Analysis of Road Traffic Accidents in UK

**Contents**

[**Introduction** 2](#_Toc172406643)

[**Analysis** 2](#_Toc172406644)

[**Task 1: Accident Analysis of Relevant Hours and Days** 2](#_Toc172406645)

[**Task 2: Motorbike Accidents by Hour and Day of the Week** 3](#_Toc172406646)

[**Task 3: Pedestrian Accidents by Hour and Day of the Week** 7](#_Toc172406647)

[**Task 4: Using the Apriori Algorithm to Explore Accident Severity** 8](#_Toc172406648)

[**Task 5: Analyzing Accident Distribution Using Clustering** 11](#_Toc172406649)

[**Task 6: Identify Unusual Entries Using Outlier Detection Method** 12](#_Toc172406650)

[**Predictions** 14](#_Toc172406651)

[**Task 7: Developing Classification Model** 14](#_Toc172406652)

[**Recommendations to Government Agencies for Improving Road Safety** 16](#_Toc172406653)

[**Conclusion** 17](#_Toc172406654)

[**References** 18](#_Toc172406655)

[**Appendix** 19](#_Toc172406656)

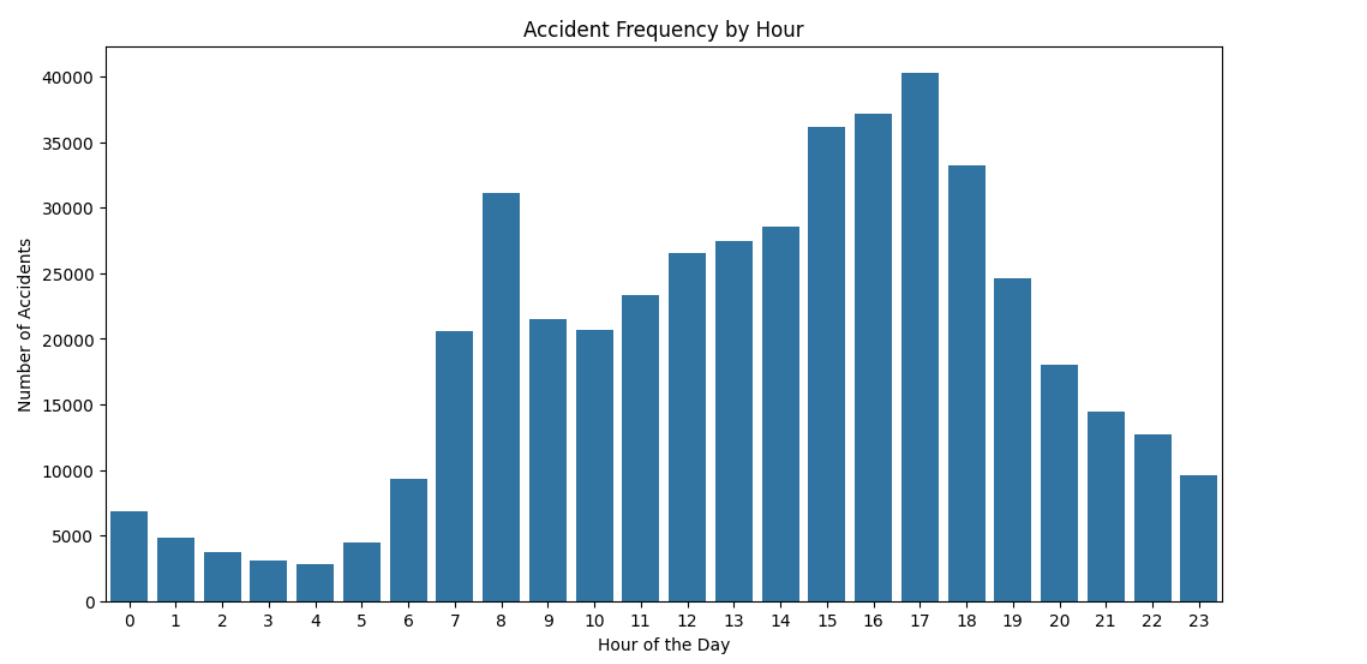
# **Introduction**

The dataset comprises comprehensive facts of street visitor's injuries in Great Britain for the year 2020, saved in a SQLite database with 4 tables: twist of destiny, casualty, car, and also. These statistics consist of particular facts approximately each incident, the motors worried, and the ensuing casualties. We aim to analyze these statistics to advice authorities companies on improving road protection and to extend predictive models for injuries and the associated accidents. By figuring out key patterns and threat elements, we goal to make contributions to centred interventions that may reduce the frequency and severity of avenue web page visitor accidents.

# **Analysis**

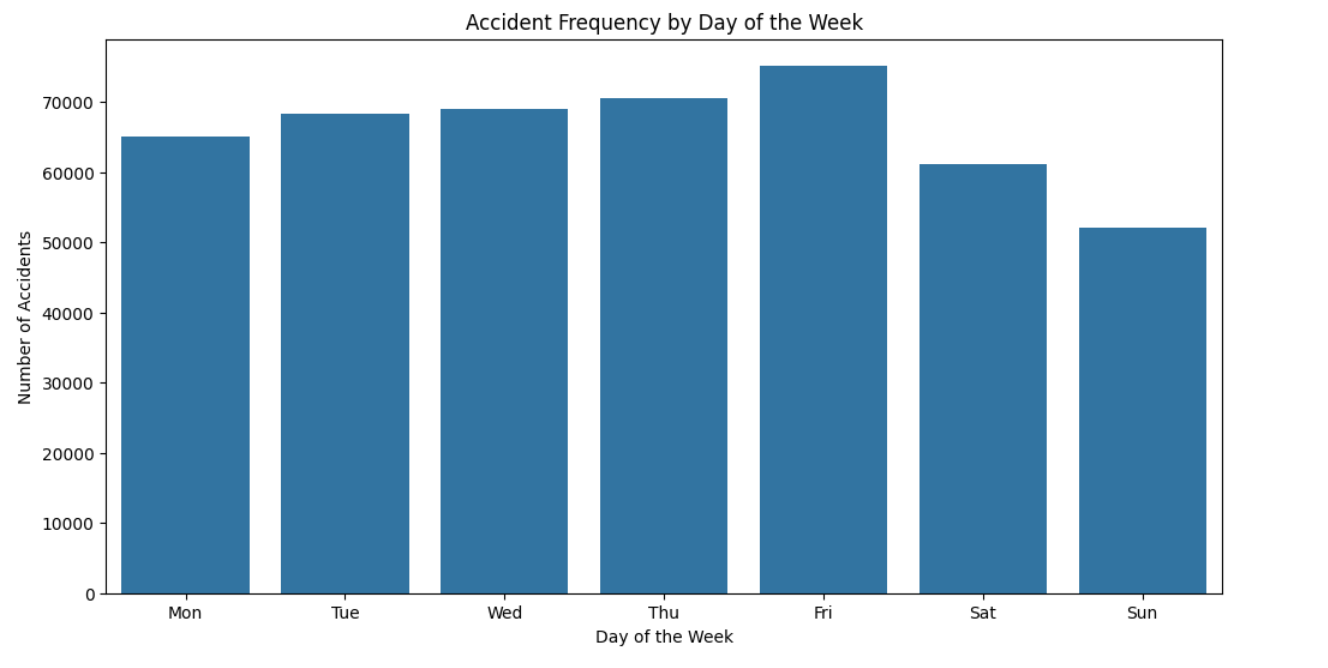
## **Task 1: Accident Analysis of Relevant Hours and Days**

The 2020 British road traffic injury survey indicates substantial variations in accident frequency by the moment, day and week. We can enforce targeted road safety measures by analyzing twist of fate data to determine specific hours and days when accidents are most common.



