National College of Ireland

Data Visualization CA1

Topic: Road accidents

By:
Sankalp Ram Saoji
X17154171
MSc. In Data analytics
School of Computing
National college of Ireland

Contents

Introduction 4
Dataset background 5
Implementation tools: 6
Tableau: 6
Power BI: 6
Data visualizations: 6
Case 1: Number of accidents in New York according to the day of week for year 2015 7
Case 2: Percentage of accident occurrences in Monroe county (New York) according to weather conditions in 2015. 8
Case 3: Percentage of fatalities to total number of car accidents in eight counties of New York for year 2015. 9
Case 4: Percentage of serious injuries to sum of total accidents in nine counties of New York respective to days of week. 10
Case 5: Number of road accidents in New York from year 206-2015. 11
Case 6: Percentage count of accidents severity in UK for year 2016. 12
Case 7: Occurrence of road accidents according to various circumstances in UK for year 2016. 13
Case 8: Number of accidents according to vehicle fuel type in UK for year 2016. 14
Case 9: Gender wise count of car accidents in UK for year 2016. 15
Case 10: Percentage of total number of accidents at various vehicle speeds in MPH, in UK for year 2016. 16
Conclusion 17
References 17
Appendix 17-18

List of Figures

Figure 1: Dataset from socrata website..... 5
Figure 2: Number of accidents in New York according to the day of week for year 2015 .. 7
Figure 3: Percentage of accident occurrences in Monroe county (New York) according to weather conditions in 2015. 8
Figure 4: Percentage of fatalities to total number of car accidents in eight counties of New York for year 2015. 9
Figure 5: Percentage of serious injuries to sum of total accidents in nine counties of New York respective to days of week. 10
Figure 6: Number of road accidents in New York from year 206-2015. 11
Figure 7: Percentage count of accidents severity in UK for year 2016. 12
Figure 8: Occurrence of road accidents according to various circumstances in UK for year 2016. 13
Figure 9: Number of accidents according to vehicle fuel type in UK for year 2016. 14
Figure 10: Gender wise count of car accidents in UK for year 2016. 15
Figure 11: Percentage of total number of accidents at various vehicle speeds in MPH, in UK for year 2016. 16

Introduction

The rise in the sales of cars globally has led to more traffic in the world. The rapidly increased traffic contributes to more number of road accidents. These fast-track increase in road accident is an area of serious concern for the society. Most of the governments in the world are trying to imply rules and regulations to avoid these road accidents or to extremely minimize them. Road accidents causes a lot of damage to life, society, property. United Kingdom government is very active and concern towards road safety. In United Kingdom there are more than 3000 fatalities caused by road accidents every year, approximately more than 35000 people are severely injured in road accidents every year (UK department of transport) (Anderson, 2009).

The world health organization nearly estimates that there were more than 1 million people who died in road accidents in the year 2013, which is really a large number in the world mortality rate. Road traffic injuries and deaths is among the 10 most leadings cause of death globally. World health organization also predicts that the rank may go up by year 2030. There are around 50 million road accident cases without fatalities. These road accidents can be classified as severe or slight respectively. Severe road accidents can also cause to death indirectly. Therefore, road accidents have become a very big public issue. These road accidents directly affect the gross domestic product (GDP) of a country, especially the under developed and developing countries. Some of the road accidents occurs just because of some small human errors or human neglections or carelessness. These amount of road accidents can surely be prevented if everyone becomes a little more conscious while driving. Thus, there is a serious need for recognition of the road accidents. Our project focuses on road accidents in two different developed countries such as USA and UK. In this project we have implemented various visualizations about the road accidents for both UK and USA. The dataset collected will be useful for the government authorities for analysis based on the visualizations. These visualizations should help authorities for the analysis of road accidents in various circumstances. The governments can plan and introduce more initiatives to overcome the problem of road safety using the following visualizations. The analysis and visualizations will be helpful to save damage to human life as well as property. Individuals should also concentrate on these visualizations of road accidents to know well about them. This is important because these visualizations can open their eyes to increase road safety, so that individuals can be more careful while driving and should follow all the safety measures. These visualizations also indicate the basic scenarios when and how most accidents occur. Thus, with respect to these visualizations people should be aware of situations in which more accidents likely to occur so, that people will be more conscious and careful at that times to minimize the chances of road accidents (Global status report on road safety by WHO).

