Division of Computing Science and Mathematics
Faculty of Natural Sciences
University of Stirling

Analysis of traffic accidents in Scotland

3301407

Dissertation submitted in partial fulfilment for the degree of
Master of Science in Big Data

September 2024

Abstract

Traffic accidents in Scotland present a substantial peril to public safety, leading to multiple deaths and enormous socio-economic consequences. This project aims to comprehensively analyse the patterns and determinants that impact traffic incidents in Scotland to suggest future strategies to mitigate these occurrences. The study utilises a comprehensive approach, using massive datasets from Kaggle to examine road and aviation traffic occurrences. The data underwent meticulous cleansing and preparation, including eliminating duplicates, correcting errors, and standardising the format to guarantee correctness. Subsequently, a range of data manipulation, visualisation, and machine learning approaches were employed for the analysis.

The findings of this study are not just informative, they are of paramount importance for understanding and addressing the issue of traffic accidents in Scotland. The study uncovers significant patterns, such as a noticeable increase in accidents during the fall and winter seasons, particularly in adverse weather conditions. Urban areas like Glasgow and Edinburgh are identified as high-risk zones, primarily due to heavy traffic and complex road networks. Furthermore, the study reveals a significant association between higher speed limits and more severe accidents, especially those resulting in fatalities, which are more likely to occur at speeds over 50 km/h. These findings underscore the research's critical role in shaping future traffic safety strategies.

When evaluating current safety measures, such as speed limits and public awareness programs, it becomes clear that while these interventions have reduced severe accidents, there is still significant potential for improvement. The results indicate that additional enhancements are not just possible but necessary. Machine learning algorithms, specifically Decision Trees and Gradient Boosting, have shown significant efficacy in predicting accident hotspots and severity, providing crucial insights for future safety efforts in Scotland.

Attestation

I understand the nature of plagiarism and know the University's policy on this.

I certify that this dissertation reports original work by me during my university project except for the following:

- The Accident and Airline Information datasets, which played a crucial role in analysing traffic accidents, were acquired via Kaggle [32] [31].
- The technological review in Section 2.5 is a testament to the wealth of knowledge in the field. It extensively relied on previous studies and research, drawing important insights from sources such as Di Milia et al. (2011) and Clarke et al. (2006) [24] [25].
- The machine learning models in Section 3.1 were created utilising established frameworks and libraries such as Python, Pandas, Matplotlib, and other referenced tools mentioned in the dissertation.
- The approach outlined in Section 3.2.1.1, which involved five steps, was derived from established methods and protocols documented in the Literature Review, including those about Geographic Information Systems (GIS) and statistical analysis [8] [9]
- The results presented in implementation section 4 were created using conventional data processing methods, guided by publicly accessible resources and documentation from Python libraries.

Signature: **Date 30/09/2024**

Acknowledgements

I express profound gratitude to my instructors, Dr Kevin Swingler and Dr Simon Powers, for their steadfast advice, knowledge, and motivation throughout this research. Despite the brevity of our time together, your guidance has been beneficial and has had a profound influence. I genuinely appreciate my incredible spouse, David Selekere, for being the pillar of strength that enabled me to achieve my goal of attaining this MSc degree. Your unwavering support and belief in me have been a formidable source of inspiration.

Dear Mr. and Mrs. Ariwayo, I am grateful for your unwavering emotional support and constant presence during these critical times. I also want to thank my classmates, Wilfred and Ayo, for their companionship and assistance throughout the MSc program. Your insightful discussions and collaborative efforts have been a significant part of my academic journey.

I am deeply thankful to Python and its libraries for providing a robust framework and clear documentation, which was pivotal in completing my research. The contribution was indispensable and greatly appreciated.

Lastly, I express my gratitude to the numerous researchers, intellectuals, and educators whose contributions have provided the basis for this thesis. I greatly appreciate your valuable and indispensable contributions to my work and your inspiration.

Table of Contents

List of Figures

1 Introduction

Examining road accidents is a crucial component of public safety research, offering a valuable understanding of these occurrence's trends, origins, and outcomes. In the UK, the primary transportation systems are maritime, aviation, rail, water and road, which serve the requirements of the people [1]. However, all these systems are susceptible to various types of accidents. The nation has witnessed a range of accidents, from minor incidents to tragic fatalities, and this circumstance has the potential to result in deaths, physical harm, and property destruction. Weather conditions, uneven road surfaces, and negligent driver behaviours frequently cause transport accidents. The emphasis on air and road accidents in this study is relevant, as these two forms of transportation are essential components of the country's transportation system. According to the Scottish Government's transportation statistics, 4,134 car crashes led to fatalities in 2022. This year, however, is one of the rare years since the 1970s when the number of accidents has been significantly lower than in 2012, marking a drop of 58 per cent. This significant decrease underscores the potential impact of the study's findings, offering hope for a safer future [2].

Air transportation in Scotland is indispensable for linking isolated regions and enabling global travel. The work of aviation safety professionals, policymakers, and researchers is crucial in ensuring air transport safety, given the potential for severe consequences in the event of an accident. An in-depth examination of air traffic accidents necessitates a sophisticated methodology considering various elements, including aircraft maintenance, pilot errors, air traffic management, and environmental conditions. Given the infrequency of air traffic accidents compared to road accidents, it is essential to examine the seriousness and outcomes of these incidents and the efficiency of current safety measures and protocols in preventing such occurrences [3].

Multiple research inquiries drive this study to comprehend the broader framework of traffic accidents in Scotland. The initial study inquiry is to ascertain the main factors contributing to road traffic accidents and how these factors differ among various locations and demographic groups. The second study inquiry investigates the influence of meteorological conditions on the occurrence and intensity of road traffic accidents, which could significantly impact road safety strategies. The varied climate in Scotland, which encompasses both mild and severe weather conditions, substantially impacts road safety [4], [19].

The third study inquiry centres on the efficacy of current road safety measures in mitigating traffic accidents. This involves assessing the significant effects of legislation, such as speed restrictions and prohibitions against driving under the influence, as well as the crucial role of enhancements to infrastructure, such as road signs and measures to reduce traffic speed [5]. The fourth research topic delves into the factors that contribute to air traffic accidents in Scotland, a crucial area of study that demands immediate attention. It compares these factors with those related to road traffic accidents, aiming to underscore the unique safety challenges in air transport, such as the impact of human errors, technical malfunctions, and environmental variables. The fifth research inquiry explores the influence of regulatory frameworks and safety measures on preventing air traffic accidents. This involves assessing the effectiveness of current laws and identifying areas that may need improvements to strengthen air transport safety, a matter of utmost importance in aviation [6].

1.2 Scope and Objectives

This thesis aims to analyse transportation mishaps involving air and road travel comprehensively. It will entail a sequential process encompassing data collection and preparation, exploratory data analysis (EDA) to reveal patterns and relationships, utilisation of machine learning models to forecast accident incidents, identification of accident hotspots through geospatial analysis, examination of time series data to understand trends and seasonal patterns, and simulation of traffic scenarios to assess potential interventions.

The objectives of this thesis are as follows:

- Conduct an in-depth review of existing research and literature on traffic accidents
- Identify patterns and trends in traffic accidents, such as the frequency of specific accident categories, the relationship between weather and accidents, and the influence of road design, including high-risk location and time on accident rates.
- Graphically illustrate the results, utilising interactive components such as heat maps, scatter plots, and dynamic charts.
- Comprehensively analyse the impact of meteorological conditions, human mistakes, and road infrastructure on air and transportation accidents.
- Influence demographic factors, such as age, gender, socioeconomic position, and geographical location, on the frequency and severity of accidents.
- Employ advanced machine learning models to predict algorithms that accurately detect and pinpoint high-risk situations and locations.

1.3 Research Questions

The following research inquiries, with the potential to significantly enhance public safety in Scotland, guide this study:

1. What are the primary elements contributing to road traffic accidents in Scotland, and how do these factors change among various areas and specific demographic groups such as age, gender, and socio-economic status?

2. How do climatic conditions impact the frequency and intensity of road traffic incidents in Scotland?

3. What is the efficacy of existing road safety measures, such as legislative interventions and infrastructural enhancements, in mitigating traffic accidents?

4. What are the contributing variables, such as weather conditions, human error, and technical malfunctions, to air traffic accidents in Scotland, and how do they contrast with the elements impacting road traffic?

5. To what extent do current regulatory frameworks and safety measures succeed in preventing air traffic accidents?

1.4 Achievements

This dissertation investigates transportation accidents in Scotland, specifically focusing on road and aviation incidents. The research begins with an introductory section that sets the stage for the comprehensive literature review, establishing the study's background and framework. The methodology section then meticulously explains the data collection and analysis techniques, including the innovative use of ML- machine learning and Geographic Information Systems (GIS) for predictive analysis. The key findings are presented in the implementation sections, highlighting the potential impact of the research. The dissertation concludes by providing a brief overview of the discoveries, acknowledging limitations, proposing recommendations for policy enhancement and further study, and underlining the topic's significance.

1.5 Overview of Dissertation

This dissertation comprehensively analyses Scotland's traffic accidents, explicitly addressing road and aviation events. The study begins with an introductory section outlining the topic and an extensive literature analysis that contextualises the research into the broader framework of traffic safety studies. The methodology chapter provides a detailed overview of the data collecting and analysis methodologies, encompassing machine learning models and GIS for predictive analysis. The results and discussion portions of the study reveal the primary discoveries, juxtaposing them with prior research. The dissertation finishes by providing a concise overview of the research findings, acknowledging the study's limitations, proposing ideas for enhancing policy, and offering proposals for future research endeavours.

2 Literature Review

Over the past century, the methodical examination of traffic incidents in Scotland has seen significant advancements, bringing hope for a safer future. The progress has been propelled by technical breakthroughs, governmental modifications, and heightened awareness of public safety concerns. In 2022, Transport Scotland undertook a thorough investigation of traffic-related occurrences. This investigation utilised advanced statistical methods such as regression analysis, data mining, and a robust methodology that assessed data collected from accident reports. The study revealed a significant decrease in recorded casualties compared to previous years, consistent with the ongoing downward trend observed over the preceding decade. The investigation identified disparities in casualty rates based on variables such as road category, meteorological circumstances, and time of day [2].

Racioppi et al. (2004) examined accidents caused by cars in the WHO European Region. They found that roughly 127,000 deaths occur each year because of road traffic accidents, with a significant impact on young adults. The study highlighted the efficacy of multisectoral road safety methods, like those implemented in Sweden, that substantially decreased fatalities and injuries. The key recommendations underscore the role of the audience, as public health professionals, policymakers, and researchers, in implementing change through solid political dedication and integrating health considerations into road safety initiatives [7]. Additionally, Shafabakhsh et al. (2017) conducted a study using Geographic Information System (GIS) tools to analyse the geographical distribution of urban traffic accidents in Mashhad, Iran. Their examination of traffic accident data in urban regions indicated locations with a significant likelihood of accidents. The report offered valuable observations on the considerable patterns of accidents, namely in densely populated areas and near intersections. This study vividly demonstrated the power of Geographic Information Systems (GIS) in visually representing and analysing areas with high traffic accident rates, making complex data easily understandable [8].

In 2020, Feng et al. conducted a groundbreaking study that utilised innovative big data analytics and machine learning approaches to analyse, visualise, and predict traffic accidents in the UK. The study used an extensive dataset from the UK's Department for Transport, comprising more than 2 million accident records from 2005 to 2017. It used clustering techniques, interactive mapping, and advanced deep learning models, including LSTM (Long Short-Term Memory) and Facebook Prophet model, to detect hotspots and predict future accident patterns. The results indicated that their technology proficiently managed extensive datasets and offered valuable insights for traffic management and accident avoidance [17].

2.1 Participating Elements

In 2019, Lewsey et al. [9] conducted a rigorous and thorough study to examine the impacts of a new legislation that reduced the permissible blood alcohol level for driving in Scotland, which examined how this legislation affected both the occurrence of road traffic accidents and levels of alcohol consumption. It utilised advanced statistical analysis methods, underscoring the scientific rigour of the research, to analyse data from road traffic accident reports and alcohol sales records. The results, which demonstrated a significant reduction in alcohol-related road traffic accidents following the implementation of the new laws, provide compelling and irrefutable evidence for the efficacy of reducing the legal blood alcohol concentration limit for driving to improve road safety and decrease alcohol intake [9]. Similarly, in 2019, Santoso and Maulina conducted a research project on human traffic accident errors, explicitly focusing on comparing the experiences of individuals operating cars and motorcycles. The study, which employed qualitative analytic methodologies, gathered data from individuals involved in traffic accidents through surveys and interviews. The data revealed a range of human errors that led to incidents for both automobile drivers and motorcyclists. Notably, the study highlighted the unique characteristics of human errors in each group, with automobile drivers often making mistakes due to a lack of focus and incorrect assessment of road conditions. These findings underscore the need for customised safety interventions and educational programs [10]. In 2009, Antov et al. embarked on a study to evaluate the influence of urban roundabouts on the lowering of vehicle speed. The research, which employed speed measurement instruments to collect data on vehicle velocities before and after the implementation of roundabouts in urban regions, followed a rigorous methodology. According to the study, vehicle speeds dropped discernibly at the entrances and exits of these circular crossings, proving that urban roundabouts are valuable tools for lowering vehicle speeds and improving road safety [11]. In 2024, McCarty and Kim accomplished a study investigating the impact of age and gender on vehicle accident rates in England. The researchers utilised regression-based machine learning models to examine demographic data from the UK census and traffic accident records. The study revealed that younger male drivers had elevated accident rates, whilst elderly groups demonstrated fewer accidents. These findings are significant to public health and transportation, as they underscore the potential of using demographic information to forecast and handle road safety hazards, emphasising the possibility of implementing focused interventions [10], [18]. Meanwhile, the primary cause of air transportation disasters is often attributed to pilot error. This highlights the critical role of human factors in aviation safety, particularly pilot decision-making and expertise levels. In 2018, a study by Kharoufah et al. delved into the human elements that influenced commercial air transport accidents and incidents. The study, which employed an exploratory research methodology, integrated qualitative data from numerous aviation safety databases and conducted quantitative analysis using Pearson's Chi-Square test. More than 200 incidents were examined and classified based on the type of operation, the operator's condition, the phase of flight, and the human component that caused them. The results highlighted the significant role of situational awareness and the impact of failure to follow procedures. The analysis also revealed that specific operational categories, such as charter operations, had a higher frequency of accidents linked to human factors [6].

