Introduction:

Road accidents significantly impacts public health and well-being, causing injuries, loss of life, and economic burdens. Data scientists analyse road statistics to improve safety measures and assist strategic decision-making. This report explores UK road statistics accident data, analysing factors like time, weather, severity determinants, and geographic distribution. The alarming statistics highlight the importance of data-driven insights in developing targeted interventions to reduce accidents, save lives, and enhance road safety..

Analysis:

My analysis is structured into three sections, each focusing on a critical aspect of road accidents: timing , location , and contextual conditions . This breakdown allows us to thoroughly examine when accidents occur, where they are concentrated, and the underlying circumstances that contribute to their unfolding.

1. **Significant Hours and Days of Accidents:**
   Data cleaning/Preprocessing

The dataset was loaded in Jupyter Notebook and converted to pandas framework for easy analysis. Four tables were loaded: accident, vehicle, casualty, and LSOA. Nan values were found in location easting and northing Osgr, and longitude and latitude columns. Cubic spline interpolation was used to replace these values, aiming for unique and close-to-original value

Analysis

I began my investigation with the timing of accidents to identify significant hours of the day and days of the week when accidents are more likely to occur. my analysis revealed, higher accident frequencies during rush hours, especially from 8:00 AM to 9:00 AM and 5:00 PM to 6:00 PM. Additionally, accidents were more frequent on Fridays.

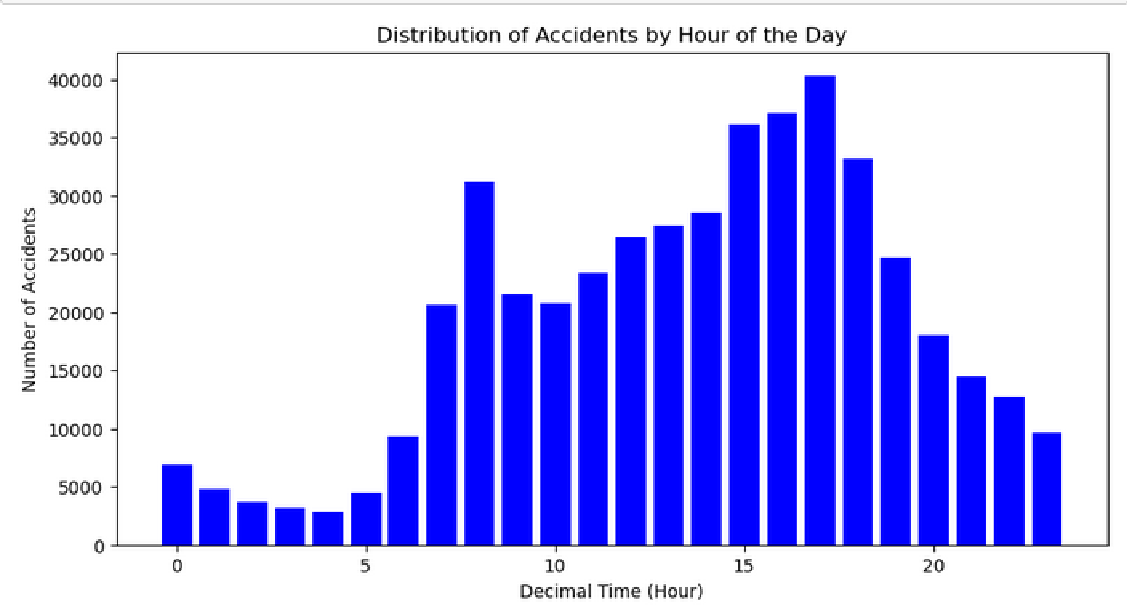

Fig 1.0: Showing the Distribution of Accidents by Hour of the Day

The chart displays a plot of accident counts against time or hours, with a spike at 8:00am due to morning rush. The peak occurs at 17:00pm, the data suggests people are trying to get home on time. This is based on a hypothesis.

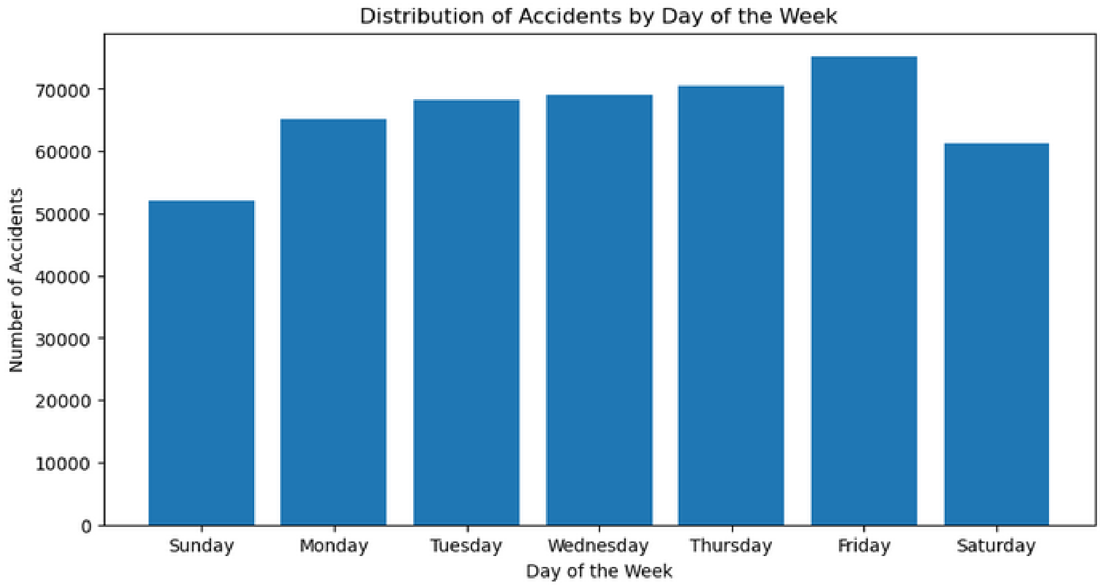

Fig 2.0: Showing the Distribution of Accidents by Day of the week

The chart displays accident counts for motorbikes rated 125cc and under (0 to 50cc), with a spike around 8am and an upward trend from 15:00 to 17:00pm. Motorcycles are less likely to be involved in accidents, as scooters, bicycles, and mobility bikes are more likely to be involved.

2.Motorbike Accidents Patterns:
Data cleaning/Preprocessing

Investigated motorbike accidents using vehicle tables and a merged dataframe, dropping duplicate columns and analyzing motorcycles based on CC grade.

Analysis

Motorcycle accidents show higher occurrences for under 125cc between 12:00 PM and 6:00 PM, while over 500cc are more prone on weekends. Accidents vary by engine capacity.

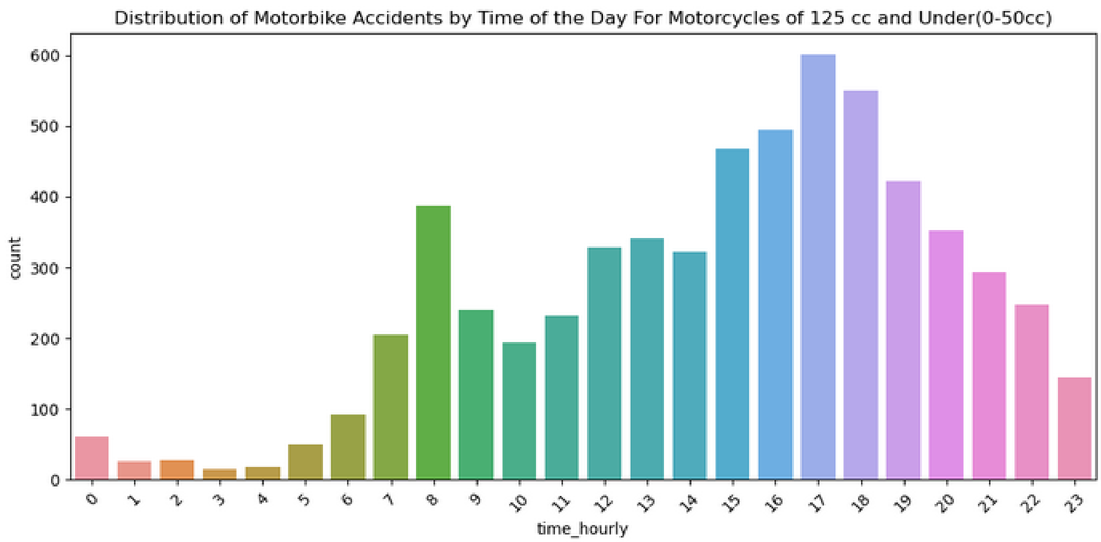

Fig 3.0: Showing the Distribution of Motorcycle Accidents by Time of the Day For Motorcycles of 125 cc and Under(0-50cc).