*Fig1: Accident Frequency by Hour*

Early or late in the day, injuries vary greatly. Statistics suggest that injuries are more common in the morning and afternoon, as well as during rush hours and high website traffic. Fig1 shows that accidents surge between 8 AM and 5 PM, when people are commuting to work or school. This sample demonstrates that peak-hour vigilance and visitor control can significantly reduce injuries.



*Fig2: Accident Frequency by day of Week*

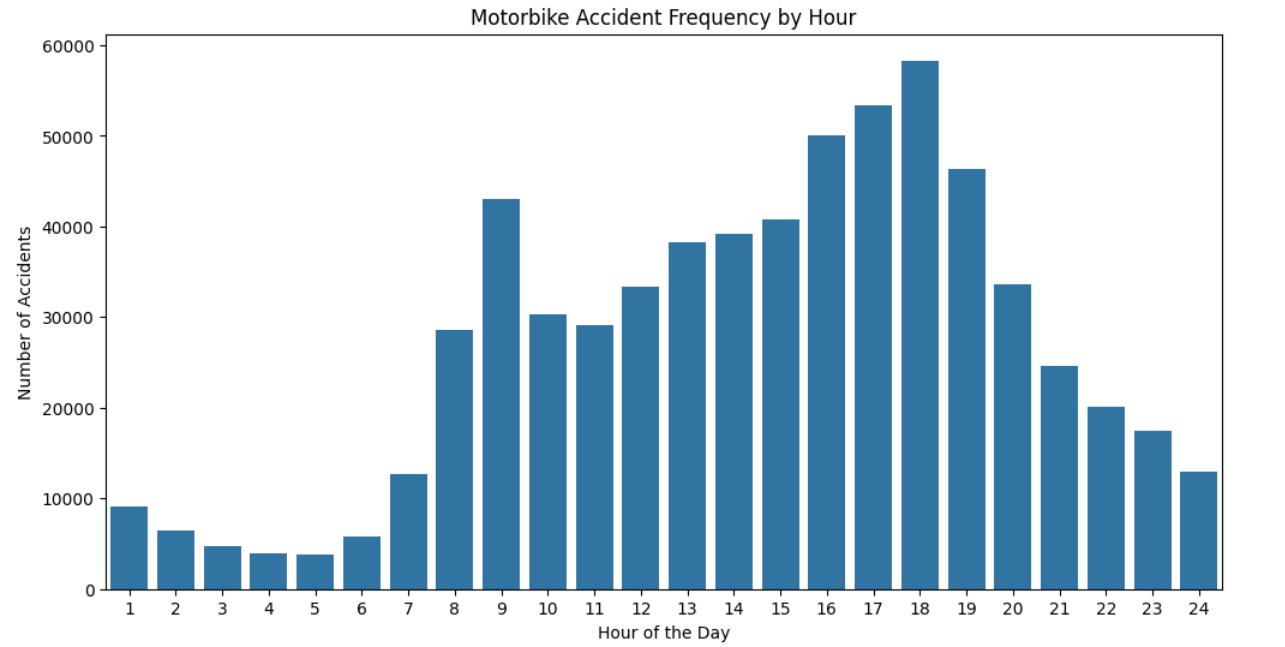
Also important in turn of fate frequency is the particular day of the week. Fig2 shows that Monday through Friday have more injuries than weekends, according to the evaluation. Weekday commutes and sporting activities increase the number of cars on the road, which may explain this trend. Weekend accidents may be lower due to fewer visitors and more leisurely tours.

## **Task 2: Motorbike Accidents by Hour and Day of the Week**

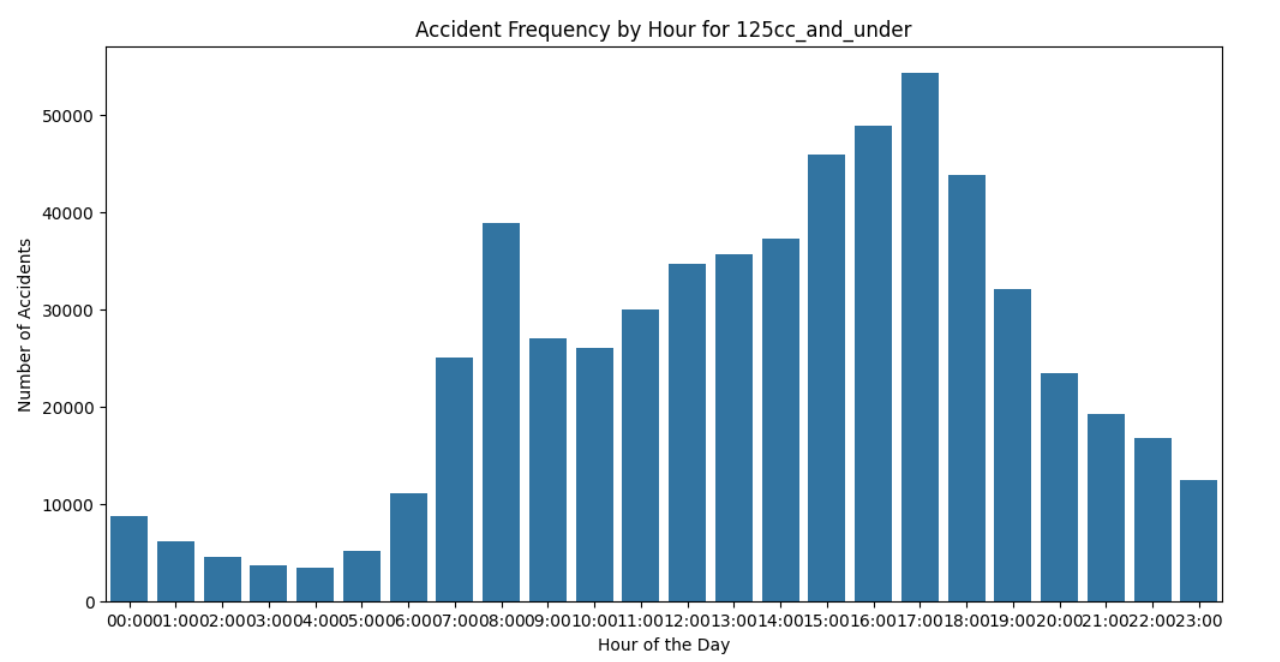
Fig 3 shows that based on engine period, motorbike injuries have beautiful patterns that vary by time of day and week. Three training classes Motorcycle 125cc and less, Motorbike over 125cc and up to 500cc, and Motorcycle above 500cc show different twist of fate rates.

In Fig4 injuries for motorcycles under 125cc will increase from noon and 2 PM. Short outings and housework may be more likely during this time. However, Fig6 shows that motorcycles over 125 cc & up to 500 use higher twist of fate rates in the afternoon and evening from 4 PM to 6 PM, likely due to peak commuting time. Fig8 shows that injury rates for motorcycles over 500cc are more distributed across the day and rise in the afternoon.

