Analyzing Road Accident Causes in London's District Areas: A Comprehensive Study Using Machine Learning and Python Tools for Proactive Road Safety Management

A Case Study in

Statistical Computing (DS122)

PROPONENTS

Carreon, Ma. Addine Anne T.

Lucas, Sydney Anne V.

Bachelor of Science in Data Science

2nd Year Level – Section A37

MR. JULYVER TOLENTINO

Instructor

2024

Table of Contents **Pages**

INTRODUCTION

The transverse network of roadways creates a growing frequency that serves as a discrete danger which constitutes a serious public safety threat. Road accidents have far-reaching consequences and implications that establish factors of road ramifications towards a district area's environment and its citizens. Every accident collision signifies an upsurge in the delicate equilibrium of everyday life. The frequency of road accidents is considered to be one of the world's major obstacles that lead to deaths and hospitalized civilians, costing nearly $500 billion globally (Ihueze & Onwurah, 2018). Road accidents are not discrete instances but threads in an intricate pattern that flows across a district's setting and its people.

Road traffic congestion is an influential factor in the transportation system, especially in the areas of London that affect high levels of traffic casualties. Central London has frequently been regarded as one of the most overflowing main Western cities that necessitates appropriate counterfactuals with such given London traffic accidents (Green et al., 2016). District areas in London contribute to the severity of road safety and the complexity of the traffic environment. The extensive road network in London imposes substantial limitations for reducing road challenges and improving safety. There are possibilities of sectors that may accord with the district area’s accidents, which need comprehensive and adaptive road management escalation. According to Prieto Curiel & Gonzalez (2018), the occurrence of road accidents is related to a combination of external factors, including impeded sight, a rapid pace by road users, or the condition of the roadways. The condition of district highways has an immense influence on road safety that delves into highlighting pavements, surfaces, inclement weather, and additional dangers that can develop from road wear.

In a district area’s structure, the complicated labyrinth of streets and different traffic patterns require flexible mechanisms that resolve potential road safety. Implications of data-driven adaptive methods such as data can deliver an understanding of road patterns to deal with resolving road safety. In a study by (Lahib et al. (2019), the accident severity in Bangladesh is observed, and the techniques serves as an analytical result through scientific domains grasp the advantage of understanding road accidents that have been a source of global problem in the means of traffic and road guidance. Examining historical data on road patterns, accident cases, and their associated variables may alter measures of movement to improve road safety in district areas.

London grapples with the persistent challenge of ensuring road safety amid a high frequency of road accidents. The current methods for analytically identifying patterns and trends need precision to address the city's dynamic and complex road circumstances. To address the road issue of a particular London district area incorporates methods of approach that deal with diverse analysis and understanding towards the relationship of variables and contributing factors.

A. PROBLEM STATEMENT

One of many threats to public safety is road accidents, in which the district areas of London have perceived varying factors contributing to levels of incidents. Inadequate involvement has compounded road safety, resulting in an elevated frequency of road accidents in London. The complexity and challenging London district area’s road network lends substantially to accident rates that present diverse challenges in aiding road safety.

B. RESEARCH OBJECTIVES

The main purpose of this case study is to identify and analyze contributing factors that lead to the London district area's road accident.

1. To determine and analyze the main contributing factors to road accidents' severity in London, focusing on the relationship of particular variables such as road surface condition, weather condition, and vehicle type.
    a. This aims to comprehend better the factors that lead to road accidents.
2. To assess the geographical distribution of road accidents throughout several district areas in London, leveraging road accidents' severity within London's district regions, incorporating variables including the geospatial information.
    a. This aim is to contribute to identifying high-risk and low-risk areas which demand targeted road safety actions.

C. HYPOTHESIS

1. Null Hypothesis

a. There is no significant association between Road_Surface_Conditions and Accident_Severity.
b. There is no significant association between Weather_Conditions and Accident_Severity.
c. There is no significant association between Vehicle_Type and Accident_Severity.

2. Alternative Hypothesis

a. There is a significant association between Road_Surface_Conditions and Accident_Severity.
b. There is a significant association between Weather_Conditions and Accident_Severity.
c. There is a significant association between Vehicle_Type and Accident_Severity.

D. SIGNIFICANCE OF THE STUDY

Given the frequency of road accidents in London's district area and various factors contributing to road challenges, this study aimed to derive its potential to assess and identify the leading potential causes for road accidents by performing an analysis of the relevant variables. The analysis's dependability will be evaluated and refined to an adaptable conclusion. The study intends to offer an insightful resource for targeting distribution deployment in road safety that focuses on identifying relationships of variables influencing road accidents.

METHODOLOGY

A. Data Collection and Overview

In the initial phase, the research method adopts a systematic approach by obtaining the "Road Accident Casualties Dataset: A Comprehensive Dataset for Traffic Incident Research" from Kaggle, authored by Sheen Nezukokamaado. Following this, the dataset undergoes a thorough cleaning process within *Jupyter Notebook*, addressing missing values, outliers, and ensuring overall data integrity. This foundational step establishes a reliable basis for subsequent analyses.

B. Exploratory Data Analysis (EDA)

The exploration phase commences with the creation of descriptive statistics to provide a comprehensive overview of the dataset. Utilizing univariate analysis, numerical and categorical variables are individually examined, providing insights into their distributions. Visualization techniques, including line graphs, bar graphs, pie charts, line charts, and scatter plots, are employed to explore variable distributions visually, identifying patterns and outliers. Furthermore, bivariate analysis is applied to investigate the relationship of accident severity with each of the three selected variables—road surface condition, weather condition, and vehicle type. This step aims to uncover potential patterns and correlations that contribute to a deeper understanding of the factors influencing accident severity.

In order to investigate intricate relationships between numerous variables, multivariate analysis is presented. By adding location-based perspectives to the investigation, geospatial analysis improves the study. In order to ensure a more nuanced interpretation of the data, outlier detection algorithms aid in the identification and comprehension of unexpected patterns within the dataset. Lastly, cross-tabulation is used to methodically examine how categorical variables interact with one another, exposing possible relationships and dependencies. By offering a comprehensive grasp of the dataset and opening the door for additional in-depth analyses in later research stages, the combined application of these many analytical techniques enhances the exploratory phase.

C. Hypothesis Testing

Using chi-squared statistical tests to examine relationships or independence between categorical variables such vehicle type, weather, and road surface condition is an essential part of the research method. A predefined significance level of 0.05 is incorporated into the formulation of hypotheses to direct the determination of the rejection region. The purpose of this statistical method is to shed light on the complex variables impacting traffic accidents in the district area of London.

D. Bootstrap Sampling

The methodology includes bootstrapping techniques to improve the stability of findings from chi-squared testing. This entails carefully resampling the dataset and then running chi-squared tests again. This strategy is iterative and seeks to clarify the intrinsic instability and unpredictability in the initial chi-squared test results. The robustness of the results is strengthened by additional testing, particularly on the resampled data, which sheds more light on the stability and variability of the initial chi-square test result.

RESULTS AND DISCUSSION

After executing the methodology and adapting the systematic approach, the researchers have obtained a dataset that suffices for the significant findings presented and discussed in this section.

The findings of this case study provide an analysis of the contributing aspects that lead to the London district area's road accidents. The dataset used highlights the dependent variables: road surface condition, weather condition, and vehicle type, with the independent variable being road accidents' severity. These variables create an intrinsic connection that identifies and analyzes the foundation contributing to London's road challenges.

A. EXPLORATORY DATA ANALYSIS

Given the data collection process that includes data overview and data cleaning, the significance of providing visualizations in both numerical and categorical aspects generates a comprehensive representation of data that enables graphical depiction which involves variable relationships and data interpretation.

1. UNIVARIATE ANALYSIS

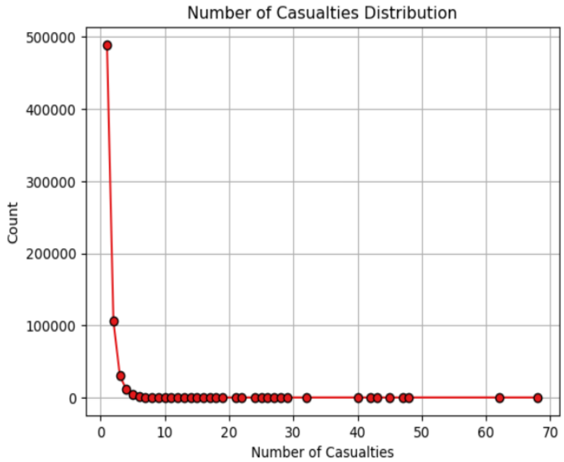
In this section, the statistical method approach is presented and focuses on understanding patterns and attributes that connect with additional variables. The analysis of univariate is to summarize the data's central inclination that creates significant implication.

1.1 Numerical

This section focuses on assessing and exploring a characteristic variable in terms of numbers. This analysis deals with quantitative data that entails components of count and observations.

1.1.1 Number of Casualties Distribution

A visual graphical presentation is used to illustrate the progression of the number of casualties that occur in London's district areas. The x-axis serves as the number of casualties while the y-axis is the denoted count. The graph presents a visible red line that signifies the indication of the accident casualty trajectory. The dotted points are the overall values of every casualty.

