UK Accidents Analysis

Mohini Patil - 015359188

Department of Applied Data Science, San Jose State University

DATA 230: Data Visualization

Instructor: Andrew H. Bond

December 12, 2021

Abstract

According to the World Health Organization (WHO), approximately 1.25 million people die in road accidents every year. Also, road traffic accidents rank 9th place as the leading cause of death. It also accounts for 2.2% of all deaths globally. So, with the help of data visualization, I wanted to develop some prevention mechanisms for UK road safety. I collected the UK traffic accident dataset from Kaggle. I used tableau prep for cleaning the data, as the dataset contains some null records, missing values, etc. Then used tableau desktop for analysis. So, I created some dashboards on the tableau desktop. Also created dashboards using the extension. In that, used sunburst chart, bar race chart, and tree representation to show the distribution of accidents. From these dashboards, we understand a few points about traffic accidents in the UK. Like the Birmingham district and Metropolitan police force is the highest prone area, so if possible UK traffic department can employ more strict rules in these areas. Also, Single carriageways can be modified to the possible extent. More awareness needs to provide in the people with age group 26 to 35 years. Public transport can be promoted on Fridays. The speed limit can be enforced strictly.

Keywords: Road safety, tableau prep, tableau desktop, accidents

Table of Contents

I. Introduction

I.1 Purpose

The purpose of analyzing road accidents data is to get insights from the data using data science methodology and also develop some prevention mechanisms for traffic accidents and road safety. To identify factors responsible for most of the accidents. To identify the cause and effect of accidents on different conditions.

I.2 Intended audience

The audience for this project will be the general public concerning road safety in the UK. Also, Highway Operators and Public Work Department Personnel.

I.3 Scope

The scope of this project is to analyze the UK Road Safety data from the year 2005-2016 to answer the following questions:

1. What district in the UK has the highest number of accidents?
2. Did the police officer attend the accident site?
3. What are the effects of different light, weather, and road conditions on accident rates?
4. What vehicles are involved in accidents by make, model, and vehicle type?
5. What is the accident rate by age of the driver?
6. What are the number of accidents that happened by year, day of the week, and time of the day?

II. Background and Objectives

Road safety is a major public health concern because there are many road crash fatalities and disabilities recognized every day. Also, according to the World Health Organization (WHO), approximately 1.25 million people die in road accidents per year, on average 3,287 deaths per day. Also, road traffic accidents rank 9th place as the leading cause of death. It also accounts for 2.2% of all deaths globally. It is very important to solve this problem using data science. The objective here is to clean data using tableau prep and then use tableau desktop to analyze. Develop dashboards and stories that will help answer some of the questions that are added in scope.

Domain Application: Traffic Accident Analysis

Domain Users: Highway Operators and Public Work Department Personnel

- **Literature Review**

Paper 1: Road Traffic Accident Data Analysis and Its Visualization

Authors: Muhammad Babar Ali Rabbani , Muhammad Ali Musarat , Wesam Salah Alaloul, Ahsen Maqsoom, Hamna Bukhari, Waqas Rafiq

Description

In this paper, they used data from Peshawar, Pakistan. After collecting the dataset, the data is first filtered, pre-processed, and then performed visualization. From the analysis, they found that more accidents were caused in the 30 to 45 years of age group.

Paper 2: Crash Prediction and Risk Evaluation Based on Traffic Analysis Zones

Authors: Cuiping Zhang, Xuedong Yan, Lu Ma, and Meiwu An

Description

In this paper, they are measuring the crash risk at the zonal level. They are building a model that will estimate significant factors for the unsafe zones. The data used for this is from the Pikes Peak Area Council of Governments, Colorado. This data has three types of datasets: the TAZ dataset, the traffic and roadway dataset, and the accident datasets.

Paper 3: A Review of Data Analytic Applications in Road Traffic Safety. Part 1: Descriptive and Predictive Modeling

Authors: Amir Mehdizadeh, Miao Cai, Qiong Hu, Mohammad Ali Alamdar Yazdi, Nasrin Mohabbati-Kalejahi, Alexander Vinel, Steven E. Rigdon, Karen C. Davis, Fadel M. Megahed

Description

In this paper, they are building a predictive model that will predict the risk of a crash based on different driving conditions. They are also finding techniques that will minimize the crash risk with the help of path or route selection and rest-break scheduling.

III. Architecture and High-Level Design

To analyze and develop a prevention mechanism, we first need to collect datasets. After collection, data need to be cleaned. Cleaning will involve removing null values, changing the data type of some variables, deleting unwanted columns, etc. Once all the datasets are cleaned, we need to combine them together. Then with the help of tableau desktop, create dashboards and stories. Finally, once the story is complete, we need to present it to the audience. Figure 1 shows the data flow for this project and Figure 2 is the high-level design.

Figure 1

Data Flow Diagram

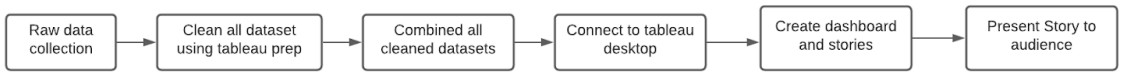
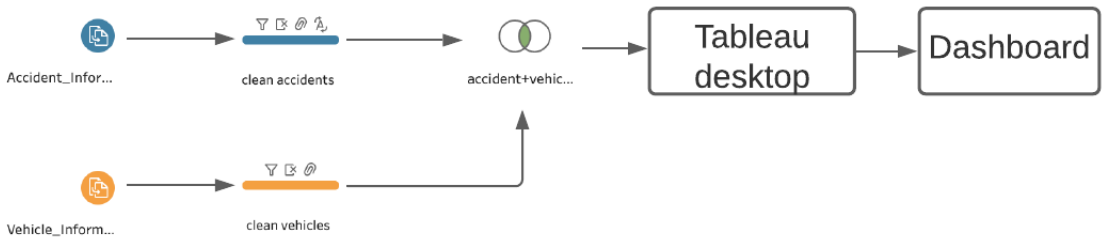


Figure 2

High-Level Design



Tools

- **Tableau Prep**

Used tableau prep for cleaning the data. Used tableau prep because it is easy to use. We don't need to write code. It is a drag and drops application. It is free for students. In tableau prep, we can combine, clean, shape, and share data with others. To combine different datasets, we can use the join or union options. Another feature is the pivot which can be helpful to see results of the subset of the dataset that we want. Then the clean features in which different options are available. These options include changing data roles, filter, aggregating and group, and replacing. Once the cleaning is done, we can run the flow to see the output and cleaned data that is created by tableau prep. Also, we can preview the cleaned data to the tableau desktop. Once we are confident that the dataset is cleaned enough then we can save the output and start analyzing it into tableau desktop. We can also publish the flow on the server and can also schedule when to run flow, how often, and who can have access to the flow. There are other features which are monitoring and alerting.

- **Tableau Desktop**

Used tableau desktop for analysis of data. Created worksheets. Created dashboard by joining different worksheets to analyze problems. Then after creating all dashboards, create a story in tableau which will be useful to tell a whole story. We used the tableau desktop because it was easy to use, and has many features. The features like it were very efficient for getting insights fast. It has many options to connect to data. It is able to give answers to deeper questions. We can also able to add data to the map.

IV. Development Process

Methodology

- **Data Collection**

The dataset was obtained from Kaggle for the 2005-2016 years. This dataset is published by the Department of Transports, under Open Government Licence. There are two datasets that we are using for this project. The first one is the Accident_Information dataset which has information about the accidents, like districts, police force areas, roads, weather, light conditions, the severity of accidents, etc. The second dataset is the Vehicle_Information which contains information about the vehicles present in the accidents like vehicles age, their makers, type, etc. First, we study and understand the features presents in both of these datasets. Table 1 shows the different variables, their data type, and a description of the features present in the Accident_Information dataset. Figure 3 shows the raw samples from this dataset. Table 2 shows the different variables, their data type, and a description of the features present in the Vehicle_Information dataset. Figure 4 shows the few raw samples from this dataset.

Table 1

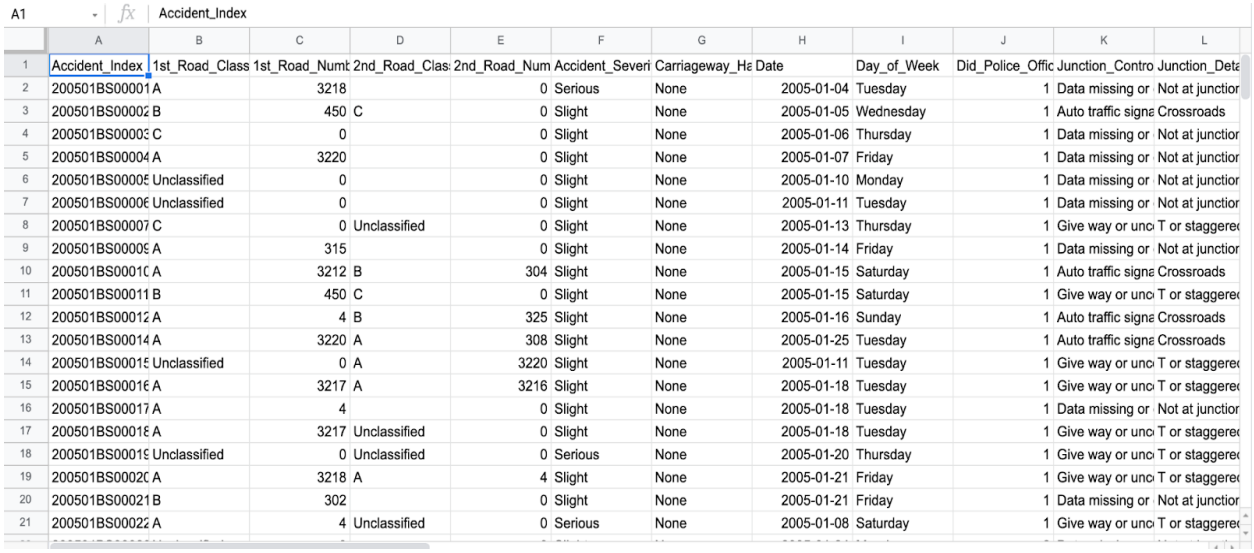
Description of Features in Accident_Information dataset

Attribute	Attribute Type	Details
Accident_Index	Ordinal	Unique identifier index of the accident
1st_Road_Class	Categorical	Road classes (6 levels)
Accident_Severity	Categorical	3 levels of severity (fatal, serious, slight)
Carriageway_Hazards	Categorical	Hazards details (7 values)
Date	Ordinal	Date of the accident
Day_of_Week	Categorical	Day of the accident
Did_Police_Officer_Attend_Scene_of_Accident	Categorical	2 values yes or no, whether police attended the accident sites or not
Junction_Control	Categorical	Junction control
Junction_Detail	Categorical	Junction details
Latitude	Quantitative	Latitude coordinate of the accidents
Light_Conditions	Categorical	Light conditions when the accident happened
Local_Authority_(District)	Categorical	District in which accident happened
Local_Authority_(Highway)	Categorical	Highway in which accident happened
Longitude	Quantitative	Longitude coordinate of the accidents
Number_of_Casualties	Quantitative	Number of casualties resulted from accidents
Number_of_Vehicles	Quantitative	Number of vehicles involved in accidents
Pedestrian_Crossing-Human_Control	Quantitative	Number of pedestrians crossing human control
Pedestrian_Crossing-Physical_Facilities	Quantitative	Number of pedestrian crossing physical facilities
Police_Force	Categorical	Police force area

Attribute	Attribute Type	Details
Road_Surface_Conditions	Categorical	5 categories (dry, wet or damp, snow, frost or ice, flood over 3 cm deep)
Road_Type	Categorical	5 types of road (single carriageway, dual carriageway, roundabout, one-way street, slip road)
Special_Conditions_at_Site	Categorical	9 categories
Speed_limit	Quantitative	The speed limit on the road
Time	Ordinal	Time of the day in 24 hours format
Urban_or_Rural_Area	Categorical	2 categories (urban or rural)
Weather_Conditions	Categorical	8 categories
Year	Ordinal	2005-2017
InScotland	Categorical	Binary (yes or no)

Figure 3

Raw Sample Data from Accident_Information File



A1 | Accident_Index

	A	B	C	D	E	F	G	H	I	J	K	L
1	Accident_Index	1st_Road_Class	1st_Road_Numb	2nd_Road_Clas	2nd_Road_Num	Accident_Severi	Carriageway_Ha	Date	Day_of_Week	Did_Police_Offic	Junction_Contro	Junction_Deta
2	200501BS00001	A	3218		0	Serious	None	2005-01-04	Tuesday	1	Data missing or	Not at junctior
3	200501BS00002	B	450	C	0	Slight	None	2005-01-05	Wednesday	1	Auto traffic signa	Crossroads
4	200501BS00003	C	0		0	Slight	None	2005-01-06	Thursday	1	Data missing or	Not at junctior
5	200501BS00004	A	3220		0	Slight	None	2005-01-07	Friday	1	Data missing or	Not at junctior
6	200501BS00005	Unclassified	0		0	Slight	None	2005-01-10	Monday	1	Data missing or	Not at junctior
7	200501BS00006	Unclassified	0		0	Slight	None	2005-01-11	Tuesday	1	Data missing or	Not at junctior
8	200501BS00007	C	0	Unclassified	0	Slight	None	2005-01-13	Thursday	1	Give way or unc	T or staggerec
9	200501BS00009	A	315		0	Slight	None	2005-01-14	Friday	1	Data missing or	Not at junctior
10	200501BS0001C	A	3212	B	304	Slight	None	2005-01-15	Saturday	1	Auto traffic signa	Crossroads
11	200501BS00011	B	450	C	0	Slight	None	2005-01-15	Saturday	1	Give way or unc	T or staggerec
12	200501BS00012	A	4	B	325	Slight	None	2005-01-16	Sunday	1	Auto traffic signa	Crossroads
13	200501BS00014	A	3220	A	308	Slight	None	2005-01-25	Tuesday	1	Auto traffic signa	Crossroads
14	200501BS00015	Unclassified	0	A	3220	Slight	None	2005-01-11	Tuesday	1	Give way or unc	T or staggerec
15	200501BS00016	A	3217	A	3216	Slight	None	2005-01-18	Tuesday	1	Give way or unc	T or staggerec
16	200501BS00017	A	4		0	Slight	None	2005-01-18	Tuesday	1	Data missing or	Not at junctior
17	200501BS00018	A	3217	Unclassified	0	Slight	None	2005-01-18	Tuesday	1	Give way or unc	T or staggerec
18	200501BS00019	Unclassified	0	Unclassified	0	Serious	None	2005-01-20	Thursday	1	Give way or unc	T or staggerec
19	200501BS0002C	A	3218	A	4	Slight	None	2005-01-21	Friday	1	Give way or unc	T or staggerec
20	200501BS00021	B	302		0	Slight	None	2005-01-21	Friday	1	Data missing or	Not at junctior
21	200501BS00022	A	4	Unclassified	0	Serious	None	2005-01-08	Saturday	1	Give way or unc	T or staggerec

