

School of Built Environment, Engineering and Computing

Leeds Beckett University

**Visualisation of Road Traffic Accidents in the UK**

By: Suman Muthukumaran

Mark Dixon

Submitted to Leeds Beckett University in partial fulfilment of the requirements for the degree of MSc Data Science

**May 2024**

# Candidate’s Declaration

I, Suman Muthukumaran confirm that this dissertation and the work presented in it are my own achievement.

Where I have consulted the published work of others this is always clearly attributed;

Where I have quoted from the work of others the source is always given. With the exception of such quotations this dissertation is entirely my own work;

I have acknowledged all main sources of help;

I have read and understand the penalties associated with Academic Misconduct.

Signed: Suman Muthukumaran

Date: 08/05/2024

Student ID No: 77338335

# Acknowledgements

I would like to express my sincere thanks to the distinguished administration and professors at Leeds Beckett University for their outstanding facilities, cutting edge labs, and consistent assistance, all of which were essential to the successful completion of this dissertation.

The sincerest thanks go to Mark Dixon, who has been an inspiration to me throughout this project with his priceless advice and steady backing.

Special thanks go out to Thalita Vergilio and Kiran Voderhobli, whose knowledgeable advice and steadfast support were crucial in helping me overcome challenges and complete the dissertation successfully.

I owe a debt of gratitude to the distinguished Computer Science department faculty members for their constant support and unmatched mentoring during my master's programme, which has greatly influenced my academic career.

TABLE OF CONTENTS

[Candidate’s Declaration 2](#_Toc166103774)

[Acknowledgements 3](#_Toc166103775)

[Abstract 6](#_Toc166103776)

[Chapter 1: Introduction 7](#_Toc166103777)

[Overview 7](#_Toc166103778)

[1.1.1 Traffic and Road Accidents 7](#_Toc166103779)

[1.1.2 Data Visualization 8](#_Toc166103780)

[1.2 Rationale 8](#_Toc166103781)

[1.3 Existing Methods 9](#_Toc166103782)

[1.4 Aim and Objective 9](#_Toc166103783)

[1.5 Outline 10](#_Toc166103784)

[Chapter 2: Literature Review 10](#_Toc166103785)

[2.1. Utilizing ML Predictions for Accident Analysis 11](#_Toc166103786)

[2.1.2. Clustering with additional ML algorithms for accident investigation 12](#_Toc166103787)

[2.1.3. Employing Different Machine Learning Algorithms for Accident Analysis. 12](#_Toc166103788)

[2.2 Utilizing spatial analysis for accident analysis. 12](#_Toc166103789)

[2.3. Utilizing data visualization and statistics for accident analysis. 13](#_Toc166103790)

[Chapter 3: Methodology 14](#_Toc166103791)

[3.1 Collecting and Preparing Data 15](#_Toc166103792)

[3.1.1 Analysis of Casualty Data: 2021-2022 15](#_Toc166103793)

[3.1.2 Analysis of Merged Collision Information: 2021–2022 15](#_Toc166103794)

[3.1.3 Analysis of Combined Vehicle Data: 2021-2022 15](#_Toc166103795)

[3.2 Visualization Methodology: Advanced Techniques and Tools 16](#_Toc166103796)

[3.3 Detailed Overview of Dashboard Metrics: 16](#_Toc166103797)

[3.3.1 Interactive Severity-Based: 16](#_Toc166103798)

[3.3.2 Gender Based Analysis 18](#_Toc166103799)

[3.3.3. Comparative Dashboard – No group or Calculation used. 18](#_Toc166103800)

[3.3.4. Time Analysis Dashboard - No group or Calculation used. 18](#_Toc166103801)

[3.4 Benefits and Limitations of the Implemented Approach: 18](#_Toc166103802)

[3.5 Ethical Considerations: 19](#_Toc166103803)

[Chapter 4: Product/Research Design and Implementation 19](#_Toc166103804)

[4.1 Excel for Data Storing: Why? 19](#_Toc166103805)

[4.1.1 Keeping Information in a Database: 19](#_Toc166103806)

[4.1.2 Keeping Information in an Excel: 19](#_Toc166103807)

[4.2 Data Fabrication 19](#_Toc166103808)

[4.3 Granting input to Tableau. 20](#_Toc166103809)

[4.4 Dashboards 20](#_Toc166103810)

[4.4.1 Things to Take into Account About While Creating a Dashboard 20](#_Toc166103811)

[4.4.2 Dashboard and Design 21](#_Toc166103812)

[4.5 Dashboard Contents 21](#_Toc166103813)

[Chapter 5: Analysis and Discussion of Research Findings 23](#_Toc166103814)

[5.1.1 Recognising Gender Differences in Traffic Accidents 23](#_Toc166103815)

[5.1.2 Understanding Age Patterns in Traffic Accidents 23](#_Toc166103816)

[5.1.3 Examining Accident Intensity by Age Group 24](#_Toc166103817)

[5.2.1 Investigating Inequalities in Gender in Traffic Situations 25](#_Toc166103818)

[5.2.2 Examining Gender Differences in Lighting 25](#_Toc166103819)

[5.2.3 Examining Gender Differences in Traffic Situations 26](#_Toc166103820)

[5.3.1 Examining Historical Patterns of Accident Severity 27](#_Toc166103821)

[5.3.2 Examining Daily Patterns in the Severity of Accidents 27](#_Toc166103822)

[5.3.3 Looking at Monthly Accident Severity Trends 28](#_Toc166103823)

[5.4 Evaluating Increases in the Severity of Traffic Accidents: A Comparison of 2021 and 2022 29](#_Toc166103824)

[Chapter 6: Project Management 31](#_Toc166103825)

[6.1 Initial Research and Topic Selection 32](#_Toc166103826)

[6.2 Data Preparation and Literature Analysis 32](#_Toc166103827)

[6.3 The Stage of Intermediate Visualisation 32](#_Toc166103828)

[6.4 The Last Stage of Visualisation 33](#_Toc166103829)

[Chapter 7: Conclusion and Future Work 33](#_Toc166103830)

[7.1 Summary of the Study 33](#_Toc166103831)

[7.2 Recommendations 34](#_Toc166103832)

[7.3 Restrictions on the Research 34](#_Toc166103833)

[7.4 Future Work 34](#_Toc166103834)

[References 35](#_Toc166103835)

[Appendices 37](#_Toc166103836)

## 

## Abstract

The well-known data visualisation software Tableau is an effective instrument for examining a wide range of datasets from different fields. The Department for Transport in the United Kingdom (UK) provided accident data for the years 2021–2022, which was analysed in this study using Tableau. Through the integration of datasets related to accidents, vehicles and casualties, this study explores the fundamental causes and patterns of accidents. Important insights are revealed through the development of interactive dashboards, such as those that are Severity-Based, Gender-Based, Comparative, and Time-Based. Through a review of 263,743 incidents over a two-year period, it can be observed that men are more likely to be involved in accidents, especially when there is good lighting. Automobiles are the most common kind of vehicle involved in collisions, and human error is the main contributing factor. The results highlight the significance of increased awareness and safe driving practices in reducing accident rates, which goes beyond simple traffic control strategies.

# Chapter 1: Introduction

## Overview

### 1.1.1 Traffic and Road Accidents

Due to their wide availability on roads, cars have been found to be involved in a disproportionately greater number of accidents when compared to other vehicle types. This phenomenon can be explained by the vast number of automobiles that use road networks compared to other forms of transportation. A crucial example that quickly and thoroughly explains the nuances of accident incidents is Figure 1.1.

A graph of a vehicle comparison

Description automatically generated

FIGURE 1.1 ROAD FATALITIES DURING 2021 AND 2022

When it comes to understanding complicated information, visualisation proves to be a crucial tool since it allows for quicker understanding and analysis than traditional statistical techniques. Its effectiveness comes from its capacity to convert unprocessed material into formats that are easy to read, which improves cognitive absorption. When one looks closer, one finds that the term "visualisation" refers to a variety of applications, from interactive dashboards to graphical representations, all of which are customised to meet the needs and preferences of individual analysts.

This multimodal technique facilitates deeper insights into complex datasets while also speeding up information interpretation.

### 1.1.2 Data Visualization

It is the standard technique for effectively communicating complicated information by making skilful use of graphical components including charts, graphs, maps, and interactive tools. There are several benefits to this multimodal method to information sharing, which are explained below:

Storytelling: Data visualisation captivates viewers by creating tales via the patterns and colours inherent in the visual representations, much as the appeal of brilliant hues in fashion, architecture, and art. This narrative quality gives the data a tangible context, which promotes greater understanding and interaction.

Accessibility: Rendering raw data in visual representations allows it to pass through cognitive and language obstacles, making a wide range of people able to access it. By guaranteeing that data-driven insights are understandable to a wide range of stakeholders, this democratisation of information promotes inclusion and transparency.

