Introduction

This study examines road accident data from 2019 to perform data analysis and gain insight into distinct factors that contribute to the occurrence of accidents, predict future accidents to advise the government to prevent more road accidents in Britain.

Cleaning and Preprocessing the Imported Data

Preprocessing entailed renaming the "Accident index" in the data and converting the "Time" column to the correct format after the data had been imported. Police Force was inferred to allow us clean Latitude, Longitude, Location Easting OSGR, and Location Northing OSGR. After combining with Local Authority (District), LSOA was cleaned using the forward fill approach, and S01007575 was used for Scotland. All -1 items were replaced with NaN across all records in all data. In addition, missing values for numeric variables were substituted with the variable set's mean score.

Analysis

QA (i) Are there significant hours of the day, and days of the week, on which accidents occur?

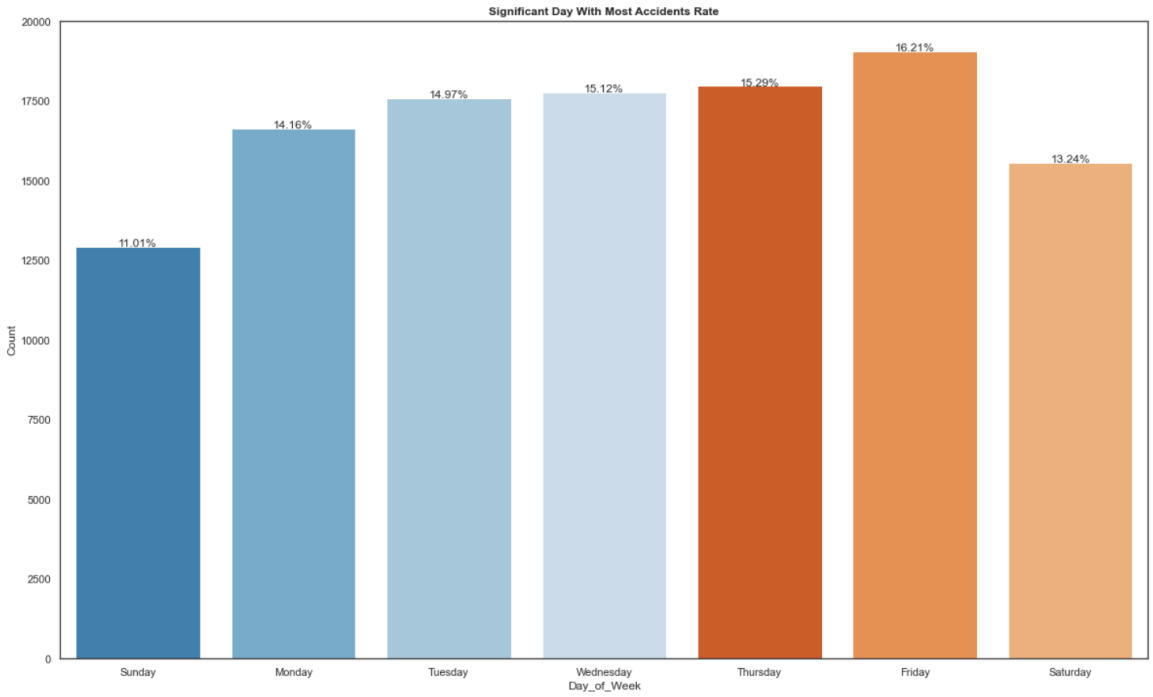

Figure 1: Accident Record on Days of the Week

Given the number of people travelling on Fridays (or racing home from work in preparation for the weekend), the graph above shows that the number of accidents occurs mostly at the same level from Tuesdays to Thursdays. However, an accident occurs mostly on Friday of the week, with Sunday being the safest day of the week on UK roads.

(ii) Are there significant hours of the day, during which accidents occur?

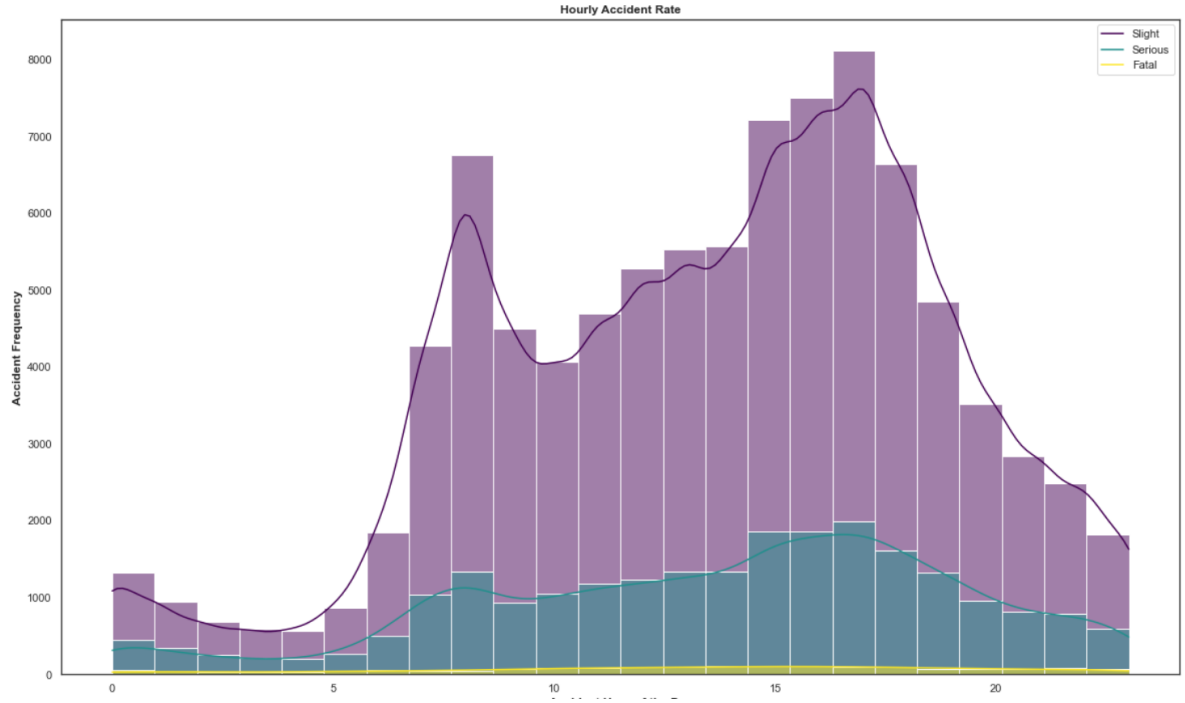

Figure 2: Accident Recorded by Hours

Most of the accidents occurred between the 17th, this was followed closely by 16th, 15th, and 18th hours of the day most common, when people are commuting home from work, 8th hour was also part of the significant hour.

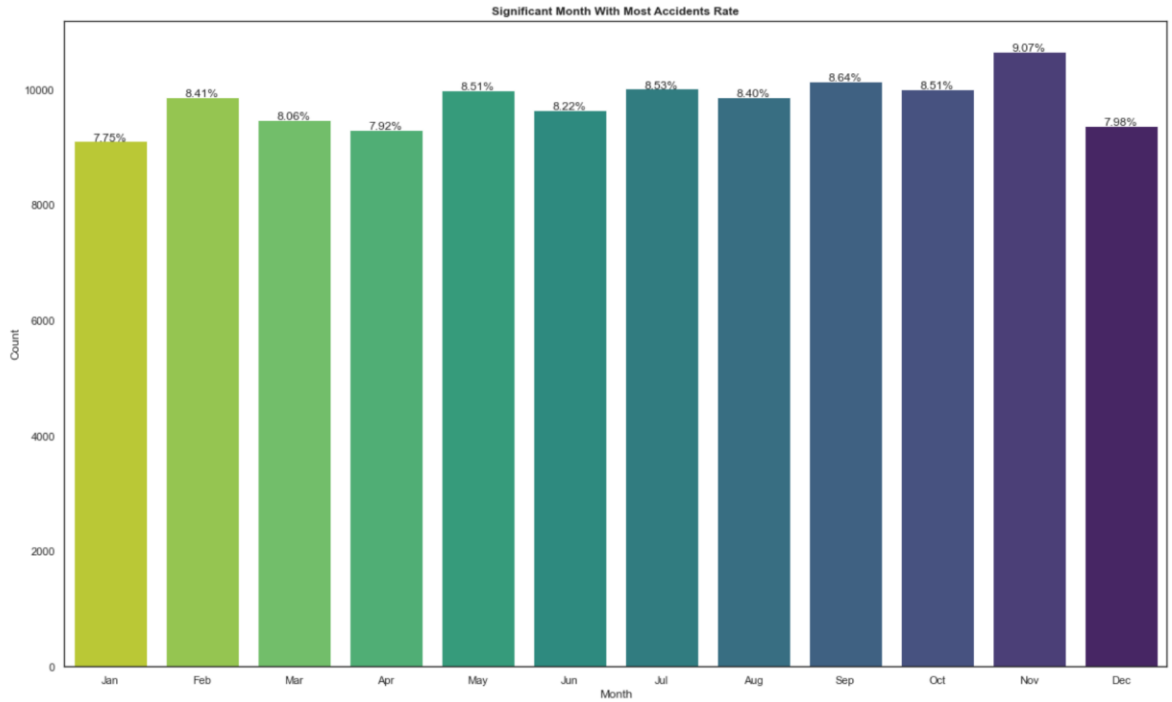

Figure 3: Significant Month with Most Accident

Further analysis reveals that most recorded accidents in the UK occurred in November, with the exact date being November 29th, which is black Friday. Many retailers got a head start on the

traditional black Friday deals, so it is expected that people will rush after work to stores to get better deals before the store closes, with a peak at 15:00 hour of the day.