Table 2

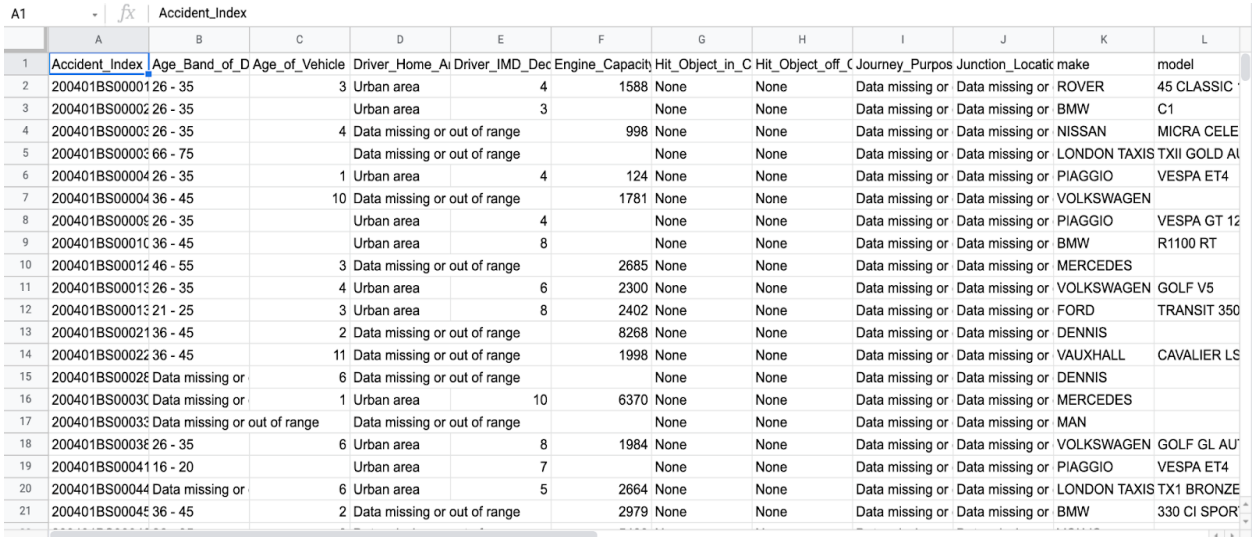
Description of Features in Vehicle_Information Dataset

Attribute	Attribute Type	Details
Accident_Index	Ordinal	Unique identifier index of the accident
Age_Band_of_Driver	Categorical	Different age bands of the driver (0 to 75)
Age_of_Vehicle	Quantitative	Age of vehicle
Driver_Home_Area_Type	Categorical	3 categories (rural, small town, urban area)
Engine_Capacity_.CC.	Quantitative	The engine capacity of the vehicle in cc
Hit_Object_in_Carriageway	Categorical	13 categories
Hit_Object_off_Carriageway	Categorical	13 categories
Journey_Purpose_of_Driver	Categorical	7 categories (commuting to/from work, riding to/from work, etc)
Junction_Location	Categorical	10 categories
Make	Categorical	308 different manufacturers of vehicle
Model	Categorical	The model name of the vehicle involved in an accident
Propulsion_Code	Categorical	10 categories (electric, gas, petrol, etc)
Sex_of_Driver	Categorical	Binary values (yes, no)
Skidding_and_Overturning	Categorical	6 categories
Towing_and_Articulation	Categorical	6 categories
Vehicle_Leaving_Carriageway	Categorical	9 categories
Vehicle_Location.Restricted_Lane	Quantitative	Lane number
Vehicle_Manoeuvre	Categorical	10 categories

Attribute	Attribute Type	Details
Vehicle_Reference	Quantitative	Reference number of the vehicle
Vehicle_Type	Categorical	16 categories (car, van, motorcycle, etc)
Was_Vehicle_Left_Hand_Drive	Categorical	Binary values (yes, no)
X1st_Point_of_Impact	Categorical	5 categories
Year	Quantitative	Year of the accident

Figure 4

Raw Sample Data from Vehicle_Information File



A1 | Accident_Index

	A	B	C	D	E	F	G	H	I	J	K	L
1	Accident_Index	Age_Band_of_D	Age_of_Vehicle	Driver_Home_Ar	Driver_IMD_Dec	Engine_Capacity	Hit_Object_in_C	Hit_Object_off_C	Journey_Purpos	Junction_Locatic	make	model
2	200401BS00001	26 - 35	3	Urban area	4	1588	None	None	Data missing or	Data missing or	ROVER	45 CLASSIC
3	200401BS00002	26 - 35		Urban area	3		None	None	Data missing or	Data missing or	BMW	C1
4	200401BS00003	26 - 35	4	Data missing or out of range		998	None	None	Data missing or	Data missing or	NISSAN	MICRA CELE
5	200401BS00003	66 - 75		Data missing or out of range			None	None	Data missing or	Data missing or	LONDON TAXIS	TXII GOLD AU
6	200401BS00004	26 - 35	1	Urban area	4	124	None	None	Data missing or	Data missing or	PIAGGIO	VESPA ET4
7	200401BS00004	36 - 45	10	Data missing or out of range		1781	None	None	Data missing or	Data missing or	VOLKSWAGEN	
8	200401BS00005	26 - 35		Urban area	4		None	None	Data missing or	Data missing or	PIAGGIO	VESPA GT 12
9	200401BS00010	36 - 45		Urban area	8		None	None	Data missing or	Data missing or	BMW	R1100 RT
10	200401BS00012	46 - 55	3	Data missing or out of range		2685	None	None	Data missing or	Data missing or	MERCEDES	
11	200401BS00013	26 - 35	4	Urban area	6	2300	None	None	Data missing or	Data missing or	VOLKSWAGEN	GOLF V5
12	200401BS00013	21 - 25	3	Urban area	8	2402	None	None	Data missing or	Data missing or	FORD	TRANSIT 350
13	200401BS00021	36 - 45	2	Data missing or out of range		8268	None	None	Data missing or	Data missing or	DENNIS	
14	200401BS00022	36 - 45	11	Data missing or out of range		1998	None	None	Data missing or	Data missing or	VAUXHALL	CAVALIER LS
15	200401BS00025	Data missing or	6	Data missing or out of range			None	None	Data missing or	Data missing or	DENNIS	
16	200401BS00030	Data missing or	1	Urban area	10	6370	None	None	Data missing or	Data missing or	MERCEDES	
17	200401BS00033	Data missing or out of range		Data missing or out of range			None	None	Data missing or	Data missing or	MAN	
18	200401BS00036	26 - 35	6	Urban area	8	1984	None	None	Data missing or	Data missing or	VOLKSWAGEN	GOLF GL AUT
19	200401BS00041	16 - 20		Urban area	7		None	None	Data missing or	Data missing or	PIAGGIO	VESPA ET4
20	200401BS00044	Data missing or	6	Urban area	5	2664	None	None	Data missing or	Data missing or	LONDON TAXIS	TX1 BRONZE
21	200401BS00045	36 - 45	2	Data missing or out of range		2979	None	None	Data missing or	Data missing or	BMW	330 CI SPOR

- **Data Cleaning**

Data cleaning is the process of detecting inaccurate, incorrect, incomplete records and then replacing, modifying, or deleting the noisy data. Noisy data is not useful for analysis. So, it is very important to clean it. For cleaning the dataset, I used tableau prep. Figure 5, and Figure 6 show the cleaned Accident_Information and cleaned Vehicle_Information dataset respectively. After cleaning both the dataset, combined them using inner join operation with common accident index column and created output of that. Figure 7 is the flow diagram that was generated by tableau prep. Figure 6 shows the join results of these two cleaned datasets.

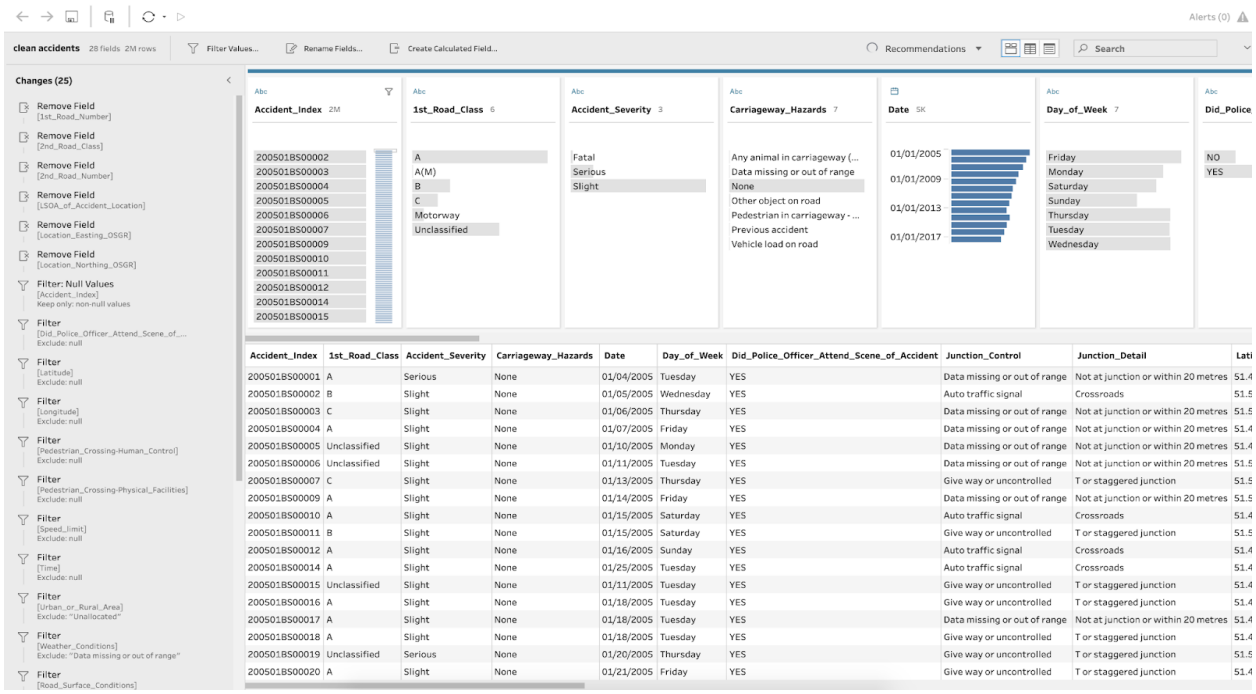
Accident_Information dataset

We did the following cleaning for the accident dataset:

- Removed unwanted columns from table like 1st_Road_Number, 2nd_Road_Class, 2nd_Road_Number, LSOA_of_Accident_Location, Location_Easting_OSGR, Location_Northing_OSGR.
- Removed null values from columns accident index, Did_Police_Officer_Attend_Scene_of_Accident, Latitude, Longitude, Pedestrian_Crossing-Human_Control, Pedestrian_Crossing-Physical_Facilities, Speed_Limit, Time, Urban_Or_Rural_Area, Weather_Conditions, Road_Surface_Conditions, Light_Conditions. As the null values for these columns were fewer, So I decided to remove them.
- Changed data type of Did_Police_Officer_Attend_Scene_of_Accident

Figure 5

Cleaned Accident_Information



Vehicle_Information dataset

We cleaned the vehicle dataset using the following steps:

- Removed null values from columns Accident-Index, model, Vehicle_Location.Restricted_Lane.
- Removed rows with “Data missing or out of range” from columns Age_Band_of_Driver, Driver_Home_Area_Type, Sex_of_Driver, Skidding_and_Overturning, Towing_and_Articulation, Vehicle_Leaving_Carriageway, Was_Vehicle_Left_Hand_Drive, X1st_Point_of_Impact.
- We removed the field named Driver_IMD_Decile.
- Removed rows with “Not known” from the Sex_of_Driver column.

Figure 6

Cleaned Vehicle_Information

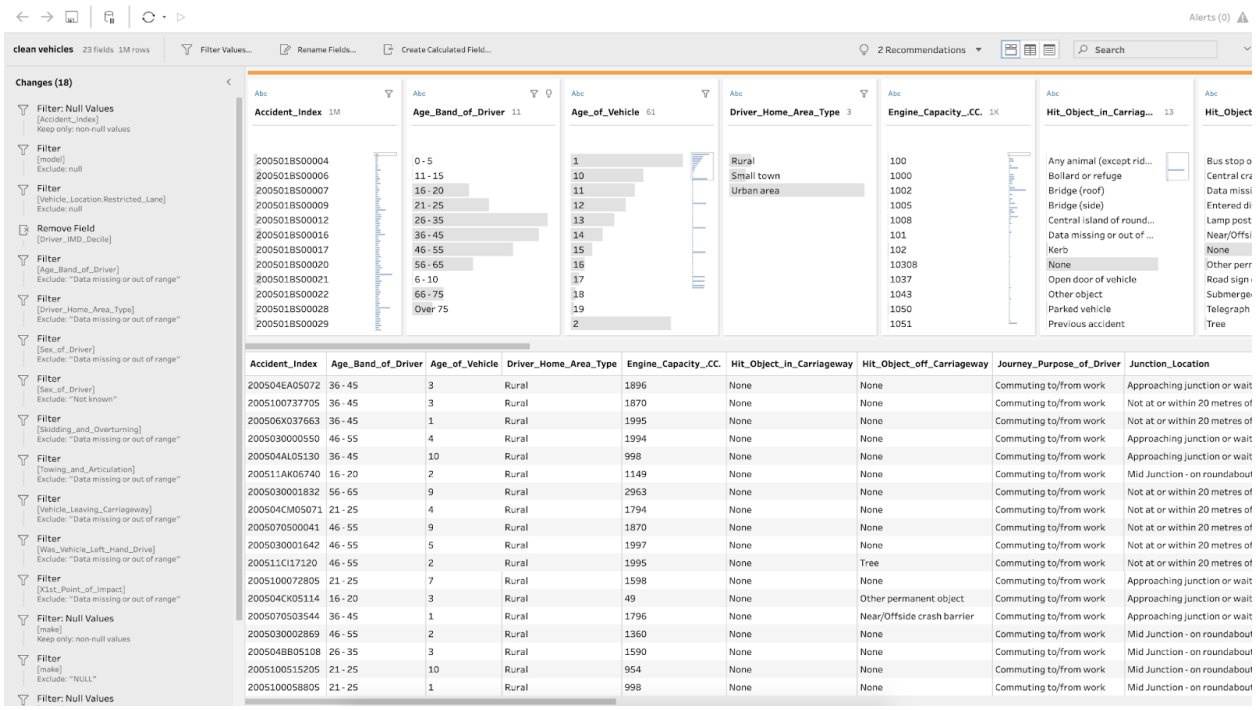
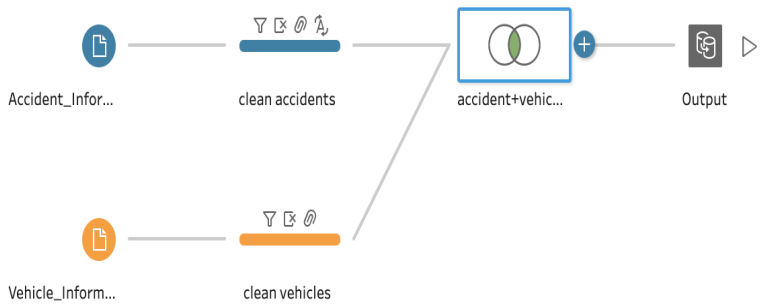