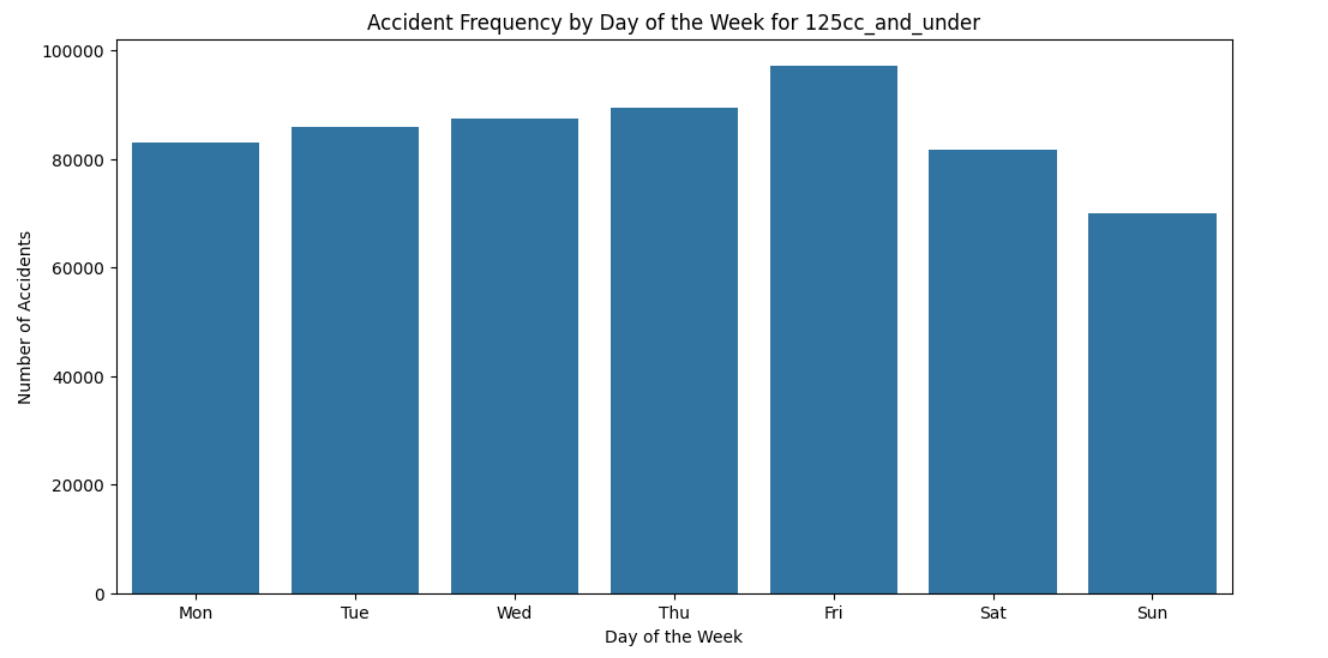
Fig5, 7, 9 shows that on weekdays, especially Monday through Friday, bike accidents and other avenue injuries are more common. Thus, this tendency suggests that motorcycle consumers' various and routine operations are similar to those of other vehicles. Monday-Friday threat occurrences show that focused street security advertising and activities, such as motorcycle repute kits and better street designs for dwelling motorbikes, are essential.



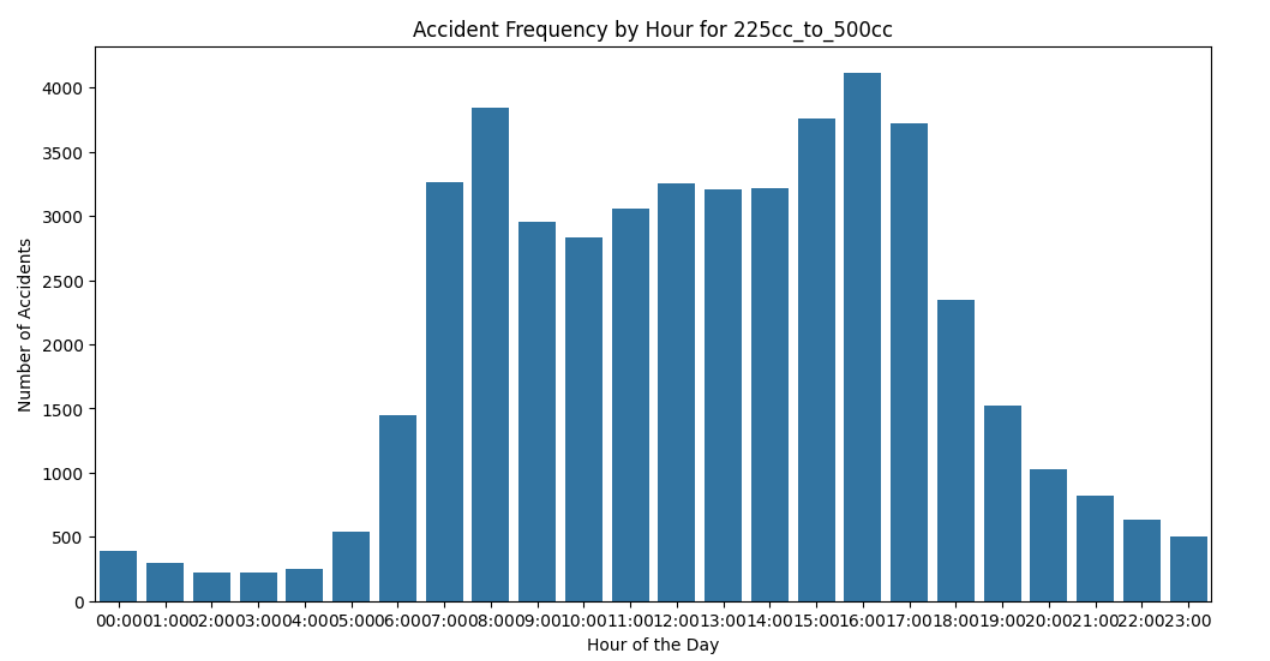
*Fig3: Motorbike Accident Frequency by Hour*



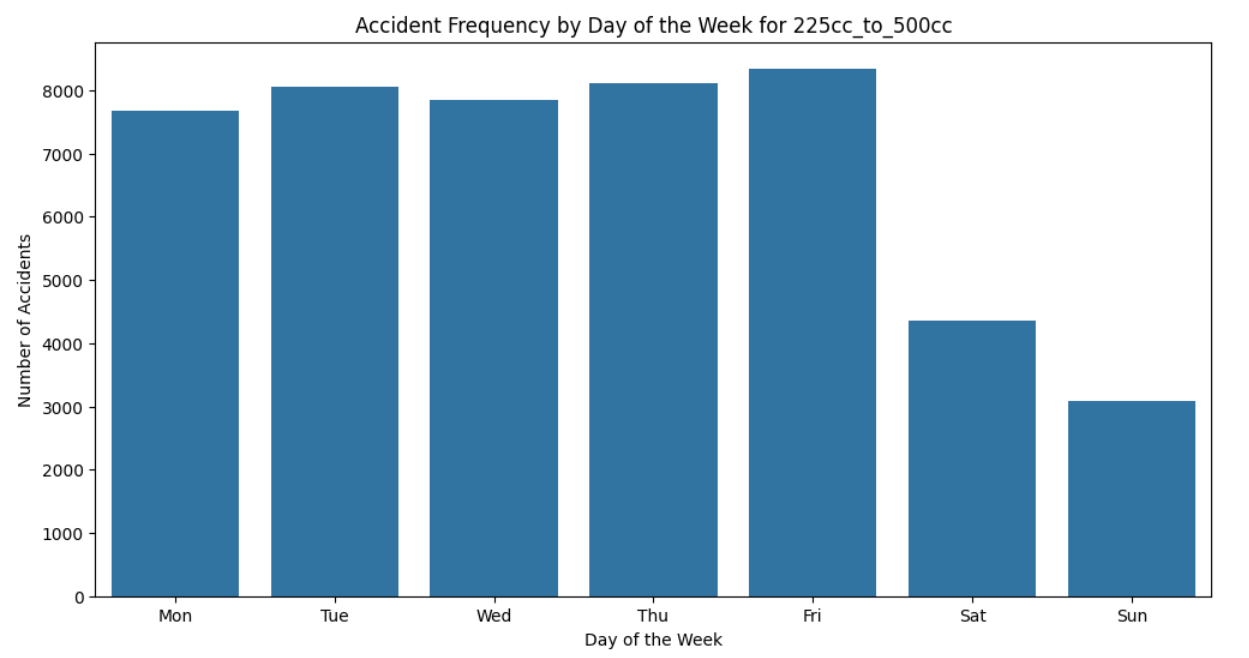
*Fig4: Motorbike Accident Frequency by Hour for 125cc and Under*