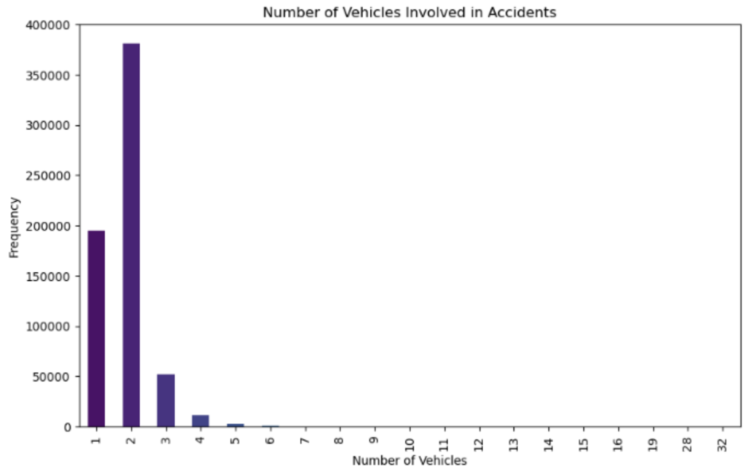


Graph 1. Number of Casualties Distribution

The graph displays the distribution of the dataset's number of casualties. This means that the graph provides a visual narrative of the number of victims that occurred. The distribution counts how many victims got into an accident or how many individuals are involved in the road accident. In this case, the surge of casualties in the range of near zeroes implies that there are more casualties in a road accident with the few to only one individual. The number of casualties that appear to a single individual occurred almost 500,00 times. The downward trend implies that the degree of accidents with many individuals is low compared to few and single people.

1.1.2 Number of Vehicles involved in Accidents.

A visual illustration is used to show the number of vehicles involved in London district area's road accidents. The x-axis represents the number of vehicles, while the y-axis represents its count or frequency. The graph only highlights the number of vehicles involved and not the severity it implies.



Graph 2. Number of Vehicles involved in Accidents.

The graph demonstrates the frequency of vehicles; the result shows that there were more accidents involving two automobiles than any other collection of vehicles. This implies that the most prevalent number of vehicles that are involved in a collision is two. While the three or more vehicles involved in the accidents show a low frequency. This indicates that the number of incidents including ten or more autos is quite minimal. The findings signify the collisions happened during the accidents due to the high frequency of two vehicles.

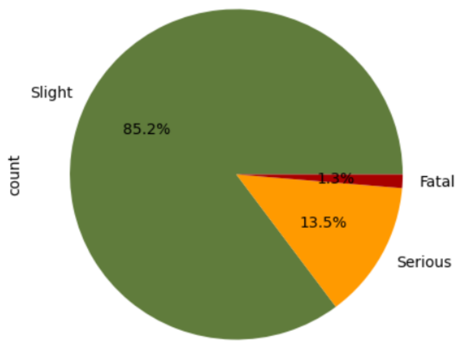
The number of casualties and the number of vehicles shows a high count and frequency in both small-scale ranges. In the number of casualties, one to four casualties are seen relatively high and in the number of vehicles, two automobiles are evidently more involved in road accidents. This implies that since there is a small scale of individuals in the number of casualties involved the tendency of vehicles involved are also fewer.

1.2 Categorical

This section focuses on comprehending the data distribution in a manner of features of categories under certain variables.

1.2.1 Distribution of Accident Severity

A pie chart visual presentation is provided that refers to how different accidents are transmitted. The severity mentioned in this context describes the extent of damage that happened in the road accident.

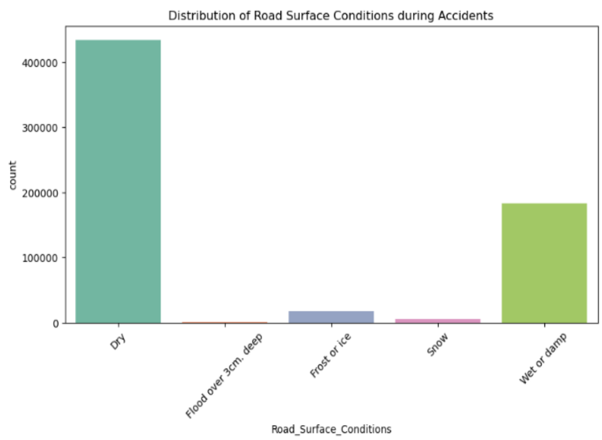


Graph 3. Distribution of Accident Severity

Based on the data findings, it is evident that the slight category is immensely high with 85.2%, followed serious with 13.5%, and coming last is the fatal with 1.3%. This implies that the severity of a road accident is mostly slight or medium to average damage. The damage can be obstructive. While there is significant value for serious and fatal. The finding indicates that the damage happening in serious severity is substantial but not as much with fatal injuries. The data derives that the road accidents happened are more slight damages that still can be apprehended to preventive measures.

1.2.2 Road Surface Conditions Distributions

The bar graph illustrates the level of road surface distribution. This representation understands the categories of road surfaces and their level of count.

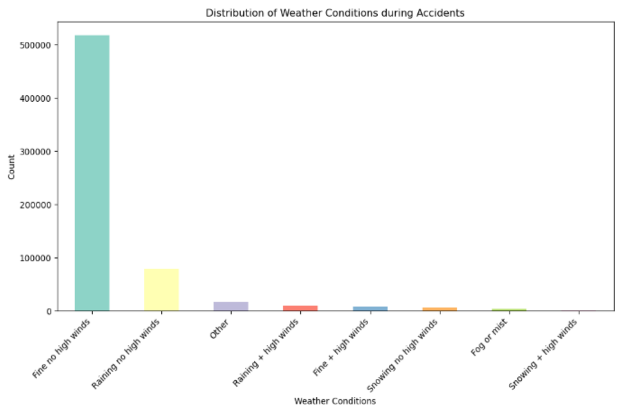


Graph 4. Road Surface Conditions Distributions

Based on the finding, it is evident that the dry road condition has the highest level of count, resulting in more than 400,000 incidents. Following dry is the wet or damp road surface, and almost the fewest or lowest level is the flood, frost and snow. The implication is that road accidents usually happen during the dry or sunny days and when the road is damp and not frequently when there is a storm. This suggests that there are more existing cars on the road during dry and wet season than during snow season, this may be due to the fact that people are avoiding stepping outside and bringing their vehicles in snow.

1.2.3 Weather Conditions Distribution

The bar graph demonstrates the distribution of weather conditions that happened in a road accident. The presentation provides a categorial analysis on which weather conditions occurred the most in road challenges.

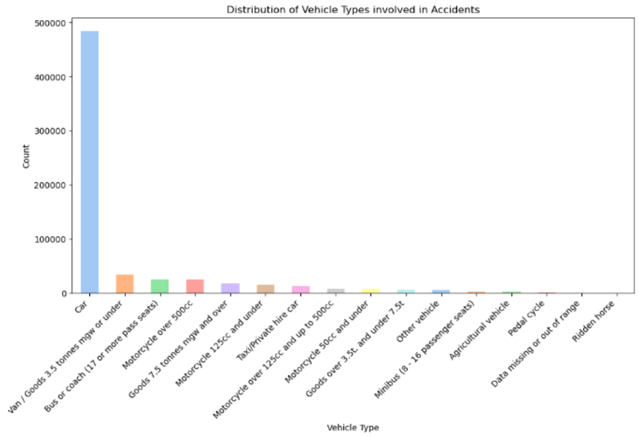


Graph 5. Weather Conditions Distribution

In this bar graph, the data of fine weather conditions, specifically, fine with no high winds, underscores the most common atmosphere exposure during road accidents with over 400,000 incidents. The following weather is the rain with no high winds that caused less than 100,000 accidents. This indicates a well significant level with the road surface conditions, in which dry road surface conditions is the highest and the highest weather condition is fine with no high winds. This means that the road accidents happen during fine weather with no high winds and a dry road surface.

1.2.4 Vehicle Type Distribution

The bar graph demonstrates the distribution of the vehicles that made contact in the road accidents. The data indicates which vehicle type has occurred more than any other car category.



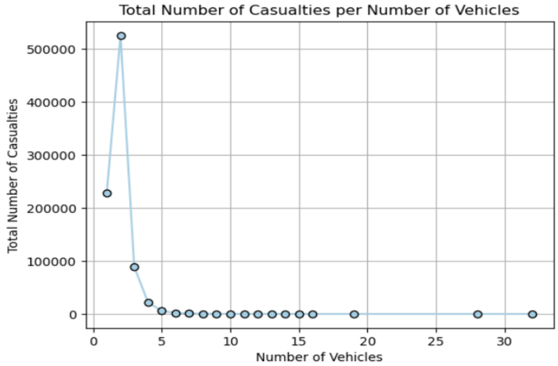
Graph 6. Vehicle Type Distribution

Based on the findings, it is highly evident that the vehicle type car has emerged with the highest count, regarding with the most common car type to be engaged in a road accident. Following the vehicle type car are vans. This implies that the most used vehicle type that has made contact with road accidents are cars.