Question B(i): For motorbikes, are there significant hours of the day on which accidents occur?

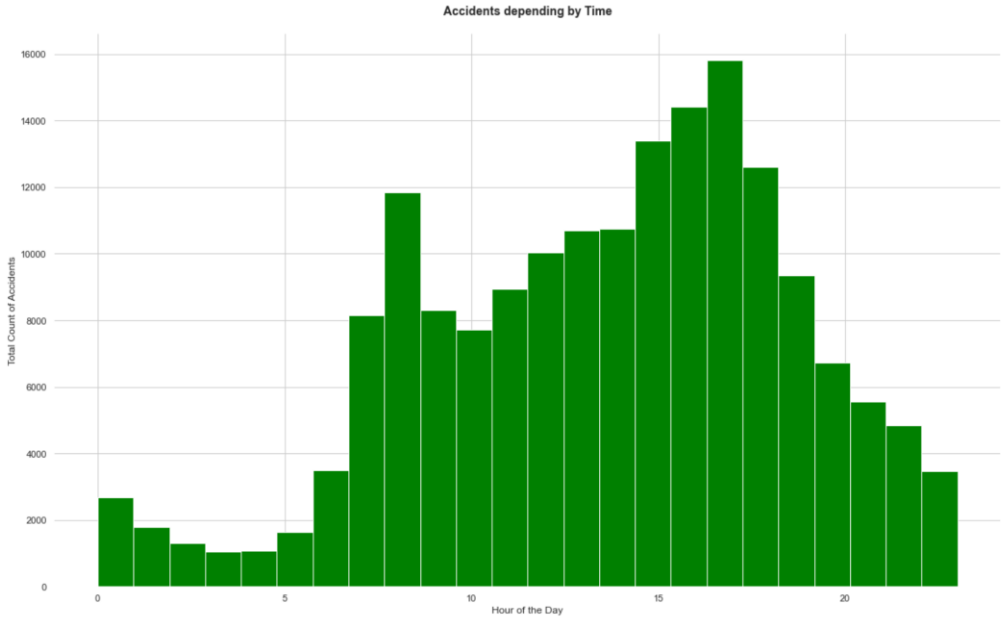

Figure 4: Motorbike Accident Records by Hour

Most accidents happened during daytime, 17th hour has the highest followed by 16th hour which is towards closing time and we have 8:00 which is towards rush hour.

Question B(ii): For motorbikes, are there significant days of the day on which accidents occur?

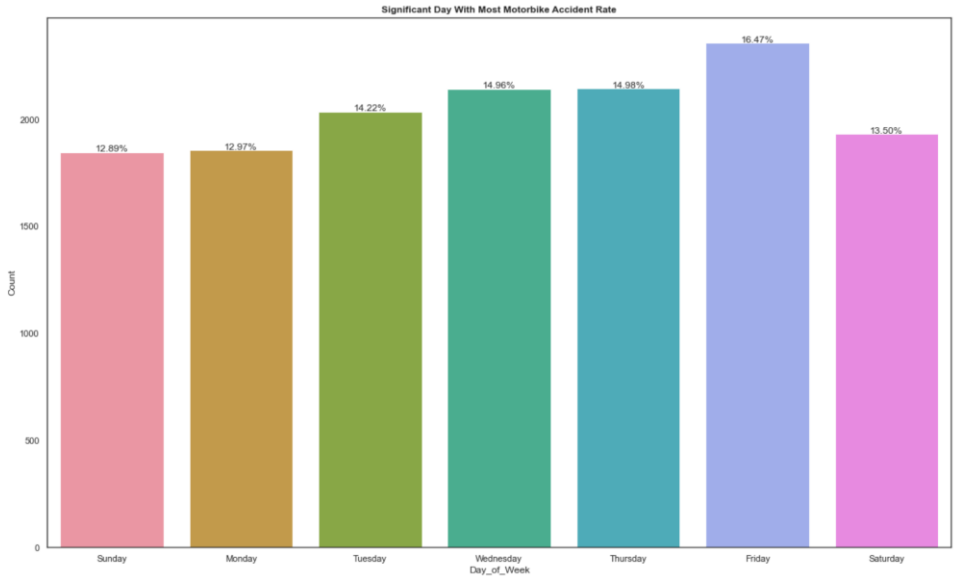

Figure 5: Motorbike Accident Records by Day

From the chart above, we can see that most of the motorbike accidents recorded in the UK happen on Friday, followed by Thursday, Wednesday, and Tuesday of the week.

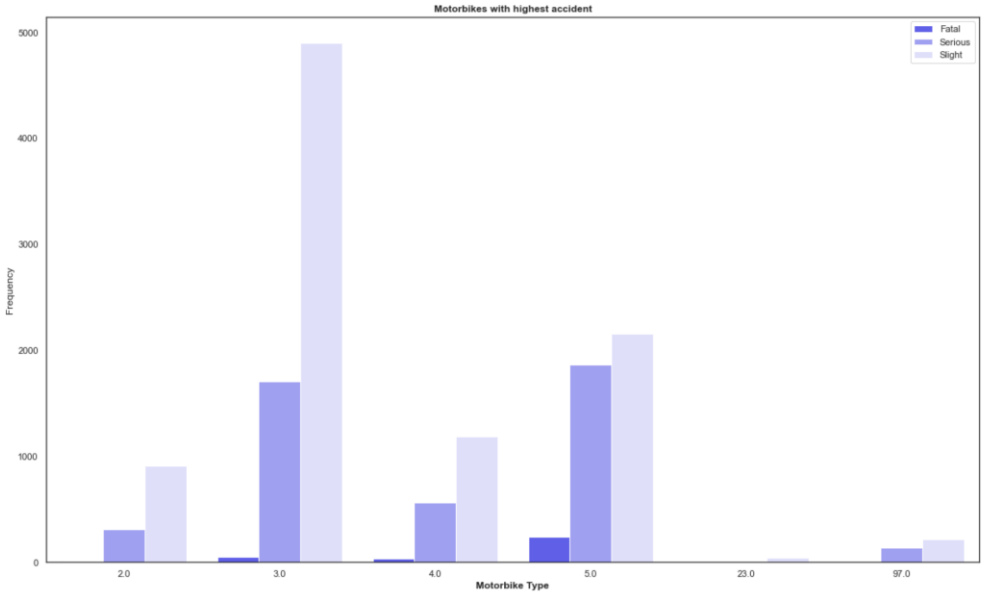

Figure 6: Motorbike Type with Highest Accident

A further analysis shows that Motorcycle 125cc has highest count of slight causality severity, Motorcycle over 500cc has the most serious causality severity which can be concluded because of the capacity.

Question C: For pedestrians involved in accidents, are there significant hours of the day, and days of the week, on which they are more likely to be involved?

