**MGMT 479 Individual Project Report**

[Deadly Traffic Accidents in the UK (2015)](https://www.kaggle.com/datasets/kwullum/deadly-traffic-accidents-in-the-uk-2015)

Zhaoming Hu

**Business Background:**

The United Kingdom is a developed country with its high-end industrialized. The transportation system is playing an important role in economic and social development. According to the UK government analysis, there are more than 1000 people dead in traffic accidents. It has become the most common cause of accidental death. In addition, traffic accidents are causing significant financial losses to the UK economy, such as medical costs, insurance compensation, road maintenance, and traffic congestion. Also, they are bringing implicit losses such as the cost of the police force and pedestrian crossing physical facilities.

I am targeting the policymakers in the transportation department. They are the people who have the ability can change the rule or ground transportation environment to improve the problem. In essence, I want to create a visualization from this dataset to make some actionable plans. By analyzing this dataset, I will provide some insight from this dataset to demonstrate that some traffic accident data in a particular area (since I want to make an actionable plan, I think I need to focus on a specific area rather than the whole UK) will correlate with some condition. Second, I will highlight some important information from this visualization which will allow policymakers to quickly understand what problem is existing in this area. At the same time, I will show my analysis process in the report. The reason is I want to keep this data transparent, which will make my conclusion creditable and convey the policymakers. Lastly, I wish my project is valuable for policymakers to make decisions.

**Problem:**

I am trying to define some questions regarding traffic accidents that can be explored. Firstly, I want to find some factors that have a coefficient with the Number of Causalities. Secondly, what is the overall trend in the Number of Causalities accidents over time, and are there any trends when broken down by factors such as Weather Conditions, Road Surface Conditions, or Urban/Rural areas? Thirdly, are there certain times of day or months when accidents are more likely to occur, and are these days related to any festivals or football game days? Fourthly, how effective have recent policy interventions been in reducing the accident rate? Fifthly, how does the presence of police officers at the scene of an accident affect the likelihood of fatalities, and is there a relationship between the police force and serious injuries? Lastly, where are the most dangerous roads and intersections in the region, and how can they be made safer?

**Data Cleaning:**

Before starting the analysis, I performed some data cleaning jobs to ensure better accessibility. One of the changes I made was to change the Date column’s default properties from day/month/year to month/day/year format. Having this change was necessary as it ensures the accuracy of the time sequence. When I tried using the Data column without this modification, the data would not follow a chronological order properly, as it would go through from January first to February first instead of moving from January first to January second. Another change I made was converting the data type of the Time column from data&time to string. This was useful as the data&time data type caused a strange representation in the visualization. By changing it to a string type, each individual time value became a separate data point. Furthermore, the data followed the 24-hour rules, and I removed the ‘am’ or ‘pm’ after the time to ensure consistency in the time values.

**Analysis:**

I created several Scatter Plot charts (Exhibit 1) to investigate the relationship between the Number of Casualties and various factors, such as Light Conditions, Road Type, Speed Limit, Number of Vehicles, Special Conditions at the Site, and Weather Conditions. From the charts, I found that only the number of vehicles showed a clear positive coefficient with a parameter value of 0.29. It is indicating that one unit increase in the Number of Vehicles involved in an accident will cause 0.29 Number of Casualties. I think it is reasonable since more vehicles in an accident mean more car drivers are at risk of getting hurt. However, all the other factors did not show a clear linear relationship in this case. I think the main reason is they are all discrete data, which can be challenging to model using single linear regression. I think using the exponential trend line will be the alternative that will be more fit for those data.

Based on the dual axis bar chart (Exhibit 2) I used to examine the relationship between the Number of Casualties and three variables: Weather Condition, Road Surface Condition, and Urban/Rural areas. In the visualization, all the bar type is the primary axis that represents the Number of Casualties, and the secondary axis shows the trend or distribution of each variable. I also added a filter to limit the data to the traffic accidents that only happened in January. The chart showed that there is a weak correlation between Weather Conditions and Road Surface. However, the Urban/Rural variable showed a clear trend, where the higher the sum of Urban/Rural areas, the higher the number of casualties. This is interesting to me, and I was the further investigation of the raw data. I discovered that the majority of the values in the Urban/Rural column were either 1 or 2. As the dataset resource on the UK.gov page was no longer available, I wanted some additional research that could be conducted to clarify the meaning behind the values. I finally found that the range of values from 0 to 3 corresponded to different levels of urbanization: 0 represented a rural area with a sparse population, 1 represented an urban area with a city or town, 2 represented an urban area with a sparse population, and 3 represented an urban area with a minor conurbation. I found out that there is 62.3% Number of Casualties happen in 1 area and 37.6% Number of Casualties happen in 2 areas. Almost all the traffic accidents happen in Urban areas.