Dataset background

In our project we have downloaded two datasets, one is from the socrata public data website known as filtered crashed car dataset which was published and updated in November 2017. Another dataset selected for this visualization project is from the open data UK government website. The dataset from UK government is about road safety accidents in UK. The dataset chosen is published in September and updated in October 2017. Both the datasets used are from public data repositories and does not contain any personal information, so it free from any ethical concern. The first dataset which is from government of UK contains 69,000 instances of road accidents and attributes contains such as accidents index number, police force in the area, number of vehicles involved in the accidents, day of week, time of incidents, latitude, longitude, road type, speed limit, light conditions, weather conditions, road surface conditions, urban and rural where accidents occur, vehicle type, towing vehicle manoeuvre, hit object, junction location, sex of driver, age of driver, vehicle fuel type and age of vehicle. Using these data attributes, we have implemented visualizations about the road accidents in UK for year 2016. Second dataset consists of more than 2,52,000 instances and contains data for years 2006 – 2015 with various attributes such as accidents case accident case ID, case year, counties in New York, municipality type, date of incident, severity, number of injuries, serious injuries and fatalities in accidents, number of vehicles crash types, light conditions, weather conditions, direction of travelling in vehicle, age of driver, day of week, closet cross street and location. The data mentioned above is used to implement visualizations on various circumstances of accidents in nine counties of New York for year 2006-2015.