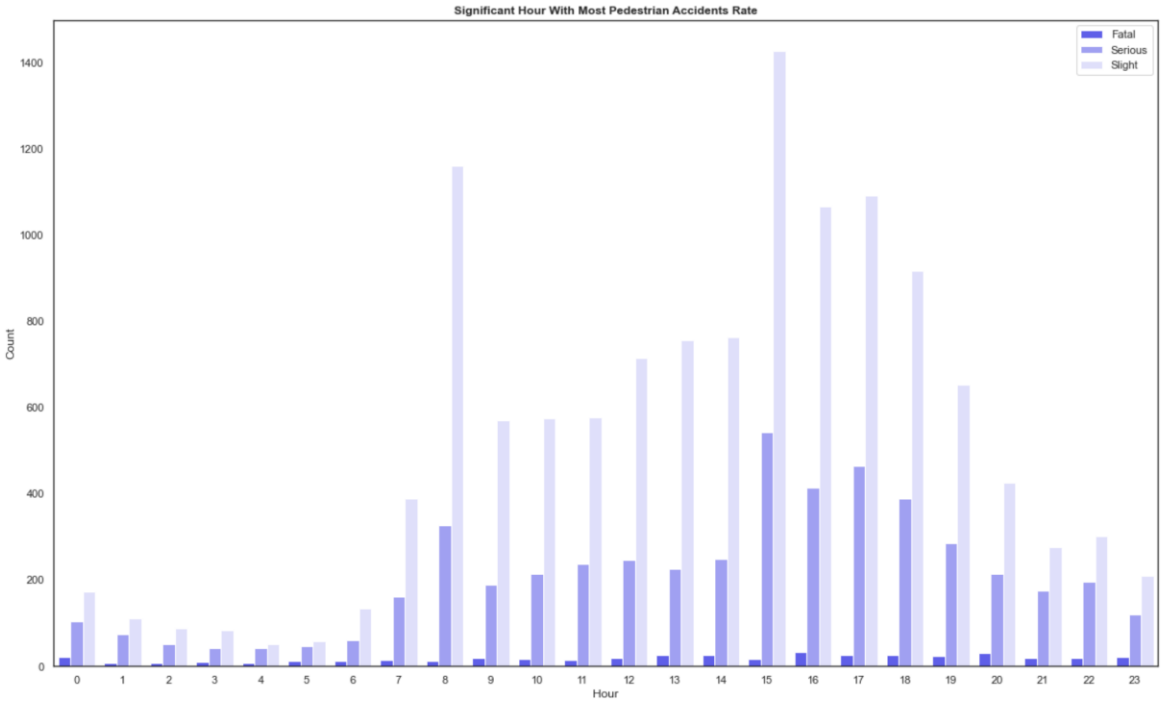

Figure 7: Hours of day for Pedestrian Accidents

From the chart above, we can see that most of the pedestrian accidents recorded in the UK happen on 15th hour of the day, followed by 8th, 16th, 17th, 18th all this falls under rush hour.

Question C(ii): Are there significant days of the week, on which they are more likely to be involved?

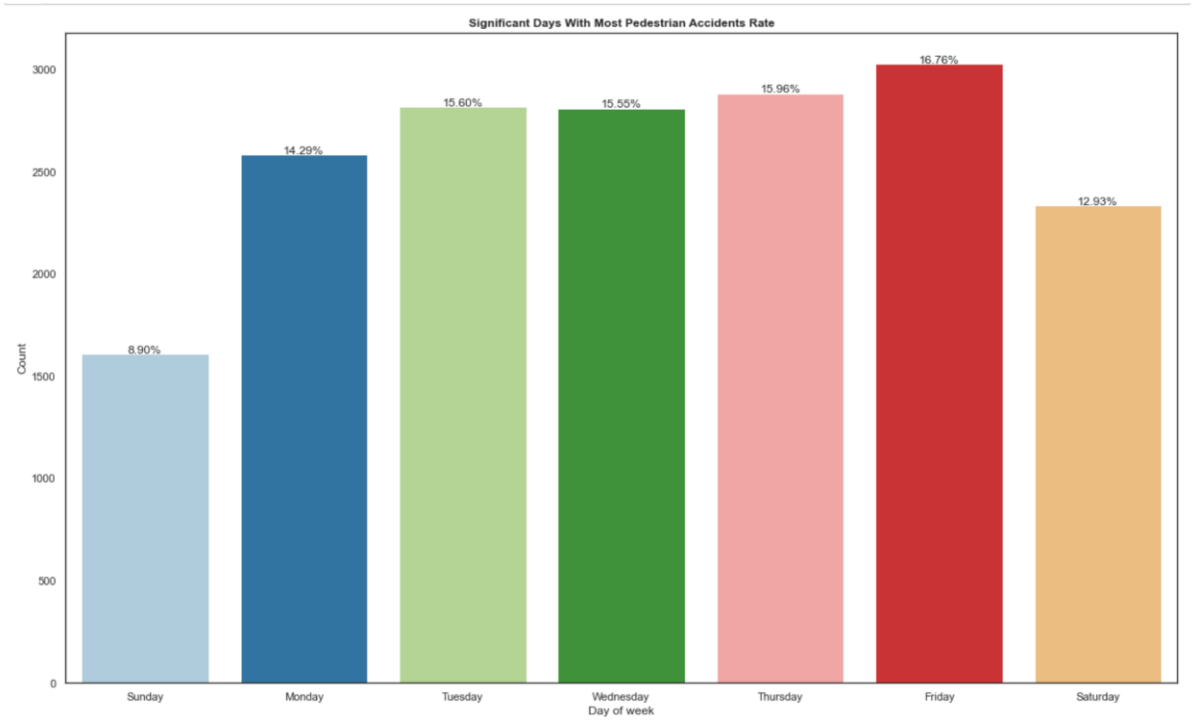

Figure 8: Days of the Week for Pedestrian Accidents

Given the number of people travelling on Fridays (or racing home from work in preparation for the weekend), the graph above shows that the number of accidents occurs mostly at the same level from Tuesdays to Thursdays. However, an accident occurs mostly on Friday of the week, with Sunday being the safest day of the week on UK roads.

Question D: What impact, if any, does daylight savings have on road traffic accidents in the week after it starts and stops?

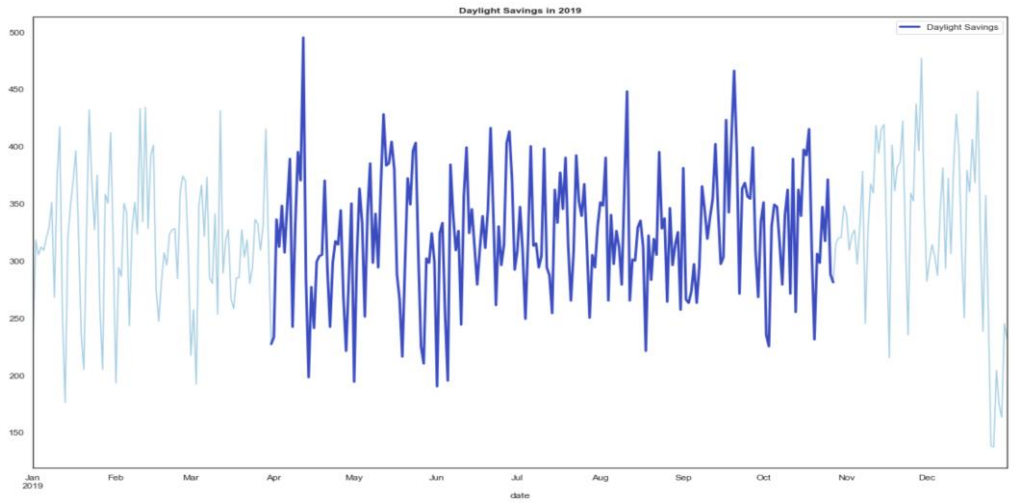

Figure 9: Impact of Daylight-Saving Time

According to *timeanddate.com* Daylight saving time 2019 in the United Kingdom began at 1:00AM on Sunday, March 31 and ended at 2:00AM on Sunday, October 27, 2019. The chart with the **dark blue** represents the period when daylight savings was observed in the UK in 2019 while the lines in **light blue** represent all timeframes from January to December. It is seen that there was one major significant spike in the number of accidents recorded per day at the early stage of the daylight-saving period. Also, there were two other spikes recorded that were beyond the usual accident frequency levels of other periods before the last quarter. With these observations, we can consider daylight saving period to have some effect on the number of accidents most especially a week after it began.

Question E: What impact, if any, does sunrise and sunset times have on road traffic accidents?

