Predict Car accident severity (Manchester)

DEBASHIS MAHUNTA

Isn’t it an intriguing observation that by stopping the travel or extending lockdown will reduce or stopped road accidents? .”

- DEBASHIS

Target Audience

Tourists and travelers who are self-driving and using a rent a car option in **Manchester** want beforehand information such as which insurance to choose during booking, or which road to choose or which spot to choose based on accident severity prediction.

Rent a car company, they will use this research and advise the tourist. Or make a recommendation of car type, insurance or accessories, etc. for travel.

Insurance companies, they can use this research to produce new product offering or decide on pricing or claim settlement planning or doing further research.

Government agencies can use this information to come up with various road safety measures

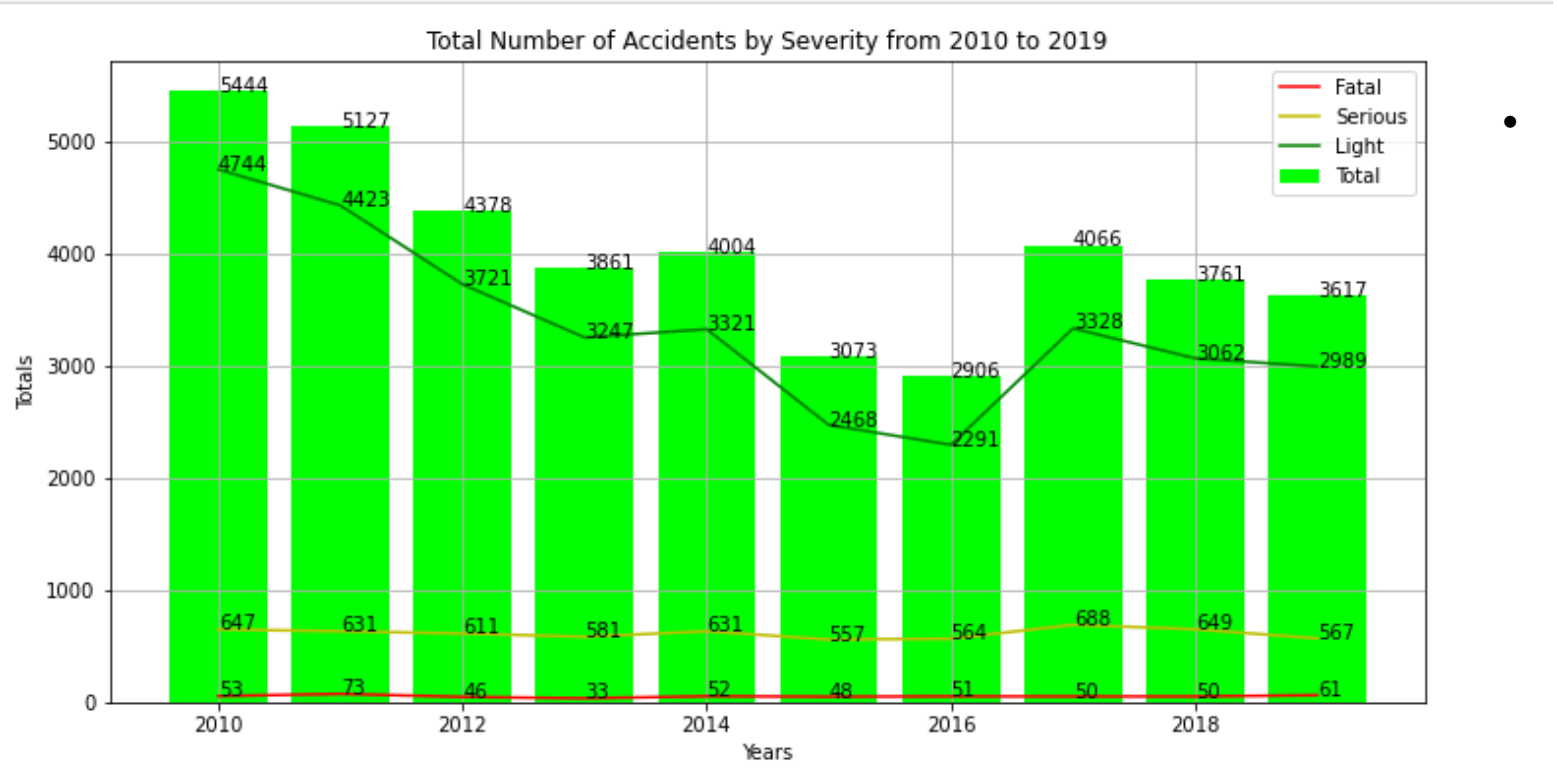
Data Source

Manchester Accidents Data is scraped from data.gov.uk.

It has Information on accidents across Manchester from the period 2010 until 2019

