VIT®

Vellore Institute of Technology

(Deemed to be University under section 3 of UGC Act, 1956)

SCHOOL OF COMPUTER ENGINEERING

PROJECT REPORT
CSE3020 – DATA VISUALIZATION

ROAD SAFETY ANALYSIS AND VISUALIZATION

GROUP 11

Team members

1.Melvin Paulsam P (20BCE0585)

2.Pranav Sehgal (20BCE0898)

3.D Harsha Vardhan (20BCE2695)

4.Mrinal Sharma (20BCI0247)

Submitted to:

Dr. Margret Anouncia S

Problem Description

1. The presentation of Road Accident Data to identify various factors that influence road accidents in the United Kingdom and the development of visualization techniques to reveal the data, which has been organized into various sections.
2. We explore various factors that influence traffic accidents on UK roads and provide various tools in order to visualize data collected from these cases.
3. Various data visualization techniques were used to analyze road accidents on the UK roads, as well as various other factors that could impact on the accidents.

SCOPE

1. To analyze the 2005-2014 road traffic accidents dataset in the UK, deploy data visualization techniques, and facilitate the formation of road safety policies.
2. The analysis and visualization techniques that are applied to the road traffic accidents dataset collected over those ten years in the UK by analyzing different data points.
3. Several computational, mathematical modeling and data visualization techniques are used for data analysis. This aims to find out how to reduce the accidents, these data analysis techniques are applied on the road accidents dataset to find out some valuable insights, thus give some valuable solutions/suggestions.

Objectives

- To analyze the road traffic accidents dataset from last 2005-2014 in the UK and deploy data visualization techniques to find pattern and insight. Several modelling techniques are used for data analysis of the same.
- To find out how to drive safer, this project could be applied on the traffic accident dataset to find out some valuable information, and hence give driving suggestion.

Datasets Used for Visualization:

Dataset 1 – Accidents0514.csv

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Accident_Index	Longitude	Longitude	Accident_Sever	Number_of	Number_c	Date	Day_of_W	Time	Speed_lim	Light_Con	Weather_	Road_Surf	Urban_or_Rural_Area
2	200501BS00001	-0.19117	0.80883	Major injuries	1	1	04-01-2005	Wednesda	17:42	30	Bright	Rainy	Wet	Urban
3	200501BS00002	-0.211708	0.788292	Minor injuries	1	1	05-01-2005	Thursday	17:36	30	Dark	Clear	Dry	Urban
4	200501BS00003	-0.206458	0.793542	Minor injuries	2	1	06-01-2005	Friday	00:15	30	Dark	Clear	Dry	Urban
5	200501BS00004	-0.173862	0.826138	Minor injuries	1	1	07-01-2005	Saturday	10:35	30	Bright	Clear	Dry	Urban
6	200501BS00005	-0.156618	0.843382	Minor injuries	1	1	10-01-2005	Tuesday	21:13	30	Dim	Clear	Wet	Urban
7	200501BS00006	-0.203238	0.796762	Minor injuries	2	1	11-01-2005	Wednesda	12:40	30	Bright	Rainy	Wet	Urban
8	200501BS00007	-0.211277	0.788723	Minor injuries	2	1	13-01-2005	Friday	20:40	30	Dark	Clear	Dry	Urban
9	200501BS00009	-0.187623	0.812377	Minor injuries	1	2	14-01-2005	Saturday	17:35	30	Bright	Clear	Dry	Urban
10	200501BS00010	-0.167342	0.832658	Minor injuries	2	2	15-01-2005	Sunday	22:43	30	Dark	Clear	Dry	Urban
11	200501BS00011	-0.206531	0.793469	Minor injuries	2	5	15-01-2005	Sunday	16:00	30	Bright	Clear	Dry	Urban
12	200501BS00012	-0.182872	0.817128	Minor injuries	1	1	16-01-2005	Monday	00:42	30	Dark	Clear	Dry	Urban
13	200501BS00014	-0.184312	0.815688	Minor injuries	2	1	25-01-2005	Wednesda	20:48	30	Dark	Clear	Wet	Urban
14	200501BS00015	-0.192366	0.807634	Minor injuries	1	1	11-01-2005	Wednesda	12:55	30	Bright	Rainy	Wet	Urban
15	200501BS00016	-0.157753	0.842247	Minor injuries	2	1	18-01-2005	Wednesda	05:01	30	Dark	Rainy	Wet	Urban
16	200501BS00017	-0.176224	0.823776	Minor injuries	1	2	18-01-2005	Wednesda	11:15	30	Bright	Clear	Dry	Urban
17	200501BS00018	-0.18022	0.81978	Minor injuries	1	1	18-01-2005	Wednesda	10:50	30	Bright	Clear	Dry	Urban
18	200501BS00019	-0.205139	0.794861	Major injuries	2	1	20-01-2005	Friday	00:15	30	Dark	Clear	Dry	Urban
19	200501BS00020	-0.171887	0.828113	Minor injuries	2	1	21-01-2005	Saturday	09:15	30	Bright	Clear	Dry	Urban
20	200501BS00021	-0.16059	0.83941	Minor injuries	2	1	21-01-2005	Saturday	21:16	30	Dark	Clear	Dry	Urban
21	200501BS00022	-0.174925	0.825075	Major injuries	1	1	08-01-2005	Sunday	03:00	30	Dark	Clear	Dry	Urban
22	200501BS00023	-0.18727	0.81273	Minor injuries	1	1	24-01-2005	Tuesday	21:45	30	Dark	Clear	Dry	Urban
23	200501BS00024	-0.204346	0.795654	Minor injuries	2	1	24-01-2005	Tuesday	17:05	30	Dark	Clear	Dry	Urban
24	200501BS00025	-0.173445	0.826555	Minor injuries	2	1	24-01-2005	Tuesday	21:30	30	Dark	Clear	Snow	Urban
25	200501BS00028	-0.167824	0.832176	Minor injuries	2	1	18-01-2005	Wednesda	17:25	30	Dark	Clear	Dry	Urban
26	200501BS00029	-0.166322	0.833678	Minor injuries	2	1	29-01-2005	Sunday	07:34	30	Bright	Clear	Wet	Urban
27	200501BS00031	-0.215629	0.784371	Minor injuries	1	1	19-01-2005	Thursday	16:35	30	Dark	Rainy	Wet	Urban
28	200501BS00032	-0.207623	0.792377	Minor injuries	2	1	30-01-2005	Monday	20:00	30	Dark	Clear	Dry	Urban
29	200501BS00033	-0.192771	0.807229	Minor injuries	2	2	29-01-2005	Sunday	13:15	30	Bright	Rainy	Wet	Urban
30	200501BS70001	-0.182872	0.817128	Minor injuries	2	1	01-02-2005	Wednesda	18:20	30	Dark	Rainy	Wet	Urban
31	200501BS70002	-0.160606	0.839394	Minor injuries	2	1	02-02-2005	Thursday	07:25	30	Dark	Clear	Dry	Urban
32	200501BS70003	-0.215545	0.784455	Major injuries	2	2	01-02-2005	Wednesda	17:30	30	Dark	Rainy	Wet	Urban