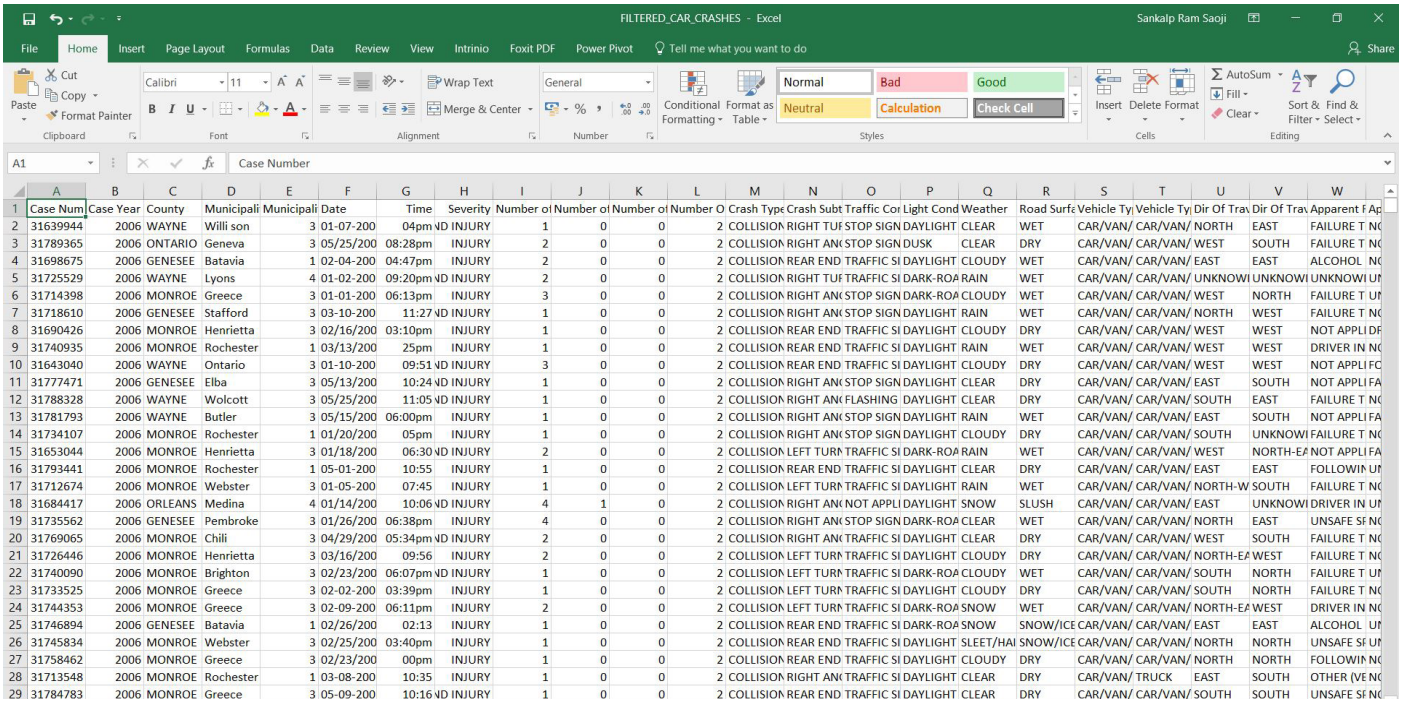

Figure 1: dataset from socrata website

Implementation tools

Tableau:

Tableau is used for visualizations in our project, tableau has good user interface as well as it makes visualizations in very interactive manner, visualizations made in tableau are easier to access and understand. Tableau has high comfortability level for implementing visualizations with its user. Tableau is compatible with almost every data types. Tableau has various number of in built graphs with high processing speed. Tableau also provides with specialized analytical and statistical features such as reference distributions, box plots, average lines with confidence intervals, trend lines and forecasting. User has whole accessibility to these tableau features.

Power BI:

Power BI is a tool by Microsoft and is used in our visualization project. Power BI is similar visualization tool as tableau but a has some little more features and number of graphs than tableau. The additional graphs in Power BI are very efficient, accurate and interactive. Power BI is more comfortable to user which are familiar with Microsoft office.

Data visualizations:

Data visualizations is termed as study of visual representations of data by Michael Friendly. Data visualization is mixture of arts and science. Data visualizations is used to create and communicate data efficiently with the use of statistical and analytical features using various tools. The visualization tool's uses various graphs, bars and box to deploy meaningful information. Visualization makes complex data in very easy readable, understandable and accessible format. Creative mind with help of easily available tools can perform and gain knowledge of many tasks using visualizations such as predictions, analysis, comparison with historical and present data, graphics (Friendly, 2009).

Visualizations

Case 1: Number of accidents in New York according to the day of week for year 2015.

This visualization is implemented in tableau and we have used bar plot for the above visualization. Bar plot is used for the visualization because it gives a clear view of data that how much accidents occur in New York according to each day of week. Thus, we can say that according to the visualization most of the accidents occur on Friday and least are on Sunday. In this visualization the color also tells us about the number of instances. Darker to faded shade tells us more number to least number respectively. We have also added an average line in the visualization to examine which days had accidents above and below average level.