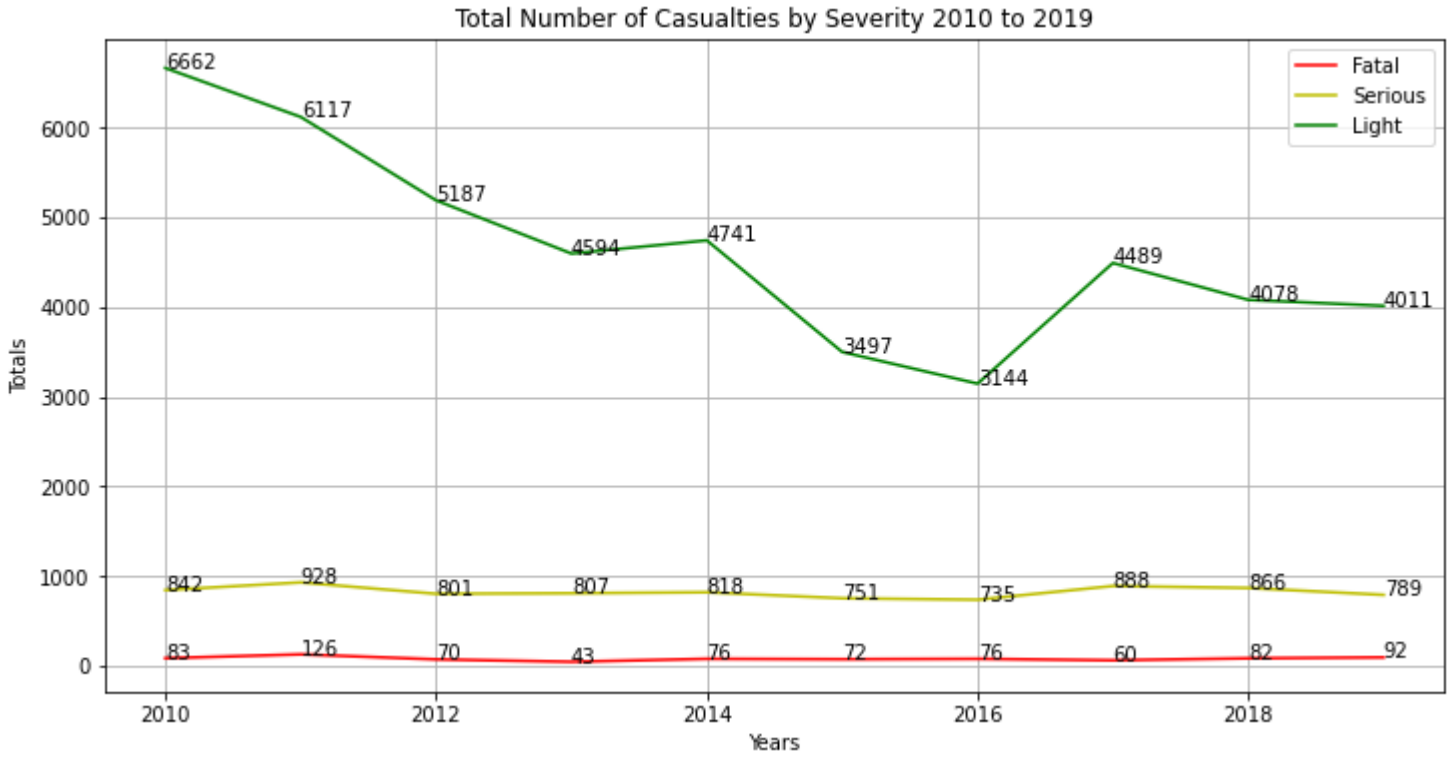
The volume of data is around 40,000 and around 27 features. Data are cleaned and converted for visualization and modeling.

Accident Trend - Manchester



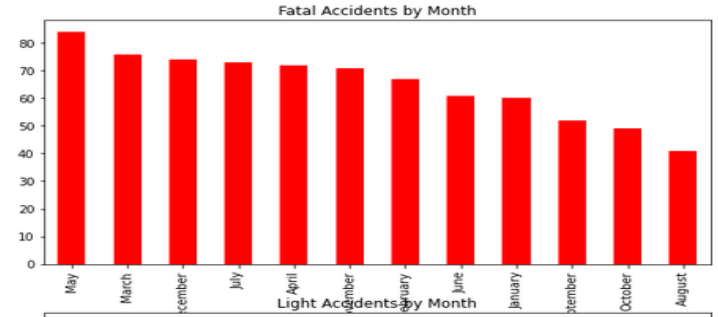
- Year 2016 is the safest year and 2010 is the unsafe year of this decade for Greater Manchester.

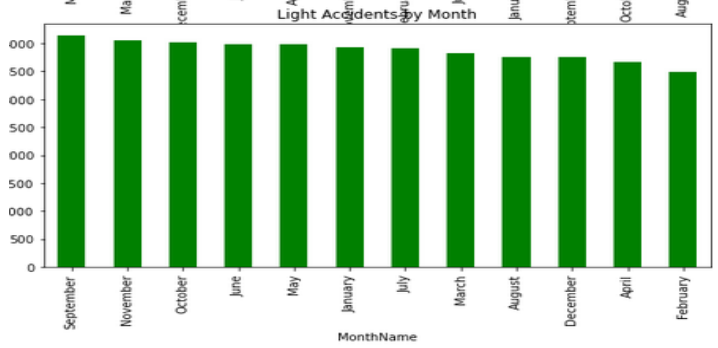
Accident Trend - Manchester

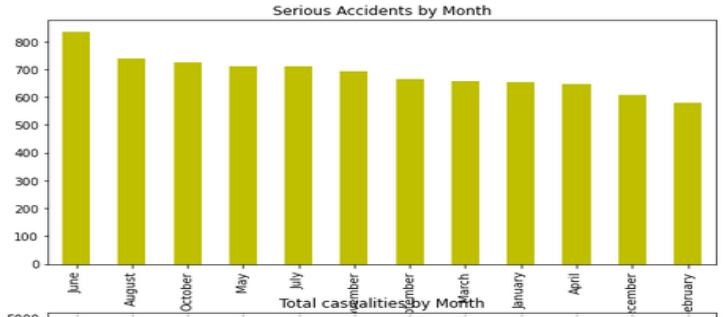


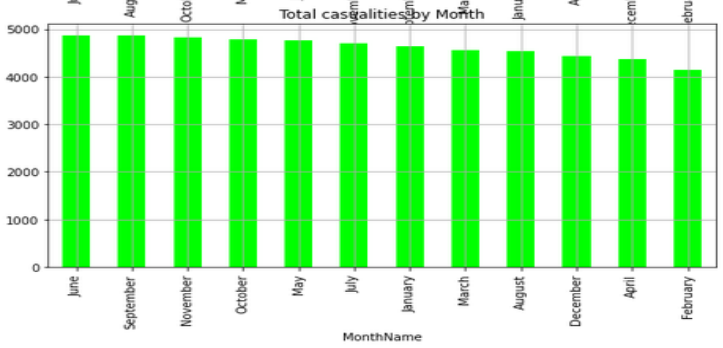
- Fatal and serious injuries remain constant over the year.
- Decreasing trend in the light type of injury.