Accidents0514

Dataset 2 - Casualties0514.csv

	A	B	C	D	E	F	G	H	I
1	Accident_I	Vehicle_Re	Casualty_F	Sex_of_Ca	Age_of_Ca	Age_Band_	Casualty_S	Bus_or_Cc	Car_Passenger
2	200501BS(	1	1	Men	37	7	Major Inju	0	0
3	200501BS(	1	1	Men	37	7	Minor Inju	4	0
4	200501BS(	2	1	Men	62	9	Minor Inju	0	0
5	200501BS(	1	1	Men	30	6	Minor Inju	0	0
6	200501BS(	1	1	Men	49	8	Minor Inju	0	0
7	200501BS(	2	1	Women	30	6	Minor Inju	0	0
8	200501BS(	1	1	Men	31	6	Minor Inju	0	0
9	200501BS(	1	1	Women	13	3	Minor Inju	0	0
10	200501BS(	1	2	Women	13	3	Minor Inju	0	0
11	200501BS(	1	1	Men	35	6	Minor Inju	0	0
12	200501BS(	2	2	Women	48	8	Minor Inju	0	0
13	200501BS(	1	1	Women	26	6	Minor Inju	4	0
14	200501BS(	1	2	Women	9	2	Minor Inju	4	0
15	200501BS(	1	3	Women	40	7	Minor Inju	4	0
16	200501BS(	1	4	Women	38	7	Minor Inju	4	0
17	200501BS(	1	5	Women	28	6	Minor Inju	4	0
18	200501BS(	1	1	Women	23	5	Minor Inju	0	0
19	200501BS(	2	1	Women	20	4	Minor Inju	0	0
20	200501BS(	1	1	Women	75	10	Minor Inju	0	0
21	200501BS(	1	1	Men	34	6	Minor Inju	0	0
22	200501BS(	1	1	Men	42	7	Minor Inju	0	0
23	200501BS(	1	2	Women	31	6	Minor Inju	0	0
24	200501BS(	1	1	Women	56	9	Minor Inju	0	0
25	200501BS(	1	1	Men	54	8	Major Inju	0	0
26	200501BS(	2	1	Men	26	6	Minor Inju	0	0
27	200501BS(	1	1	Men	33	6	Minor Inju	0	0
28	200501BS(	1	1	Men	10	1	Major Inju	0	0
29	200501BS(	1	1	Men	49	8	Minor Inju	0	0
30	200501BS(	2	1	Women	25	5	Minor Inju	0	0
31	200501BS(	2	1	Men	59	9	Minor Inju	0	0
32	200501BS(	1	1	Men	40	7	Minor Inju	0	0

Casualties0514

Dataset 3 - Vehicles0514.csv

	A	B	C	D	E	F	G	H
1	Accident_Index	Vehicle_Reference	Vehicle_Type	Sex_of_Driver	Age_of_Driver	Age_Band_of_Driver	Age_of_Vehicle	
2	200501BS00001	1	9	Women	74	10	1	
3	200501BS00002	1	11	Men	42	7	3	
4	200501BS00003	1	11	Men	35	6	5	
5	200501BS00003	2	9	Men	62	9	6	
6	200501BS00004	1	9	Women	49	8	4	
7	200501BS00005	1	3	Men	49	8	10	
8	200501BS00006	1	9	Men	51	8	1	
9	200501BS00006	2	3	Women	30	6	2	
10	200501BS00007	1	3	Men	31	6	1	
11	200501BS00007	2	9	Men	41	7	4	
12	200501BS00009	1	9	Men	68	10	16	
13	200501BS00010	1	9	Men	35	6	13	
14	200501BS00010	2	9	Women	48	8	1	
15	200501BS00011	1	11	Men	42	7	2	
16	200501BS00011	2	90	Others	18	-1	1	
17	200501BS00012	1	9	Men	34	6	6	
18	200501BS00014	1	9	Women	19	4	8	
19	200501BS00014	2	3	Women	20	4	1	
20	200501BS00015	1	9	Women	47	8	2	
21	200501BS00016	1	9	Men	34	6	2	
22	200501BS00016	2	9	Men	33	6	4	
23	200501BS00017	1	5	Men	42	7	2	
24	200501BS00018	1	11	Men	57	9	1	
25	200501BS00019	1	5	Men	54	8	6	
26	200501BS00019	2	9	Men	44	7	1	
27	200501BS00020	1	9	Men	27	6	7	
28	200501BS00020	2	3	Men	26	6	1	
29	200501BS00021	1	5	Men	33	6	1	
30	200501BS00021	2	9	Men	31	6	5	
31	200501BS00022	1	9	Men	28	6	1	

Vehicles0514

DESIGN & METHODOLOGY

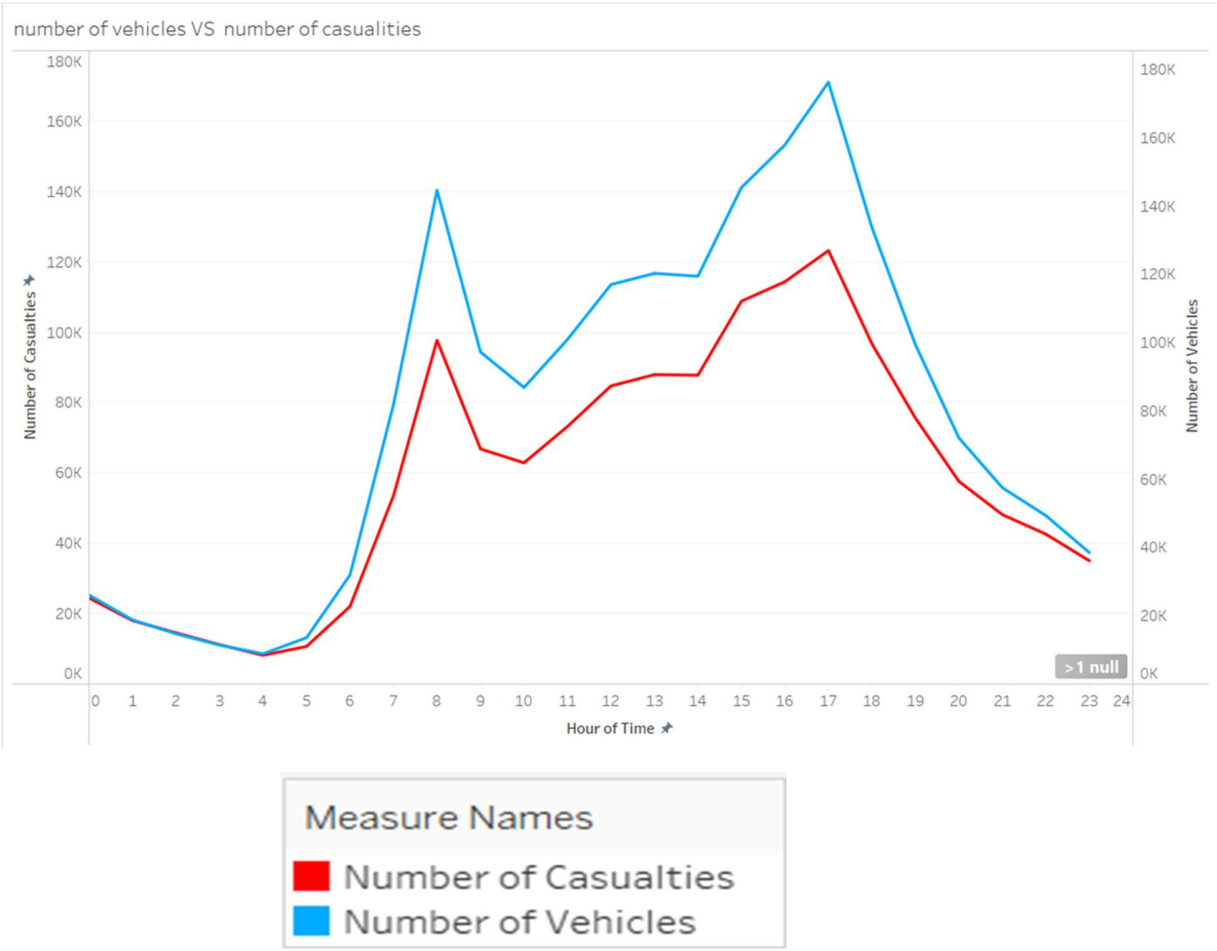
DESIGN JUSTIFICATION

- Line graphs have been used for correlation between 2 specific attributes, as well as for temporal analysis of data.
- Dot map have been used for representation of geospatial data over a fixed area.
- Treemap has been utilised for representation of hierarchical data in the standard dataset.
- For a clear analysis of the attributes as well as representation of the data, bar graphs have been used as well.

