# Capstone Project:
# Vehicle Accident Severity Analysis

By: M.Maharoof
Github repo: https://github.com/ShazGit/Coursera_Capstone.git

## Introduction

Around the world many road accidents take place daily. Some accidents can be minor while some can be severe. Analyzing variables associated with the cause of severe accidents can help develop an action plan to reduce severe accidents and thereby reduce loss of life caused from severe accident. This project analysis aims to predict how severity of accidents can be reduced based on identifying factors that are likely to cause the accidents.

This analysis will be beneficial for stakeholders such as the Public Development Authority of Seattle which for improving their understanding for key factors surrounding accidents so that necessary precautions can be taken to reduce the accident frequency rate.

## Data

A sample set of data collected by the Seattle Police Department recorded over a time period from 2004 till present shall be used to analyzed and used to make judgments to the variables associated with severe accidents. The sample data set contains information such as severity, location, collision type, weather conditions, road conditions, and light conditions, among others. For the data analysis portion of this project, Github will be used for a repository and IBM Watson Studio to host the notebook. The Python libraries utilized throughout the analysis were Pandas, Numpy, Matplotlib, and Seaborn are intended to be used.

## Methodology

I used Jupyter Notebooks to conduct that analysis and imported relevant Python libraries such as Numpy, Matplotlib, and Seaborn. The data was mostly categorical so I stuck to graphical representation to see correlation between various variables. I started by importing the csv file and I used df.types in order to see what type of data was contained within the spreadsheet. This generated the following result.

```
SEVERITYCODE        int64
X                 float64
Y                 float64
OBJECTID            int64
INCKEY              int64
COLDETKEY           int64
REPORTNO           object
STATUS             object
ADDRTYPE           object
INTKEY            float64
LOCATION           object
EXCEPTRSNCODE      object
EXCEPTRSNDESC      object
SEVERITYCODE.1      int64
SEVERITYDESC       object
COLLISIONTYPE      object
PERSONCOUNT         int64
PEDCOUNT            int64
PEDCYLCOUNT         int64
VEHCOUNT            int64
INCDATE            object
INCDTTM            object
JUNCTIONTYPE       object
SDOT_COLCODE        int64
SDOT_COLDESC       object
INATTENTIONIND     object
UNDERINFL          object
WEATHER            object
ROADCOND           object
LIGHTCOND          object
PEDROWNOTGRNT      object
SDOTCOLNUM        float64
SPEEDING           object
ST_COLCODE         object
ST_COLDESC         object
SEGLANEKEY          int64
CROSSWALKKEY        int64
HITPARKEDCAR       object
      dtype: object
```

After determining the types of data, I performed value count for the Location in order to get an understanding of where the accident points are. This produced the table below

| | LOCATION |
|---|---|
| BATTERY ST TUNNEL NB BETWEEN ALASKAN WY VI NB AND AURORA AVE N | 276 |
| BATTERY ST TUNNEL SB BETWEEN AURORA AVE N AND ALASKAN WY VI SB | 271 |
| N NORTHGATE WAY BETWEEN MERIDIAN AVE N AND CORLISS AVE N | 265 |
| AURORA AVE N BETWEEN N 117TH PL AND N 125TH ST | 254 |
| 6TH AVE AND JAMES ST | 252 |
| AURORA AVE N BETWEEN N 130TH ST AND N 135TH ST | 239 |
| ALASKAN WY VI NB BETWEEN S ROYAL BROUGHAM WAY ON RP AND SENECA ST OFF RP | 238 |
| RAINIER AVE S BETWEEN S BAYVIEW ST AND S MCCLELLAN ST | 231 |
| ALASKAN WY VI SB BETWEEN COLUMBIA ST ON RP AND ALASKAN WY VI SB EFR OFF RP | 212 |
| WEST SEATTLE BR EB BETWEEN ALASKAN WY VI NB ON RP AND DELRIDGE-W SEATTLE BR EB ON RP | 212 |
| AURORA BR BETWEEN RAYE ST AND BRIDGE WAY N | 190 |
| ALASKAN WY VI NB BETWEEN SENECA ST OFF RP AND WESTERN AV OFF RP | 164 |
| 1ST AVE BETWEEN BLANCHARD ST AND BELL ST | 161 |
| 5TH AVE AND SPRING ST | 160 |
| RAINIER AVE S BETWEEN S HENDERSON ST AND S DIRECTOR N ST | 152 |
| RAINIER AVE S BETWEEN S DEARBORN ST AND S CHARLES N ST | 146 |

I also performed a value count on Light Condition to see the breakdowns of accidents occurring during the different light conditions. These counts can be seen below.

| | 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 |

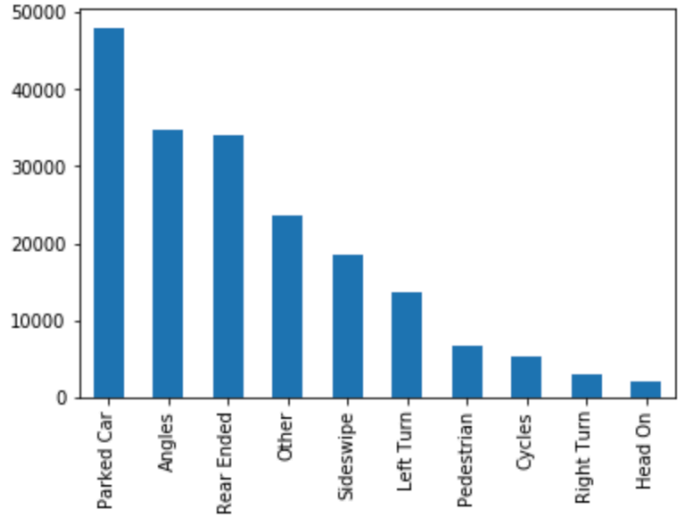
Thereafter I generated various graphs to analyze the data set. To perform this I imported matplotlib and seaborn libraries to conduct graphical analysis.