The chart displays accident counts for motorbikes rated 125cc and under (0 to 50cc), with a spike around 8am and an upward trend from 15:00 to 17:00pm. Motorcycles are less likely to be causes of accidents, as scooters, bicycles, and mobility bikes are more likely to be in this category.

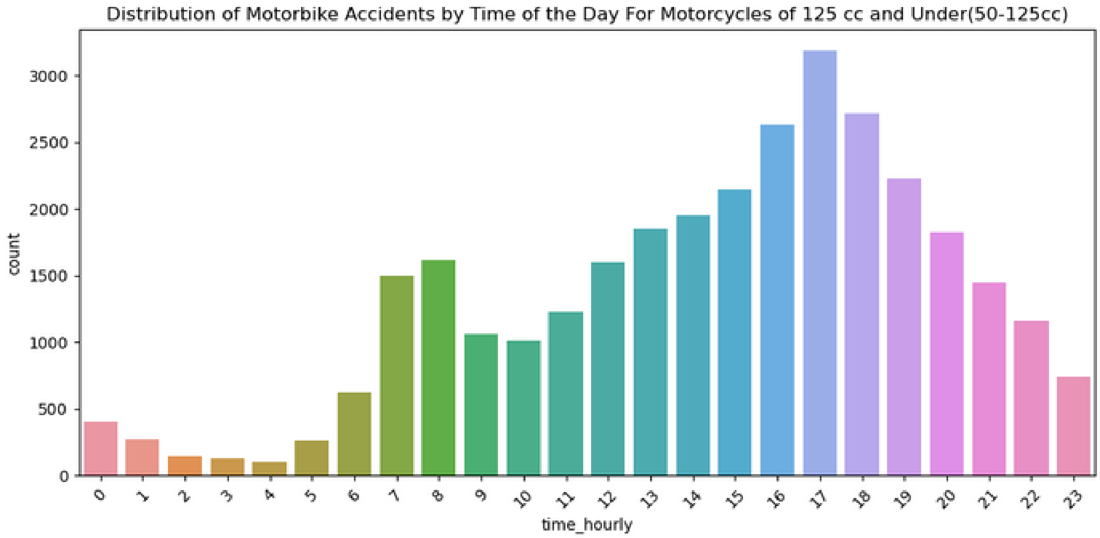

Fig 4.0: Showing the Distribution of Motorcycle Accidents by Time of the Day For Motorcycles of 125 cc and Under(50-125cc).

The chart displays accident counts for motorbikes rated 125cc and under between 50 and 125cc. The initial spike starts at 7am and increases from 11:00am to 17:00pm, dropping to 18:00pm. This suggests that motorcycles rated 125cc and under are more likely to cause accidents. Top speeds ranges from 80-65mph for these motorbikes.

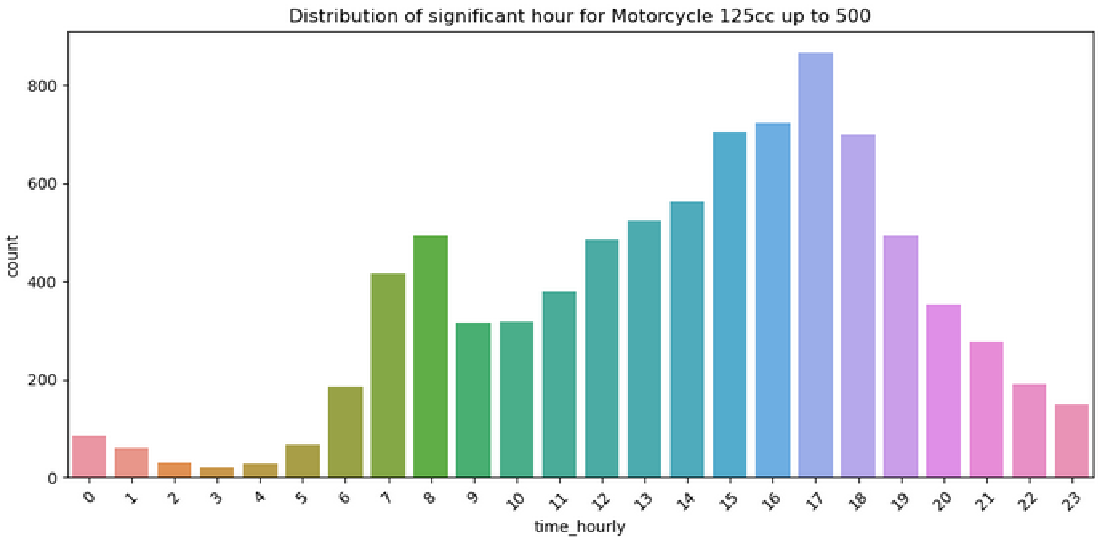

Fig 5.0: Showing the Distribution of Motorcycle Accidents by Time of the Day For Motorcycles of 125 cc and up to 500cc.

The chart displays accident counts for motorbikes ranging from 125cc to 500cc. The initial spike starts at 7am and increases from 11:00am to 17:00pm, dropping to 18:00pm. This suggests that motorcycles rated 125cc up to 500cc are more likely to cause accidents. Top speeds for these motorbikes are 155-130mph, making them a contributing facto

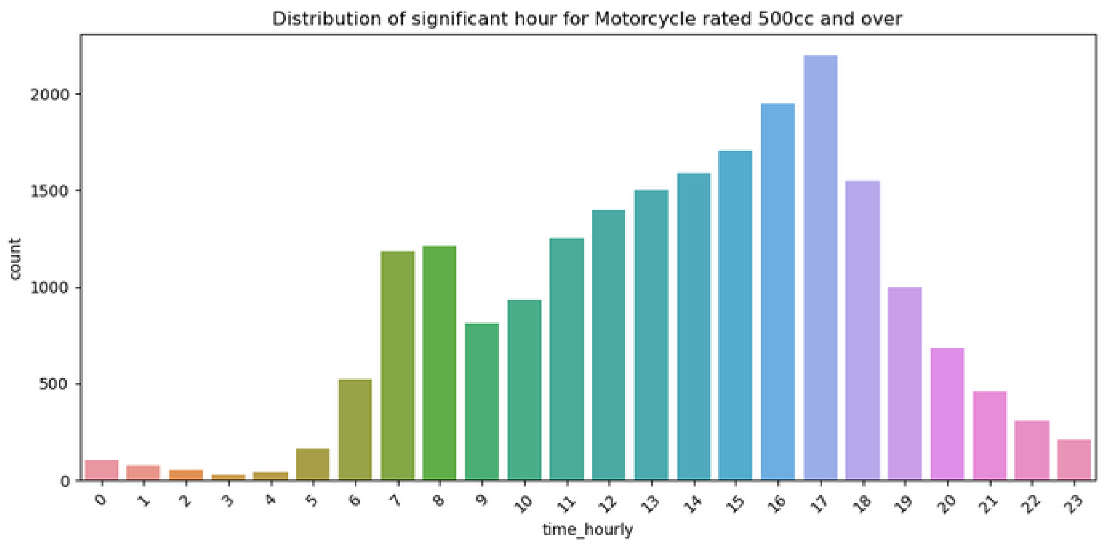

Fig 6.0: Showing the Distribution of Motorcycle Accidents by Time of the Day For Motorcycles of 500cc and over.

