Analysis of Car Accidents and Population Demographics: Informing Road Safety Strategies in the United Kingdom

GROUP 2

Sahithi Vidiyala 11717331
Sathwika Karingu 11708107
Lakshmi Siri Chandana Jasti 11704788
Rohit Ibrahimpatnam 11719429

Introduction

Car accidents remain a significant problem in the UK, resulting in injuries, fatalities, and economic losses. The main cause of accidents is driver behavior, including over speeding, distracted driving, and drunken driving. Safe driving habits need to be promoted to reduce accidents. Road infrastructure, vehicle characteristics, and environmental factors also contribute to accident rates.

This project will focus on the UK's accident rate to establish trends and understand accident patterns and locations. The report will investigate car accidents within the UK, including their occurrence, patterns, and consequences. The aim is to help policymakers design interventions and policies to lower the incidence of car accidents and promote safer road environments.

Domain

Introduction to Domain: Health and Transportation Development
The focus of the project is the overlap between health and Transportation, specifically targeting the considerable consequences of car accidents in UK. This subject explores the significant effects that car accidents has on transportation health.
and the overall economic health.

Motivation

The inspiration behind this project is due to high rate of car accidents that are happening in the United Kingdom. Through the report, we can come up traffic policies that are based on evidence by investigating the data obtained from car accidents so that we can successfully mitigate the accidents and reduce their number and also improve road safety features.

Significance

This project is of importance since it aims to tackle the issue of road safety in transportation. Through data analysis and data visualization, we can find out the main features, patterns, and risk factors that lead to car accidents. The final project can be utilized to inform evidence-based road safety measures which are in turn used to reduce the impacts of road accidents and to improve road safety initiatives. These findings will lead to the development of safe transport systems in the long run.

Objectives

1. To identify key trends, distributions & risk factors involved in accidents.
2. Exploring demographic data in order to gain insights into how different populations influence car accident rates & patterns.
3. Utilizing data visualization techniques & tools to create visually appealing and informative visualizations.
4. To extract insights from our findings.

Data Abstraction

Car Accident Dataset

Source link: https://www.kaggle.com/datasets/nextmillionaire/car-accident-dataset

Source: Kaggle

Size: Number of rows: 312456, Number of columns: 21

Types and attributes: {'Name': 'Car Accident Dataset', 'Type': 'CSV', 'Attributes': ['ï»¿Accident_Index', 'Accident Date', 'Day_of_Week', 'Junction_Control', 'Junction_Detail', 'Accident_Severity', 'Latitude', 'Light_Conditions', 'Local_Authority_(District)', 'Carriageway_Hazards', 'Longitude', 'Number_of_Casualties', 'Number_of_Vehicles', 'Police_Force', 'Road_Surface_Conditions', 'Road_Type', 'Speed_limit', 'Time', 'Urban_or_Rural_Area', 'Weather_Conditions', 'Vehicle_Type']}

Data types of each column as a list:

[dtype('O'), dtype('O'), dtype('O'), dtype('O'), dtype('O'), dtype('O'), dtype('float64'), dtype('O'), dtype('O'), dtype('O'), dtype('float64'), dtype('int64'), dtype('int64'), dtype('O'), dtype('O'), dtype('O'), dtype('int64'), dtype('O'), dtype('O'), dtype('O'), dtype('O')]

Data Abstraction

Population Dataset

Source link: (https://www.kaggle.com/datasets/rajkumarpandey02/2023-world-population-by-country/)

Source : kaggle

Size: Number of rows: 234, Number of columns: 19

{'Name': 'Population Dataset', 'Type': 'CSV', 'Attributes': ['country', 'rank', 'area', 'landAreaKm', 'cca2', 'cca3', 'netChange', 'growthRate', 'worldPercentage', 'density', 'densityMi', 'place', 'pop1980', 'pop2000', 'pop2010', 'pop2022', 'pop2023', 'pop2030', 'pop2050']}

Data types of each column as a list:

[dtype('O'), dtype('int64'), dtype('float64'), dtype('float64'), dtype('O'), dtype('O'), dtype('float64'), dtype('float64'), dtype('float64'), dtype('float64'), dtype('float64'), dtype('int64'), dtype('int64'), dtype('int64'), dtype('int64'), dtype('int64'), dtype('int64'), dtype('int64'), dtype('int64')]