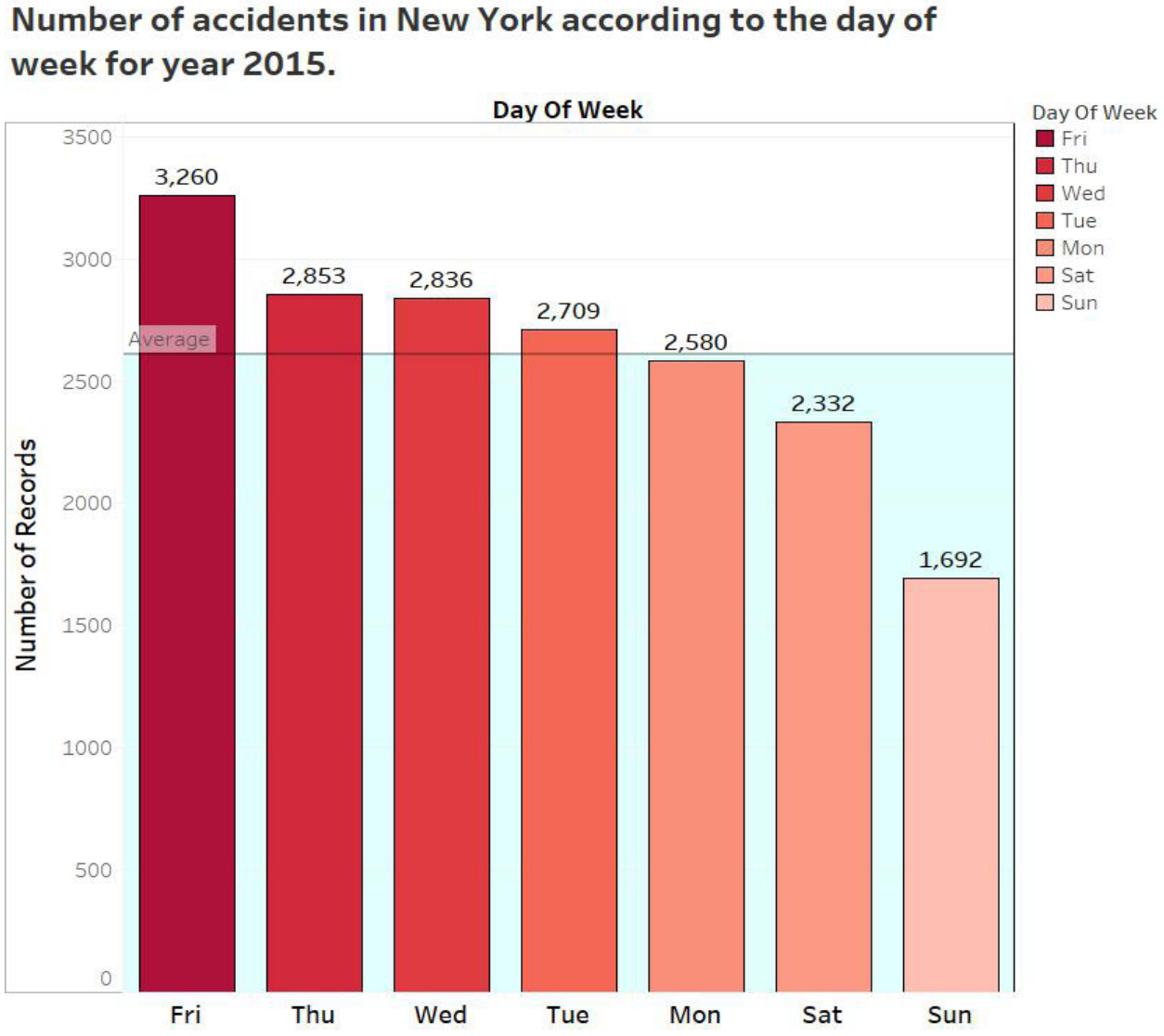

Figure 2: Number of accidents in New York according to the day of week for year 2015.

Visualizations

Case 2: Percentage of accident occurrences in Monroe county (New York) according to weather conditions in 2015.

The above visualization is all about the weather conditions. In this visualization it clearly tells us about the percentage of how much accidents occur in different weather conditions. Here, we can see that more than 50% accidents occur in clear conditions which Is denoted by yellow, 31% in cloudy which Is denoted by sky blue, 10% in rain which is dark blue and 5% in snow which is pink. We have chosen pie chart because sum of all the weather conditions comes up to 100%.

