Capstone Project – Accident Prediction Model

Introduction:
Build a model to warn the driver in advance by predicting the reason for queue and if any accident ahead in the route, then it should analyze the data and alert the driver either to drive carefully to avoid accident or provide an alternate best route to the destination.

Data Description:
To build the model we will use sample data set made available by open sources (i.e. govt open portal or some research groups) in this case we use Data_Collisions.csv file. We will have to first balance the data set to avoid creating a biased model. Some of the key data elements like Vehicle location , Weather Condition, Car Speed, Light Condition, Road Condition, Junction near the incident, No of People Involved in the incident and No of Vehicles Involved in the incident will be used for building the model.

Models Used:
After analyzing the complete data set, we selected key data elements which were then used for building the model. The selected data elements were analyzed for duplicates and the selected data elements were analyzed for null records.

Model Analysis Used

- K Nearest Neighbour
- Decision Tree
- Support Vector Machine
- Logistic Regression

Data Elements Considered

ADDRTYPE
COLLISIONTYPE
LIGHTCOND
ROADCOND
WEATHER
SDOT_COLCODE
UNDERINFL
HITPARKEDCAR
INCDTTM
SEVERITYCODE

Conclusion:
The F1 score and Jaccard score is calculated for all the models, for the data that I selected the best KNN value was found to be 14. The corresponding F1 scores and Jaccard scores are shown here. Also a heatmap plot was also performed for the selected data elements.

```
#KNN
```

