**Analysis of Car Accident Severity at Seattle**

9/14/2020

1. **Introduction**

Seattle is the largest city in the state of Washington, as well as the largest in the Pacific Northwest. The current metro area population of Seattle in 2020 is **3,433,000**, a 0.79% increase from 2019 [1]. Seattle residents get around by car, trolley, streetcar, public bus, bicycle, on foot, and by rail; and a crash happened in Washington every 4.5 minutes, Seattle is the 8th most dangerous city for car accidents in USA.

For this project, we want to analyze existing collision data provided by SPD of Seattle, to understand what the most common causes and their correlation are; and to prevent it with better care in future.

1. **Data**

This collisions data is been provided by Seattle Police Department and weekly recorded by Traffic Records from 2004 to present.

It contains information such as severity code, address type, location, collision type, weather, road condition, speeding, whether collision was due to inattention, whether or not a driver involved was under the influence of drugs or alcohol.

Data cleaning is required. To accurately build a model, these attributes with high ~% null value of data can’t be used for this analysis.

eg: In total 194673 cases, INATTENTIONIND has 164868 cases of null value; and SPEEDING has 185340 cases of null value.

For rest of the attributes, we must decide how to deal with null value of data for each of them.

*Column Name # of null value*

SEVERITYCODE 0

X 5334

Y 5334

OBJECTID 0

INCKEY 0

COLDETKEY 0

REPORTNO 0

STATUS 0

ADDRTYPE 1926

INTKEY 129603

LOCATION 2677

EXCEPTRSNCODE 109862

EXCEPTRSNDESC 189035

SEVERITYCODE.1 0

SEVERITYDESC 0

COLLISIONTYPE 4904

PERSONCOUNT 0

PEDCOUNT 0

PEDCYLCOUNT 0

VEHCOUNT 0

INCDATE 0

INCDTTM 0

JUNCTIONTYPE 6329

SDOT\_COLCODE 0

SDOT\_COLDESC 0

INATTENTIONIND 164868

UNDERINFL 4884

WEATHER 5081

ROADCOND 5012

LIGHTCOND 5170

PEDROWNOTGRNT 190006

SDOTCOLNUM 79737

SPEEDING 185340

ST\_COLCODE 18

ST\_COLDESC 4904

SEGLANEKEY 0

CROSSWALKKEY 0

HITPARKEDCAR 0

1. **Methodology**

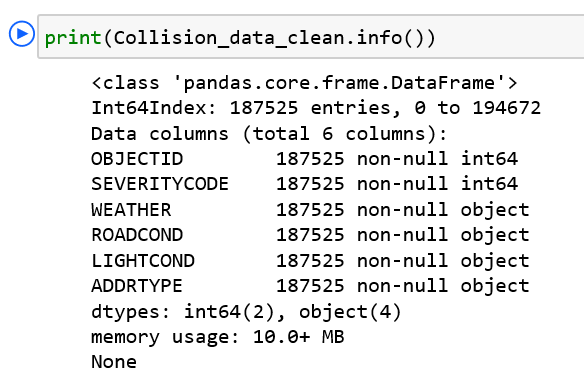
Jupyter Notebooks was been used to conduct this analysis. And basic Python libraries Pandas, Numpy, Matplotlib, and Seaborn were also been imported and used throughout this project.

Using pd.read\_csv command to import Data-Collisions.csv, and Pandas DataFrame shape function showed there were 194,673 rows and 38 columns in this csv file.

After initial reviewed the data, cleaned up the dataset by removing columns that are not informative for visualization, and columns of OBJECTID, SEVERITYCODE, WEATHER, ROADCOND, LIGHTCOND and ADDRTYPE had been selected to use for this analysis.

One of the most common data pre-processing steps is to check for null values in the dataset, WEATHER, ROADCOND, and LIGHTCOND columns all have < 3% of missing values, since it would have little impact on the overall result so that these rows with null value are been removed.

