Our primary data source is a csv file. This CSV file having information of accidents since last 16 years. From this data we can also analyze the weather, road condition, street, type of collision as well as street light type. During this analysis, we can not predict anything properly. So, that we need to drop the rows having inadequet information. For example, There is no way we can predict the weather of any day or we can predict the condition of read of any particular day. Because these parameters can never be constant.

**Data Analysis**

We need to use several factors individually in order to see its impact on number of accidents. We can analyze the weather impact on number of accidents.

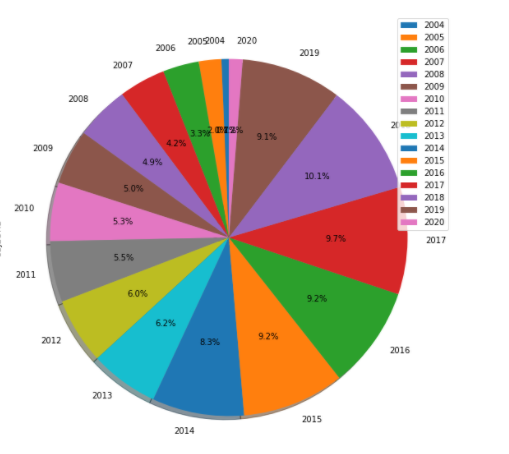
We can easily say that most of accidents are in clear weather. More than half of the accidents in Clean weather.



There are nearly similar accidents in Raining and Overcast season. In other weathers there are less accidents. So we can not have any proper conclusion from weather.

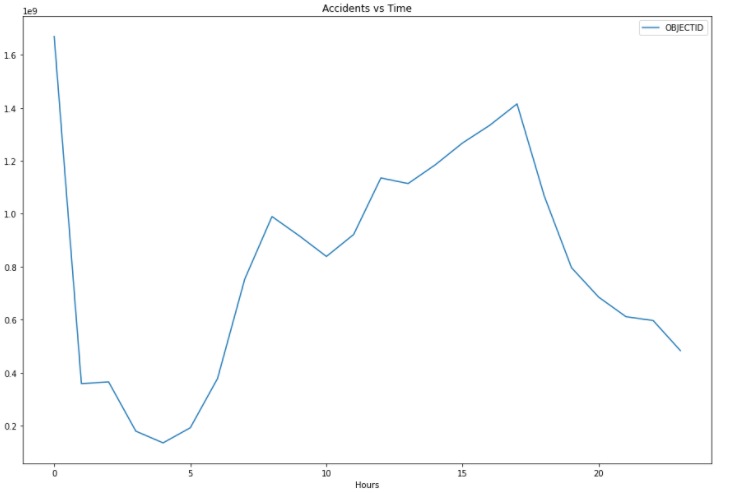
Here we can analyze the accidents year wise by using this pie chart. By this we can easily conclude that

After 2013 their a huge spike in number of accidents each year till 2019. In 2019 the number of accidents are nearly 10% of last 15 years. We can see less accdeints of 2020 because we could not have proper data of 2020 the current year.



We can easily predict the time and accident relation using this graph.

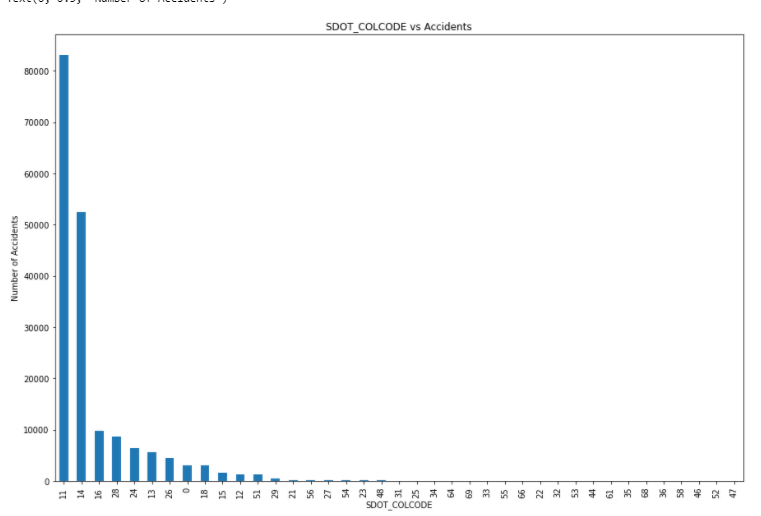
By this graph, we can easily check that mostly accidents occur in night 12 to 1 time. The chances of accidents are also high during noon 12-5. We can ask people to avoid using vehicles during that time. We can also ask government to put more protection during this time.