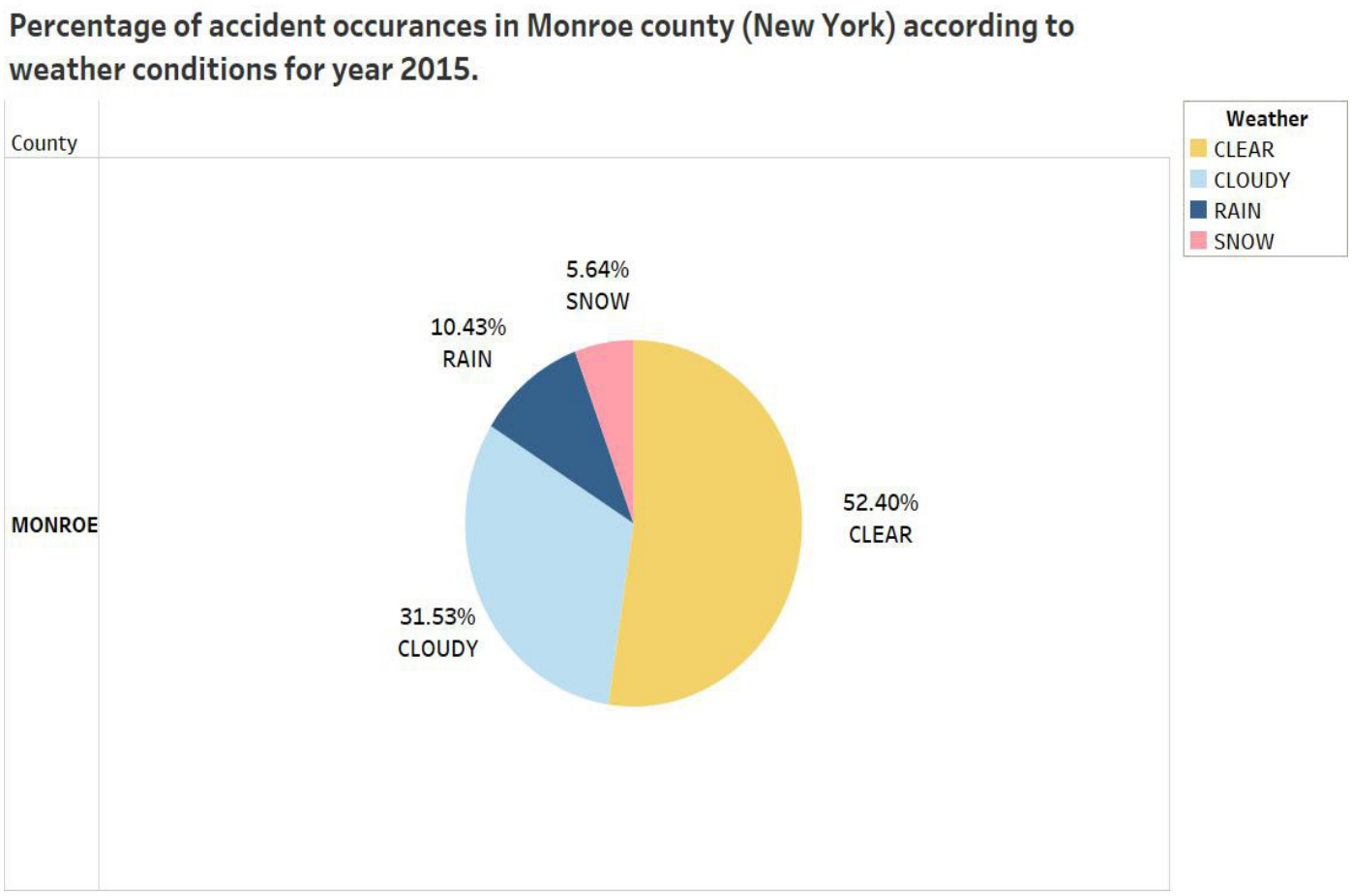

Figure 3: Percentage of accident occurrences in Monroe county (New York) according to weather conditions in 2015.

Visualizations

Case 3: Percentage of fatalities to total number of car accidents in eight counties of New York for year 2015.

We have chosen bubble chart for this visualization as bubble chart is easy to understand and remember. Here, we are representing percentage of fatalities to total number of accidents in eight counties of New York. The size and darker shade in the bubble chart describe the number of fatalities, larger the size more the number of fatalities. Hence, from the above bubble chart we can see that yates have maximum number of fatalities and whereas Orleans has least number of fatalities.

