IBM Data Science Capstone: Car Accident Severity Report

Meenal Lodha

Introduction | Business Understanding

This project is based on the case study which is to predict the severity of an accident. Say you are driving to another city for work or to visit some friends. It is rainy and windy, and on the way, you come across a terrible traffic jam on the other side of the highway. Long lines of cars barely moving. As you keep driving, police car start appearing from a far shutting down the highway. Oh, it is an accident and there's a helicopter transporting the ones involved in the crash to the nearest hospital. They must be in critical condition for all of this to be happening. Now, wouldn't it be great if there is something in place that could warn you, given the weather and the road conditions about the possibility of you getting into a car accident and how severe it would be, so that you would drive more carefully or even change your travel if you are able to. Well, this is exactly the problem we have to find solution for. In an effort to reduce the frequency of car collisions in a community, an algorithm must be developed to predict the severity of an accident given the current weather, road and visibility conditions. When conditions are bad, this model will alert drivers to remind them to be more careful.

Data Understanding

Our predictor or target variable will be 'SEVERITYCODE' because it is used measure the severity of an accident and other attributes used to weigh the severity of an accident are 'WEATHER', 'ROADCOND' and 'LIGHTCOND'.

Extract Dataset

First, import the required libraries and then read dataset and load data.

In [1]:
```
import itertools
import numpy as np
import matplotlib.pyplot as plt
from matplotlib.ticker import NullFormatter
import pandas as pd
import numpy as np
import matplotlib.ticker as ticker
from sklearn import preprocessing
%matplotlib inline
```

About dataset

Second, read .csv file and load data.

In [2]:
```
df=pd.read_csv("https://s3.us.cloud-object-storage.appdomain.cloud/cf-courses-data/CognitiveClass/DP0701EN/version-2/Data-Collisions.csv")
```

```
/opt/conda/envs/Python36/lib/python3.6/site-packages/IPython/core/interactiveshell.py:3020: DtypeWarning: Columns (33) have mixed types. Spe
cify dtype option on import or set low_memory=False.
  interactivity=interactivity, compiler=compiler, result=result)
```

In [3]:
```
df.head()
```

Out[3]:

	SEVERITYCODE	X	Y	OBJECTID	INCKEY	COLDETKEY	REPORTNO	STATUS	ADDRTYPE	INTKEY	...	ROADCOND	LIGHTCOND	PEDROWNOTGRNT
0	2	-122.323148	47.703140	1	1307	1307	3502005	Matched	Intersection	37475.0	...	Wet	Daylight	NaN

Then, I began choosing columns to use from the data frame that I created. The columns that I chose were SEVERITYCODE, which assigns a crash a value of 1, which means no injury, and 2, indicating injury, COLLISIONTYPE, which

describes the type of crash, WEATHER, which describes the weather at the time of crash, ROADCOND, which describes the condition of the road at the time of crash, LIGHTCOND, which describes the light conditions at the time of crash, INATTENTIONIND, which describes whether the driver was distracted, and UNDERINFL, which describes whether the driver was under the influence.

In [6]:
```
df=df[['SEVERITYCODE','WEATHER','ROADCOND','LIGHTCOND','COLLISIONTYPE','INATTENTIONIND','UNDERINFL']]
df.head()
```

Out[6]:

	SEVERITYCODE	WEATHER	ROADCOND	LIGHTCOND	COLLISIONTYPE	INATTENTIONIND	UNDERINFL
0	2	Overcast	Wet	Daylight	Angles	NaN	N
1	1	Raining	Wet	Dark - Street Lights On	Sideswipe	NaN	0
2	1	Overcast	Dry	Daylight	Parked Car	NaN	0
3	1	Clear	Dry	Daylight	Other	NaN	N
4	2	Raining	Wet	Daylight	Angles	NaN	0

In [8]:
```
df.dtypes
```

Out[8]:
```
SEVERITYCODE       int64
WEATHER           object
ROADCOND          object
LIGHTCOND         object
COLLISIONTYPE     object
INATTENTIONIND    object
UNDERINFL         object
dtype: object
```

Methodology

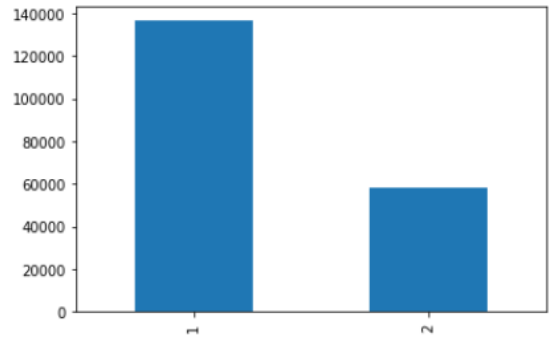
For the data analysis portion of this project, Github was used for a repository and I chose to use IBM Watson Studio to host the notebook. To host the large collision csv file, it was uploaded to IBM Cloud.

The Python libraries utilized throughout the analysis were Pandas, Numpy, Matplotlib, and Seaborn.

Once the data from the csv file was read, I used df.types in order to see what type of data was contained within the spreadsheet. I have considered weather condition ,type of collision and Light condition on streets in my analysis. Because I was trying to determine the amount of accidents occurring with the light condition of “Dark- No Street Lights”, I chose to focus on the Severity Code also.