In 2013, the Scottish Safety Camera Program performed a comprehensive study to evaluate the effort of red-light cameras (RLCs) in reducing traffic accidents. The study employed rigorous statistical analysis techniques to scrutinise extensive data on traffic events, including accident rates, both before and after the deployment of RLCs in various regions of Scotland. According to the Reported Road Casualties in Scotland 2011, there was a 41% decrease in severe and fatal accidents on all roads in Scotland between 2001 and 2011. Furthermore, the number of accidents of all severities decreased by 32% during the same period. However, the study revealed that the impact of RLCs on enhancing traffic safety was minimal, with specific locations even witnessing an increase in incidents, suggesting that RLCs had limited effectiveness as a safety measure and raising concerns about their widespread use, highlighting potential risks that need to be addressed in the future [12].

2.2 Road Users at Risk

Bicycles, motorbikes, and pedestrians are classified as vulnerable road users since they are more prone to becoming injured in case of an accident. Suggestions have been made to enhance the safety of road users, such as implementing dedicated bike lanes and pedestrian crossings. However, it is equally important to emphasise the role of public awareness campaigns in promoting safe driving habits for drivers and other individuals at risk on the road [13]. In 2012, Freeman et al. disclosed a lengthy assessment examining vulnerable road user categories, including younger drivers, motorcyclists, and senior drivers. The findings identified specific risk variables connected with each group, such as increased propensity for risk-taking in younger drivers and greater physical susceptibility in older drivers. The study clearly showed the necessity of customised road safety measures to target the hazards encountered by these susceptible populations [13].

2.3 Factors that impact air traffic accidents

Air traffic occurrences are commonly linked to multiple sources, including technological failures, human errors, and environmental conditions. Human error is a primary factor contributing to aircraft accidents, encompassing factors such as pilot fatigue, ineffective communication, and insufficient training. Studies have shown that fatigue in pilots can impact their cognitive functions and response times, increasing the likelihood of errors during critical stages of a flight. In 2012, Oster et al. [14] executed a globalised study examining aviation safety's economic analysis. The study used accident data analysis and safety performance records from different commercial aviation sectors. The study revealed notable aviation safety advancements, particularly in developed nations. However, discrepancies persist in less developed places. The findings highlighted the significance of employing reactive and proactive safety strategies to improve worldwide aviation safety standards. Imposing stringent regulations on the number of hours pilots can work and the amount of rest they must have been essential for minimising this risk. Another significant peril to aviation safety is mechanical malfunction, encompassing issues with aircraft engines, navigation systems, and structural components. Regular inspections and maintenance are crucial for promptly identifying and addressing mechanical problems, providing a sense of security, and helping mitigate the risk of accidents. Adverse weather conditions, collisions with birds, and volcanic ash can disrupt flight operations and increase the risk of accidents. Low visibility conditions, such as fog or severe weather, can complicate take-off and landing operations, while bird attacks can potentially damage an aircraft's engines [14], [15]. In 2013, Joni K. Evans conducted an extensive study investigating aviation incidents associated with turbulence, wind shear, and thunderstorms. By employing statistical analysis of more than 800 mishaps, the study classified them into seven separate air danger categories. The study's results have significant practical consequences for aviation safety. They uncover the frequency of clear air turbulence and low-altitude wind shear in aviation accidents, especially in certain geographic regions and during specific flight operations [16].

2.4 Variations in Accident Frequencies and Trends Over Time

Scotland has seen a consistent decline in its accident rate in recent decades, a trend attributed mainly to implementing more stringent traffic regulations, improving road infrastructure, and enforcing better car safety standards. The research by Transport Scotland [2] indicates a substantial decrease in traffic fatalities over the past decade, with a specific decline of 58 % in the last five years. This decline in vehicle accidents can be credited to various factors, including the implementation of advanced driver-assistance systems (ADAS), electronic stability control (ESC), and anti-lock braking systems (ABS) in cars [20]. In the 2014 study, Bengler et al. thoroughly examined the development and potential future directions of driver assistance systems (DAS) spanning three decades. The study discovered that initial DAS systems included proprioceptive sensors, which later progressed to include exteroceptive sensors like radar and lidar. This trend exemplified the gradual shift towards autonomous and collaborative driving. The study addressed the growing incorporation of Driver Assistance Systems (DAS) into automobiles and stressed the necessity for more investigation into automated driving and the interaction between humans and machines. The results indicate that although there has been notable advancement, there are still considerable obstacles in attaining completely self-driving vehicles [21]. Additionally, there have been instances where accident rates have notably risen due to circumstances like heavy holiday traffic, which frequently causes congestion and driver tiredness, or unfavourable weather conditions, which decrease visibility and grip. Time series analysis was used by Razzaghi et al. (2013) to thoroughly examine the trend and seasonality in traffic accidents [22].

Furthermore, there have been improvements in safety standards and aviation technology and a decrease in air transport accidents. The integration of the Enhanced Ground Proximity Warning System (EGPWS) and Traffic Collision Avoidance System (TCAS) has been crucial in averting mid-air crashes and incidents of controlled flight into terrain (CFIT), In 2015, Lu and Yang conducted research to assess the efficacy of Ground Proximity Warning Systems (GPWS) in mitigating Controlled Flight into Terrain (CFIT) incidents. The study discovered that the compulsory implementation of GPWS substantially reduced CFIT accidents. The study demonstrated that although GPWS has enhanced safety, additional restrictions are necessary to encompass smaller aircraft to mitigate the dangers of Controlled Flight into Terrain (CFIT) [23].

2.5 Factors Related to Demographics Affecting Accident Rates

Accident rates are significantly influenced by these demographic variables: age, gender, and socioeconomic position [24]. Di Milia et al. (2011) conducted a thorough and comprehensive study to determine the relationship between demographic characteristics, weariness, and accident risk in transport systems. The study meticulously analysed existing literature to examine these connections, particularly emphasising age, gender, socioeconomic status, employment arrangements, and individual variations. The findings emphasised notable methodological constraints in the field and advocated for further multidisciplinary research. The study revealed that weariness is a complex phenomenon influenced by various demographic characteristics, with age and sex playing particularly influential roles, underscoring their importance in the study [24]. A similar study initiated by Clarke et al. (2006) aims to investigate the impact of age, driving experience, and time of day on accidents involving young drivers in the UK. They employed statistical approaches such as regression analysis and chi-square tests to examine the data and find common factors contributing to accidents. The study also revealed that younger drivers, especially males, were more susceptible to accidents under specific circumstances, such as driving at night and navigating country curves. The results emphasise the necessity of implementing focused measures to reduce accidents among young drivers, including those associated with driving at night and navigating rural road conditions [25]. Williams (2006) conducted a study that examined the factors that increase the likelihood of accidents among young drivers. The study also assessed several strategies implemented to decrease the number of crashes in this age group. The study employed data from prior studies, crash statistics, and assessments of treatments such as graduated licensing and driver education programs. The results suggested that conventional driver education programs had not successfully decreased crash rates, but graduated licensing programs showed a decrease in crashes by 20–30%. The study highlighted the crucial role of community-based initiatives, which involve the collaboration of parents and law enforcement, in improving adherence to license requirements [26].

2.6 Safety Procedures and Their Performance

Many safety initiatives have been widely implemented to mitigate traffic accidents in Scotland. Typical solutions include enhancing road infrastructure, such as building roundabouts, implementing public awareness campaigns, and installing speed cameras. Also, the aviation sector has experienced a rise in safety due to many measures, including enhanced maintenance processes, upgraded pilot training programs, and the implementation of revolutionary navigation technology. Furthermore, the Federal Aviation Administration (FAA) has implemented various measures to enhance safety consciousness among pilots and air traffic controllers [15], [16].

2.7 Technological Advancement's effect on the Study of traffic accidents

Technological progress has dramatically enhanced the analysis and prevention of road accidents. Flight data recorders, also known as 'black boxes', are essential instruments in aviation for investigating accidents. They provide comprehensive and detailed information on aircraft performance and pilot activities, which is instrumental in identifying the causes of accidents and improving safety measures [27]. The Air Accidents Investigation Branch (AAIB) employs these technologies with witness testimony and site evaluations to establish global safety standards and provide input for regulatory modifications [27].

Using GPS and GSM technologies in road traffic has improved emergency response times by promptly alerting services about accidents, potentially resulting in preserving lives [28], [29]. Furthermore, simulation models such as AIMSUN and VISSIM play a crucial role in evaluating elements such as weather and driver behaviour, aiding in predicting the effects of different traffic management techniques. This comprehensive analysis is instrumental in understanding the complexity of road safety and improving responses to accidents [30]. By combining Artificial Intelligence (AI) and Machine Learning (ML) with Geographic Information Systems (GIS), the analysis of traffic accidents is significantly improved. This integration allows for the identification of patterns and the prediction of accident hotspots.

3 Research Methodology

This thesis employs a quantitative approach to investigating accidents because it can systematically measure and analyse extensive datasets, such as records of traffic accidents. The study incorporates the following components: examination of the current system, preparation and processing of data, data analysis, creation of data models, and visualisation. Every individual stage is crucial in ensuring thorough and precise data comprehension, resulting in practical insights and recommendations directly applicable to transportation and public safety.

3.1 Current System

These new systems heavily rely on machine learning techniques and predictive analytics methods such as logistic regression, decision trees, random forests, and gradient-boosting machines, which are utilised to create proactive prediction models. These models use data on weather conditions, traffic volume, and time of day to predict locations where accidents are likely to occur. Geospatial analysis, facilitated by Geographic Information Systems (GIS), enables the mapping of accident locations and the identification of areas with a high risk of accidents. GeoPandas and Folium are software tools that can spatially display accident data, emphasising areas that need attention.

The data cleaning procedure addresses problems such as missing values, duplicate records, and inconsistent data formats. This approach guarantees the quality and dependability of the information.

Exploratory Data Analysis (EDA) reveals underlying patterns, trends, and anomalies within the dataset. Heatmaps, scatter plots, histograms, and correlation analysis are analytical techniques that offer valuable insights into the relationships between road types and weather conditions that influence the severity of accidents.

3.2 Road Traffic Accident Dataset Collection

The primary data sources for this study are the public datasets on road safety in the United Kingdom, accessed via the Kaggle. Data collection is needed to ensure the validity of the analysis and findings. These datasets include comprehensive and in-depth data on traffic accidents, including essential characteristics like the date, time, and location of events, the severity of the collisions, the kinds of cars involved, the state of the roads, and the weather at the time of the incident etc. The CSV files "Accident Information.csv" comprise the datasets used in this study.

Accident_Index	1st_Road_Class	1st_Road_Number	2nd_Road_Class	2nd_Road_Number	Accident_Severity	Carriageway_Ha
200501BS00001	A	3218		0	Serious	None
200501BS00002	B	450	C	0	Slight	None
200501BS00003	C	0		0	Slight	None
200501BS00004	A	3220		0	Slight	None
200501BS00005	Unclassified	0		0	Slight	None
200501BS00006	Unclassified	0		0	Slight	None
200501BS00007	C	0	Unclassified	0	Slight	None
200501BS00009	A	315		0	Slight	None
200501BS00010	A	3212	B	304	Slight	None

Figure 1. Sample images from the accident dataset

3.2.1 Airline Accident Dataset Collection

The National Transportation Safety Board (NTSB) investigation data for aircraft accidents was the source of the data obtained from Kaggle to create the air traffic accident analysis. These datasets include characteristics such as date, location, country, injury severity, aircraft registration number, category, and total fatal, minor, serious, and unserious injuries.

Event Id	Investigation Type	Accident Number	Event Date	Location	Country	Latitude
20080125X00106	Accident	SEA08CA056	12/31/2007	Santa Ana, CA	United States	33.675556
20080206X00141	Accident	CHI08WA075	12/31/2007	Guernsey, United Kingdom	United Kingdom	49.435
20080129X00122	Accident	CHI08CA057	12/30/2007	Alexandria, MN	United States	45.866111
20080114X00045	Accident	LAX08FA043	12/30/2007	Paso Robles, CA	United States	35.542222
20080109X00032	Accident	NYC08FA071	12/30/2007	Cherokee, AL	United States	34.688611
20080129X00118	Accident	DEN08CA045	12/29/2007	Gunnison, CO	United States	38.533889
20080214X00193	Accident	CHI08CA056	12/29/2007	Abingdon, IL	United States	40.799722
20080215X00200	Accident	CHI08CA058	12/29/2007	Crystal Falls, MI	United States	46.8
20071231X02014	Accident	DFW08FA053	12/29/2007	Venice, LA	United States	28.958056

Figure 2. Airline Dataset

3.2.1.1 Data Preparation and Processing

Multiple crucial processes are conducted while preparing traffic accident data for analysis to guarantee that the dataset is error-free, maintains uniformity, and is ready for subsequent analysis. The procedure commences by importing the dataset into a Panda DataFrame, a versatile framework for manipulating data. After loading the data, the initial step is to filter the dataset to include accidents that occurred exclusively in Scotland.