Identification of Patterns: Visualisation is an effective technique for identifying complex patterns and trends that are present in datasets. Data-driven decision-making is facilitated when analysts can quickly detect connections, anomalies, and emerging phenomena by condensing complicated data into understandable visualisations.

Furthermore, the field of data visualisation comprises an extensive number of approaches and procedures, each customised to meet unique analytical goals and modalities. These consist of but are not restricted to:

Pie Charts: Pie charts are a great way to show proportionate connections between categorical data and provide an eye-catching visual representation of distribution patterns.

Bar charts: Known for their adaptability and simplicity, bar charts are a powerful tool for communicating trends and comparisons across discrete categories or time periods.

Maps: Spatial visualisation of data is made easier by geographic information systems (GIS), which help analysts identify spatial patterns and linkages.

Clustering: Primarily utilised in exploratory data analysis, clustering methods divide information into coherent groups according to fundamental similarities or differences.

Dashboards: Users may explore and query data in real-time via interactive dashboards, which combine several visualisations into unified interfaces.

Scatter Plots: Scatter plots are well recognised for their effectiveness in displaying correlations between two continuous variables by contrasting data points to reveal underlying linkages.

Line Graphs: Used in temporal data analysis, line graphs show oscillations and trends throughout time, making longitudinal comparisons and trend interpretation easier.

## 1.2 Rationale

Road accidents have a wide range of effects on society as well as the persons who are directly affected and their families. To reduce the likelihood of accidents and lessen their effect, a careful investigation is required to determine the financial and societal costs associated with them.

Even if the most current validated data relates to events that occurred in 2022, it is crucial to place this information in the perspective of larger historical trends and patterns. Over time, there are variations in accident rates due to a variety of factors, including changes in driving behaviours, infrastructural advancements, and economic swings.

An extensive analysis of accident data over several years is essential to extract useful insights and develop focused solutions. By carefully examining data from 2021 and 2022, investigators can find reoccurring patterns, broad trends, and underlying causes of accidents.

This extensive approach not only improves our awareness of accident dynamics but also gives stakeholders and policymakers the ability to create evidence-based plans for raising road safety. The goal of this project is to promote a more secure and resilient transportation environment by using advanced data analytic methods and smart interpretation.

## 1.3 Existing Methods

Numerous approaches used in accident analysis have been made clear by extensive study, highlighting the multidisciplinary character of this discipline. Studies by (Wahab,2019; Wen,2021; Kushwaha,2021; Geyik,2020) show that machine learning (ML) algorithms have become effective tools for identifying the complex elements impacting accident severity. These projects make use of cutting-edge computational methods to reveal subtle patterns and correlations in big datasets, providing priceless insights into the fundamental causes of accidents.

Simultaneously, methods of spatial analysis have attracted a lot of interest due to their effectiveness in identifying accident hotspots and clarifying spatial patterns of risk. The use of spatial analysis in detecting regional clusters of accidents has been shown by (Soltani, 2017; Lakshmi, 2019; Hazaymeh,2022). This information informs targeted actions and resource allocation methods to reduce risk and improve road safety.

Furthermore, by providing simple visual representations of complex data structures, the incorporation of data visualisation techniques has completely changed accident analysis. Research conducted by (Sunkpho,2020; Lokala,2017; Sodikov,2018) demonstrate how visualisation approaches may significantly impact decision-making and policy development by revealing patterns, anomalies, and causal links in accident datasets.

It aims to determine the best methodology for accident analysis by synthesising the body of available literature. It attempts to provide a strong analytical framework capable of revealing the complex dynamics of traffic accidents and guiding evidence-based initiatives to improve traffic safety by utilising insights from a variety of backgrounds.

## 1.4 Aim and Objective

Aim:

Analysing several contributing variables to road accidents in the UK in 2021 and 2022 is the goal of this research.

Objectives:

• To carry out a critical review of literature on current studies in the field and to develop an in-depth knowledge of it.

• To gather and clean up UK accident data for the 2021 and 2022 financial years.

• To look for any clear patterns or connections between the severity of traffic events and related demographic traits.

• To examine how environmental elements, such the state of the weather and the quality of the road system, affect the number and severity of accidents.

• To investigate the efficacy of current traffic safety protocols and pinpoint opportunities for enhancement through the analysis of accident data.

• To look at the relationship between the time of day, day of the week, and accident frequency to develop focused intervention plans.

## 1.5 Outline

The format of the dissertation is as follows:

Chapter 2: Literature Review This chapter explores a collection of literature related to accident analysis. It looks at several research approaches and discusses the advantages and disadvantages of each to determine which is best for accident analysis.

Chapter 3: Methodology This section explains the project's selected methodology and the reason behind it. The chapter describes the development process in depth, highlighting the advantages and disadvantages of the selected strategy.

Chapter 4: Implementation and Management of Data This section describes how the data is handled and how the project is then put into action. It gives a general overview of the steps involved in creating and utilising analytical tools as well as data processing techniques.

Chapter 5: Results and Interpretation The outcomes of the project are shown and analysed in this chapter. The relevance and significance of the results are brought out in the extensive discussion that follows, with a focus on the analysis.

Chapter 6: Viewpoints on Project Management The project's management during its development is covered in this section. Obstacles that arise during the procedure are discussed, as well as the tactics used to overcome them.

Chapter 7: Final Thoughts and Prospects for the Future The last chapter provides a thorough synopsis of the whole project, summarising the main conclusions and realisations. It also suggests possible directions for further study and project-related enhancement projects in the future.

Chapter 2: Literature Review   
  
The main goal of this chapter is to analyze and discuss important research projects related to accident analysis in detail. These studies have analysed accidents using a wide range of approaches. These include using Geographic Information System (GIS) technology to identify hotspots, utilizing Machine Learning (ML) for predictive accident modelling, utilizing time series analysis for accident occurrence forecasting, carrying out thorough statistical analyses, utilizing sophisticated visualization techniques, and more.  
  
Why is this relevant to my project?  
  
Accident analysis research offers the fundamental information and techniques necessary for producing precise and perceptive visualizations, it is extremely relevant to visualize traffic incidents in the UK. Visualization tool fully depicts the complex nature of road safety concerns by having a thorough understanding of the components that contribute to accidents, such as weather patterns, road infrastructure, and driver behaviour. Prioritizing visualization approaches, choose pertinent data variables to visualize, and spot important patterns and trends in accident data by utilizing research findings. Furthermore, research provides useful strategies that improve the analytical capabilities of your visualization tool, such as spatial analysis techniques and machine learning algorithms.

Through improved understanding of accident frequency, hotspots, and risk factors, stakeholders may make more informed decisions and implement focused interventions to increase road safety. In the end, including research results into your visualization project guarantees that it will be successful in tackling UK road safety issues by giving stakeholders the information and understanding they need to encourage safer roads and lower accident rates.  
  
Why does it influence me?

Choosing to start on the project of visualizing road accidents in the UK is influenced by the insightful and useful approaches that accident analysis research offers. Research shows the need to tackle road safety issues in a comprehensive way by clarifying the elements that lead to accidents. This knowledge inspires us to develop a visualization tool that effectively conveys the complicated nature of traffic incidents, supporting initiatives to encourage safer driving and lower the number of accidents in the UK.

## 2.1. Utilizing ML Predictions for Accident Analysis

2.1.1 Forecasting Accident Severity  
  
To forecast the severity of injuries in traffic accidents, research by Sameen and Pradhan, 2017 provides a deep learning model that uses a Recurrent Neural Network (RNN) and 1130 accident records from 2009 to 2015. Within the TensorFlow framework, a Stochastic Gradient Descent approach was used to train the model. With a validation accuracy of 71.77%, the model performed better than the Multilayer Perceptron (MLP) and Bayesian Logistic Regression (BLR) models. According to the study, the RNN model shows promise as a tool for estimating the severity of injuries sustained in traffic accidents.

A Recurrent Neural Network (RNN) model for forecasting the severity of injuries from traffic accidents in Malaysia is presented in this research. A methodical grid search and the dropout technique were used to optimize the model to minimize complexity. Compared to other models, the model's validation accuracy was 71.77%, suggesting that performance might be enhanced by more temporal and contextual data. The profile technique was used to calculate the model's impact, and the results showed that the two most important criteria in forecasting the severity of injuries were collision timing and road bound circumstances. Additionally, the model projected that drivers would be more likely to get injured in low light and on dry surfaces.

The study by Alkheder et al. (2017) employed an artificial neural network (ANN) to predict the severity of injuries sustained in traffic accidents in Abu Dhabi. Between 2008 and 2013, 5973 accidents are provided for the data. Two classifiers were employed for training and verifying the ANN classifier, which was constructed using the WEKA data-mining programme. The outcomes demonstrated that the ANN classifiers, with overall model prediction performance of 81.6% and 74.6%, respectively, can predict accident severity. Using a k-means algorithm, traffic accident data was divided into three groups to increase accuracy, with notable gains. The dependent variable was converted from an ordinal to a numerical form using an ordered probit model as a comparison benchmark.