The chart displays accident counts for motorbikes ranging from 125cc to 500cc. The initial spike starts at 7am and increases from 11:00am to 17:00pm, dropping to 18:00pm. This suggests that motorcycles rated 125cc to 500cc are more likely to cause accidents, with a visible increase in accidents. Top speeds for these motorbikes are 155-130mph.

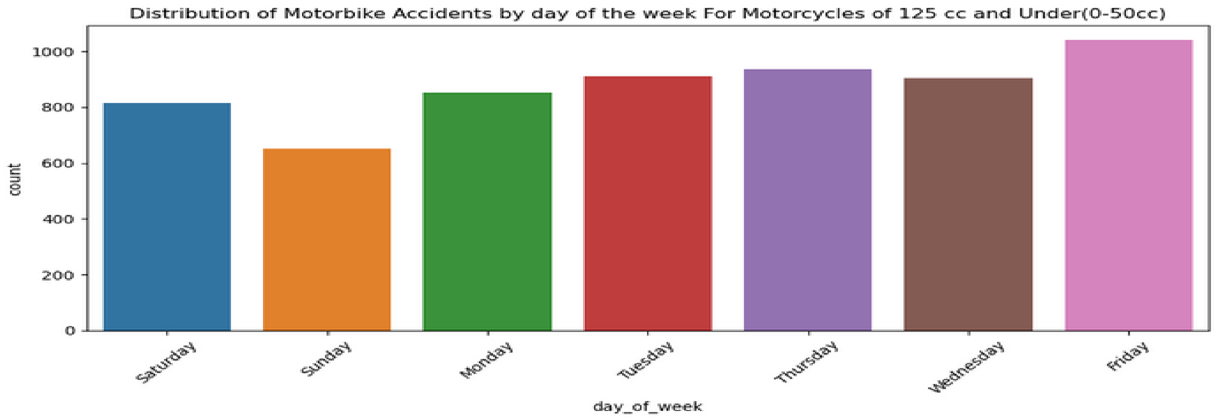

Fig 7.0: Showing the Distribution of Motorcycle Accidents by Day of the week For Motorcycles of 125 cc and Under(0-50cc).

The chart displays accident counts for motorbikes rated 125cc and under, 0 to 50cc, and the day of the week its likely to occur. It shows a trend of fewer accidents involving motorcycles, as scooters, bicycles, and mobility bikes which are more likely to be in this category.

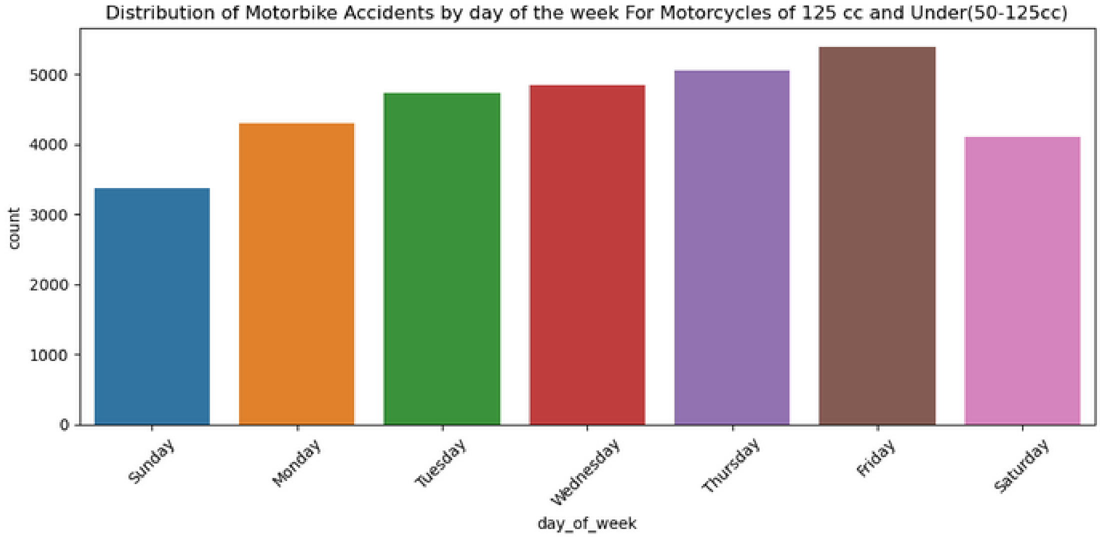

Fig 8.0: Showing the Distribution of Motorcycle Accidents by Day of the week For Motorcycles of 125 cc and Under(50-125cc).

The chart displays accident counts for motorbikes rated 125cc and under between 50 and 125cc. The mid-week difference between the bars is negligible, suggesting a higher likelihood of accidents on weekdays. Top speeds for these motorcycles range from 80-65mph, making them more likely to cause accidents.

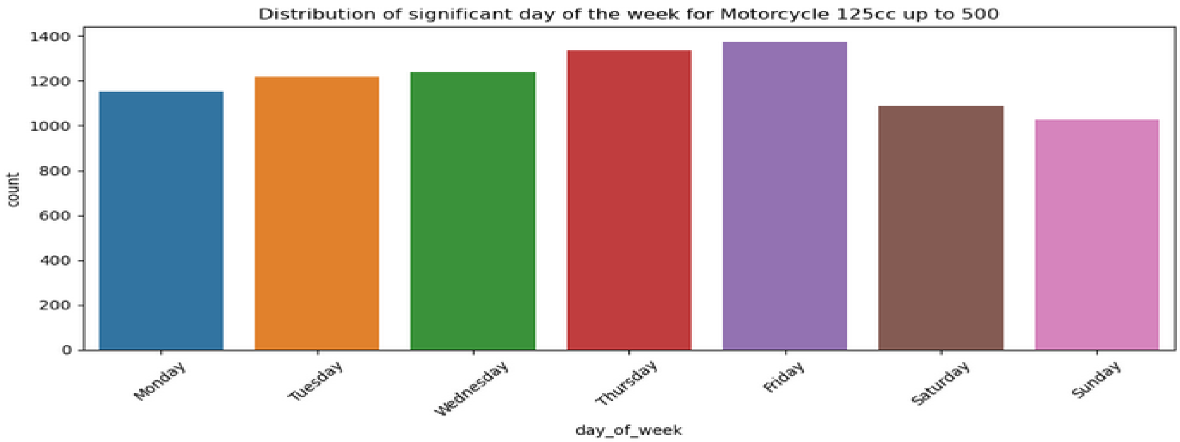

Fig 9.0: Showing the Distribution of Motorcycle Accidents by Day of the week For Motorcycles of 125 cc up to 500cc.

The chart displays an increase in accident count for motorbikes rated 125cc to 500cc, with a significant day analysis being Sunday also showing a negligible difference between the mid-week bars. This indicates a higher percentage of accidents occurring on weekdays, possibly due to the presence of motorcycles rated 125cc to 500cc.