2. BIVARIATE ANALYSIS

Given that numerical and categorical values are distributed with interpretations, the relationship and connections between variables are then identified and elaborated. The provided analysis will determine the level and nature of association between the variables.

2.1 Number of Casualties per Number of Vehicle

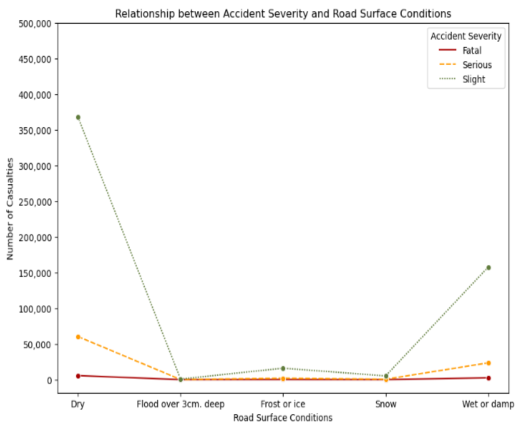


Graph 7. Number of Casualties per Number of Vehicle

The graph exhibits the totality of the number of casualties per number of vehicles. The x-axis represents the number of vehicles, and the y-axis represents the total number of casualties. Based on the data presentation, the graph explores a correlation. This means that the ranges of fewer vehicles involved in a road accident obtain a high number of casualties. This suggests that incidents involving a smaller number of cars may have an excessive influence on casualties. Additionally, the fewer number of vehicles involved in road accidents may acquire a high level of speed that causes higher casualties than with congested vehicles on the road.

2.2 Relationship between Accident Severity and Road Surface Conditions

In these data findings, a relationship of two variables is examined, which are the accident severity and road surface conditions. The x-axis symbolizes the road surface conditions, while the y-axis shows the count of casualties. Accordingly, the three lines exhibited in the graph represent the accident severity types.

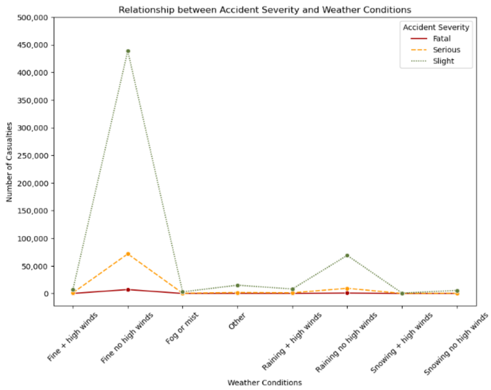


Graph 8. Relationship between Accident Severity and Road Surface Conditions

In the road surface condition data visualization, it was mentioned that dry roads are most likely to occur in road accidents. Similarly, the category slight in accident severity appears to be the highest type. Hence, this implies that the acquired number of casualties during a dry road showcases a high count with a slight accident severity. The casualties appearing during dry roads indicate a low fatal severity and a less average severity but a high slight severity. Meanwhile, following the dry road is the wet or damp road that emphasizes the same implication with a high slight severity and a minimal serious and fatal severity. The other road conditions have low casualties in the three types of severity. The graph shows that during the dry and wet roads the casualties are evident with mostly slight severity than with snow roads.

2.3 Relationship between Accident Severity and Weather Conditions

In this data finding, the relationship of accident severity with weather conditions. The x-axis symbolizes the weather conditions, while the y-axis shows the count of casualties. Accordingly, the three lines exhibited in the graph represent the accident severity types.

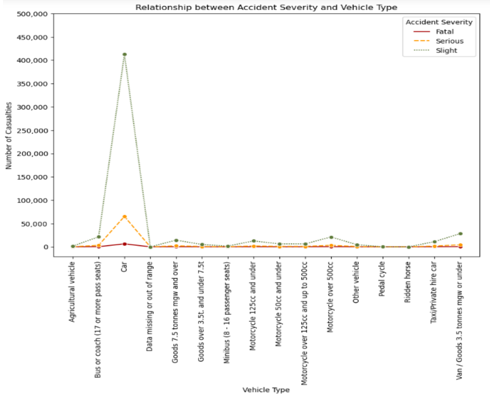


Graph 9. Relationship between Accident Severity and Weather Conditions

In the illustration for weather conditions, the fine with no high winds are evidently high among any other weather conditions. Similarly to this presented graph, the highest condition with a high number of casualties is the fine weather. This implies that the high number of casualties during fine weather with no high winds acquires a slight severity with less average serious damage and little to no fatal accidents. In the weather condition graph, raining with no high winds also appears to be the second highest weather type that has a similar correlation with a high slight severity of casualty and little to less serious and fatal accidents. This graph implies that during fine weather the casualties are beyond average but have a slight severity. Additionally, the graph showed that during rainy weather, the casualties are evident but not too many with also a slight severity.

2.4 Relationship between Accident Severity and Vehicle Type

In this data finding, a relationship of two variables is examined, which are the accident severity and vehicle type. The x-axis symbolizes the vehicle type, while the y-axis shows the count of casualties. Accordingly, the three lines exhibited in the graph represent the accident severity types.



Graph 10. Relationship between Accident Severity and Vehicle Type

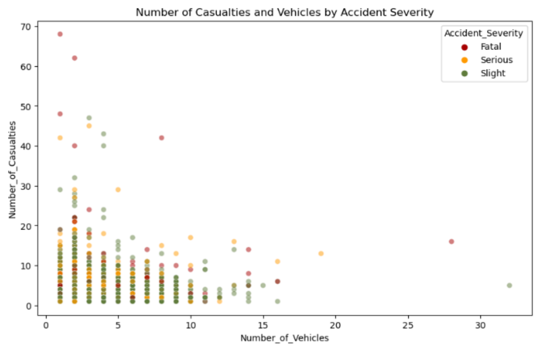
In the illustration of vehicle type, the highest count corresponds to cars then following are vans. This graph shows that the vehicle type, car, has the highest count of casualties. The car type corresponds to obtaining a highest slight severity with low serious severity and few fatal severities. The rest of the cars appear to have almost the same distribution of low number of casualties. However, it is evident that the degree of severity of every vehicle type has a slight damage. This graph shows that the commonly employed vehicle type by the individuals are cars with it being the highest number of casualties including a slight severity.

3. MULTIVARIATE ANALYSIS

In this section, the analysis of more than two variables is examined. This analysis focuses on simulating the relationship of a quantity of variables to understand the factors of one another.

3.1 Number of Casualties and Vehicles by Accident Severity

In this data finding, a relationship of variables is examined, which are the accident severity, number of casualties and vehicles. The x-axis symbolizes the number of vehicles, while the y-axis shows the count of casualties. Accordingly, the dotted points exhibited in the graph represent the accident severity types.



Graph 11. Number of Casualties and Vehicles by Accident Severity

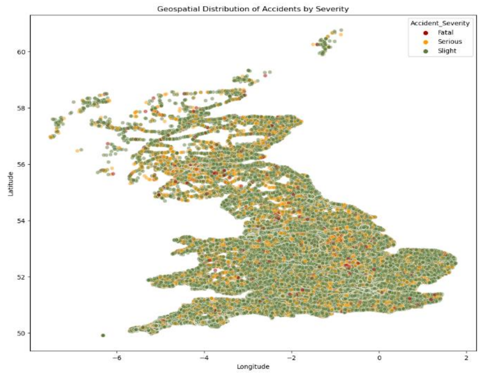
This graphing illustration underlines the relationship of how the number of vehicle types correlates to the severity and number of casualties in a road accident. Given that in the number of vehicles involved in the accident graph, two has the highest count corresponding to the number of casualties. This means that the graph shows that with two vehicles being involved the number of casualties has increased but with a slight severity of road accident damage. While, with a higher number of vehicles involved the lesser casualties occurred with almost no sight accident severity.

4. GEOSPATIAL ANALYSIS

This case study aims to identify high-risk and low-risk areas in the district areas of London. Hence, this section provides an analysis on visualizing data that has geographic components to locate coordinated road safety percussion.

4.1 Geographical Distribution of Accidents

In this data finding, the utilization of coordinates of certain area districts from the dataset is acquired. The x-axis symbolizes longitude, while the y-axis shows latitude. Accordingly, the dotted points exhibited in the graph represent the accident severity types.



Graph 12. Geographical Distribution of Accidents

With the visualization findings, the graph indicates that most of the district areas in London obtained a result of almost slight accident severity. This correlates to the pie chart distribution of accident severity, where the slight severity is the most abundant among the three categories. This graph shows that the majority of district areas have a low or slight level of accident severity than serious and fatal damage. The results shown in the graph and associated pie chart reveal a comparable trend of almost minor accident severity throughout London district regions.

B. HYPOTHESIS TESTING

In this section, this case study employs hypothesis testing along with the chi-squared statistical tests to identify the independence between categorical variables. In acquiring a hypothesis testing, the needed null and alternative hypothesis in determining the relationship of the variables is presented.

1. ROAD SURFACE CONDITIONS

After executing a cross tabulation, the numerical data are utilized to compute the p value. Results showed that the p-value for road surface conditions is 8.8421e-77. Since the p-value is less than the alpha, it will reject the null hypothesis. Therefore, the findings occurred that in road surface conditions there is a significant association between Road_Surface_Conditions and Accident_Severity.