Using WEKA data-mining software, the model was examined. Its predicted performance for training data was 81.6%, while for testing data it was 74.6%. The accuracy of the model was 94.5%, 0%, 78.4%, and 82% for accidents with death, severe, moderate, and minor severity, respectively. The UAE Traffic Agency may utilize the study to enhance traffic safety and provide a tool for the Abu Dhabi Emirate Traffic Office to help with decision-making.

## 2.1.2. Clustering with additional ML algorithms for accident investigation

The goal (Sohn and Lee, 2003) is to increase each classifier's precision in estimating the severity of traffic accidents in Korea. Three methods are used: the Dempster-Shafer algorithm, the Bayesian process, the logistic model, arcing and bagging, and the k-means algorithm. Neural networks and decision trees are used. The empirical findings illustrate the significance of precise data analysis for road accident prevention measures by demonstrating that the clustering-based classification method is the most successful for classifying road traffic accidents in Korea.  
  
The study compares fusion algorithms, ensemble algorithms, and clustering algorithms in terms of classification accuracy and discrimination power using road accident data. Fusion algorithms show better discrimination power than single classifiers, with the Bayesian procedure being the best. The Dempster-Shafer algorithm improves classification accuracy for both individual algorithms and fusion algorithms, but marginally. Ensemble algorithms like bagging and arcing show improvement in accuracy but not as much as clustering-based classification. The study suggests clustering accident data sets first and fitting a classification model for each cluster, ensuring homogeneous safety policies. Further research is needed to generalize comparisons and use multivariate algorithms as fusion tools. Comparing fusion and clustering performances to human experts is also recommended.

## 2.1.3. Employing Different Machine Learning Algorithms for Accident Analysis.

Traffic accidents on the roads are a fundamental problem that result in fatalities and injuries. For documenting important social statistics like traffic volumes and auto accidents, the traffic control system is essential. The study by Priyanka (2014) examines the road traffic databases in Coimbatore city, considering the large volume of traffic and cars. Decision support system analysis uses classic machine learning techniques like SMO, J48, and IBK. A sample database of more than 5,000 items with 29 accident characteristics was used to test the algorithms. The SMO algorithm produced 94% correct findings, according to the results.  
  
The purpose of this study (Priyanka and Sathiyakumari, 2014) is to create a classification model for Coimbatore city road traffic accident data using data mining techniques. In addition to helping traffic officers comprehend driver behavior, accident mode, time, road and weather conditions, and associated factors causing accidents, the model would also help traffic police make judgements when conducting traffic control tasks. Using SMO techniques, several models were constructed, with the top-performing classifier reaching 94% accuracy. The study indicated that there is a significant likelihood of traffic accidents in Coimbatore, indicating the need for improved traffic safety management measures.

## 2.2 Utilizing spatial analysis for accident analysis.

Irumba (2014) examines the reasons behind construction mishaps in Kampala, Uganda, by means of a cross-sectional analysis of 201 sizable construction projects. There are 3797 injuries for every 100,000 workers, and 84 deaths. Falling items, height falls, and mechanical risks are the most frequent causes. Accidents are caused in part by traffic and densely populated areas. Policies on building density, maintenance, safety equipment, and working hours are recommended to reduce occurrences.

Vitianingsih et al. (2021) propose a hybrid estimating model integrating artificial neural networks (ANNs) and multicriteria decision-making (MCDM) is proposed as a framework for geographical analysis. The model will be used to validate and categorize traffic accident-prone roads through the application of metaheuristic optimization techniques. The correctness of the framework is essential for the creation of effective GIS software, and the outcomes may be used by policymakers for preliminary planning.

Briz-Redón et al. (2019) used a linear network to simulate more than 30 km of urban road structure to analyze traffic safety data in Valencia, Spain. The study concentrated on the directionality of traffic flow and traffic accidents at road crossings. Traffic safety at road junctions and spatial variability were considered using zero-inflated negative binomial count models. Along the network, hotspots and cold spots were identified, and the models verified that the existence of road junctions, overdispersion, and spatial heterogeneity explain accident numbers. Additionally, hotspot detection showed that accident numbers at the road segment level may be influenced by ambiguous factors.

Full Bayes (FB) hierarchical models of accident data from Pennsylvania from 1996 to 2000 are compared with conventional negative binomial estimates (Aguero-Valverde and Jovanis, 2006). According to the results, there is a larger collision risk in counties with higher rates of poverty, older age groups, and longer roads. Spatial correlation in collision data is shown by FB injury crash models, and this connection can be addressed to lessen its impact.

By using spatial analytic methodologies, policymakers may develop evidence-based plans for accident prevention and mitigation by gaining a thorough understanding of accident patterns and their underlying causes. Collectively, these studies highlight how crucial it is to consider spatial aspects when creating successful intervention strategies meant to improve public safety and lower accident rates. Examples of these factors include traffic flow dynamics, features of the road infrastructure, and socioeconomic situations.

## 2.3. Utilizing data visualization and statistics for accident analysis.

Shaadan et al. (2021) use secondary data from PDRM and MIROS to investigate traffic incidents in Shah Alam, Malaysia. According to the results, accidents happen more frequently in Selangor, and they happen most often during rush hour. On municipal and straight-ahead roads, motorcyclists (90-250cc), vehicles, and trucks are the leading causes of accidents. Collisions resulting from straying, digressing, skidding, rear, and side collisions are the most frequent types. Driver demographics are linked to the severity of accidents; men are more likely to be the cause of deaths.

According to Sakib et al. (2019), driving beyond the speed limit, age, location, environmental variables, and road types are some of the elements that contribute to traffic accidents and casualties. It creates a dashboard on the UK's traffic accident information from 2014 to 2016 using data visualization techniques. According to the dashboard, most incidents happened on dry, single-carrier ways, metropolitan locations, pleasant weather, and speed restrictions of less than 30 mph. To lower the number of fatalities in traffic accidents, this research can assist UK traffic management authorities in identifying underlying causes and spotting patterns.

Every year, vehicle accidents result in 20–50 million injuries and 1.2 million deaths, mostly in underdeveloped nations like Pakistan. Experts must examine previous data to address this concerning tendency. Using data from Peshawar, Pakistan, Rabbani et al. (2021) discovered that most accidents happen during the day and among those who have not received enough traffic education. The age range of 30 to 45 years old is more likely to cause accidents, which calls for more research. Authorities will be able to reduce the number of traffic accidents by using the study's effective techniques.

Effective preventative measures may be developed by using data visualization and statistics to analyze accidents. This approach provides significant insights into accident patterns and contributing variables.

Shaadan et al. (2021) draw attention to traffic incidents how often are in Shah Alam, Malaysia, especially during rush hours and involving cars, lorries, and motorcycles on city roadways. Driver demographics are related to accident severity, with men more likely to be the cause of deaths. In the meanwhile, dashboard visualization is used by Sakib et al. (2019) to highlight the importance of speed, age, location, and road conditions in traffic accidents in the United Kingdom. Finding underlying causes and patterns is important for focused therapies, and our research helps with that. Furthermore, Rabbani et al. (2021) provided insight into the significance of age demographics and traffic education in relation to accident frequency in Peshawar, Pakistan, and advocated for additional study and efficient use of results to lower accident rates. All these studies highlight how important it is to use statistics and data visualization to improve road safety policies throughout the world.

Jia et al. (2018) provide a spatial clustering technique that makes use of publicly available point-of-interest data to analyze macro-level traffic accidents. The approach calculates regression models for Suzhou Industrial Park, China, and evaluates land use parameters using kernel density estimation. The results indicate that whereas businesses, restaurants, and entertainment venues have little bearing on traffic collisions, residential density, banks, and hospitals do. This shows that crowded places with public facilities may make driving more dangerous.

The performance of two spatial autocorrelation measures based on a Road Safety Risk Index (RSRI) in Tunisia is compared for three areas and three time periods (Ouni and Belloumi, 2019). It looks for likely hot spots and attempts to forecast upcoming traffic accidents. Different geographic and temporal criteria distinguish the known and suspected hot zones, most of which are along rural roadways in the Northwest and Center-West areas. The report offers guidance for Tunisia's safety regulations.

The study chapter begins by taking a broad approach to accident analysis and examining a variety of approaches and strategies employed in this area. Geographic Information System (GIS) technology, Machine Learning (ML), time series analysis, statistical approaches, and visualization techniques are all included in this extensive investigation. It establishes the framework for the next talks by giving this detailed overview.

It progressively narrows down on the topic of visualizing traffic incidents in the UK as the chapter goes on. It can lead readers through the logical flow of the study through this shift from an overview to a more in-depth analysis of the selected issue. It gives readers a clear blueprint of the chapter's flow by beginning with a broad overview and then focusing in on topics.

Also, it makes sure that every subject covered in this shift is directly related to the main goal of the study. It keeps the story coherent and relevant by tying the broader investigation of accident analysis approaches to the particular focus on visualizing traffic incidents in the UK.