Subsequently, the data cleaning procedure addresses the issue of missing values, which is common in extensive datasets. The 'drop () 'function is crucial in eliminating rows with missing values. This decision is necessary to ensure the data's robustness and integrity.

The 'Date' column is subsequently transformed from a string format to a datetime format using the pd.to_datetime() function. This translation is essential for conducting time-based analysis, such as detecting patterns across specific years. A vital part of this transformation is the 'errors='coerce'' argument, which is crucial in managing erroneous date formats by transforming them into NaT (Not a Time) values. This ensures the accuracy of the data transformation process. Following the conversion process, the data is examined for any NaT (Not a Time) entries to verify that all dates have been effectively changed. Next, the year is extracted from the 'Date' column, and a new column called 'Year' is created. Following the feature engineering process, the data is reevaluated to identify any missing values in the recently generated 'Year' column and verify that the feature engineering procedure did not introduce any discrepancies.

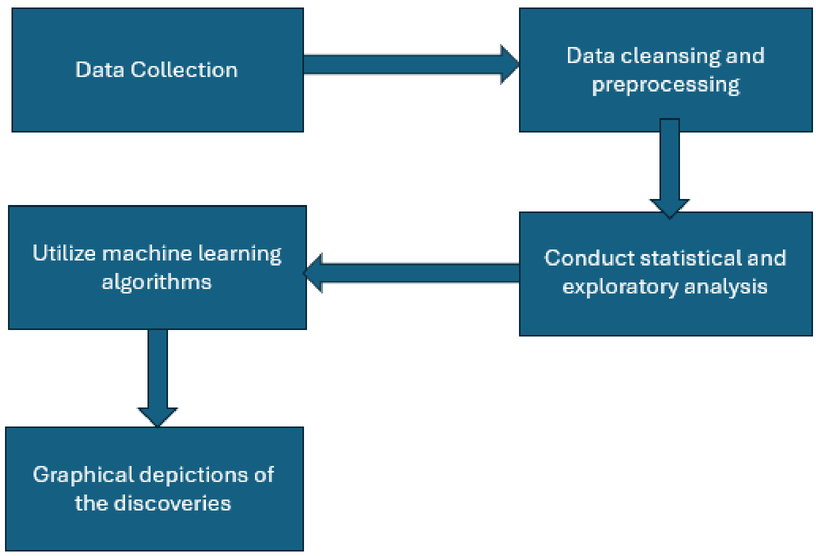

Figure 3. Analysis Workflow

3.3 Data Analysis

This research uses statistical and computational approaches to extract essential insights from the processed data. Python and its robust libraries, such as Matplotlib, Plotly, and Pandas, are utilised for this. Data analysis facilitates the identification of the main reasons contributing to traffic accidents, the geographical regions most impacted, and the influence of numerous variables such as weather conditions and road classifications.

- **Descriptive and Inferential Statistics**

Descriptive statistics, such as variability and central tendency measures, provide a comprehensive view of the data, reassuring understanding of the distribution of accident severity and the average number of accidents per year. GIS technologies are then utilised to create maps of accident locations throughout Scotland. Heatmaps are employed to investigate the relationships between different elements contributing to accidents.

- **Models for Machine Learning**

Machine learning (ML) methods are becoming more prevalent in predicting traffic accidents and identifying high-risk locations. Algorithms like logistic regression, decision trees, random forests, and gradient-boosting machines are used to create prediction models. These models use many variables, including traffic volume, weather, and time of day, to predict the likelihood of accidents. Feature engineering, which comprises creating new predictive variables from raw data, is essential in increasing model accuracy.

- **Training and Assessment of Theories**

The dataset is partitioned into separate testing and training sets to evaluate the model's performance. The testing set assesses the model's accuracy once it has been built using the training set. Accuracy, precision, recall, F1-score, and AUC-ROC are crucial metrics for evaluating performance.

- **Visualization of Traffic Accidents on the Road**

Among the visuals used to represent automobile accidents are:

- Plots time series: Displaying the historical pattern of accidents.
- Accident hotspots are highlighted on maps using geographic data.
- Bar charts: Evaluating how frequently accidents occur on various kinds and conditions of roads.

- **Visualization of Airline Accidents**

Among the air traffic accident visualisations are:

- Pie charts: Showing how accident causes are distributed.
- Line graphs: Showing patterns in the number of accidents over time.
- Scatter Plots: Displaying the correlation between accident severity and flight phases.

- **Tools and Technology**

The below tools and technologies are utilised in this thesis:

- Python is the primary programming language used for data analysis.
- Pandas: For analysis and data manipulation.
- Matplotlib was used to create static visualisations.

- Plotly: For producing interactive visuals.
- Jupyter Notebooks: To record and distribute the analytic

4 Implementation and Documentation

This investigation aims to gain a thorough understanding of traffic incidents in the UK to discover patterns, causes, and the efficacy of existing safety measures, aiming to decrease the frequency and severity of accidents. The analysis addresses inquiries regarding the temporal and spatial distribution of accidents, the leading causes and contributing factors, the efficacy of current safety measures, locations with a high risk of accidents, and prediction insights for future accidents.

4.1 Road Scope of Analysis

The analysis meticulously examines road accidents in the United Kingdom, specifically those recorded from 2005 to 2010. The dataset, sourced from Kaggle, covers a range of accidents classified by severity—fatal, serious, and slight—and spans numerous geographical regions. It includes driver demographics, such as age and gender, and provides detailed information about the circumstances of each accident, instilling confidence in the thoroughness of the analysis. The preprocessing procedures ensure the quality of the data by addressing missing values, guaranteeing uniform formats, and eliminating duplicates. Also, the data was manipulated by transforming categorical characteristics into numerical values, combining data for analysis over time, and generating additional attributes such as time of day and season. Descriptive and inferential statistics concisely summarise the data and reveal connections between variables.

4.2 Descriptive Statistics

- **Participation in road-related activities**

The data shows a wide range of road numbers associated with traffic accidents, with an average of 1011.997 for the first and 387.0004 for the second. The median results (125 for the first road number and 0 for the second) suggest that a substantial number of accidents occur on principal roads without any involvement of secondary roads.

- **Police presence**

The statistics reveal the pivotal role of police officers in attending to traffic incidents, with an average attendance value of 1.198669. Their extensive participation in this process is instrumental in ensuring the reliability and comprehensiveness of the recorded accident reports.

- **Geographical Distribution**

The geographical coordinates (latitude and longitude) indicate that accidents are evenly dispersed throughout Scotland. The average latitude is 52.57297, and the average longitude is -1.454537, within the predicted range.

- **Loss of life and transportation equipment**

The average number of casualties in each traffic collision is 1.359930, while the average number of vehicles involved is 1.835030. The median value for both criteria is 1, which suggests that most incidents lead to only one casualty and involve two cars.

- **Speed limits**

The average speed restriction at accident sites is around 39.43654 km/h, with a standard deviation of 14.31206. This implies that many accidents happen in areas with moderate speed limits, although there is considerable variation, with speed limits ranging from 10 km/h to 70 km/h. This diversity underscores the complexity and variety of the issue, demonstrating the occurrence of accidents in various roadways, ranging from low-speed urban areas to higher-speed rural or arterial routes.

- **Duration of Time Covered**

The dataset covers 2005 to 2010, with an average year of 2007328. This analysis's temporal dimension allows for a comprehensive examination of accident patterns and trends across several years, enabling the evaluation of the long-term impact of safety actions.

4.3 Analysis of Road Traffic Accident Trends

The chart in Figure 4 Vividly illustrates the prevalence of traffic accidents according to severity, categorised into Slight, Serious, and Fatal. This insightful distribution provides a valuable understanding of the characteristics and consequences of road accidents. Most traffic accidents, more than 600 occurrences, are classified as 'minor'. While less severe, it occurs more frequently and can lead to minor injuries or damage. There is a notable disparity between the number of accidents categorised as 'Serious' and 'Slight,' with around 100 instances documented in the former category. This suggests severe accidents occur less frequently but have significant hazards and consequences. Fatal accidents, with a count approaching zero, are infrequent. The frequent occurrence of minor accidents can be ascribed to numerous low-speed crashes, efficient safety measures such as seatbelts and airbags, and advancements in vehicle safety regulations.

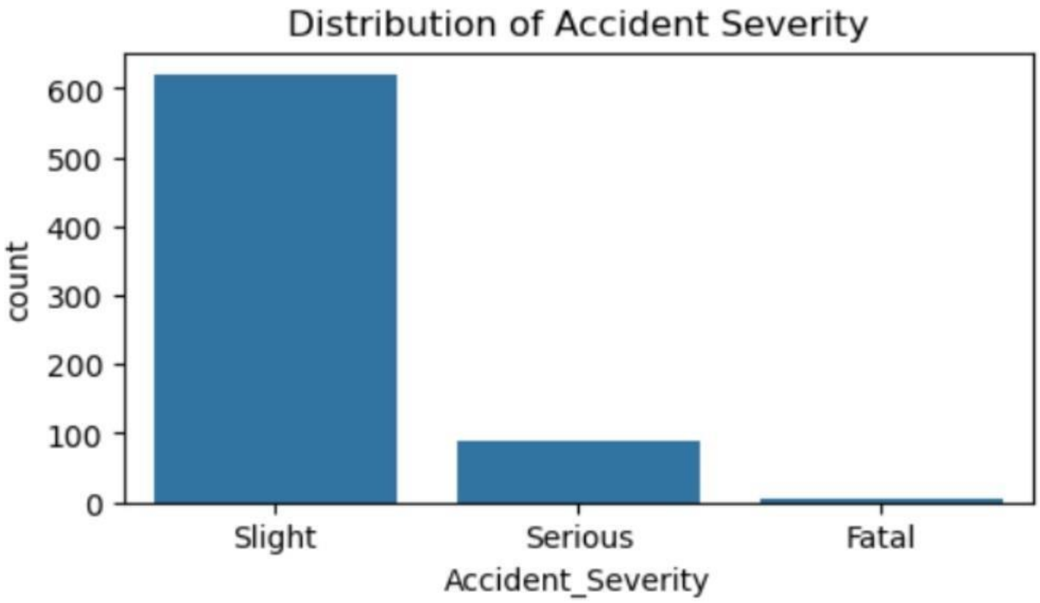

Figure 5. The Accident severity

- The line chart in Figure 5 illustrates the yearly occurrence of traffic accidents between 2005 and 2010. In 2005, the number of accidents was over 200,000; however, this figure consistently declined in the following years. In 2010, the incidence of road accidents decreased to around 150,000. The continuous decrease in traffic accidents indicates the successful execution of road safety measures, advancements in car safety technology, and potentially strengthened traffic law enforcement. However, the most significant factor contributing to this decline is the role of public awareness programs in promoting a safety-oriented culture on the roadways. The data shows a steady and consistent improvement in traffic safety, and the findings underscore the effectiveness of current initiatives to decrease traffic accidents.

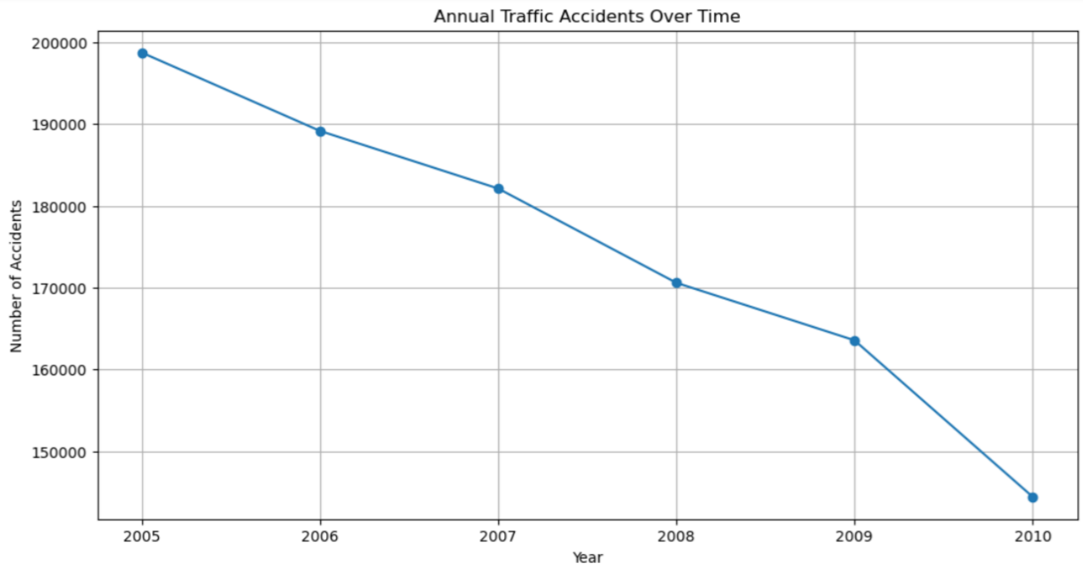

Figure 6. Accidents trend over time.

