Nysa P Ginu

October 2020

CAR ACCIDENT SEVERITY- SEATTLE

BUSINESS PROBLEM

- The objective of this capstone project is to analyze the factors leading to a collision from a dataset of Car accident Severity in the city of Seattle, USA and using Machine Learning Methodology, create a system that can successfully predict the Severity of the car accident.

  This will help in answering the business problem: What are the conditions that would lead to a car accident and how severe it would be?

DATA

- **Data Sources:**

  The dataset taken is the one given in the earlier modules of the course. This dataset can be found here.

- **Data Cleaning:**
- The data downloaded can be loaded into a dataframe. On observing the dataframe, I felt that a lot of the attributes(columns) are redundant i.e. repeating the same information or are irrelevant to the problem in hand. Therefore, I selected only a few attributes for working on the project i.e. SEVERITYCODE, COLLISIONTYPE, VEHCOUNT, INATTENTIONIND, UNDERINFL, WEATHER, ROADCOND, LIGHTCOND, SPEEDING. I compiled them into another dataframe named “new”.
- Again a few of the columns had variable values like INATTENTIONIND, UNDERINFL, SPEEDING.
- So we will convert them into a more correct form.
- Most of the columns of the dataframe are of the type object, when we need it to be numerical. Therefore, The dataframe in this form is not suitable for analysis. To make it suitable, we will label encoding on the data.

METHODOLOGY

The models that we used are:

- **K-Nearest Neighbor(KNN):**

  KNN will help us predict the severity code of an outcome by finding the most similar to data point within k distance.

- **Logistic Regression:**

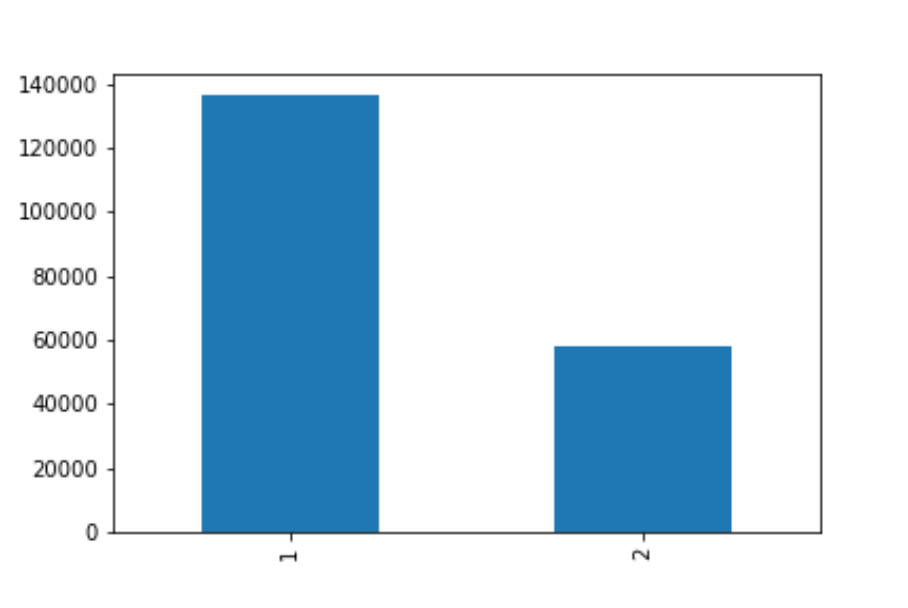
  Because our dataset only provides us with two severity code outcomes, our model will only predict one of those two classes. This makes our data binary, which is perfect to use with logistic regression

- **Decision Tree:**

  A decision tree model gives us a layout of all possible outcomes so we can fully analyze the consequences of a decision. It context, the decision tree observes all possible outcomes of different weather conditions.