In the end, the top-down strategy helps to gain a thorough grasp of the study issue in an organized manner. It successfully engages readers and encourages them to go further into the subject matter by progressively shifting the attention from issues that are distantly linked to ones that are closely associated.

# Chapter 3: Methodology

The methodology chapter, which outlines the methodical technique used to analyse both quantitative and qualitative data, is the foundation of the study. Although the study mostly relies on quantitative analysis, qualitative insights are included in an organised way to offer an in-depth understanding of the investigated topic.

Raw Information (From the Official Website) -> Data Pre-processing (Removing Unnecessary Information) -> Visualisation (Dashboard, Graph, and Charts)

Figure 2 Flowchart of Conducted Analysis

What Makes Visualisation Necessary?

Chen (2014) When asked what the goal of visualisation is, "gaining insight" is typically the first thing that comes to mind for visualisation researchers. But, unless the definition of "insight" is expanded to include all thinking processes, it is not entirely true that all applications of visualisation are for developing a profound knowledge. It might be challenging to determine what, how much, or how accurately insight is received even when the goal of a visualisation activity is insight.

It makes complicated data easier to understand, produces insights, improves communication, identifies trends and abnormalities, and supports decision-making, visualisation is essential.

## 3.1 Collecting and Preparing Data

The datasets for each year are available in CSV Excel format and include information on accidents, vehicles, and casualties. Despite having a shared Accident\_Reference column, these datasets are kept apart because of the complex nature of incidents involving several automobiles and individuals. It would be risky and inaccurate to combine them into a single dataset. The datasets 2021 and 2022 are combined into separate CSV files to simplify analysis while maintaining data consistency and integrity. This tedious procedure establishes a solid foundation for analysis, enabling an in-depth investigation of accident patterns and relevant variables.

### 3.1.1 Analysis of Casualty Data: 2021-2022

Overview: The 2021–2022 casualty dataset offers important information about developments in traffic safety. Using a sample of the 263,690 registered casualties, an in-depth analysis was carried out.

Important characteristics: Essential variables like accident index, year, reference, sex, age, age band, severity, and kind of casualty are included in the dataset's 19 columns. To simplify the research and concentrate on key factors that are essential for comprehending casualty patterns, eleven elements that were deemed less significant were eliminated.

Points of Interest: It provides a casualty trend over the course of two years by focusing on factors such as accident index, year, reference, casualty's sex, age, age band, severity, and casualty type.

### 3.1.2 Analysis of Merged Collision Information: 2021–2022

Overview: An Excel file containing 207,092 rows was created by consolidating the collision dataset for the years 2021 and 2022. Only a portion of the features were chosen, and the other sixteen traits were left out to simplify the study.

Important Elements Chosen: An accident's index, year, reference, longitude, latitude, severity, number of cars involved, number of casualties, date, day of the week, time, kind of road, speed limit, light, weather, and road surface conditions are among its features.

### 3.1.3 Analysis of Combined Vehicle Data: 2021-2022

Overview: The Excel file contains 379,989 rows from the combined vehicle dataset for the years 2021 and 2022. Unnecessary columns were eliminated from the dataset to make it more manageable and feature-focused for analysis.

Important Elements Chosen: Essential variables including the accident index, year, reference, vehicle type, driver's purpose of travel, driver's sex, driver's age, driver's age band, and vehicle age are all retained in the dataset.

## 3.2 Visualization Methodology: Advanced Techniques and Tools

Important data cleansing and archiving are followed by the visualization stage. The richness of the data is captured in an intricate collection of graphs, charts, and tables. To improve the depth of analysis and enable smooth understanding, dynamic dashboard development becomes essential.

Dashboards are all-inclusive visual works that hold several graphs and charts in one consistent interface. Utilizing modern software, like Tableau, which is well-known for its adeptness in dashboard development, takes the visualization process to previously unheard-of heights of sophistication and effectiveness.

In simple terms, the visualization stage is more than just presenting data and it is a diligent combination of art and science that turns complicated datasets into clear and understandable visual stories.

Various kinds of dashboards were created to look at:

* Detailed Accident Count: Find out the number of incidents there were overall throughout the course of the two years. This research gives a comprehensive picture of accident patterns and delivers insightful information about overall traffic safety.
* Gender Perspectives: Start an analysis based on gender to find differences in accident involvement. Through analyzing accident data according to gender, it can be able to identify complex trends and logical causes. Figuring out the Influence of
* Environmental Elements: Analyze collisions under various lighting, weather, and road conditions. This thorough analysis clarifies the ways in which environmental factors affect the frequency and severity of accidents.
* Severity Dynamics Over Time: Investigate at how accident severity relates to time-related variables including day of the week, time of day, and month.

## 3.3 Detailed Overview of Dashboard Metrics:

### 3.3.1 Interactive Severity-Based:

Current Year (CY) Metrics:

CY Accidents: The formula: Total This statistic counts the total number of accidents that happened in the current year (IF [Accident Year] = [Current Year] THEN [Accident Year] END). The algorithm counts the number of accident reports when the accident year coincides with the year that was chosen.

CY Casualties: Compound: SUM The total number of casualties for the current year is determined by this metric: (IF [Accident Year] = [Current Year] THEN [Casualty Reference] END). It summarizes the accident-related casualty references for the chosen current year.

CY Fatal Casualties: Equation: COUNT(IF [casualty\_severity] = 1 and ([Accident Year]) = [Current Year] THEN [casualty\_severity] END) This indicator concentrates on fatalities that have occurred in the current year. The number of cases where the accident year corresponds to the chosen current year and the casualty severity is classified as "fatal" is counted.

CY The formula for serious casualties is COUNT(IF [casualty\_severity] = 2 and ([Accident Year]) = [Current Year] THEN [casualty\_severity] SHOULD END). This measure computes the total number of major injuries that have occurred in the current year. Cases where the accident year corresponds to the chosen current year and the casualty severity is classified as "serious" are counted.

CY Slight Casualties: Formula: COUNT(IF [casualty\_severity] = 3 and ([Accident Year])= [Current Year] THEN [casualty\_severity] END) This measure reflects the total amount of minor injuries that have occurred in the current year. Instances when the accident year corresponds to the chosen current year and the casualty severity is marked as "slight" are counted.

Previous Year (PY) Metrics:

PY Accidents: Formula: COUNT(IF [Accident Year] = [Previous Year] THEN [Accident Year] END). The result determines how many accidents there were overall in the preceding year. When the accident year corresponds with the chosen prior year, the accident data are counted.

PY Casualties: Formula: SUM(IF [Accident Year] = [Previous Year] THEN [Casualty Reference] END). This measure counts and adds up all the injuries that occurred in the preceding year.

PY Fatal Casualties: Formula: COUNT(IF [casualty\_severity] = 1 and ([Accident Year])= [Previous Year] THEN [casualty\_severity] END) This measure is dedicated to the fatal casualties that occurred in the last year. Cases where the accident year corresponds to the chosen prior year and the casualty severity is classified as "fatal" are counted.

PY Serious Casualties: Formula: COUNT(IF [casualty\_severity] = 2 and ([Accident Year])= [Previous Year] THEN [casualty\_severity] END) This measure determines the overall number of fatalities that occurred in the preceding year. Cases where the accident year corresponds to the chosen prior year and the casualty severity is classified as "serious" are counted.

PY Slight Casualties: Formula: COUNT(IF [casualty\_severity] = 3 and ([Accident Year])= [Previous Year] THEN [casualty\_severity] END) This measure represents the total amount of minor injuries that occurred in the preceding year. Cases where the accident year corresponds to the chosen prior year and the casualty severity is classified as "slight" are counted.

Year-on-Year (YoY) Metrics Calculation Formulas:

YoY Accidents: Formula: ([CY Accidents] - [PY Accidents]) / [PY Accidents] This measure determines the percentage change in the total number of accidents from the prior year (PY) to the current year (CY). It gauges how many accidents occur each year in comparison to the overall number of accidents in the year prior.

YoY Casualties: Formula: ([CY Casualties] - [PY Casualties]) / [PY Casualties] The percentage change in the total number of casualties between the current year (CY) and the prior year (PY) is calculated using this statistic. It shows how the overall number of casualties in the previous year varies with the number of casualties in the subsequent year.

YoY Fatal Casualties: Formula: ([CY Fatal Casualties] - [PY Fatal Casualties]) / [PY Fatal Casualties] This indicator compares the percentage change in the total number of fatalities from the previous year (PY) to the current year (CY). It calculates the variation in fatalities between years in relation to the overall number of fatalities in the preceding year.

YoY Serious Casualties: Formula: ([CY Serious Casualties] - [PY Serious Casualties]) / [PY Serious Casualties] This indicator evaluates the percentage difference between the total number of significant casualties in the preceding year (PY) and the current year (CY). It shows how the overall number of serious casualties in the previous year varies with the number of serious casualties in the subsequent year.