Data Abstraction

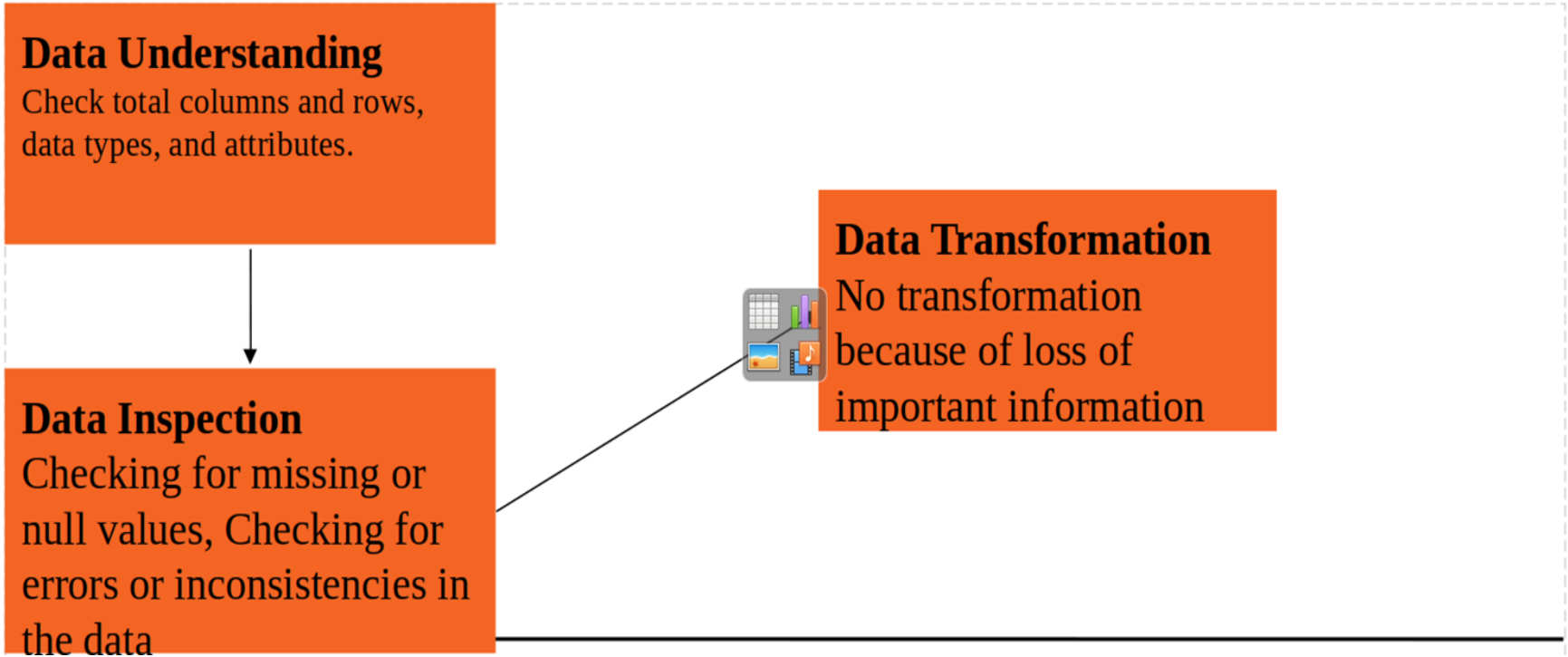
Regional traffic
Source link: https://www.data.gov.uk/dataset/208c0e7b-353f-4e2d-8b7a-1a7118467acc/gb-road-traffic-counts/datafile/b5a2e377-d4f5-40da-a840-71bd54a6a0ac/preview
Source: data.gov.uk
Number of rows: 89494
Number of columns: 28

{'Name': 'Car Accident Dataset', 'Type': 'CSV', 'Attributes': ['collision_index', 'collision_year', 'collision_reference', 'vehicle_reference', 'vehicle_type', 'towing_and_articulation', 'vehicle_manoeuvre', 'vehicle_direction_from', 'vehicle_direction_to', 'vehicle_location_restricted_lane', 'junction_location', 'skidding_and_overturning', 'hit_object_in_carriageway', 'vehicle_leaving_carriageway', 'hit_object_off_carriageway', 'first_point_of_impact', 'vehicle_left_hand_drive', 'journey_purpose_of_driver', 'sex_of_driver', 'age_of_driver', 'age_band_of_driver', 'engine_capacity_cc', 'propulsion_code', 'age_of_vehicle', 'generic_make_model', 'driver_imd_decile', 'driver_home_area_type', 'lsoa_of_driver']}

Data Transformation

Datasets were not transformed. Dropping duplicates or other transformations would lead to a loss of important information or distort the representation of the dataset.