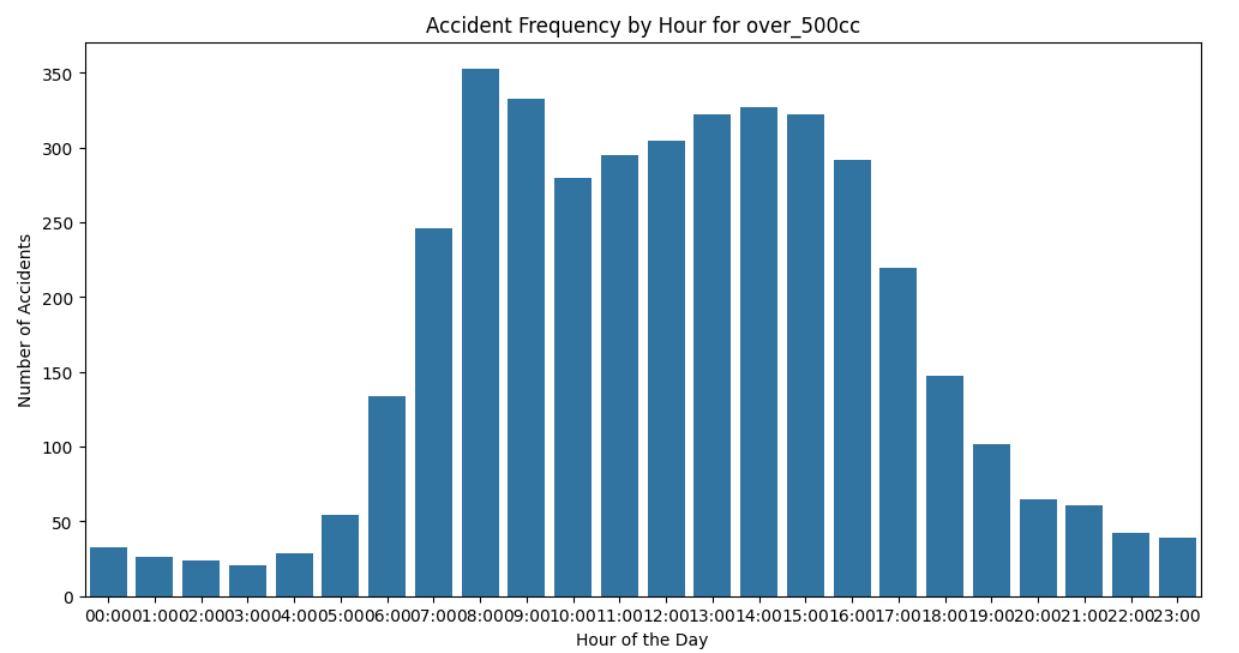
*Fig5: Motorbike Accident Frequency by week of day for 125cc and under*



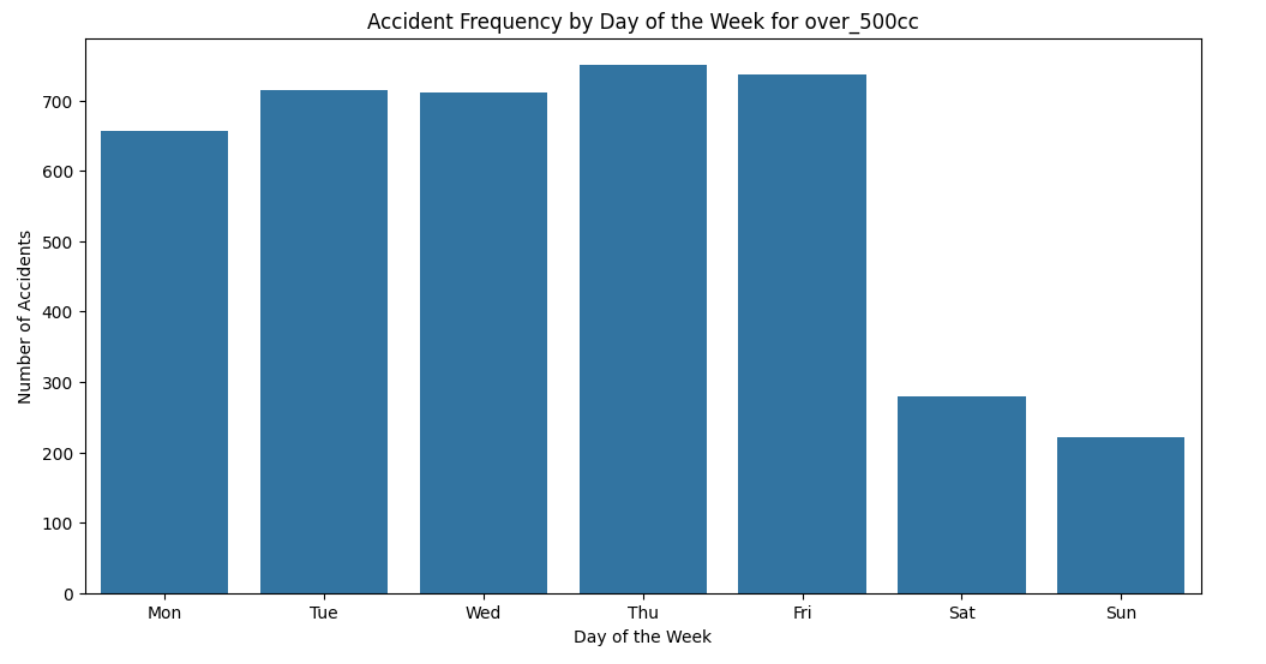
*Fig6: Motorbike Accident Frequency by hour for 225cc to 500cc*

****

*Fig7: Motorbike Accident Frequency by week of day for 225cc to 500cc*

****

*Fig8: Motorbike Accident Frequency by hour for over 500cc*

****

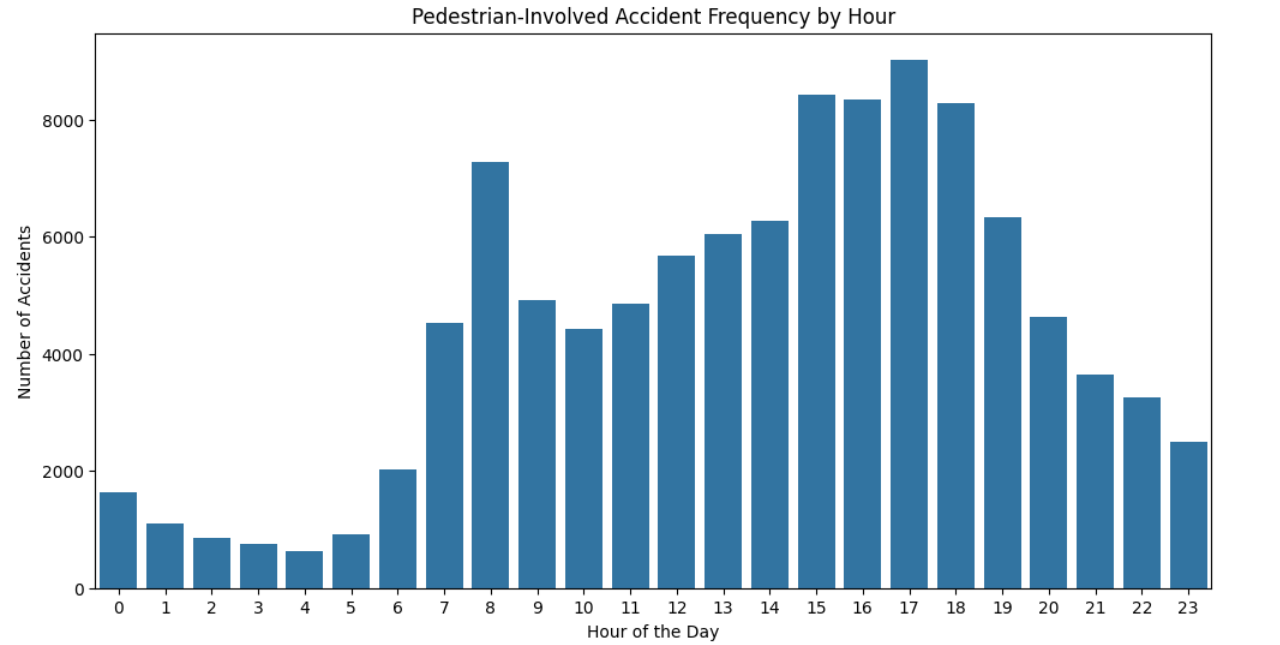
*Fig9: Motorbike Accident Frequency by week of day for over 500cc*