YoY Slight Casualties: Formula: ([CY Slight Casualties] - [PY Slight Casualties]) / [PY Slight Casualties] This indicator calculates the percentage change in the total number of minor casualties from the prior year (PY) to the current year (CY). It measures the variation in minor casualties across years in relation to the overall number of minor casualties in the preceding year.

### 3.3.2 Gender Based Analysis

For better visualization and understanding, the age of the casualties was divided into distinct age groups to do a thorough gender-based analysis. These were the definitions of the age groups:

Age Group: 0–9 Description: People in the age range of 0 to 9 years.

Age Group: 10–19 Description: People in the age range of 10 to 19.

Age Group: 20–29 People in the age range of 20 to 19.

Age Range: 30-39 Description: People in the age range of 30 to 39.

Age Group: 40-49 Persons in the age range of 40 to 49.

Age Group: 50-59 Description: People in the 50-59 age range.

Age Group: 60-69 Description: People in the 60-69 age range.

Age Group: 70-79 Description: People in the age range of 70 to 79.

### 3.3.3. Comparative Dashboard – No group or Calculation used.

### 3.3.4. Time Analysis Dashboard - No group or Calculation used.

## 3.4 Benefits and Limitations of the Implemented Approach:

It will effectively acquire a large amount of data and produce comprehensive dashboards that offer a thorough understanding of the reasons that contribute to accidents by utilising visualisation techniques. The dashboards have a range of graphical forms that facilitate multi-way accident data analysis. The dashboard's adaptability is one of its key advantages; it is simple to add new datasets, ensuring that the knowledge gathered will always be pertinent.

It is important to understand, nevertheless, that while visualisation similarities could reveal correlations between factors, they do not always indicate cause. It is important to do more in-depth analysis than only looking at visual uniformity since correlations between variables might exist even in the absence of visual parallels.

To address these shortcomings, the research aims to do comprehensive and detailed analysis, investigating not only the primary accident records but also supplemental datasets and other causes. The research seeks to identify complex contributing variables to accidents by looking at a wider range of data sources, providing a more thorough knowledge of road safety dynamics.

## 3.5 Ethical Considerations:

The study complies with strict ethical guidelines and only uses data from publicly accessible databases on the UK Government website. Because of this, there are no in-person encounters during the study procedure, guaranteeing ethical compliance. Moreover, complete ethical permission was gained from Leeds Beckett University, under which the study is carried out. The integrity of the research process is protected, and the study complies with ethical guidelines thanks to this institutional control.

# Chapter 4: Product/Research Design and Implementation

The main objective of this section is to provide thorough explanations of the techniques used to bring data into Tableau. It also explores the task of creating dashboards, revealing the methods and approaches used to create visually appealing displays of the data. Furthermore, it offers a breakdown of the volume and variety of dashboards created, illuminating the range and depth of analytical investigation made possible by the tableau environment.

## 4.1 Excel for Data Storing: Why?

Spreadsheets and databases are essentially the two primary methods used for data storage. This section explores the reasons for using Excel as the data storing technique.

### 4.1.1 Keeping Information in a Database:

Databases work best when handling linked datasets. They are excellent at upholding consistency, guaranteeing data integrity, and handling intricate systems. But storing in a database is not the best option for this project, as different datasets, such accident causes, are not immediately tied to the main dataset.

### 4.1.2 Keeping Information in an Excel:

Excel is the preferred tool for storing data when it comes to non-complex information. It is a flexible tool that allows you to import, browse, clean, analyse, and visualise data on a single platform. The user-friendly design of Excel makes data handling and updating simple. Excel is the sensible option for managing and storing data because of the project's variety of separate datasets.

## 4.2 Data Fabrication

Several of the variables in our datasets were denoted by numbers rather than by meaningful labels. Gender codes, for instance, may have been 1 for male, 2 for female, and 9 or -1 for uncertain. We decided to manage these modifications in Tableau rather than making the necessary adjustments to the data directly in the original Excel file. This is the reason why:

* Lower Error Risk: Making direct changes to the data in the source might have resulted in mistakes, particularly in huge datasets where it was difficult to identify problems by hand. It can easily assign different names to variables with Tableau's alias function. To reduce the possibility of mistakes, any changes to these alias names might be performed immediately inside Tableau.
* Preventing Data Type problems: Reading the data may have resulted in data type problems if the source had changed the data types. For example, altering this directly in the source would have made it more difficult to tally the overall number of cars involved in accidents by summing the "number\_of\_vehicles" column. We were able to avoid these problems by using Tableau's aliases and by preserving the original data.
* Maintaining Cleanliness and Consistency in the Dataset: Although it eliminates excessive columns and null data from the dataset, it did not manipulate the data to classify fields to guarantee consistency across the dataset. Confusion during operations may have resulted from mixing data types in the altered dataset, such as strings and integers. We were able to preserve a clean and consistent dataset by managing these adjustments in Tableau.
* Futureproofing with Additional Data: Managing new data additions easier by leveraging Tableau's built-in alias function. Aliases in Tableau made data manipulation easier when new data was supplied, guaranteeing smooth integration of data from subsequent years.

Using Tableau's data manipulation and aliasing tools, all the above minimises the chance of errors and ensures that the dataset stays clean, consistent, and adaptable to future changes.

## 4.3 Granting input to Tableau.

The datasets were imported into Tableau after being cleansed and saved. In this phase, just had to choose the source file from among the many formats that Tableau offers, including Excel, text files, PDFs, and Microsoft Access. In this case, the "Microsoft Excel" option under the "To a File" category was selected to import the data.

This procedure was designed to make thorough analysis in Tableau easier. They were able to take use of Tableau's strong analytical capabilities and allow effective data exploration and visualisation by importing the cleaned datasets.

Relationships between the three datasets were discovered after linking them to Tableau. Relationships provide a dynamic and adaptable way to combine data for analysis from several tables. Relationships enable complete outer join capabilities and can manage many-to-many connections, unlike joins, which have different types. The variable "accident\_reference," which acts as a common identifier for the three main datasets, was used to construct these linkages. By ensuring that pertinent data points from several datasets can be examined together, this connection makes it possible to analyse the underlying patterns and insights more thoroughly.

## 4.4 Dashboards

### 4.4.1 Things to Take into Account About While Creating a Dashboard

When dashboards were being created, the emphasis was on combining several graphs and visualisations in one place to give a thorough rundown of the data. Four unique dashboards were created, each focusing on a different dataset aspect:

1. Interactive Severity-Based Dashboard: The initial dashboard was designed to provide a basic grasp of the severity levels of accidents and their consequences by offering insights into casualties based on severity.
2. Gender-Based Analysis: Looking at variables including age range and severity, the second dashboard examined accidents according to gender. Finding patterns and trends pertaining to incidences that are particular to a gender was made easier by this approach.
3. Dashboard for Comparing Accidents with the gender: The third dashboard allowed users to examine different accident-related indicators and trends according to gender. The data might be explored and understood more thoroughly due to this dynamic method.
4. Dashboard for Time Analysis: The fourth dashboard concentrated to time-related analysis and included comparisons of accidents from month to month and year to year. This dashboard gave users a thorough grasp of accident occurrences over time by revealing temporal trends and patterns.

### 4.4.2 Dashboard and Design

We have four dashboards in the design. Tableau has Horizontal and Vertical containers to arrange the visual components in an efficient manner. These containers functioned as flexible frames that allowed for easy size alterations for better readability and presentation. Every dashboard was made from a combination of vertical and horizontal containers that were arranged to hold various graphs and visualisations. To accommodate the various graph sizes and contents, container sizes were adjusted both inside and across dashboards. Larger graphs need larger containers to guarantee correct alignment and readability.

## 4.5 Dashboard Contents

Accidents that are occurred in 2021 and 2022 is visualized, with the following dashboards:

1. Interactive Severity-Based Dashboard
2. Gender-Based Analysis
3. Comparative Dashboard
4. Time Analysis Dashboard

##### 1.Interactive Severity-Based Dashboard

This dynamic dashboard offers a thorough overview of accident data and lets users switch between three severity levels: fatal, serious, and slight. Key features consist of:

* Toggle Between Severity Levels: To assess accident data according to its level of severity—Fatal, Serious, or Slight—users may effortlessly navigate between severity levels.
* Total Accidents: Gives the total number of incidents that have been reported in the chosen degree of severity.
* Total Casualties: Shows the total number of people killed in accidents that happened at the chosen degree of severity. Fatal Casualties: Shows the total number of deaths that have been reported at the chosen severity level.
* Serious Casualties: Indicates how many serious injuries were reported at the chosen severity level.
* Slight Casualties: Shows how many minor injuries were reported at the chosen severity level.
* Casualties by Vehicle Type: Provides information about how casualties are distributed according to the kinds of vehicles that are involved in collisions.
* Percentage of Casualties by Weather: Shows the percentage of deaths attributable to various weather patterns.
* Road Surface: Indicates the percentage of casualties related to different road surface conditions.
* Percentage of Casualties by Road Type: Gives a breakdown of casualties according to the various types of roads.
* Casualty by Location: Provides information on accident hotspots and areas of concern by displaying a geographical distribution of casualties.