- This graph in Figure 6 presents the average monthly frequency of traffic accidents, revealing significant patterns that change with the seasons. Notably, January, March, May, June, and October show significantly higher average accident rates, ranging from around 14,500 to 15,000. With the highest average number of accidents, November stands out, approaching about 16,000. Conversely, February and December show lower mean values, with February recording the minimum number of accidents, approximately 12,500. These findings underscore the importance of understanding seasonal variations in traffic accidents. These statistics reveal a higher occurrence of accidents during specific months, particularly in the autumn and early winter months of October and November. This tendency can be ascribed to many variables, encompassing shorter periods of sunshine, adverse weather conditions, and increased travel activity during vacation. In contrast, despite being in the winter season, February and December show lower accident frequencies, potentially due to a reduced number of travel days or the effectiveness of public safety initiatives. These findings suggest successful interventions can significantly reduce traffic accidents, offering hope for improved road safety. The region with the highest number of accidents in Scotland is 32834, totalling 19081 incidents. The area with the lowest number of accidents in Scotland is 17062, with only one incident.

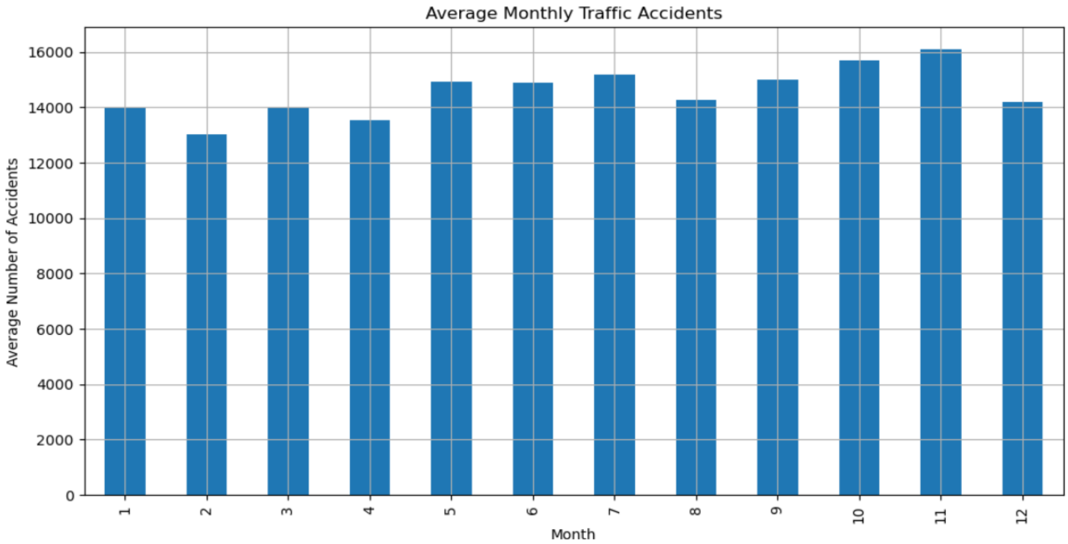

Figure 7. Monthly trend

- The bar chart in Fig 7 presents a significant insight into the ten regions in the UK with the highest frequency of traffic accidents. Birmingham stands out with the highest incidence of accidents, exceeding 20,000. Leeds has the second highest number of accidents, totalling over 13,000, while Manchester is a close third with slightly over 10,000. Bradford, Liverpool, and Sheffield all report comparable accident statistics, with each city recording figures ranging from 9,000 to 10,000. Westminster, Glasgow City, and Bristol City are near one other, with each town reporting just under 9,000 accidents. Kirklees completes the list with a total of slightly more than 7,000 accidents. The data indicates a dramatic decrease in accident numbers following Birmingham, an unexpected anomaly. Although Leeds and Manchester also have somewhat elevated accident rates, the other places demonstrate a steadier decrease, indicating a more equitable distribution of traffic accidents throughout these regions.

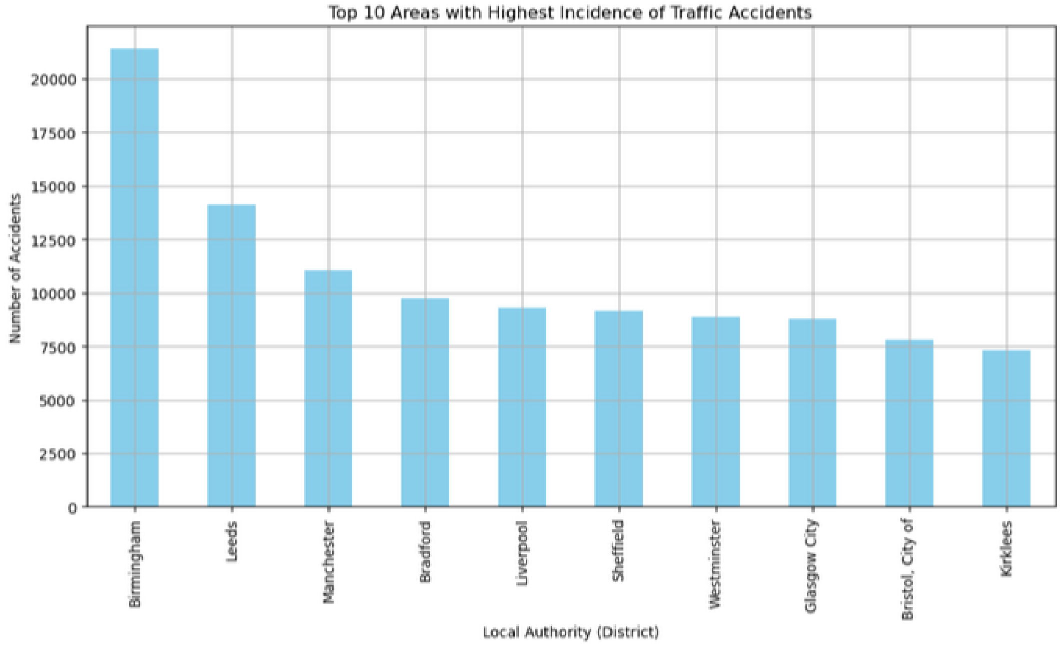

Figure 8.Highest areas with traffic accidents trend

4.3.1 Categorical Analysis of Road Traffic Accidents

The data reveals a significant trend in the frequency of weather conditions during traffic accidents. Notably, the most common condition, 'Fine no high winds,' Accounts for over 500 incidents. This highlights the need for caution even in seemingly safe weather, as most accidents occur when the weather is clear and no adverse conditions are present. The data also points to 'Raining with no high winds' as the second most common weather condition during traffic accidents. While it occurs less frequently than in good weather, it's crucial to emphasise the potential risks even in mild weather. Even without strong winds, rain significantly increases the likelihood of traffic accidents, necessitating heightened awareness and preparedness. It's important to note that additional weather conditions such as 'Unknown,' 'Other,' 'Fine + high winds,' 'Fog or mist,' 'Snowing no high winds,' and 'Raining + high winds' have far lower frequencies.

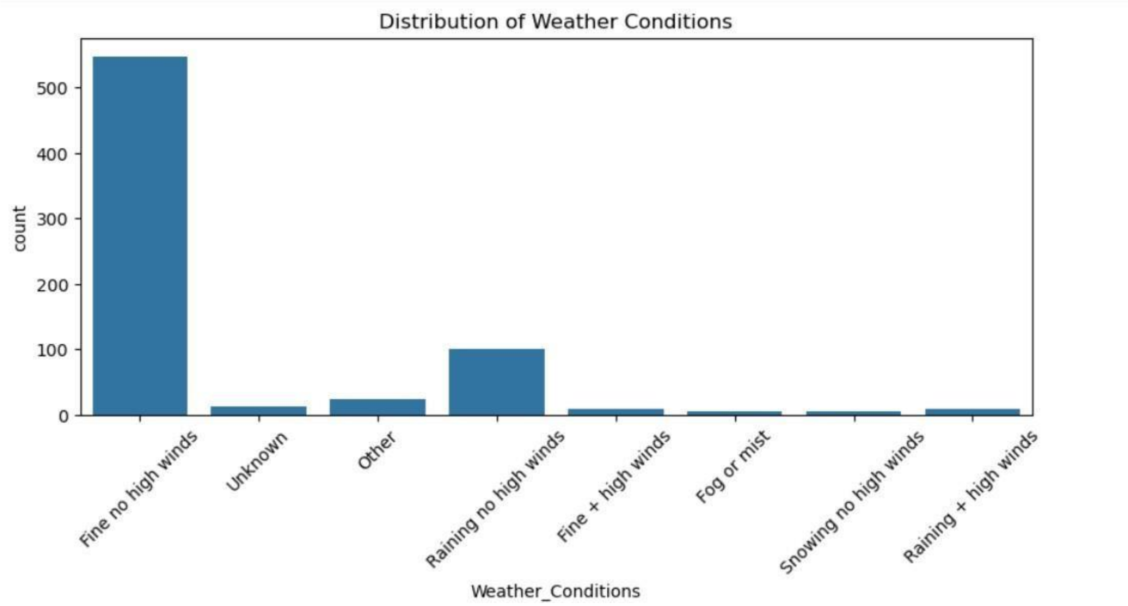

Figure 9.The Weather Conditions

4.3.2 Severity of Road Traffic Accidents

The box plot in Fig 9 effectively demonstrates the correlation between speed limits and the severity of traffic accidents, a finding that could significantly influence road safety policies. The median speed limit for minor accidents is approximately 30 km/h, most occurring within the 30-40 km/h range. Several data points exceed 50 km/h, indicating infrequent minor incidents at greater velocities. The median speed limit for severe accidents is slightly elevated, approximately 40 km/h, with a comparable interquartile range to minor accidents. This suggests that these accidents also occur between the 30-50 km/h range, with a few exceptional cases above 50 km/h. On average, fatal accidents happen at a speed restriction of approximately 60 km/h. The range of speeds for these incidents is quite broad, spanning from 30 to 70 km/h. This suggests that fatal accidents are likelier to happen at higher speed limits. The result demonstrates that speed is a crucial factor influencing the magnitude of traffic accidents. Deadly accidents occur at significantly higher speed limits than severe and slight accidents. This underscores the essential influence of speed on the severity of accidents. Increased velocities reduce the driver's response time and increase the impact force, leading to more severe consequences. The presence of outliers in both minor and major accidents suggests that while these accidents mainly occur at lower to moderate speeds, they can still happen at higher speeds, indicating that factors other than speed, such as road conditions, driver behaviour, and vehicle safety features, can also influence the severity of accidents.

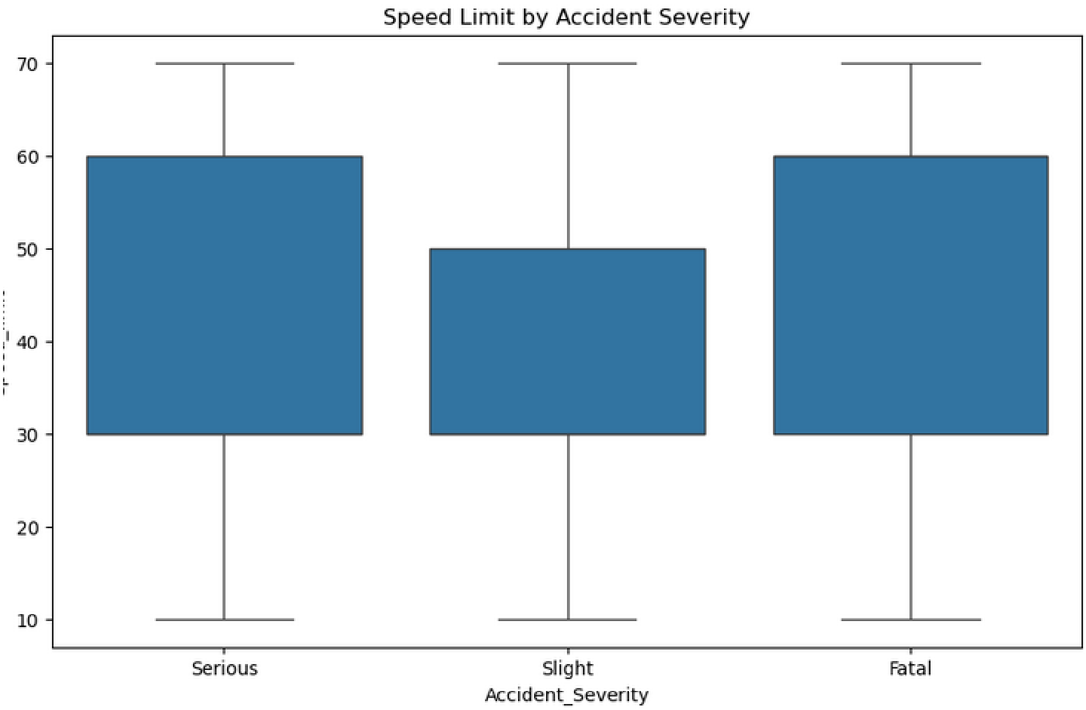

Figure 10.The speed limit

4.3.3 Exploratory Data Analysis of Road Traffic Accidents

Using latitude and longitude coordinates, the scatter plot is a crucial tool in identifying high-risk areas by displaying the geographical positions of traffic incidents, classified according to their severity (slight, serious, and fatal). The plot demonstrates that minor incidents, represented by blue dots, are evenly distributed over the area without any noticeable grouping. Severe accidents, indicated by orange, occur less frequently but tend to concentrate in regions with high traffic volume. Fatal accidents, shown by the colour green, are infrequent and are dispersed over the region, with a small number of easily identifiable areas with a higher risk. The distribution indicates that some regions, especially those with higher traffic levels or intricate road systems, are more susceptible to severe and deadly accidents. This emphasises the importance of the scatter plot in identifying and resolving these high-risk areas.