F1 Score: 0.7112975224027157
Jaccard: 0.7374598690124566

```
#Decision Tree
```

F1 Score: 0.6880245532349181
Jaccard: 0.7518171311159625

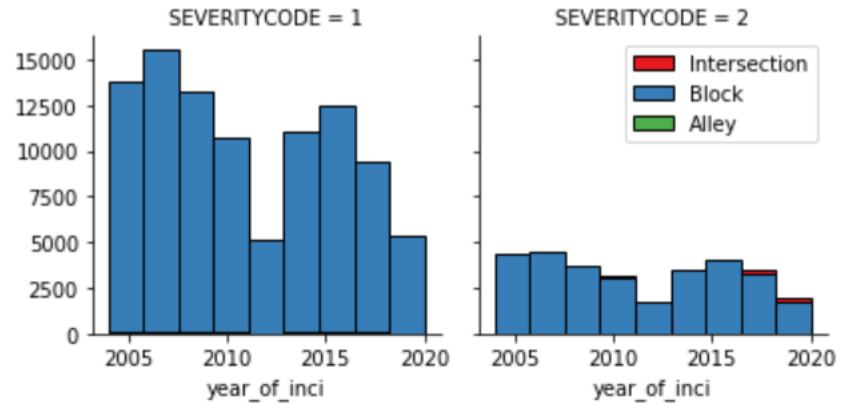
```
#Support Vector Machine
```

F1 Score: 0.6854165505731017
Jaccard: 0.7487350712726338

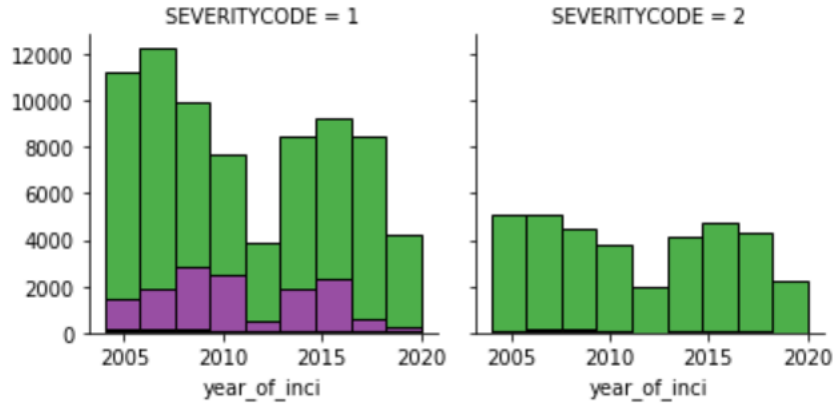
```
#Logistic Regression
```

F1 Score: 0.6585175565700814
Jaccard: 0.7256709901117246

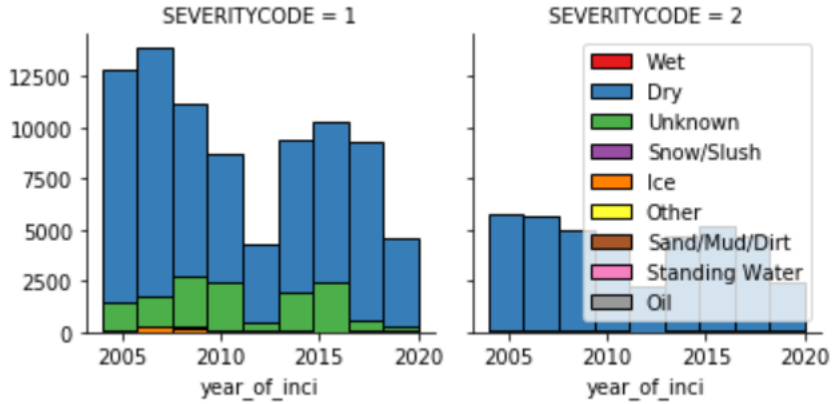
```
#analyzing the severity based on collision address
type against each year
```



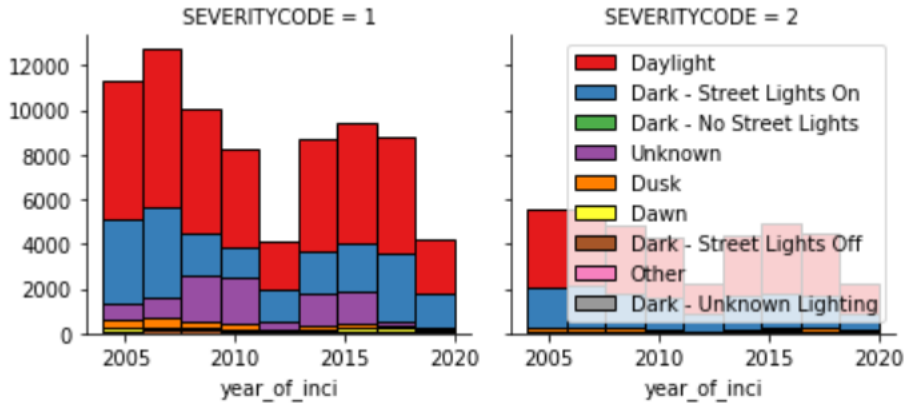

```
#analyzing the severity based on collision address
type against each year
```



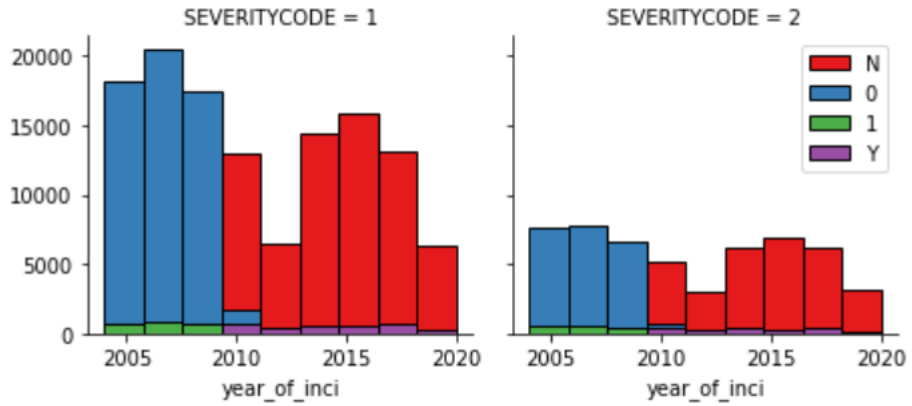

```
#analysing the severity based on road condition
against each year
```