## **Task 3: Pedestrian Accidents by Hour and Day of the Week**

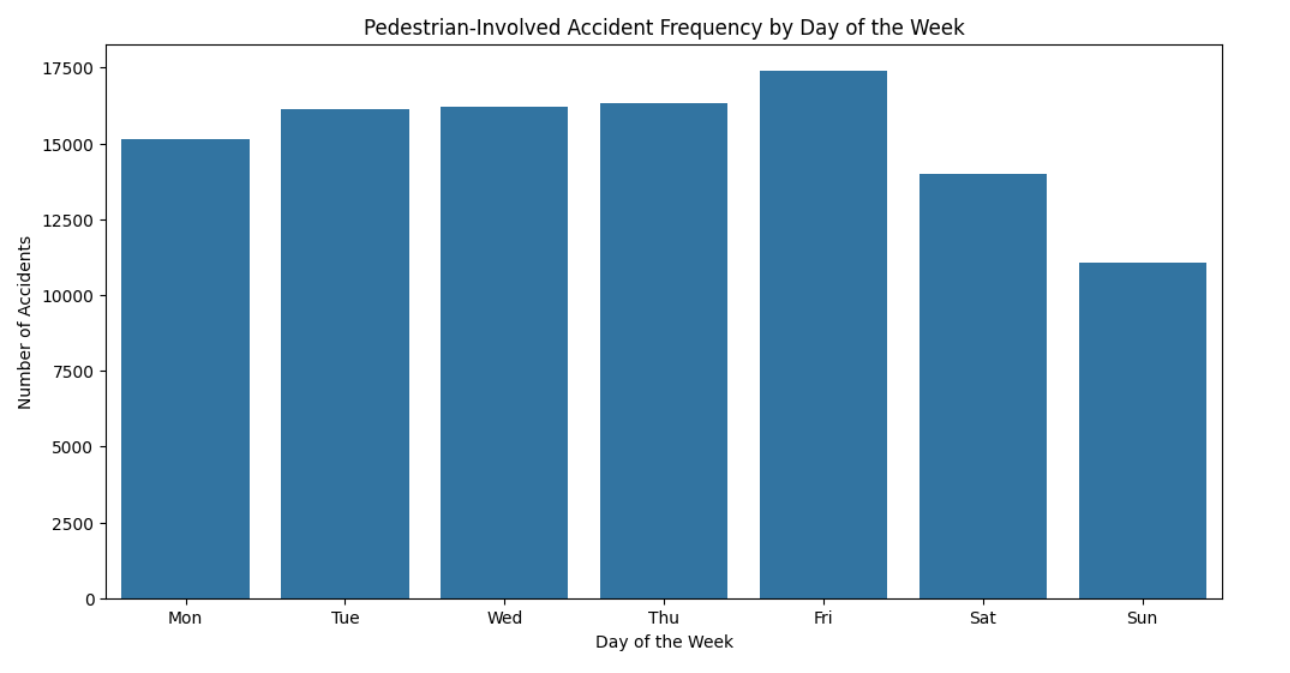
Pedestrian accidents have distinct timing characteristics that can safeguard vulnerable road users. The data also shows how pedestrian accidents are more prevalent during certain hours and days.

Fig10 shows that on Most roadway incidents involve pedestrians between 8 AM and 9 AM and 3 PM and 5 PM. These times are also whenever pedestrians are probably to be on roadways heading to work or school. These are conceivable because numerous children and older people are on the streets, requiring safer pedestrian crossings and police surveillance.

In Fig11 similar to the general accident trend, weekday pedestrian accidents are more frequent than weekend accidents. This pattern supports weekday safety interventions because to the high frequency of causalities. Installing traffic lights, well-lit zebra crossings, and public awareness of pedestrian crossing dangers can reduce risk.



*Fig10: Pedestrian Accident Frequency Hour*



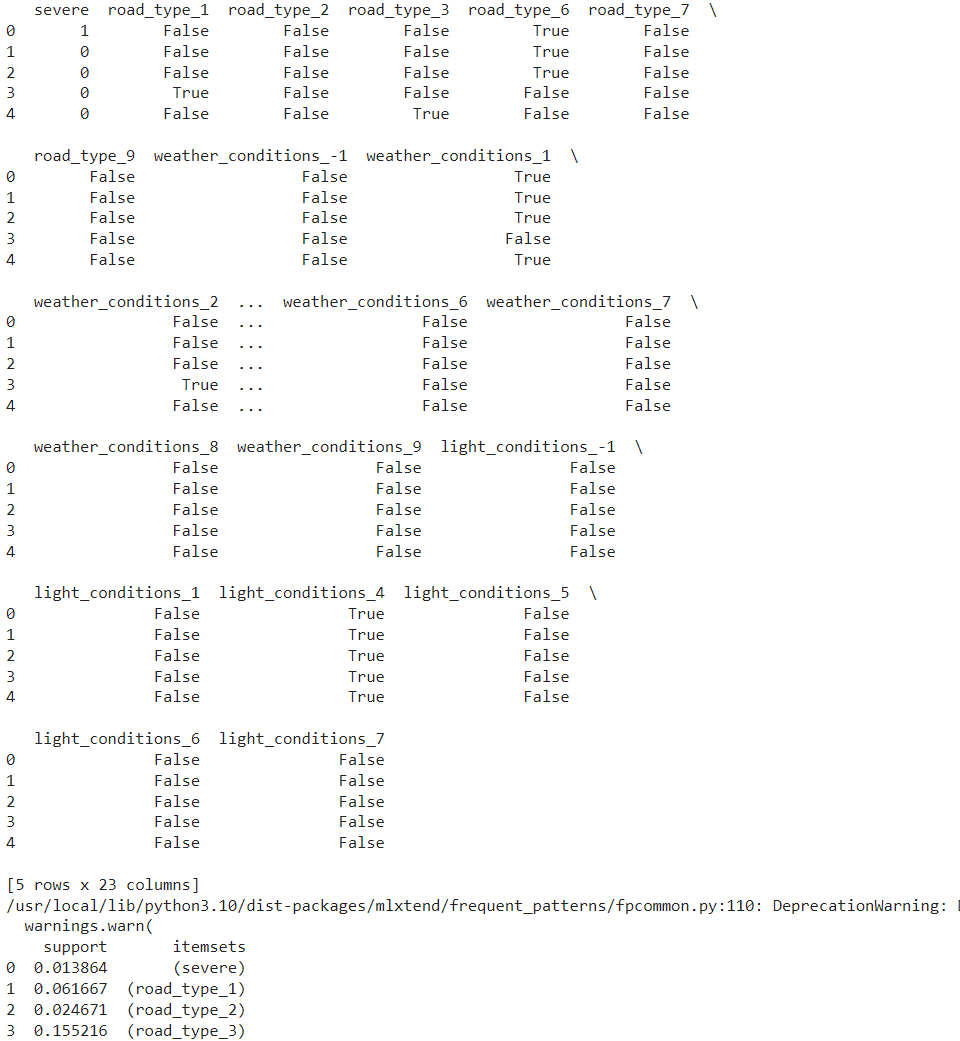
*Fig11: Pedestrian Accident Frequency by week of day*