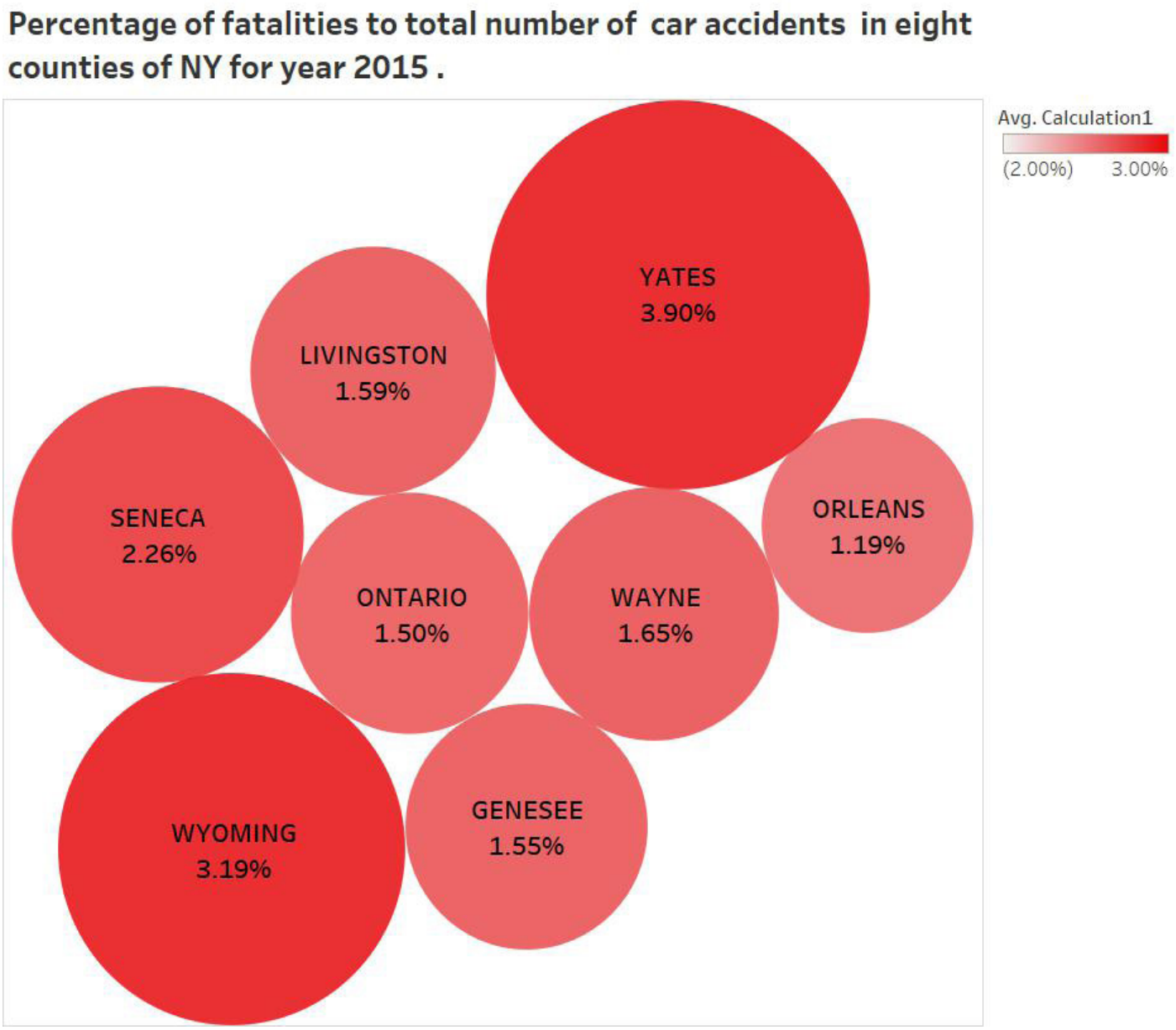

Figure 4: Percentage of fatalities to total number of car accidents in eight counties of New York for year 2015.

Visualizations

Case 4: Percentage of serious injuries to sum of total accidents in nine counties of New York respective to days of week.

The tree map chart tells us about serious injuries to total number of accidents according to days of week. The maximum size and darker shade tells us the maximum number of serious injuries, lighter shade and smaller size tells us less number of injuries. Thus, Sunday has maximum number of serious injuries despite of less number of accidents. Thursday has least number of serious injuries as compared to other days of week.

