**SHORT AND LONG TERM DATA ANALYTICS & RECOMMENDATIONS USING THE 2016-2020 VEHICLE ACCIDENT DATA OF THE UNITED KINGDOM**

***Authors:***

*KOLLA BALAJI*

*AJAY MADHUKAR DHAKNE*

*DOROTA KUZMA*

*SNEHA SUSAN JOJI*

*DEGBEDZUI DEREK KWEKU*

*AMOAKO STANLEY*

# 

# Abstract

The project discusses the analysis of road accidents in the United Kingdom. We had the road accident dataset for 2016-2020. The data spanned a period of 5 years which gives us a better understanding of the situation and how instituted efforts can be made to address the cost involved in the severity. The data was analyzed using Wolfram code in Mathematica. The problem was treated as a prediction problem, with additional exploratory analysis. The results have been visualized using bar graphs, line charts, etc. By using statistical methods patterns were examined within various circumstances of the accidents e.g., day of the week, weather conditions. Attention was also given to a specific short period in 2016-2020 to examine the immediate effects of lockdown on vehicle accident frequency.

# 

# Contents

# Introduction

# Data Model & Source

# Methodology

3.1 Data Pre-Processing

3.2 Data Analysis

# Result

4.1 Number of accidents yearly and effect of lockdown

4.2 Day-wise Trend

4.3 Number of accidents based on the severity

4.4 Analysis of accident severity based on weather & light conditions and road type

4.3.1 Weather Conditions

4.3.2 Light Conditions

4.3.3 Road Type

4.5 Accidents by Speed Limit and its effect on accident severity

# Conclusion

# References

# 

# 1. Introduction

The UK government has put its efforts into building an aggregated collection of accident records in the England, Scotland, and Wales regions over a years. The available data includes millions of accident instances. Having access to this, one of the most comprehensive data sets on traffic, we work towards uncovering patterns and answering questions related to the potential causes and trends of accidents. This report is focused on data from 2016-2020 as we believe the advancements in technology related to motor vehicles from this period are similar, allowing for direct comparisons without additional adjustments. The data was obtained by downloading directly onto the researcher’s computers and imported into Wolfram Mathematica where the analysis was conducted.

**Long term analysis:**

This section aimed to examine various trends of vehicle accidents between 2016 and 2020 using exploratory analysis. Previous studies have shown that accident risk can be influenced by weather conditions [1], day of the week [2], speed, road type, and accident consequences [3], among others. Therefore, we wanted to examine whether our findings would fall in line with previous literature when looking at the most recent period in the UK. Patterns were examined mainly in variables (i) day of the week, (ii) accident severity (Fatal, Serious, Slight), (iii) weather, light, road type conditions, (iv) accident severity and environmental conditions, (v) speed limit and accident severity, (vi) severity and light conditions.

**Short term close-up analysis:**

This section aimed to examine whether there were significant differences in the number of vehicle accidents in the UK before the Covid-19 pandemic and after it developed in 2020. It was reported before that road accidents were generally globally reduced when Covid-19 restrictions were in place [4] therefore we wanted to examine whether this was noticeable immediately after lockdown implementation. It has been suggested before that vehicle accident rates are affected by the Holiday period in December [5], therefore the earliest starting date for our comparison was January 1st.

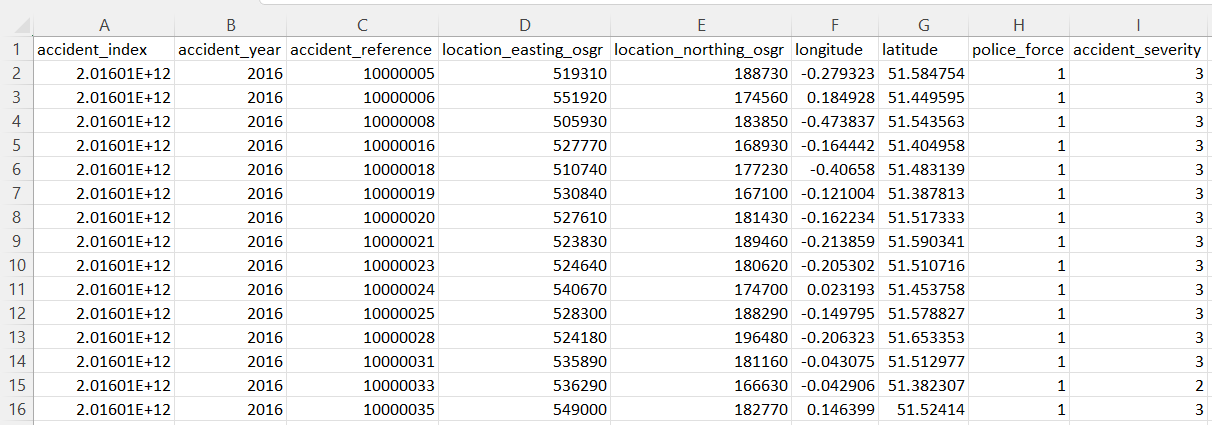
The first 84 days (85 in leap years) of 2016-2020 were compared with the 85 days immediately after the lockdown was introduced in the UK, which was on March 26th in 2020 [6]. The reason for looking at these periods in years before Covid-19 was to make sure the lower number of accidents was not a typical characteristic for this time of year. The objectives were to compare (i) the number of vehicle accidents in years 2016-2020, (ii) the first 85 days of 2020 with 85 days right after March 26th, 2020, (iii) the same periods as in (ii) in 2019, and broadly in 2016-2018. Exploratory descriptive statistics and independent sample t-tests were used for the analysis.