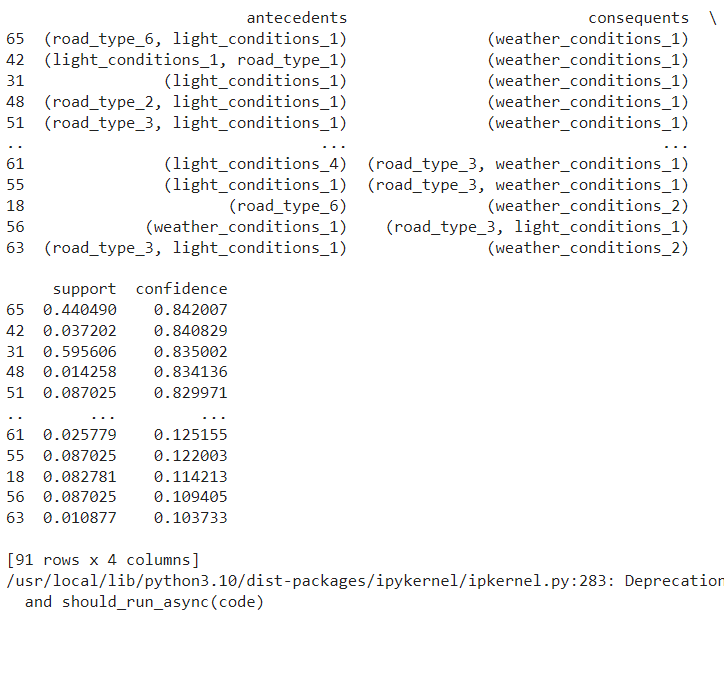
## **Task 4: Using the Apriori Algorithm to Explore Accident Severity**

The apriori method finds institutions with distinctive skills in the dataset, but it loses modest on accident severity under various settings. We focused on avenue type, weather, moderate circumstances, and coincidence severity for this investigation. We turned accident severity into a binary variable, where '1' denoted excessive accidents, then one-warm encoded the specific variables for preparing the data set for the apriori set of rules.

Using the apriori rules with a 0.01 help threshold, we found many common item units. Road\_type\_6 and light\_conditions\_1 were closely related to weather\_conditions\_1. Positive a combination of avenue types, climate conditions, and mild conditions substantially affected the twist of fate severity, according to the affiliation criteria from the common object units.

The results showed that road\_type\_6 (motorways) accidents in certain moderate and climate circumstances were more likely to be intense. The self-notion policies also showed that intense accidents often involved light\_conditions\_1 (daylight hours) and weather\_conditions\_1 (extraordinary without high winds). These findings suggest that even in seemingly safe conditions, some road types can cause serious accidents. This emphasizes the need for traffic safety measures, especially in those scenarios.



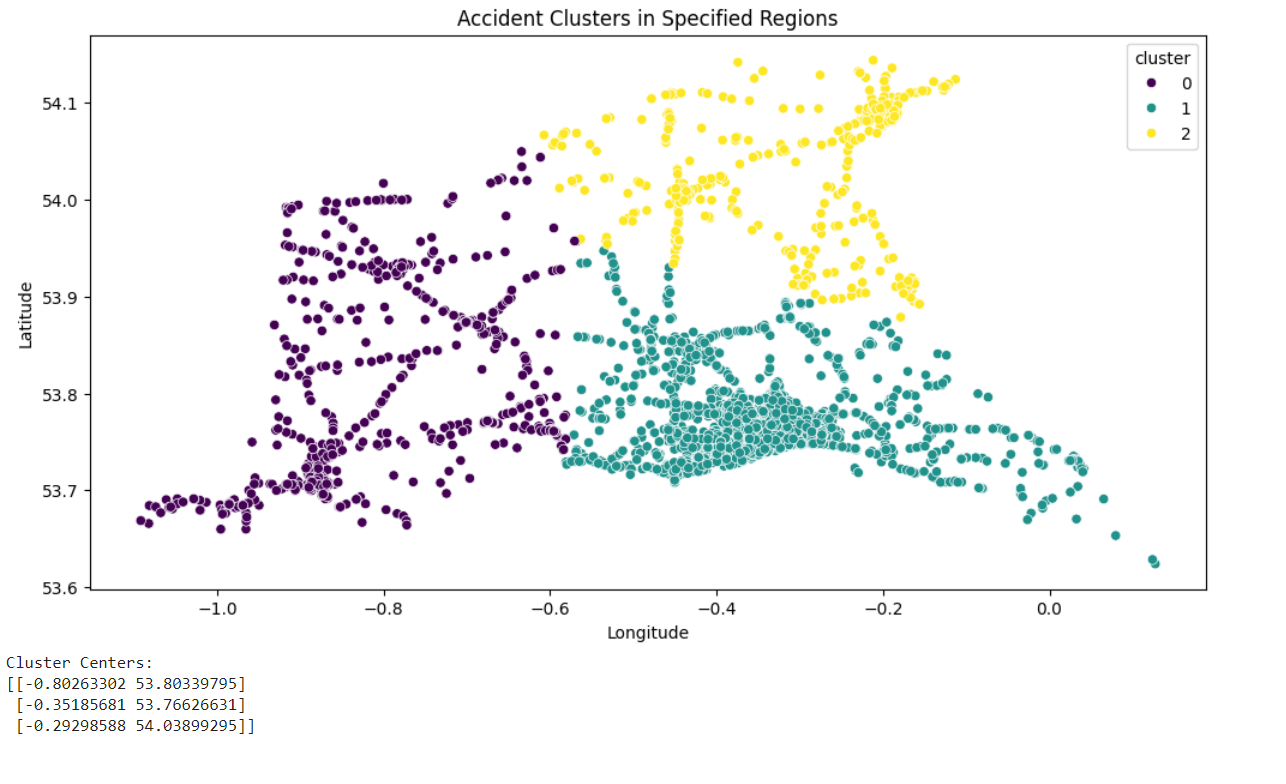


## **Task 5: Analyzing Accident Distribution Using Clustering**

Fig12 shows that to evaluate the event pattern in Kingston upon Hull, Humberside, and East Riding of Yorkshire, twist of fate surveys were coupled with LSOA statistics based on common codes. K-Means clustering with 3 clusters was used to identify accident distribution patterns by filtering by injuries inside the stated locations and grouping by longitude and latitude.

Scatter plots confirmed clustering and projected twist of destiny hotspots. Cluster centres calculated longitude and latitude, showing accident-prone locations. One cluster middle became a round (-0.89), another a square (-0.37), and the third a square (-0.44).

These clusters found that positive locations in the considered areas had higher twist of destiny densities, perhaps due to factors like headquarters site visitors, road layout, or local riding infrastructure. It showed where injuries were most severe, helping local authorities develop targeted interventions. Increased signage, avenue layout adjustments, enforcement, and assurance on such routes must greatly reduce twist of fate events in such hot regions.