Rejecting the null hypothesis implies that the observed distribution accidents in various categorical road surface conditions are inconceivable to have happened by chance alone. In concrete terms, it implies that road surface conditions have a major effect on the severity of accidents. One possible explanation is that road surface conditions, such as dry or wet roads, have association with the increased risk of slight, serious, and fatal accidents. This might be related to decreased momentum and longer braking distances. The expectancy of road surface condition is highly inclined with the potential severity of road accidents. In the Relationship between Accident Severity and Road Surface Conditions graphing illustration, it highlights that dry and wet roads have dealt with an emerging number of casualties among other road surface conditions with slight accident severity.

2. WEATHER CONDITIONS

After executing a cross tabulation, the numerical data are utilized to compute the p value. Results showed that the p-value for road surface conditions is 4.7677e-13. Since the p-value is less than the alpha, it will reject the null hypothesis. Therefore, the findings occurred that in road surface conditions there is a significant association between Weather_Conditions and Accident_Severity.

Rejecting the null hypothesis implies that the observed distribution accidents in different categorical types of weather conditions are unlikely to have happened by chance alone. It highly indicates that weather conditions serve an important influence in determining the severity of accidents. The hypothesis implies that increased risks of accidents depending on weather conditions is evident that may be the cause by varying factors such as changing driving dynamics or road visibility. In the Relationship between Accident Severity and Weather Conditions, the evident weather type is a fine one with no high winds as the highest contributing factor to accidents, which implies a sunny or cloudy atmosphere have the most influence on road accidents with a slight or low severity. This implies that even in good weather conditions, road accidents still emerge with slight severity.

3. VEHICLE TYPE

After executing a cross tabulation, the numerical data are utilized to compute the p value. Results showed that the p-value for road surface conditions is 0.4690. Since the p-value is greater than the alpha, it will fail to reject the null hypothesis. Therefore, the findings occurred that in road surface conditions there is no significant association between Vehicle_Type and Accident_Severity.

Failing to reject the null hypothesis, the observed distribution accidents in various types of vehicles involved are relatively likely to have transpired by chance. Similarly with the findings, the data provided fails to provide sufficient evidence that will support the notion between vehicle types involved and accident severity. The type of vehicle used by an individual does not have a significant factor that contributes to the severity of accidents. This entails that the vehicle type that is being used by any individual does not fully contribute to the level of accident that will happen.

C. BOOTSTRAPPING

In this section, the executed bootstrapping is utilized to estimate the distribution to generate samples from the provided data. This identifies the potential uncertainty occurring in the chi-square test statistic.

1. ROAD SURFACE CONDITIONS

The bootstrap sample 5 results indicate the outcome of the data findings after the execution of bootstrap. Along with the bootstrap sample is a generated heat map that is the same basis with the data result found in the table diagram with a level of accident severity.

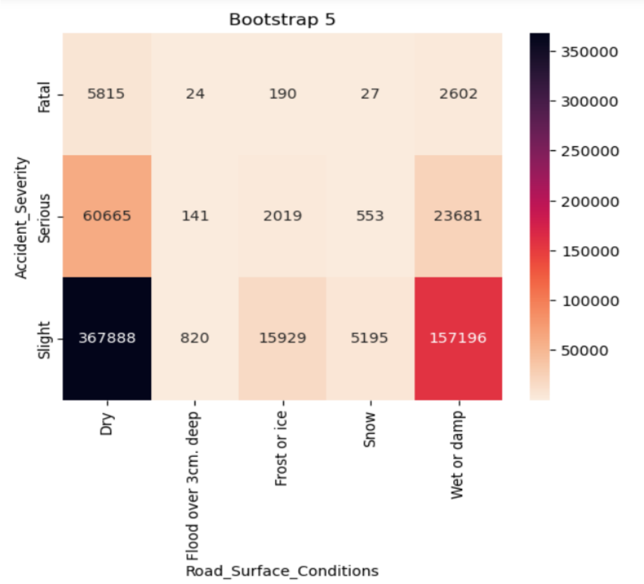
1.1 Table Diagram Road Surface Conditions Bootstrap Sample

Bootstrap Sample 5 Results:

Accident_Severity	Dry	Flood over 3cm. deep	Frost or ice	Snow	Wet or damp
Fatal	5815	24	190	27	2602
Serious	60665	141	2019	553	23681
Slight	367888	820	15929	5195	157196

Table 1. Road Surface Conditions Bootstrap Sample

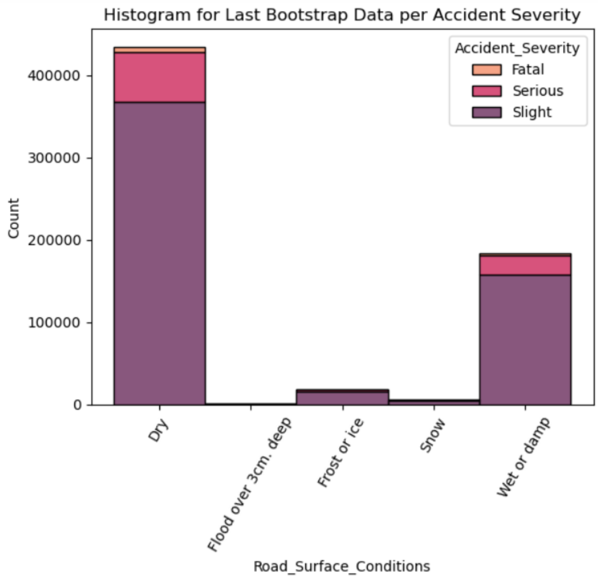
1.2 Heat Map Road Surface Conditions Bootstrap Sample



Graph 13. Road Surface Conditions Bootstrap Sample

The given table diagram along with its heat map presentation showcases the categorical road surface conditions and its correlation to the accident severity. The findings indicate that the quantity of accidents on dry roads increased more than double on wet or damp roads resulting in an implication that dry roads are more risky than wet roads. Simultaneously, dry roads also account for nearly twice as many fatal, serious, and slight accident severity than wet or damp roads. While the other road surfaces emphasize low data values acquiring small quantity findings. Observing the heatmap, the numerical findings are evidently high on both dry and wet roads. However, dry roads exhibit more road accidents and accident severity than wet roads.

1.3 Histogram of Road Surface Conditions Bootstrap Data



Graph 14. Road Surface Conditions Bootstrap Data

The findings of this histogram correlate effectively with the heat map data results. After the execution of bootstrapping, it implies that there is a higher count in the dry road surface conditions than any other category. Additionally, the dry road surface exhibits a high rating for slight accident severity. In a study of Ismael and Razzaq (2017), their study mentioned that between 2011 to 2015, there were more accidents on dry roads than on rainy days in Manchester-UK. Their results show that there are more collisions happening in dry road weather. Which significantly correlates to this study findings, where dry road condition has more accident occurred than in wet roads.

1.4 Road Surface Conditions Bootstrap Decision

After the bootstrap execution, the numerical data are utilized to compute the chi-square and the p-value. Results showed that the chi-squared is 366.4403 while the p-value for road surface conditions is 2.7950e-74. Therefore, the findings occurred to reject the null hypothesis that there is significant association between Road_Surface_Conditions and Accident_Severity. The association between road surface conditions and accident severity shows significant influence that contributes to the level of potential road accidents.

2. WEATHER CONDITIONS

The bootstrap sample 5 results indicate the outcome of the data findings after the execution of bootstrap. Along with the bootstrap sample is a generated heat map that is the same basis with the data result found in the table diagram with a level of accident severity.

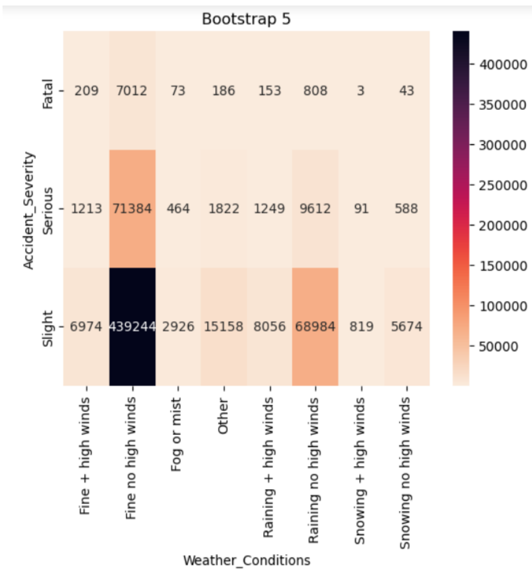
2.1 Table Diagram Weather Conditions Bootstrap Sample

Bootstrap Sample 5 Results:

Accident_Severity	Fine + high winds	Fine no high winds	Fog or mist	Other	Raining + high winds	Raining no high winds	Snowing + high winds	Snowing no high winds
Fatal	209	7012	73	186	153	808	3	43
Serious	1213	71384	464	1822	1249	9612	91	588
Slight	6974	439244	2926	15158	8056	68984	819	5674