# 

# ​​2. Data Model & Source

Data was downloaded directly from the UK Department for Transport [7]. In particular, files “Road Safety Data – Accidents 2020”, “Road Safety Data – Accidents 2019”, “Road Safety Data – Accidents last 5 years” were used. The “Road Safety Data – Accidents last 5 years” file contained 36 pieces of detailed information about each of 597 973 vehicle accidents that took place in the UK between 2016-2020. The data in the first two files were also contained in the third file, the only reason for separation was a more convenient analysis. The dataset gives a detailed information of each accident that has taken place during 2016-2020 including information like the date, location, severity of the accident, casualties in each accident, weather and light condition in the spot of accidents and more.

**The structure of the main data “Accident Dataset” is as follows:**



*Figure 1. Structure of the Accident Dataset*

**Long term analysis:**

For this analysis, the Accident dataset 2016-2020 was used. It had a total of 36 columns out of which only the columns required for analysis were taken into consideration. These included: “accident\_year”, “accident\_severity”, “day\_of\_week”, “road\_type”, “speed\_limit”, “light\_conditions”, “weather\_conditions”.

Ongoing through the data, quite a few of the columns were found to be mentioned in integer values which otherwise should be in string form. Also, some columns have values like “Data Missing” or “Unknown”. These have not been excluded because then proper proportions for visual presentation of results might be lost, or during analysis, there might be confused with the total values. Following are the list of all such columns and the meaning of their corresponding integer values:

accident\_severity:1*→*Fatal,2*→*Serious,3*→*Slight

day\_of\_week:1*→*Sunday,2*→*Monday,3*→*Tuesday,4*→*Wednesday,5*→*Thursday,6*→*Friday,7*→*Saturday

road\_type:1*→*Roundabout,2*→*One way street,3*→*Dual Carriageway,6*→*Single Carriageway,7*→*Slip Road,9*→*Unknown,12*→*One way Street\Slip road,-1*→* Data missing

light\_condition:1*→*Daylight,4*→*Darkness-light lit,5*→*Darkness-lights unlit,6*→*Darkness-no lightings,7*→*Darkness-lighting unknown,-1*→*Data missing

weather\_condition:1*→*Fine no high winds,2*→*Raining no high winds,3*→*Snowing no high winds,4*→*Fine + high winds,5*→*Rain + high winds,6Snow + high winds,7*→*Fog or Mist,8*→*Other,9*→*Unknown,-1 *→*Data missing

**Short term close-up analysis:**

The only data used for this analysis were the year of the accident from column 2 “accident\_year”, and the exact date of each accident from column 12 “date”.

# ​​3. Methodology

The data files used were all CSV files. After a visual inspection in Excel, they were directly imported into Mathematica using ’Import’, or ‘SemanticImport’ which automatically turned the data into a Dataset within the programme. The data was not separated by vehicle type for the analysis. Some of the vehicle types included e.g., car, motorcycle, minibus.

## 3.1 Data Pre-Processing

**Long term analysis:**

The appropriate column for each of the analyses was taken from the “Road Safety Data – Accidents last 5 years” file. The required columns were taken using the ‘Part’ function. The ‘Count’ function was used to take the total count for each of the scenarios which could be displayed in the chart used. Chart enhancement functions like ‘ChartLegends’, ‘ChartLabels’, ‘PlotLabels’, ‘PlotLegends’, ‘LabelingFunction’, and ‘AxesLabel’ each depending on the type of chart taken for analysis.

**Short term close-up analysis:**

The appropriate columns “accident\_year” and “date” were selected using the ‘Part’ function from the “Road Safety Data – Accidents last 5 years” file and the accidents were sorted by date and counted using ‘SortBy’ and ‘Tally’. The mean number of accidents in the first and after March 26th 84/85 days of the years 2016-2018 was counted using ‘Mean’, and by separating the appropriate rows of each year using ‘Part’. The same means for 2019 and 2020 were calculated later from their separate files i.e. “Road Safety Data – Accidents 2020”, “Road Safety Data – Accidents 2019”.

The total number of accidents in the years 2016-2020 was calculated using ‘Length’, while the number of accidents per year was calculated using ‘Count’. Then the yearly total numbers of accidents were presented visually using ‘BarChart’.

From the imported separate accidents’ files for 2019 and 2020, only the “date” column was selected, then the dates were sorted and the number of accidents on each day of those years was done by using ‘Sort’ and ‘Tally’. Pre-processing for before and after lockdown analyses for 2019 and 2020 was done by selecting the first 84/85 days and the 84/85 days after March 26th and using ‘Part’. The data was turned into lists using ‘Normal’. Then a plot showing daily accidents from January 2019 to December 2020 was made using ‘DateListPlot’.

## 3.2 Data Analysis

**Long term analysis:**

For the analysis of the year 2016-2020, the first thing done was a yearly trend displaying the number of accidents that have occurred each year that can help to understand the frequency of accidents. A similar analysis was done for days of the week too, helping to understand the days with the least recorded accidents, or in other words, the safest day of the week to travel.