*Fig12: Accident Clusters in a Specific Region*

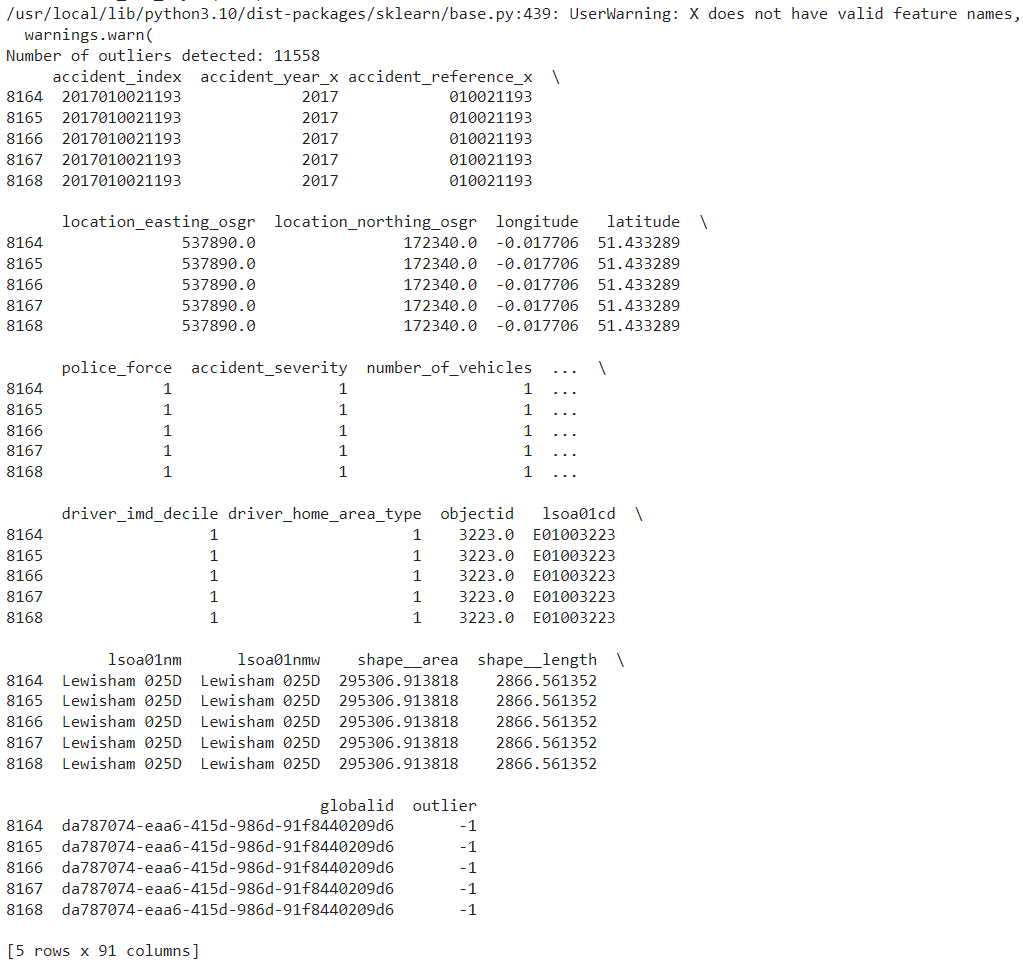
## **Task 6: Identify Unusual Entries Using Outlier Detection Method**

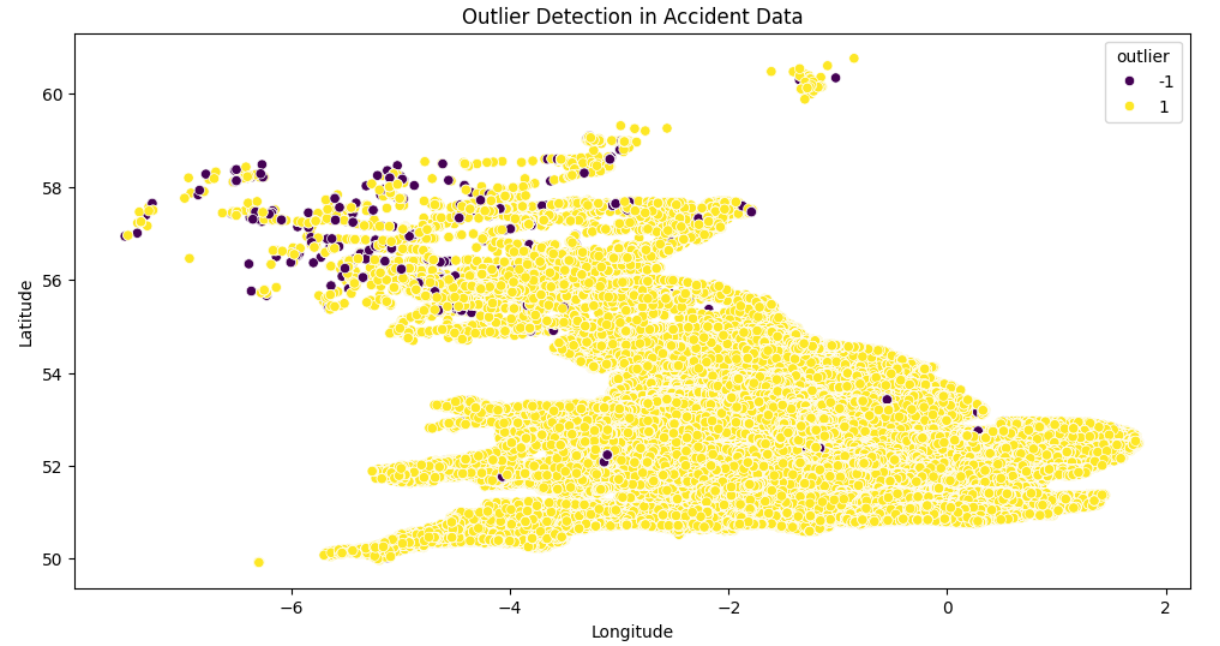
Outlier analysis is needed because anomalies are rare and outside the normal range. The Isolation Forest approach was used to find outliers in the accident database. Thus, our analysis used Longitude, Latitude and the number of vehicles, casualties, and accident severity.

Missing values were either eliminated from the dataset or filled with the average value before feeding it to the Isolation Forest model to reduce noise. The model's contamination rate was 0.01, indicating cleanliness, identified 11,558 outliers new and rare records in the dataset.

Fig13 shows the scatter plots outlier locations. These may involve unusual accident conditions, data gathering or recording errors, or other out-of-range conditions. Therefore, various longitude and latitude outliers may vividly show places with abnormal accident likelihood.

Whether to keep outliers in the dataset depends on their nature. If they show rare but true accidents, they can help improve road safety in severe instances. If it's attributable to data entry variables, they must be fixed or eliminated to improve data accuracy. These outliers should be examined more closely to distinguish between these two circumstances and choose the best course of action.





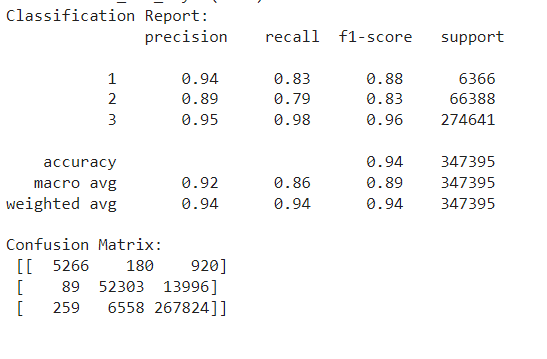
*Fig13: Outlier Detection in Accident Data*