Table 2. Weather Conditions Bootstrap Sample

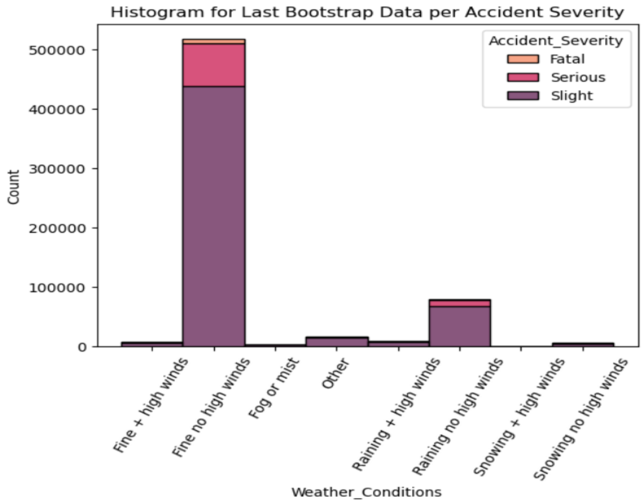
2.2 Heat Map Weather Conditions Bootstrap Sample



Graph 15. Weather Conditions Bootstrap Sample

The given table diagram along with its heat map presentation showcases the categorical weather conditions and its correlation to the accident severity. The findings indicate that the quantity of accidents on fine no high winds are twice the count for raining with no high winds. While the other weather conditions appear to have a fewer quantity range. The fine weather with no high winds accumulates an intense level of slight severity. This implies that during fine weather, the accidents are evident but with slight severity.

2.3 Histogram of Weather Conditions Bootstrap Data



Graph 16. Weather Conditions Bootstrap Data

The findings of this histogram correlate effectively with the heat map data results. After the execution of bootstrapping, it implies that there is a higher count on fine weather with no high winds than any other category. Additionally, the fine weather condition exhibits a high rating for slight accident severity. Contrasting to that is the study of Malin et al. (2019) contrasts these findings. In their study, the poor weather conditions such as ice and slippery are extremely hazardous conditions that raise corresponding accident risk. However, in our data findings, snow weather is relatively low in road accidents and its severity. This implies that there may be a complex factor that contributes to weather conditions. For instance, there could be drivers that are more experienced in navigating snow conditions that leads to a lower chance of accidents.

2.4 Weather Conditions Bootstrap Decision

After the bootstrap execution, the numerical data are utilized to compute the chi-square and the p-value. Results showed that the chi-squared is 628.0314 while the p-value for road surface conditions is 5.7206e-125. Therefore, the findings occurred to reject the null hypothesis that there is significant association between Weather_Conditions and Accident_Severity. The association between weather conditions and accident severity shows significant influence that contributes to the level of potential road accidents.

3. VEHICLE TYPE

The bootstrap sample 5 results indicate the outcome of the data findings after the execution of bootstrap. Along with the bootstrap sample is a generated heat map that is the same basis with the data result found in the table diagram with a level of accident severity.

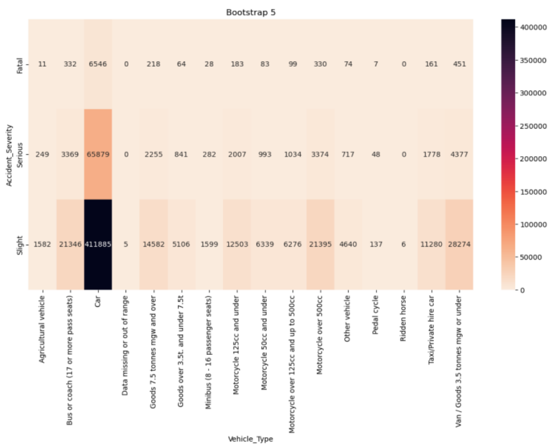
3.1 Table Diagram Vehicle Type Bootstrap Sample

Bootstrap Sample 5 Results:

Accident_Severity	Agricultural vehicle	Bus or coach (17 or more pass seats)	Car	Data missing or out of range	Goods 7.5 tonnes mgw and over	Goods over 3.5t. and under 7.5t	Minibus (8 - 16 passenger seats)	Motorcycle 125cc and under	Motorcycle 50cc and under	Motorcycle over 125cc and up to 500cc	Motorcycle over 500cc	Other vehicle	Pedal cycle	Ridden horse	Taxi/Private hire car	Van / Goods 3.5 tonnes mgw or under
Fatal	11	332	6546	0	218	64	28	183	83	99	330	74	7	0	161	451
Serious	249	3369	65879	0	2255	841	282	2007	993	1034	3374	717	48	0	1778	4377
Slight	1582	21346	411885	5	14582	5106	1599	12503	6339	6276	21395	4640	137	6	11280	28274

Table 3. Vehicle Type Bootstrap Sample

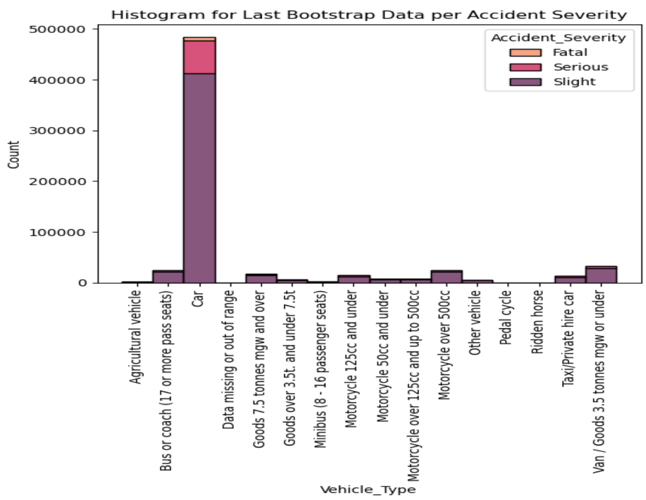
3.2 Heat Map Vehicle Type Bootstrap Sample



Graph 17. Vehicle Type Bootstrap Sample

The given table diagram along with its heat map presentation showcases the categorical vehicle types and its correlation to the accident severity. The findings indicate that the quantity of accidents on car types are immensely the highest among the other vehicle types. This suggests that the utilization of cars is more engaging to road accidents with a level of slight severity.

3.3 Histogram of Vehicle Type Bootstrap Data



Graph 18. Vehicle Type Bootstrap Data

The findings of this histogram correlate effectively with the heat map data results. After the execution of bootstrapping, it implies that there is a higher count on the vehicle type of car than any other category. Additionally, the car category exhibits a high rating for slight accident severity. With a similar study finding of Chand et al. (2021), where they assert that two wheelers, heavyweight automobiles, and passenger cars present an elevated danger of road accidents than cycling, using of pedestrians, and utilizing public transportation systems. This implies that vehicle type cars have significant correlations towards potential road damages.

3.4 Vehicle Type Bootstrap Decision

After the bootstrap execution, the numerical data are utilized to compute the chi-square and the p-value. Results showed that the chi-squared is 61.8632 while the p-value for road surface conditions is 0.0005.

Therefore, the findings occurred to reject the null hypothesis that there is significant association between Vehicle_Type and Accident_Severity. The association between the vehicle type and accident severity shows significant influence that contributes to the level of potential road accidents. This finding implies that the employment of an individual in using a vehicle type has a correlation towards potential road accidents with a slight severity. Comparing the bootstrap and the hypothesis outcomes implies that with multiple data findings, it is evidently sufficient that there is a significance in what type of vehicle is being used. The relationship displayed by the vehicle type and accident severity means that the qualities or features of a vehicle greatly influence the likelihood of road accidents and the level of accident severity, which specifically in this case, car has a significant contribution to a slight accident severity.

CONCLUSION AND RECOMMENTATION

CONCLUSION

The case study meticulously explored the intricate web of factors influencing road accidents in the London district, focusing on road surface conditions, weather patterns, and vehicle types.

1. Initial analysis using traditional p-values unveiled significant associations between accident severity and road surface conditions, as well as weather patterns. However, the relationship with vehicle types exhibited a more nuanced connection.
2. The pivotal inclusion of bootstrapping in the analysis played a transformative role, refining and affirming the findings' robustness. It underscored the substantial impact of both road surface conditions and weather patterns on accident severity, emphasizing the necessity of advanced statistical techniques in comprehensive assessments.
3. Rigorous testing of null hypotheses revealed a rejection in the case of road surface conditions and weather patterns, indicating a significant association with accident severity. Conversely, the null hypothesis related to vehicle type and accident severity was retained, suggesting a less conclusive relationship.
4. The alternative hypotheses, postulating significant associations between these variables and accident severity, found support in the study's findings.
5. Dry roads and fine weather without high winds emerged as high-risk contributors, predominantly resulting in accidents with slight severity.

6. The prevalence of cars in accidents underscored their impact on casualty counts, highlighting a noteworthy association between vehicle types and accident severity.
7. Geospatial analysis added a spatial dimension, offering targeted insights for effective road safety measures in specific areas.