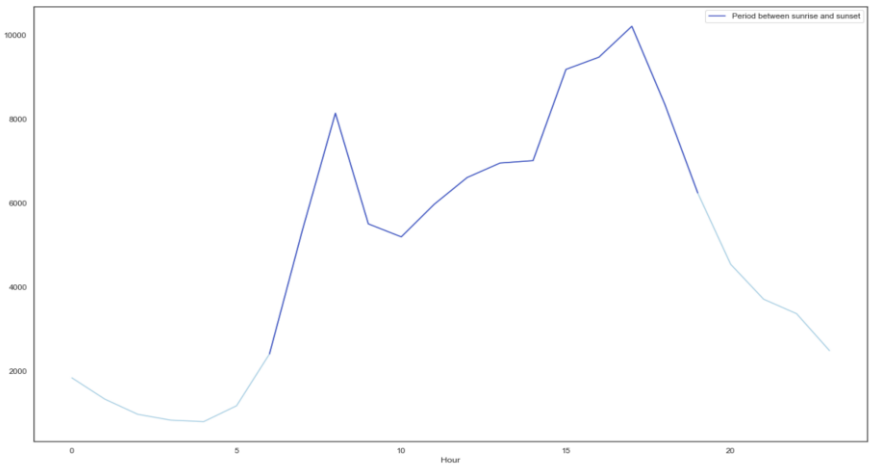

Figure 10: Sunrise and Sunset Accident Occurrence

From the graph above, sunrise time was pegged to 6th hour while sunset time to the 19th hour. During this period, we can see a steady rise in the number of road accidents as the day progressed. The highest number of accidents recorded was also between sunrise and sunset. This clearly shows that sunrise and sunset had a high significance on road accidents in the UK in 2019.

Question F: Are there particular types of vehicles (engine capacity, age of vehicle, etc.) that are more frequently involved in road traffic accidents?

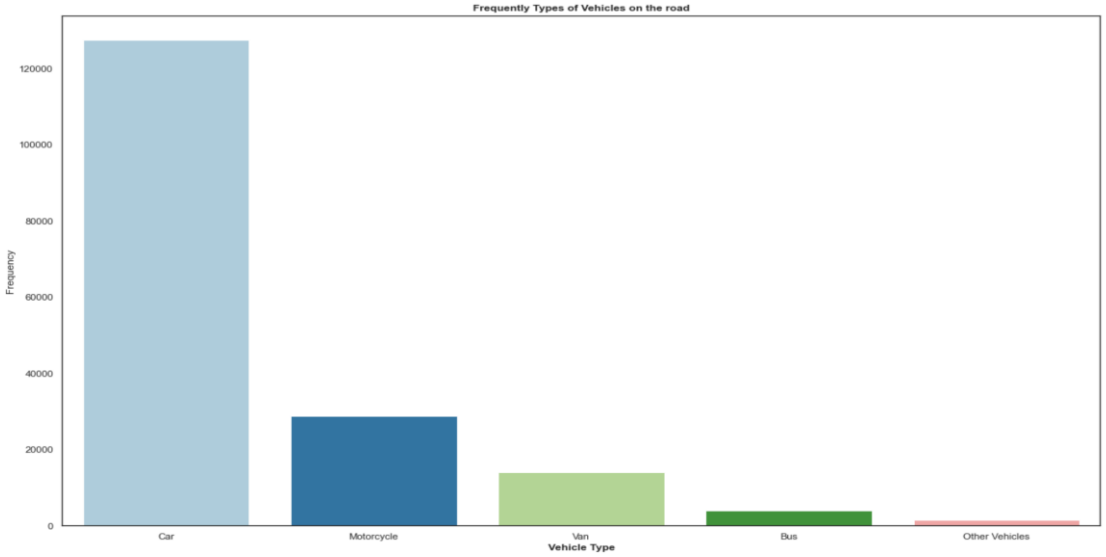

Figure 11: Vehicle Type

From the chart above, vehicle type (cars) is more frequently involved in road traffic accidents with 127,538 occurrences.

Question F(ii): Are there particular types of vehicle engine capacity that are more frequently involved in road traffic accidents?

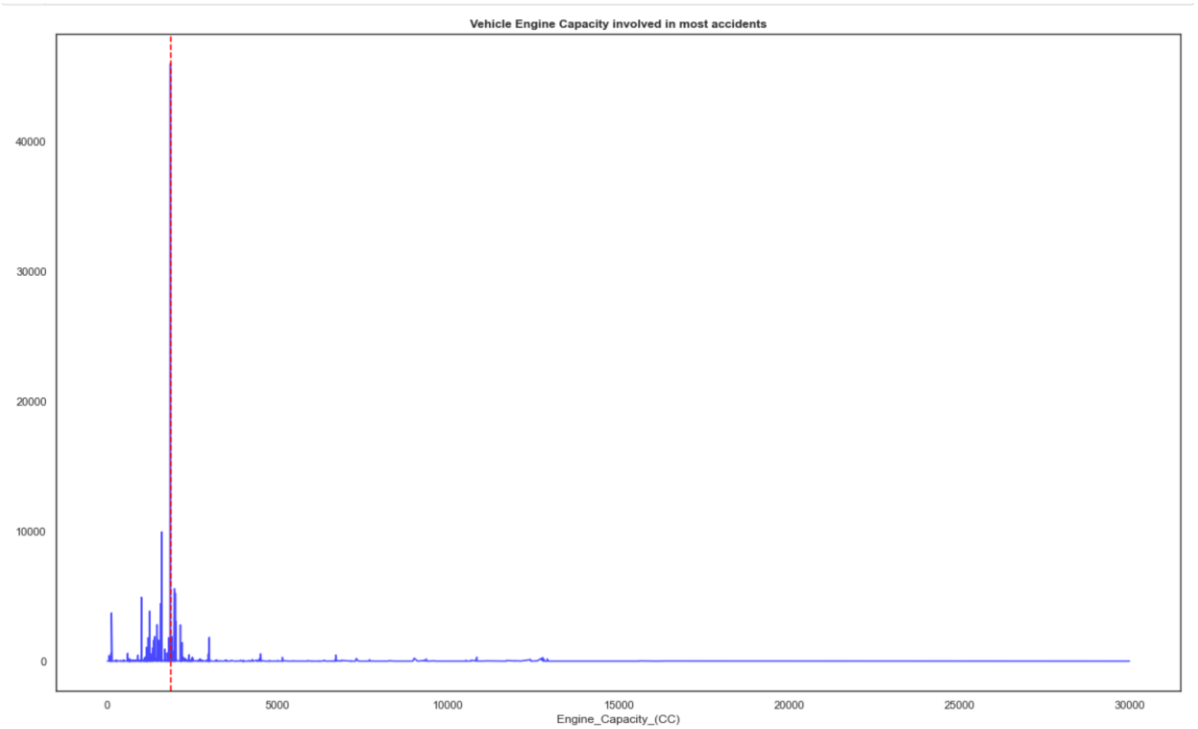

Figure 12: Engine Capacity

From the line chart above, Vehicles with Engine Capacity of 1849 which are more frequently involved in road accidents have 46,012 occurrences.

Question F(iii): Vehicle age that are more frequently involved in road traffic accidents?

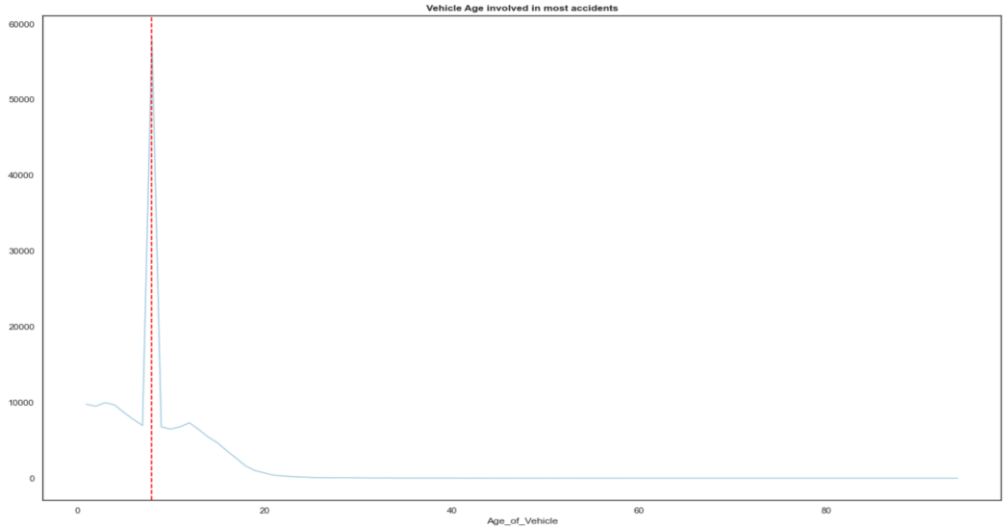

Figure 13: Vehicle Age

From the line chart above, Vehicles aged 8years are more frequently involved in with 58,202 occurrences.

Question G: Are there particular conditions (weather, geographic location, situations) that generate more road traffic accidents?