I visualized the data in the form of bar graphs. I filtered out the columns I wanted from the provided .csv and then called value_counts to graph the mostly categorical data.

```
In [11]: df_Conditions_balanced['SEVERITYCODE'].value_counts().plot(kind = 'bar')
Out[11]: <matplotlib.axes._subplots.AxesSubplot at 0x7f6166a136a0>
```



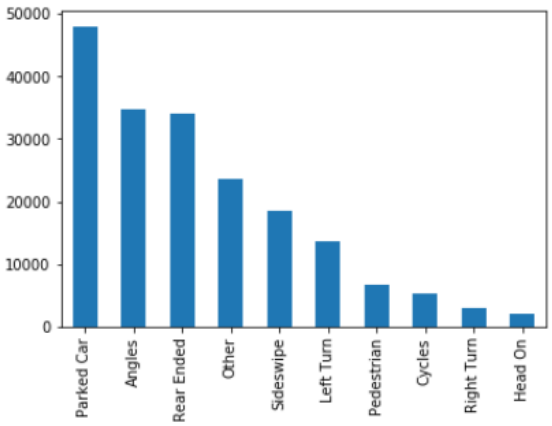
SEVERITYCODE, which assigns a crash a value of 1, which means no injury, and 2, indicating injury

This code corresponds to the severity of the collision :

1 means property damage but no injury

2 means injury

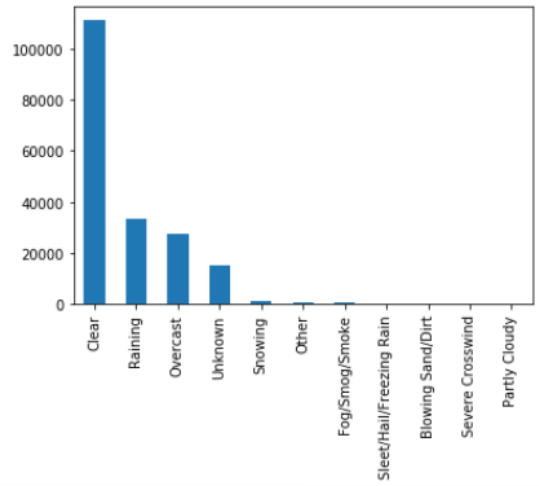
```
In [9]: df_Conditions_balanced['COLLISIONTYPE'].value_counts().plot(kind = 'bar')
Out[9]: <matplotlib.axes._subplots.AxesSubplot at 0x7f6166d96518>
```



This graph show the type of collision with indicates that most of collision occurs with parked vehicles.

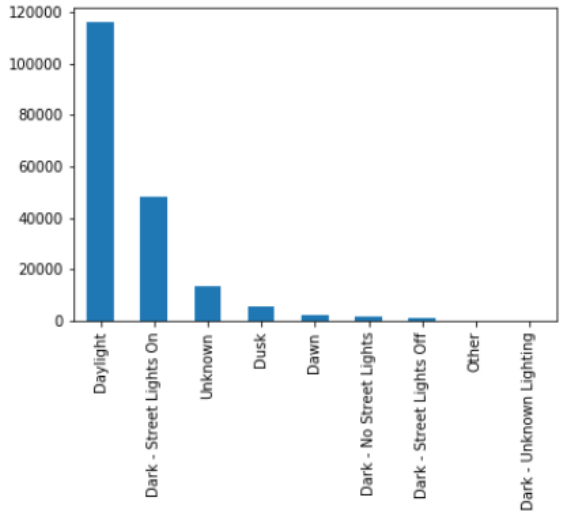
```
In [10]: df_Conditions_balanced['WEATHER'].value_counts().plot(kind = 'bar')
```

```
Out[10]: <matplotlib.axes._subplots.AxesSubplot at 0x7f6166a8bcc0>
```



Weather conditions also plays important role during driving. Driver needs to be more careful as most of crashes happened in clear, raining, and bright conditions. Most days are clear, raining, and bright, so it's no surprise that most car crashes occur under these conditions.

```
In [13]: df_Conditions_balanced['LIGHTCOND'].value_counts().plot(kind = 'bar')
Out[13]: <matplotlib.axes._subplots.AxesSubplot at 0x7f61669ea588>
```



Light conditions also have major impact on road accidents. Most accidents takes place in daylight as well as in night where lights are on.

Result

1. Most crashes happened in clear, dry, and bright conditions. Most days are clear, dry, and bright, so it's no surprise that most car crashes occur under these conditions. I also found out that crashes with a distracted driver or an impaired driver are statistically more likely to result in injury, which is also not a surprise. The results of the data indicate to city officials that they should ask drivers to be more alert in ideal conditions.
2. Based on historical data from weather conditions pointing to certain classes, we can conclude that particular weather conditions have a

somewhat impact on whether or not travel could result in property damage (class 1) or injury (class 2).

3. The lack of lighting in areas does not appear to have a big impact or correlation to accidents at all. The location where there is no street light that had the most accidents, as well as streets with lights also have quite a number of accidents.

Discussion

With the given set of data, that contains information such as severity, location, collision type, weather conditions, road conditions, and light conditions, among others the task was to identify a problem and who would be interested in a solution to this problem. Based on my preliminary review of the data, I decided to explore the amount of accidents that occur in areas labeled as "Dark- No Street Lights" and "Weather conditions".

Based on that I realized that streets with no lights and zero visibility due to weather conditions are more prone to accidents.

As well during day light or with street lights but bad weather chances of collisions are more.

Weather conditions are not predictable but precautions can be taken during such conditions to avoid any type of causality.

Conclusion

Based on analysis I came to conclusion that more precautions should be taken during harsh conditions and the results of the data indicate to city officials that they should ask drivers to be more alert in ideal conditions.