# Results

The first graph generated provided insights to the types of collisions. From the graphs it can be seen that most of the collisions took place with parked cars.

```
In [8]: df_dataset['COLLISIONTYPE'].value_counts().plot(kind='bar')
```

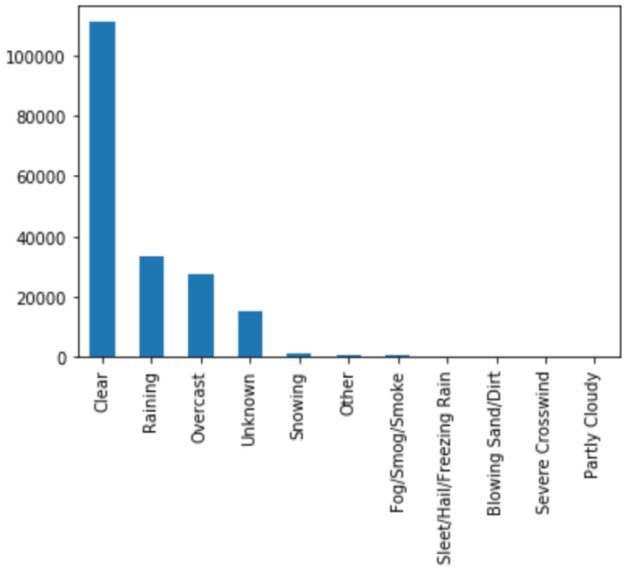
```
Out[8]: <matplotlib.axes._subplots.AxesSubplot at 0x7fbf4c32d780>
```




The next graph generated was to identify if weather had contributed to the accidents. But it can be seen that majority of the accidents took place during clear weather

```
In [9]: df_dataset['WEATHER'].value_counts().plot(kind='bar')
```

```
Out[9]: <matplotlib.axes._subplots.AxesSubplot at 0x7fbf11a80128>
```



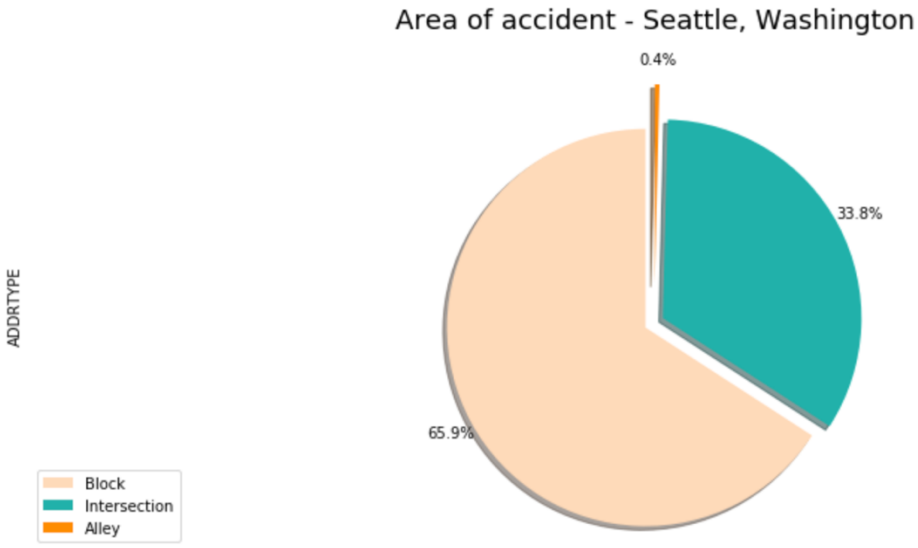
The following pie chart was generated to identify the areas of accident.

In [14]:
```
explode_list = [0.05, 0.05, 0.2]
color_list=['peachpuff','lightseagreen','darkorange']
addtype=df_dataset['ADDRTYPE'].value_counts()

addtype.plot(kind='pie',
             figsize=(15, 6),
             autopct='%1.1f%%',
             startangle=90,
             shadow=True,
             labels=None,
             pctdistance=1.12,
             colors=color_list,
             explode=explode_list)


plt.title('Area of accident - Seattle, Washington', fontsize=18, y=1.05)
plt.axis('equal')
plt.legend(labels=addtype.index, loc='lower left')

plt.show()
```




## Discussion

At the start of our analysis, I was trying to figure out the severity and frequency of road accidents based on weather conditions, road conditions, and other factors. Most crashes happened in clear, dry, and bright conditions and therefore the weather was not an important factor for the cause of the accidents. The results of the data indicate to city officials that they should ask drivers to be more alert in ideal conditions.

# Conclusion

In conclusion I can say that most of the accidents may have been cause due to driver negligence. Therefore, I would recommend the following measures to be taken in response to this data analysis insights:

- Install safety signs on the roads and ensure that all precautions are being taken by people within the area
- Launch development projects for those areas where most severe accidents take place in order to minimize the effects of these two factors

Although this analysis has given some valuable insight, there needs to be a closer inspection of certain other variables. It seems like a lot of these accidents are minor and avoidable.