Accident Trend - Manchester



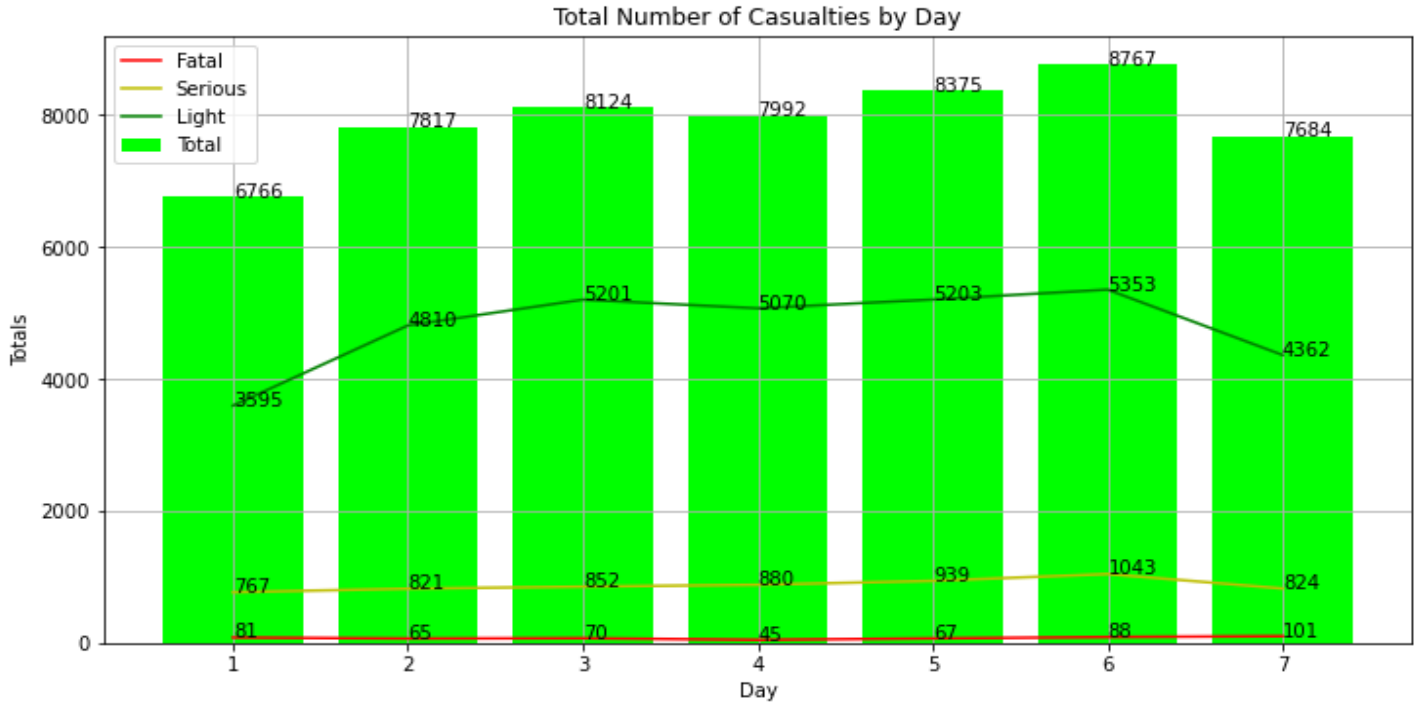






- The month of June is the most unsafe and February is the safest month

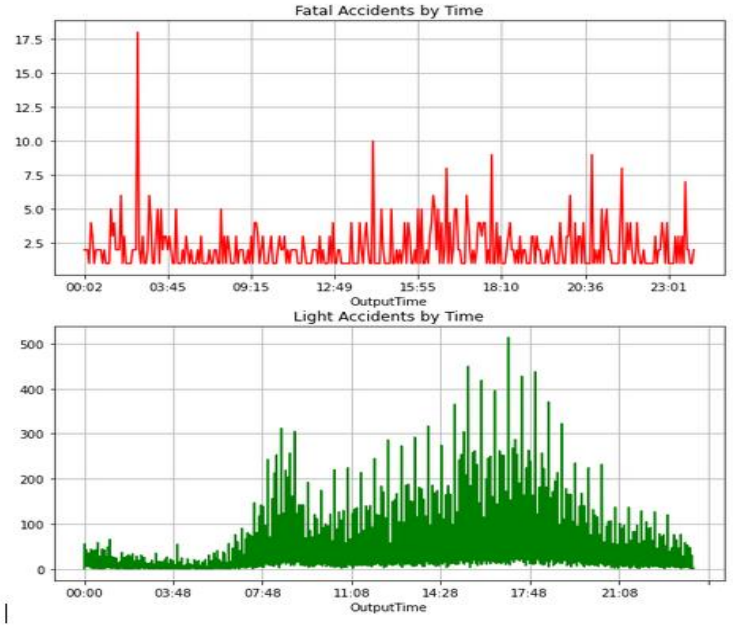
Accident Trend - Manchester

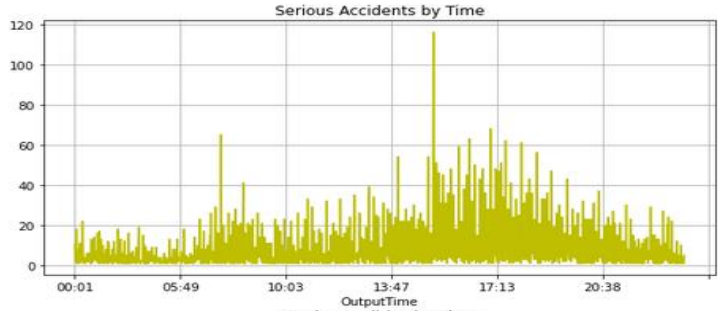


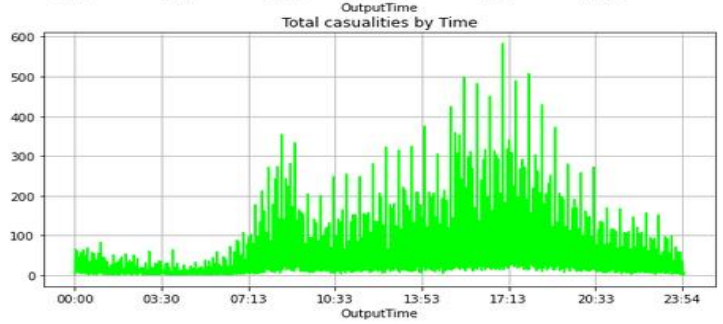
- The safest day is Sunday, and the unsafe day is on Friday

P.S : Value descriptions are : (1,Sunday), (2,Monday), (3,Tuesday), (4,Wednesday), (5,Thursday), (6,Friday), (7,Saturday).