Workflow diagram



Task Abstraction

- Severity of Accidents:

Target Variable: Accident severity (minor, moderate, severe).

Analysis Goal: Identify trends and risk factors associated with different levels of accident severity.

- Accident Frequency:

Target Variable: Number of accidents per time period (weekly).

Analysis Goal: Analyze trends and distributions of accident frequency over time or across different regions.

Task Abstraction continuation

- Types of Accidents:

Target Variable: Type of accident.

Analysis Goal: Identify the distribution of different types of accidents and explore factors contributing to specific types.

- Accident Causes:

Target Variable: Root cause of accidents

Analysis Goal: Investigate the main causes of accidents and assess their impact on accident rates.

Task Abstraction continuation

- Accident Risk Factors:

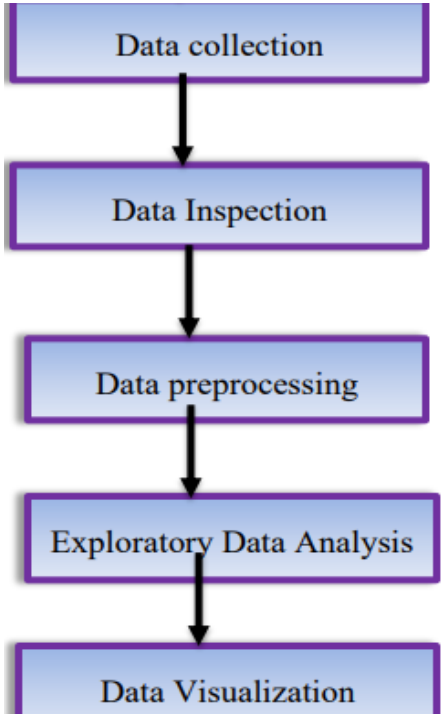
Target Variable: Factors contributing to accidents (speed, weather conditions, road conditions).

Analysis Goal: Identify and analyze risk factors that increase the likelihood of accidents occurring.

- Accident Locations:

Target Variable: Geographic locations or areas where accidents occur.

Analysis Goal: Analyze spatial distributions of accidents and identify high-risk zones.



Workflow

- Data Collection: This is the first step in the process, where we gathered from kaggle.
- Data Inspection: We checked for missing or null values, outliers or anomalies, errors or inconsistencies in the data, distribution of the data (mean, median, mode, standard deviation) and data type.

Workflow

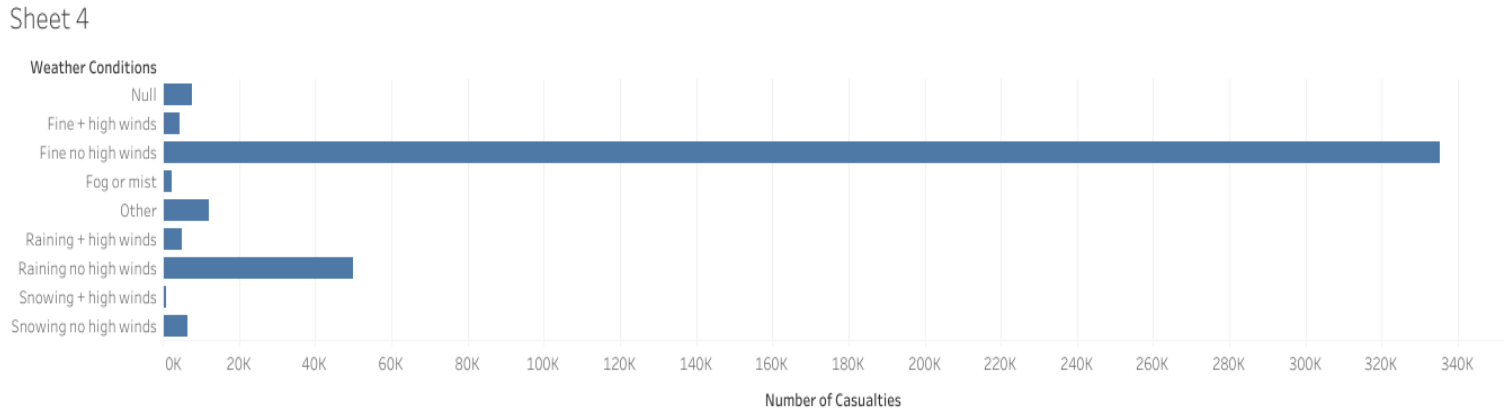
- Data Preprocessing: After inspecting the data, the next step is to preprocess it. This might involve cleaning the data (e.g., removing duplicates or outliers), transforming it into a format that's easier to work with, or normalizing it so that all data points are on the same scale. We did not transform data in this project because it would lead to a loss of important information or distort the representation of the dataset.
- Exploratory Data Analysis (EDA): Once the data has been preprocessed, it's time to start analyzing it. EDA involves using statistical methods and visualization techniques to explore the data and identify patterns or trends. We used python libraries(pandas, seaborn and matplotlib)
- Data Visualization: Finally, the results of the EDA can be visualized using charts, graphs in D3js and Tableau. This helped to communicate the results of the analysis and also help to identify any further areas for investigation