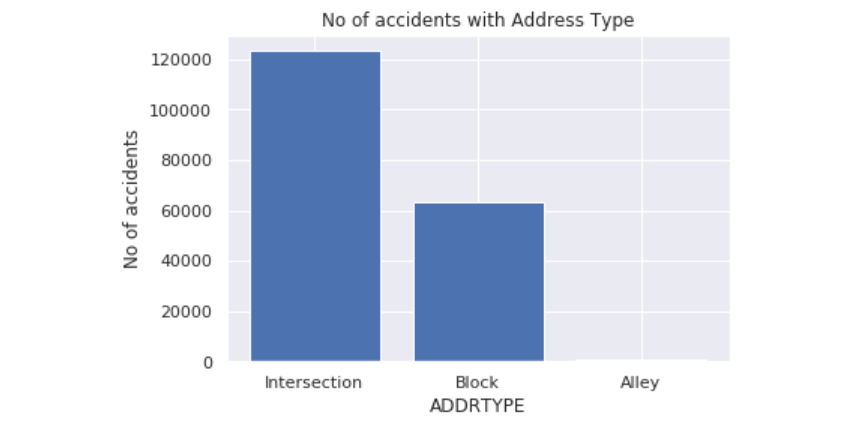
Finally gathered some information about each column in this dataFrame.

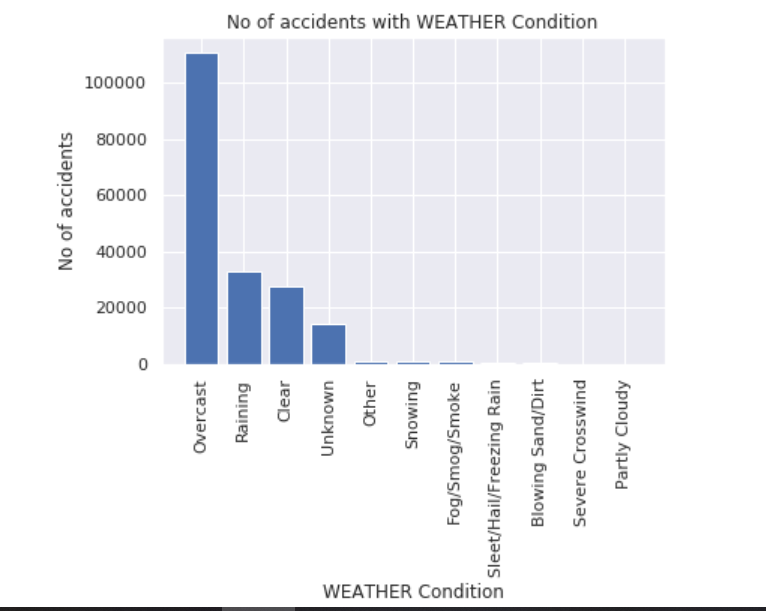


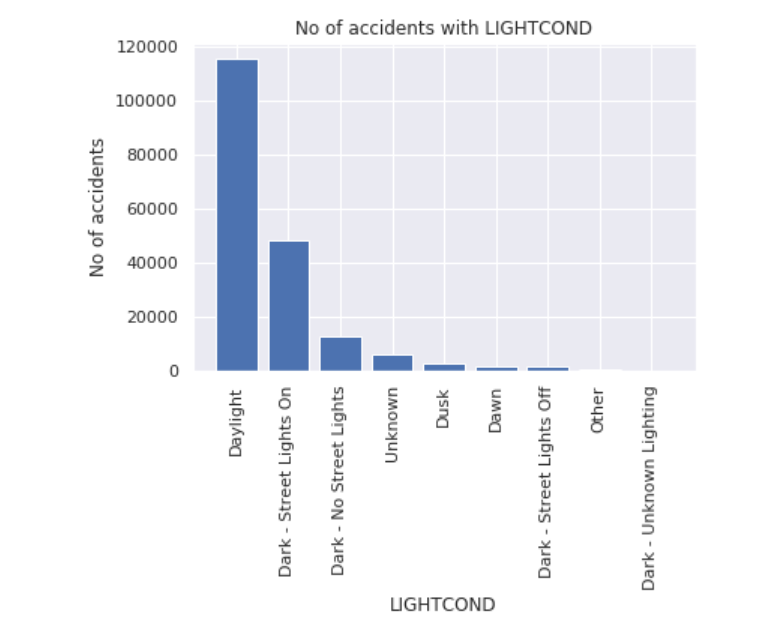
This dataset was mostly categorical, graph algorithm was be used to analyze the collision data; to discover patterns and correlation between various variables.