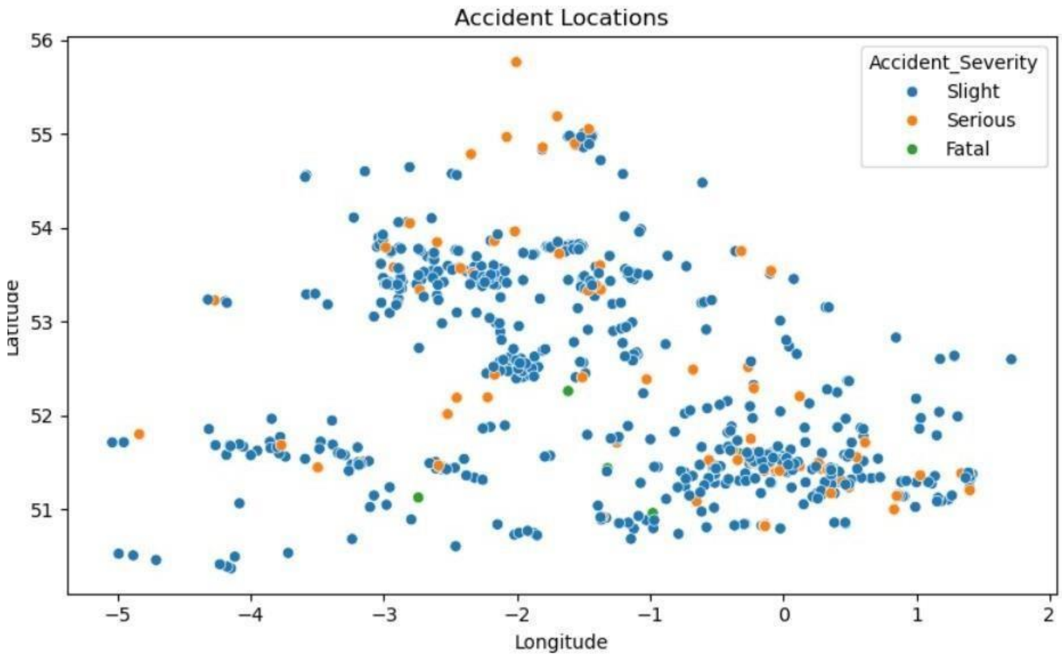

Figure 11.Accident locations with scatter plot

- The heatmap in Figure 11. is centred at the coordinates [52.377852731468536, -1.294142151048951], which are the geographical centres under analysis, offering a comprehensive picture of the area. Urban regions exhibit the greatest concentration of accidents, primarily due to the higher traffic congestion. The map shows a dense concentration of incidents in major metropolitan areas, suggesting that these locations are high-risk zones characterised by frequent accidents. By contrast, rural areas present a starkly different picture. They have fewer locations with high temperatures on the heatmap, indicating a significantly reduced occurrence of accidents. The lower density in these locations can be ascribed to decreased traffic volumes, fewer intersections, and generally lowered speeds, highlighting the difference in traffic safety between urban and rural areas. The map also features heat points, which are specific locations with a high concentration of accidents in Scotland, enabling the identification of places with the highest and lowest incidence rates. Urban areas in Scotland, such as Glasgow and Edinburgh, are prone to a higher concentration of accidents due to their large population and high traffic volume. Conversely, the Highlands and Islands, which are inaccessible and have a limited population, will likely have the lowest density of accidents.

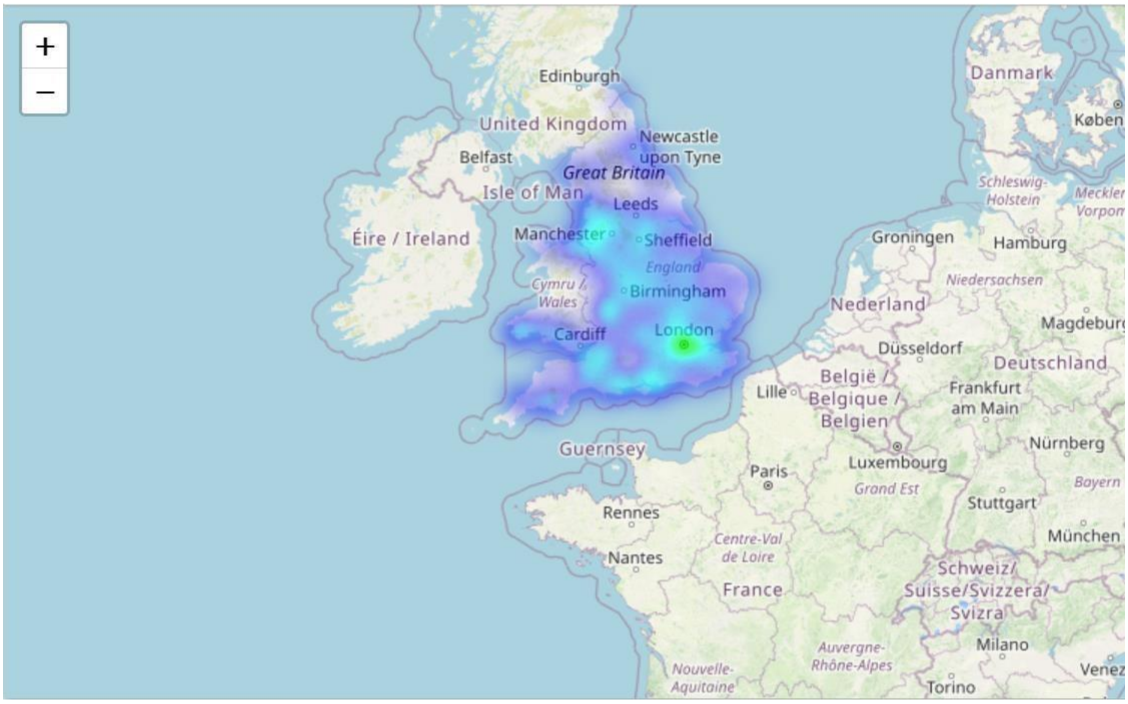

Figure 12.Accident heatmap

4.4 Model Evaluation

Evaluating the Gradient Boosting model's performance, as indicated by the classification report and confusion matrix, unveils various valuable observations. The model has strong performance in accurately predicting high-severity accidents (Class 2) with a precision of 0.85, recall of 1.00, and an F1-score of 0.92, effectively classifying many of these events. However, it encounters considerable difficulties when dealing with accidents of low and medium intensity (Class 0 and Class 1). These classes precision, recall, and F1 scores exhibit a significant deficiency, suggesting a substantial frequency of misclassification. This highlights the need for further research and improvement in these areas. The overall accuracy is 0.85, principally influenced by accurately identifying high-severity accidents. The confusion matrix reveals that low and medium-severity accidents are often mislabelled as high-severity. This mislabelling results from a distinct disparity in class distribution, where high-severity accidents far outnumber the other classes. This skewed distribution significantly impacts the model's performance, highlighting a fundamental limitation.

	precision	recall	f1-score	support
0	0.02	0.00	0.00	4374
1	0.12	0.00	0.01	41542
2	0.85	1.00	0.92	268657
accuracy			0.85	314573
macro avg	0.33	0.33	0.31	314573
weighted avg	0.75	0.85	0.79	314573

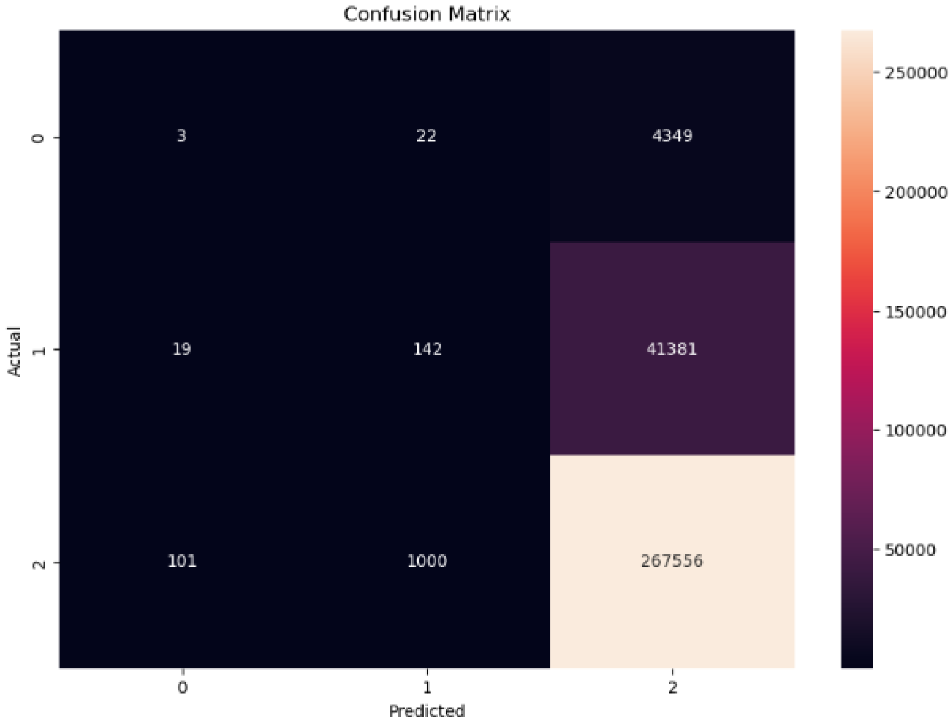

Figure 13.Model Performance

4.5 Scope of Air Analysis and Findings

The analysis thoroughly examines globalised air traffic incidents and accidents from Kaggle, and it determines the distribution of accident causes, scrutinises accident trends over time, investigates the association between accident severity and flight phases, and assesses the significance of various parameters in forecasting accident severity. The use of graphical representations such as pie charts and line graphs, along with a feature importance analysis using a Gradient Boosting model, ensures the thoroughness of the study. The insights gained, including the distribution of accident causes, trends in accidents over time, and the relationships between severity and flight phases, are invaluable in understanding air traffic incidents and accidents.

4.5.1 Analysis of Air Traffic Accident Trends

The study depicted in the pie chart highlights a notable disparity across various weather variables linked to these occurrences. The data illustrates that most air mishaps, around 86.9%, occurred in a particular weather state (presumably indicating clear or favourable weather). In contrast, just 13.1% occurred in a contrasting situation (presumably terrible weather). This implies that although unfavourable weather circumstances play a role in air mishaps, they are not as widespread as expected. The significant prevalence of accidents happening in seemingly ideal meteorological conditions underscores the necessity of considering additional factors, such as human mistakes (like pilot error), mechanical failure (like engine malfunction), or operational processes (like communication breakdown), in the incidence of air mishaps.

The occurrence of a significant proportion of accidents under non-adverse weather circumstances suggests that pilots and aircraft may encounter difficulties that are unrelated to the weather. Furthermore, it implies that aviation safety procedures and processes should encompass many parameters beyond just meteorological conditions, such as pilot training, aircraft maintenance, and operational protocols, to mitigate the likelihood of accidents effectively.

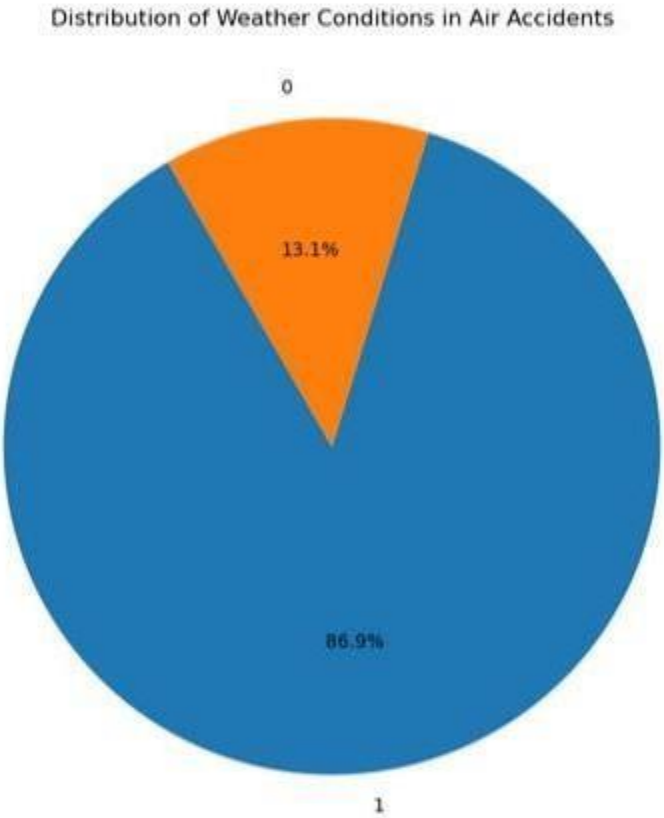

Figure 14.Air Analysis

- The pie chart below illustrates the distribution of flying phases in air accidents, revealing various mishaps occurring during flight periods. The graphic demonstrates that air mishaps are not limited to a single flight period but occur throughout multiple stages. Approximately 25.2% of accidents happen at a particular phase of the flight, such as landing or departure, which is particularly crucial and known to be high-risk. The second most significant group, representing 19.6% of the accidents, likewise suggests another critical phase of the flight, possibly the ascent or descent. This pattern indicates that the transitional stages of flight, such as take-off, climb, descent, and landing, are most susceptible to accidents. This observation is consistent with the overall findings on aviation safety in general. Additional stages, such as cruising, accounting for 15.0%, and the first or final approach, indicated by 13.6%, also substantially contribute to the total distribution of incidents. This emphasises that although the cruising phase is often considered safer because of the flight's stability, it still carries some danger, especially in situations involving mid-air crashes, mechanical breakdowns, or severe weather encounters. The remaining parts of the chart, which reflect smaller proportions, are likely associated with other phases, such as taxiing, ground handling, or specialised operations inside the flight. These phases contribute to the overall accident statistics, albeit to a lesser degree. The variation in the allocation underscores that each stage of flight entails distinct difficulties and hazards, necessitating targeted safety protocols customised for each phase.