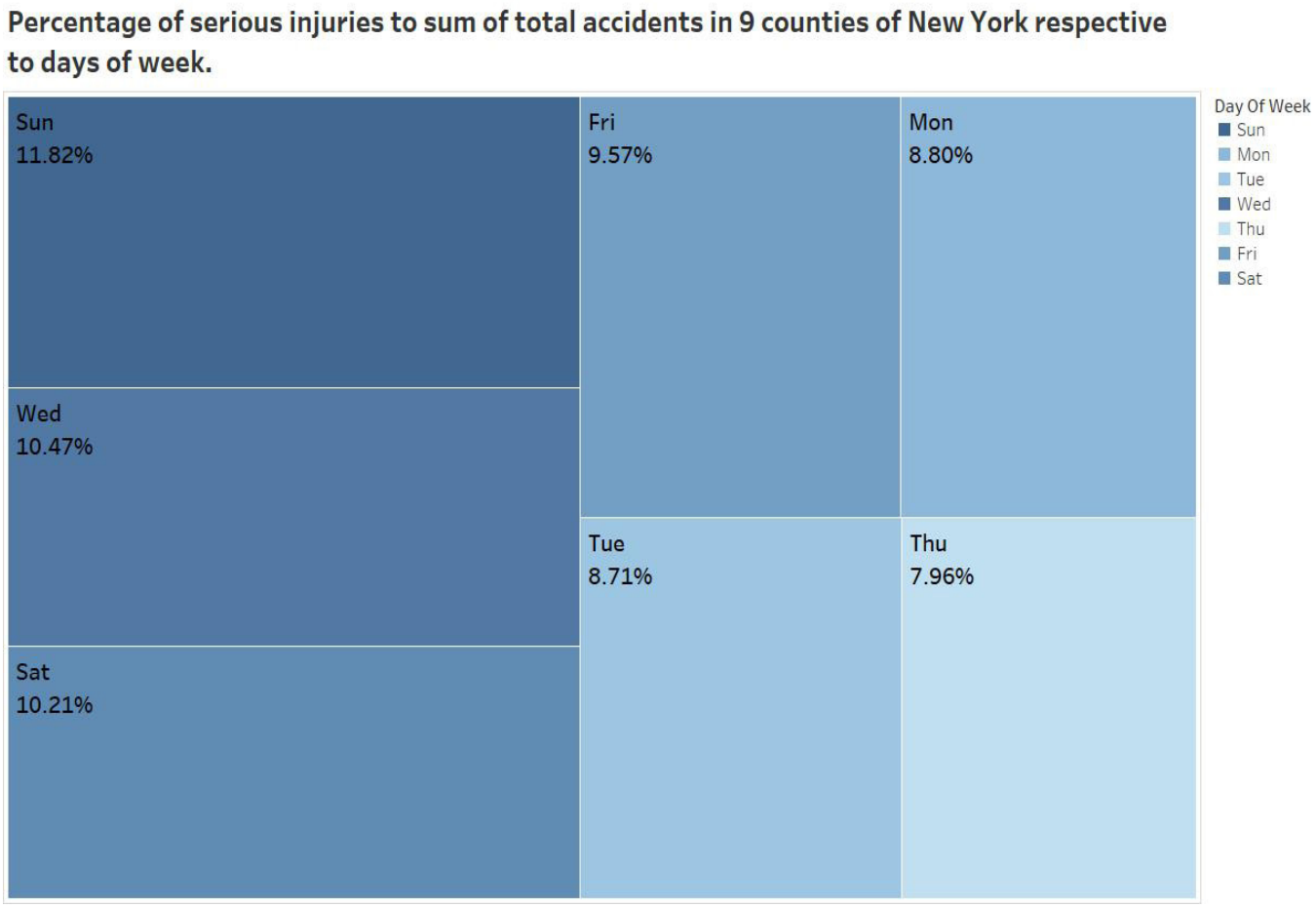

Figure 5: Percentage of serious injuries to sum of total accidents in nine counties of New York respective to days of week.

Visualizations

Case 5: Number of road accidents in New York from year 2006-2015.

This is the simplest and most easily understandable graph. This visualization tells us about the number of accidents occur from year 2006-2015 in New York. Therefore, we have chosen line graph as it shows the trend that in which year there were most number of accidents and in which year it dropped down. In our graph we can see that 2010 was having highest amount of accidents and then it dropped till 2014 and then was a slight increase in 2015.