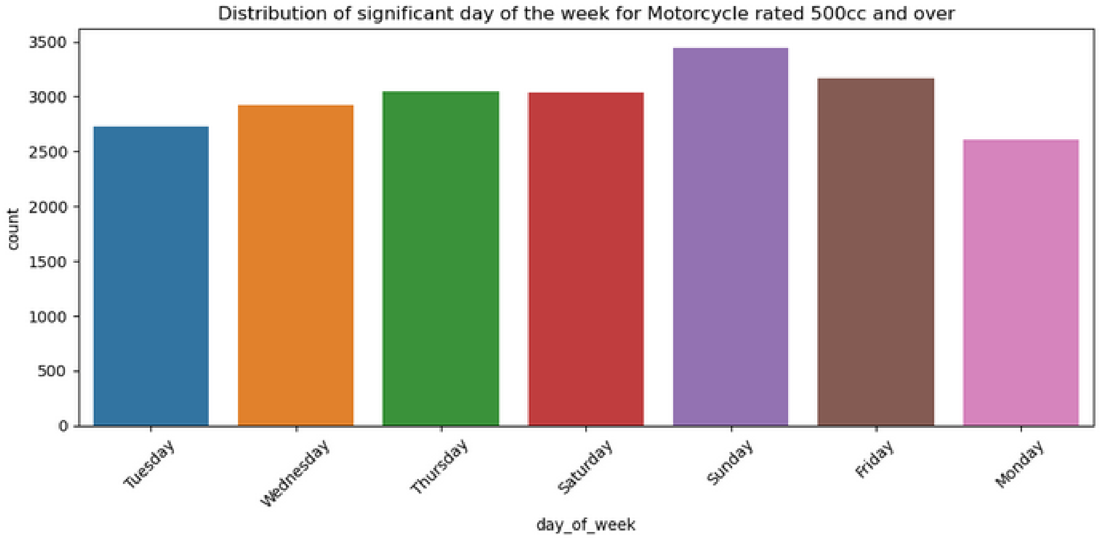

Fig 10.0: Showing the Distribution of Motorcycle Accidents by Day of the week For Motorcycles of 500cc and over.

The chart displays accident counts for motorbikes ranging from 125cc to 500cc, with Friday showing a clear trend. The other days follow closely, suggesting a higher percentage of accidents involving motorcycles rated 125cc to 500cc. Top speeds for these motorbikes are 155-130mph, making them a potential contributing factor

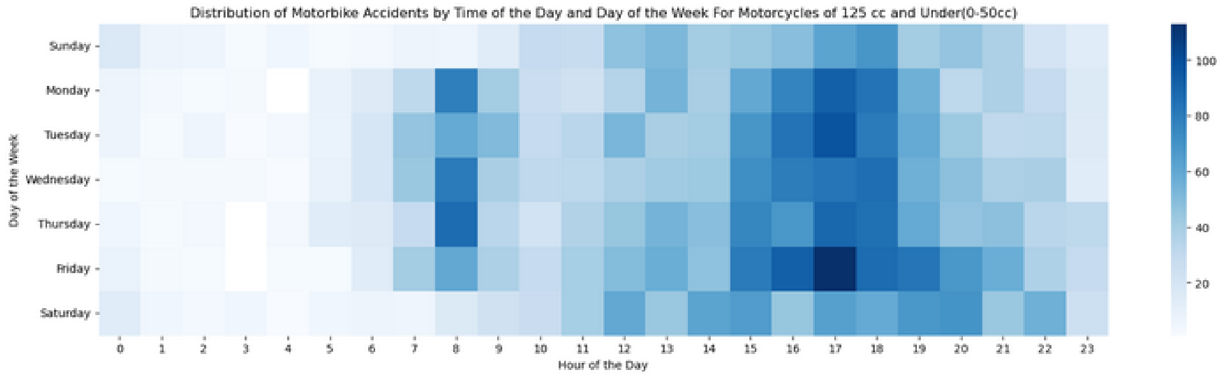

Fig 11.0: Showing the combined Distribution of Motorcycle Accidents by Day of the week and hour of the day For Motorcycles of 125 cc and Under(0-50cc).

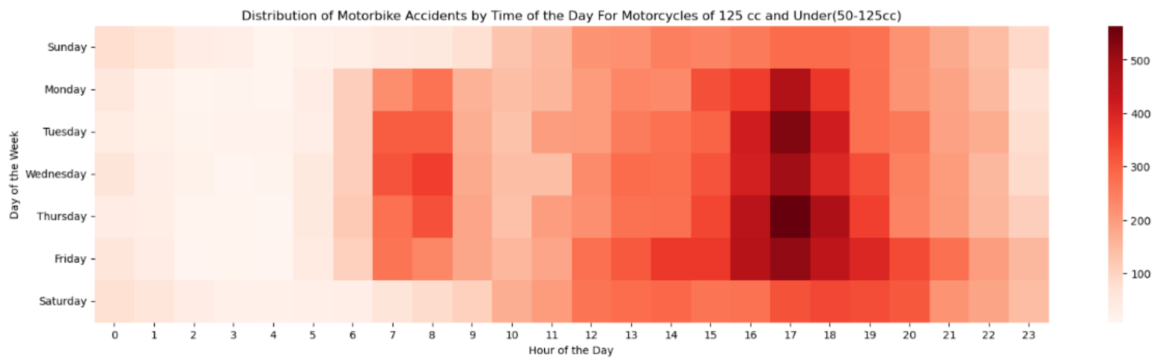

Fig 12.0: Showing the combined Distribution of Motorcycle Accidents by Day of the week and hour of the day For Motorcycles of 125 cc and Under(50-125cc).

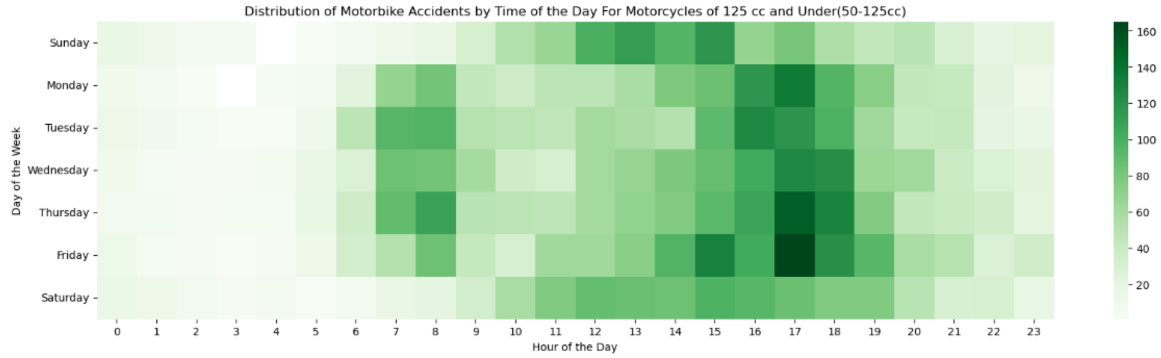

Fig 13.0: Showing the combined Distribution of Motorcycle Accidents by Day of the week and hour of the day For Motorcycles of 125 cc up to 500cc.

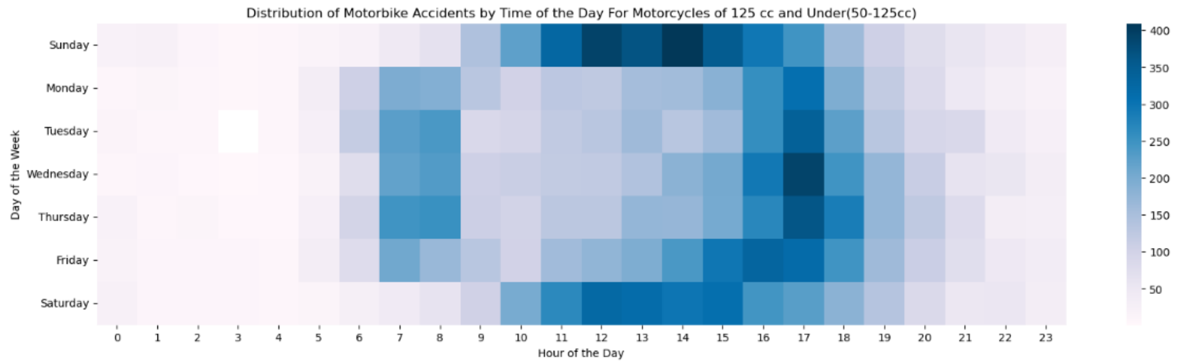

Fig 14.0: Showing the combined Distribution of Motorcycle Accidents by Day of the week and hour of the day For Motorcycles of 500cc and over.