Figure 7

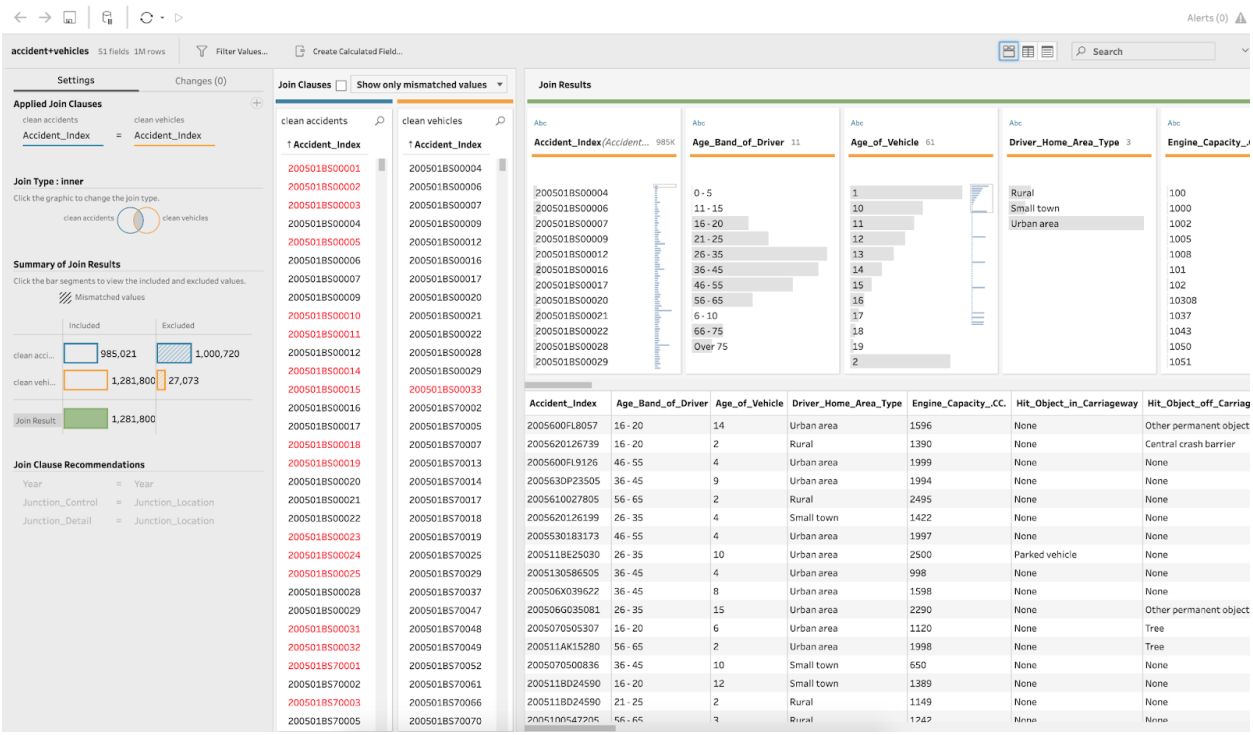
Tableau Prep Flow Diagram



Note. The above figure is generated from tableau prep after completing the cleaning.

Figure 8

Combined cleaned Accident_Information and Vehicle_Information dataset



Note. The above figure shows some sample data from combined cleaned Accident_Information and Vehicle_Information. The screenshot is from tableau prep.

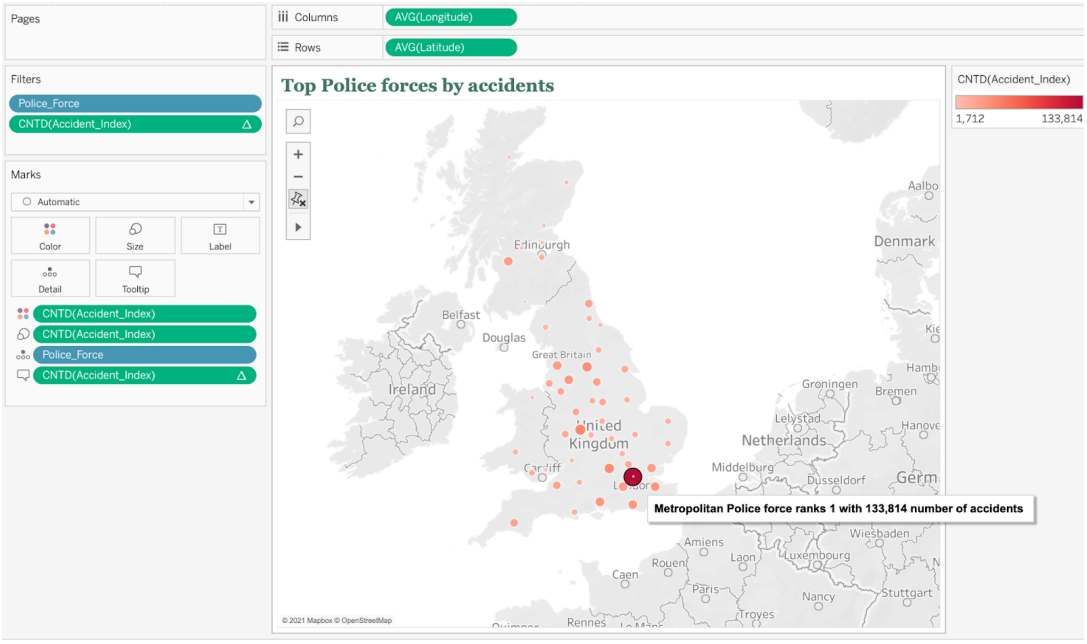
V. Deliverables

After cleaning and combining the dataset, the next step is to do data visualization with the help of a tableau desktop. I am using preview in the tableau desktop to analyze the cleaned dataset. Once clicked on the preview option, it will create two files .hyper and .tds. .hyper is the tableau extract file and .tds is the tableau data source file. We connected the .hyper file to the tableau desktop and try to visualize it. In tableau desktop, created different worksheets to analyze the dataset. Created different dashboards that will give answers to our questions. Added screenshots in the following figures to describe different worksheets and dashboards that were created in tableau desktop.

- Use case 1: Top areas in accidents

Figure 9

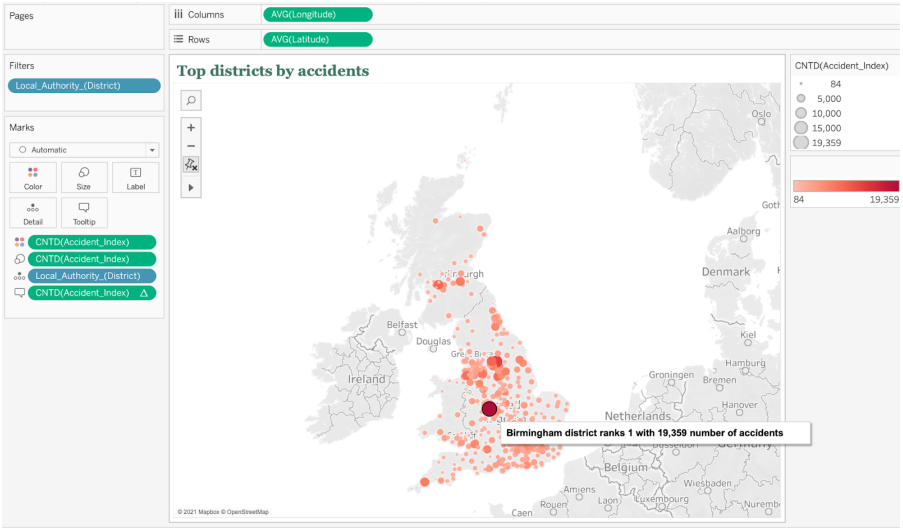
Worksheet for Top Police Force by Accidents



Note. The above worksheet shows the metropolitan police force area rank one in a number of accidents than other areas.

Figure 10

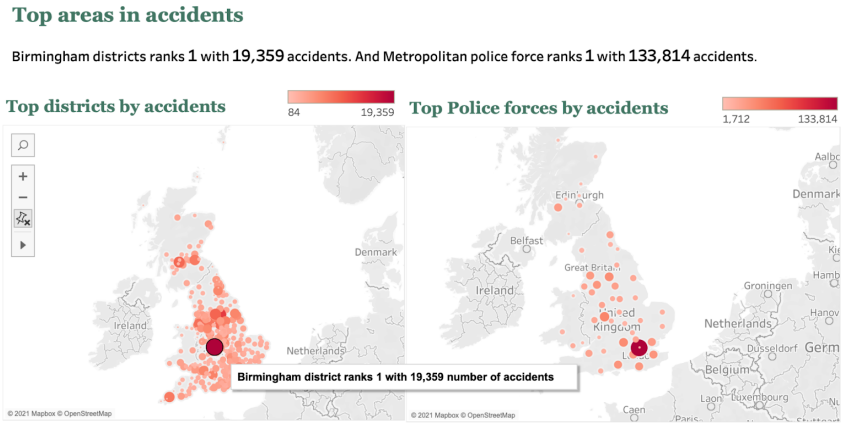
Worksheet for Top Districts by Accidents



Note. The above worksheet is to show different district areas by the number of accidents. As we can see Birmingham ranks number one.

Figure 11

Dashboard for Top districts and Police Forces by Accidents

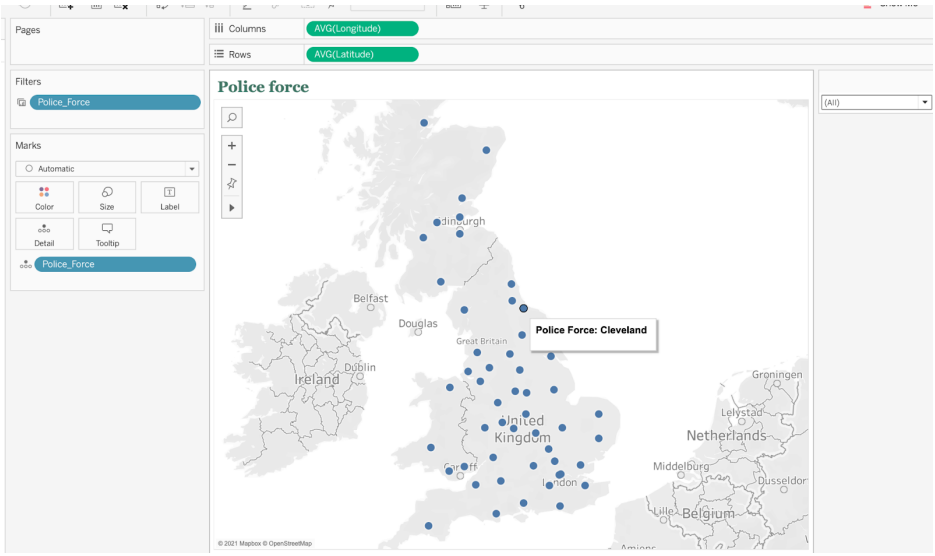


Note. The above dashboard shows the distribution of accidents in top district areas and top police force areas. From above, we can see Birmingham and Metropolitan ranks number one in district and police force respectively.

- Use case 2: Accidents by police forces, road type, and severity

Figure 12

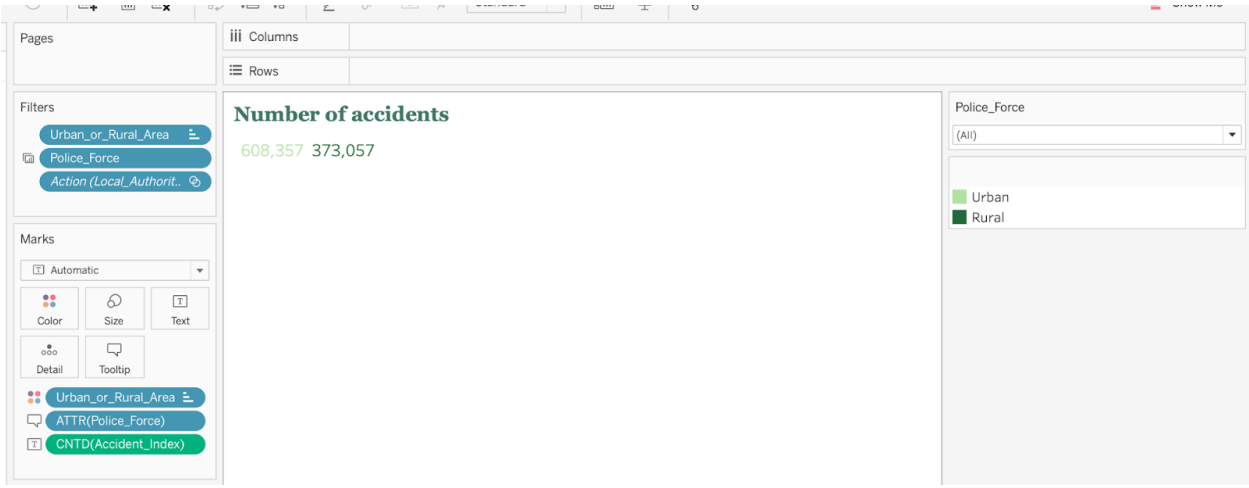
Worksheet for Police Force Areas



Note. This worksheet is to show different police force areas where accidents happened in the UK.

Figure 13

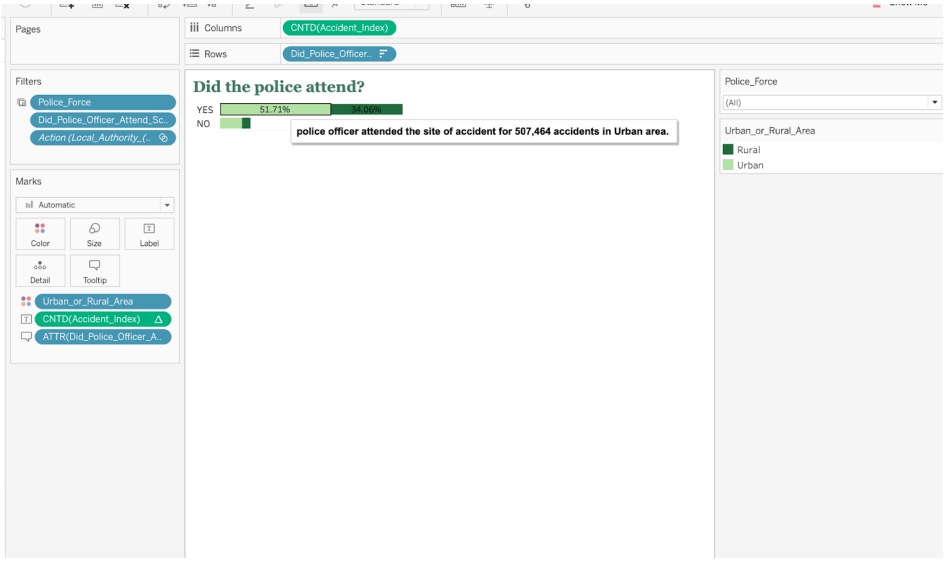
Worksheet for Number of Accidents in Urban and Rural Areas



Note. The above worksheet is to see the number of accidents in urban and rural areas. Light and dark green colors represent the urban and rural areas respectively throughout the visualization.