Accident Trend - Manchester

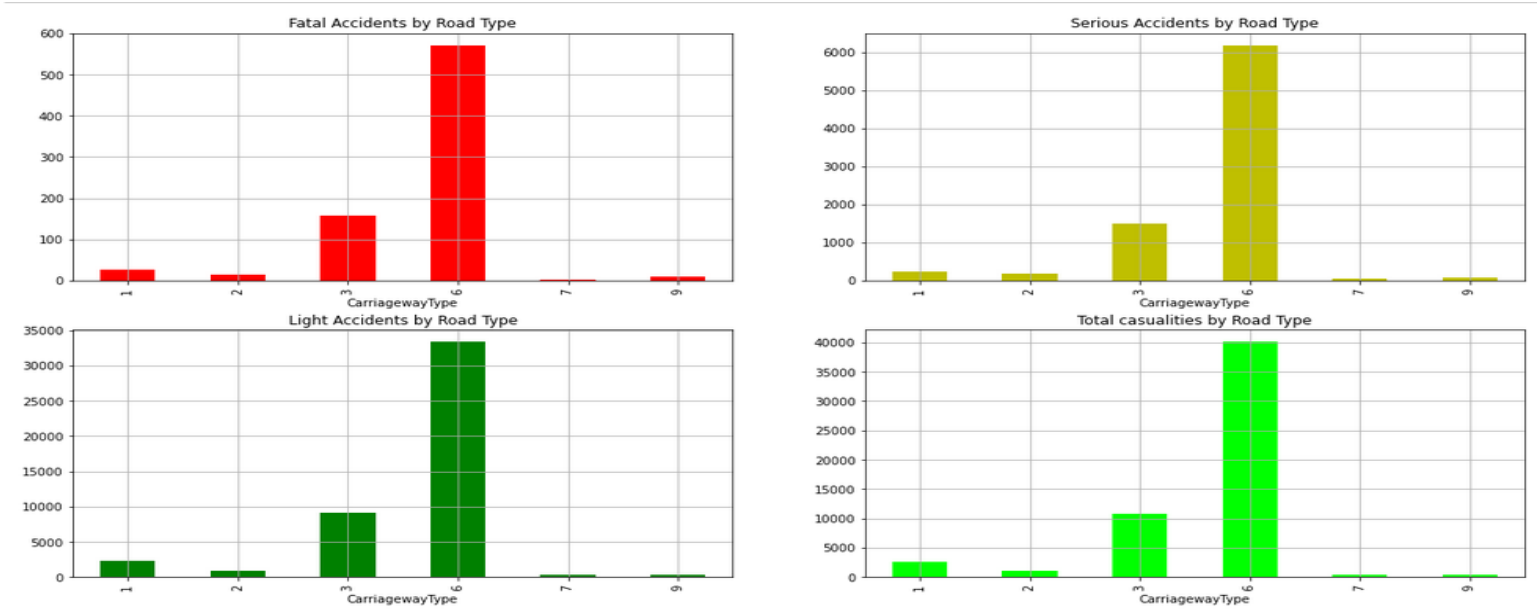






- 'Fatal' casualties are high at midnight between 2 AM to 3 AM
- 'Serious' casualties are high between 3 PM to 6 PM.
- 'Light' casualties high between 5 PM to 9 PM.

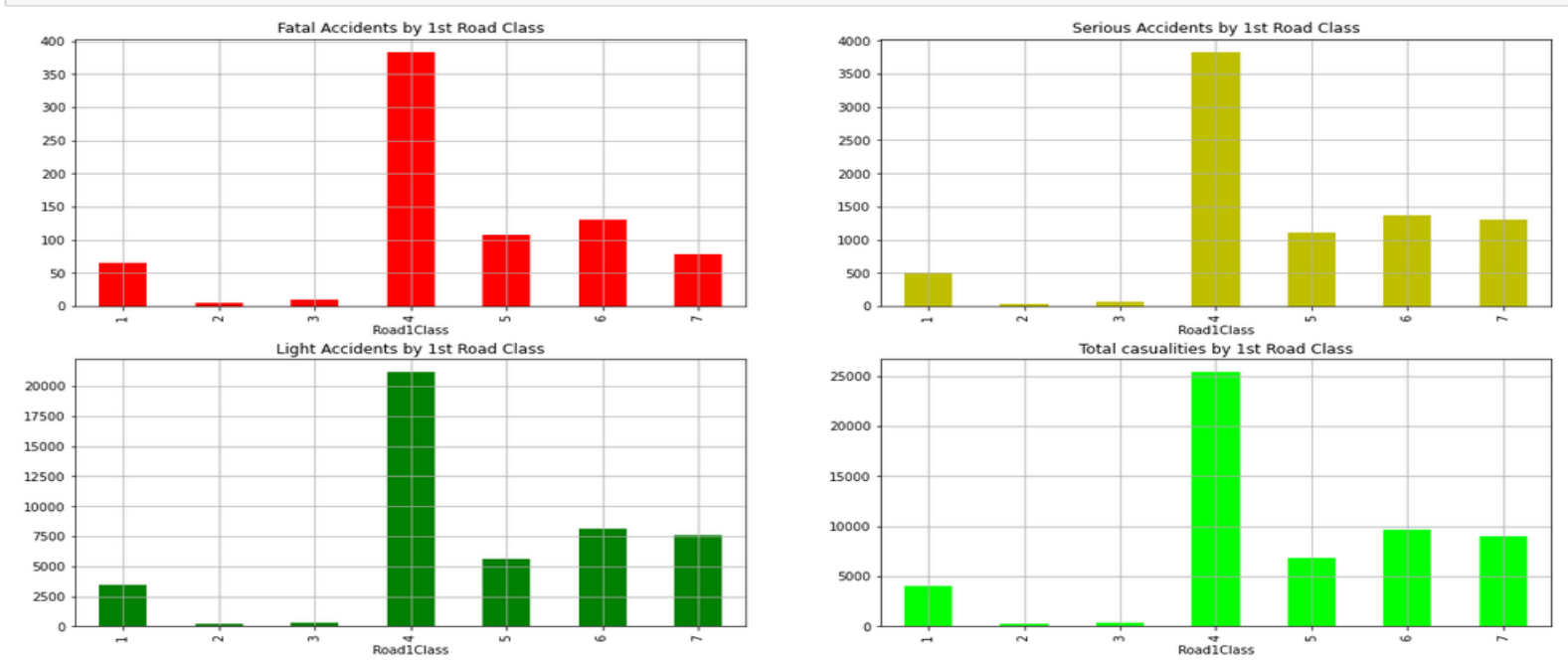
Accident Trend - Manchester



- 'Fatal', 'Serious', and 'Light' casualties are high in Single Carriageway(6) followed by Dual Carriageway (3) and Roundabout(1).