##### 2.Gender-Based Analysis

This detailed dashboard gives users an in-depth analysis of accident statistics according to gender, offering insightful information on a variety of variables affecting road safety. Vital features consist of:

* Total Casualties by Gender: Provides a summary of all casualties broken down by gender, allowing users to see how injuries are distributed across the sexes.
* Comparison of Casualties by Age Range and Year: This feature lets users examine casualties across several years in a variety of age groups, giving them insights into age-related patterns and changes in accident rates over time.
* Yearly Comparison of Casualties by Gender: Offers a breakdown of casualties by year so that users may examine patterns in accident rates over time that are particular to a certain gender.
* Analysis of the association Between Accident Severity and Age Range: Provides information on the association between accident severity and age range, allowing users to pinpoint the age groups most susceptible to certain accident severity levels.

##### 3. Comparative Dashboard

By comparing several gender-related characteristics and how they interact with numerous contextual and environmental elements, this comparative dashboard offers an insightful perspective. Important attributes consist of:

* Gender-Based Comparison Throughout Week, Road Surface, Light, and Weather Conditions: It enables users to examine trends and differences in accident frequency according to gender over various time periods (week), road surface conditions, light conditions, and weather conditions.
* Road Surface Conditions and Gender: This provides a thorough analysis of accident rates according to various road surface conditions, offering insights into the connection between gender and road conditions.
* Light Conditions: Examines how gender and light conditions are related, showing how differing lighting conditions affect accident rates.

##### 4. Time Analysis Dashboard

This significant dashboard provides extensive details about the correlation between accident severity and timing variables. Essential traits consist of:

* Time and Severity: Offers a thorough examination of the trends in accident severity over time, allowing users to spot trends and variances in accident severity across various time periods, including days of the week, hours of the day, and months of the year.
* Day and Severity: Examines the connection between the severity of an accident and particular days of the week, enabling users to compare the differences in severity levels between weekdays and weekends.
* Month and Severity: Displays how severity levels change over the course of the year and provides insights into the seasonal fluctuations in accident severity.

# Chapter 5: Analysis and Discussion of Research Findings

A detailed review of the findings from the analysis is provided in this chapter.

### 5.1.1 Recognising Gender Differences in Traffic Accidents

Understanding how gender affects the number of victims in accidents was a key goal of the data study on road safety. Gender-specific total casualties provided some interesting new information.

The distribution of casualties by gender for the years 2021 and 2022 is shown in the pie chart. It is evident that men were involved in the greatest number of incidents, which led to 162,139 fatalities overall. As an example, women were responsible for 99,448 reported casualties, a somewhat smaller but still considerable figure. Interestingly, there were cases in which the gender of the participants was unclear, contributing to the total of 2,102 deaths.

A pie chart with text

Description automatically generated

Figure 3 Overall Casualty According to gender

### 5.1.2 Understanding Age Patterns in Traffic Accidents

An analysis of road safety statistics from 2021 and 2022 revealed a strong curiosity in the relationship between age and the number of accidents. By comparing the number of casualties between different age groups, several interesting conclusions were drawn.

The information provides a thorough analysis of casualties for the years 2021 and 2022 broken down by age group. Interestingly, most victims were always people in their 20s and 30s - 29,701 in 2021 and 29,446 in 2022. Subsequently, the 30-39 age group also suffered a significant number of casualties (24,920 in 2021 and 25,464 in 2022). In addition, the number of casualties included those in the 40–49 age range (17,829 in 2021 and 18,464 in 2022). Remarkably, there were also a substantial number of casualties among the 10–19 age group—16,232 in 2021 and 17,941 in 2022.

This data makes it clear that younger people especially those in their 20s and 30s are disproportionately impacted more casualties and there are frequently a lot of fatalities from these incidents in both years.

A graph of a number of people

Description automatically generated

Figure 4 Casualty comparing with age range and Year

### 5.1.3 Examining Accident Intensity by Age Group

The important insights by analysing the distribution of casualties classified as Fatal, Serious, and Slight. To comprehend how accident severity differs among various age ranges.

The information showed a severity-based breakdown of casualties for two age groups: 20–29 and 30-39. There were 611 fatal accidents, 20,300 severe accidents, and 145,158 minor accidents caused by people between the ages of 20 and 29. In the same manner, those in the 30-39 age range were responsible for 24,839 minor accidents, 16,520 serious accidents, and 511 fatal accidents.

These statistics made it clear that a sizable percentage of casualties across all severity levels were associated with both age groups. On the other hand, compared to people aged 30-39, those aged 20-29 caused somewhat more casualties in all severity groups.

A graph of different colored bars

Description automatically generated

Figure 5 Accident Severity and age range

### 5.2.1 Investigating Inequalities in Gender in Traffic Situations

The investigation analysed the relationship between gender and four different road conditions: dry, snow/frost, wet, and unknown. The data on road safety for the years 2021 and 2022 were evaluated. The information about how men and women use different types of roads was provided by the statistics.

On dry road surfaces, men were involved in 106,017 accidents, while under wet conditions, they were involved in 31,786 incidents. On dry roads, however, women were involved in 62,529 incidents, whereas in rainy weather, 19,691 accidents occurred.

Males were involved in more accidents than females in both dry and wet road conditions, according to the statistics.

A screenshot of a graph

Description automatically generated

Figure 6 Gender and Road Condition

### 5.2.2 Examining Gender Differences in Lighting

Through the analysis of road safety data, to determine the correlation between gender and various light situations, including daylight, darkness with and without illumination, darkness with and without illumination, and darkness with no streetlight. The findings provided fascinating new information on the ways that different light conditions impact males and females.

Comparing the light circumstances, it shows that 64,112 casualties were caused by females and 100,014 by men. In addition, men caused 122,744 casualties in situations where streetlights were present and lighted, while women were accountable for 59,288 casualties.

This data suggests that, relative to females, men were engaged in more casualties in both daytime and dark settings with streetlights present and illuminated environments. This shows that there may be a gender difference in the results of road safety in different lighting scenarios, and it implies that specific interventions are needed to address the risk factors associated with male drivers in different lighting scenarios.

A screenshot of a graph

Description automatically generated

Figure 7 Gender and Light Condition

### 5.2.3 Examining Gender Differences in Traffic Situations

Comparing the effects of gender on other elements, including the day of the week, road surface conditions, lighting conditions, and weather conditions, is a method of analysing data related to road safety. The information showed some intriguing patterns in the ways that certain situations impact men and women differently.

Friday has the greatest number of casualties when compared to other days depending on gender. In this scenario, 9,866 casualties were caused by females, while 14,269 casualties were caused by men. On Sunday, on the other hand, there were the fewest casualties—9,907 from men and 5,975 from women.

Based on this data, it can be determined that weekdays, especially Fridays are more accident-prone for both men and women than weekends. In addition, compared to females, men often account for a greater number of casualties under all circumstances.

A screenshot of a graph

Description automatically generated

Figure 8 Gender based Comparison with Week, Road Surface, Light and Weather Condition

Age Group: Due to variables including inexperience, risk-taking behaviour, and overconfidence, young drivers (usually between the ages of 20 and 29) can be more likely to be involved in accidents. Drivers in their middle years (between 30 and 39 years old) may also be more vulnerable, especially if they are balancing job and family obligations, which might result in distractions or exhaustion.

Gender: Historically, there has been a greater incidence of accidents involving men than women. This might be explained by the fact that men are more prone than women to drive recklessly, including speeding or driving aggressively.

Reasons for Increased Road and Weather-Related Casualties:

Road Situation: The possibility of accidents can be raised by wet or slick road conditions, particularly in the event of rainy or snowy weather. Vehicles may slide due to drivers losing control of their cars. Particularly at night, dimly lit or unlit roadways can decrease visibility and raise the risk of crashes, especially if drivers don't modify their speed and driving style accordingly.

Weather Analysis: Rain, snow, or fog are examples of adverse weather conditions that can limit traction and impair vision, making it more difficult for drivers to manoeuvre their cars safely. A build-up of ice or snow on a road can make driving dangerous and raise the possibility of incidents like skidding or sliding off the road.

In conclusion, when the road and weather are bad, younger, and middle-aged drivers especially men may be more likely to cause fatalities. Various factors, including inexperience, a knack for taking risks, and poor visibility or traction on the road, are responsible for the increased accident rates among these populations.

### 5.3.1 Examining Historical Patterns of Accident Severity

Observed crucial information on the timing of accidents and their severity levels by examining patterns in Fatal, Serious, and Slight events.