DATABASE CONNECTIVITY

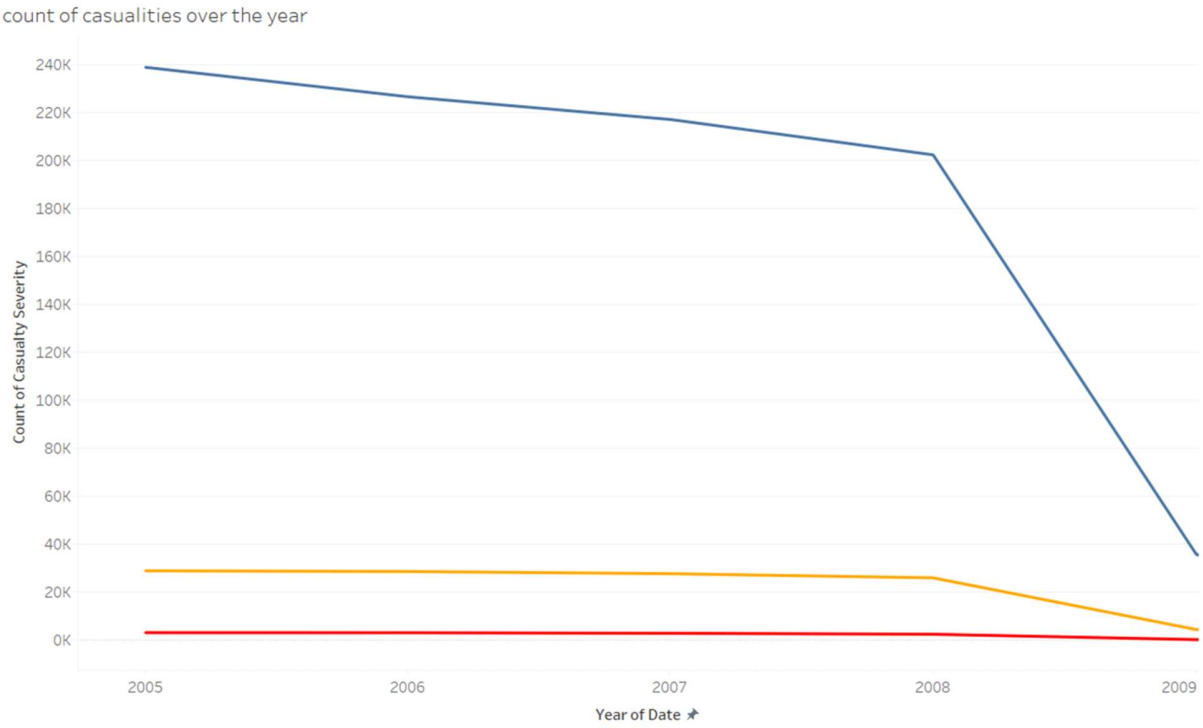
Visual Encoding

1.Line Graph



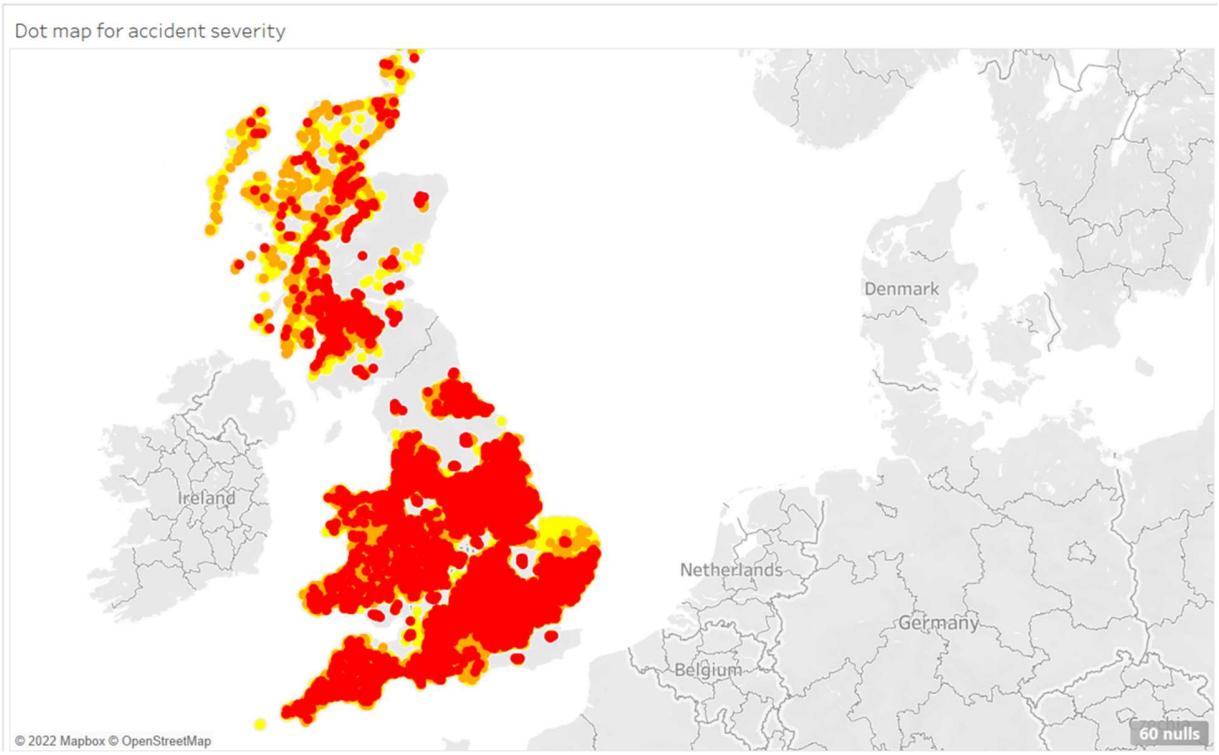
The graph shown is helpful to visualize the correlation of total number of casualties vs the total number of vehicles involved at different time of day.

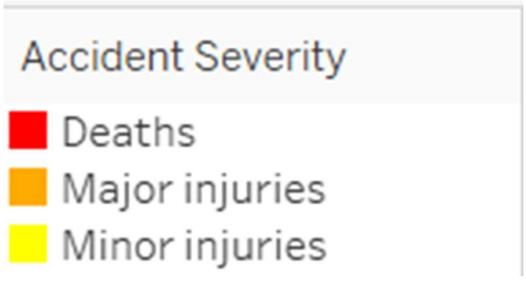
2.Temporal Analysis



This line chart is to visualize the temporal data that is the total casualties took place over the years 2005-2009.