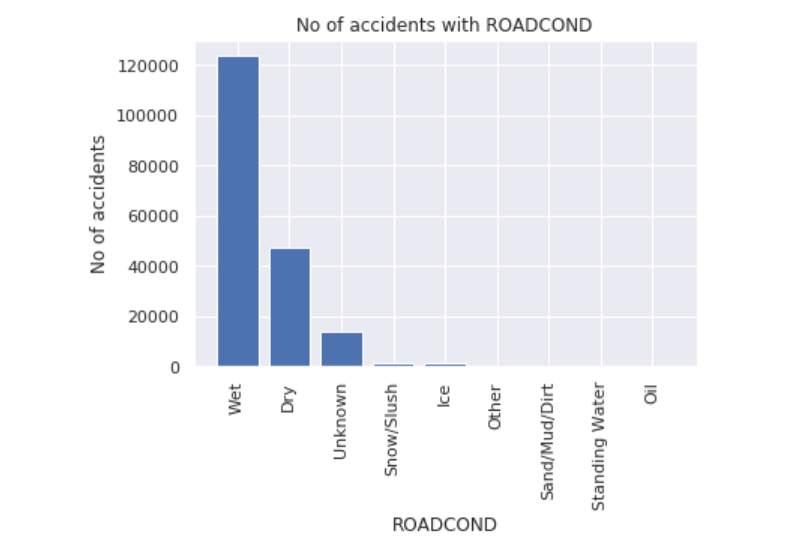
1. **Results and Discussion**

By using Python plotting library matplotlib, we generated the bar charts based on address type, weather, road, and light condition. The result showed the car accident happened more at intersection location; at overcast weather condition; and when road condition is wet; but it seems that the dark – streetlights on or no streetlights condition has no significant impact for car accidents.









Seattle is one of the cloudiest large cities in the United States, where over half the days a year has cloud covers more than three quarters of the sky [2]. From this analysis, the data showed in the overcast weather, the cat accident happened much more than other weather condition, this information can be used for Seattle government to see if improvement can be applied.

| Cloudiest major US cities, where over half the days a year cloud covers more than three-quarters of the sky | | |
| --- | --- | --- |
| **City** | **Days of Heavy Cloud** | **% of Days** |
| Seattle, Washington | 226 | 62 |
| Portland, Oregon | 222 | 61 |
| Buffalo, New York | 208 | 57 |
| Pittsburgh, Pennsylvania | 203 | 56 |
| Cleveland, Ohio | 202 | 55 |
| Rochester, New York | 200 | 55 |
| Columbus, Ohio | 190 | 52 |
| Cincinnati, Ohio | 186 | 51 |
| Detroit, Michigan | 185 | 51 |



Speeding and INATTENTIONIND (Whether or not collision was due to inattention) missed over 85% of data. If these data can be carefully collected, it may help Seattle to find out the top causes for car accidents.

1. **Conclusion**

As we expected, the analysis supported the poor weather and road condition, intersection place increased the risk of car accidents but surprising the light didn’t impact much.

Seattle transportation department may need to collect more data for speeding drive and drunk driving, and further analyze if they have correlation with car accidents.

Most accidents involve some form of driver error, but proper safety instruction will prevent and reduce the accidents and save the life.

**6. References**

[1] https://www.macrotrends.net/cities/23140/seattle/population#:~:text=The%20current%20metro%20area%20population%20of%20Seattle%20in,was%203%2C379%2C000%20%2C%20a%201.2%25%20increase%20from%202017.

[2] https://www.currentresults.com/Weather-Extremes/US/cloudiest-cities.php