Implementation using tools

For this project we used Python as the tool for analysis of both the datasets, for data preprocessing and analysis. Data cleaning and wrangling were performed with Pandas. NumPy was also be employed. Static graphs were be created with Matplotlib and Seaborn packages.

The visualizations were implemented using D3.js and Tableau. With the effective study of the problem, we applied evidence-based analysis and recommendations that put safety and people first in the UK.

Visualizations

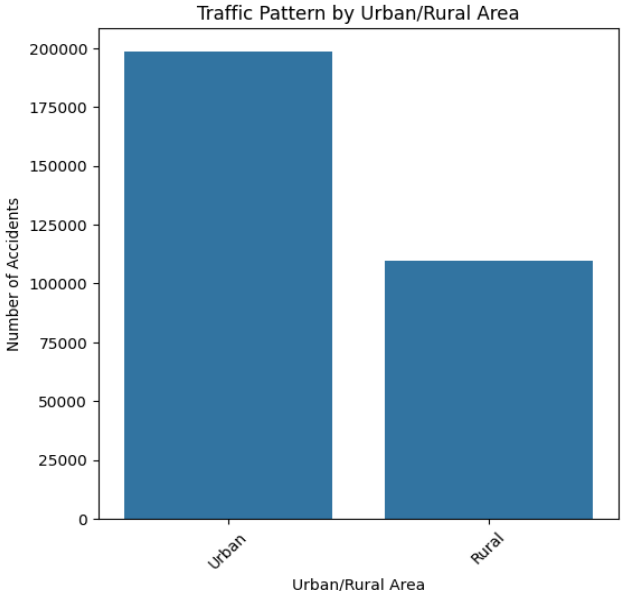


Number of Casualties by Weather Conditions

The relationship between the weather conditions and the number of casualties is complex and may be influenced by other factors.

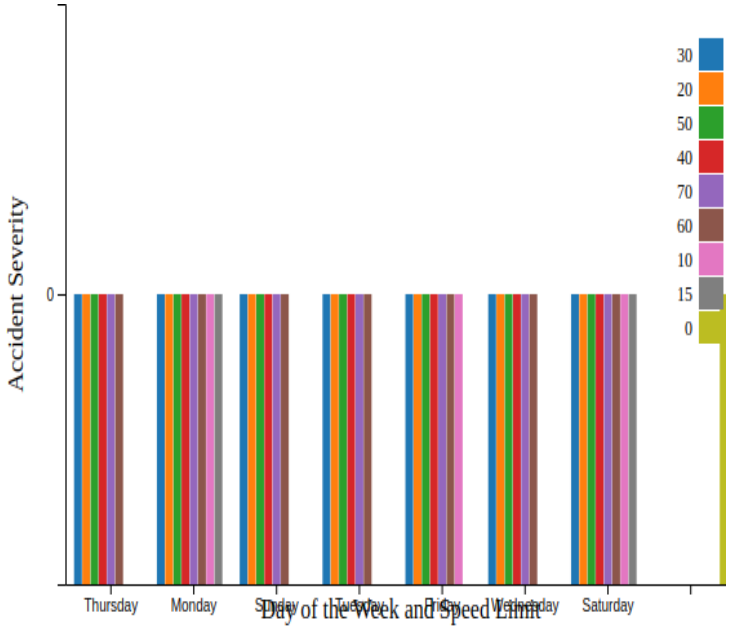
However, the trends in the data suggest that high winds and precipitation may be associated with a higher number of casualties.