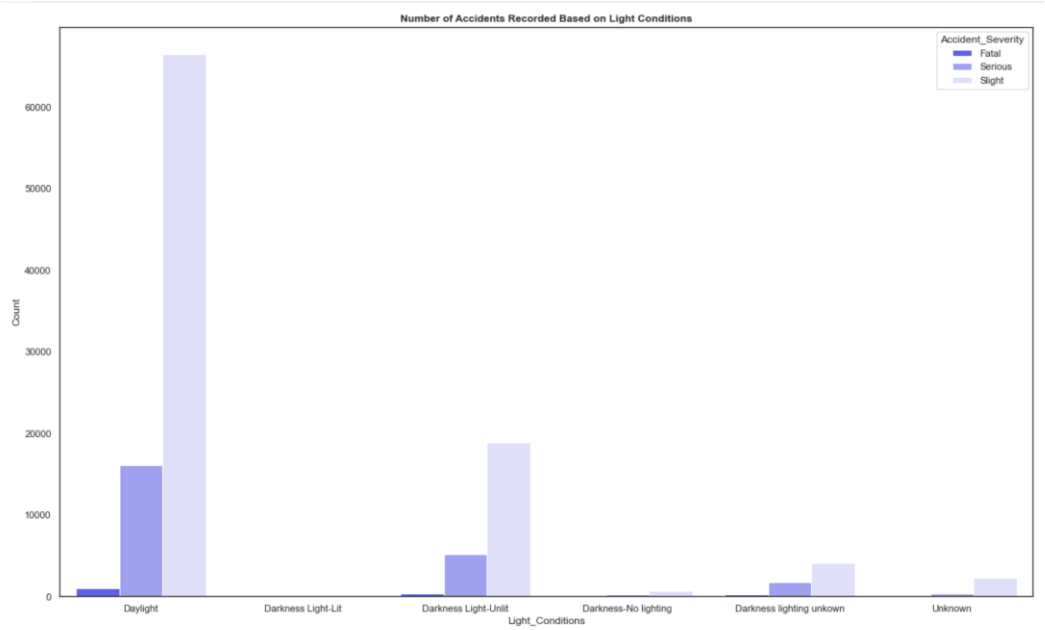

Figure 14: Light Condition and Accident Severity

From the chart above, we can see that the UK records more accidents during daylight under light conditions people aren't conscious enough as other weathers although the accident severity is slight.

Question G(ii): Weather Conditions

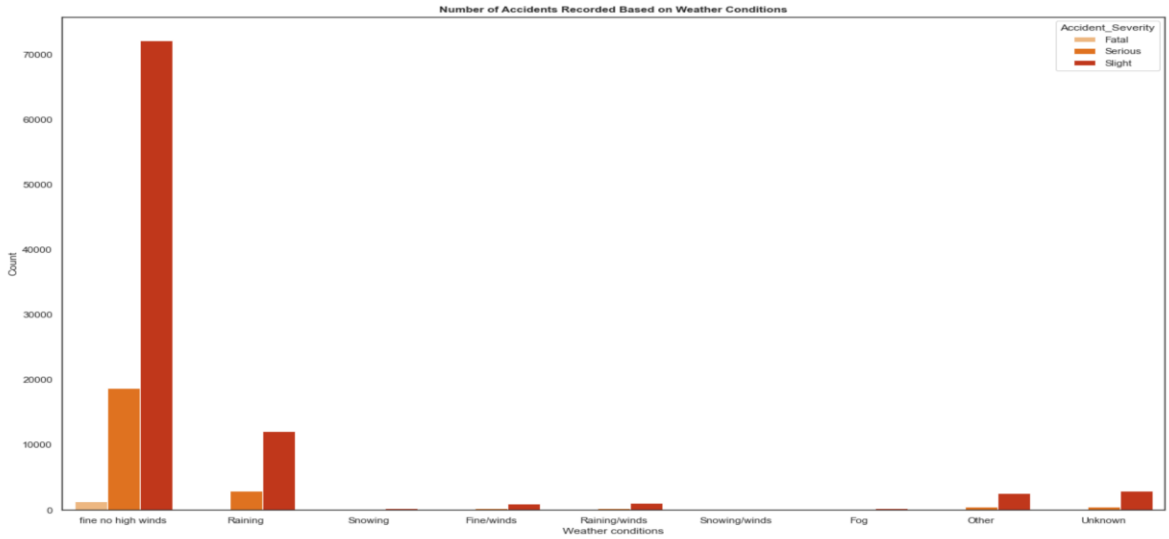

Figure 15: Weather Condition and Accident Severity

From the chart above, we can see that the UK records more accidents during fine no high winds under weather conditions people aren't conscious enough as other weathers although the accident severity is slight.

Question G(ii): Road Surface Conditions

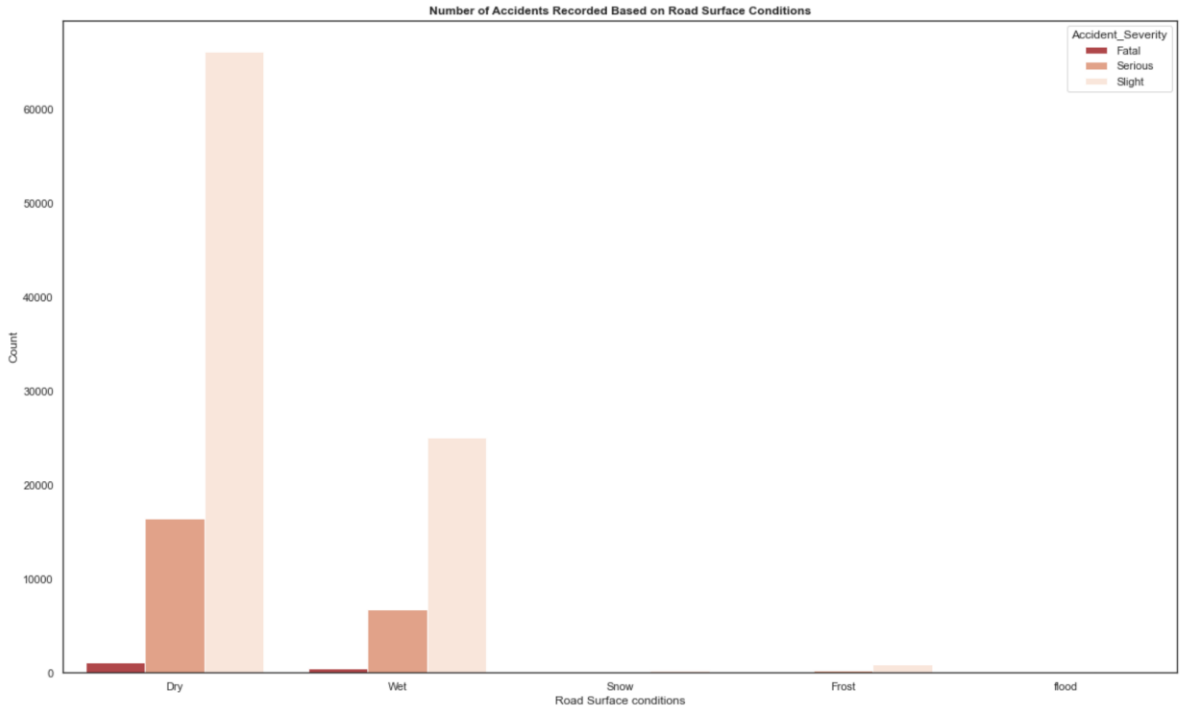

Figure 16: Road Surface Condition and Accident Severity

From the chart above, we can see that the UK records more accidents when road surface is dry this could be because there are more cars on the road although the accident severity is slight.

Question G(iii): Special Conditions at Site

Figure 17: Special Conditions at Site and Accident Severity

From the chart above, we can see that the UK records more accidents when there are no special conditions at site, so special conditions does not have any influence on accident rate.