Figure 14

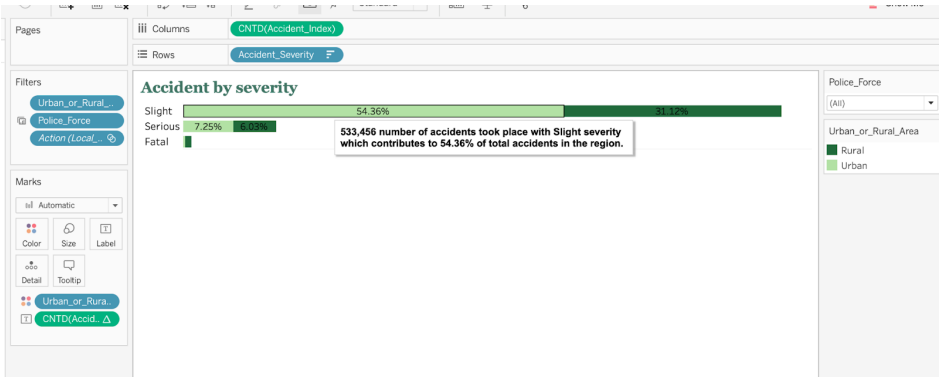
Worksheet for Did the Police Attend



Note. The above worksheet shows the percentage of accidents when police attended the accident site and also when they did not attend. From this, we can see most of the time police attended the accident site.

Figure 15

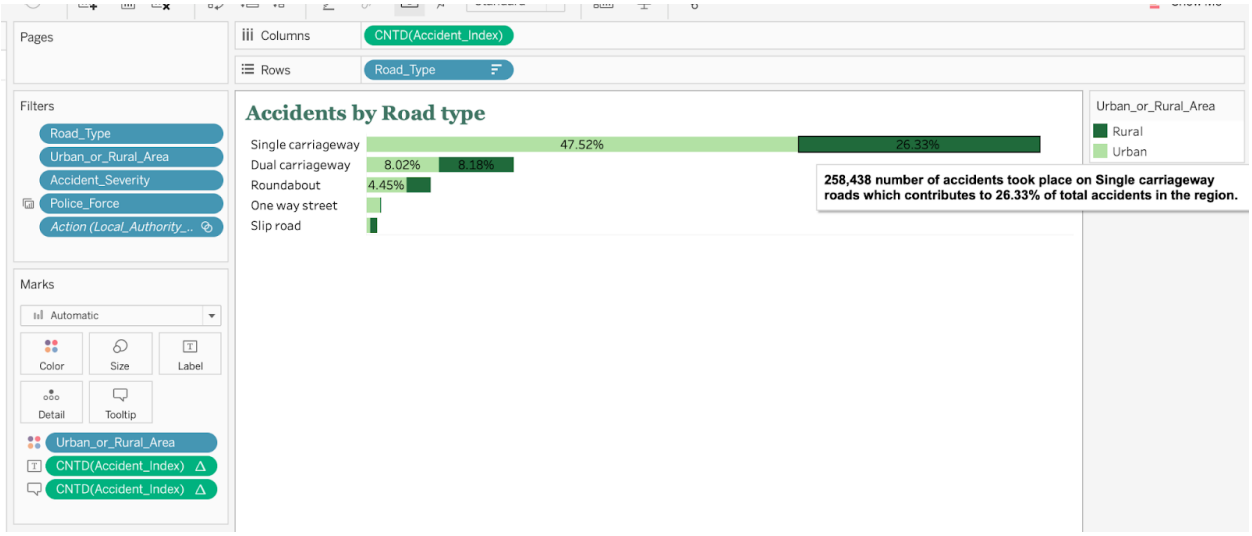
Worksheet for Accident by Severity



*Note***.** The above worksheet shows the percentage of accidents by severity type. From above, we can see a slight type of severity has the highest percentage of accidents which is not bad. Also, the fatal type has a low percentage.

Figure 16

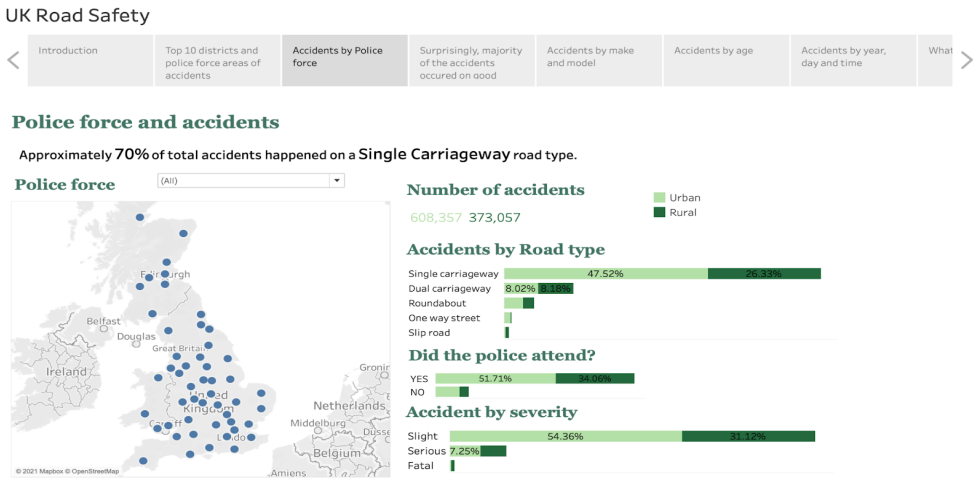
Worksheet for Accidents by Road Type



Note. The above worksheet shows the percentage of accidents by different road types. As we can see from above, single carriageway road type has the highest percentage.

Figure 17

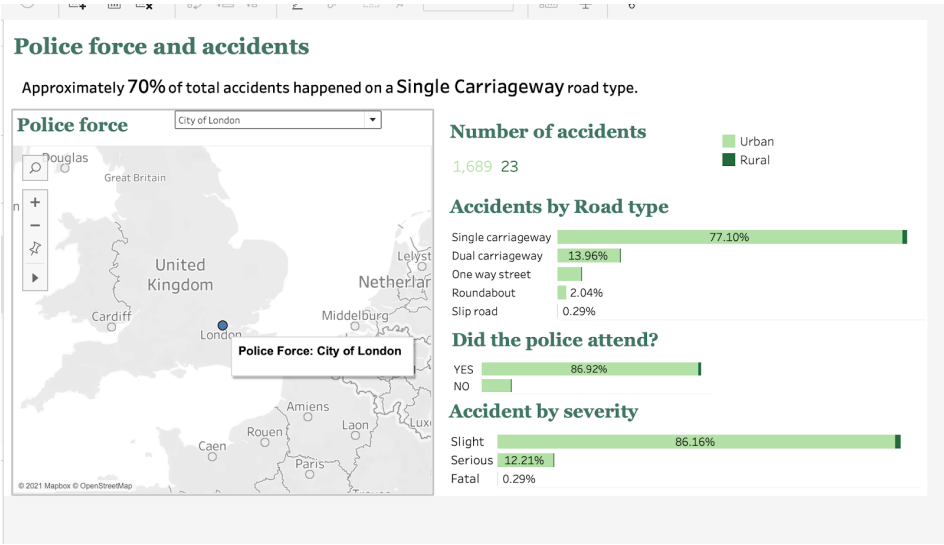
Dashboard for Accidents by Road Type, Police Force Areas in Urban and Rural Areas



Note. The above figure shows the percentage of accidents by road type, police forces, severity, and did the police attend the accident site or not. We also applied filters for urban or rural areas. Overall single carriageways contribute approximately 70% of total accidents.

Figure 18

Number of Accidents for City of London

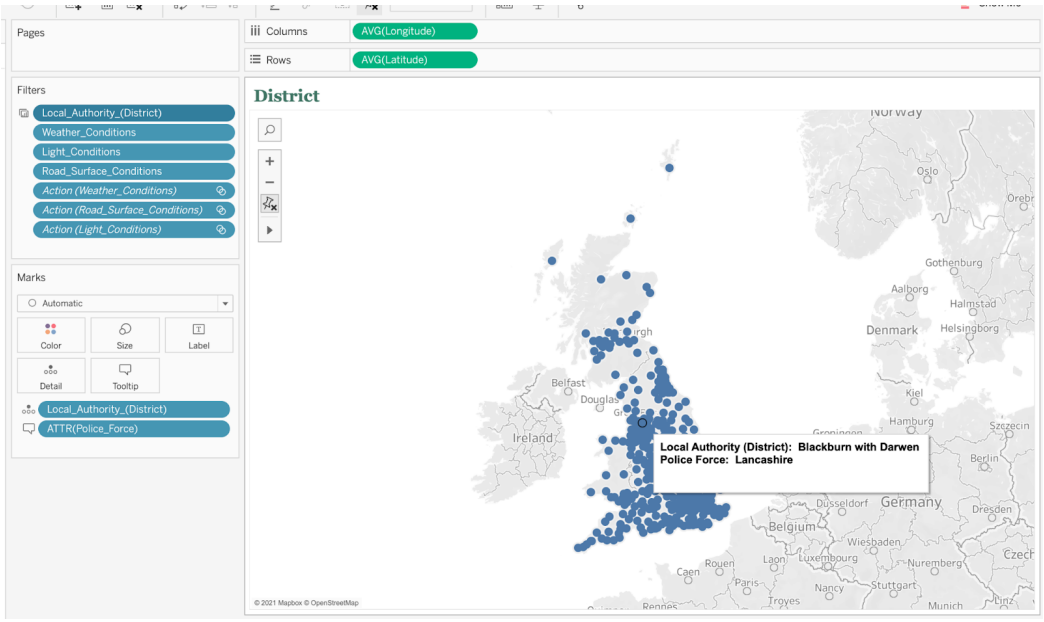


Note. From the above figure, we can see the accident rate for the police force area City of London.

- Use case 3: Accidents by districts and road, weather, and light conditions

Figure 19

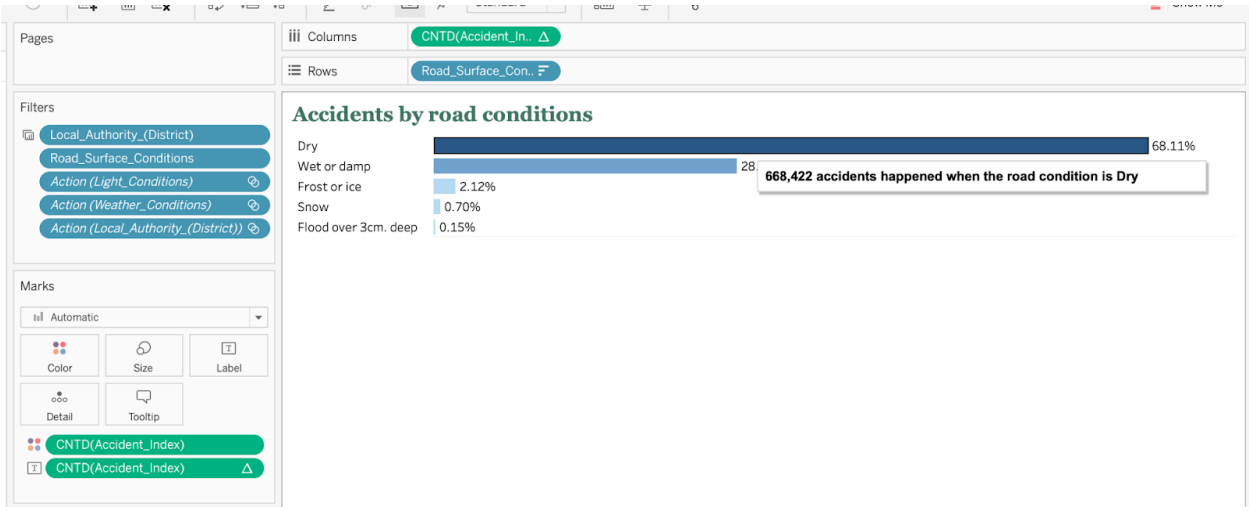
Worksheet for the Districts Involved in Accidents



Note. The above worksheet shows the different districts in the UK involved in accidents.

Figure 20

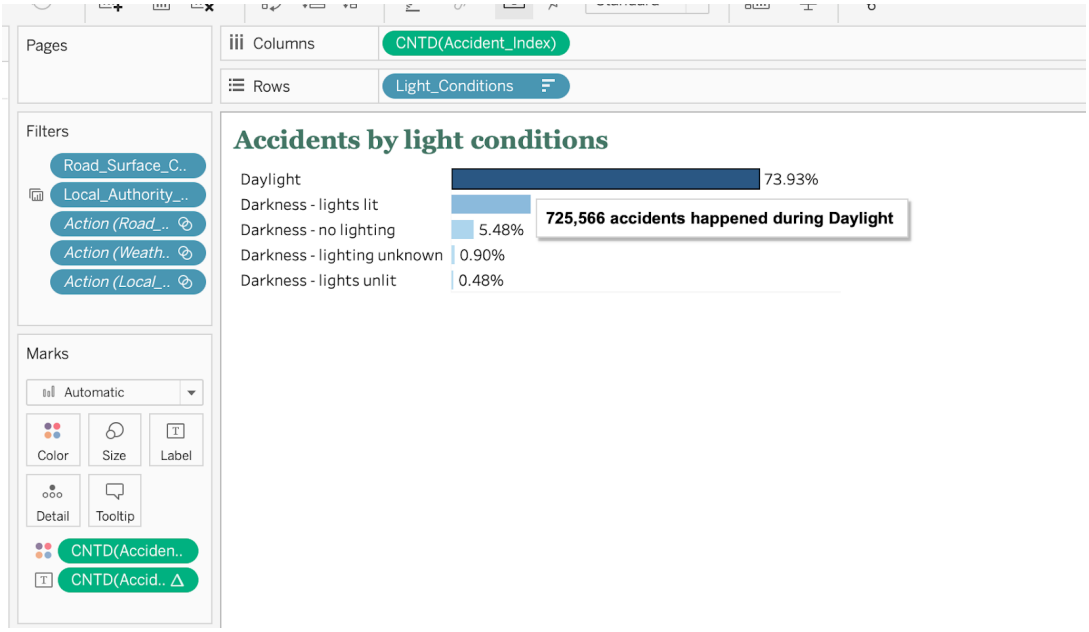
Worksheet for Accident by Road Conditions



Note. The above worksheet shows the percentage of accidents for different road conditions in the UK. As we can see, Dry-type of road conditions has the highest percentage of accidents.