RECOMMENDATION:

1. Utilize advanced statistical techniques, such as bootstrapping, in future comprehensive assessments to ensure robust findings.
2. Implement a multi-faceted approach, including infrastructure improvements, tailored awareness campaigns for specific weather conditions, enhanced vehicle safety measures, continuous road condition monitoring, stakeholder collaboration, and further research.
3. Recognize and target the distinct and significant effects of road surface conditions and weather patterns on accident severity in interventions.
4. Specifically address dry roads and fine weather without high winds as high-risk contributors to accidents with slight severity.
5. Tailor interventions to the prevalence of cars in accidents, highlighting their impact on casualty counts and emphasizing the association between vehicle types and accident severity.
6. Leverage geospatial analysis for targeted insights and effective road safety measures in specific areas.
7. Acknowledge the study's contribution to the field of road safety by providing a comprehensive understanding of the factors influencing accident severity in the London district area.

REFERENCES

Ihueze, C. C., & Onwurah, U. O. (2018). Road traffic accidents prediction modelling: An analysis of Anambra State, Nigeria. Accident Analysis & Prevention, 112, 21–29. doi:10.1016/j.aap.2017.12.016

Green, C. P., Heywood, J. S., & Navarro, M. (2016). Traffic accidents and the London congestion charge. Journal of Public Economics, 133, 11–22. doi:10.1016/j.jpubeco.2015.10.005

Prieto Curiel, R., González Ramírez, H., & Bishop, S. R. (2018). A novel rare event approach to measure the randomness and concentration of road accidents. PLOS ONE, 13(8), e0201890. doi:10.1371/journal.pone.0201890

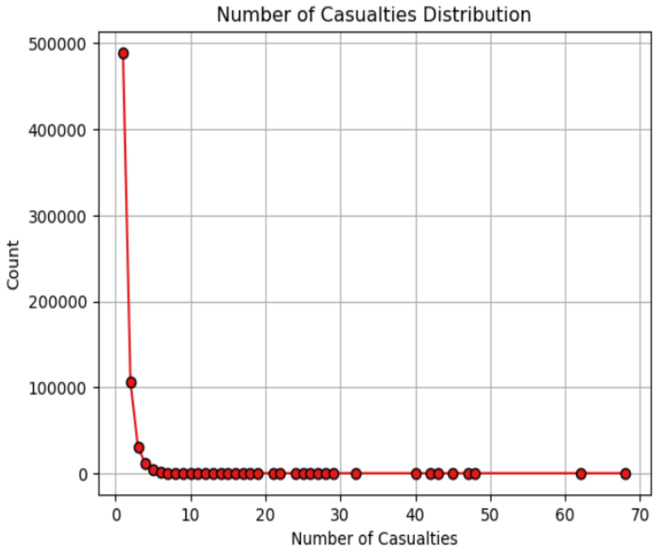
Labib, M. F., Rifat, A. S., Hossain, M. M., Das, A. K., & Nawrine, F. (2019). Road Accident Analysis and Prediction of Accident Severity by Using Machine Learning in Bangladesh. 2019 7th International Conference on Smart Computing & Communications (ICSCC). doi:10.1109/icscc.2019.8843640

Ismael, K. S., & Razzaq, N. A. (2017). Traffic Accidents Analysis on Dry and Wet Road Bends Surfaces in Greater Manchester-UK. Kurdistan Journal of Applied Research, 2(3), 284-291. https://doi.org/10.24017/science.2017.3.51

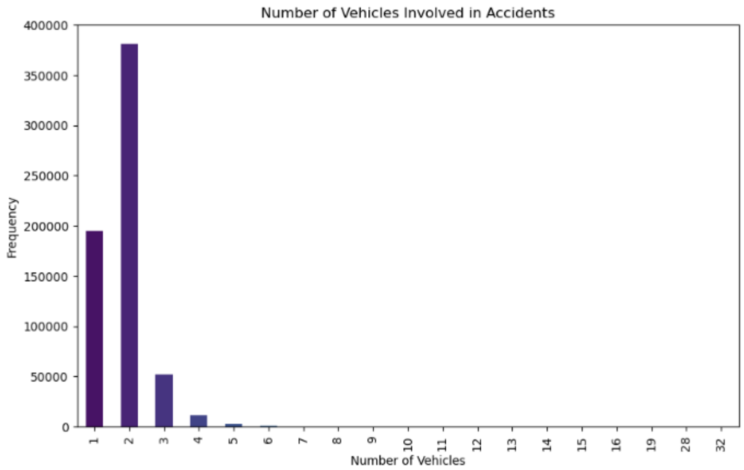
Malin, F., Norros, I., & Innamaa, S. (2019). Accident risk of road and weather conditions on different road types. Accident Analysis & Prevention, 122, 181–188. doi:10.1016/j.aap.2018.10.014

Chand, A., Jayesh, S., & Bhasi, A. B. (2021). Road traffic accidents: An overview of data sources, analysis techniques and contributing factors. Materials Today: Proceedings. doi:10.1016/j.matpr.2021.05.415

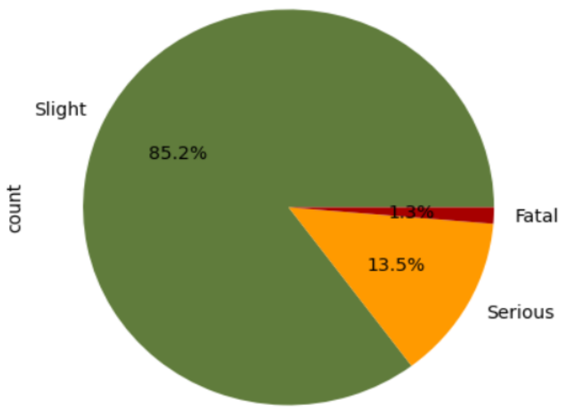
APPENDIX A: Graphs



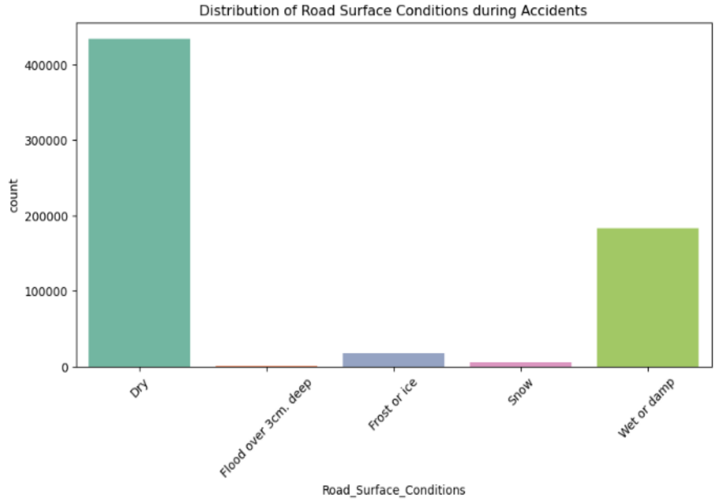
Graph 1. Number of Casualties Distribution



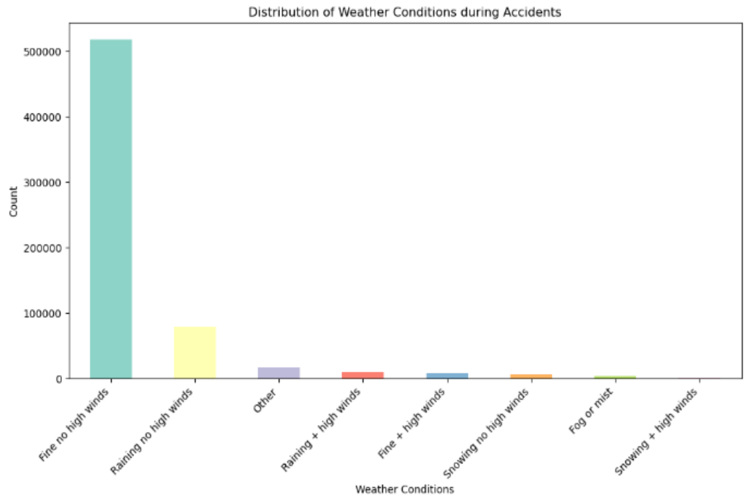
Graph 2. Number of Vehicles involved in Accidents.