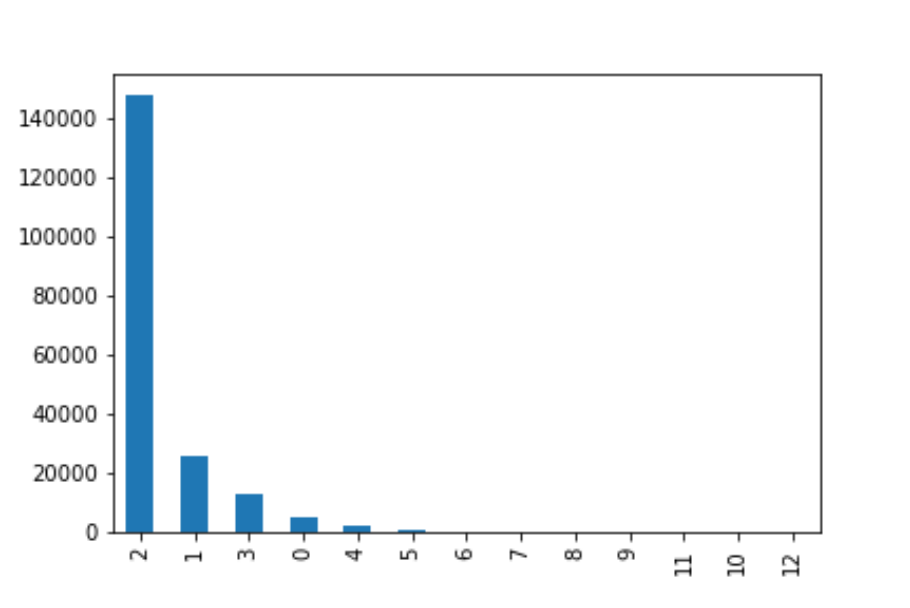
OBSERVATIONS

- On observing the graph for SEVERITYCODE, we can see that most of the accidents were of type-1 i.e. Property damage rather than type-2 i.e. Injury.



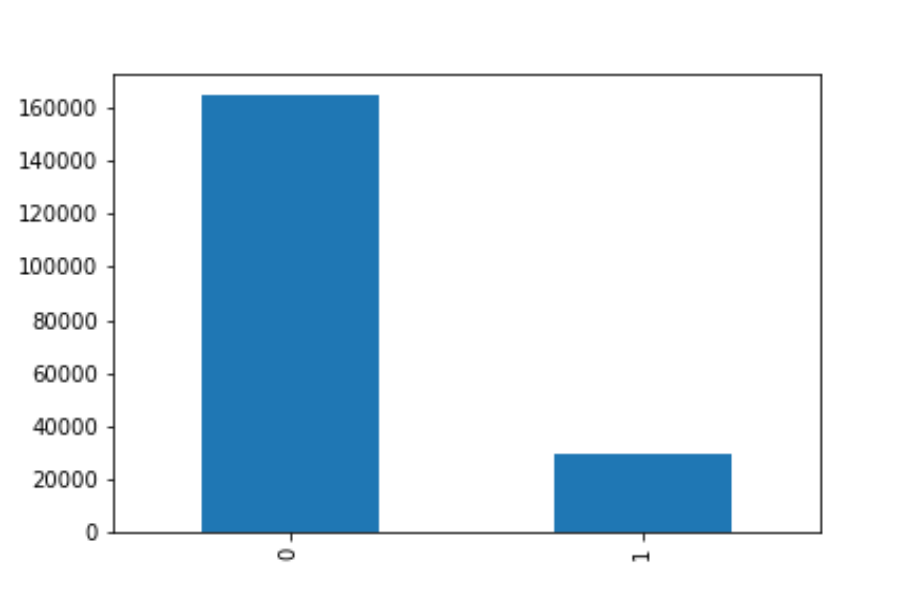
OBSERVATIONS

- Here we can see that most of accidents involved 2 vehicles. So, this eliminates the option of multi-car collisions.