Figure 21

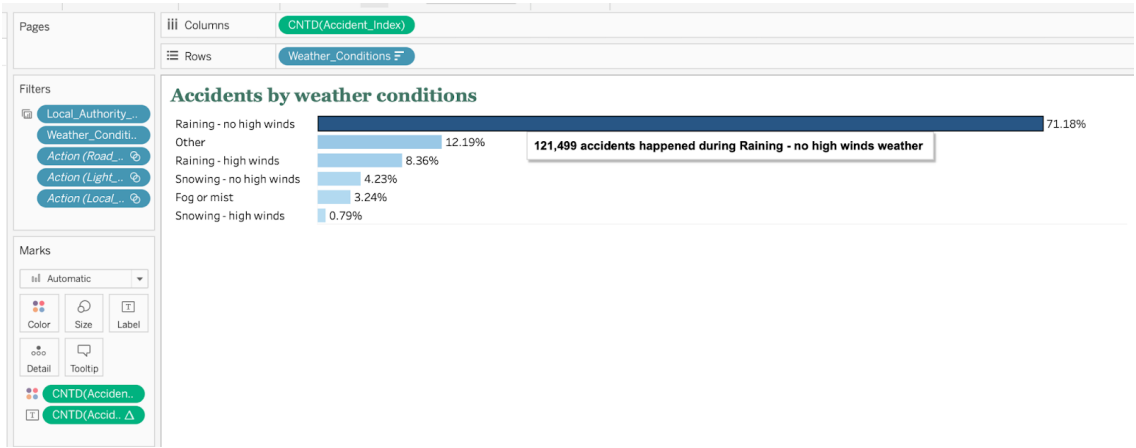
Worksheet of Accident by Light Conditions



Note. The above worksheet shows the accidents by different light conditions. As we can see from above, most accidents happened in daylight.

Figure 22

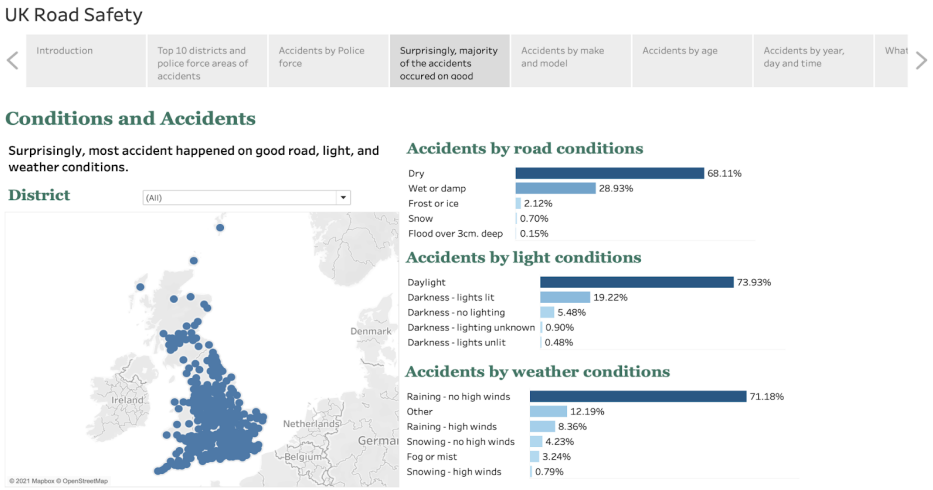
Worksheet of Accident by Weather Conditions



Note. The above worksheet shows the accidents by different weather conditions. As we can see from above, most accidents happened in rain with no high winds conditions.

Figure 23

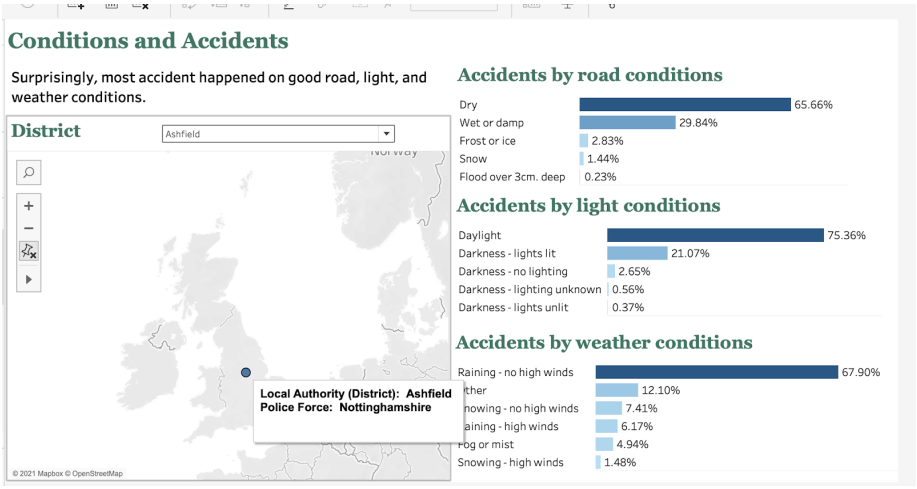
Dashboard for Different Conditions



Note. The above dashboard shows the accident rate by a different road, weather, and light conditions. As we can see, for all good types of conditions, there were more accidents. We need to find more information about this.

Figure 24

Different conditions for Ashfield District

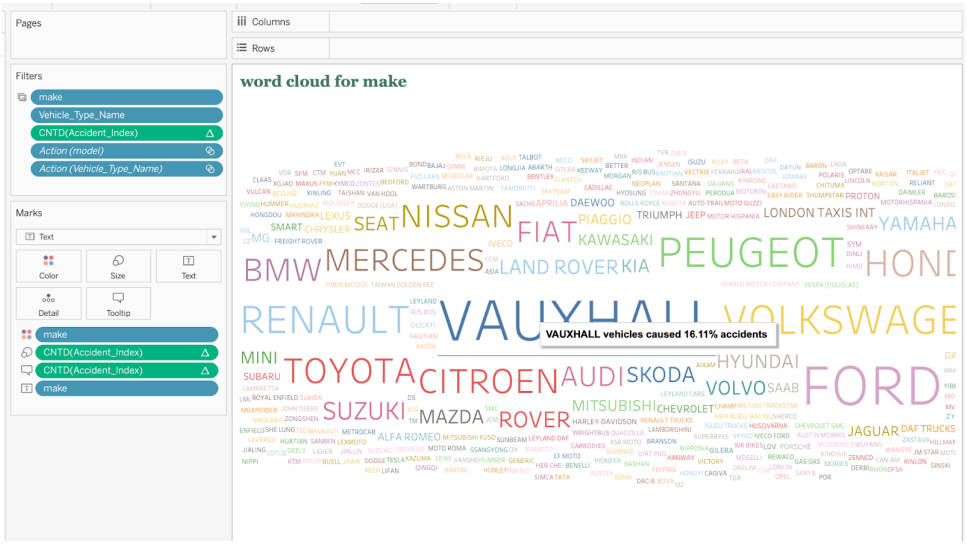


Note. In the above dashboard, select the Ashfield district from the drop-down option to see the accident rates for different conditions.

- Use case 4: Accidents by Vehicles maker, vehicle type, age of the vehicle

Figure 25

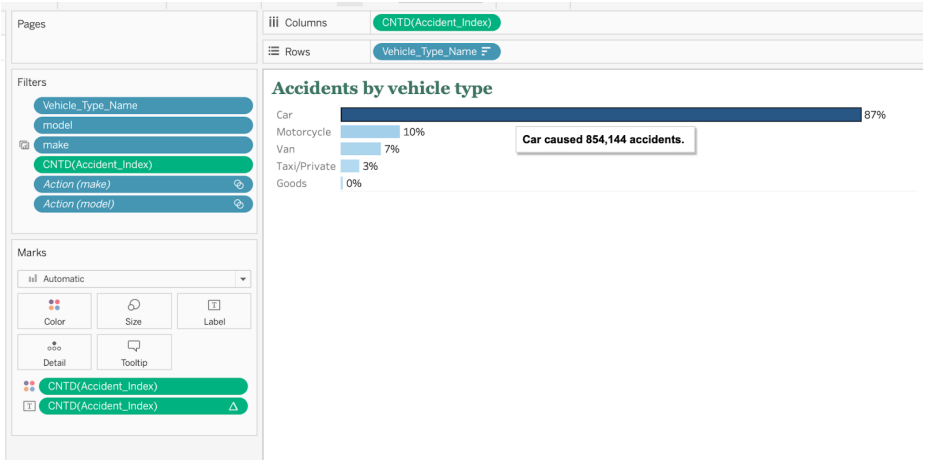
Worksheet for Word Cloud for Make



Note. The above worksheets show the different makers of vehicles in the UK. As we can see Vauxhall has the highest percentage of accidents and after that Ford vehicle comes.

Figure 26

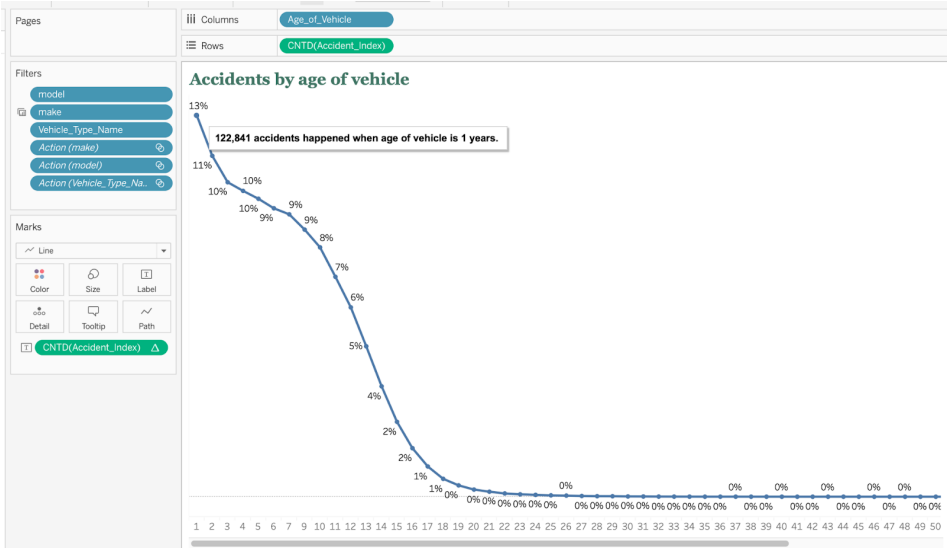
Worksheet Accidents by Vehicle Type



Note. The above worksheet shows the Accidents by different vehicle types. As we see car type of vehicle causes more accidents than any other type.

Figure 27

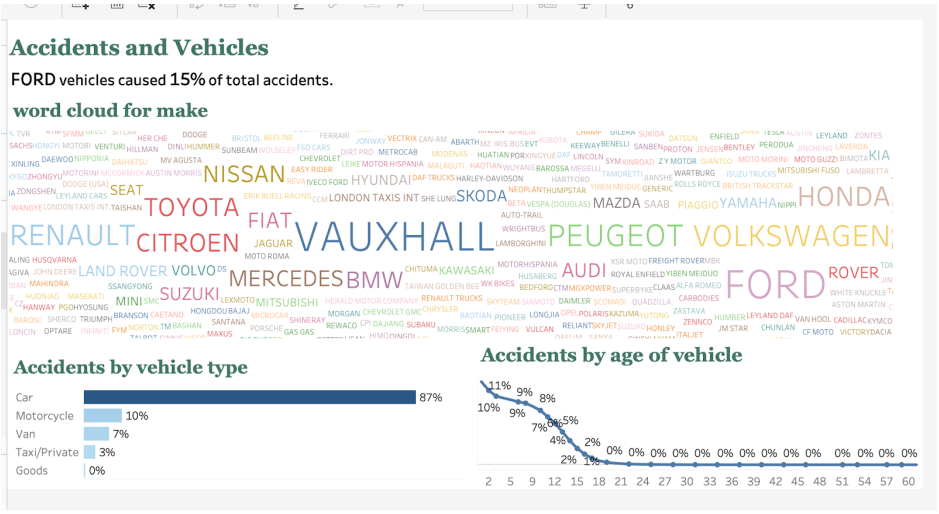
Worksheet for Accidents by Age of the Vehicle



Note. The above worksheet shows the accidents by age of vehicle. As we can see from above, new vehicles caused more accidents than older vehicles.

Figure 28

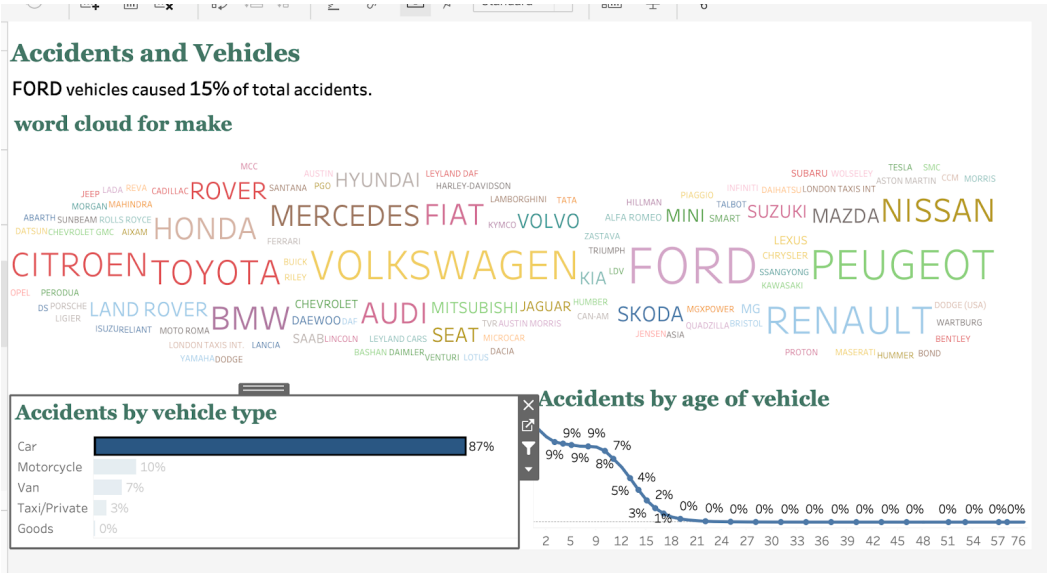
Dashboard for Accidents and Vehicles



Note. The above dashboard shows the accident by different makers of vehicles, different type of vehicles involved in accidents, and how the age of the vehicles affect accidents in the UK.

Figure 29

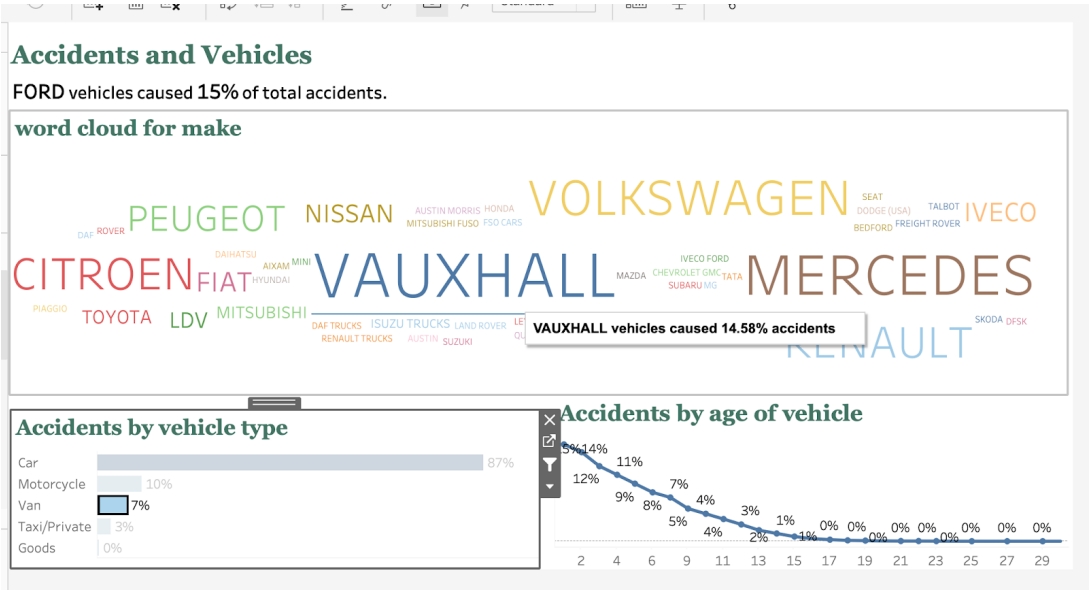
Accidents for Car Type of Vehicle



Note. From above, we can see what type of makers are there for car types of vehicles. Ford makers caused more accidents in car type.