The statistics showed that after eight in the morning, all accident severity levels climbed progressively. At 16:00, there were 205 fatal events, marking the highest recorded number of the total. In close succession, at 17:00, the number of casualties from Serious incidents peaked by 3,779, while the number of casualties from Slight incidents peaked by 13,960. The graph shows that after this peak, the number of accidents progressively declined. It suggests that the probabilities of accidents, of any kind, tends to rise with the length of the day and peaks in the late afternoon.

A graph with numbers and a line

Description automatically generated

Figure 9 Time and Severity Comparison

### 5.3.2 Examining Daily Patterns in the Severity of Accidents

This is the correlation between the day of the week and the severity of accidents - Slight, Serious, and Fatal.

According to the data, Friday had the largest number of accidents, especially about Slight and Serious injuries. There were 8,136 serious casualties and 35,890 slight casualties on this day. On the other hand, Saturday saw the greatest number of recorded fatalities - 550.

Conversely, Saturday saw the fewest accidents both minor and serious of any day. On this day, there were only 24,479 recorded slight wounds and 6,440 reported serious injuries. Tuesday had the fewest number of fatalities 402 instances were reported in contrast.

A screenshot of a graph

Description automatically generated

Figure 10 Day and Severity Comparison

### 5.3.3 Looking at Monthly Accident Severity Trends

According to the statistics, November in both 2021 and 2022 had the greatest number of accidents overall, regardless of severity. November 2022 saw 7,386 minor injuries, 2,073 serious injuries, and 135 fatal injuries. Similar numbers were recorded in November of 2021: 7,711 minor injuries, 2,127 serious injuries, and 138 fatalities.

On the other hand, February consistently recorded the fewest accidents in both years. Figured that November has a history of having a higher-than-average risk of accidents, especially when it comes to minor, serious, and fatal incidences. Numerous variables, including the weather, more traffic, or vacation travel, might be to blame for this.

February has fewer accidents than other months, which might be brought on by things like shorter days, colder temperatures, or fewer outside activities.

A screenshot of a graph

Description automatically generated

Figure11 Month and Severity Comparison to the year 2021 and 2022

The following hour, day of the week, and month had the highest number of casualties:

* Time: At 17:00, the number of victims reached its peak.
* Day of the week: The greatest number of fatalities was recorded on Friday.
* The month with the most casualties was November.

Cause of Increased Deaths: Time (17:00): There are several reasons why accidents occur most frequently, including rush hour traffic, drivers who are more tired after a long day at work, lowered visibility at dark, and possible distractions from the road. Additionally, there may be more accidents around this period due to variables like aggressive driving, speeding, and intoxicated driving.

Day of the week (Friday): As people get ready for the weekend and make their way home from work, Fridays usually experience a spike in traffic. There can be more cars on the road at this time, which increases the risk of accidents. The increased accident rate on Fridays may also be attributed to drivers' tendency for hazardous behaviour or exhaustion following a lengthy workweek.

Month (November): A few seasonal factors may lead to an increase in accident rates in November. For instance, November often announces the arrival of winter in many areas, which is characterised by fewer daylight hours, bad weather (snow or rain), and possibly dangerous driving conditions.

### 5.4 Evaluating Increases in the Severity of Traffic Accidents: A Comparison of 2021 and 2022

Reason: To identify patterns and changes in the severity of accidents, it is carefully scrutinised data from 2021 and 2022 to comprehend the dynamics of road safety. They aimed to provide insight into how the situation of traffic accidents changed over the course of the two years by carefully reviewing the interactive severity dashboard.

Results and Interpretation:

Overall Rise in Accidents and Casualties: Between 2021 and 2022, the number of accidents increased by 5.67%, signalling a worrying increase in traffic-related occurrences. Over the same year, the total number of casualties increased by 7.70%, indicating a concerning pattern of increasing traffic-related injuries. The percentage of casualties increased by 9.82%, which is a concerning increase in accidents that pose a hazard to life. Furthermore, there was a notable 10.52% rise in serious casualties, indicating an increased probability of severe injuries. The percentage of slight injuries increased somewhat by 4.51%, suggesting a worrying trend towards an increase in minor accidents.

A screenshot of a graph

Description automatically generated

Figure 12 Comparison of Casualty between the year 2021 and 2022

Analysis according to Severity:

Fatal Severity: Cars were the leading cause of fatal accidents, accounting for 30,874 fatalities; closely behind them, with 19,231 deaths, were freight vehicles. These occurrences mostly happened on dry roads, with 3494 deaths on single carriageways and 1025 on multiple carriageways, and in pleasant weather (62.47% and 56.69% of fatalities, respectively).

A screenshot of a computer

Description automatically generated

Figure 13 Fatal Severity Dashboard

Serious Severity: Cars once again accounted for 438,423 casualties in serious accidents, with freight vehicles coming in second with 155,957 injuries. These occurrences were most common on dry roads, with 37,091 casualties on single carriageways and 7,736 on dual carriageways, and in good weather, accounting for 57.99% and 61.54% of major accidents, respectively.

A screenshot of a computer screen

Description automatically generated

Figure 14 Serious Severity Dashboard

Slight Severity: With 1,904,145 fatalities, cars were the main cause for slight accidents. Goods vehicles came in second with 613,272 injuries. These incidents mostly occurred on dry roads, with 90,172 casualties on single carriageways and 24,336 on multiple carriageways. They also occurred mostly in good weather circumstances, accounting for 49.93% and 61.27% of slight crashes, respectively.

A screenshot of a computer screen

Description automatically generated

Figure 15 Slight Severity Dashboard

Conclusion:

From 2021 to 2022, there was a worrying increase in traffic accidents and fatalities across all severity categories, according to the thorough investigation. The frequency of automobile-related incidents, especially during favourable weather and on dry roads, highlights the pressing need for focused interventions and improved road safety protocols.

# Chapter 6: Project Management

This chapter explains the supervision of the research and the communication between the author and the supervisor.

A screenshot of a computer

Description automatically generated

Figure 13 Gantt chart for the project creation timeline

## 6.1 Initial Research and Topic Selection

The study subject was selected in this phase, with a focus on accident analysis for the years 2021–2022. The programme (Tableau) and dataset were finalised. In addition to developing the study concept, the first steps of collecting data and cleaning were started. The supervisor provided insightful input that helped determine the study's course and the best visualisation tools to be used.

## 6.2 Data Preparation and Literature Analysis

To close the research gap and to select the best approach, more research was done. After the data was cleaned, the first visualisations were produced. The supervisor reviewed the literature review and provided recommendations to improve the coherence and connections between the various parts.

## 6.3 The Stage of Intermediate Visualisation

This stage involved an in-depth examination of the accident data and the consideration of qualitative data sources to enhance the study. Email exchanges were used to get the supervisor's input. Along with Time Analysis, four dashboards were made: Severity-Based, Gender-Based, and Comparative. Obtained insightful input on how to make the report's explanation of each dashboard's promote better.

## 6.4 The Last Stage of Visualisation

Changes were implemented in response to the input gathered. Under the supervisor's direction, the dissertation writing advanced concurrently, bringing the project to its current state.

# Chapter 7: Conclusion and Future Work

## 7.1 Summary of the Study

An overview of the results After the accident data was analysed, several important trends emerged. First, an increased risk of accidents towards the conclusion of the week was indicated by the observation that most accidents happened on Fridays. On the other hand, a disproportionately high number of fatal accidents were reported on Saturdays in comparison to other days, suggesting that the severity of accidents peaked on these days.

A closer look indicated a gender difference in accident participation, with male drivers responsible for most of the incidents. Part of the reason for this tendency is that male drivers are more likely to be involved in accidents since they often travel farther than female drivers.

It is interesting to note that environmental elements like illumination, weather, and road surface quality did not seem to have a major impact on the frequency of accidents. Remarkably, most accidents happened in situations that were considered safe for driving, either in the daytime or at night when lamps were on. This implies that other variables, such driving habits or the quality of the roads, can have a greater influence on accident rates.

##### Potential Causes of the Weekend Accident

There may be a connection between alcohol use and related driving behaviours and the reported rise in accidents on Fridays and Saturdays. According to research, one in four people frequently drink no- or low-alcoholic beverages at least once a month. This represents a significant share of the adult population. These drinks, which are promoted as healthier substitutes, are frequently more expensive than traditional alcoholic beverages, which may lead to differences in consumption habits between socioeconomic groups.

Furthermore, the market share of well-known alcohol brands in the no/low drink category highlights how widely accessible and well-liked these goods are. This accessibility, together with rising off-trade sales of alcoholic drinks, raises the possibility of a relationship between weekend accident rates and alcohol intake.

In conclusion, the evidence points to the possibility that alcohol use, especially on the weekends, may be a factor in the higher frequency of accidents on Fridays and Saturdays. Targeted initiatives aiming at lowering alcohol-related accidents and encouraging safer driving practices can be informed by an understanding of these trends.

