IBM/Coursera
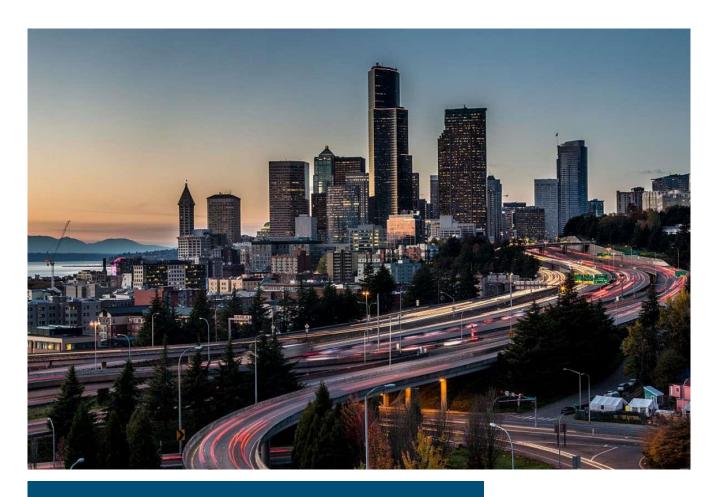
Capstone Project for Applied Data Science

Seattle Car Accident Severity

By: Jose Arias

Introduction or Business Problem:

Statistics:

Seattle is a city in Washington State, as of 2020, it has a total metro area population of 3.4 million residents. In the year of 2016, an average of 450,000 vehicles was announce in the State Census. Even with two metro system providers in the city in an overpopulated city this is not enough. The traffic situation motivated the population to buy car and as a result use this method of transportation to commute. The increase in car ownership raised the numbers of accidents on the road. Seattle is surrounded by water and mountains, factors that brings severe weather changes like rain, fog and snow into the city, statically, this weather factors turns out in more car accidents

Worldwide situation:

Car accidents turns out in a very serious situation, bring to reality a lot of issues, example, deaths health conditions, an financial struggles to the insurance companies, governments and population. According to 2017 WSDOT data, a car accident occurs every 4 minutes and a person dies due to a car crash every 20 hours in the state of Washington. The Seattle Severity Car Accident project pretends to predict how severity of accidents can be reduced based on a few factors.

Stakeholders:

Minimizing the number of car accidents can be beneficial for the population, insurance companies, and local governments. A lot of local governments work in a daily basis to prevent this king of accidents, announcing weather conditions in the area, conditions in the road or even announcing car accidents ahead, in this way they help to prevent more accidents.

Data:

Using the data provided by Coursera on Seattle Traffic, I will investigate the connection between severity of car accidents and weather conditions. This data provides history of more than sixteen years in the city Seattle.

Methodology:

IBM Watson Studio was used to create the notebook and a Repository from Github. Python libraries used for data analysis was Pandas.
A csv file were provided by Coursera, I used df.dtypes to see what type of data are the columns in the file. To investigate the connection between car accidents based in the conditions I have choose, my selections are INATTENTIONIND (obj); LIGHTCOND (Object); ROADCOND (Condition); SPEEDING (object); UNDERINFL (object) and WEATHER (Object).

From the Notebook:

Lack of attention or distraction.

```
In [7]:  ▶  df['INATTENTIONIND'].value_counts().to_frame()
```

Out[7]:

	INATTENTIONIND
Y	29805

LIGHTCOND:
The light conditions during the collision.

```
In [10]:  ▶  df['LIGHTCOND'].value_counts().to_frame()
```

Out[10]:

	LIGHTCOND
Daylight	116137
Dark - Street Lights On	48507
Unknown	13473
Dusk	5902
Dawn	2502
Dark - No Street Lights	1537
Dark - Street Lights Off	1199
Other	235
Dark - Unknown Lighting	11

ROADCOND:

The condition of the road during the collision.

```
In [11]: ▶ df['ROADCOND'].value_counts().to_frame()
```
Out[11]:

	ROADCOND
Dry	124510
Wet	47474
Unknown	15078
Ice	1209
Snow/Slush	1004
Other	132
Standing Water	115
Sand/Mud/Dirt	75
Oil	64

SPEEDING:

Whether or not speeding was a factor in the collision. (Y/N).

```
In [12]: ▶ df['SPEEDING'].value_counts().to_frame()
```
Out[12]:

	SPEEDING
Y	9333

UNDERINFL:

Whether or not a driver involved was under the influence of drugs or alcohol.

In [13]: ▶ df['UNDERINFL'].value_counts().to_frame()

Out[13]:

	UNDERINFL
N	100274
0	80394
Y	5126
1	3995

WEATHER:

A description of the weather conditions during the time of the collision.

In [14]: ▶ df['WEATHER'].value_counts().to_frame()

Out[14]:

	WEATHER
Clear	111135
Raining	33145
Overcast	27714
Unknown	15091
Snowing	907
Other	832
Fog/Smog/Smoke	569
Sleet/Hail/Freezing Rain	113
Blowing Sand/Dirt	56
Severe Crosswind	25
Partly Cloudy	5

This is a comparation dataframe considering: Weather, Roadcond and Lightcond.

```
In [19]: ▶ df_Weather=df.filter(['SEVERITYCODE', 'WEATHER', 'ROADCOND','LIGHTCOND'], axis=1)
            df_Weather.head()
```

Out[19]:

	SEVERITYCODE	WEATHER	ROADCOND	LIGHTCOND
0	2	Overcast	Wet	Daylight
1	1	Raining	Wet	Dark - Street Lights On
2	1	Overcast	Dry	Daylight
3	1	Clear	Dry	Daylight
4	2	Raining	Wet	Daylight

My perspective:

Weather, Light and Road conditions

There are more occurrences of severe collisions during daylight whereas during the night with the lights on, accidents tend to be less risky. The reason for this, may be related to a more cautious driving during the night which predispose car users to an aware state. Dusk and dawn tend to be related to more severe collisions, maybe because of the visibility reduction while facing the sun directly in the vision zone.

Conclusion:

Extremely dangerous weather and road conditions do not produce a quite significant accident rate, such as snow and ice. However, caution have to be taken with rainy weather and wet roads, since after clear days and dry roads, these are the following conditions in order of importance.

Based on the data on collisions in Seattle from 2004 to the present, there are no relationship between bad weather conditions and wet road conditions that affected collisions. From the data, we see that there were a lot more collisions that happened on dry roads and clear weather conditions. There are much less collisions that happen when weather and road conditions are not that great.