2. **Pedestrian Involvement in Accidents:**

   Data cleaning/Preprocessing

Investigated pedestrian accidents using casualty table and daily accident data, merging to create a new dataframe, dropping duplicate columns, and focusing on casualty class pedestrians.

Analysis

Analysing pedestrian accidents, I uncovered that pedestrian involvement peaks during the afternoon rush hours, particularly at 3:00 PM. There was a notable rise in pedestrian accidents on Fridays, aligning with increased weekend activity.

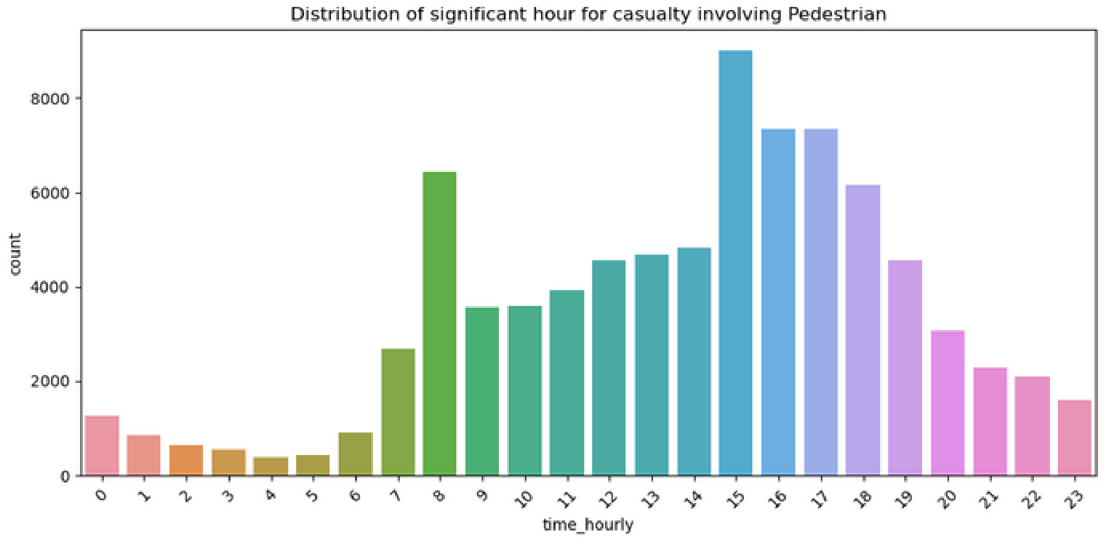

Fig 15.0: Showing the Distribution of significant hour of the day for casualty involving Pedestrian.

The chart shows pedestrian accident counts, with a spike around 8am and a sharp drop at 15:00. The 15pm spike may be due to children returning from school.

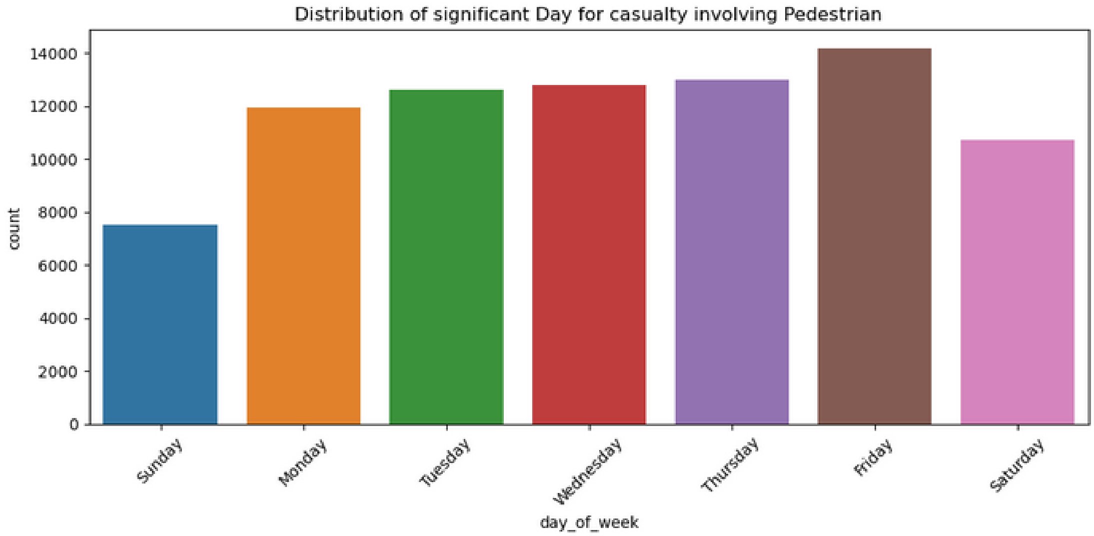

Fig 16.0: Showing the Distribution of significant day of the week for casualty involving Pedestrian.

The chart above shows the accident count for pedestrian, looking at the chart we observe the same trend observed in the significant day analysis looking at the chart we observe the same trend observed in the significant day analysis which is the difference between the bars mid-week is almost negligible which signifies the likelihood of accidents occurring weekdays.

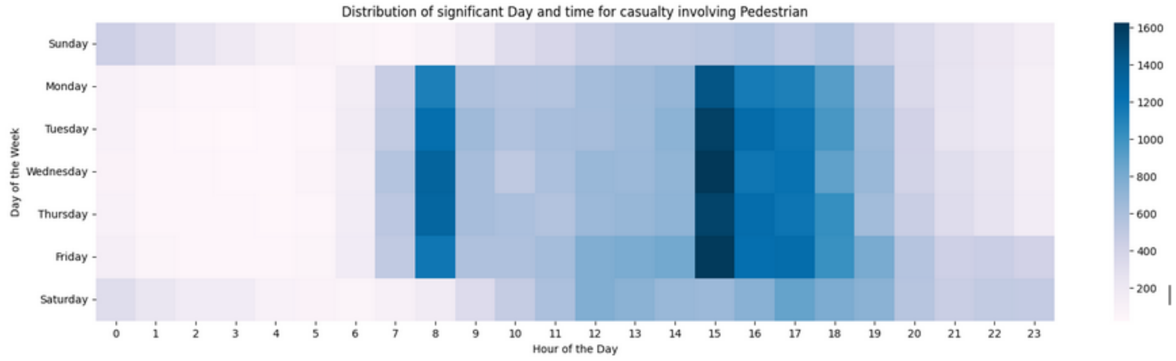

Fig 17.0: Showing the combined Distribution of pedestrian Accidents by Day of the week and hour of the day.

3. **Impact of Selected Variables on Accident Severity:**
   Data cleaning/Preprocessing

Investigating accident severity using selected variables, concatenated accident, vehicle, and casualty tables, and querying. Dropping -1 column and filtering initial rules to higher support and confidence values.