## 7.2 Recommendations

Several focused questions arise to improve road safety and lessen the risks connected to drunk driving to address the alarming trends in alcohol-related incidents.

* How can widespread awareness efforts, especially on weekends, successfully inform people about the dangers of driving after drinking?
* What steps may be taken to make driving while intoxicated more strictly enforced, particularly on Fridays and Saturdays when the highest number of accidents occur?
* What tactics may be used to increase accessibility to other modes of transportation, such ride-sharing services, to give those who drink alcohol safe alternatives?
* What are some ways to encourage community involvement in road safety measures, such as outreach programmes and educational events centred around responsible driving practices?
* What improvements to the infrastructure, including better lighting, can be made to increase road safety in regions that are prone to accidents?
* Which interventions, such as those for young drivers or people from lower socioeconomic backgrounds, can be created to meet the special requirements of the demographic groups who are more likely to be involved in alcohol-related accidents?

Extensive public awareness efforts can educate people about the dangers of driving after drinking. Improving police presence during high-accident times, particularly on Fridays and Saturdays, discourages drunk driving. Increasing the availability of ride-sharing services provides safe travel choices. Initiatives for responsible driving are supported by community involvement. Improvements to the infrastructure, such as better lighting, increase traffic safety. Specific interventions cater to the requirements of high-risk populations.

## 7.3 Restrictions on the Research

Challenges in Gathering Data: Due to different sources, integrating additional datasets like those on traffic and accident cause difficulties that affect the research's qualitative analysis.

Causation vs. Correlation: It is important to understand that the connection of features does not imply dependency on one another, nor does a visual relationship between variables suggest causation.

Lack of Validated 2023 Dataset: The year data could not be used for analysis because the data source did not provide a validated 2023 dataset ready. This could have affected the research's fullness.

## 7.4 Future Work

1. Incorporating Validated 2023 Dataset:

Improved Trend Analysis: Upon validation of the 2023 dataset's availability, it is integration will provide a more thorough comprehension of changing accident trends, offering insights into recent advancements and modifications in accident patterns.

Temporal Trend Analysis: By analysing the 2023 data in conjunction with pre-existing records, it will be possible to analyse accident occurrences temporally and spot patterns and trends that change over time.

2. Integration of GIS and Visualisation:

Enhancement of Spatial Analysis: By combining Geographic Information Systems (GIS) with visualisation methods, accident data may be analysed more thoroughly and spatially to investigate the elements that lead to accidents in various geographic regions.

Identification of Patterns: Using spatial visualisation of accident data, patterns and clusters may be found, which helps identify high-risk regions and formulate focused initiatives aimed at enhancing road safety.

3. Employing Supplementary Datasets:

Data Enrichment: By adding further characteristics and factors that can affect accident occurrences, the analysis can be made more comprehensive by utilising the Department for Transport (DFT), UK, additional datasets.

Deeper Understanding of the Complex Interplay of Factors Causing Accidents: Complex Analysis can be facilitated by including varied datasets, such as information on road infrastructure, driver demographics, and traffic flow statistics.

# References

Aguero-Valverde, J. and Jovanis, P.P., 2006. Spatial analysis of fatal and injury crashes in Pennsylvania. *Accident Analysis & Prevention*, *38*(3), pp.618-625.

Alkheder, S., Taamneh, M. and Taamneh, S., 2017. Severity prediction of traffic accident using an artificial neural network. *Journal of Forecasting*, *36*(1), pp.100-108.

Briz-Redón, Á., Martínez-Ruiz, F. and Montes, F., 2019. Spatial analysis of traffic accidents near and between road intersections in a directed linear network. *Accident Analysis & Prevention*, *132*, p.105252.

Chaudhary, U., Patel, A., Patel, A. and Soni, M., 2021. Survey paper on automatic vehicle accident detection and rescue system. In Data Science and Intelligent Applications: Proceedings of ICDSIA 2020 (pp. 319-324). Springer Singapore.

Chen, M., Floridi, L. and Borgo, R., 2014. What is visualization really for?. *The Philosophy of Information Quality*, pp.75-93.

Geyik, B. and Kara, M., 2020, June. Severity prediction with machine learning methods. In 2020 international congress on human-computer interaction, optimization and robotic applications (HORA) (pp. 1-7). IEEE.

Hazaymeh, K., Almagbile, A. and Alomari, A.H., 2022. Spatiotemporal analysis of traffic accidents hotspots based on geospatial techniques. *ISPRS International Journal of Geo-Information*, *11*(4), p.260.

Irumba, R., 2014. Spatial analysis of construction accidents in Kampala, Uganda. Safety Science, 64, pp.109-120.

Jia, R., Khadka, A. and Kim, I., 2018. Traffic crash analysis with point-of-interest spatial clustering. Accident Analysis & Prevention, 121, pp.223-230.

Kushwaha, M. and Abirami, M.S., 2021, September. Comparative analysis on the prediction of road accident severity using machine learning algorithms. In *International Conference on Micro-Electronics and Telecommunication Engineering* (pp. 269-280). Singapore: Springer Nature Singapore.

Lakshmi, S., Srikanth, I. and Arockiasamy, M., 2019. Identification of traffic accident hotspots using geographical information system (GIS). *International journal of engineering and advanced technology*, *9*(2), pp.4429-4438.

Lokala, U., Nowduri, S. and Sharma, P.K., 2017. Road accidents bigdata mining and visualization using support vector machines. World Academy of Science, Engineering and Technology-International Journal of Computer and Systems Engineering, 10(8).

Murray, Å., 2003. Decreasing number of young licence holders and reduced number of accidents: A description of trends in Sweden. *Accident Analysis & Prevention*, *35*(6), pp.841-850.

Ouni, F. and Belloumi, M., 2019. Pattern of road traffic crash hot zones versus probable hot zones in Tunisia: A geospatial analysis. Accident Analysis & Prevention, 128, pp.185-196.

Priyanka, A. and Sathiyakumari, K., 2014. A comparative study of classification algorithm using accident data. *International Journal of Computer Science & Engineering Technology (IJCSET)*, *5*(10), pp.1018-1023. 

Rabbani, M.B.A., Musarat, M.A., Alaloul, W.S., Maqsoom, A., Bukhari, H. and Waqas, R., 2021. Road traffic accident data analysis and its visualization. Civil engineering and architecture, 9(5), pp.1603-1614.

Sakib, A., Ismail, S.A., Sarkan, H., Azmi, A. and Mohd Yusop, O., 2019. Analyzing traffic accident and casualty trend using data visualization. In Recent Trends in Data Science and Soft Computing: Proceedings of the 3rd International Conference of Reliable Information and Communication Technology (IRICT 2018) (pp. 84-94). Springer International Publishing.

Sameen, M.I. and Pradhan, B., 2017. Severity prediction of traffic accidents with recurrent neural networks. *Applied Sciences*, *7*(6), p.476.

Shaadan, N., Suhaimi, M.A., Hazmir, M.I. and Hamzah, E.N., 2021, July. Road accidents analytics with data visualization: a case study in Shah Alam Malaysia. In Journal of Physics: Conference Series (Vol. 1988, No. 1, p. 012043). IOP Publishing.

Sodikov, J., 2018. Road traffic accident data analysis and visualization in R. International Journal of Computer Science Engineering and Information Technology Research (IJCSEITR), 8, pp.25-32.

Sohn, S.Y. and Lee, S.H., 2003. Data fusion, ensemble, and clustering to improve the classification accuracy for the severity of road traffic accidents in Korea. *Safety Science*, *41*(1), pp.1-14.

Soltani, A. and Askari, S., 2017. Exploring spatial autocorrelation of traffic crashes based on severity. *Injury*, *48*(3), pp.637-647.

Sunkpho, J. and Wipulanusat, W., 2020. The role of data visualization and analytics of highway accidents. *Walailak Journal of Science and Technology (WJST)*, *17*(12), pp.1379-1389.

Vitianingsih, A.V., Suryana, N. and Othman, Z., 2021. Spatial analysis model for traffic accident-prone roads classification: A proposed framework. IAES International Journal of Artificial Intelligence, 10(2), p.365.

Wahab, L. and Jiang, H., 2019. A comparative study on machine learning based algorithms for prediction of motorcycle crash severity. *PLoS one*, *14*(4), p.e0214966.

Wen, X., Xie, Y., Jiang, L., Pu, Z. and Ge, T., 2021. Applications of machine learning methods in traffic crash severity modelling: current status and future directions. *Transport reviews*, *41*(6), pp.855-879.

# Appendices

Meeting 1 Name of the meeting: Supervisory Meeting

Date and Time: Monday 5 February 2024 at 12:30 – 13:00.

Place: Online MS Teams Meeting

Attendees: Suman Muthukumaran & Mark Dixon

Meeting 2 Name of the meeting: Supervisory Meeting

Date and Time: Friday 5 April 2024 at 12:30 – 13:00.

Place: Online MS Teams Meeting

Attendees: Suman Muthukumaran & Mark Dixon