*P.S : Value descriptions are : (1,Roundabout),(2,One way street),(3,Dual carriageway),(6,Single carriageway),(7,Slip road),(9,Unknown).

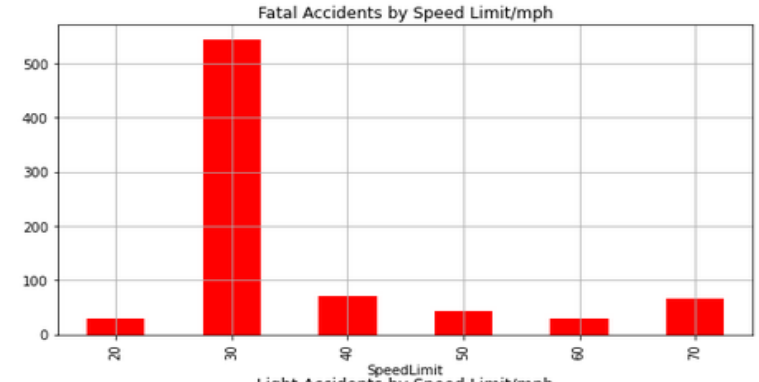
Accident Trend - Manchester

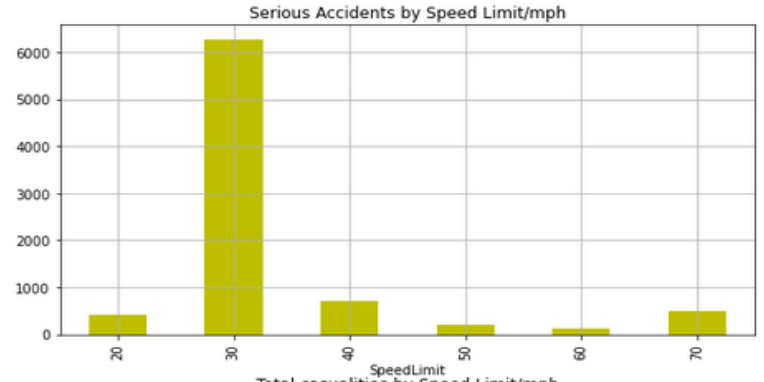


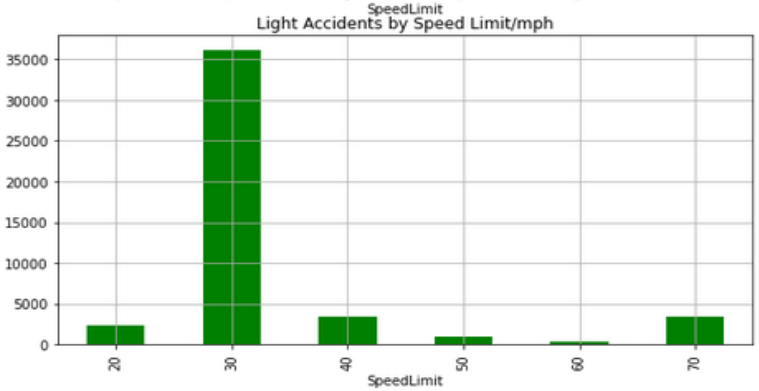
- Casualties are significantly high in Road Class B(4)

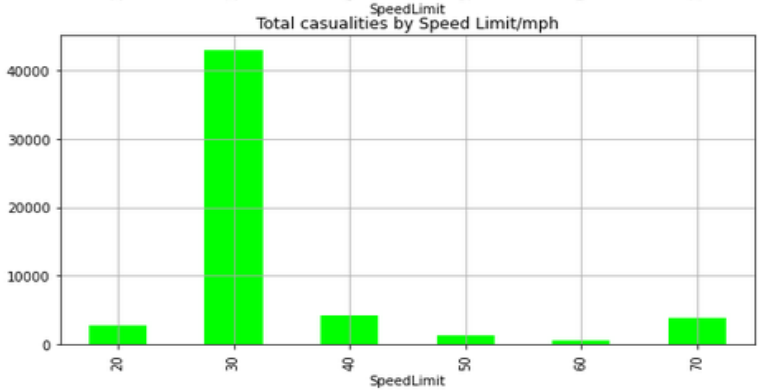
*P.S : Value descriptions are : (1,Motorway),(2,A(M)),(3,A),(4,B),(5,C),(6,Unclassified),(7,Unknown).