Question G(iv): Speed Limit

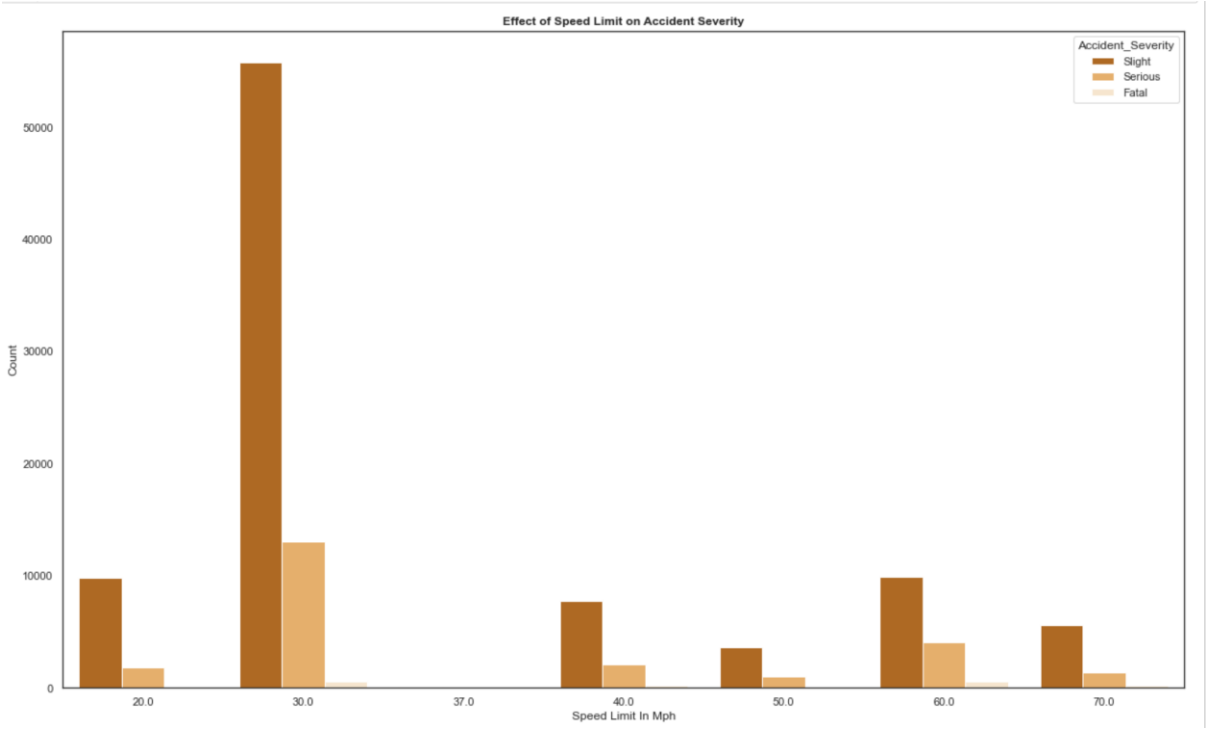

Figure 18: Speed Limit and Accident Severity

Most collisions occur at 30 mph, which is the most prevalent speed limit on UK roadways. The national speed limit of 60 mph also sees a lot of accidents. Surprisingly, more deadly accidents occur at 30 mph than at 60 mph, demonstrating that speed isn't the primary determinant in major road accidents.

Question G(v): Vehicle Maneuver

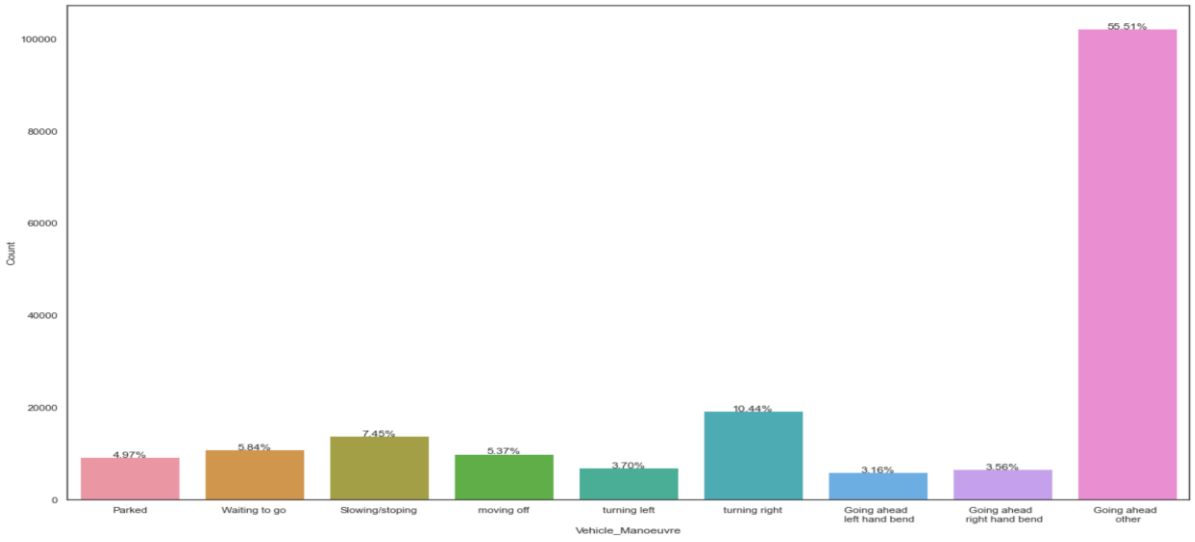

Figure 19: Vehicle Maneuver and Accident Frequency

According to the chart above, going ahead others have a greater influence on the UK accident rate than others, due to our inability to maneuver our vehicle in dangerous or emergency conditions while attempting to get ahead of others.

Question G(vi): Junction Control

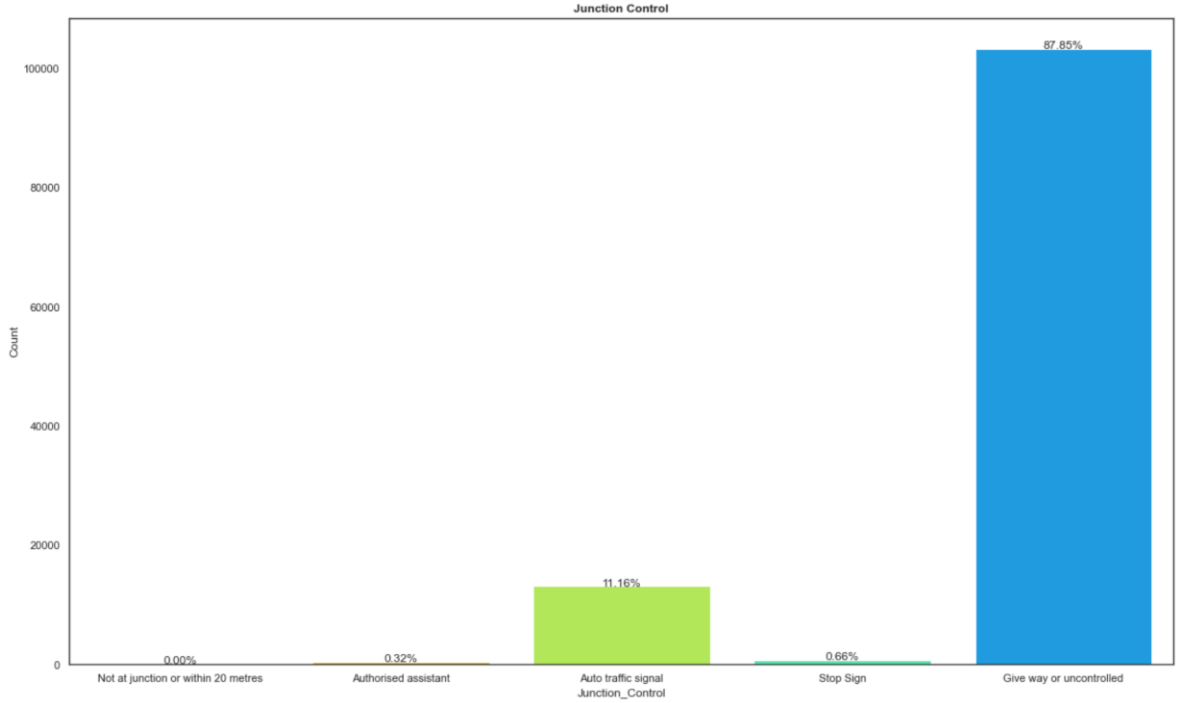

Figure 20: Junction Control and Accident Frequency

Many give way junctions only have the white lines painted on the road. The warning sign and the triangle may or may not be present, sometimes you must look very closely to see which markings are present at a junction a lot of times drivers do not see this which can lead to more accidents.