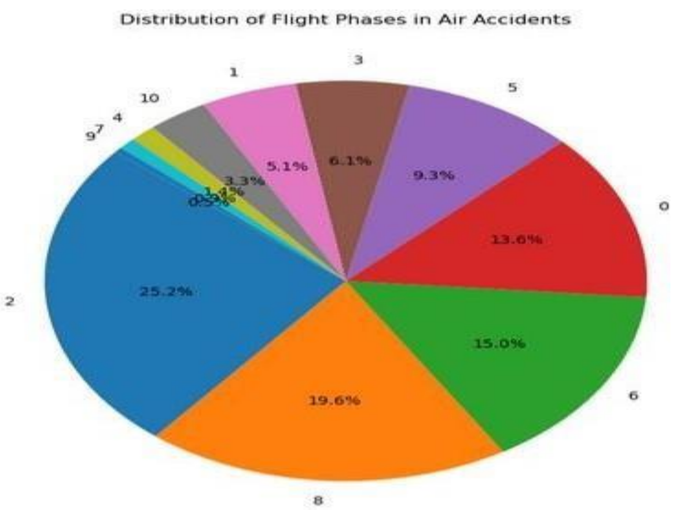

Figure 15.Flight Phases

- The line chart illustrating aviation accident frequency over time reveals several significant patterns. There was a substantial surge in accidents between the late 1990s and early 2000s, with a particularly noteworthy peak in 2000. This peak, a critical point in the data, may be linked to specific events or conditions that led to a temporary increase in accidents. Following this peak, the number of accidents plummeted, almost reaching zero from around 2002 to approximately 2006. This decline can be attributed to the successful adoption of enhanced safety protocols, legislative changes, or technological advancements, which have significantly reduced the occurrence of mishaps in recent years. However, beginning in 2006, there has been a substantial and swift surge in the number of accidents, with the pattern persistently escalating throughout 2007 and into 2008. The increase in accidents over this period may indicate a revival of risk factors or the emergence of new issues in the aviation industry. The rise in air traffic, probable safety procedure failures, older aircraft, or other operational problems could contribute to this spike. The figure demonstrates the cyclical nature of air accidents and underscores the critical need for ongoing vigilance and adaptation in aviation safety practices. The fluctuations observed during specific periods suggest that external events or industry changes can significantly impact accident frequency.

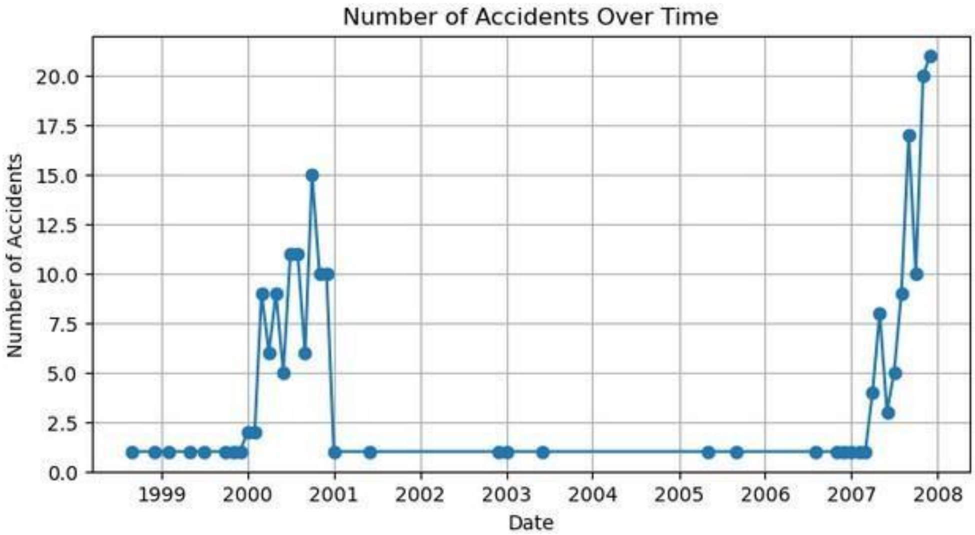

Figure 16.Flight Accidents Trend

4.5.2 Analysis of Air Model Evaluation

The results demonstrate the efficacy of four distinct machine learning models, Logistic Regression, Decision Tree, Random Forest, and Gradient Boosting, in accurately forecasting accident severity using various characteristics associated with air accidents. These findings validate the potential of machine learning in aviation safety. The performance of each model was assessed using commonly used measures such as accuracy, precision, recall, and F1 score. Additionally, an effort was made to compute the AUC ROC for multi-class classification.

The Logistic Regression model attained an accuracy of 86.05%, exhibiting a precision of 89.73% and a recall of 86.05%, yielding an F1 score of 83.05%. Unfortunately, the AUC ROC could not be computed due to a discrepancy between the number of classes in the proper labels (y_true) and the number of columns in the projected scores (y_score). The confusion matrix reveals that the model accurately predicted the primary classes but encountered challenges in correctly classifying the less prevalent minor courses.

```
Logistic Regression Performance:
Accuracy: 0.8604651162790697
Precision: 0.8972868217054264
Recall: 0.8604651162790697
F1 Score: 0.8305517555859553
AUC ROC could not be calculated: Number of classes in y_true not equal to the number of columns in 'y_score'
[[17  0  0  0  0  0  0]
 [ 0 14  0  0  0  0  0]
 [ 0  2  6  0  0  0  0]
 [ 0  0  2  0  0  0  0]
 [ 0  0  1  0  0  0  0]
 [ 0  0  0  0  0  0  1]
 [ 0  0  0  0  0  0  0]]
```

Figure 17.Logistic Performance

The Decision Tree model outperformed Logistic Regression, achieving an accuracy rate of 97.67%. What's more impressive are the precision and recall rates, both soaring at 100% and 97.67%, respectively. This led to a high F1 score of 97.67%, indicating the Decision Tree's significant role in accurately classifying the data and recognising almost all instances. Like logistic regression, the calculation of AUC ROC was impossible due to mismatched classes.

```
Decision Tree Performance:
Accuracy: 0.9767441860465116
Precision: 1.0
Recall: 0.9767441860465116
F1 Score: 0.9767441860465116
AUC ROC could not be calculated: Number of classes in y_true not equal to the number of columns in 'y_score'
[[17  0  0  0  0  0  0]
 [ 0 14  0  0  0  0  0]
 [ 0  0  8  0  0  0  0]
 [ 0  0  0  2  0  0  0]
 [ 0  0  0  0  1  0  0]
 [ 0  0  0  0  0  0  1]
 [ 0  0  0  0  0  0  0]]
```

Figure 18.Tree Performance

The Random98.45% forest performed strongly, achieving an accuracy of 95.35%, precision of 98.45%, and recall of 95.35%. This balance between precision and recall, resulting in an F1 score of 94.42%, makes the Random Forest a reliable classifier. The confusion matrix further supports this, showing minimal errors and suggesting the model's exceptional proficiency in performing the classification assignment.

```
Random Forest Performance:
Accuracy: 0.9534883720930233
Precision: 0.9844961240310078
Recall: 0.9534883720930233
F1 Score: 0.944186046511628
AUC ROC could not be calculated: Number of classes in y_true not equal to the number of columns in 'y_score'
[[17  0  0  0  0  0  0]
 [ 0 14  0  0  0  0  0]
 [ 0  0  8  0  0  0  0]
 [ 0  0  0  2  0  0  0]
 [ 0  0  0  1  0  0  0]
 [ 0  0  0  0  0  0  1]
 [ 0  0  0  0  0  0  0]]
```

Figure 19.Forest Performance

The Gradient Boosting model achieved the same level of performance as the Decision Tree, with an accuracy rate of 97.67% and flawless precision, recall, and F1 scores. These results demonstrate that Gradient Boosting, like Decision Tree, was highly proficient in predicting accident severity. Regarding the AUC ROC results, the calculation was constrained to select classes due to the problem's multi-class nature and class-specific distribution. In the classes where AUC ROC could be calculated, it achieved a perfect score of 1.0, demonstrating a flawless distinction between the positive and negative examples. However, inadequate data did not calculate the AUC ROC score for classes 5, 7, and 10. This highlights the complexity of the task, as these models did not have enough examples of these classes to calculate the AUC ROC score accurately. The Decision Tree and Gradient Boosting models displayed exceptional performance across all metrics, while the Random Forest model also showed robust performance. Although Logistic Regression was helpful, it did not match the robustness of the tree-based models. The AUC ROC results underscore the excellent classification performance, particularly in cases where it could be calculated. However, they also highlight the challenges faced when dealing with classes with limited samples, a common issue in multi-class classification problems. The results reaffirm the suitability of tree-based models, particularly Gradient Boosting and Decision Trees, for this classification task, providing reassurance to the safety of Aviation.

```
Gradient Boosting Performance:
Accuracy: 0.9767441860465116
Precision: 1.0
Recall: 0.9767441860465116
F1 Score: 0.9767441860465116
AUC ROC could not be calculated: Number of classes in y_true not equal to the number of columns in 'y_score'
[[17  0  0  0  0  0  0]
 [ 0 14  0  0  0  0  0]
 [ 0  0  8  0  0  0  0]
 [ 0  0  0  2  0  0  0]
 [ 0  0  0  0  1  0  0]
 [ 0  0  0  0  0  0  0]
 [ 0  0  0  0  0  1  0]]
```

Figure 20.Gradient Performance

5 Conclusion, Summary, Limitation, and Recommendation

5.1 Conclusion, Summary and Limitation

This study provides a comprehensive examination of traffic accidents in Scotland, explicitly emphasising occurrences occurring on both roads and in the air. Notable trends were recognised, including the impact of environmental factors, human behaviour, and governmental interventions on the frequency and severity of accidents. Using sophisticated machine learning models and Geographic Information Systems (GIS) has proven beneficial in precisely identifying locations with a high risk of accidents and forecasting patterns of accidents, offering essential knowledge for improving public safety.

Nevertheless, the study contains various constraints. It mainly depends on publicly accessible datasets, which may not completely describe the current dynamics of traffic safety. Moreover, the emphasis on Scotland restricts the applicability of the results to other areas. The machine learning models, although usually efficacious, exhibited certain constraints in correctly forecasting accidents of low and medium severity. These considerations indicate that future studies should broaden the geographical range, incorporate more up-to-date and varied data, and improve predictive models to increase accuracy.

To summarise, although the research greatly enhances its understanding of traffic safety in Scotland, there is an evident requirement for ongoing enhancements and more comprehensive investigations to tackle the intricacies of traffic accidents and safety measures in various settings comprehensively.

5.4 Recommendation

To increase traffic safety, policymakers should prioritise continuous infrastructure enhancements, particularly in areas with a high risk of accidents, and deploy adaptable safety measures to address different weather conditions. Enhancing public awareness efforts regarding safe driving behaviours and the hazards associated with distracted driving is crucial. In addition, it is critical to uphold rigorous adherence to air traffic safety rules. Frequent training for pilots to minimise human error is essential to ensuring the audience feels secure and protected. Utilising cutting-edge technology like machine learning and GIS for routine safety evaluations will yield more precise forecasts and facilitate proactive responses.

5.5 Future Work

While this study is subject to numerous constraints, it also opens exciting possibilities for future research. Broadening the geographical coverage, incorporating up-to-date and varied data, and boosting the accuracy of predictive models, including additional variables such as economic conditions and real-time traffic data, holds promise for further advancements in the field.

Moreover, while this research has effectively uncovered significant accident patterns, comparing more comprehensively with similar studies conducted in different geographical areas would be beneficial. This would help to contextualise the findings and highlight unique or standard patterns. While using GIS and machine learning models was a significant advantage, there is still room for improvement in the methodology. Further investigation into the existing safety protocols and their effectiveness could provide a deeper understanding for policymakers. Future research should overcome these limitations, refine the methodology, and extend the study's applicability to broader contexts.