After this analysis was done to understand how severe have the accidents been over the last 5 years. This section also helped to understand how various conditions like weather, light, and road type affects the severity of the accidents. An analysis like this can help people to understand what conditions need to be avoided for safe travel and also can help the government to make improvements or changes in services which can be a step aiming to reduce the frequency of accidents. Functions ‘BarChart’, ‘ListLinePlot’ and ‘SectorChart3D’ were used to visually demonstrate the patterns in the various accident-related characteristics.

**Short term close-up analysis:**

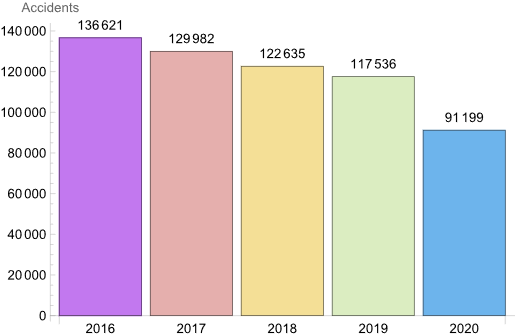
Data analysis was firstly focused on comparing the first 85 days to the 85 days after March 26th, when lockdown started, in 2020. The means and standard deviations were calculated using ‘Mean’ and ‘StandardDeviation’, and then an independent samples t-test of those means was run using ‘Ttest’. The same method of analysis was then applied to (a) the first 85 days of 2020 and the first 84 days of 2019, and (b) the 85 days after March 26th in both 2019 and 2020. Then a bar chart was created using ‘BarChart’ to demonstrate the mean number of daily accidents in the first 84/85 days and the 85 days after March 26th in both 2019 and 2020. Lastly, two more bar charts were created using ‘BarChart’ to demonstrate the mean number of daily accidents in the same two time periods as specified before, but across all years 2016-2020.

# 

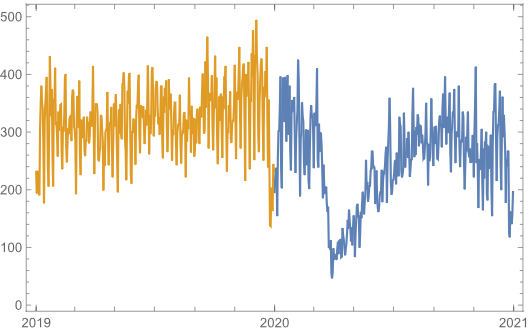
# 4. Results

## 4.1 Number of Accidents Yearly and effect of Lockdown

The total number of accidents in the UK between 2016 and 2020 was 597 973 (see Figure 2 for yearly totals). It can be seen that there was a steady decline in the number of accidents year to year, with a steeper decline In 2020. The detailed plot of daily vehicle accidents in 2019 and 2020 can be seen in Figure 3.

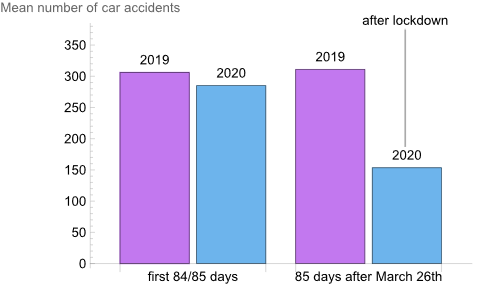


*Figure 2. The total number of vehicle accidents in each year between 2016-2020.*



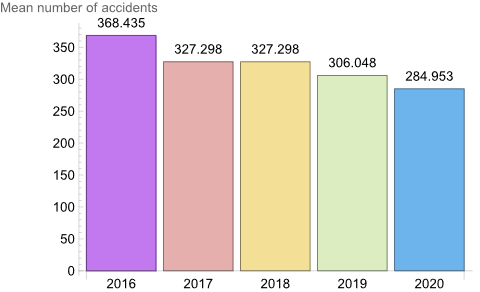
*Figure 3. Daily vehicle accidents between 2019 and 2020; y-axis displays number of accidents, x-axis is the passage of time from the beginning of 2019 to the end of 2020.*

An independent samples t-test showed that there was a significant difference in the number of vehicle accidents in the first 85 days of 2020 before lockdown (M = 285, SD = 71.43), and the 85 days after lockdown was introduced (M = 153.45, SD = 55.22); *t*(168) = 13.43, *p* < 0.001. Therefore, the number of vehicle accidents during the 85 days after the lockdown was introduced in 2020 was significantly lower than at the beginning of that year (Figure 4).



*Figure 4. Bar chart showing the mean number of accidents in the first 84 days of 2019 and first 85 days of 2020 (additional day because 2020 is a leap year), and the mean number of accidents in the 85 days after March 26th in 2019 and 2020, wherein 2020 that was the day lockdown was introduced.*

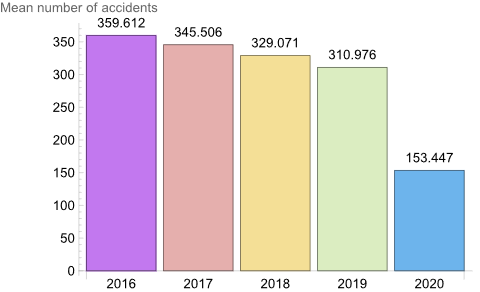
Furthermore, an independent samples t-test showed that there was a significant difference between the number of vehicle accidents in the first 84 days of 2019 (M = 306.05, SD = 55.93) and the first 85 days of 2020 before lockdown was introduced (M = 285, SD = 71.43); *t*(167) = -2.14, *p* = 0.034. Therefore, the number of vehicle accidents in early 2020 was significantly lower than during the same period in 2019. As the t-value was not very large, indicating more similarity between the samples, this result might be explained by the steady decline in the number of accidents observed in the past 5 years (see Figure 2), rather than the effects of the beginning of the pandemic. Indeed, this is further supported by looking at the mean daily accidents early in the year in the period 2016-2020 (see Figure 5). Apart from 2017 and 2018, every year there was a decline in the number of accidents early in the year.