Number of Casualties by Urban or Rural Area



There are more casualties in urban areas than in rural areas, regardless of density. However, the number of casualties in urban areas decreases as the density decreases. This suggests that population density may be a contributing factor to the number of casualties in urban areas. In contrast, the number of casualties in rural areas does not appear to be affected by density.

Accident Severity vs. Number of Casualties

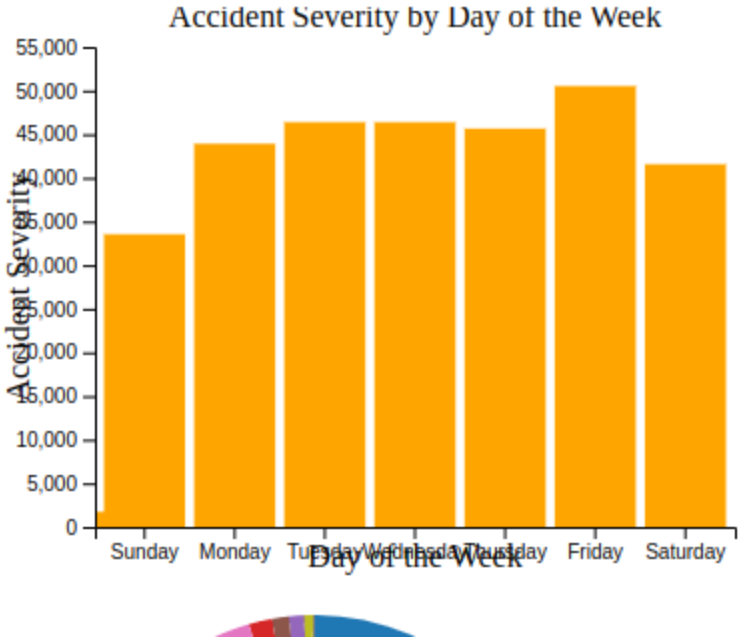


- From the table, we can see that the accident severity categories are not mutually exclusive with the number of casualties. For example, there are cases where the accident severity is "Serious" with 20, 40, and even 50 casualties. Similarly, there are cases where the accident severity is "Slight" with 10 or 20 casualties.

- However, we can observe some general trends in the data. For instance, as the severity of the accident increases, the number of casualties also tends to increase. Specifically, we can see that the "Fatal" accident severity category has the highest number of casualties, with a range of 30 to 50.

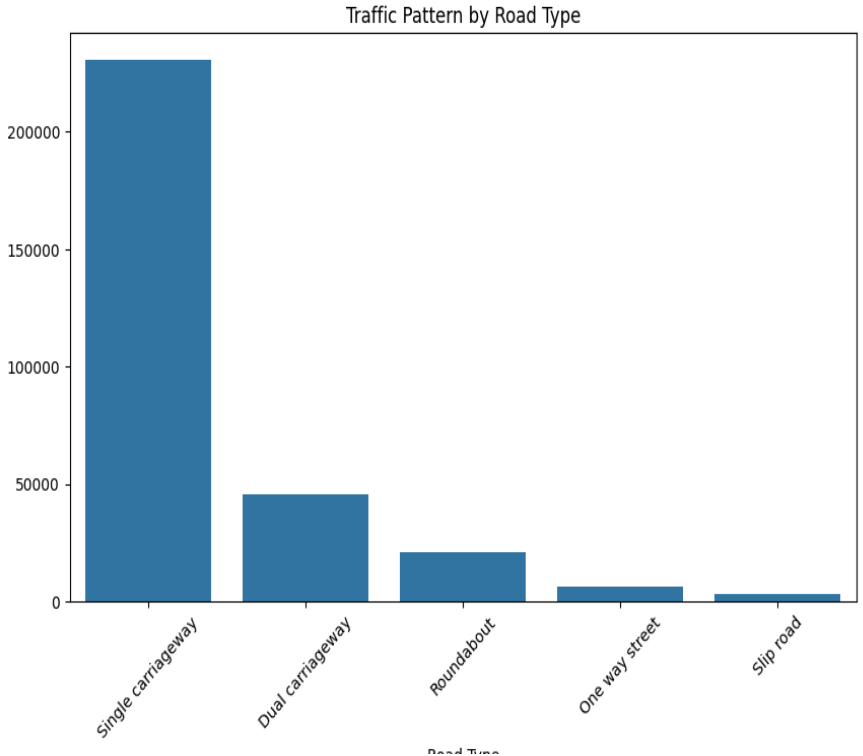
Continuation...