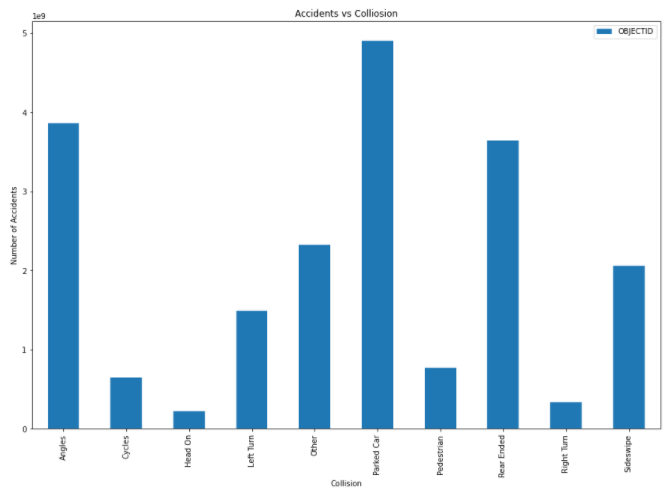
This is the most interesting graph of word cloud for this project. By looking at this word cloud we can easily see that mostly accidents happened with “Parked car”. This is twist in this project.

We can also see that “rear ended: and “left turn” are also important collision type for accidents in this region.



By using this graph we can easily find that MOTOR VEHICLE STRUCK MOTOR VEHICLE, FRONT END AT ANGLE and MOTOR VEHICLE STRUCK MOTOR VEHICLE, REAR END are major reasons for accidents. Nearly more than 80% accidents are because of these reasons . We can eaily eliminate it using different methods.

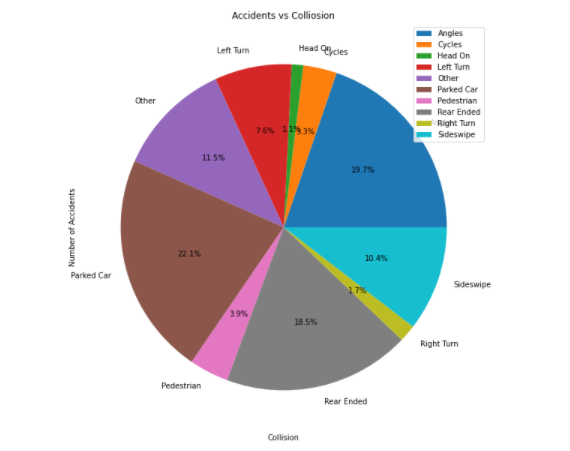
One more interesting fact is that most of the collision with the parked car in this region. So parked car are also creating issues for divers. Here is the graph to show that data .

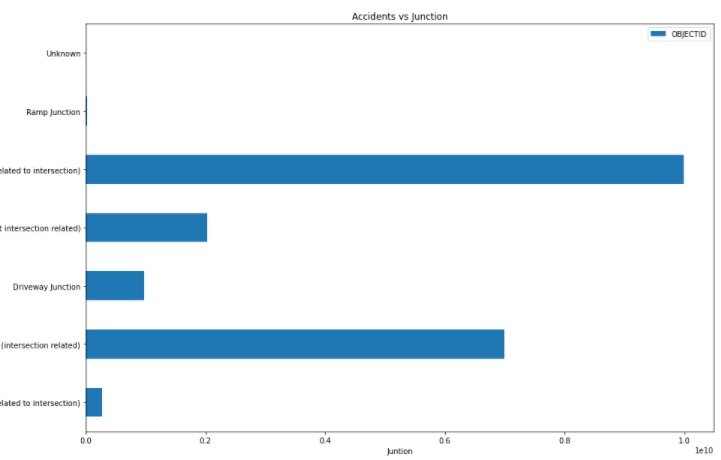


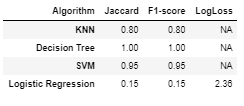
I have also used knn neighbours method to predict the accuracy when SDOT\_COLCODE is 14.

Here are two graphs only for 14.

Here is the collision type for SDOT\_COLCODE 14. We can say that Parked cars and angles are the highest for it.







I have used certain parameters for Machine Learning Results. I also used SDOT\_COLCODE value as 14 for all. And predictable value is hour.

I have used various different features for calculating it. It clearly shows us the accuracy level from the graph. From this graph you can easily notice the decrease in accuracy level as the features value changing.

**Conclusion**:

At certain values of SDOT\_COLCODE the accidents numbers are really high. We can eaily downsize it.

I can also say that weather condition, Light condition and road condition having less impact for road accidents. At certain timing the accidents are really high. Accidents with park vehicles are high and we can easily eradicate it.

**Future Prediction:**

Accidents on certain time is really high so we can easily deacrese it with the help of government and society. Accidents in last five years are really high as compare to previous 15 years. So we need to careful about it. Accidents with parked vehicles are really higher which can be easily decrease with the help of proper rules. Front end at angle and rear end are the reasons for most of accidents which can be easily downsize.