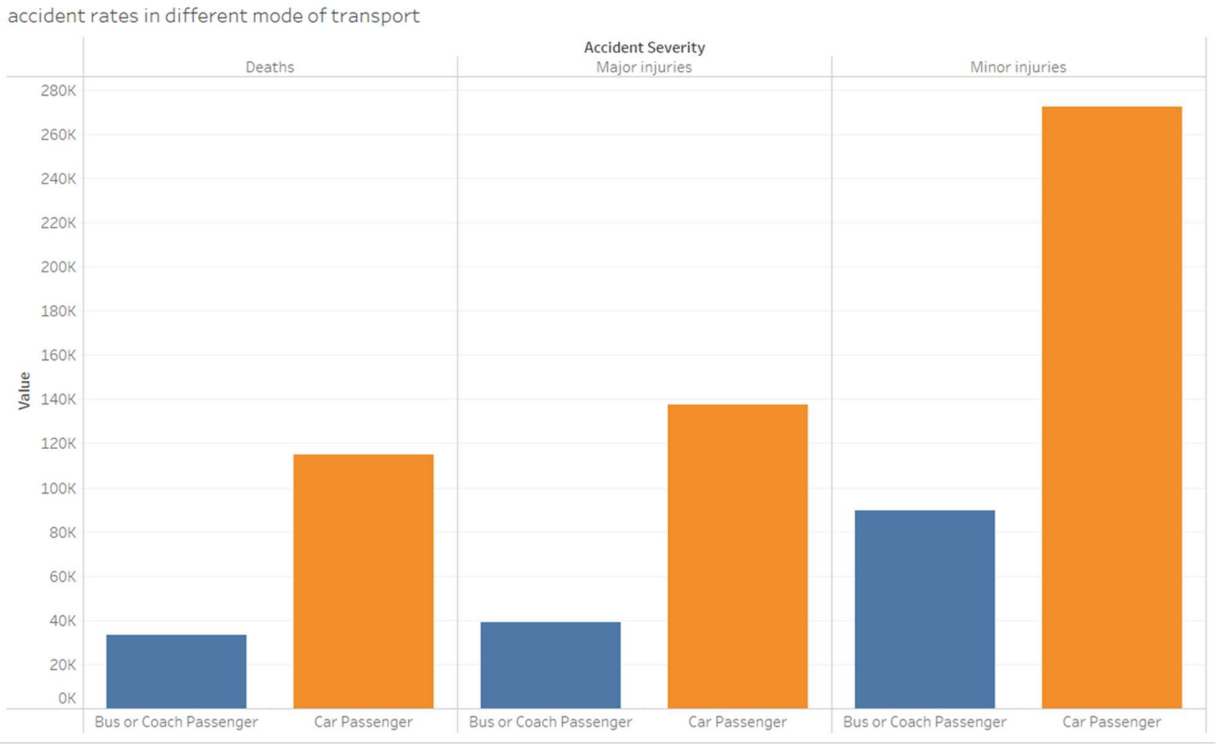
3.Dot Map

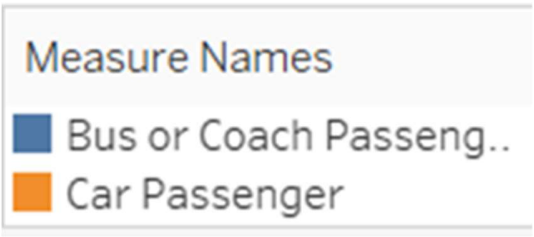




The dot map shown here is to represent the geospatial data of type of accident severity over the map of United Kingdom.