The Tree map (Exhibit 3) indicated that July had the highest Number of Casualties, whereas February has the lowest. I think there is one plausible explanation for this finding could be related to the Daylight-Saving Time policy, which extends daylight hours in the summer season, and consequently, people tend to spend more time outside. Moreover, I think this trend could also be linked to climatic conditions as summer is more favorable and comfortable for outdoor activities than winter. Additionally, I created the Heat map (Exhibit 4), which implied that the highest Number of Causalities occurred at two peaks: 8:00 and 17:00. I think this observation could be explained by the working schedule of people. Typically, people start their work at 8:00 and return home from work at 17:00. Hence, there could be a higher number of Vehicles on the road during these periods, leading to an increased risk of accidents.

In the dashboard (Exhibit 5), I utilized the map to filter out the data based on the speed limit in the selected area and visualized the corresponding Number of Casualties in the following bar chart. I chose the Westminster and Waterloo are in the UK map filter, which could be considered the downtown of London City. I saw that the bar chart showed that almost all accidents occurred at a speed limit of 30 miles. Based on what I thought higher speed limits might increase the likelihood of accidents, but this is not reflected in the graph. I think one possible explanation for this situation is that the speed limit is already set to a relatively safe level, and further reducing the limit might not have a significant impact on reducing accidents. In my opinion, the way to effectively reduce the Number of Casualties, policy makers could focus on improving Road Surface Condition, and educating drivers on safe driving practices.

In bar chart (Exhibit 6), I created an area chart to display the sum Number of Casualties (blue) against the average Police Force (green). I also created a calculated filed to select only the Number of Casualties that either below 300 or above 610. The two number represented the data points that are significantly below or above the mean of Number of Casualties. The chart reveals that the average policy force remains relatively constant between 26 to 32, regardless of the extreme number that I picked for casualties. I think this area chart suggested that a high average Police Force may not necessarily lead to a significant reeducation in the Number of Casualties. I think policy makers need to consider using an efficient average force could be an effective strategy to reduce the number of casualties while optimizing the Police Force. In Exhibit 7, I used the 1st Road Number dimension to color the police force filed and the Number of Casualties shown in bar chart. I chose the 1st Road Number because it is the most used by UK citizens. To balance out the Number of Casualties and Police Force across different road numbers, I used the quick calculation method with the percent of the total. Since the most road numbers have a very low number of casualties except road 0, using the percent of the total helps me to show a more balanced picture. From the visualization, there a dark blue of the Police Force in July and December. This could be due to increased police presence during holiday seasons and other major events. Also, I think the Number of Casualties is similar across different months. Therefore, I suggest policy makers could consider implementing targeted interventions during high-risk periods to reduce the Number of Casualties.

The last map (Exhibit 8) displayed a map that illustrated where the Number of Casualties over 8. The map also included Weather Condition in size mark and Road Surface Condition as color. Upon examining this graph, I conducted some research on the highways in the UK and found that those serious accidents frequently occur near main cities or highways. Additionally, the graph shows that several serious accidents occurred in high Weather condition and Road Surface Condition which indicates that the simultaneous presence of these two factors increases the likelihood of a severe accident.

**Conclusion:**

The next step that the policy maker can do is having an actionable plan. There are some steps that I recommend to the policy makers. Firstly, they can utilize the insights from the analysis to allocate the police force more efficiently during high-risk seasons and prioritize patrols during peak times of the day and month when accidents are most likely to occur. Additionally, the policy makers can work on implementing road improvements specifically on the sections with Road Surface Conditions rated above 6, as there are more prone to accidents.

Since the Weather Conditions are beyond our control, I think it is also advisable for policy makers to launch awareness campaigns to educate drivers on safe driving practices during adverse weather conditions. This could include informing the public about the dangers of driving during bad weather.

Another important action plan I would suggest that policy makers can consider is the implementation of technology-based solutions such as intelligent transportation systems and automated driving assistance systems. These systems can help to improve road safety by providing real-time information on road conditions, reducing driver error, and improving overall road efficiency.

**Appendix:**

**Exhibit 1**

**A picture containing text, line, screenshot, diagram

Description automatically generated**

**Exhibit 2**

**A picture containing text, screenshot, font, diagram

Description automatically generated**

**Exhibit 3**

**A screenshot of a computer screen

Description automatically generated with low confidence**

**Exhibit 4**

**A screenshot of a computer screen

Description automatically generated with low confidence**

**Exhibit 5**

**A screenshot of a graph

Description automatically generated with medium confidence**

**Exhibit 6**

**A picture containing text, screenshot, plot, diagram

Description automatically generated**

**Exhibit 7**

**A picture containing screenshot, rectangle, plot, line

Description automatically generated**

**Exhibit 8**

A map of the united kingdom

Description automatically generated with low confidence