Accident Trend - Manchester



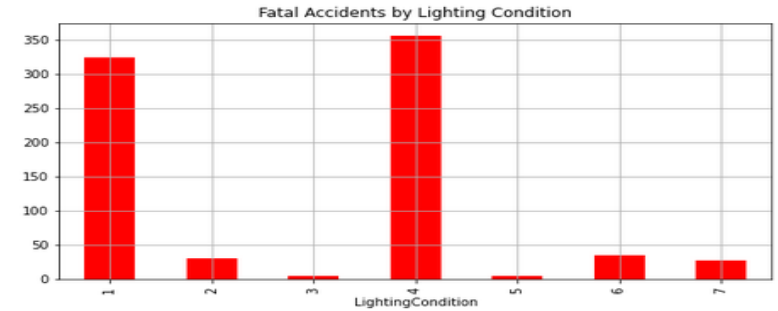


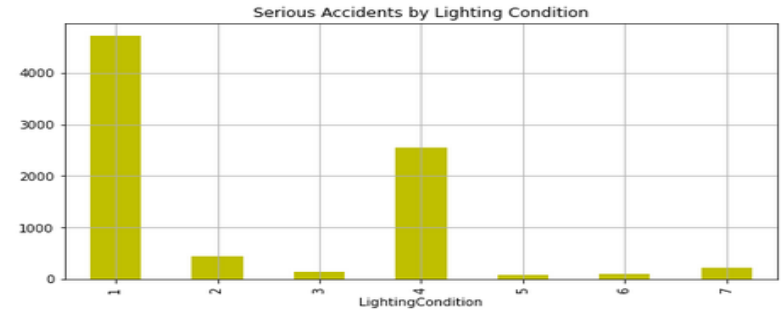


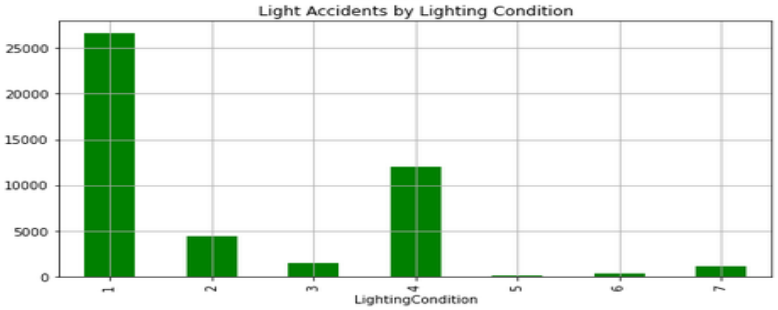


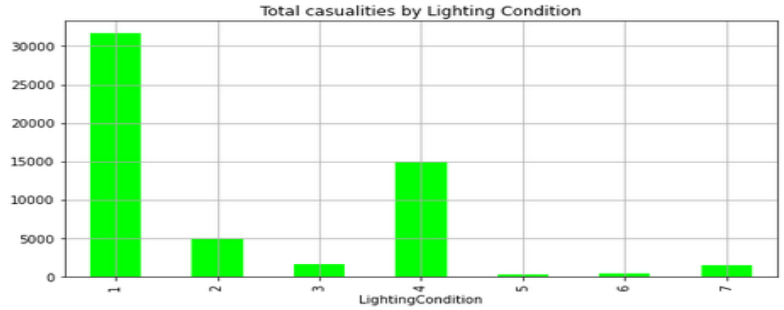
- Casualties are significantly high where the speed limit is 30 mph(80%)

Accident Trend - Manchester





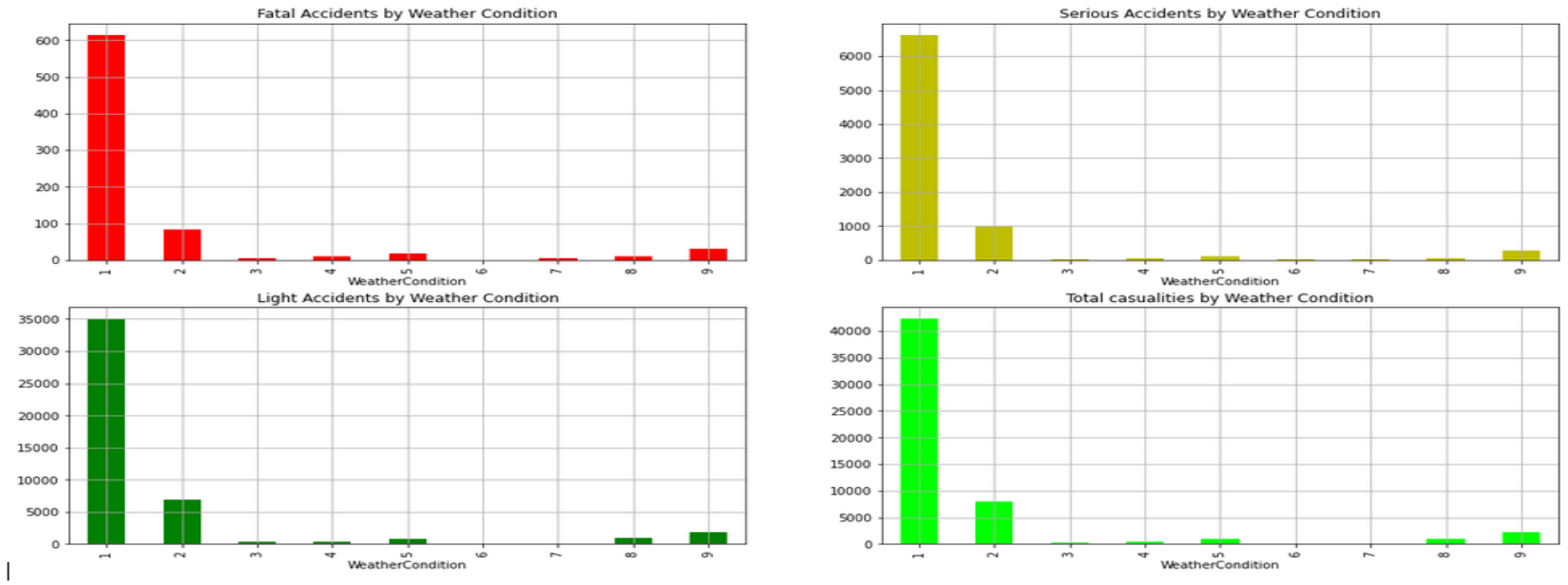




- Most of the casualties have occurred in daylight(75%),

**P.S: Value descriptions are :(1,2,3 - Daylight),(4, Darkness: street lights present and lit) (5, Darkness: street lights present but unlit),(6, Darkness: no street lighting),(7,Darkness: street lighting unknown)*