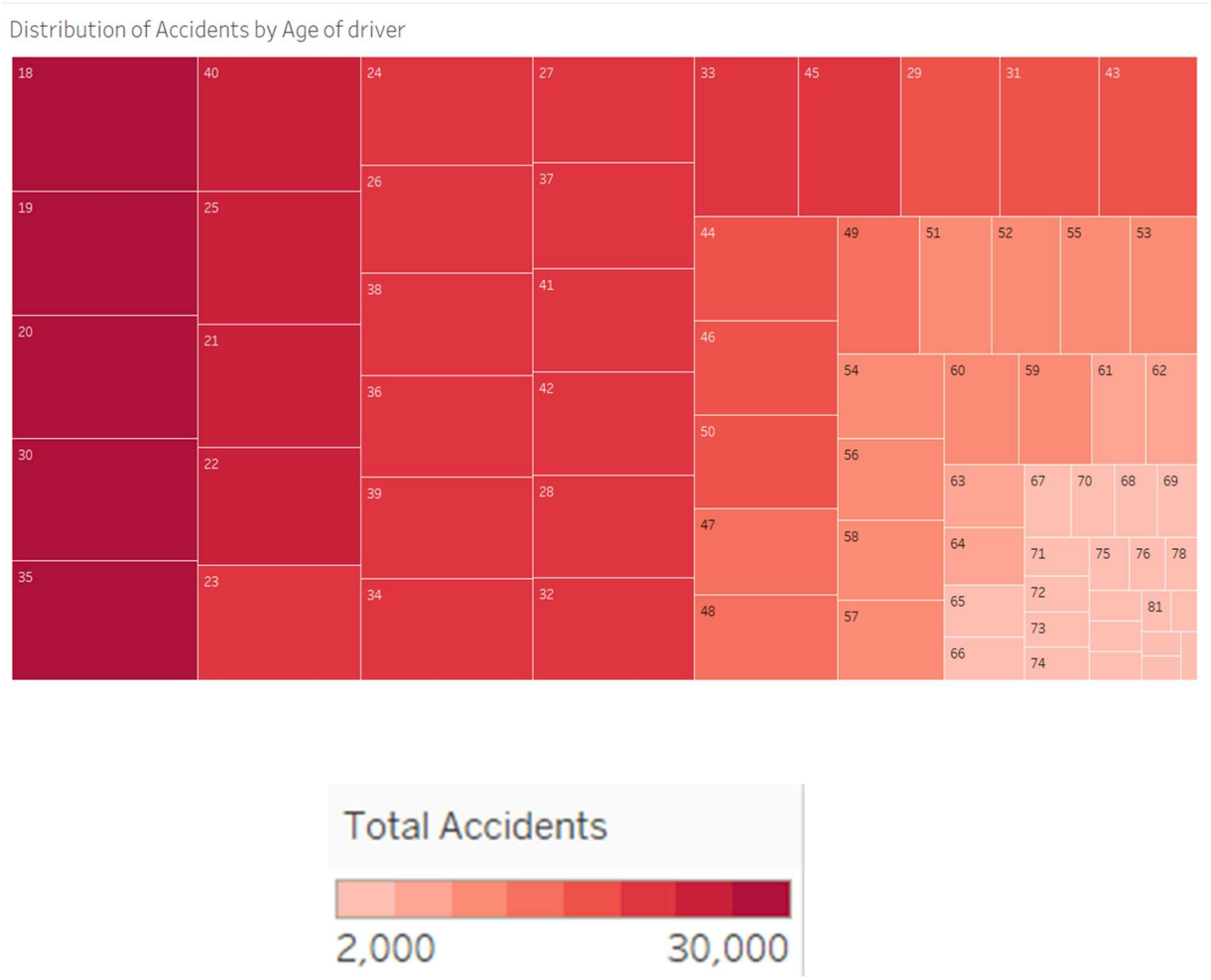
4.Bar Chart





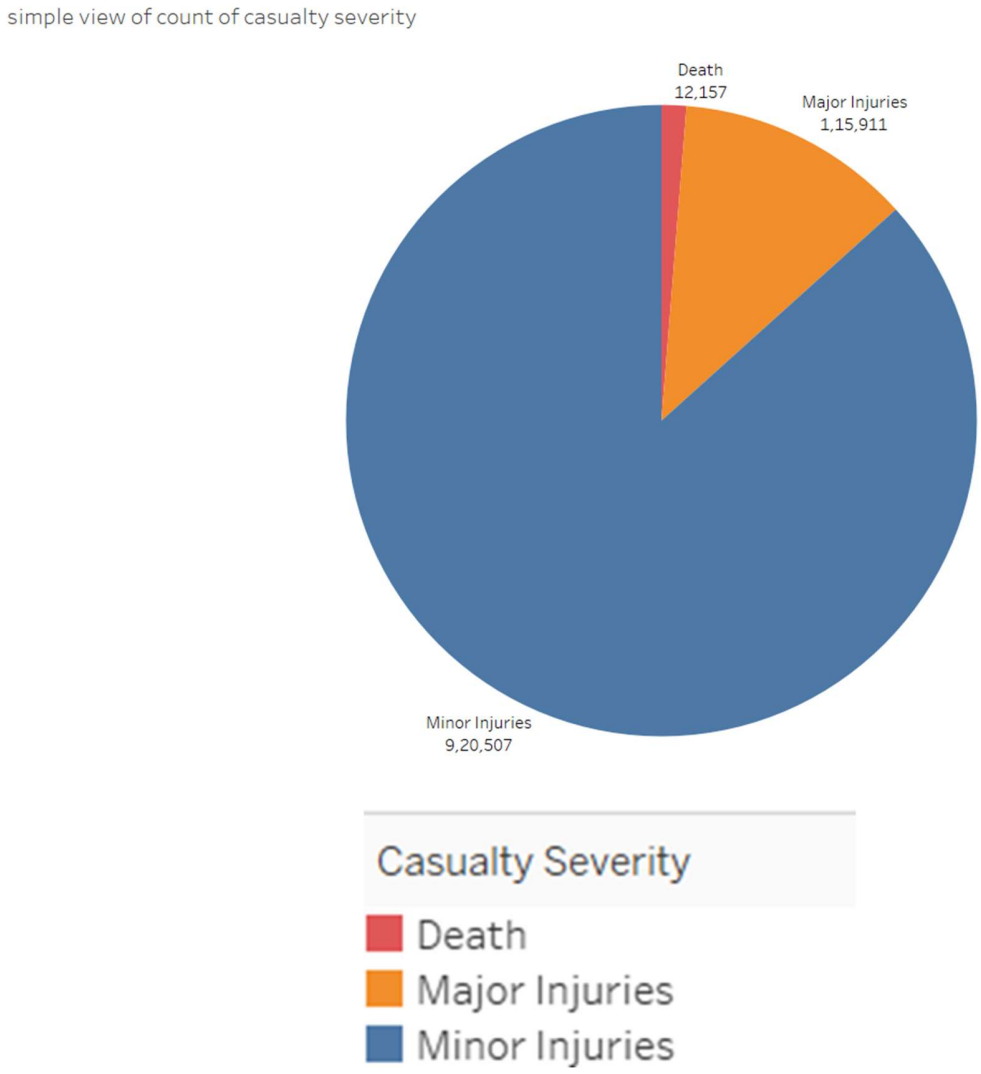
This bar chart represents the casualty occurred in accidents to car passengers and bus/coach passenger in accidents.

5.Tree Map



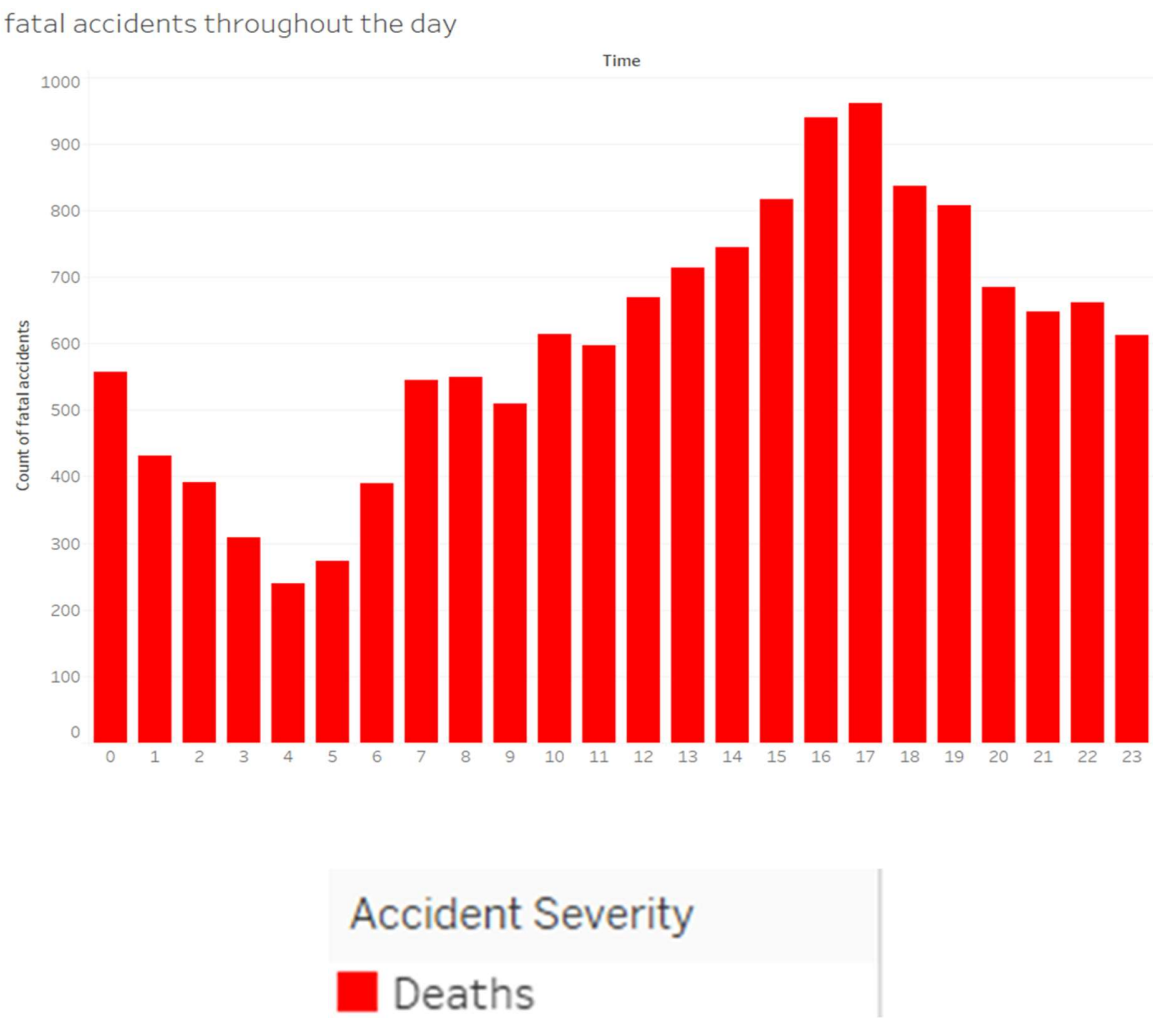
The Tree map shown here helps us to understand the distribution of accidents over the age of driver.

6.Pie Chart



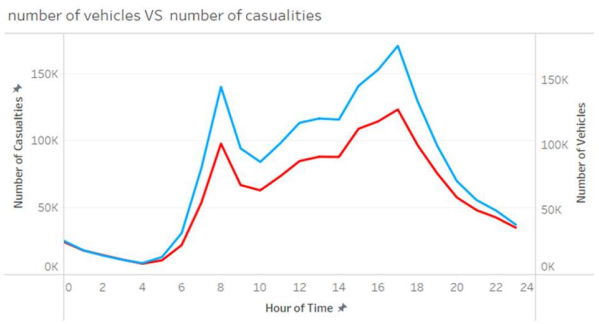
The pie chart shown represents the distribution of casualties based on the severity that is, Death, Major Injuries and Minor Injuries.