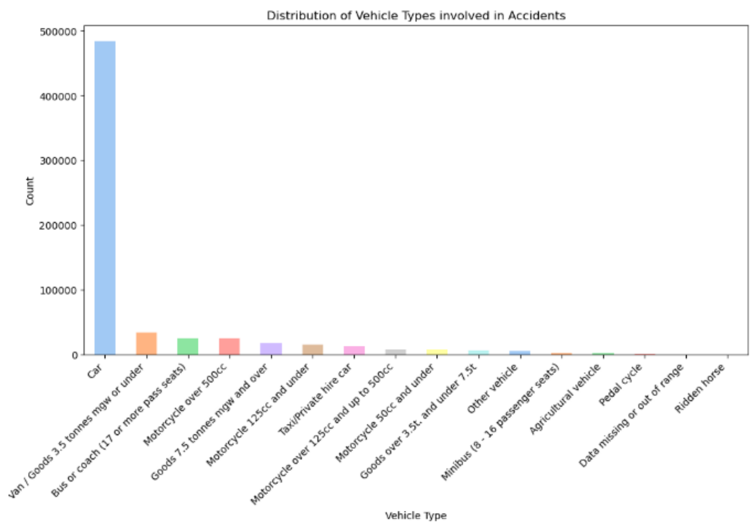
Graph 3. Distribution of Accident Severity



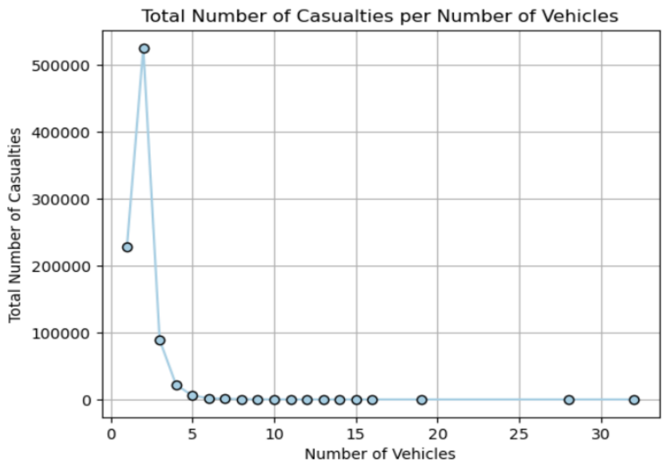
Graph 4. Road Surface Conditions Distributions



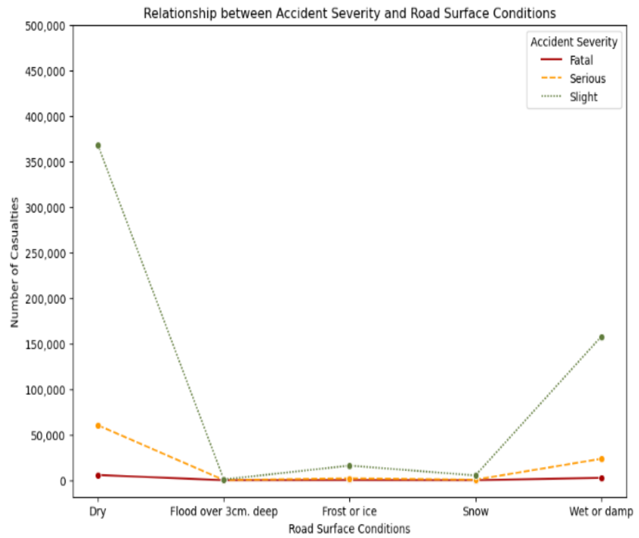
Graph 5. Weather Conditions Distribution



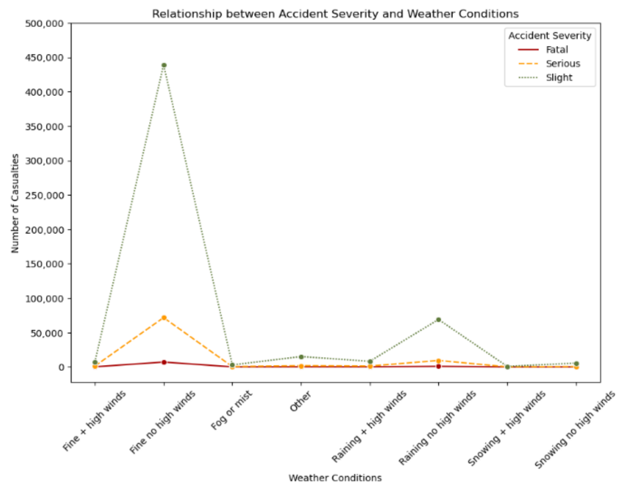
Graph 6. Vehicle Type Distribution



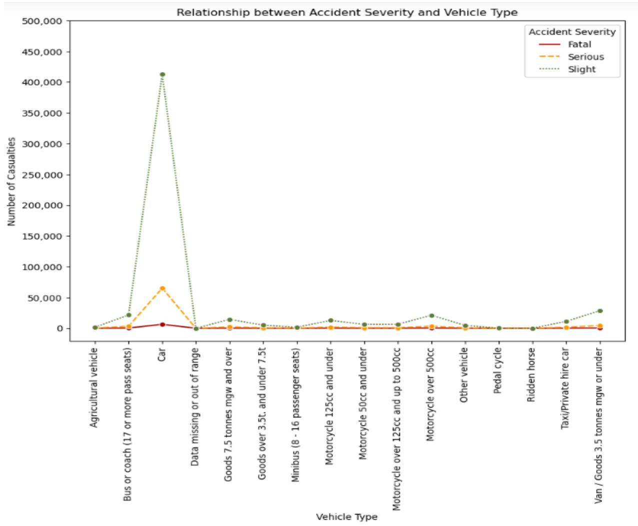
Graph 7. Number of Casualties per Number of Vehicle



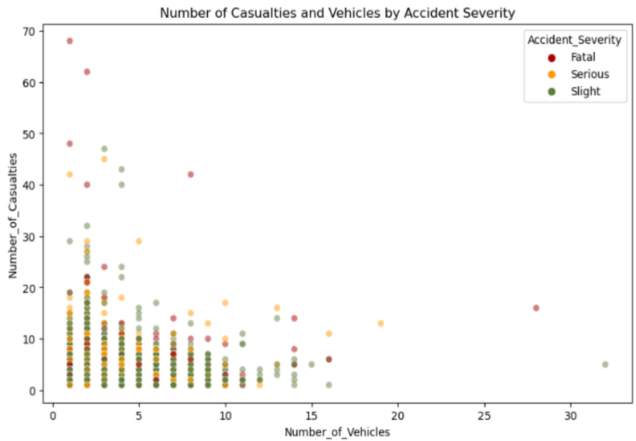
Graph 8. Relationship between Accident Severity and Road Surface Conditions