Accident Trend - Manchester

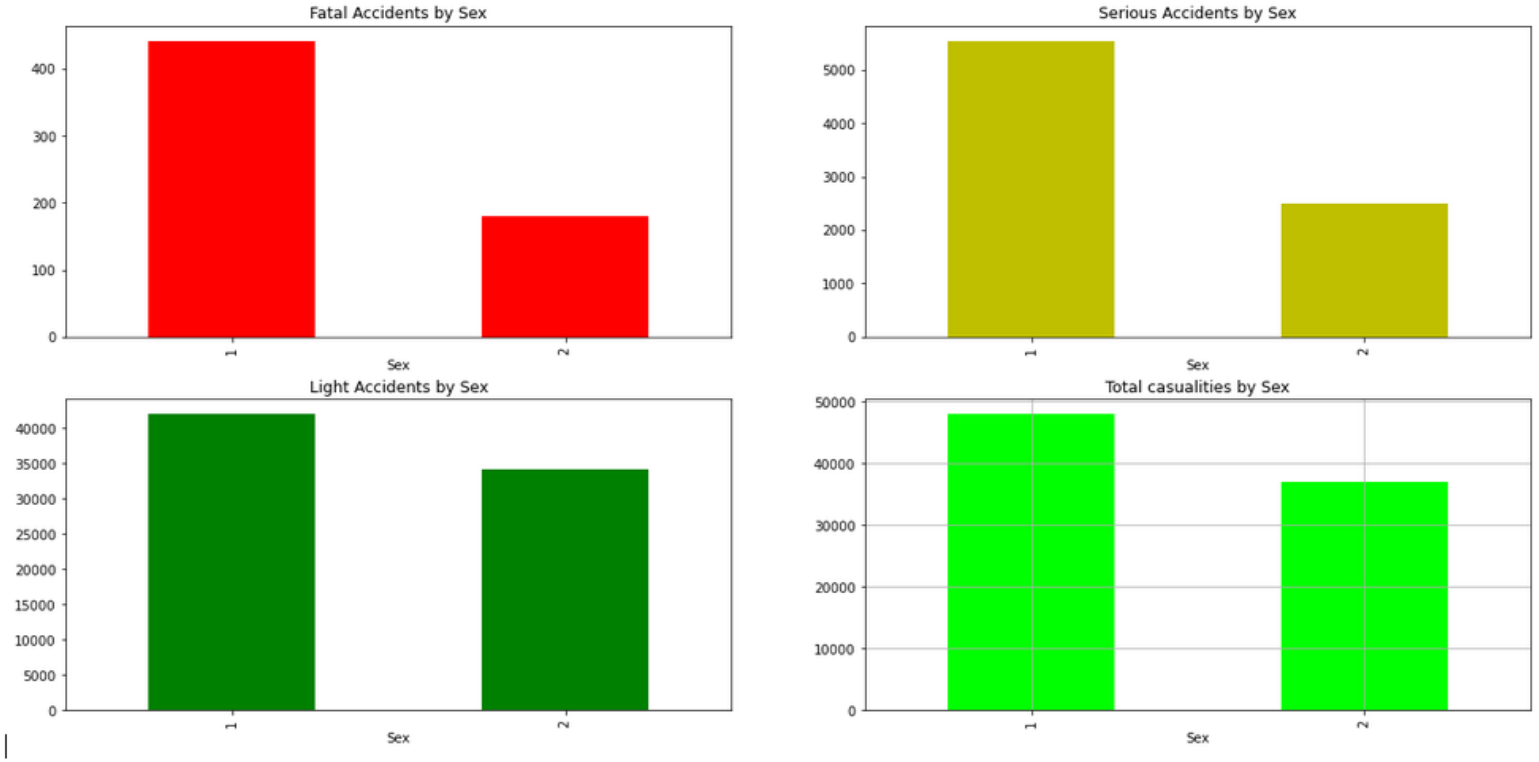


Casualties have occurred in fine weather without high winds(76%),

Around 14% of cases casualties have occurred during rain without high winds.

P.S: Value descriptions are :(1, Fine without high winds),(2, Raining without high winds) (3, Snowing without high winds),(4, Fine with high winds),(5,Raining with high winds),(6,Snowing with high winds),(7,Fog or mist — if hazard),(8,Other),(9,Unknown)

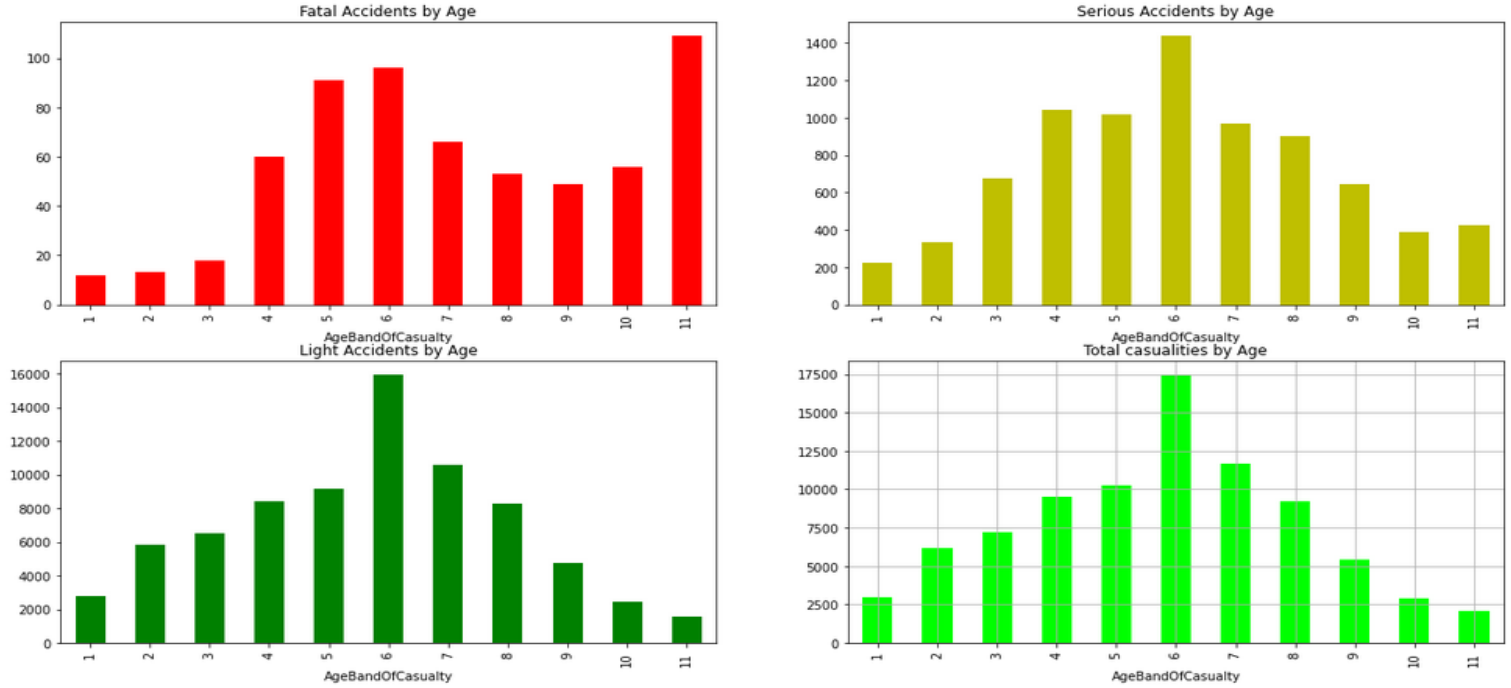
Accident Trend - Manchester



casualties are high for males (56%) and low for females (44%)

P.S: Value descriptions are :(1,Male),(2, Female)

Accident Trend - Manchester



Fatal casualty is high for age group 75 and above however, overall between age group, 26 to 35 accident casualties are the highest(20%).