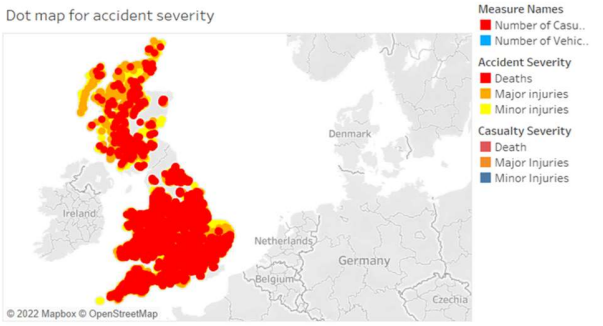
7.Bar Chart

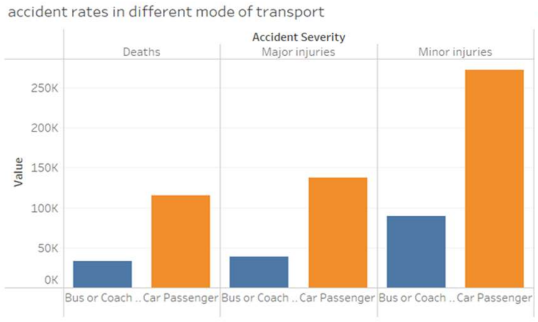


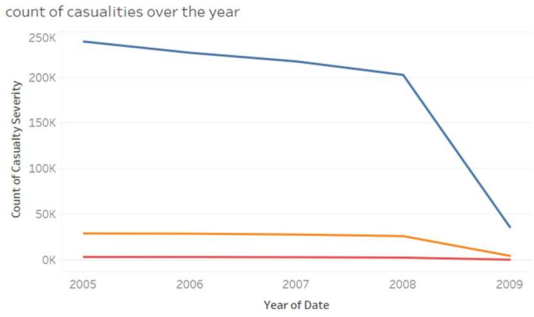
This bar chart is used to visualize the distribution of fatal accidents over the 24 hours in the day.