- On the other hand, the "Slight" accident severity category has the lowest number of casualties, with a range of 0 to 20.
- Additionally, we can see that there are some cases where the number of casualties is relatively high, even for lower severity accidents. For example, there are cases where the accident severity is "Slight" with 20 casualties, which is relatively high compared to other "Slight" accidents.
- We can see that as the severity of the accident increases, the number of casualties also tends to increase, although there are some exceptions to this trend.



Analysis of the Correlation between Accident Severity and Day of the Week

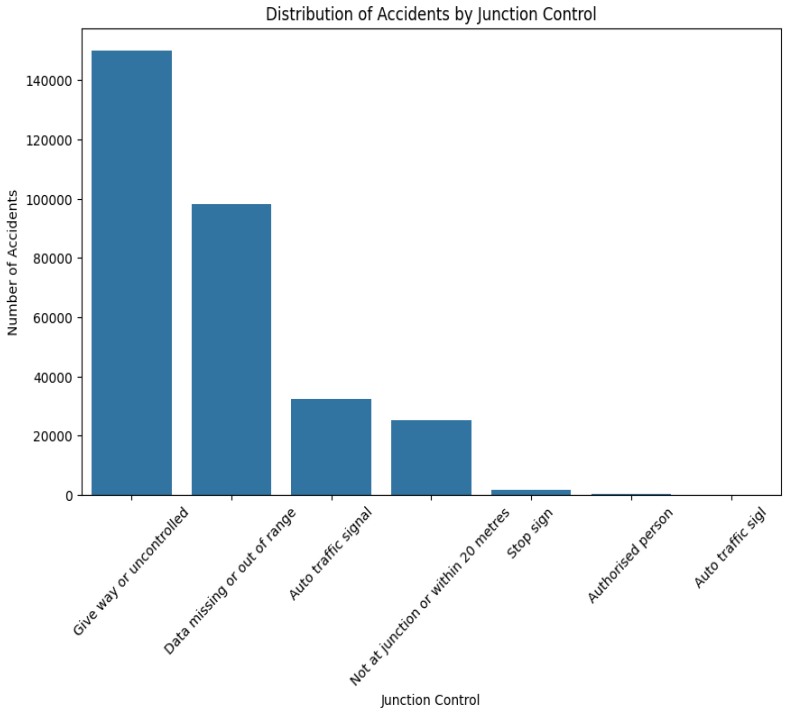
- The number of accidents varies throughout the week, with the highest number of accidents occurring on Monday (40,000) and the lowest number of accidents occurring on Sunday (10,000).
- The "Slight" category has the highest number of accidents on Monday (20,000) and Tuesday (15,000), while the "Serious" category has the highest number of accidents on Monday (15,000) and Friday (10,000).
- The "Fetal" category has the highest number of accidents on Wednesday (5,000), and the "Fatal" category has the highest number of accidents on Monday (5,000).



Analysis of the Distribution of Accidents by Road Type and Severity

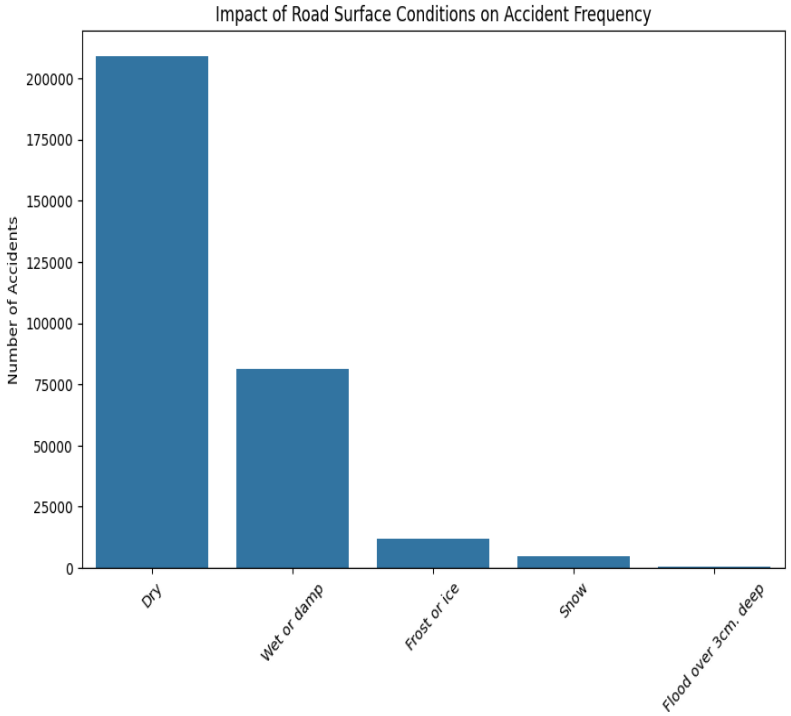
- The distribution of accidents by road type and severity shows that the majority of accidents are slight in severity, regardless of the road type.
- However, the percentage of accidents in the "Serious" and "Fatal" categories is higher on single carriageways compared to other road types.