*Figure 5. Mean the number of accidents in the first 85 days of the year between 2016-2020. For years 2017, 2018, and 2019 it is the mean number of accidents in the first 84 days, as those were not leap years.*

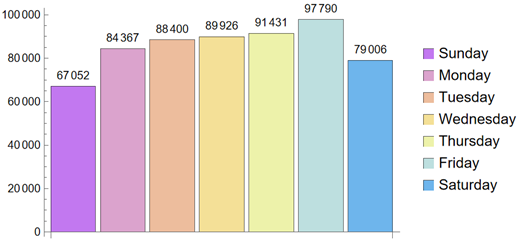
Additionally, an independent samples t-test showed that there was a significant difference between the number of vehicle accidents in the 85 days after March 26th in 2020 (M = 153.45, SD = 55.22) and the same 85 days in 2019 (M= 310.98, SD = 49.38); *t*(168) = -19.61, *p* < 0.001. Therefore, the number of vehicle accidents in the 85 days after the lockdown was introduced on March 26th, 2020 was significantly lower than during the same time of year in 2019. For a comparison of all 4 aforementioned means see Figure 4.

As the t-value was way larger than when comparing yearly days of 2019 and 2020, this suggests that the difference between the frequency of accidents was large, therefore lockdown might have affected the number of vehicle accidents in this period. Furthermore, as can be seen in Figure 6, the decline in the number of accidents right after March 26th in 2020 was steep when compared to not only 2019, but also all the way to 2016. Overall, these findings fall in line with Yasin et al. [4] who found that when Covid-19 restrictions were in place, road accidents were generally reduced on a global scale, but additionally they show that these effects could already be seen immediately after lockdown implementation.



*Figure 6. Mean number of daily accidents in the 85 days after March 26 (when lockdown started in 2020) in years 2016-2020.*

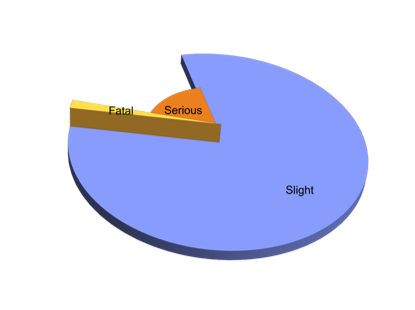
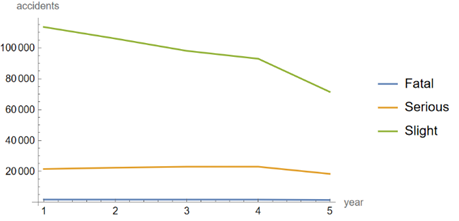
## 4.2 Day-wise Trend:



*Figure 7. Number of accidents during each day of the week in 2016-2020.*

Figure 7 shows that the highest number of accidents have been recorded to be on Friday (97790) and the lowest on Sunday (67052), indicating that the end of the work week seems to be the busiest and most accident-prone day and the end of the weekend as the least busy. This can also be considered as an analysis to know the safest day for people to travel. These findings differ from previous literature. For example, in the US the most accident-prone day was found to be Saturday [8].

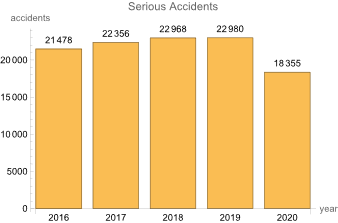
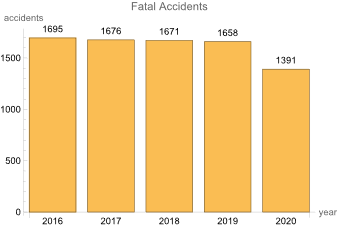
## 4.3 Number of accidents based on the severity of the accidents

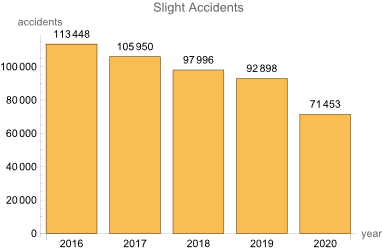
 

*Figure 8. Overall proportion of accident count based on severity, and accidents severity over the years.*

Figure 8 shows that the majority of the accidents have been of slight severity and this has been consequently decreasing over the last 5 years, with an immense reduction during the year 2020. The number of serious accidents seems to have increased during 2016-2019 but there has been a subsequent reduction in these accidents during the year 2020.