Figure 30

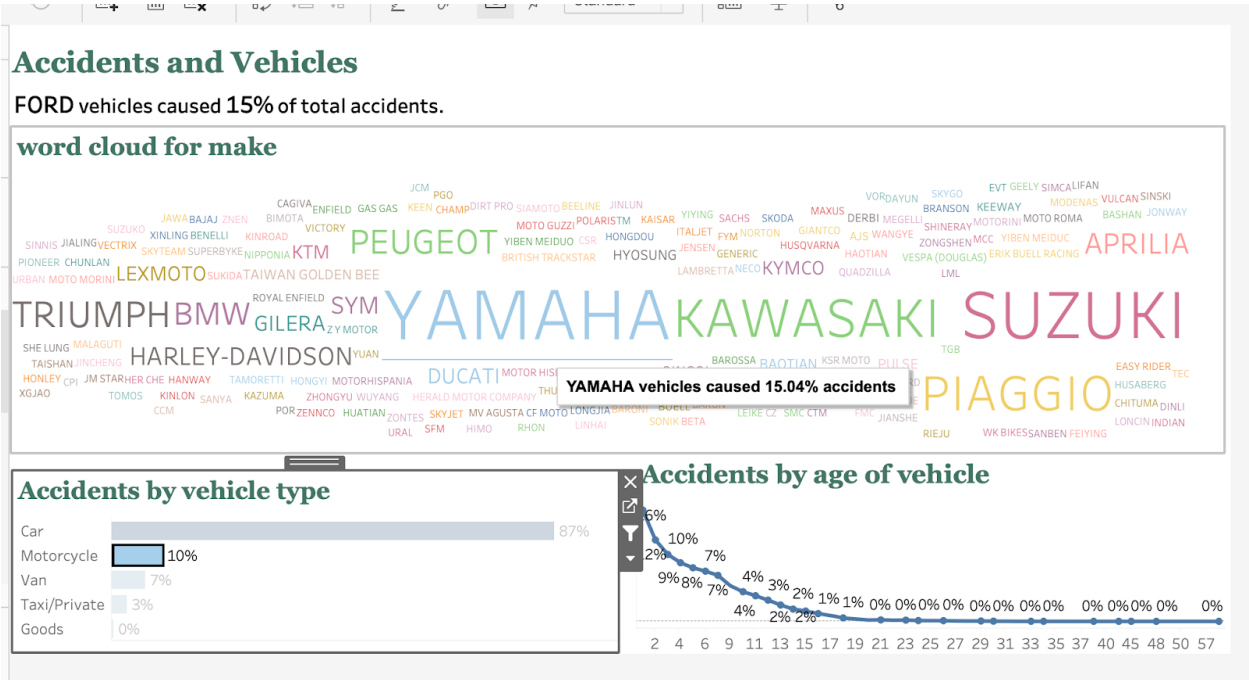
Accidents for Van Type of Vehicle



Note. In the van type, Vauxhall caused around 15% of total accidents.

Figure 31

Accidents for Motorcycle Type of Vehicle

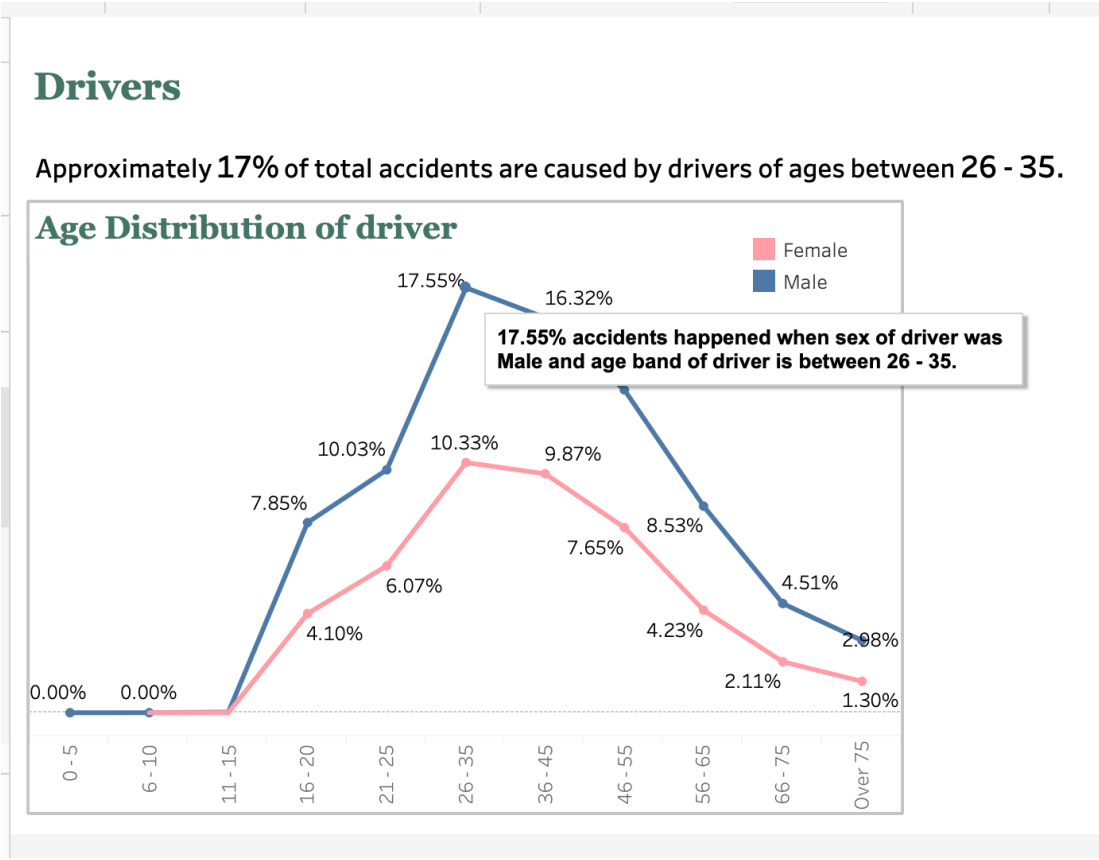


Note. In motorcycles, Yamaha caused 15% of accidents.

- Use case 5: Accidents by Age of the Driver

Figure 32

Dashboard for Accidents by Drivers Age

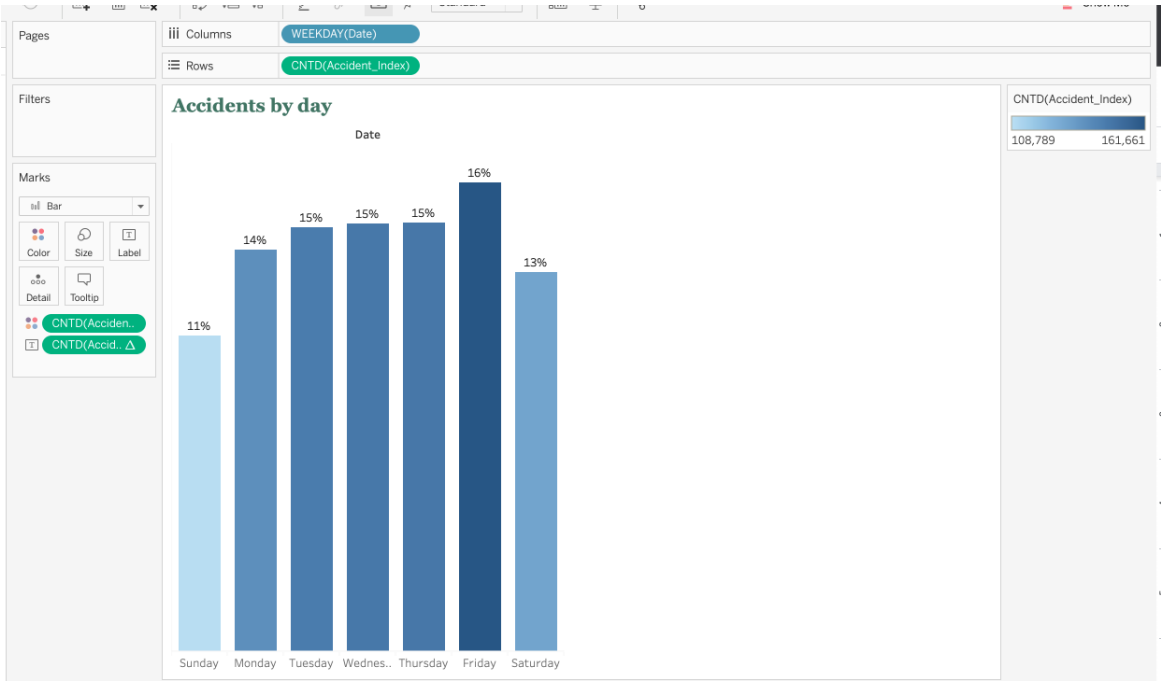


Note. The above dashboard is to show for accidents by different driver age for both male and female. Pink and Blue color are used to distinguish female and male drivers respectively. Both a female and male age group of 26-35 years are involved in most the accidents.

- Use case 6: Accidents by Day, year, and Time of the day

Figure 33

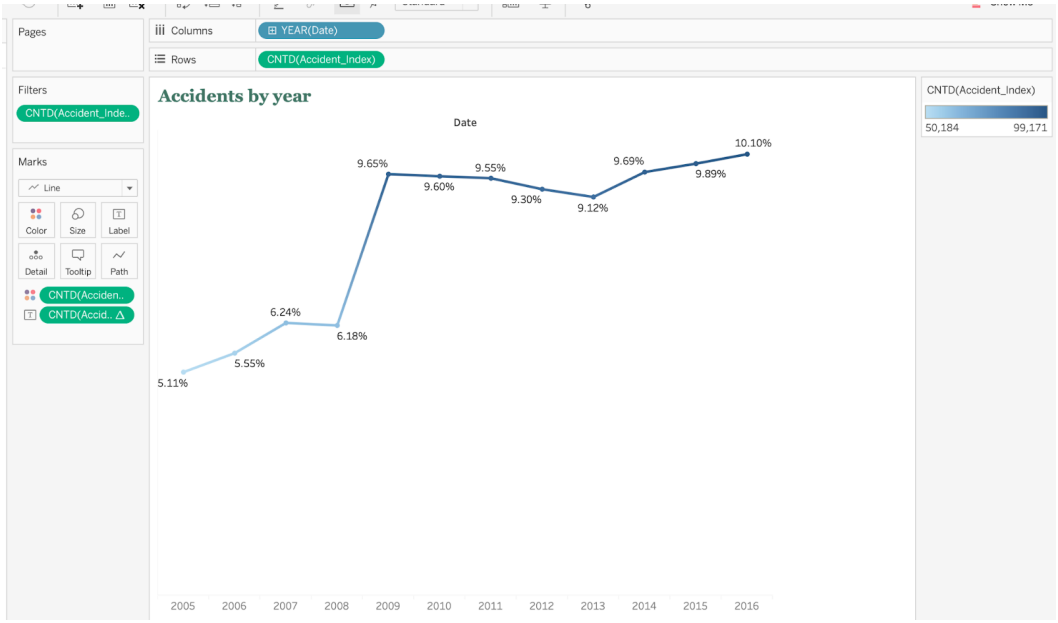
Worksheet for Accidents by Day



Note. The above worksheet shows the percentage of accidents for different days. As we can see Friday has the highest rate of accidents. In general, the weekend seems to have less number of accidents, and on weekdays the distribution of accidents is the same.

Figure 34

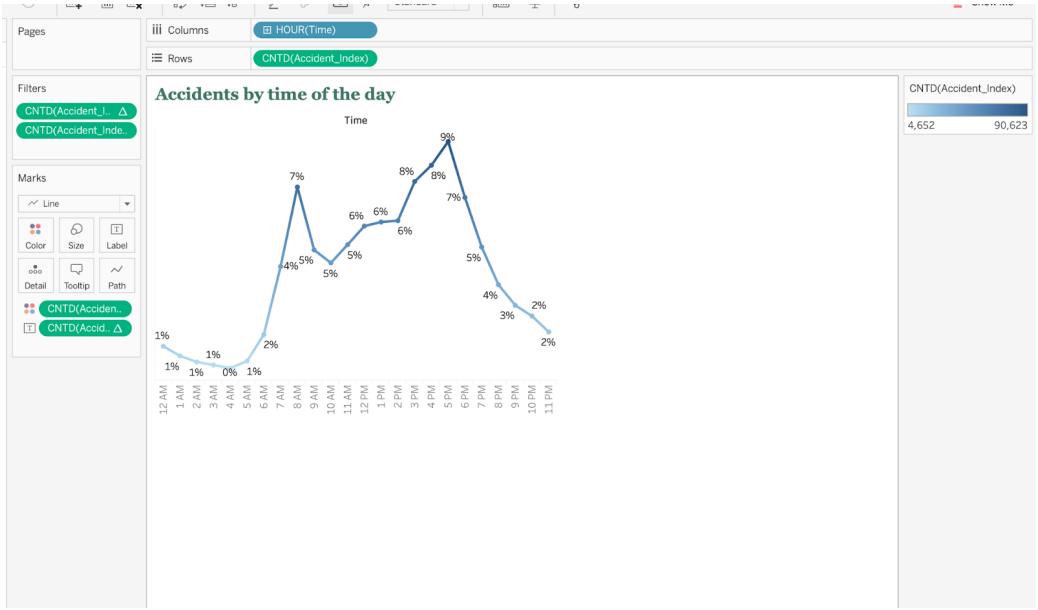
Worksheet for Accidents by Year



Note. The above worksheet shows the percentage of accidents for different years. As we can see, 2016 has the highest rate for accidents but from 2009-2016, it shows an approximately uniform distribution.