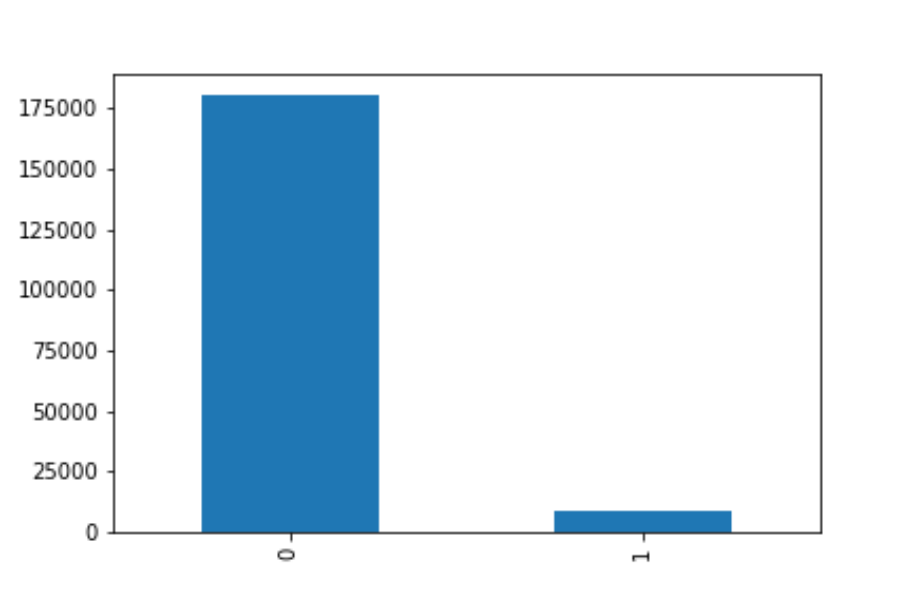
OBSERVATIONS

- This shows that majority of the drivers involved were inattentive.



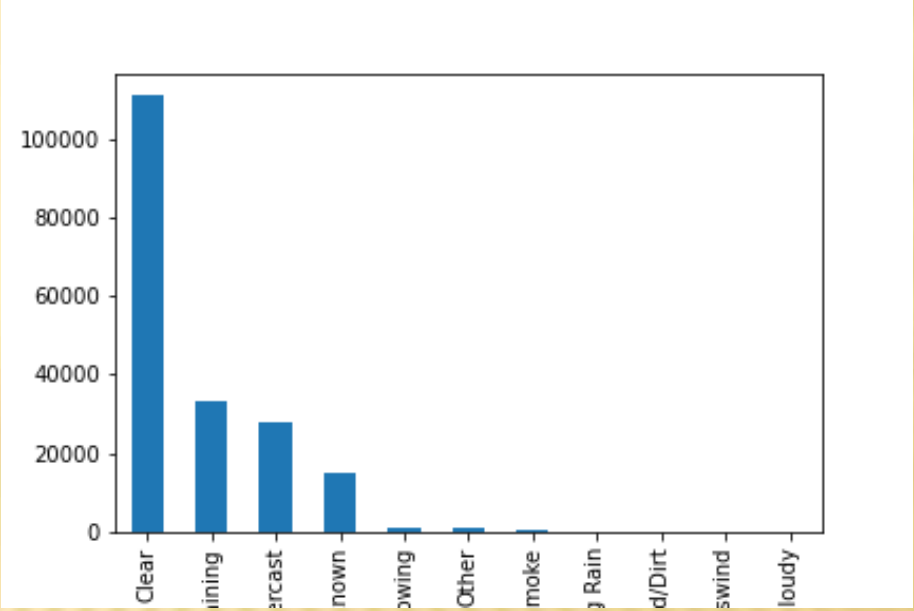
OBSERVATIONS

- Graph clearly depicts the role played by drivers under the influence of drugs or alcohol in causing the accidents.



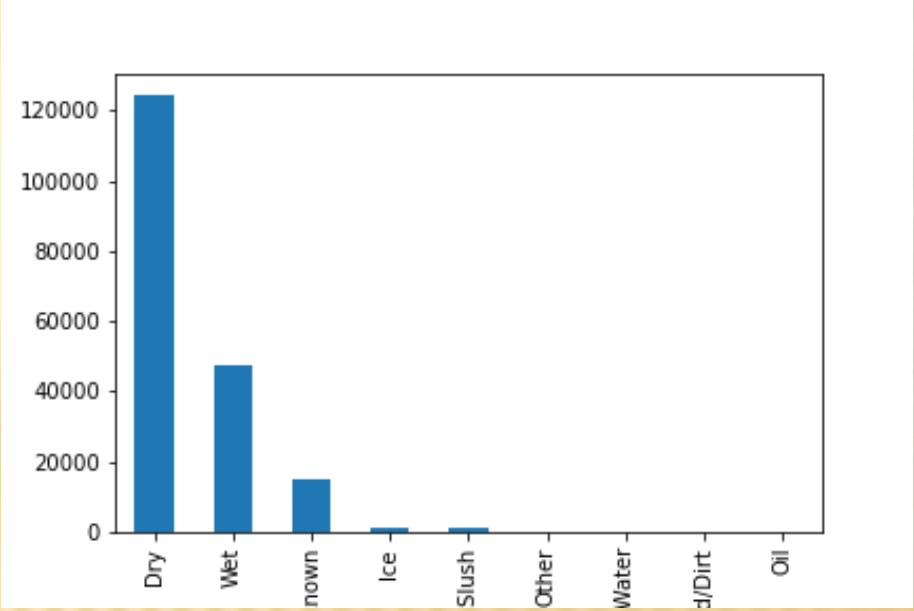
OBSERVATIONS

- According to this graph, most of the accidents happened in a clear weather.