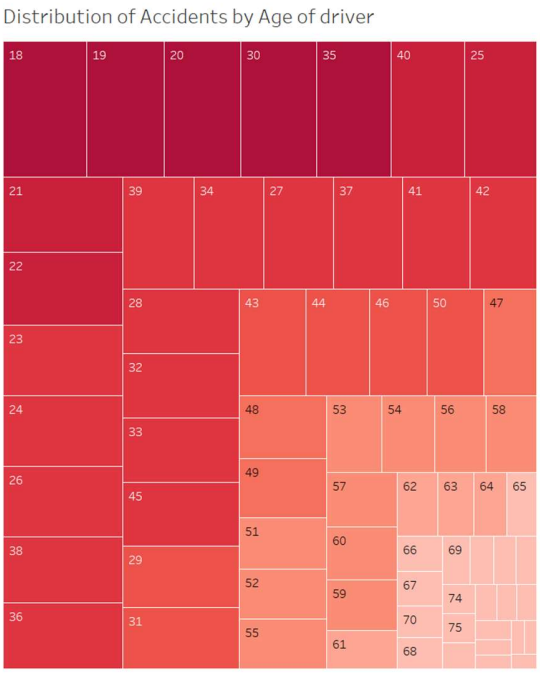
USER INTERFACE

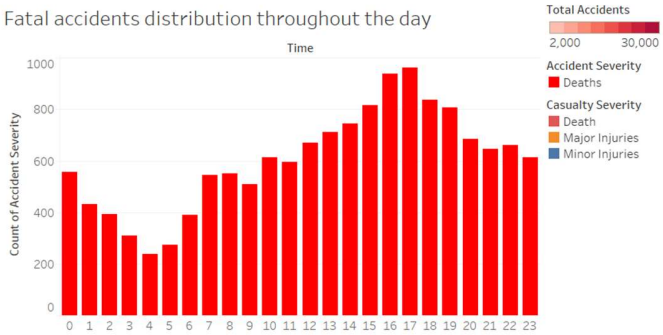


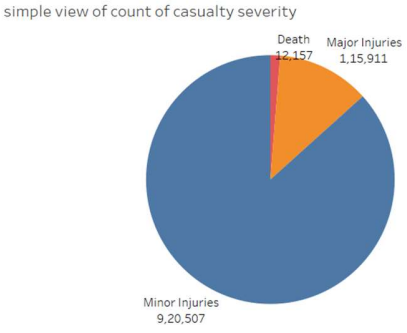








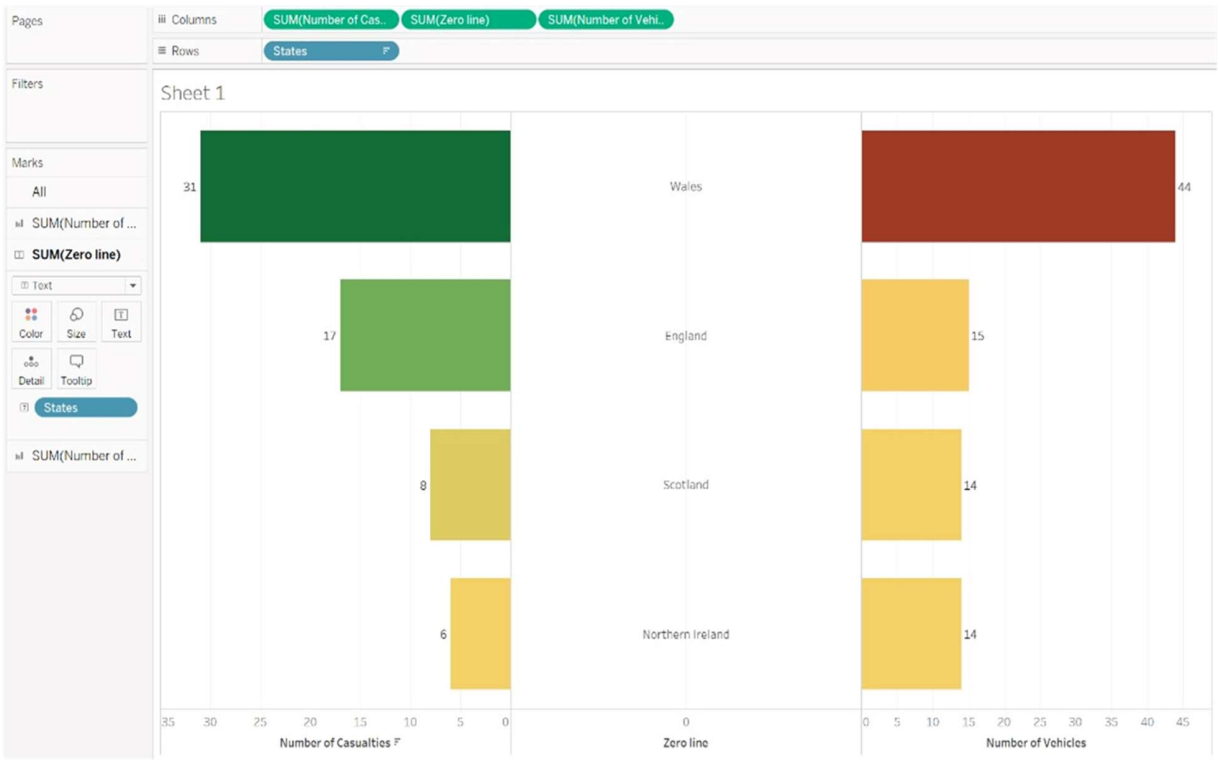


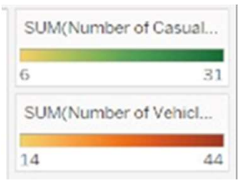


CUSTOMIZED DATASET

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Accident_Index		States	Accident_	Number_c	Number_c	Date	Day_of_W	Time	Vehicle ty	Speed_lim	Pedestriar	Weather_conditions	
2	200501BS00001		England	Minor	1	1	4/1/2022	Sunday	17:40	LMV	30	1	Snow	
3	200501BS00002		England	Major	1	1	5/1/2022	Wednesda	17:32	HMV	30	5	Rainy	
4	200501BS00003		Wales	Major	2	1	6/1/2022	Thursday	0:15	LMV	50	0	Rainy	
5	200501BS00004		England	Major	1	1	7/1/2022	Friday	10:35	LMV	50	0	Snow	
6	200501BS00005		England	Major	1	1	18/5/2022	Monday	21:13	LMV	30	0	Clear	
7	200501BS00006		Northern I	Major	2	1	3/4/2022	Sunday	12:40	LMV	30	0	Snow	
8	200501BS00007		Northern I	Major	2	1	13/01/202	Thursday	20:40	LMV	50	0	Rainy	
9	200501BS00009		England	Major	1	2	14/01/202	Friday	17:35	HMV	30	0	Snow	
10	200501BS00010		Wales	Major	2	2	15/01/202	Tuesday	22:43	LMV	30	5	Rainy	
11	200501BS00011		Wales	Major	2	5	15/01/202	Tuesday	16:00	LMV	25	8	Snow	
12	200501BS00012		England	Major	1	1	16/01/202	Saturday	0:42	LMV	30	5	Rainy	
13	200501BS00014		Wales	Death	2	1	25/01/202	Sunday	20:48	LMV	30	5	Rainy	
14	200501BS00015		England	Death	1	1	3/2/2020	Thursday	12:55	Two whee	25	1	Snow	
15	200501BS00016		Wales	Death	2	1	18/01/202	Sunday	5:01	Two whee	25	0	Rainy	
16	200501BS00017		England	Death	1	2	18/01/202	Sunday	11:15	HMV	25	0	Snow	
17	200501BS00018		England	Major	1	1	18/01/202	Sunday	10:50	LMV	25	1	Snow	
18	200501BS00019		Scotland	Minor	2	1	20/01/202	Thursday	0:15	LMV	25	0	Rainy	
19	200501BS00020		Wales	Major	2	1	21/01/202	Friday	9:15	LMV	30	0	Snow	
20	200501BS00021		Wales	Major	2	1	21/01/202	Friday	21:16	LMV	30	0	Rainy	
21	200501BS00022		England	Minor	1	1	8/1/2020	Tuesday	3:00	LMV	30	0	Rainy	
22	200501BS00023		England	Major	1	1	24/01/20(	Monday	21:45	LMV	30	0	Rainy	
23	200501BS00024		Wales	Major	2	1	24/01/202	Monday	17:05	LMV	30	0	Rainy	
24	200501BS00025		Wales	Major	2	1	24/01/202	Monday	21:30	HMV	30	0	Rainy	
25	200501BS00028		Wales	Major	2	1	18/01/202	Sunday	17:25	LMV	45	0	Rainy	
26	200501BS00029		Wales	Major	2	1	29/01/202	Tuesday	7:34	HMV	30	5	Snow	
27	200501BS00031		England	Major	1	1	19/01/202	Wednesda	16:35	LMV	45	1	Rainy	
28	200501BS00032		Wales	Major	2	1	30/01/202	Saturday	20:00	LMV	30	0	Rainy	
29	200501BS00033		Scotland	Major	2	2	29/01/202	Tuesday	13:15	LMV	30	5	Snow	
30	200501BS70001		Scotland	Major	2	1	1/2/2021	Sunday	18:20	HMV	35	5	Rainy	
31	200501BS70002		Scotland	Major	2	1	2/2/2021	Wednesda	7:25	LMV	30	1	Rainy	
32	200501BS70003		Wales	Minor	2	2	1/2/2021	Sunday	17:30	Two whee	30	0	Rainy	
33	200501BS70004		Wales	Major	2	1	3/2/2021	Thursday	12:30	LMV	30	1	Snow	

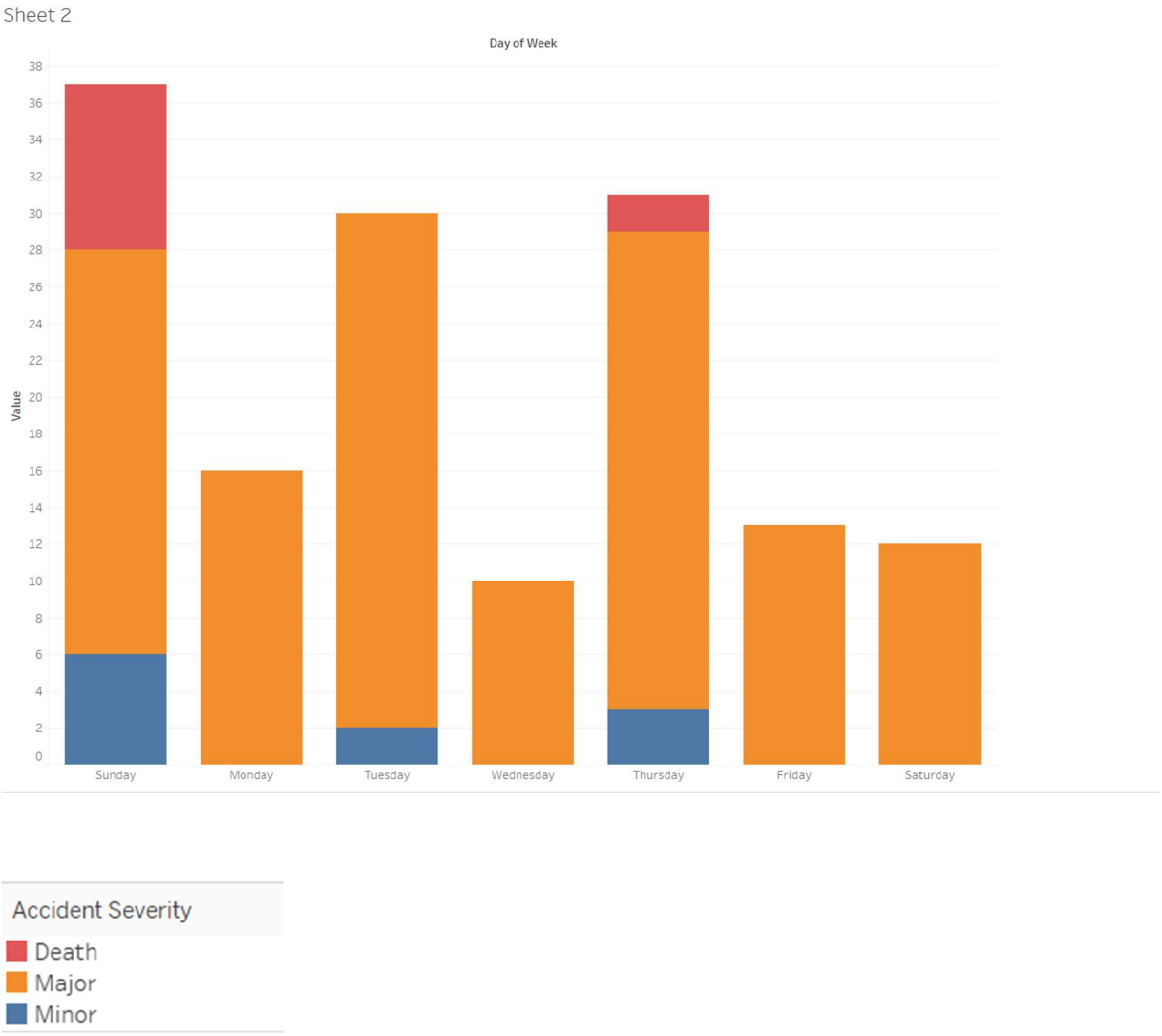
1.Butterfly Chart





A Butterfly Chart is a type of bar chart where two sets of data series are displayed side by side. It gives a quick glance of the difference between two groups with same parameters. Comparing no of vehicles and causalities in the regions.