Analysis

Employing the apriori algorithm, we explored the influence of selected variables on accident severity. Our analysis showcased how factors like weather conditions, speed limits, and road surface contribute to the severity of accidents, highlighting areas for targeted safety measures

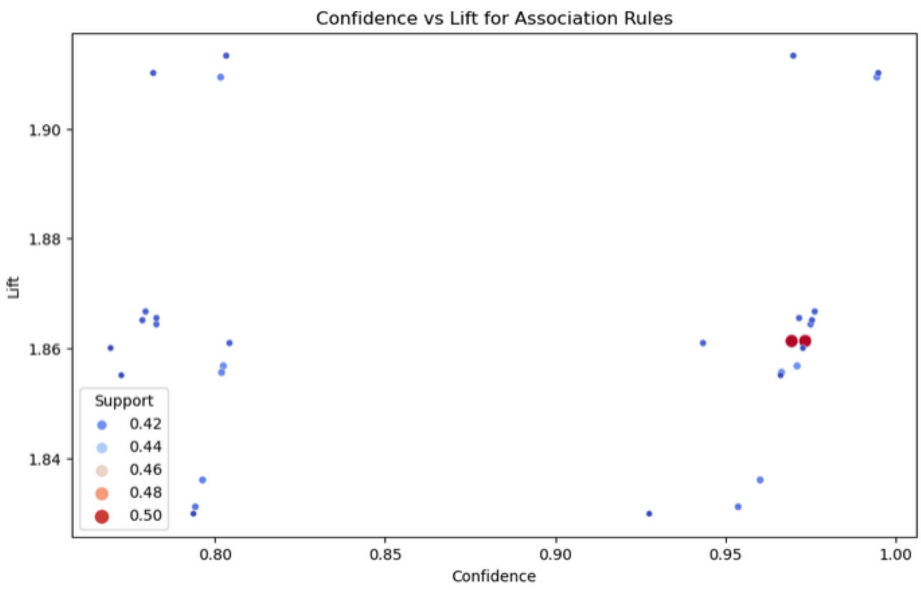

Fig 18.0: Showing Confidence vs Lift for Association Rules.

The chart displays confidence and lift values for accident severity, speed limit, weather, time, road type, surface conditions, special conditions at site, light, pedestrian crossing facilities, human control, and day of the week. Red dots indicate support for these variables.**Clustering Analysis for**

4.Regional Insights:

Data cleaning/Preprocessing

Investigating regional insight using clustering, imported libraries, and grouped data by location in Humberside, east riding of Yorkshire, Kingston upon hull.

Analysis

By clustering accidents in Kingston upon Hull, Humberside, and the East Riding of Yorkshire, I unveiled distinct spatial patterns. Clusters revealed accident hotspots in urban areas and potential correlations with specific road features, providing insights for focused interventions. I plotted two set of charts.

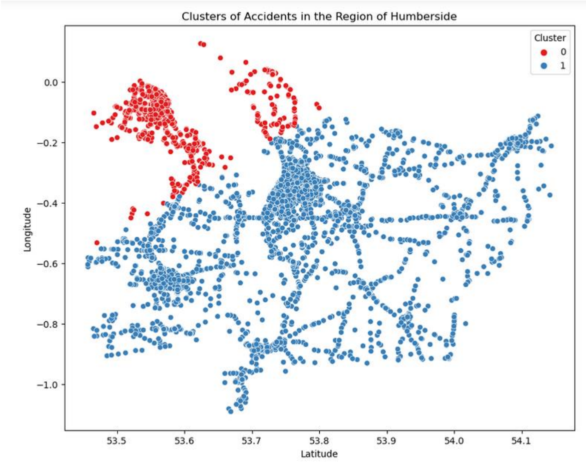

Fig 19.0: Showing Clusters of Accidents in the Region of Humberside(without other conditions).

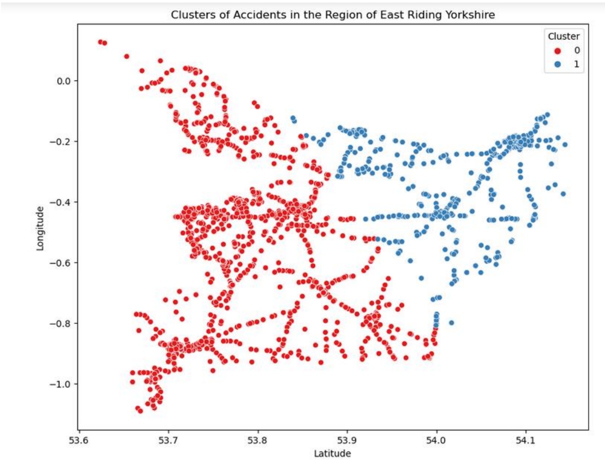

Fig 20.0: Showing Clusters of Accidents in the Region of east riding of yorkshire(without other conditions).

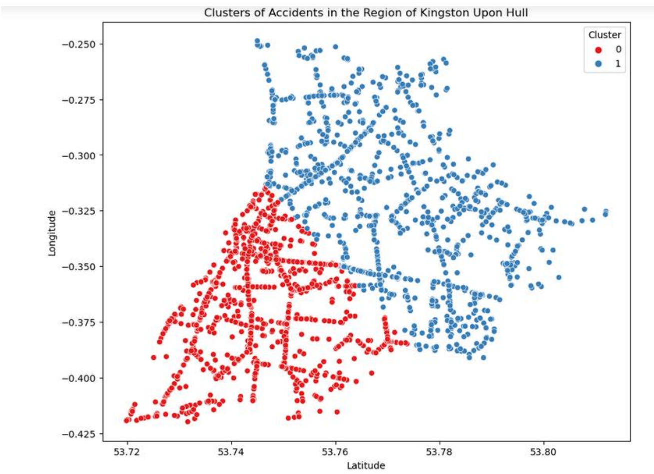

Fig 21.0: Showing Clusters of Accidents in the Region of Kingston upon hull(without other conditions).

And the other chart having conditions number of casualties, number of vehicles , urban or rural area we see the clusters take on a different form

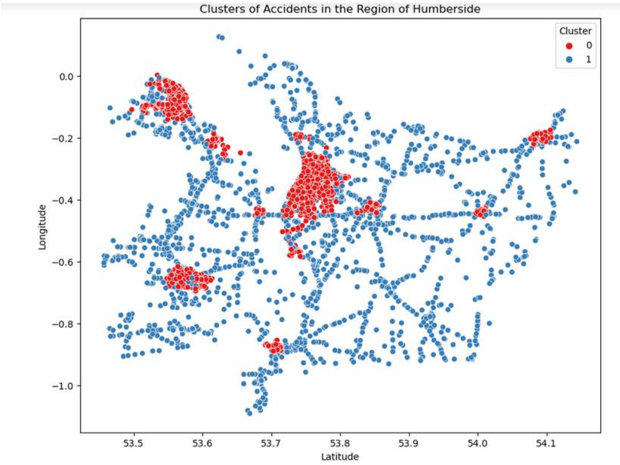

Fig 22.0: Showing Clusters of Accidents in the Region of Humberside(with other conditions).

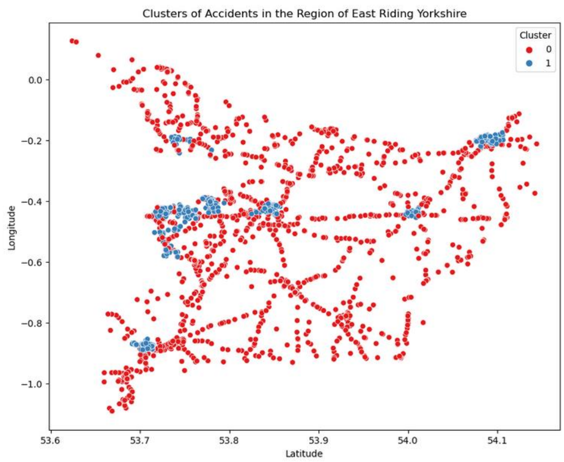

Fig 23.0: Showing Clusters of Accidents in the Region of east riding of yorkshire (with other conditions).

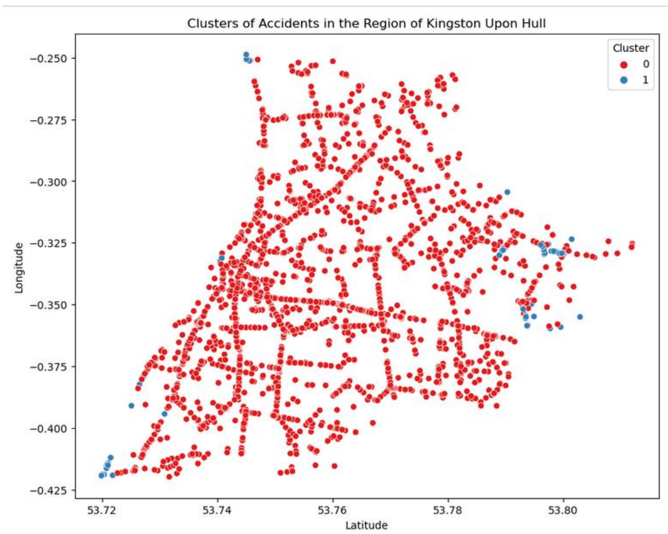

Fig 24.0: Showing Clusters of Accidents in the Region of east riding of yorkshire (with other conditions).