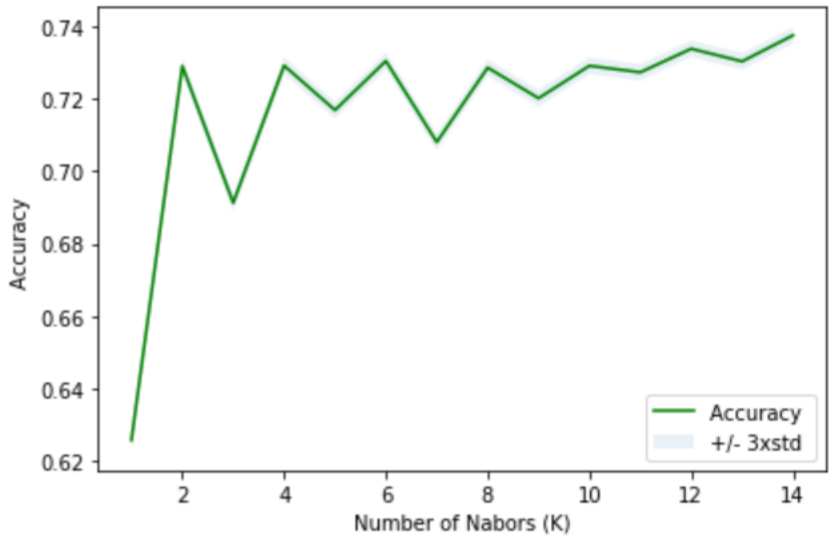

```
#analysing the severity based on light condition
against each year
```



```
#analysing the severity based on influence
condition against each year
```