Figure 35

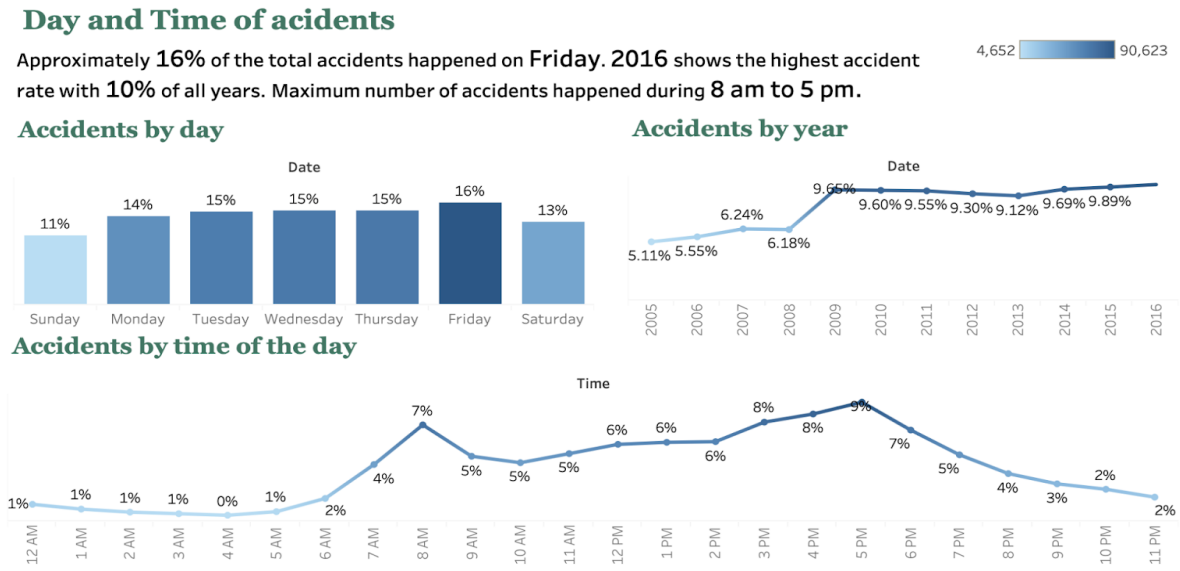
Worksheet for Accidents by Time of the Day



Note. The above worksheet shows the percentage of accidents by the time of the day. We can see from above, from 8 am to 5 pm is a dangerous timing for accidents. Because most of the people were outside doing their work.

Figure 36

Dashboard for Accidents by Day, Year, and Time of the Day



Note. The above dashboard shows the percentage of accidents for the day, year, and time of the day. We can say that On Friday it seems to have more number of accidents. 2016 shows the highest rate. 8 am to 5 pm is a dangerous time for accidents.

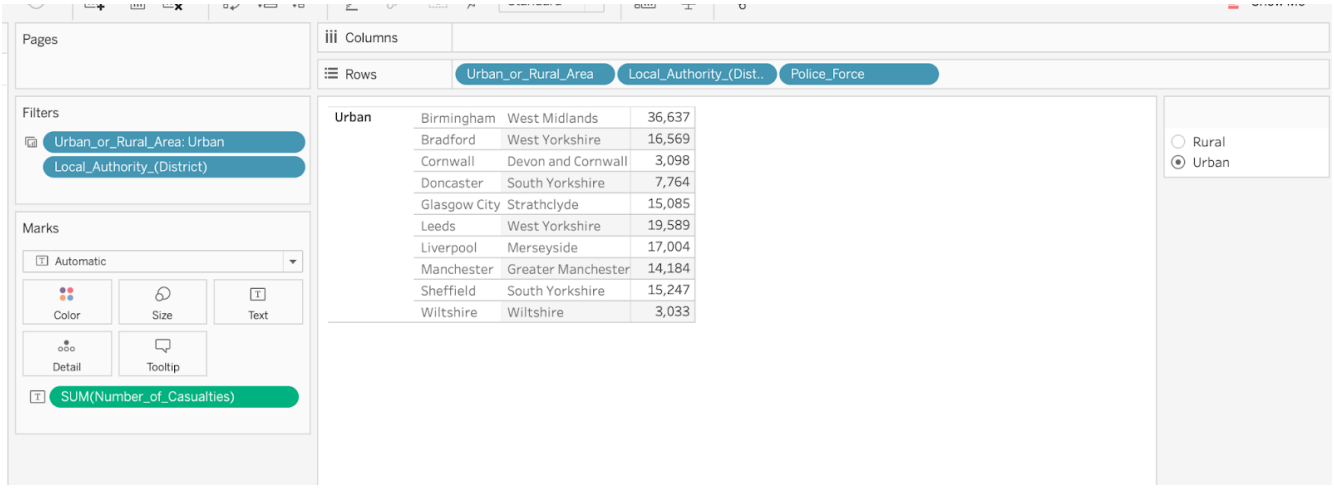
- Dashboards Using Interactive Extensions
- Sunburst Chart Extension

  The sunburst chart is also called the ring chart or radial treemap. It is useful to visualize the hierarchical dataset. Different concentric rings are used to show the level in the hierarchy. Each ring also represents its proportion.

- Use case 7: Number of casualties in urban or rural areas

Figure 37

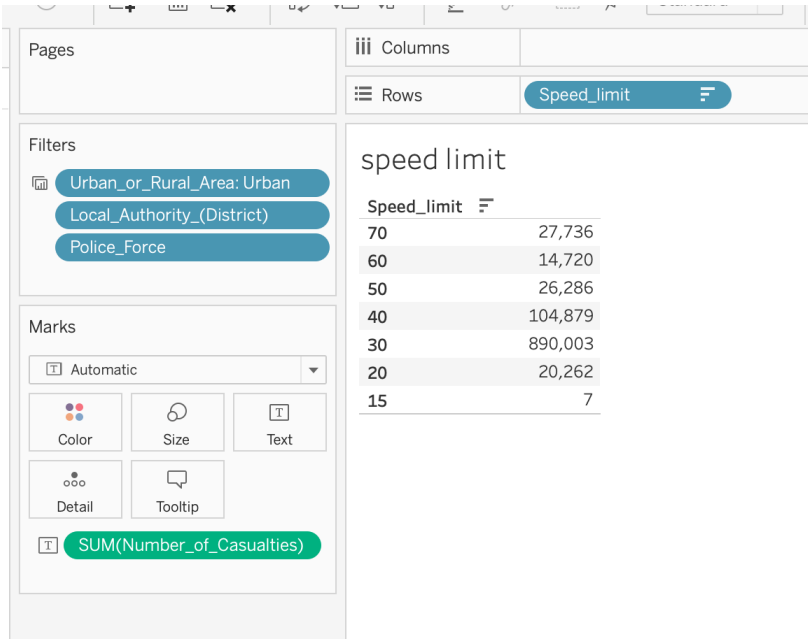
Worksheet for Number of Casualties in Urban and Rural Areas of District and Police Forces



Note. The above worksheet shows the number of casualties for different urban and rural areas of district and police forces in the UK.

Figure 38

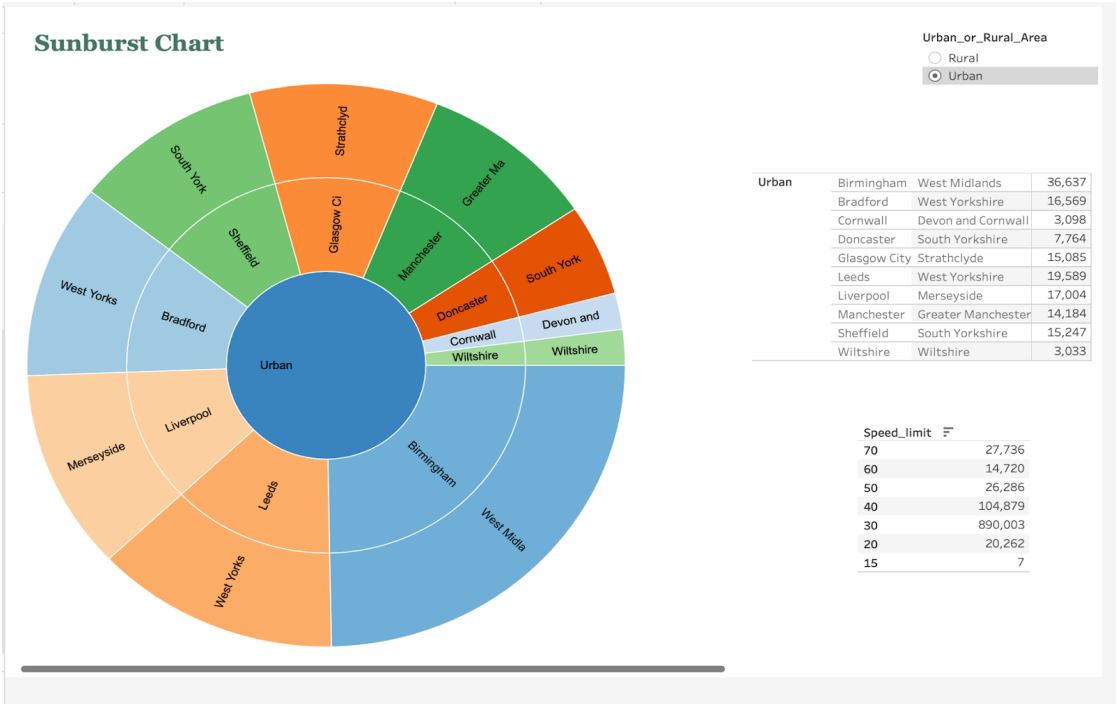
Worksheet for Number of Casualties for Speed Limit



Note. The above worksheet shows the number of casualties for different speed limits.

Figure 39

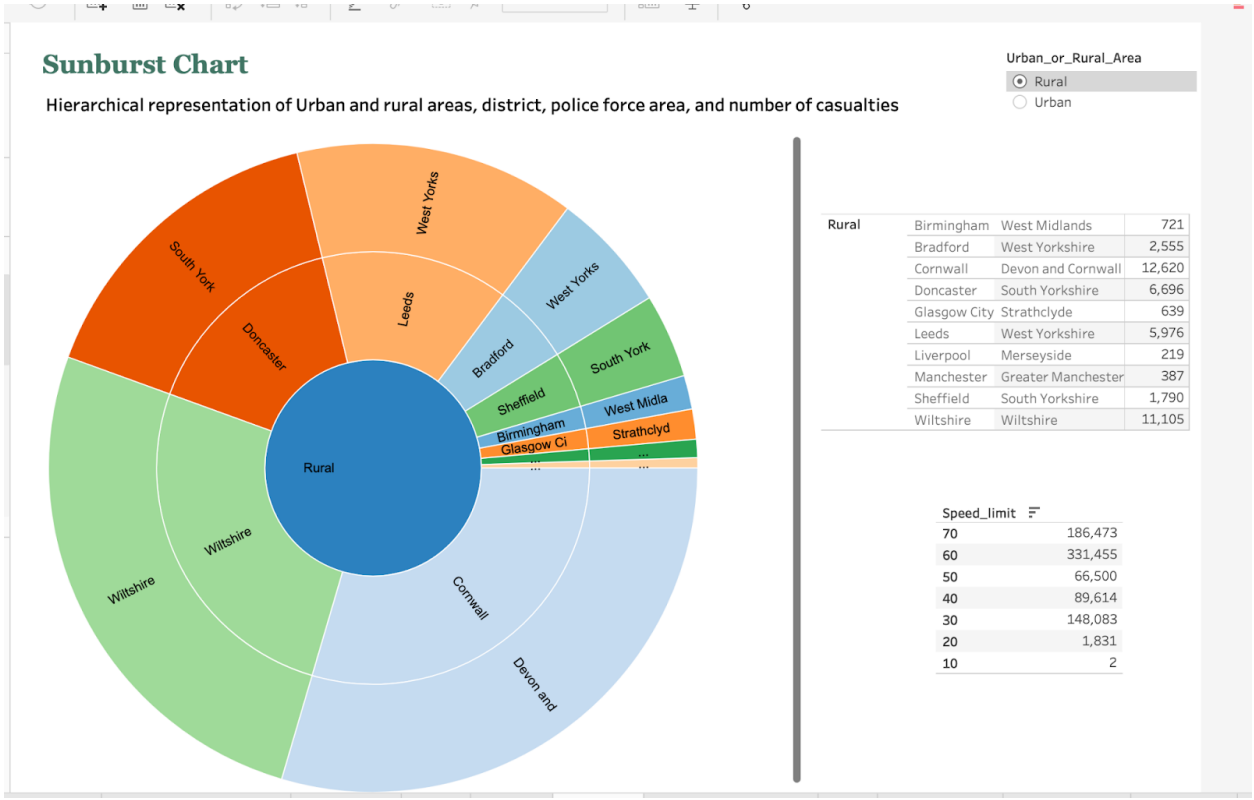
Sunburst Chart for Urban Areas



Note. Sunburst chart using tableau extension shows the number of casualties for urban or rural areas of districts and police force areas. Here added filter for urban or rural areas. If we select rural then the sunburst chart will update accordingly.

Figure 40

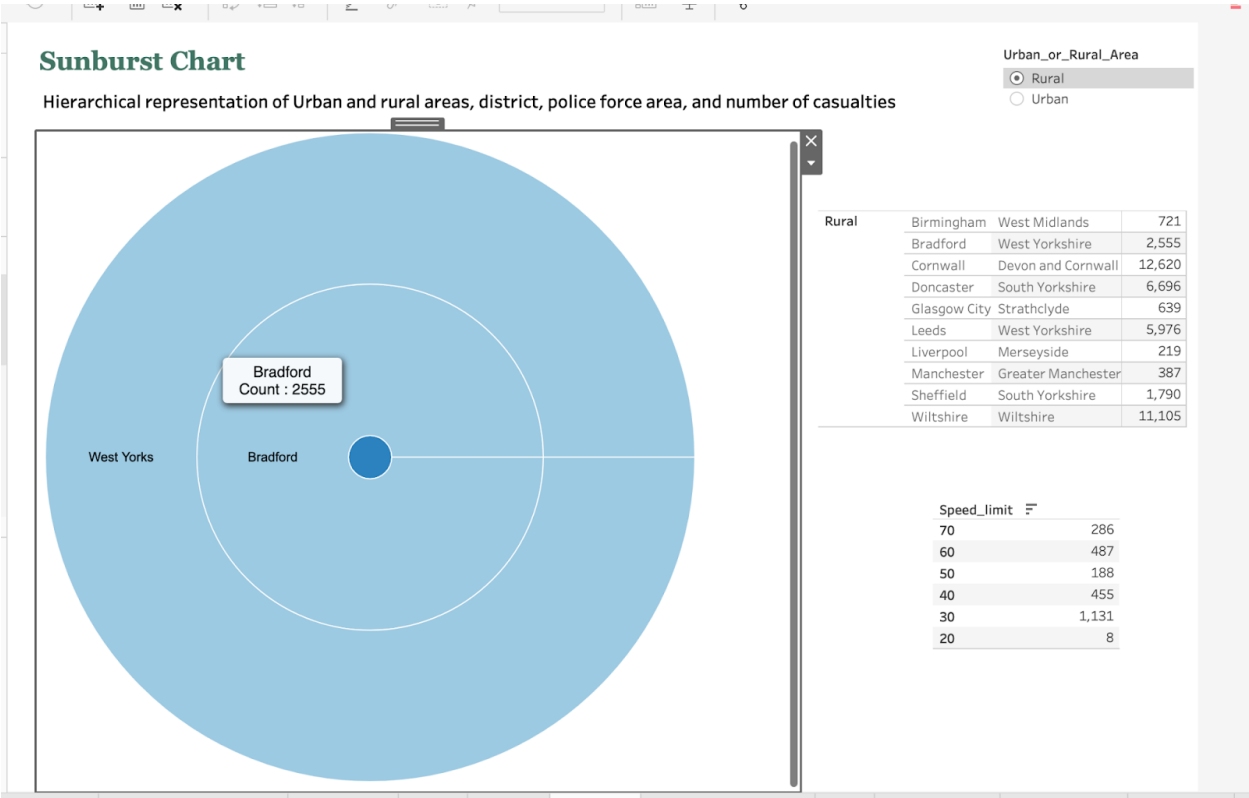
Sunburst Chart for Rural Areas



Note. Sunburst chart and table updated accordingly for rural areas.