References

[1] A. Bhawkar, “Severe Traffic Accidents in United Kingdom ,” 2018 [Online]. Available: https://www.researchgate.net/publication/330676135. [Accessed: 2023]

[2] C. A. Transport Scotland, “key-reported-road-casualties-scotland-2023,” Nov. 01, 2023 [Online]. Available: Reported Road Casualties Scotland 2022 (transport.gov.scot)

[3] D. K. Yadav and H. Nikraz, “Implications of evolving civil aviation safety regulations on the safety outcomes of air transport industry and airports,” 2014 [Online]. Available: https://www.tandfonline.com/doi/abs/10.3846/16487788.2014.926641

[4] M. J. Koetse and P. Rietveld, “The impact of climate change and weather on transport: An overview of empirical findings,” 2009, doi: 10.1016/j.trd.2008.12.004. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S136192090800165X

[5] J. Lewsey, H. Haghpanahan, D. Mackay, E. McIntosh, J. Pell, and A. Jones, “Impact of legislation to reduce the drink-drive limit on road traffic accidents and alcohol consumption in Scotland: a natural experiment study,” 2019 [Online]. Available: https://ueaeprints.uea.ac.uk/id/eprint/71606/

[6] H. Kharoufah, J. Murray, G. Baxter, and G. Wild, “A review of human factors causations in commercial air transport accidents and incidents: From to 2000–2016,” 2018 [Online]. Available: https://pdf.sciencedirectassets.com/271425/1-s2.0-S0376042118X00049/1-s2.0-S0376042117302154/main.pdf?X-Amz-Security-Token=IQoJb3JpZ2luX2VjEFAaCXVzLWVhc3QtMSJHMEUCIBGiz%2F1aLUBaswauWGg%2Bgqa2puijk9va1ddYW%2FkF-tUcjAiEA2BDAqfF7Za3J2byQMnBIrKekWyTRGlmz8mXcoVUjajkqswUIGRAFGgwwNTkwMDM1NDY4NjUiDOohKQa71FRdsyTFfyqQBcpGtrT5wtT7UsgiwNoc2v5KBYJT%2FoFOothDGZDzb5JxHKoMCn4wdru3Qo1uo6NyI1E5OqOOUdvnOcHtH1KJOHCeW-bbRzxJCyB7euS2Dicph0wCpnz0ULPQUC9L4rjzLzpzYatjN0qWtqODjiy%2BdOvOd96rXSx5vzVj5kbYltUcPPKBlcH0UgwBD33il%2FaYVqdDY6ZgZjKyNERCa-HUPC9HF8N6HDeSjVIfJd6PBXf%2BxX0g0XFJso6JrlcCWzUnhkfuZdZ7EWnlXdLBfkfeVbj9nw3fcjU0USNCR3IHE7pw2Iy02IjOhVH2NGel9QZPf5k77UUdGSH8YtvdFkRRoLt-lKvTlu99E0REC4EVaN4Y3eoUKREogfaYVUcJRd%2FsZOFc5mGjJ28oxGVbB9Tc2j1XfVyitpzoeY1OOI15f6tjCgKXLTk0lTLgLhfTCeG9TtE7Fs8XiTN%2FBL3S%2FcxVgHZllL0ZIE5vY2kXtvCiDd45SKFk%2FgkND2g%2B2bt%2FCmh0pQDni-LACbP2iiQM2IR%2BSR4%2F11A8r95uQrG1XKtsAT0yIY%2Bov4XmBTZP4k8Mpnmh9orxbMP2ccEiBt8w0ER%2FN3ciZxolmIpAMIroHToaH%2BBxyHM-ZOL3Hs1WeaFG6kBosXkpzjqikL2OtlKjHtMld2X05c7RvEyzubqPhn%2FbmOukstQZ9C0B%2Be8xQU5FIyHgvAxB2HFC0qaXPmibcJ%2BNyFHQZ%2FDc2vSNNjvEH-lKDHw2OyebhrTGGLLokdePgz8I863VPfr94kMvqQOCLTrDFFiPCF5T4thNKtnOzCItcIulch%2F78DbD2N%2B%2BDma2im-DFjg4tyRD0J4LXQPngBl7ViNLDYNw%2BcsYbcDV2W3vFY3eZuJlTsvMLiRsLQGOrEB8MvGFAwrdpOPdLo2tw%2FChu3JAZvM0RgHXVk27TRFoYbhka-ZhyDjFmIkPje%2FBa62w8TgcQvUrxRzFY2LR1H2RqX4nte1qyp5anD1DyAVgWfc2USkPlsw6ueq%2FkRHewlkhZ1tvJLu5Uny8DagmYi%2Fivd51ia16ez6T71GpEX-PeLrl7X%2FIVhRCe6OsJlG0QrPbJTgGo39StYJuHcFxZYi%2FI2lXqCp1J6drhSu524RXGeVoE&X-Amz-Algorithm=AWS4-HMAC-SHA256&X-Amz-Date=20240708T162727Z&X-Amz-SignedHeaders=host&X-Amz-Expires=300&X-Amz-Credential=ASIAQ3PHCVTYWR2ZWUSQ%2F20240708%2Fus-east-1%2Fs3%2Faws4_request&X-Amz-Signature=a99d195bd0d2fcde441e5cd208ce9b5d225a738bfdbaaa7ba16648a38b78ed8c&hash=b54da

84672ca6d5ebf4ba31e0090d7804584be7d13381f4779404002244a003a&host=68042c9435910
13ac2b2430a89b270f6af2c76d8dfd086a07176afe7c76c2c61&pii=S0376042117302154&tid=sp
df-9a9ec559-d881-47b2-8609-
0715991e6dc3&sid=2b20b3c4823c774f3539bd91b6692f56d16fgxrqb&type=cli-
ent&tsoh=d3d3LnNjaWVuY2VkaXJlY3QuY29t&ua=05045c0657575856548&rr=8a0171156f43
b316&cc=gb
[7] F. Racioppi, L. Eriksson, C. Tingvall, and A. Villaveces, Preventing Road Traffic Injury: A Public Health Perspective for Europe, World Health Organization, 2004. Available:
https://iris.who.int/bitstream/handle/10665/107554/9789289010931-eng.pdf?sequence=1&isAl-
lowed=y
[8] G. A. Shafabakhsh, A. Famili, and M. S. Bahadori, “GIS-based spatial analysis of urban traffic accidents: Case study in Mashhad, Iran,” 2017 [Online]. Available: https://pdf.sciencedirectas-
sets.com/312331/1-s2.0-S2095756417X00044/1-s2.0-S2095756417301988/main.pdf?X-Amz-
Security-Token=IQoJb3JpZ2luX2VjEHQaCXVzLWVhc3QtMSJIMEYCIQCUjXFgfvu-
DkKHJDvh7fjErM7mS2vAZOdAGzSmSGcN8tgIhAN2FdfLRisL6V1xlf%2BtJgrKRy9lRT1gR
C5T8FLX1Gl%2FTKrM-
FCE0QBRoMMDU5MDAzNTQ2ODY1IgylHMutTmR4SAw01NYqkAUUNOFtyy336Vlo1eY
CQ65xPnUU1IDEKF8ywHbwUHFJs4rR4Glwok%2FDsP4vZkg0r9rjKorlZBlUHB31NwMMT
p7TfVCb7SW3NA5hjyevPBgHtwlMu4EKbI995ix3a%2FTDJ%2F8Av36aXnsSVcYpQThM%2
BD46Xe1Oyy4ixMh%2FMefPg6w4F4C66hB6jM4nysczC8R5g3BsBvETb07xPIN-
WwZexZUPnrASNEbZIJSUl%2BPCW0VDguG%2Bwu%2FotX7TDk9dn%2BvciaFhF6QD8v
VgKiY8jJRVmzLryyRgX9Zway-
FKCw7RyEYb5ny37MxN6ZHONwOxH%2FpCmO1s7YO06LM6LO1BWJocLeQELzAkOwp
iaUOXd3BNMtYPto%2BgIJo0IeJnZsxuyN2REfGTJQrcDT0lP99beU-
WYCHxdnXDbo5wy%2FelTtWB%2FqO7p05rRDW0jhtcePTty8oeg1NKZ20K4UudRy5TXxU
WtEvqLTW04DP1gjBMGOyr7SjFnzv%2Be0ds2PASqo36QKxMzZbIFk3SAVM%2FkoUhlsK
puuxeCR3FdSvZ6apSHhtstHhFzuX2Vuha-
oMiJHcabl1ZlhynBs0KER%2BeBc9x%2FZ0%2BUzbNKKVvlN51PtKHV2lDEa4fJ9Ck3pslxq
0RJIQlT%2BBQaxYub3XMp5DGfgLflZVJIalhh8iNfPBIunOo8aTNuFF%2Bbuk%2FbwByNQ
OIxFNGhAB%2BGCGoSIClJMHPuleIYnOllWo56AhZ89P667suA8GIw24TRKK99a7Mnezli-
idO5bvu2w3plRaFbCpdwcnQ0MydfybO8%2FHOqvG%2FdbkBYaKKGsembaP%2B3Fxhtd7
VnQGvIYeZGIVy%2FBj%2FmH9tTktxk4vwNuqdImk5MEUOtNFXNYMBVnLO-
cEQjjCmsPC0BjqwAa7dUtLH02hexZrHWWVi77kmXEeETu3VJPCh87%2B%2FfSpq%2FOq
kUeiPGoslOgSOnjnH2eMaPgMX5ilyNAwYVYdoghqplmUMrteFyeput-
tizP2adakj%2Bzfq14cUhHeAhaVLr4i5Y%2BH6LCbZoW0M6jWFtaT5%2BeTN%2B%2B%2
F4iqNFRcfUs0YP3DqNgiOLqeXSr6nRTlBms-
gOYEJu%2FTkkQfGez2TCusvuI7PU96nXznd%2Fmgad8%2F37%2Bm&X-Amz-
Algorithm=AWS4-HMAC-SHA256&X-Amz-Date=20240720T211656Z&X-Amz-SignedHead-
ers=host&X-Amz-Expires=299&X-Amz-
Credential=ASIAQ3PHCVTY4QUD74ET%2F20240720%2Fus-east-1%2Fs3%2Faws4_re-
quest&X-Amz-
Signa-
ture=fe3f5eda2f83fd83402d19bb152be602f792563ed3467e56da553cc198484849&hash=19180
78c2bc491068e06e6caf209bff3dc7ae3b52109b49fc4d7e0faddee2673&host=68042c943591013
ac2b2430a89b270f6af2c76d8dfd086a07176afe7c76c2c61&pii=S2095756417301988&tid=spdf-
42a0b64f-a212-45c5-8e66-

4d49035ae3ec&sid=275f61a1101d704881295c363488d7412024gxrqb&type=client&tsoh=d3d3LnNjaWVuY2VkaXJlY3QuY29t&ua=05045c065153095e04&rr=8a65f9a59909beec&cc=gb

[9] J. Lewsey, H. Haghpanahan, D. Mackay, E. McIntosh, J. Pell, and A. Jones, "Impact of legislation to reduce the drink-drive limit on road traffic accidents and alcohol consumption in Scotland: a natural experiment study," 2019 [Online]. Available: https://ueaeprints.uea.ac.uk/id/eprint/71606/

[10] G. P. Santoso and D. Maulina, "Human error in traffic accidents: Differences between car driver and motorcyclist experiences," 2019 [Online]. Available: https://download.garuda.kemdikbud.go.id/article.php?article=1649026&val=15109&title=Human%20Errors%20in%20Traffic%20Accidents%20Differences%20Between%20Car%20Drivers%20and%20Motorcyclists%20Experience

[11] D. Antov, K. Abel, P. Sürje, H. Rõuk, and T. Rõivas, "Speed reduction effects of urban roundabouts," 2009 [Online]. Available: https://bjrbe-journals.rtu.lv/bjrbe/article/view/1822-427X.2009.4.22-26/111

[12] M. Maher, "Statistical Advice: Scottish safety camera programme," 2013 [Online]. Available: https://thenewspaper.com/rlc/docs/2013/uk-scotcritique.pdf

[13] J. Freeman, B. Scott-Parker, I. Wong, and N. Haworth, "Vulnerable road user groups: a review of younger drivers, motorcyclists and older drivers," 2012 [Online]. Available:https://www.tandfonline.com/doi/epdf/10.3402/vgi.v3i0.14889?needAccess=true

[14] C. V. Oster Jr., J. S. Strong, and C. K. Zorn, "Analyzing aviation safety: Problems, challenges, opportunities," *Research in Transportation Economics*, vol. 43, pp. 148-164, 2013. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0739885912002053

[15] E. T. Wong and W. Y. Man, *Smart Maintenance and Human Factor Modeling for Aircraft Safety*. Springer, 2023 [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-031-21232-1_2

[16] J. K. Evans, "An examination of aviation accidents associated with turbulence, wind shear and thunderstorm," 2013 [Online]. Available: https://ntrs.nasa.gov/api/citations/20130013459/downloads/20130013459.pdf

[17] M. Feng, J. Zheng, J. Ren, and Y. Liu, "Towards big data analytics and mining for UK traffic accident analysis, visualization & prediction," presented at the Proceedings of the 2020 12th International conference on machine learning and computing, 2020 [Online]. Available: https://dl.acm.org/doi/pdf/10.1145/3383972.3384034?casa_token=4yvi-NuXPnWwAAAAA:ZrnQRqFx9qRfxgljEYWdhnd8uho0xDyC2aonWRtCZlkSRF1KUKfquoIKs3Hb8aMm1HFzZJhjva6jgQ

[18] D. McCarty and H. W. Kim, "Risky behaviors and road safety: An exploration of age and gender influences on road accident rates," 2024 [Online]. Available: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0296663

[19] M. J. Koetse and P. Rietveld, "The impact of climate change and weather on transport: An overview of empirical findings," 2009, doi: 10.1016/j.trd.2008.12.004. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S136192090800165X

[20] A. Linder *et al.*, "Methods for the evaluation of traffic safety effects of Antilock Braking System (ABS) and Electronic Stability Control (ESC): A literature review," 2007.