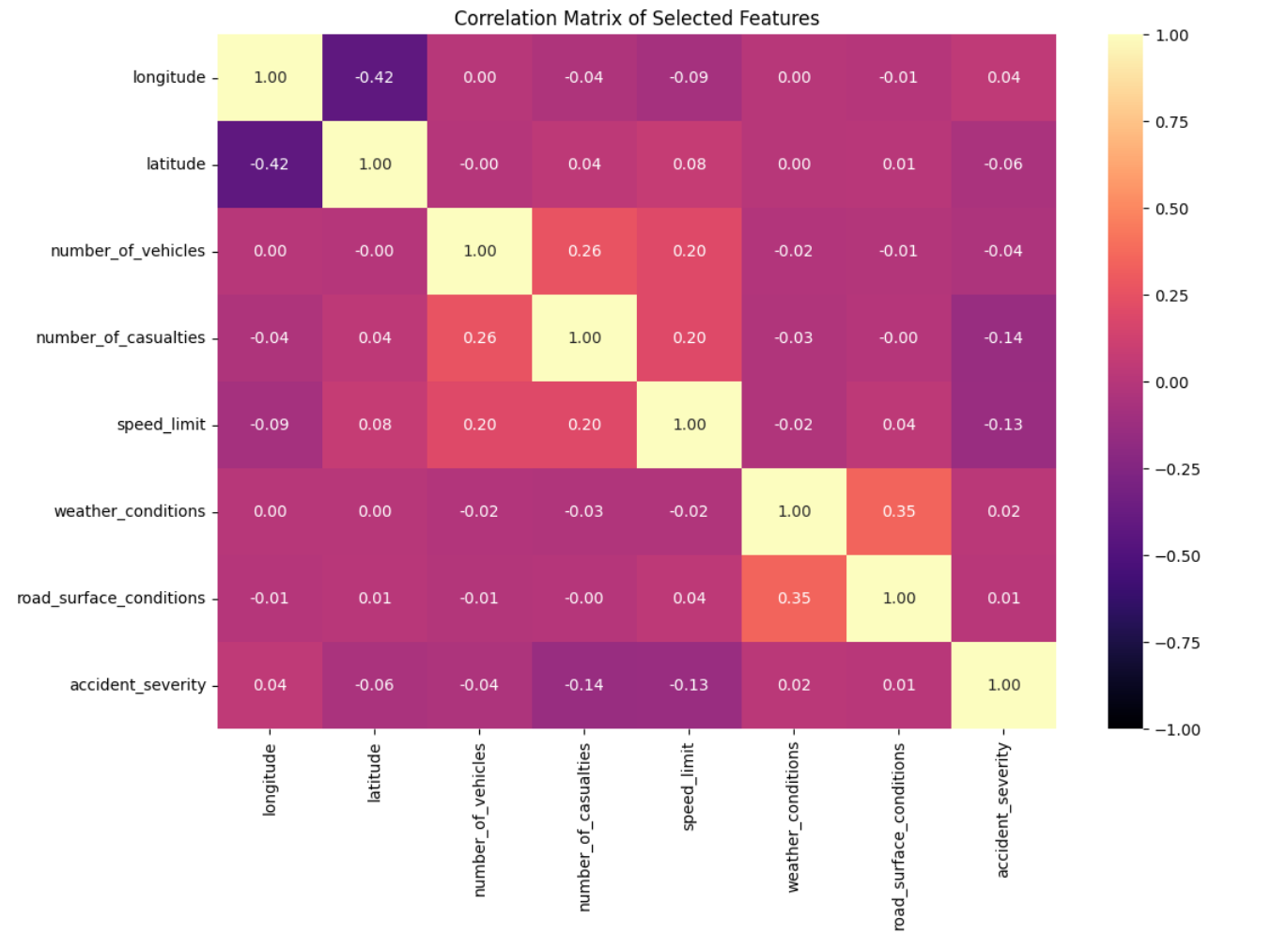
# **Predictions**

## **Task 7: Developing Classification Model**

Regarding the data used to train the model that would classify the kind of accidents as fatal or not fatal, we used RandomForestClassifier. Some of the key factors included the geographical coordinates of the accident, the number of vehicles and people, the speed limit, the type of weather at the time of the accident, the type of road surface, and the target variable which was the severity of the accident. The missing values were dropped and measurements of them were taken, and the categorical variables were fit by using LabelEncoder.

Dividing the collected data into the training subset and the test one was followed by feature scaling using the StandardScaler tool. The RandomForestClassifier gave a relatively impressive test set accuracy of about 94 per cent. Thus, the following statement of the classification report recollects the severe level one fatal accident and its precision and recall 0. 94 and 0. 83, respectively. From the confusion matrix, it was observed that there was a good classification and very few mistakes were committed in the categorization of the severe accidents thus the model can go a long way in assisting in matters concerning road safety.





*Fig14: Correlation matrix of Selected Features*

# **Recommendations to Government Agencies for Improving Road Safety**

1. Increase road construction and improve the luminosity of street lamps in areas which have many fatal accidents to help bring down the death rates.
2. Create response strategies about poor weather conditions such as black ice by programs warning and advancement of road works during such harsh weather to avoid disasters.
3. Concentrate on specific areas like Kingston upon Hull, Humberside, and East Riding of Yorkshire and increase traffic patrols, better road surface and campaigns on road safety.
4. Record group specimens to identify particular safety problems and to develop proper safety precautions.
5. Join the war on dangerous driving, and speed, and implement awareness creation to minimize accident severity.

# **Conclusion**

Therefore in understanding the real patterns, and risk factors relating to road traffic accidents, the analysis of the statistical data of 2020 and previous years in Great Britain is crucial. The following are some of the findings; there was a high number of accident occurrences during the peak business hours and weekdays, there is a difference between motorbike and pedestrian accidents and lastly; clustering resulted in the identification of hot spots of accidents. Accident severity and any strange manifestations were explained by analyzing the results of the apriori algorithm and the outlier indication. The RandomForestClassifier that was developed indicated high accuracy in cases of fatal accidents. There are specific ways on how the risk of accidents can be brought down and good road safety made a reality: targeted improvements such as better designed roads, properly developed weather-responsive plans, and enhanced public awareness are some of the ways that can be used to drive down the rates of accidents and generally improve road safety.

# **References**

*Reported road casualties in Great Britain: notes, definitions, symbols and conventions* (2022). <https://www.gov.uk/government/publications/road-accidents-and-safety-statistics-notes-and-definitions/reported-road-casualties-in-great-britain-notes-definitions-symbols-and-conventions>.

Zhou, X., Lu, P., Zheng, Z., Tolliver, D. and Keramati, A., 2020. Accident prediction accuracy assessment for highway-rail grade crossings using random forest algorithm compared with decision tree. *Reliability Engineering & System Safety*, *200*, p.106931.

Sinha, A., Vu, V., Chand, S., Wijayaratna, K. and Dixit, V. (2021). A Crash Injury Model Involving Autonomous Vehicle: Investigating of Crash and Disengagement Reports. *Sustainability*, 13(14), p.7938. doi: <https://doi.org/10.3390/su13147938.>

Bucsuházy, K., Matuchová, E., Zůvala, R., Moravcová, P., Kostíková, M. and Mikulec, R., 2020. Human factors contributing to the road traffic accident occurrence. *Transportation research procedia*, *45*, pp.555-561.

Shi, L., Yang, X., Chang, X., Wu, J. and Sun, H. (2023). An improved density peaks clustering algorithm based on k nearest neighbors and turning point for evaluating the severity of railway accidents. *Reliability Engineering & System Safety*, [online] 233(129734521582002), p.109132. doi: <https://doi.org/10.1016/j.ress.2023.109132.>

Mannering, F., Bhat, C.R., Shankar, V. and Abdel-Aty, M., 2020. Big data, traditional data and the tradeoffs between prediction and causality in highway-safety analysis. *Analytic methods in accident research*, *25*, p.100113.

Ahmed, A., Sadullah, A.F.M. and Yahya, A.S. (2017). Errors in accident data, its types, causes and methods of rectification-analysis of the literature. *Accident Analysis & Prevention*, 456(213954). doi: <https://doi.org/10.1016/j.aap.2017.07.018.>

Esenturk, E., Wallace, A.G., Khastgir, S. and Jennings, P. (2022). Identification of Traffic Accident Patterns via Cluster Analysis and Test Scenario Development for Autonomous Vehicles. *IEEE Access*, 10(5673290), pp.6660–6675. doi: <https://doi.org/10.1109/access.2021.3140052.>

# **Appendix**