Analysis of the Distribution of Accidents by Junction Control and Severity



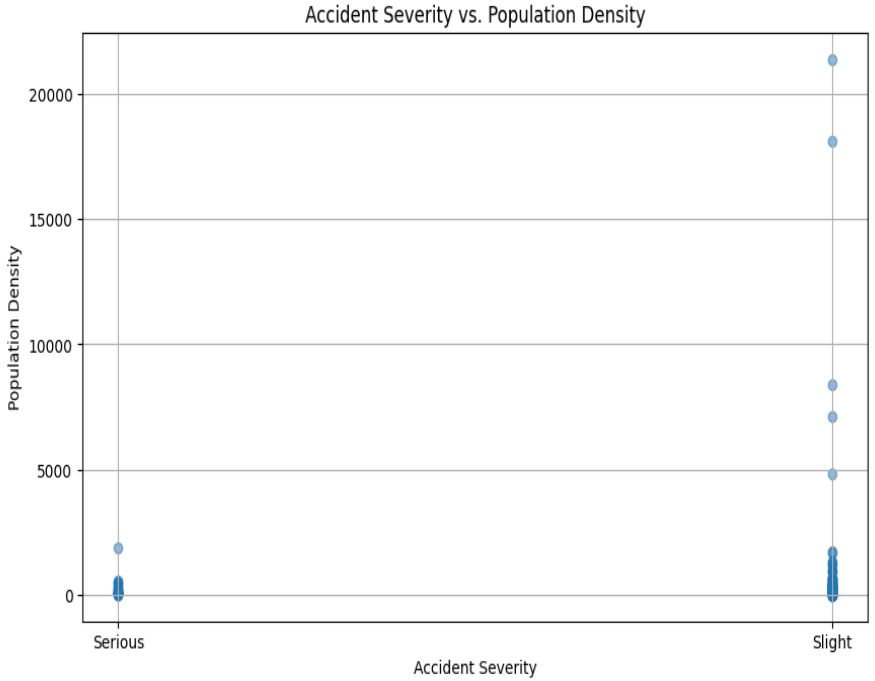
- The distribution of accidents by junction control and severity shows that the majority of accidents are slight in severity, regardless of the junction control.
- However, the percentage of accidents in the "Serious" and "Fatal" categories is higher at junctions with give way or uncontrolled control compared to other junction controls.

Analysis of the Relationship between Accident Severity and Road Surface Conditions



- The percentage of accidents in the "Slight" category is highest on all road surface conditions, ranging from 50% on wet or damp roads to 60% on roads with frost or ice.
- The percentage of accidents in the "Serious" category is second-highest on all road surface conditions, ranging from 25% on wet or damp roads to 20% on roads with frost or ice.

Accident severity vs population density

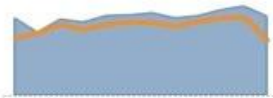


This visualization shows a striking correlation between accident severity and population density. Areas where accidents tend to be somewhat severe have significant population concentrations. This indicates that a higher population density can cause more frequent but less severe accidents, possibly due to increased traffic and infrastructure. In contrast, areas with severe accidents have a lower population density. This could indicate factors such as fewer vehicles on the road or better safety measures in less populated areas. Overall, the data show a complex interplay between population density and crash severity, highlighting the need for tailored safety measures in different urban environments..