Question G(vi): Area Type

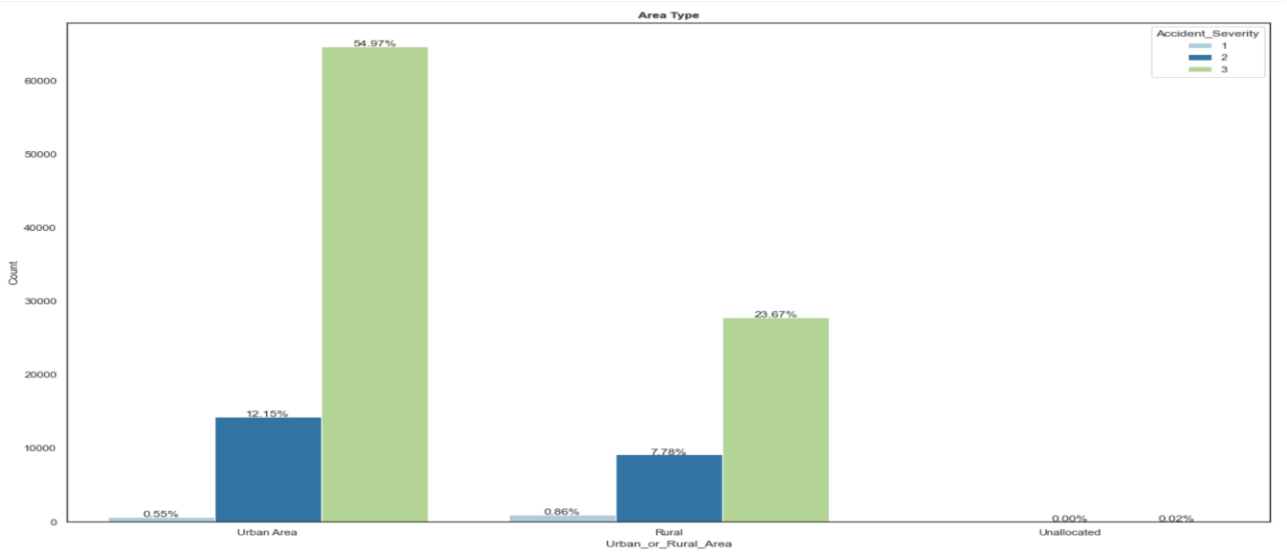

Figure 21: Area Type and Accident Frequency

In metropolitan areas, the number of serious and slight accident severity is higher, with 79,533 on urban roads and 37,975 on rural roads. These results reveal that while the number of incidents is larger in metropolitan areas, rural roads have a higher risk of fatalities, accounting for 7.78 percent of serious accidents.

Question H: How does driver related variables affect the outcome (e.g., age of the driver, and the purpose of the journey)?

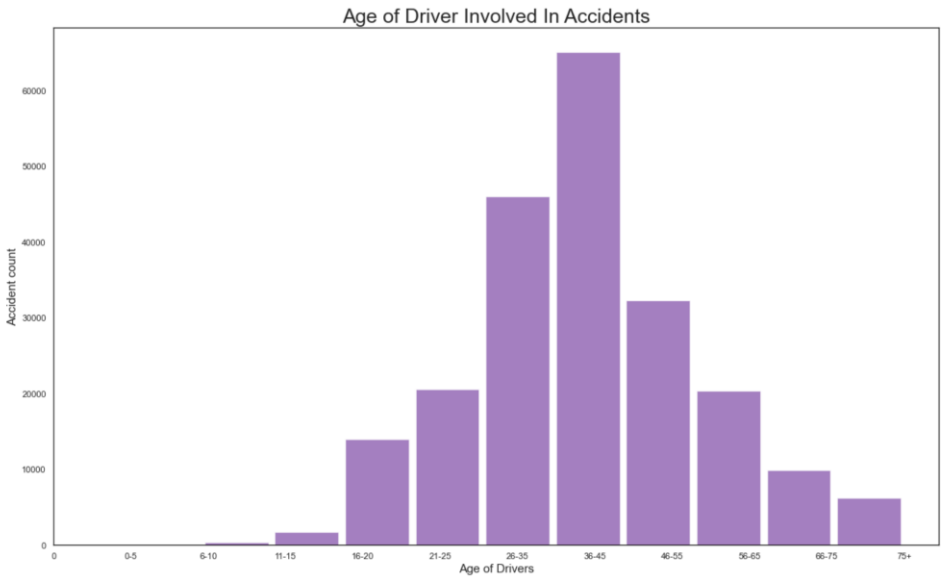

Figure 22: Age of Drivers and Accident Frequency

From the chart above, drivers who fall within the age band of (36-45years) have the most frequency of road accidents in the UK, You can start driving a car when you're 17 according https://www.gov.uk/driving-lessons-learning-to-drive however we have age band 6-10 and 11-15 we can assume they were probably driving other type of vehicle e.g. Pedal Cycle

Question H(ii): How does driver related variables affect the outcome (the purpose of the journey)?

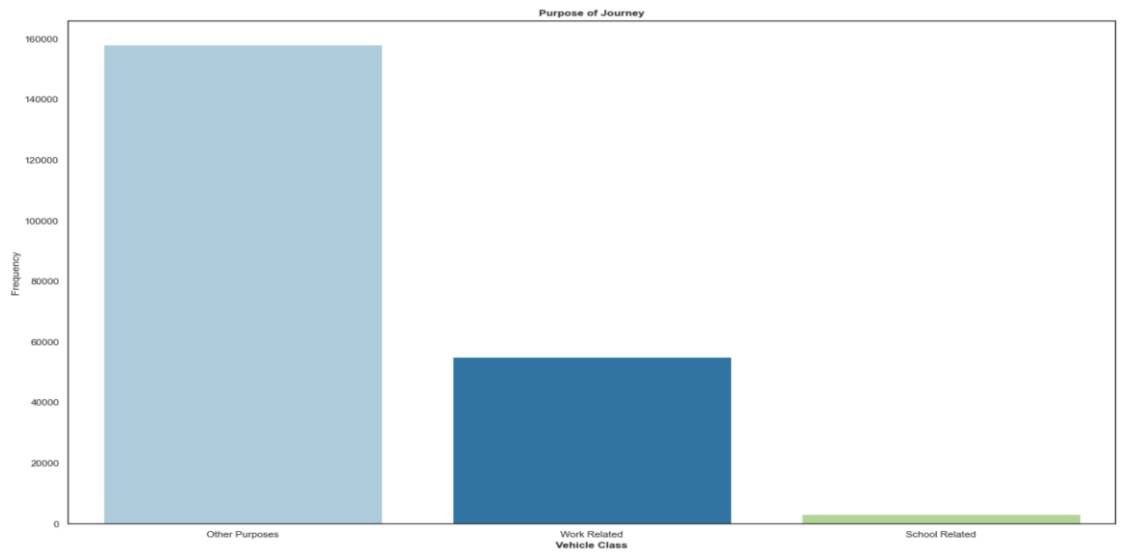

Figure 23: Purpose of the journey and Accident Frequency

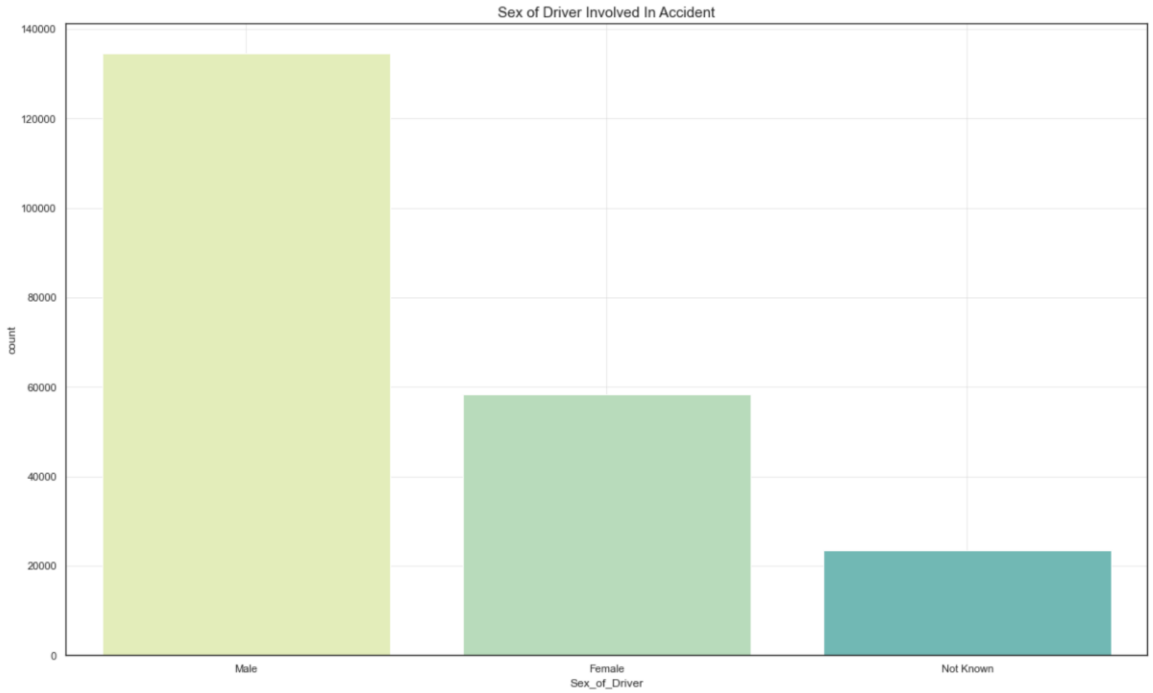

Figure 24: Sex of Driver and Accident Frequency

From the chart above, there are more Male involved driver involved in accident compare to Female in UK.