I presume this is due to the urban and rural column as these have more cars or population density in comparison to rural areas.

4. Outlier Detection and Relevance:

Data cleaning/Preprocessing

Beginning investigation for outlier detection, I first imported the necessary libraries, I proceeded to group my data by set based on the specified location of Humberside, east riding of Yorkshire and Kingston upon hull using the police force and local authority code to filter them out

Analysis

Utilizing outlier detection methods, we identified unusual entries within the dataset. While outliers may contain valuable information, careful consideration is necessary to determine their impact on analysis. Decisions on retaining or excluding outliers should be made in accordance with specific objectives.

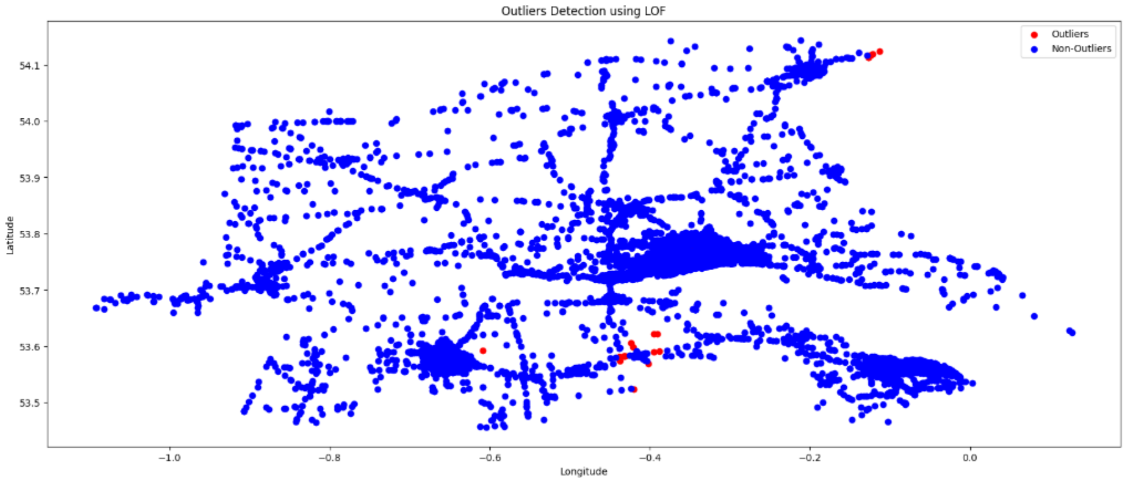

Fig 25.0: Showing Outliers Detection using LOF for Humberside

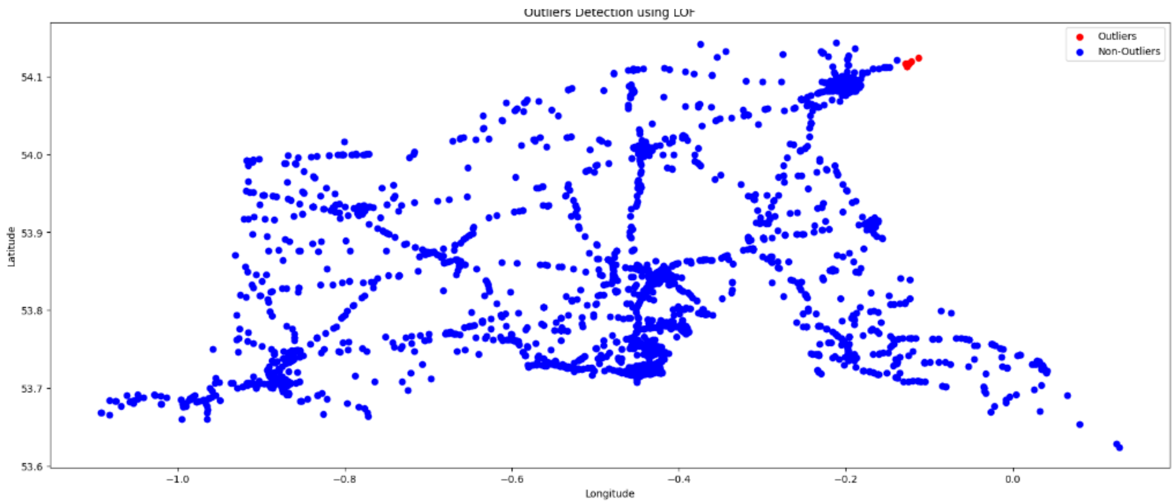

Fig 26.0: Showing Outliers Detection using LOF for east riding of Yorkshire

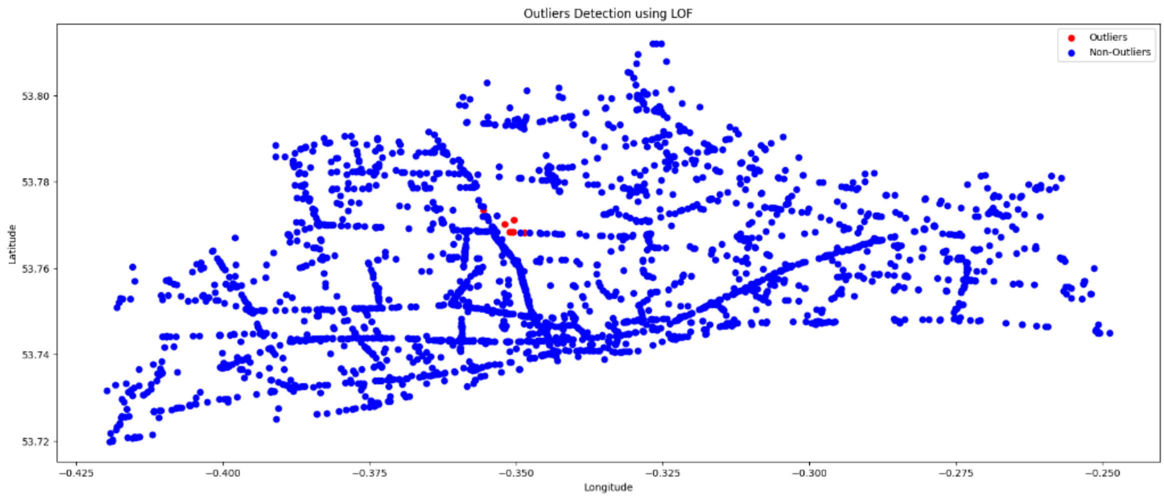

Fig 27.0: Showing Outliers Detection using LOF for Kingston upon hull

Charts display outliers in Humberside, East Yorkshire, and Kingston, containing only longitude and latitude. LOF algorithm separates anomalies, it indicates unique circumstances contributing to accident likelihood by separating them by dots. Blue dots indicate consistent data points with similar patterns, red unique or different circumstances.

5.Predictions:

Data cleaning/Preprocessing

Beginning investigation for outlier detection, I first imported the necessary libraries, I proceeded to group my data by set based on the specified location of Humberside, east riding of Yorkshire and Kingston upon hull using the police force and local authority code to filter them out . I proceeded to merge the accident table, vehichle table and casualty table.

Analysis

To address the prediction of fatal injuries, we developed a classification model using the provided data. Our model accurately predicts fatal injuries sustained in road traffic accidents by considering variables such as weather conditions, vehicle types, and speed limits. This predictive capability offers potential for informed decision-making and improved road safety strategies.