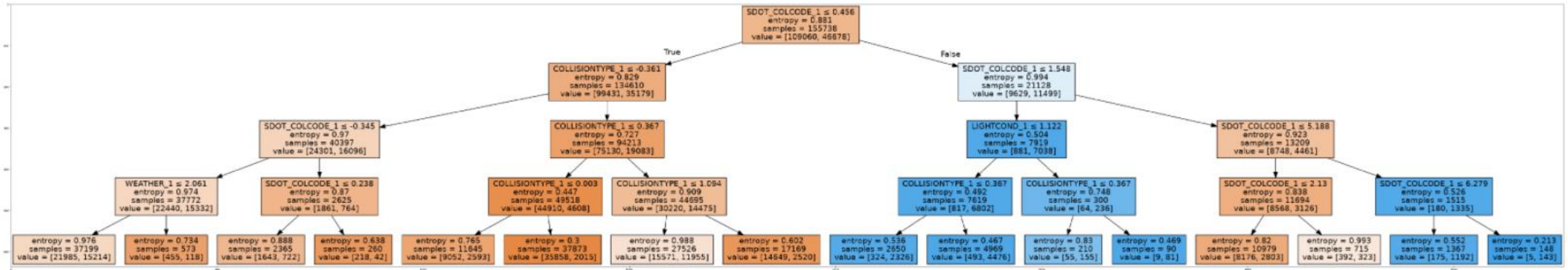



```
#K Nearest Neighbour
```

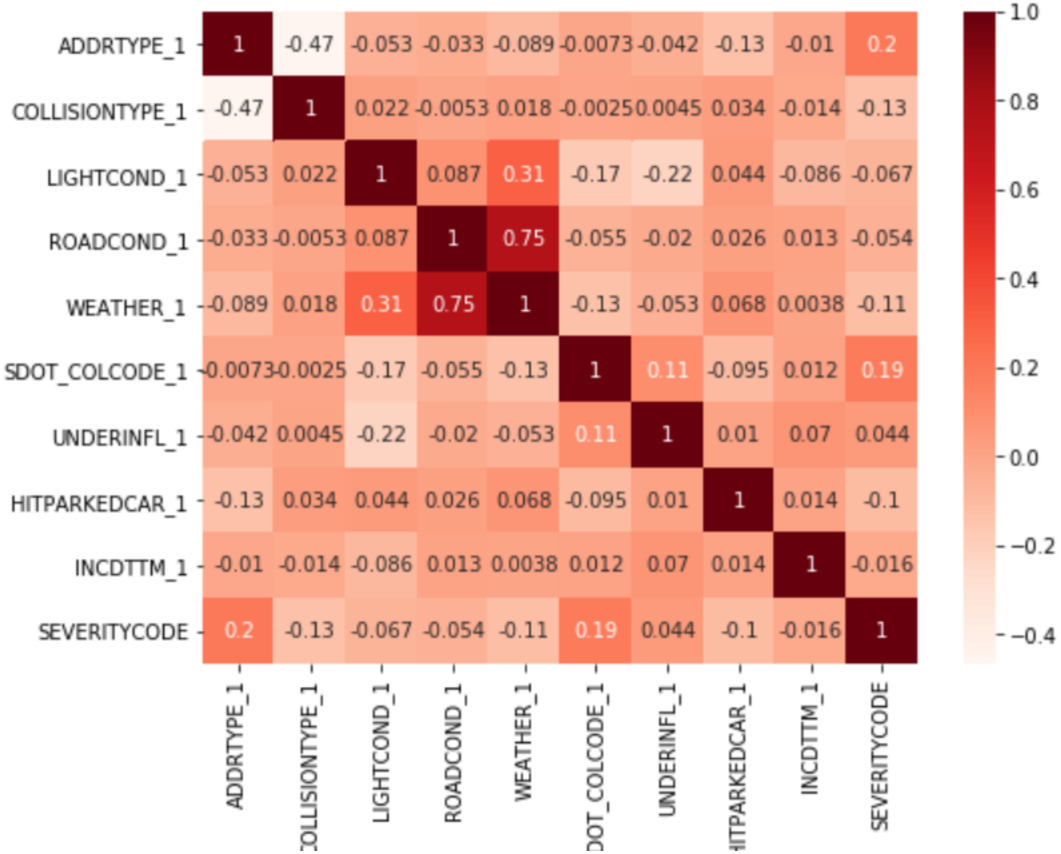



```
#Decision Tree Model
```

Out[45]: <matplotlib.image.AxesImage at 0x7fcd70727da0>



```
#Heat Map
```


ADDRTYPE_1 1 -0.47 -0.053 -0.033 -0.089 -0.0073 -0.042 -0.13 -0.01 0.2
COLLISIONTYPE_1 -0.47 1 0.022 -0.0053 0.018 -0.0025 0.0045 0.034 -0.014 -0.13
LIGHTCOND_1 -0.053 0.022 1 0.087 0.31 -0.17 -0.22 0.044 -0.086 -0.067
ROADCOND_1 -0.033 -0.0053 0.087 1 0.75 -0.055 -0.02 0.026 0.013 -0.054
WEATHER_1 -0.089 0.018 0.31 0.75 1 -0.13 -0.053 0.068 0.0038 -0.11
SDOT_COLCODE_1 -0.0073 -0.0025 -0.17 -0.055 -0.13 1 0.11 -0.095 0.012 0.19
UNDERINFL_1 -0.042 0.0045 -0.22 -0.02 -0.053 0.11 1 0.01 0.07 0.044
HITPARKEDCAR_1 -0.13 0.034 0.044 0.026 0.068 -0.095 0.01 1 0.014 -0.1
INCDTTM_1 -0.01 -0.014 -0.086 0.013 0.0038 0.012 0.07 0.014 1 -0.016
SEVERITYCODE 0.2 -0.13 -0.067 -0.054 -0.11 0.19 0.044 -0.1 -0.016 1
ADDRTYPE_1
COLLISIONTYPE_1
LIGHTCOND_1
ROADCOND_1
WEATHER_1
SDOT_COLCODE_1
UNDERINFL_1
HITPARKEDCAR_1
INCDTTM_1
SEVERITYCODE
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4