We can have a detailed analysis for each of these severities for each year as shown in Figure 9:



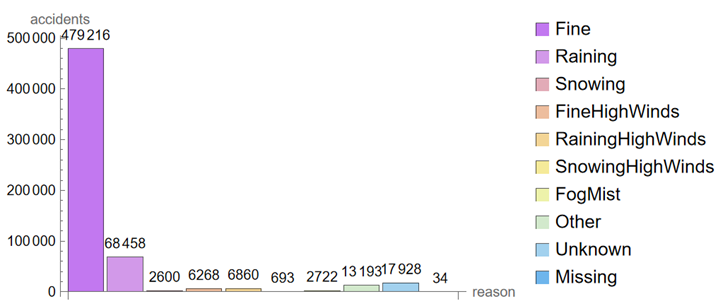


*Figure 9. Fatal, Serious and Slight Severity accidents during 2016-2020*

The above figures show year wise breakdown of the number of accidents based on accident severity. It can be seen that although all types of accidents have reduced manifold in the year 2020, only the slight severity of accidents have been reduced by quite a number consequently in the last five years.

## 4.4 Analysis of accident severity based on weather & light conditions and road type

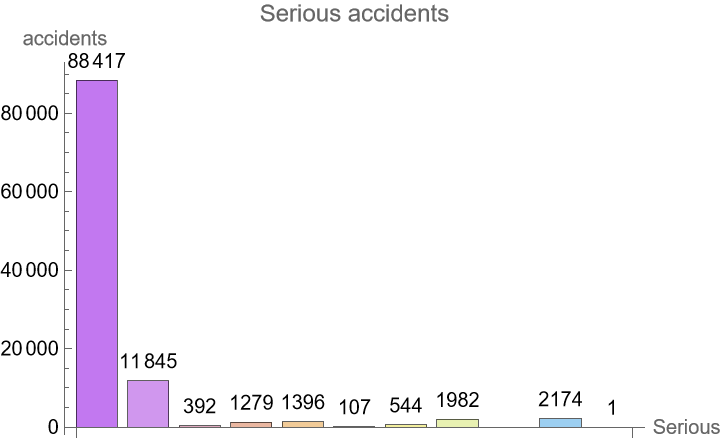
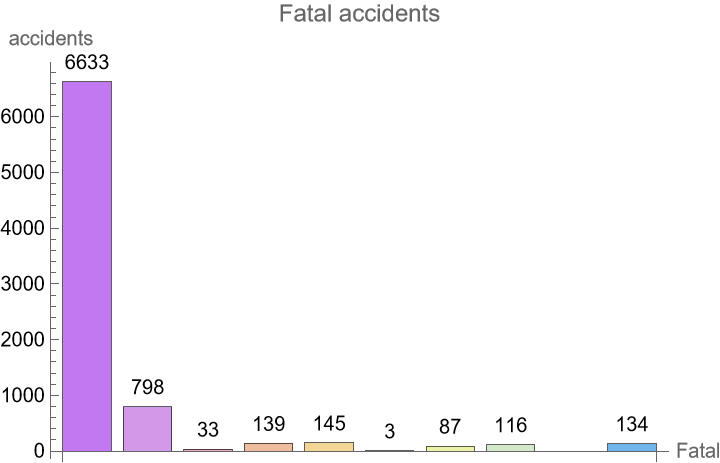
### 4.4.1 Weather Conditions

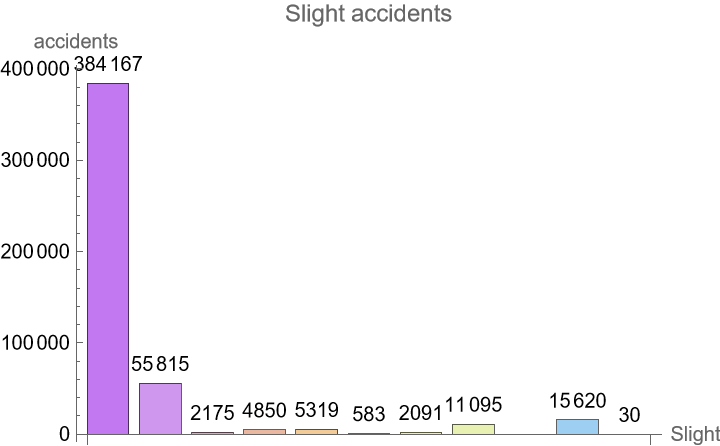
****

*Figure 10. Overall accidents caused by each weather condition in the 5 years.*

From Figure 10, we conclude that the number of accidents that happen during perfectly fine weather is much higher in count compared to the conditions which could adversely affect the normal traveling conditions. This analysis makes it evident that adverse weather conditions are not the root cause of accidents.

Now analyzing the effect of these weather conditions on the severity of an accident.