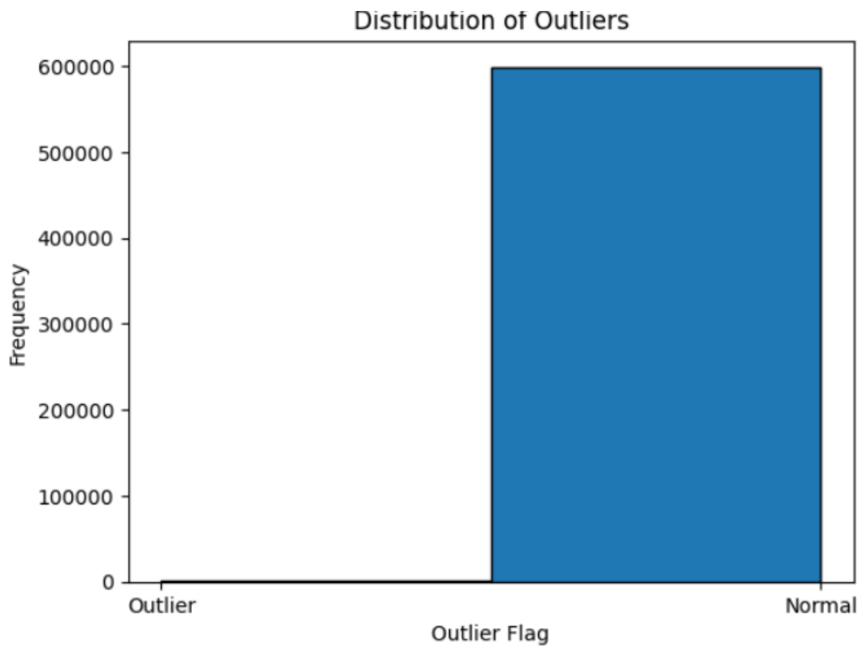

Fig 28.0: Showing Distribution of Outliers

The chart above shows the outliers present in the data

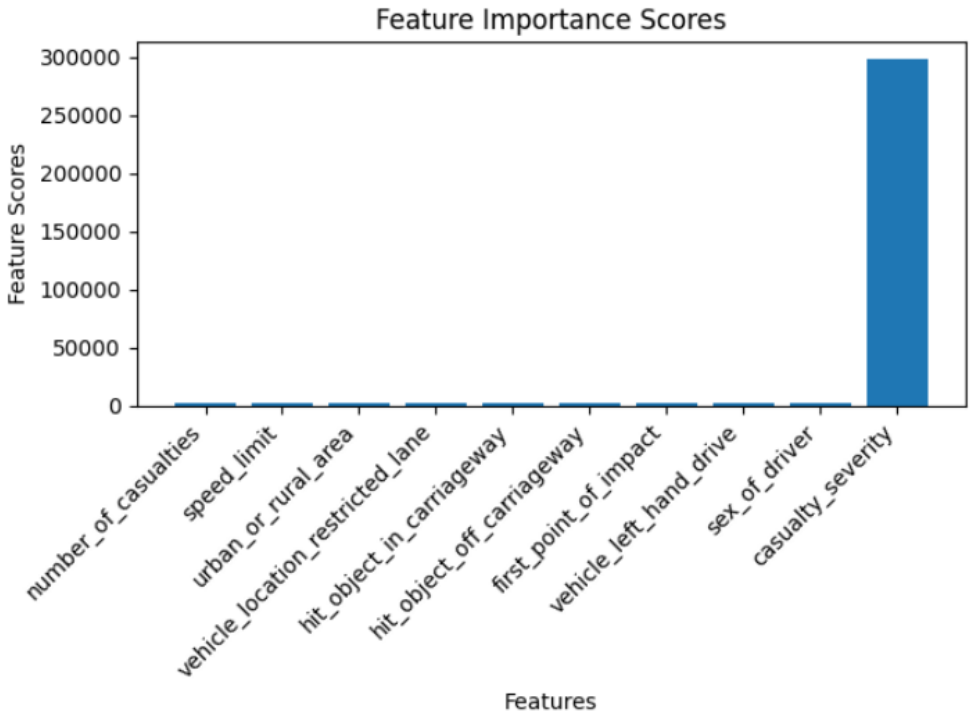

Fig 29.0: Showing Feature Importance Scores

The chart above shows the scores of the different selected features

```
Accuracy: 0.9059866775935616
Classification Report:
              precision    recall  f1-score   support

       False       0.87      0.95      0.91     23686
        True       0.94      0.86      0.90     23903

    accuracy                           0.91     47589
   macro avg       0.91      0.91      0.91     47589
weighted avg       0.91      0.91      0.91     47589
```

Fig 30.0: Showing Classification Report

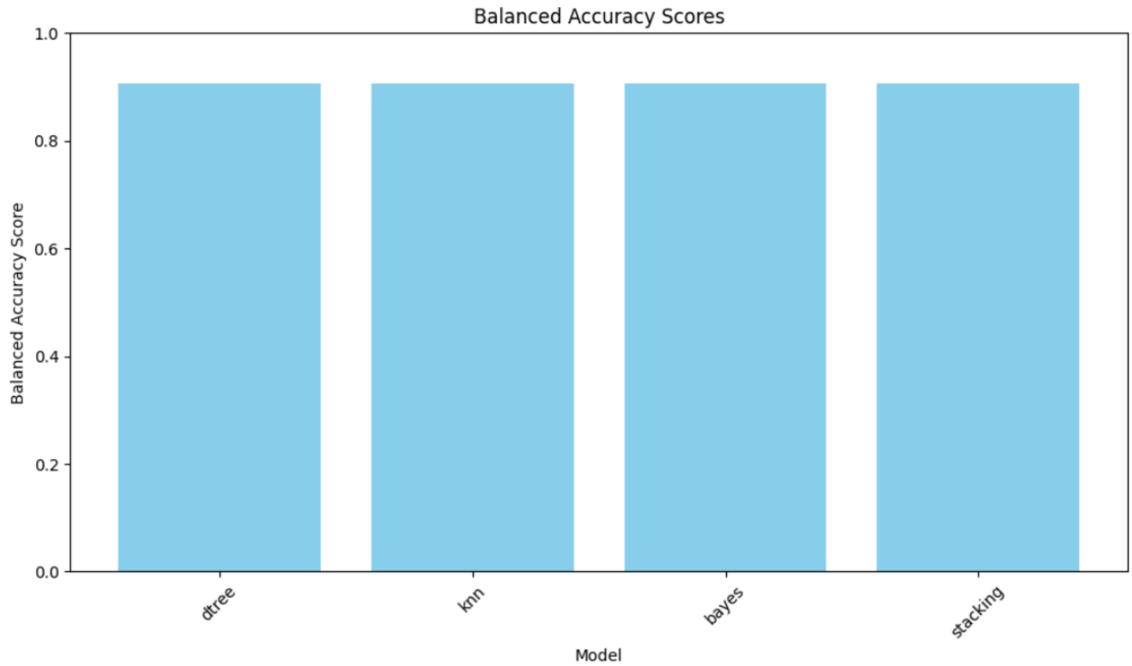

Fig 31.0: Showing Accuracy scorce

Recommendations:

- Strengthen enforcement
- public awareness campaigns
- infrastructure improvements
- targeted measures to protect vulnerable road users.
- Increase speed cameras
- increase public awareness
- invest in safer intersections
- improve lighting and designated pedestrian zones.

By combining data-driven predictions with strategic recommendations, we can foster safer road environments and mitigate accidents risks.

Sources:

- Department for Transport: https://www.gov.uk/government/collections/road-accidents-and-safety-statistics
- Brake: https://www.brake.org.uk/get-involved/take-action/mybrake/knowledge-centre/uk-road-safety
- https://begin-motorcycling.co.uk/125cc-top-speed/
- https://en.wikipedia.org/wiki/Spline_interpolation