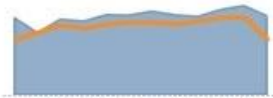
UNITED KINGDOM ACCIDENT DASHBOARD

Current Year: 2022 | Previous Year: 2021 | Accident Severity: All

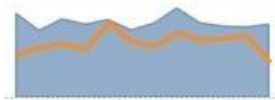
Total Accidents
144,419
↓11.70% YoY



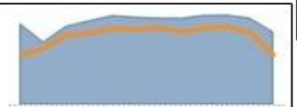
Total Casualties
195,737
↓11.89% YoY



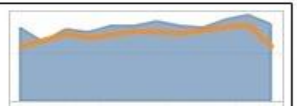
Fatal Casualties
2,855
↓26.40% YoY



Serious Casualties
27,045
↓16.30% YoY



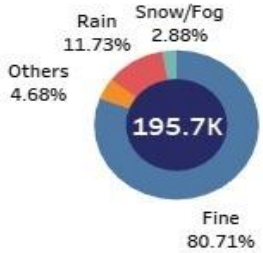
Slight Casualties
165,837
↓10.82% YoY



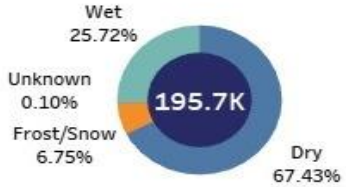
All Casualties By Vehicle Type.

15,579	6,573	155,804	1,477	399	15,905
↓13.89%	↑5.59%	↓12.31%	↓24.14%	↓36.97%	↓9.46%

All Casualties By Weather



All Casualties By Road



All Casualties By Location

All Casualties By Road Type

Recommendations

- Weather Conditions: While the relationship between weather conditions and the number of casualties is complex, it's clear that high winds and precipitation may be associated with a higher number of casualties.
- Therefore, it's recommended to increase safety measures during such weather conditions, such as reducing speed limits, increasing visibility measures, and encouraging drivers to be more cautious.

Recommendations continuation

- Urban vs Rural Areas: Since there are more casualties in urban areas than in rural areas, regardless of density, it's recommended to implement stricter traffic regulations in urban areas.
- This could include stricter enforcement of speed limits, more traffic signals and signs, and increased pedestrian safety measures.

Recommendations continuation

- Population Density: Given that the number of casualties in urban areas decreases as the density decreases, it's recommended to consider population density when planning urban development.
- This could involve creating more green spaces to reduce density, or implementing traffic management measures in high-density areas.

Recommendations continuation

- Accident Severity and Casualties: Implement measures to reduce accidents and casualties across all severity categories, with a particular focus on fatal accidents. Enhance emergency response and medical services to improve outcomes for severe accidents.
- Day of the Week: Since the number of accidents varies throughout the week, with the highest number of accidents occurring on Monday, it's recommended to increase safety measures and traffic enforcement on Mondays.

Recommendations continuation

- Light Conditions: Given that the number of accidents varies depending on the light conditions, with the highest number of accidents occurring in darkness with lighting unknown, it's recommended to improve lighting conditions on roads, especially in areas with high accident rates.
- Road Type and Junction Control: Since the percentage of accidents in the "Serious" and "Fatal" categories is higher on single carriageways and at junctions with give way or uncontrolled control, it's recommended to implement stricter traffic regulations and safety measures on these road types and junctions.

Recommendations continuation

- Speed Limit: Given the relationship between the accident severity and the speed limit, it's recommended to enforce speed limits more strictly, especially at a speed limit of 60 kph
- Road Surface Conditions: Since the percentage of accidents in the "Serious" and "Fatal" categories is higher on roads with frost or ice, it's recommended to improve road maintenance during winter months and to encourage drivers to be more cautious in such conditions.

Recommendations continuation

- Vehicle Type: Given the relationship between the vehicle type and the accident severity, it's recommended to implement stricter safety regulations for motorcycles over 500cc and agricultural vehicles.
- This could include stricter licensing requirements, more rigorous vehicle inspections, and increased safety features on these vehicles.

Project management

NAME	STUDENT ID	CONTRIBUTION
Sahithi Vidiyala	11717331	Data inspection and pre-processing
Sathwika Karingu	11708107	Data visualization using d3
Lakshmi Siri Chandana Jasti	11704788	Data visualization using tableu
Rohit Ibrahimpatnam	11719429	Data visualization using python

References

- https://www.freecodecamp.org/learn/data-analysis-with-python
- https://help.tableau.com/current/pro/desktop/en-us/analyze.htm
- https://www.freecodecamp.org/news/d3js-tutorial-data-visualization-for-beginners/