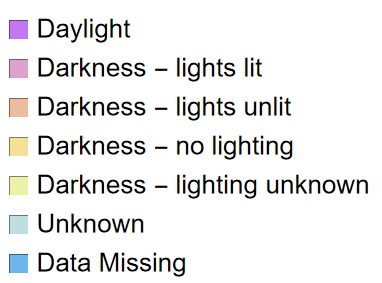
 

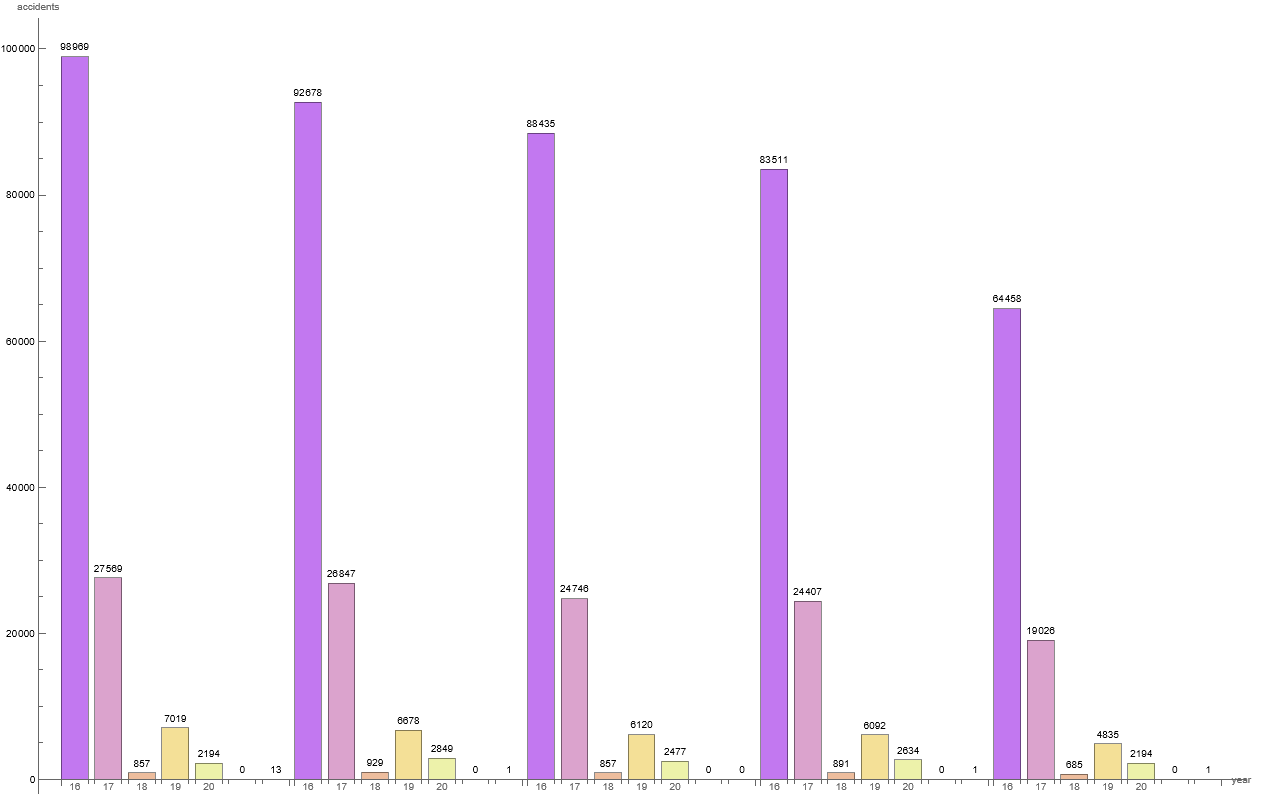
*Figure 11. Severity of an accident caused by weather conditions in the 5 years.*

As understood from the above charts, the number of slight severity accidents is highest irrespective of the weather conditions. This helps to conclude that the weather conditions have no impact on the severity of the accidents.

### 4.4.2 Light Conditions

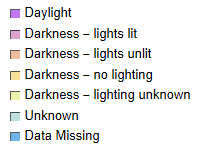
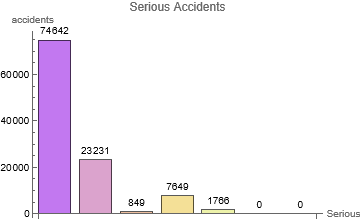
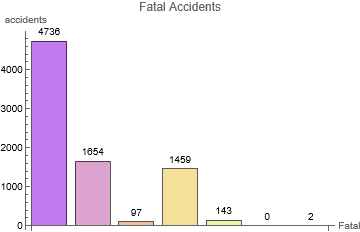
We discovered that most accidents happen either in the daylight or when the streetlights are present and lit in the darkness (Figure 12). This helps to conclude that lack of light conditions cannot be considered a reason for the accidents taking place.

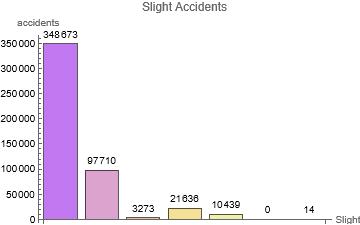
****

****

*Figure 12. Overall number of accidents by light conditions from 2016 to 2020.*

Now we can analyze how much the light conditions contribute to the severity of the accidents.