P.S: Value descriptions are :(1,1 - 5),(2,6 - 10),(3,11 - 15),(4,16 - 20),(5,21 - 25),(6,26 - 35),(7,36 - 45),(8,46 - 55),(9,56 - 65),(10,66 - 75),(11,Over 75)

ML Model- K Nearest Neighbor(KNN)

KNN Model Evaluation using Test set

```
y_knn = neigh.predict(X_test)
print("KNN Model Accuracy: ",'\n',"Metrics Accuracy: ", metrics.accuracy_score(y_test, y_knn),'\n','Jaccard_similarity_sco
#let's find jaccard and F-score value to a data frame
# initialize list of lists
data1 = [['KNN',jaccard_score(y_test, y_knn,average="weighted"),f1_score(y_test, y_knn,average="weighted"),'NA']]
# add result to dataframe for summarization.
df_score = pd.DataFrame(data1,columns = ['Algorithm','Jaccard','F1-score','LogLoss'])
```

```
KNN Model Accuracy:
 Metrics Accuracy:  0.8279485904991835
 Jaccard_similarity_score:  0.6950309684054108
 F1-score:  0.7626832450161618
```

Model Accuracy
82%

ML Model- Decision Tree

Decision Tree Model Evaluation using Test set

```
y_tree = dtree.predict(X_test)
print("Decision Tree Model Accuracy: ",'\n',"Metrics Accuracy: ", metrics.accuracy_score(y_test, y_tree),'\n','Jaccard Sco
# initialize list of lists
df_score.loc[1] = ['Decision Tree', jaccard_score(y_test,y_tree,average="weighted"),f1_score(y_test,y_tree,average="weight
```

```
Decision Tree Model Accuracy:
 Metrics Accuracy:  0.8361144642476745
 Jaccard Score:  0.6990873973241758
 F1-score:  0.7614856382176787
```

Model Accuracy
83%

ML Model- Support Vector Machine(SVM)

SVM Model Evaluation using Test set

```
#let's predict
y_svm = svm_cl.predict(X_test)
print("SVM Model Accuracy: ",'\n',"Metrics Accuracy: ", metrics.accuracy_score(y_test, y_svm),'\n','Jaccard Score: ', jacc
# initialize list of lists
df_score.loc[2] = ['SVM', jaccard_score(y_test,y_svm,average="weighted"),f1_score(y_test,y_svm,average="weighted"),'NA']
```

```
SVM Model Accuracy:
 Metrics Accuracy:  0.8361144642476745
 Jaccard Score:  0.6990873973241758
 F1-score:  0.7614856382176787
```

Model Accuracy
83%

ML Model- Logistic Regression

Logistic Regression Model Evaluation using Test set

```
#let's predict
y_logr = lr.predict(X_test)
y_logr_prob = lr.predict_proba(X_test)
print("Logistic Regression Model Accuracy: ",'\n',"Metrics Accuracy: ", metrics.accuracy_score(y_test, y_logr),'\n','Jacca
# add evaluation to lists
df_score.loc[3] = ['Logistic Regression', jaccard_score(y_test,y_logr,average="weighted"),f1_score(y_test,y_logr,average="
```

```
Logistic Regression Model Accuracy:
 Metrics Accuracy:  0.8361144642476745
 Jaccard Score:  0.6990873973241758
 F1-score:  0.7614856382176787
 Log Loss:  0.48625105404152175
```

Model Accuracy
83%

ML Model Evaluation

Algorithm	Jaccard	F1-score	LogLoss
KNN	0.695031	0.762683	NA
Decision Tree	0.699087	0.761486	NA
SVM	0.699087	0.761486	NA
Logistic Regression	0.699087	0.761486	0.486251

All the model are nearly the same precision.
Hence whichever model we select we can predict with an 70% accuracy score about the severity of accidents.

Recommendation

Visitors or Residence Self Driving in Manchester

Please be careful while driving between 7 AM to 10 AM and later between 5 PM to 8 PM. There is a higher probability of accidents during this period.

Fatal casualties are highest at the midnight between 2 AM to 3 AM, so please try to avoid driving in late-night or at least be extra careful during driving.

Please be careful in T junction and be extra alert in those junctions where there is no human or physical control in place.

Recommendation

Rent a Car Companies or Insurance agencies

Your customers might be complacent during fine weather or in the daytime and likely careless while driving or choosing an insurance product. But as we observed casualties are high in fine weather or daylight. So please advise your customer accordingly.

Alternatively do not misguide or overcharged customers during bad weather or darkness. As we observed bad weather is a significantly less contributing factor for casualties.

The age group between 26 to 45 is at high chances of getting into a fatal or serious casualty. Hence, please provide all the necessary help in educating the customer group between the ages of 26 to 45 about safe driving.

Recommendation

Government agencies

We observed that single carriageways are more vulnerable for casualties, wherever possible identify those roads and place enough safety measures.

Identify T junction where casualties are high and if possible, install some physical or human control to minimize casualties.

It seems people are not observing speed limit specific to places where it is 30 mph. If possible, bring in more major in those places to minimize casualties.

Further Study

Further study is needed to map the Easting, and Nothing coordinate to identify places that are contributing casualties.

Due to limitations on the dataset not able to find out how the negligence of drivers or pedestrians is factoring casualties. To solve it we may need to reach out to the city authority to provide details about the same.

References

[Coursera] https://www.coursera.org

[Similar Study] https://www.kaggle.com/phip2014/ml-to-predict-accident-severity-pa-mont

[Other blog on driving in UK] (https://www.visitbritainshop.com/world/articles/guide-to-driving-in-the-uk/

[UK Road Classification] https://www.eastriding.gov.uk/environment/roads-streets-traffic-and-parking/roads-pavements-and-traffic/classification-of-roads/#primary-route-network