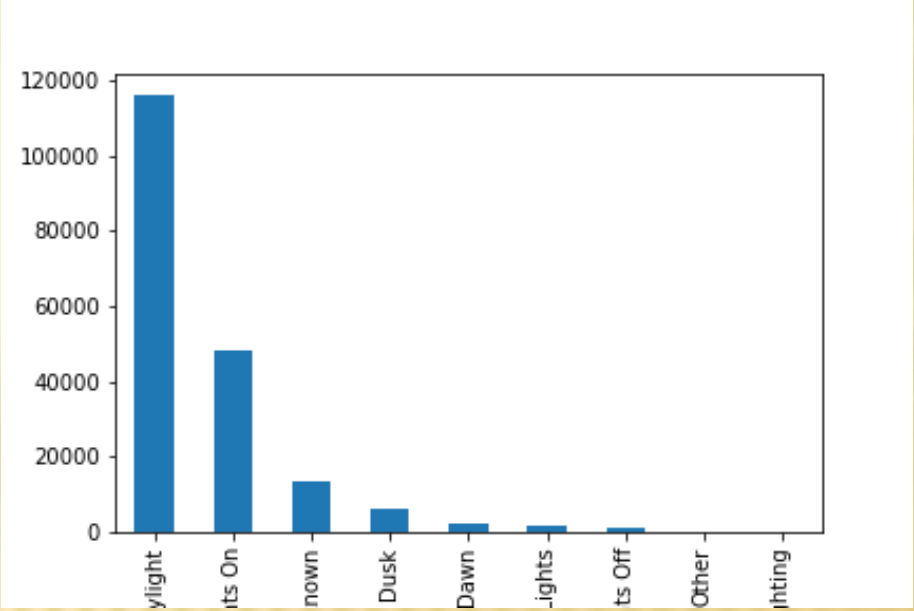
OBSERVATIONS

- Dry road conditions were observed during majority of the accidents.



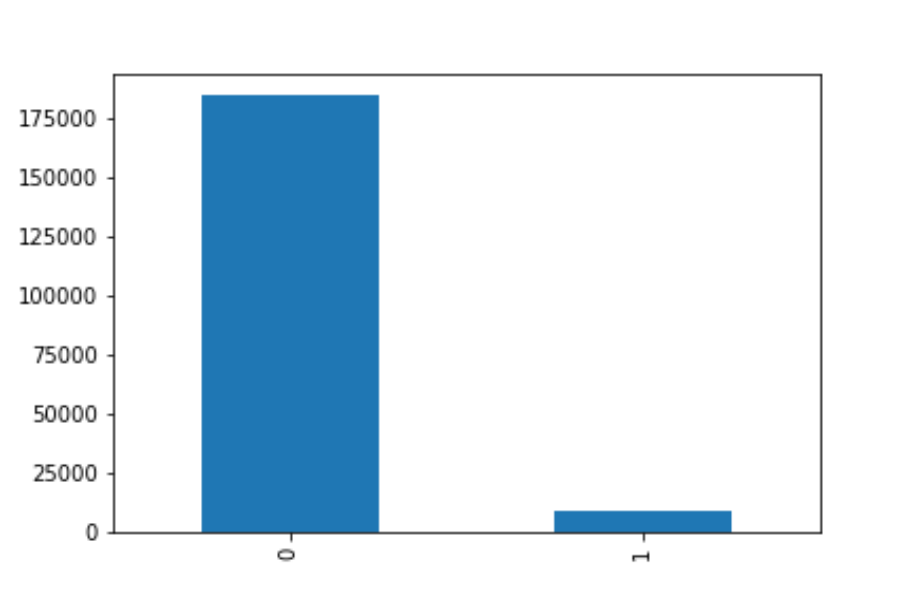
OBSERVATIONS

- The days were bright when maximum of the collisions happened.



OBSERVATIONS

- The role played by driver speeding the car can be seen from this graph.



CONCLUSION

- In this study, I analyzed the factors that can cause an accident. I identified that most crashes happened in clear, dry and bright conditions while the attentiveness, under the influence or speeding did not play a vital role. Most days are clear, dry and bright, so it's no surprise that car crashes occur under these conditions. I also created a model to predict the severity of the car accident using machine learning algorithms.

THANK YOU!