Question I(i): Can we make predictions about when and where accidents will occur, and the severity of the injuries sustained from the data supplied to improve road safety? How well do our models compare to government models??

Four different predictions were carried out: When (The Hour) Accidents will happen, Location of where accidents will occur (Latitude and Longitude), Accident Severity and a prediction to evaluate the performance of our model with that of the government.

a) Accident Hour Prediction

For when accidents will occur, the evaluation metric used is mean absolute error and the value obtained was 3.576

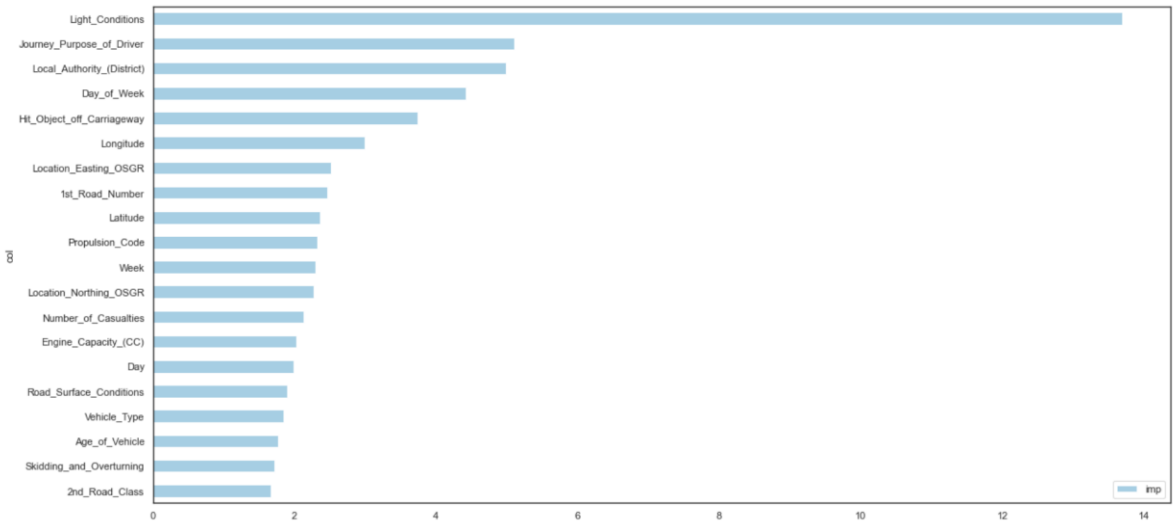

Figure 25: Important Features that impact when accidents will happen

The **light condition** is the most important feature that predicts the hour an accident will occur.

b) Location Prediction

The table below represents sample point predictions of where accidents will happen in the UK after training our model with a 70% training set. MAE score for Latitude prediction is 0.058 while the MAE score for Longitude prediction is 0.092

Table 1: Latitude and Longitude

	Latitude	Longitude
0	51.045437	-2.958735
1	55.964633	-3.788900
2	51.251641	-0.305450
3	52.395060	-0.812230
4	53.754715	-0.420467
...	...	...
32851	52.635517	1.323875
32852	56.472226	-3.534749
32853	53.778274	-1.532183
32854	53.488707	-2.220567
32855	51.641933	-0.075883

The Local_Authority_(District) is the most important feature for predicting where an accident will occur.

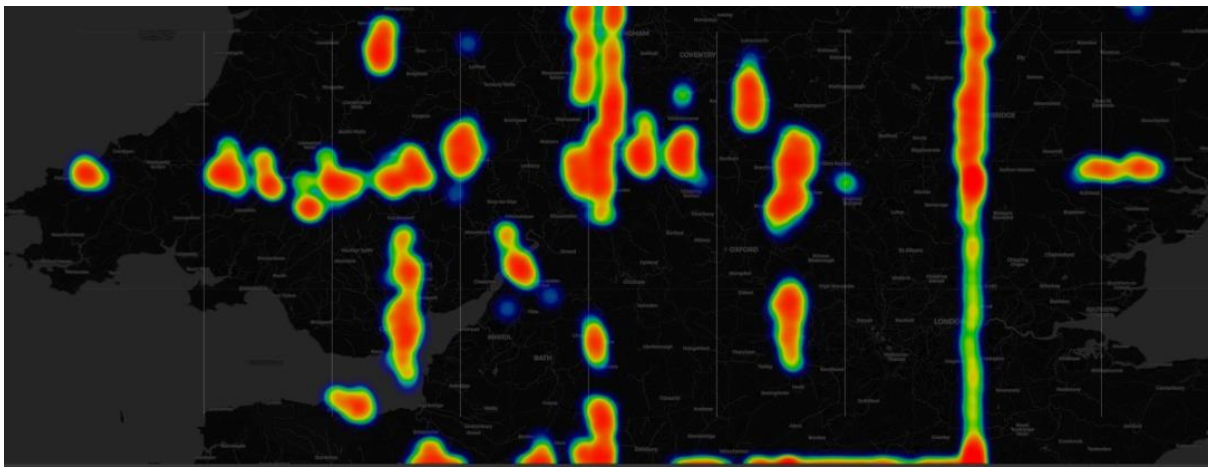

Figure 27: Heat map showing some locations where accidents will happen in the UK

Figure 28: Open map showing some locations where accidents will happen in the UK

c) Predicting Accident Severity

```
1 naive = model(X_train, y_train, X_test, y_test, GaussianNB())

Training Accuracy is: 93.35
Test Accuracy is: 93.4
              precision    recall  f1-score   support

           2       0.88      0.98      0.93     15227
           3       0.98      0.89      0.94     18342

    accuracy                           0.93     33569
   macro avg       0.93      0.94      0.93     33569
weighted avg       0.94      0.93      0.93     33569
```

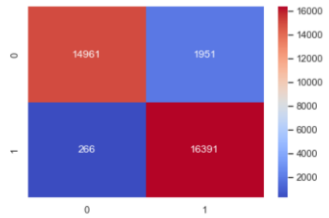

Figure 29: Naïve Bayes Model Output

```
1 log = model(X_train, y_train, X_test, y_test, LogisticRegression())

Training Accuracy is: 81.4
Test Accuracy is: 81.57
              precision    recall  f1-score   support

           2       0.80      0.83      0.81     16440
           3       0.83      0.81      0.82     17129

    accuracy                           0.82     33569
   macro avg       0.82      0.82      0.82     33569
weighted avg       0.82      0.82      0.82     33569
```

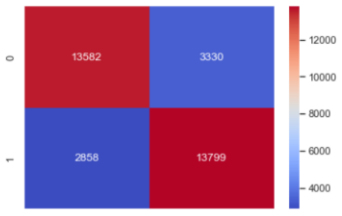

Figure 30: Logistics Regression Model Output

The accident severity using The GaussianNB model has an accuracy score of 93.35 percent and a recall score of 98 percent, implying that the model delivered relevant results, whereas the Logistics regression model has an accuracy score of 81.4 percent and a recall score of 83 percent, which is similarly good.

This result was gotten after balancing our data with SMOTE.

d) Comparing the performance of the accident severity model with that of the government.
Comparing government model with our model shows that it's better as it gives 94.7% which is higher than our two modes.

Recommendations

- Regular remodeling and comparison of government models with models made by private establishments should be made so the government can keep in checks factors that influence road accidents
- Adequate preventive measures should be made by the government to regulate drivers' decisions, especially on the high-risk factors that affect road accidents like getting to understand and mitigate factors that cause an increase in road accidents in the 17th hour
- To keep our roads safe, it's critical to maintain their surfaces, which involves patching potholes, crack sealing, and applying new surface treatments or sealants. Regular maintenance improves the road's durability and minimizes the likelihood of an accident.
- Transporters and engineers to identify dark spots areas that require attention and also rank dangerous locations and timings for improvement.

Reference

Time and Date
https://www.timeanddate.com/time/change/uk?year=2019[27/04/2022]

https://www.sunrise-and-sunset.com/en/sun/united-kingdom/bristol [27/04/2022]

https://www.gov.uk/driving-lessons-learning-to-drive [27/04/2022]

https://www.kaggle.com/code/joseconomy/data-analysis-accidents-in-uk [27/04/2022]

STATISTICA 2020. Factors leading to road accidents in Great Britain. Available at: https://www.statista.com/statistics/323079/contributing-factors-leading-to-road-accidents-in-great-britain-uk/ [27/04/2022]