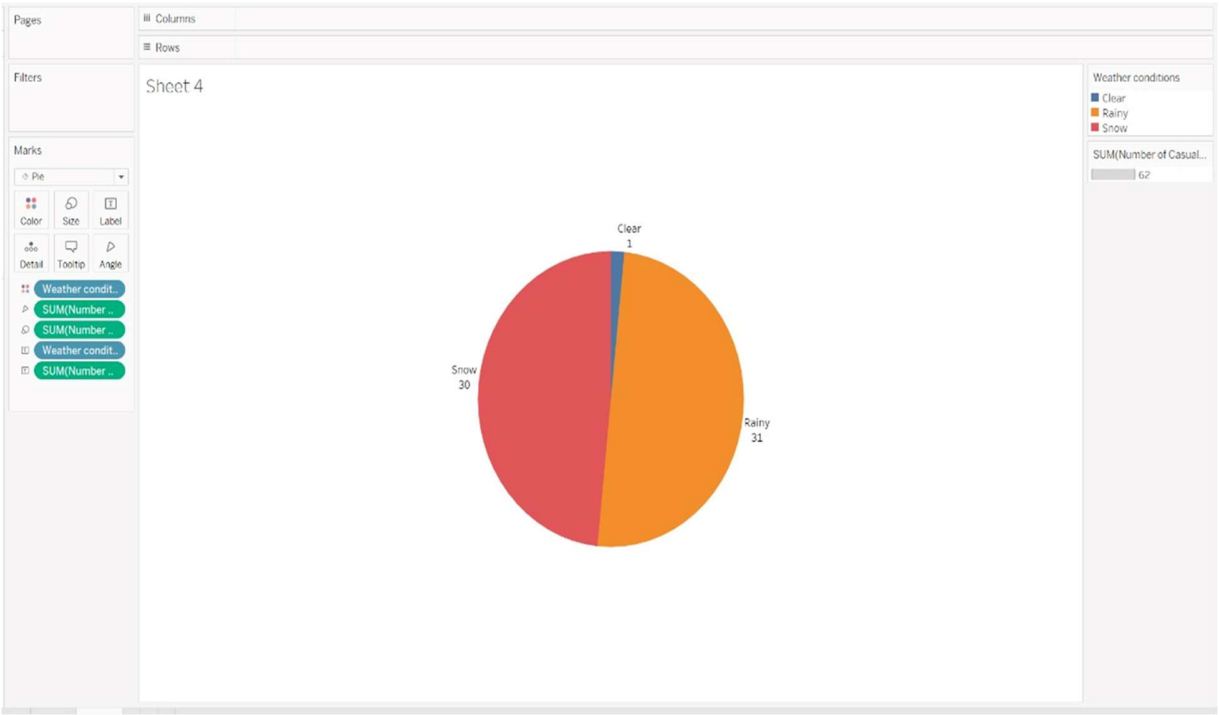
2.Bar Stacked Graph

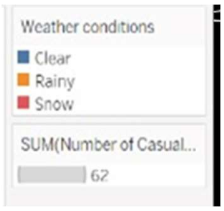


A stacked bar graph is a chart that uses bars to show comparisons between categories of data, but with ability to break down and compare parts of a whole.

This shows the number of casualties over different days and also the severity of accident.

3.Pie Chart

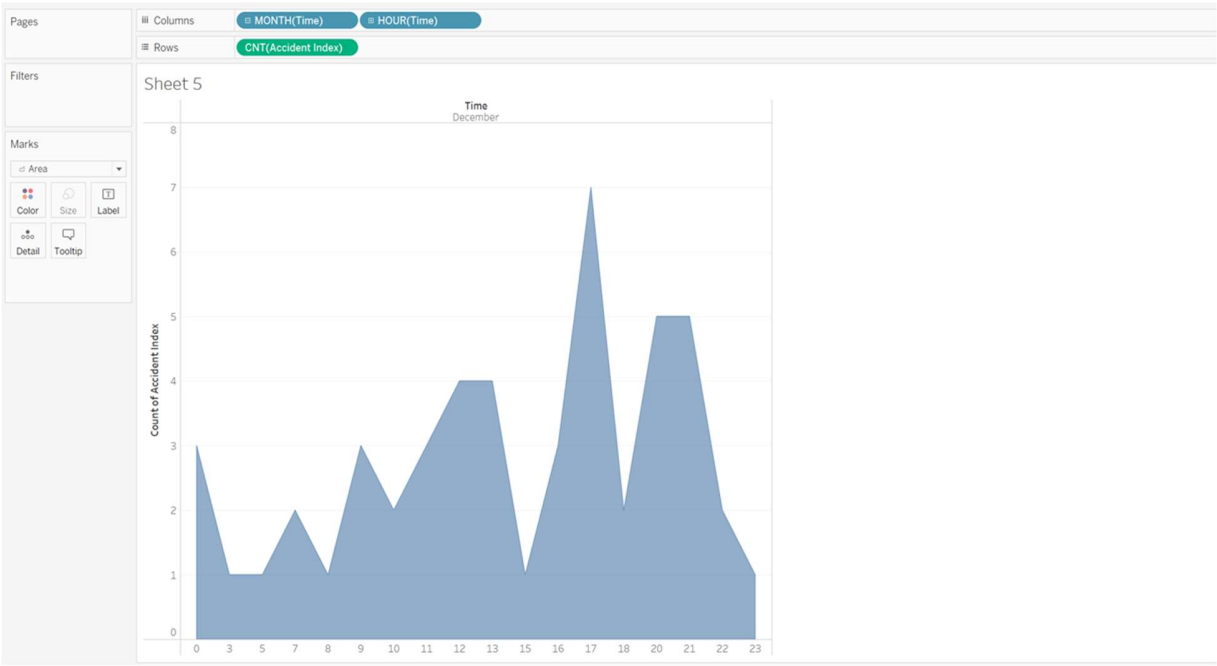




A pie chart is a circular statistical graphic, which is divided into slices to illustrate numerical proportion.

This shows the proportion of number of causalities in each weather condition.

4.Area Chart



An area chart or area graph displays graphically quantitative data. It is based on the line chart.

Here we show temporal visualization of the time and the count of accidents.

USER INTERFACE

For the customized dataset.

Visualization Tools Used:

Tableau Desktop was used for this Project, due to the following reasons:

- Remarkable Visualization Capabilities, along with thriving community and forums.
- Makes it easier to link geographical data to real-world maps.
- Provides a wide range of Visualizations to better analyze the data
- Multiple Data Source Connections.

Result and Conclusion

The results of our analysis include association rules among the variables, clustering of states in the UK on their populations and number of fatal accidents, and classification of the regions as being high or low risk of fatal accident. We have used Kaggle for modelling and visualization and used tableau for interactive dashboard.

Some commonly used statistical techniques imply assumptions that are often violated by the special properties of time series data, namely serial dependency among disturbances associated with the observations. The objective of this paper is to demonstrate the impact of such violations upon the applicability of standard methods of statistical inference, which leads to an under or overestimation of the standard error and consequently may produce erroneous inferences.

As seen in statistics, association rule mining, and the classification, the environmental factors like roadway surface, weather, and light condition do not strongly affect the fatal rate, while the human factors like being drunk or not, and the collision type, have stronger effect on the fatal rate.

Through the task performed, we realized that data seems never to be enough to make a strong decision. If more data, like non-fatal accident data, weather data, mileage data, and so on, are available, more test could be performed thus more suggestion could be made from the data.

Moreover, having established the adverse consequences of ignoring serial dependency issues, the paper aims to describe rigorous statistical techniques used to overcome. There are plans to extend the measurements to rate the probability of a collision for the road. These ratings are being used to inform planning and authorities' targets. For example, in Britain two-thirds of all road deaths in Britain happen on rural roads, which score badly when compared to the high-quality motorway network; single carriageways claim 80% of rural deaths and serious injuries, while 40% of rural car occupant casualties are in cars that hit roadside objects, such as trees. Improvements in driver training and safety features for rural roads will hopefully make this statistic less unfortunate.

References

- Kaggle Dataset Link - https://www.kaggle.com/code/sohamsave/accident-data-analysis/data
- Mind map online - https://www.mindmeister.com/map/2257726605