Figure 41

Zooming in to See Bradford District



Note. Here another functionality of the sunburst chart is to zoom in and zoom out. For zoom in, we need to click on a specific district or police force and for zoom out, we can click on center.

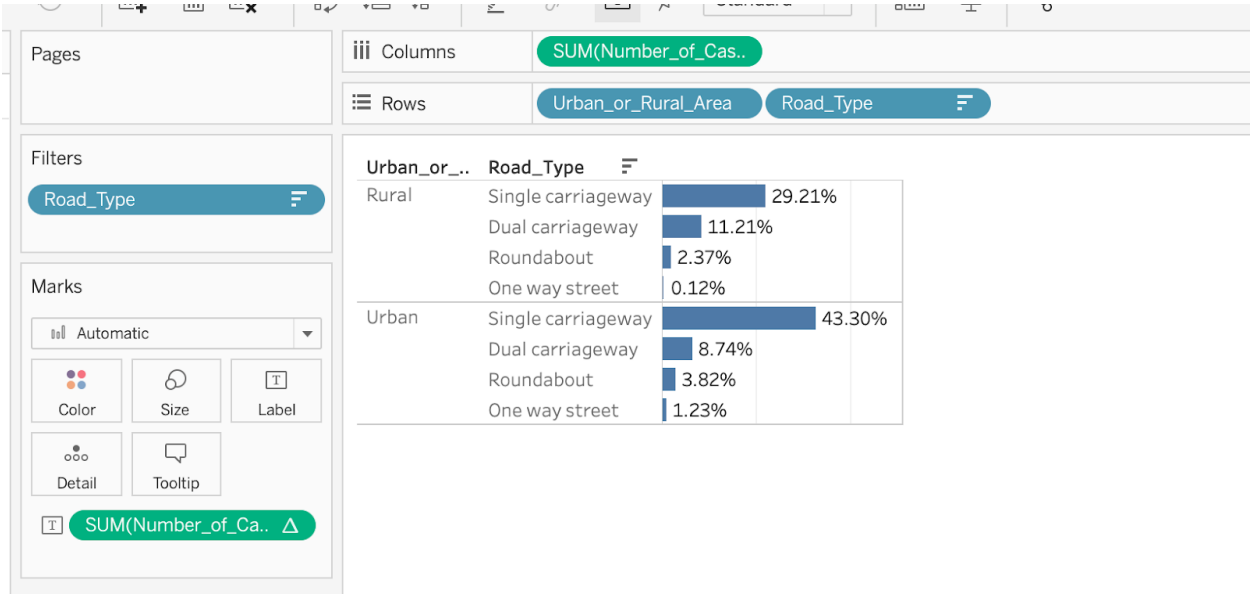
- Tree Diagram Extension

  This type of extension shows the hierarchical representation between different entities.

- Use case 8: Number of casualties in urban and rural areas for different road types

Figure 42

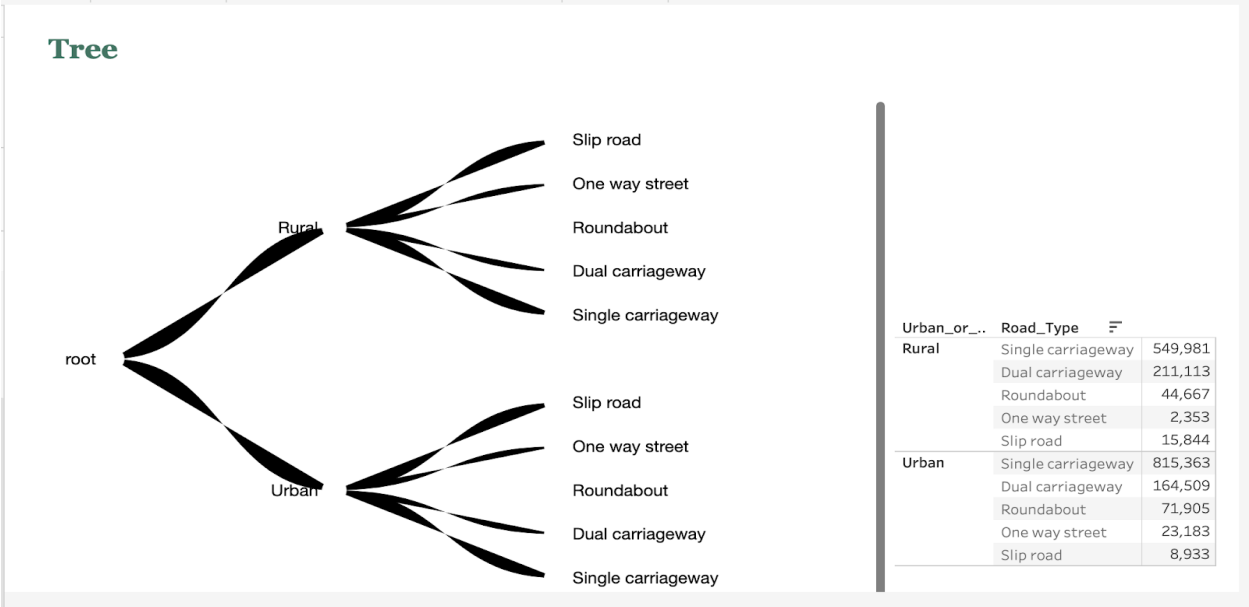
Worksheet of Casualties by Vehicle Type



Note. The above worksheet shows the number of casualties for different road types in urban and rural areas.

Figure 43

Dashboard of Tree Diagram Extension



Urban_or_..	Road_Type	
Rural	Single carriageway	549,981
	Dual carriageway	211,113
	Roundabout	44,667
	One way street	2,353
	Slip road	15,844
Urban	Single carriageway	815,363
	Dual carriageway	164,509
	Roundabout	71,905
	One way street	23,183
	Slip road	8,933

Note. Tree diagram using tableau extension showing number of casualties for road type in urban or rural areas.

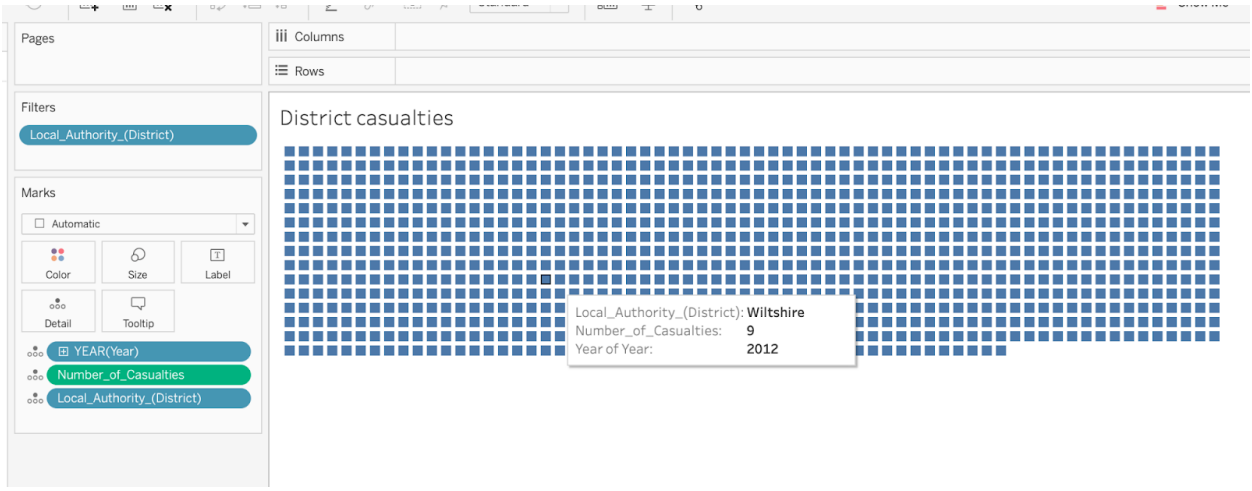
- Bar Race Chart Extension

  This interactive extension helps to visualize the change over time. Because of its animation, we can clearly understand the trend. Also, we can pause and see for a specific time period how is the trend. We can move forward and backward in time.

- Use case 9: Top 12 districts with the highest number of casualties through 2005-2016.

Figure 44

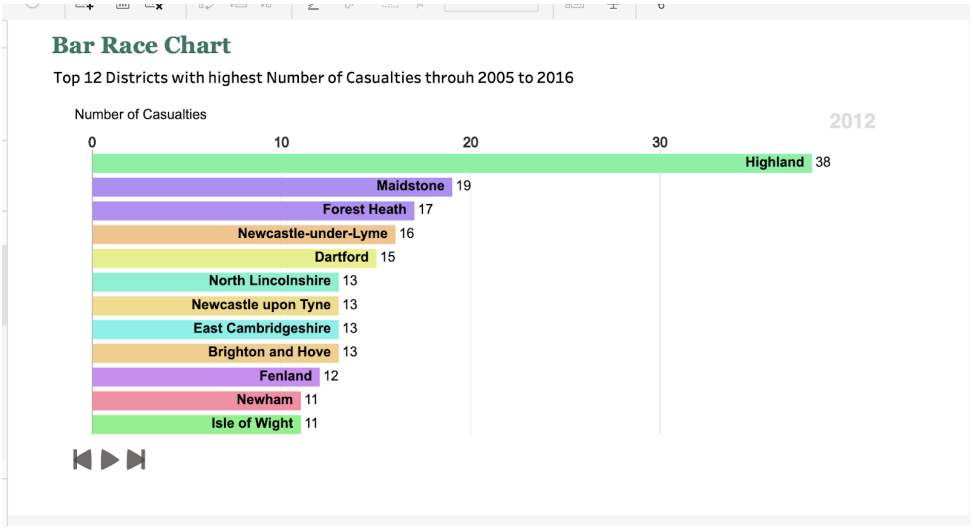
Worksheet for District Casualties



Note. The above worksheet is created for the bar race chart extension which shows the number of casualties in districts of the UK.

Figure 45

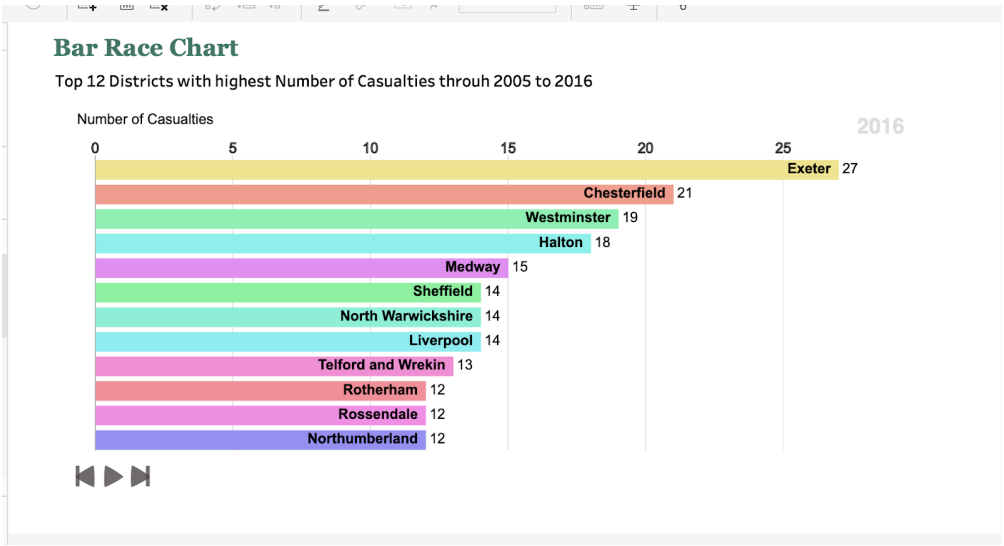
Bar Race Chart



Note. Bar race chart using tableau extension shows the number of casualties in the top 12 districts. There are buttons in the UI for next, previous, and pause. We can use this button to see results for different years. For 2012, the Highland district has the highest casualties.

Figure 46

Bar Race Chart for the 2016 Year



Note. Bar race chart for the 2016 year. We can see for 2016, the Exeter district has more casualties.

Conclusion

Based on the visualization few suggestions can be made for the prevention of accidents, as listed below.

- Birmingham being the highest prone area, more police force can be employed with strict rules. The speed limit should be enforced strictly.
- Single carriageways can be modified to the possible extent.
- Also, we saw those young people of age 26 to 35 years are involved in most of the accidents, so more awareness is needed to control and reduce the accidents.
- Fridays were found to be more accident-prone days. One solution for this case is if possible to promote public transport on Fridays to control the rate of accidents.
- The speed at the peak hours should also be controlled. More medical facilities can be provided till the conditions get better.
- More light facilities can be provided to prevent accidents in the dark.
- The design of the vehicles causing most accidents can be reviewed to prevent accidents.

Future Work

Using this analysis, we can build a predictive model to accurately determine the severity of accidents based on some factors. We can also try to build a recommendation system for the UK Department of Transport. This system will help to improve road safety and also will help to reduce accidents.

References

Thanasis, (2018). *UK Road Safety: Traffic Accidents and Vehicles*. Kaggle. https://www.kaggle.com/tsiaras/uk-road-safety-accidents-and-vehicles

Muhammad, B. A. R., Muhammad, A. M., Wesam, S. A., Ahsen, M., Hamna, B., & Waqas, R. (2021, July 19). Road Traffic Accident Data Analysis and Its Visualization. *Civil Engineering and Architecture.* 9(5), 1603-1614. DOI: 10.13189/cea.2021.090530. https://www.hrpub.org/download/20210830/CEA30-14823813.pdf

Cuiping Zhang, Xuedong Yan, Lu Ma, Meiwu An, "Crash Prediction and Risk Evaluation Based on Traffic Analysis Zones", *Mathematical Problems in Engineering*, vol. 2014, Article ID 987978, 9 pages, 2014. https://doi.org/10.1155/2014/987978 https://www.hindawi.com/journals/mpe/2014/987978/#conclusions-and-discussions

Mehdizadeh, A., Cai, M., Hu, Q., Alamdar Yazdi, M. A., Mohabbati-Kalejahi, N., Vinel, A., Rigdon, S. E., Davis, K. C., & Megahed, F. M. (2020). A Review of Data Analytic Applications in Road Traffic Safety. Part 1: Descriptive and Predictive Modeling. *Sensors (Basel, Switzerland)*, *20*(4), 1107. https://doi.org/10.3390/s20041107

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7070501/

https://medium.com/@rawanme/u-k-traffic-accidents-data-analysis-10-years-c81293180ee5

https://data.gov.uk/dataset/cb7ae6f0-4be6-4935-9277-47e5ce24a11f/road-safety-data

Dataset link: https://www.kaggle.com/tsiaras/uk-road-safety-accidents-and-vehicles

Data and tableau files can be downloaded from:

https://drive.google.com/drive/folders/1egscbYlJdEsoP95KgnnvjdFFfrYl9qn-?usp=sharing

Tableau Public:

https://public.tableau.com/app/profile/mohini.patil4996/viz/UKAccidentAnalysis/UKRoadSafety?publish=yes