****

****

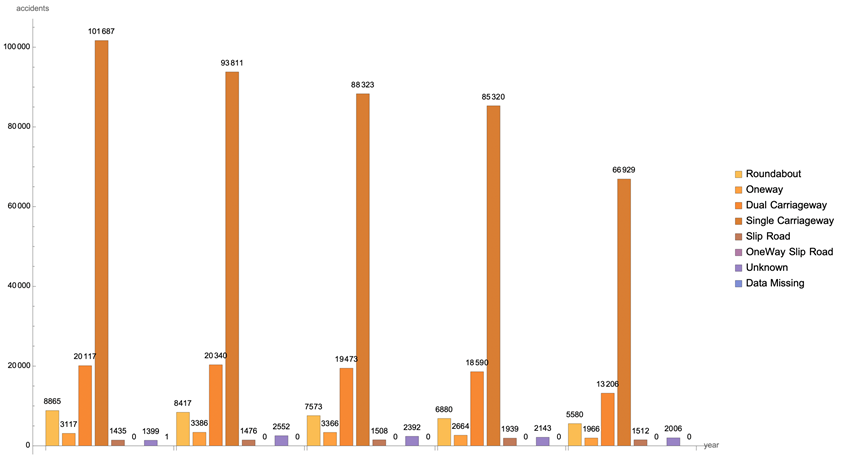
*Figure 13. Severity of an accident caused by light conditions in the 5 year period.*

Knowing the severity of an accident in different light scenarios can help to understand if the frequency of these accidents are due to lack of service or human error. In the figures above, in all severity cases, the accidents due to daylight are highest followed by the accidents due to lights lit when dark. One thing that can be observed here is that the number of fatal accidents due to light conditions is less compared to the fatal accidents caused by weather conditions.

### 4.4.3 Road Type

The type of road has a huge impact on the severity of the accidents caused.

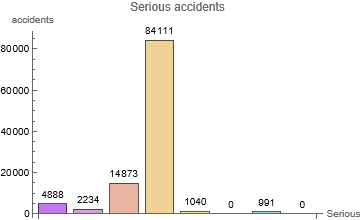
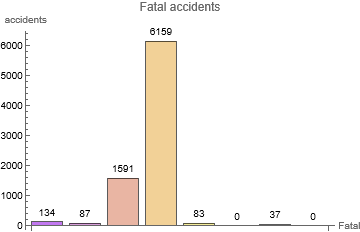
The figure below shows the frequency of accidents due to each road type.

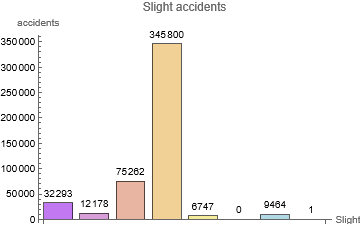
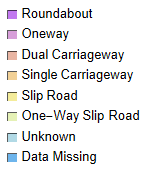


*Figure 14. Overall number of accidents by road type from 2016 to 2020.*

Figure 14 shows that the maximum number of accidents are caused in ‘Single Carriageway’ with 101687 accidents in 2016, and then consistently reducing over the years to 66929 in 2020. ‘Dual Carriageway’ being the second highest in causing the most accidents shows a slight increase in 2017 and then reduced each year moving forward. ’Slip Road’ has not caused many accidents over the past 5 years but needs to be put into consideration as there has been a slight increase each year in the number of accidents caused by it. See Mathematica notebook for detailed bar charts with better visibility of values for each year.

Further analysis of the severity of the accidents due to the different road types is shown below:

****

** **

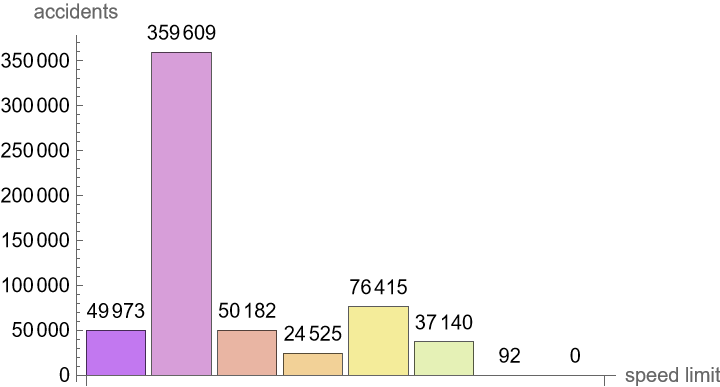
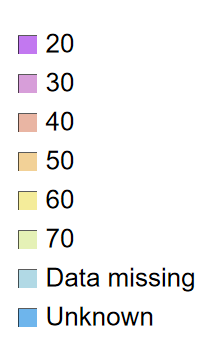
*Figure 15. Severity of the accidents due to the different road types in the 5 years.*

As observed above in Figure 15 (Fatal, Serious & Slight accidents), in all of the cases ‘Single Carriageway’ causes the highest number of accidents causing 6159 fatal accidents, 84111 serious accidents, and 345800 slight accidents. This can help the travelers to understand which is the type of road that isn't safe to travel especially when still learning. Further analysis can be done on this suggesting which road would be the best to take in a particular weather condition.

Considering the various factors, that is, weather, light, and road type, the weather condition is the main reason for fatal accidents, followed by light conditions and the reason affecting the least among these three is road type. This gives an idea of what factors to be extra careful about when on road.