[21] K. Bengler, K. Dietmayer, B. Farber, M. Maurer, C. Stiller, and H. Winner, "Three decades of driver assistance systems: Review and future perspectives," 2014 [Online]. Available: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6936444

[22] A. Razzaghi, A. Bahrampour, M. R. Baneshi, and F. Zolala, "Assessment of trend and seasonality in road accident data: an Iranian case study," 2013 [Online]. Available: Assessment of Trend and Seasonality in Road Accident Data: An Iranian Case Study (ijhpm.com)

[23] C.-H. Yang, Y.L. Alex, “Did technology improve safety? An empirical study of controlled flight into terrain accidents,” no. 1, 2015 [Online]. Available: https://jats.aviationsociety.gr/index.php/JATS/article/view/62/49
[24] L. Di Milia, M. H. Smolensky, G. Costa, H. D. Howarth, M. M. Ohayon, and P. Philip, “Demographic factors, fatigue, and driving accidents: An examination of the published literature,” 2011 [Online]. Available: https://pdf.sciencedirectassets.com/271664/1-s2.0-S0001457510X00082/1-s2.0-S0001457509003327/main.pdf?X-Amz-Security-Token=IQoJb3JpZ2luX2VjENv%2F%2F%2F%2F%2F%2F%2F%2F%2FwEaCXVzLWVhc3QtMSJHMEUCIQDbzqsmdXSlzG3JK4GQNGougPbHUFI6MHQvynexyVyiAIgKQHQr9HLjgCRyIL9lUlB1rZmKttxw%2FNIDUIAfUG459MqvAUIk%2F%2F%2F%2F%2F%2F%2F%2F%2F%2F%2F%2FARAFGgwwNTkwMDM1NDY4NjUiDLd%2FIkCSdks%2FOxVEjiqQBWQeO8M%2F5X6trpR0iJNp8Z4PJGzej0SC23OEghkl5B1FPz8OE64Q6l%2B%2FvlOuHxrlU6d07DULmiseZlJ2PnRm2pN%2Bm9gP5siqV5hUj1L0ssRwBozJKZuCh%2FflxCqYpU5ab%2B3q7lGTqZfEmvLEH9CynHCJ72x5H3R62a3ERXA9kdvTWvzC%2FvumsBBZiUaUdHsGZuUJXS%2BR1nc6f%2FPa0pDfmowjiIDY%2B31cegiMRUnPs%2FxFnqHqnrBHWETPjWSNg6ml5YpfAeqZW1QX7PRSFd9N1rsdfe%2BvGEVks2%2BNFvzu-frfSYo7rL5u1ylLQK4WSCO%2BHFYpXFHAhasAnLgmVgpm4Qk1rzXbcAnax%2BprOplfZVd5u9KgrsFE1quFUVAdn1djU9atv2H9EfNhLQlWtrsEB-wElMo9jhU58XSeDIG0nRrYMU%2BYQSJXK8K6SR%2FXHJuw0BH0vCUlvfYfxo1Uq19UKAXfesgENbpOQamHP1VMCu78tbZ%2FpSqMg2Z2DoGkP5YsJAAJNYLenDY-BUELPLXKeQweZaHu1zTAwDn%2BREVaWlonbIusEDAV8MiaBkRexWefhwCRIzi4PwuK%2B3%2BATLoTIR-WAKCdq50VoUFvgj7qtkJLyqy3SslH93BM1Ookon5opnFh2KHafyt54%2BqLVDftvpIvWxgLCeTtb2O2FGepMSV8iHXVxSnH2BRmrPj6zbGILogeL2UUsbz4gM6J3S4xRtujgk3lat-fSejKBzPd2pRA87OIA6bFpfgfEYuLEpPP1VhaFejTS6eUKOMjeUFKsgxnp6P3AbYqwugBJUxpfrE2cwqtslyShMgMOFiGiOy%2F%2FtoQiJY2D3%2BesYoM0cm%2BgdDqkjCPBL1HNI-yWaKvxieHW0cXadMK6plrQGOrEBMw1914T%2B5OqqYEna6CHmmshiI88bickd6cUO4v3tBTvYw9hbxsmMUdM8jlO-FeFSJeMw0XpmY7milGLTC2bT8P8Kh7xF08L%2FNSTmQkD0zDbfjd5D9EbYyhHL5mZqgcBrfuS9hM3kOilBvPIZ57qE7fiTQIJjm8gKabJY-YifdWQQ8FlGD%2FBtY42saCa%2Bdza7drhURHBMz65lPWbd1uJZYfk3fyU9tu4PRwnkhcc1WIdAj3&X-Amz-Algorithm=AWS4-HMAC-SHA256&X-Amz-Date=20240703T190337Z&X-Amz-SignedHeaders=host&X-Amz-Expires=300&X-Amz-Credential=ASIAQ3PHCVTYVH7K2QSY%2F20240703%2Fus-east-1%2Fs3%2Faws4_request&X-Amz-Signature=6f9f7c4684a9791bb3588f2c30f8e6d487adbd736ed904f8627b6079c3d3d90c&hash=394699d0faeb1b7fdf425c89f7e7566f4892674b38b48911decb582280afced3&host=68042c943591013ac2b2430a89b270f6af2c76d8dfd086a07176afe7c76c2c61&pii=S0001457509003327&tid=spdf-3dbdd199-cf01-4842-a8e3-bda735bb67a7&sid=444981a01dd3984cdc-b73e-24864048c65bgxrqb&type=client&tsoh=d3d3LnNjaWVuY2VkaXJlY3QuY29t&ua=05045c5e035f5d5503&rr=89d922facbbebec3&cc=gb&kca=eyJrZXkiOiJGa1k3WGQ4WGpzN3hTcTBE-djZwUzBaK2drRks4a1B0NlB1ZjRGTTZLRUNyL3JMQ1Ira1J3Z1NaNzR6QjhGU09WWFhHaUt6K091TS85OFZraEpidG-ZuWVhhN0ZNUGhuQmFmRCs3ZzFCa0NvbUc0NUdjc3JQZHVGc1NUTmM3VGpGSE1tNkpKOUpJQUJpVmpHWUErSFg1T2Y0aUdlVXBsa0lqelBUaFU2a21pU0RWZXN3PSI-sIml2IjoiNmQxMzc0Y2JjM2M2YTc3ZGU5MDNiM2ZhNjJjZjg0NTIifQ==_1720033466776
[25] D. D. Clarke, P. Ward, C. Bartle, and W. Truman, “Young driver accidents in the UK: The influence of age, experience, and time of day,” 2006 [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0001457506000285
[26] A. F. Williams, “Young driver risk factors: successful and unsuccessful approaches for dealing with them and an agenda for the future,” 2006 [Online]. Available: https://injuryprevention.bmj.com/content/injuryprev/12/suppl_1/i4.full.pdf
[27] Air Accidents Investigation Branch (AAIB), "Annual Safety Review," AAIB, 2021.

[Online]. Available: https://www.gov.uk/government/organisations/air-accidents-investigation-branch
[28] M. S. Amin, J. Jalil, and M. B. I. Reaz, “Accident detection and reporting system using GPS, GPRS and GSM technology,” presented at the 2012 International Conference on Informatics, Electronics & Vision (ICIEV), 2012 [Online]. Available: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6317382
[29] V. S. Kushwaha, D. Yadav, A. Topinkatti, and A. Kumari, “Car accident detection system using GPS and GSM,” 2015 [Online]. Available: https://pnrsolution.org/Datacenter/Vol3/Issue3/135.pdf
[30] S. S. Mahmud, L. Ferreira, M. S. Hoque, and A. Tavassoli, “Micro-simulation modelling for traffic safety: A review and potential application to heterogeneous traffic environment,” 2019 [Online]. Available: https://pdf.sciencedirectassets.com/280270/1-s2.0-S0386111219X00023/1-
s2.0-S0386111217302133/main.pdf?X-Amz-Security-Token=IQoJb3JpZ2luX2VjE-
HUaCXVzLWVhc3QtMSJHMEUCICNhckOKCZhKxC0wEdchfuus0Qs5gc1Bs%2FRqDrRIU
C%2FDAiEAujvxeglwiLOPmdXh38fKpASwTB8iE04tGC%2Fjl14E12oqswUIThAFGg-
wwNTkwMDM1NDY4NjUiDDgPfzBAJi90C7DDViqQBXy%2F5oiOgfCsOgtHpFPifEFucSpg
HZeHWmFnCW2QYdPQndJVc1B6xh4OpC%2FVPF5P9oedKlW-
brOATAW1U9FL9fYtKjBXcJ3hoeX2Pyz9lahuPDUeiBaJ2f%2FqAIH%2B2TCYXhEkautsrgU
o8aA1EcoF1POwQluLHoYQAUSvJj7WcspbKLwnC14OPUm1gyiPjsDOriVbot-
lAa8sV0STlu6Rc6otsG9leEyu0MGlAUm4as9YB9sonHupztyWEHLEUNjNHGNd%2BrEAlD
L0BLN0OJ7v024OHXuNKzshcaBzJ%2BGGRwdM6CXfMx4DnvAHeM1Bd1Sz31TjxgtlskR
m9o7Na0IGTc2Baqq4GUwHnNC6b8eme4ZLg%2Bk0LY5bl0ijPuTzmKjlFw8o7JpBtMiz0Rse
BxtVxn%2FuLt-
doLPfM3O%2FMK80x9LpN%2B6iCVI2bVduJqexL9GxknJTMOrNC8HPZmPBxJ%2FSv7eS
NFH0U3AFYHHss-
wFPXH2ciq9hylvff78sFcEW8VcmNMlbmYBZE0oEXew8%2FDQA%2Ba5Igt0jQ6FHQ%2B
HCAjGGhNNIl0ZCUB2FJmU5Raqz6dex-
ADXpsbAqKuqEe2kG9ZC9ML0YSNsZyvc%2Bz2wdmE2lKzYR7472mFDmuBZgkp1k%2Bb
6s9MhSNcufBF%2BuZMpFnVeBSTkE6R8iYgUrGIUL3RecuhhEx-
HaTNr%2BNKsw792fbKafigrn1BW4yY81wMhnsyBCMKGC5raaMaogKN34YLY%2BEnFm
1mRlmnO8jRRrXB5U%2BzNKjyNpH1iqjreD%2FzyiKs7b%2FDtCdePpZXYTgH-
dlv7S173idhwXaCp841wNNPLVgsD8sRCppnWzGPuYCyN2GNBSPoihxDT695HUVZ3NZU
BVbAhvxCoaTMLO78LQGOrEBRpM4GBnixlwwQNobE3MNYZ5vPCSCRtacgDZMmNY-
crDsLP9FS4ST19DzUJfeHdxn4SAidVVz4wYyIq%2BhOjHYlXsqnENHOnLklnbazdoBwRD
X4Qe4tXxpY%2FxS3x0Sx19uXgtk9jTHCdnPHuYnArU0sbn-
ByXkqqgcZIqx4KO5pHK4HSsk48pkblUZ7HG7jSOPtZPbAQjhsSSi%2FgQYDS9yx75lTFlH
H%2B%2B9tuDJyxn5hT4IgH&X-Amz-Algorithm=AWS4-HMAC-SHA256&X-Amz-
Date=20240720T213810Z&X-Amz-SignedHeaders=host&X-Amz-Expires=300&X-Amz-Credential=ASIAQ3PHCVTYQOHEPGKO%2F20240720%2Fus-east-
1%2Fs3%2Faws4_request&X-Amz-Signature=42104cfe55dd4203629557a9e8f7f36cab4caeab73af30a61fffebb625a633bd&hash=1d07fcf
8f991dc48f47f54dc35f1e781d869d787c485a97f5a650576b671c052&host=68042c943591013ac
2b2430a89b270f6af2c76d8dfd086a07176afe7c76c2c61&pii=S0386111217302133&tid=spdf-
37fc71f9-4a41-4fef-bfac-
931210d73e95&sid=275f61a1101d704881295c363488d7412024gxrqb&type=client&tsoh=d3d3LnNjaWVuY2VkaXJlY3QuY29t&ua=05045c0651505e5f07&rr=8a6618bd9b6d
4596&cc=gb
[31] P. Sharma, "Aviation Accidents and Incidents (NTSB, FAA, WAAS)," Kaggle, Accessed: Jul. 27, 2024. [Online]. Available: https://www.kaggle.com/datasets/prathamsharma123/aviation-accidents-and-incidents-ntsb-faa-waas?select=airline_accidents.csv
[32] T. Tsiaras, "UK Road Safety: Traffic Accidents and Vehicles," Kaggle, Accessed: Jul. 27, 2024. [Online]. Available: https://www.kaggle.com/datasets/tsiaras/uk-road-safety-accidents-and-vehicles?select=Accident_Information.csv
[33] “Pandas Documentation”, Available: https://pandas.pydata.org/docs/user_guide/index.html.

[34] “Scikit-learn: Machine Learning in Python”, Available: https://scikit-learn.org/stable/user_guide.html.
[35] “NumPy Documentation”, Available: https://numpy.org/doc/stable/user/index.html.
[36] “Matplotlib: Visualization with Python”, Available: https://matplotlib.org/stable/users/index.
[37] "XGBoost Documentation," *XGBoost: Scalable, Portable, and Distributed Gradient Boosting (GBDT, GBRT or GBM) Library*, Available: https://xgboost.readthedocs.io/en/stable/.
[38] “Plotly Documentation”, Available: https://plotly.com/python/getting-started/.
[39] “Seaborn: Statistical Data Visualization”, Available: https://seaborn.pydata.org/installing.html#quickstart.
[40] “GeoPandas Documentation”, Available: https://geopandas.org/en/stable/getting_started/introduction.html.
[41] “Folium: Python Data, Leaflet.js Maps”, Available: https://python-visualization.github.io/folium/latest/getting_started.html.
[42] W. McKinney, *Python for Data Analysis*, 2nd ed. Sebastopol, CA: O'Reilly Media, 2017.
[43] P. Bruce and A. Bruce, *Practical Statistics for Data Scientists*, 1st ed. Sebastopol, CA: O'Reilly Media, 2017.

Click here to enter text.