Graph 9. Relationship between Accident Severity and Weather Conditions

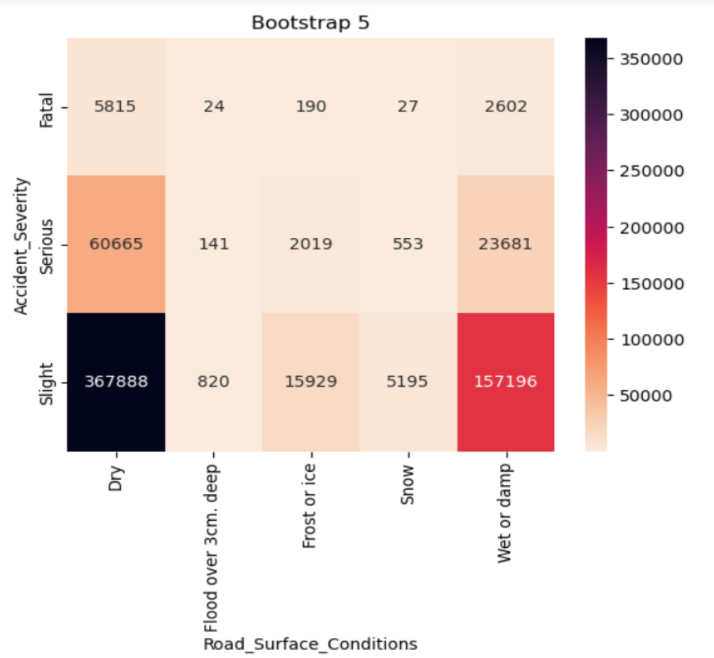


Graph 10. Relationship between Accident Severity and Vehicle Type

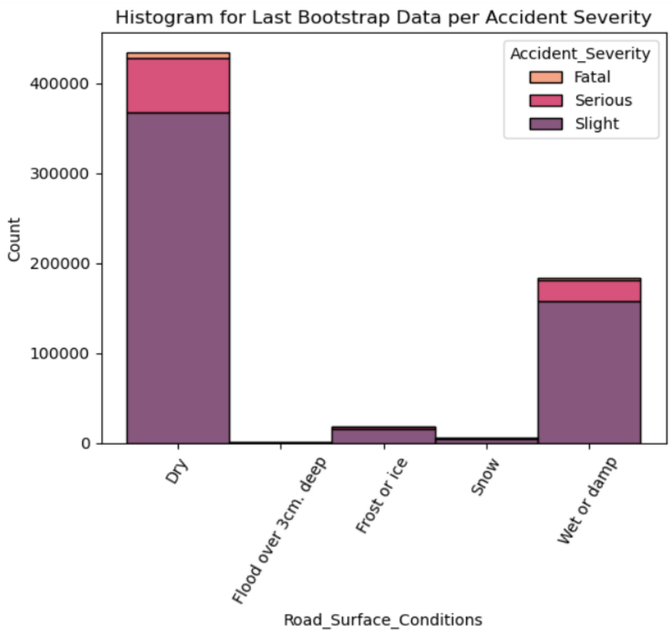


Graph 11. Number of Casualties and Vehicles by Accident Severity

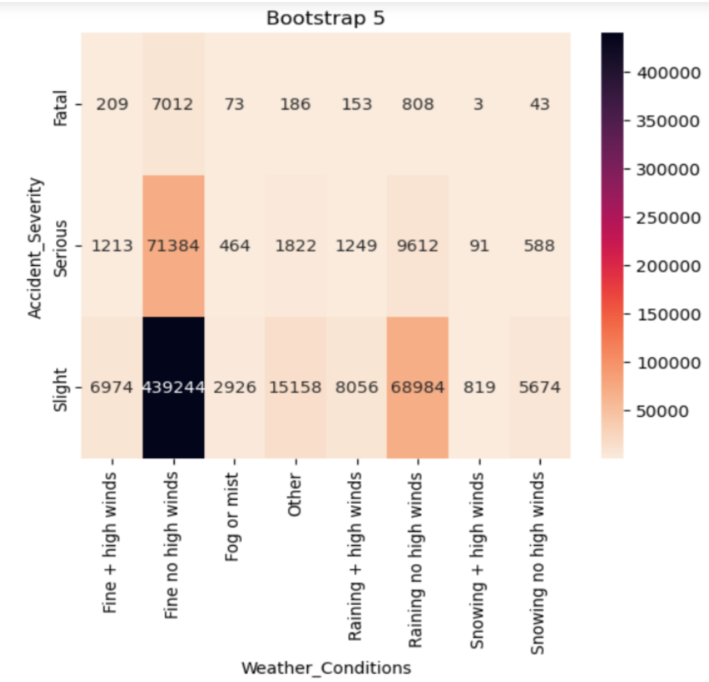
Graph 12. Geographical Distribution of Accidents



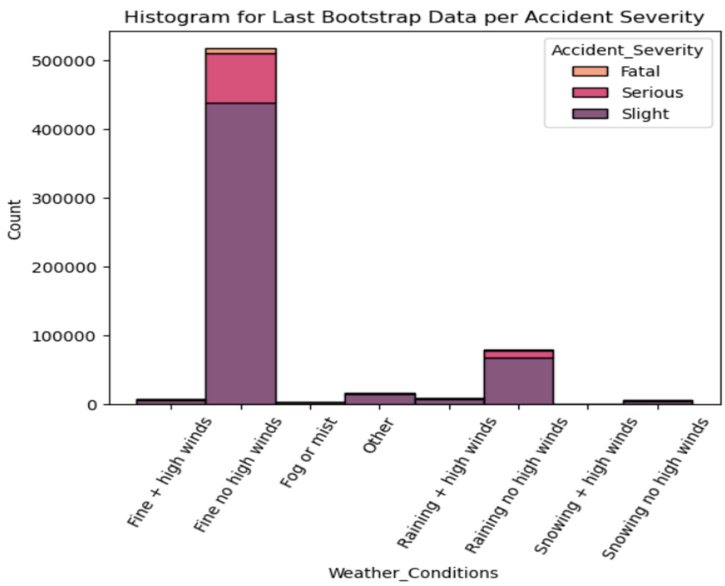
Graph 13. Road Surface Conditions Bootstrap Sample



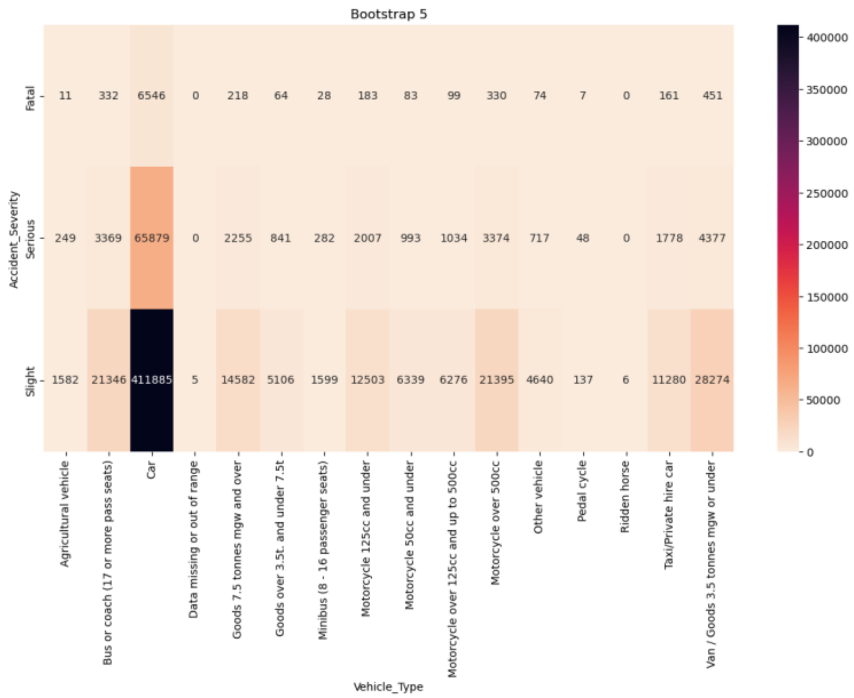
Graph 14. Road Surface Conditions Bootstrap Data



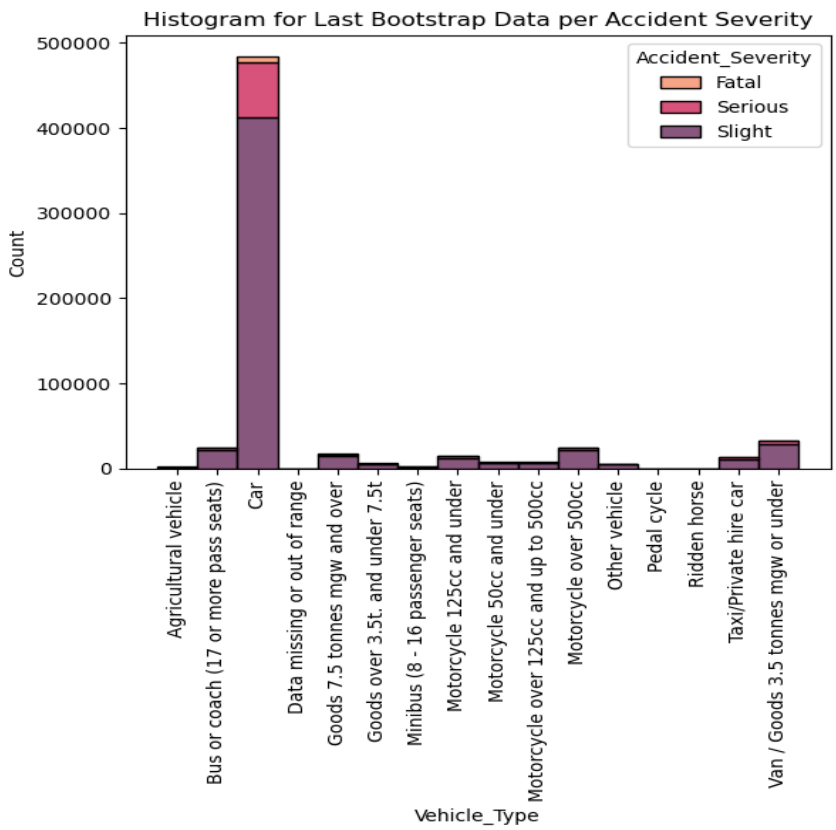
Graph 15. Weather Conditions Bootstrap Sample



Graph 16. Weather Conditions Bootstrap Data



Graph 17. Vehicle Type Bootstrap Sample



Graph 18. Vehicle Type Bootstrap Data

APPENDIX B: Table

Bootstrap Sample 5 Results:

Accident_Severity	Dry	Flood over 3cm. deep	Frost or ice	Snow	Wet or damp
Fatal	5815	24	190	27	2602
Serious	60665	141	2019	553	23681
Slight	367888	820	15929	5195	157196

Table 1. Road Surface Conditions Bootstrap Sample

Bootstrap Sample 5 Results:

Accident_Severity	Fine + high winds	Fine no high winds	Fog or mist	Other	Raining + high winds	Raining no high winds	Snowing + high winds	Snowing no high winds
Fatal	209	7012	73	186	153	808	3	43
Serious	1213	71384	464	1822	1249	9612	91	588
Slight	6974	439244	2926	15158	8056	68984	819	5674

Table 2. Weather Conditions Bootstrap Sample

Bootstrap Sample 5 Results:

Accident_Severity	Agricultural vehicle	Bus or coach (17 or more pass seats)	Car	Data missing or out of range	Goods 7.5 tonnes mgw and over	Goods over 3.5t. and under 7.5t	Minibus (8 - 16 passenger seats)	Motorcycle 125cc and under
Fatal	11	332	6546	0	218	64	28	183
Serious	249	3369	65879	0	2255	841	282	2007
Slight	1582	21346	411885	5	14582	5106	1599	12503

Accident_Severity	Motorcycle 50cc and under	Motorcycle over 125cc and up to 500cc	Motorcycle over 500cc	Other vehicle	Pedal cycle	Ridden horse	Taxi/Private hire car	Van / Goods 3.5 tonnes mgw or under
Fatal	83	99	330	74	7	0	161	451
Serious	993	1034	3374	717	48	0	1778	4377
Slight	6339	6276	21395	4640	137	6	11280	28274

Table 3. Vehicle Type Bootstrap Sample