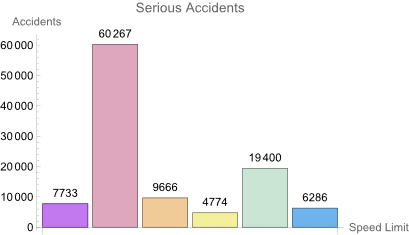
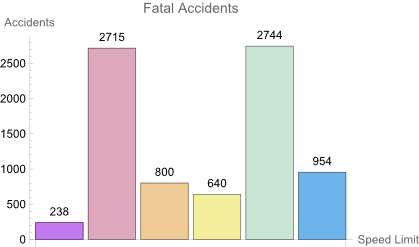
### 4.4.4 Accidents by Speed Limit

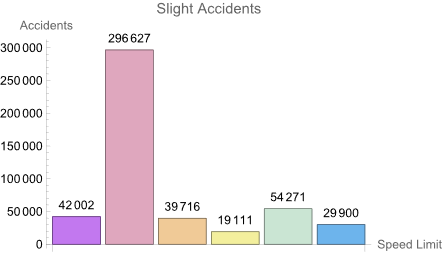
Though not a very directly understood reason, Speed Limits can be considered to have quite an influence on the number of road accidents. Figure 16 below shows the number of accidents caused due to each of the speed limits on particular roads.

*Figure 16. Overall number of accidents by speed limit in miles per hour (mph) in the 5 years.*

Figure 16 displays that the maximum road accident is at a speed limit of 30mph and the lowest at 50mph. We can further analyze the severity of the accidents caused by different speed limits.





*Figure 17. The severity of the accidents is due to the different speed limits in the 5 years.*

The fatal accidents are seen to be highest on roads having a speed limit of 60mph and lowest at 20mph. On the other hand, serious and slight accidents are seen to be highest on roads having a speed limit of 30mph and lowest at 50mph (Figure 17). This makes it quite a discussion as there seems to be some relationship between the severity of the accident being slight or serious to the speed limit of the road.

# 5. Conclusion and Further Work

In conclusion, this analysis provided us with useful insights on different ways showing how the UK government can use its collective data to improve road safety. From the short-term analysis it has been found that between 2016-2019 there was a steady decline in the number of yearly vehicle accidents, while in 2020 there was a comparatively large drop in the number of accidents. Furthermore, the decrease in the number of accidents immediately after lockdown implementation was significant when compared to the early days of 2020 or similar periods in 2019. Therefore it appears that vehicle accidents decreased in the UK immediately after lockdown. However, the data we had was from all of the UK and while the first lockdown was implemented all over around March 26th, as time went on local restrictions differed in e.g. Scotland and England. Therefore, future analysis should take the location of accidents into account when trying to examine the effects of lockdowns and restrictions.

Long-term analysis has revealed various patterns in accident-related circumstances. The largest number of accidents in 2016-2020 was found to happen on a Friday. Most accidents were of slight severity, with large reductions of serious and fatal accidents in 2020. Furthermore, most accidents took place during fine weather and appropriate lighting conditions, suggesting that adverse weather and lack of light is not the main cause of accidents; however, weather conditions seem to affect the numbers of fatal accidents more than light conditions. In regards to road conditions, most accidents happen on Single Carriageway and Dual Carriageway roads, with Single Carriageway accidents being the most fatal of roads. When it comes to speed limits, most accidents happen in areas with a time limit of 30mph, but fatality is highest at the speed limit of 60mph.

Further analysis of this data could focus on running more advanced correlations with multiple accident-related circumstances to draw more detailed conclusions and ways of preventing accidents in the future. As our data included information about various accidents, in future analysis it would be important to separate the accidents by type of vehicle. As for the short-term analysis, after data is collected for 2021, looking at the impacts of prolonged lockdowns and restrictions could be further insightful in a way that could not have been measured ever before.

# 

# 6. REFERENCES

[1] Malin, F., Norros, I., & Innamaa, S. (2019). The accident risk of road and weather conditions on different road types. *Accident Analysis & Prevention*, *122*, 181-188. “https://www.sciencedirect.com/science/article/pii/S0001457518308455”.

[2] No author. Crashes by Time of Day and Week. National Safety Council website. Accessed on November 19, 2021 “<https://injuryfacts.nsc.org/motor-vehicle/overview/crashes-by-time-of-day-and-day-of-week/>”.

[3] Pikūnas, A., Pumputis, V., & Sadauskas, V. (2004). The influence of vehicle speed on accident rates and their consequences. *Transport*, *19*(1), 15-19.

[4] Yasin, Y. J., Grivna, M., & Abu-Zidan, F. M. (2021). Global impact of COVID-19 pandemic on road traffic collisions. *World journal of emergency surgery*, *16*(1), 1-14. “<https://doi.org/10.1186/s13017-021-00395-8>”.

[5] No author. Holiday Traffic Fatality Estimate. National Safety Council website. Accessed on November 19, 2021

“[https://injuryfacts.nsc.org/motor-vehicle/holidays/christmas-day/data-details](https://injuryfacts.nsc.org/motor-vehicle/holidays/christmas-day/data-details/)”.

[6] No author. Timeline of UK government coronavirus lockdowns. Institute for Government website. Accessed on November 19, 2021

**“**<https://www.instituteforgovernment.org.uk/charts/uk-government-coronavirus-lockdowns>”.

[7] Road Safety Data. Department for Transport. Accessed on November 19, 2021 “<https://data.gov.uk/dataset/cb7ae6f0-4be6-4935-9277-47e5ce24a11f/road-safety-data>”.

[8] No author. Crashes by Time and Day of Week. National Safety Council website. Accessed on November 19, 2021

“<https://injuryfacts.nsc.org/motor-vehicle/overview/crashes-by-time-of-day-and-day-of-week/> ”.