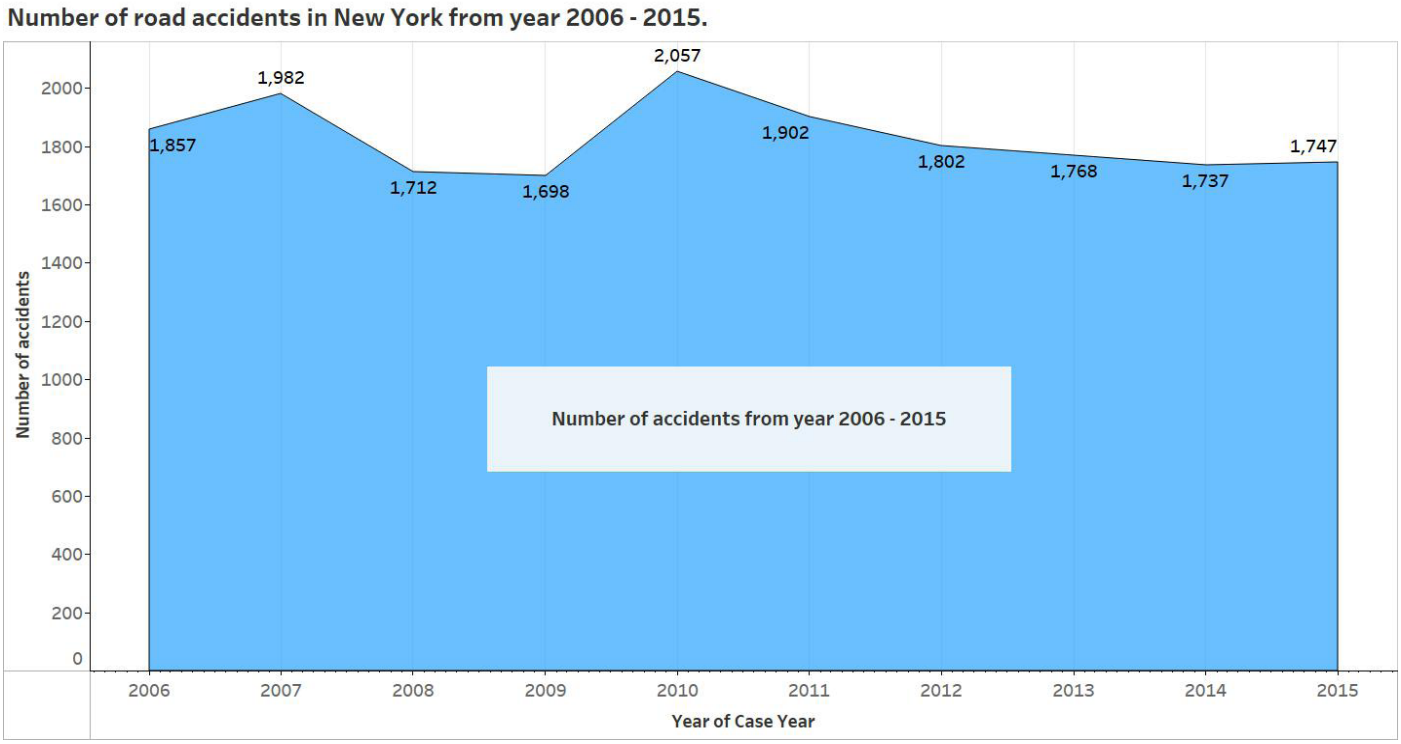

Figure 6: Number of road accidents in New York from year 2006-2015.

Visualizations

Case 6: Percentage count of accidents severity in UK for year 2016.

This graph tells us about the percentage of accident severity to the total number of accidents in UK for year 2016. The total accidents are divided into three categories such as fatalities, serious injuries and slight accidents. Total number of accidents sum up to 100%.

We have chosen donut chart for this visualization and is implemented in power BI. Therefore, we conclude that 82% accidents were slight and color used is blue, 1.24% were fatal and color used in red and 16 % were serious injuries which is denoted by black color, in UK for year 2016.

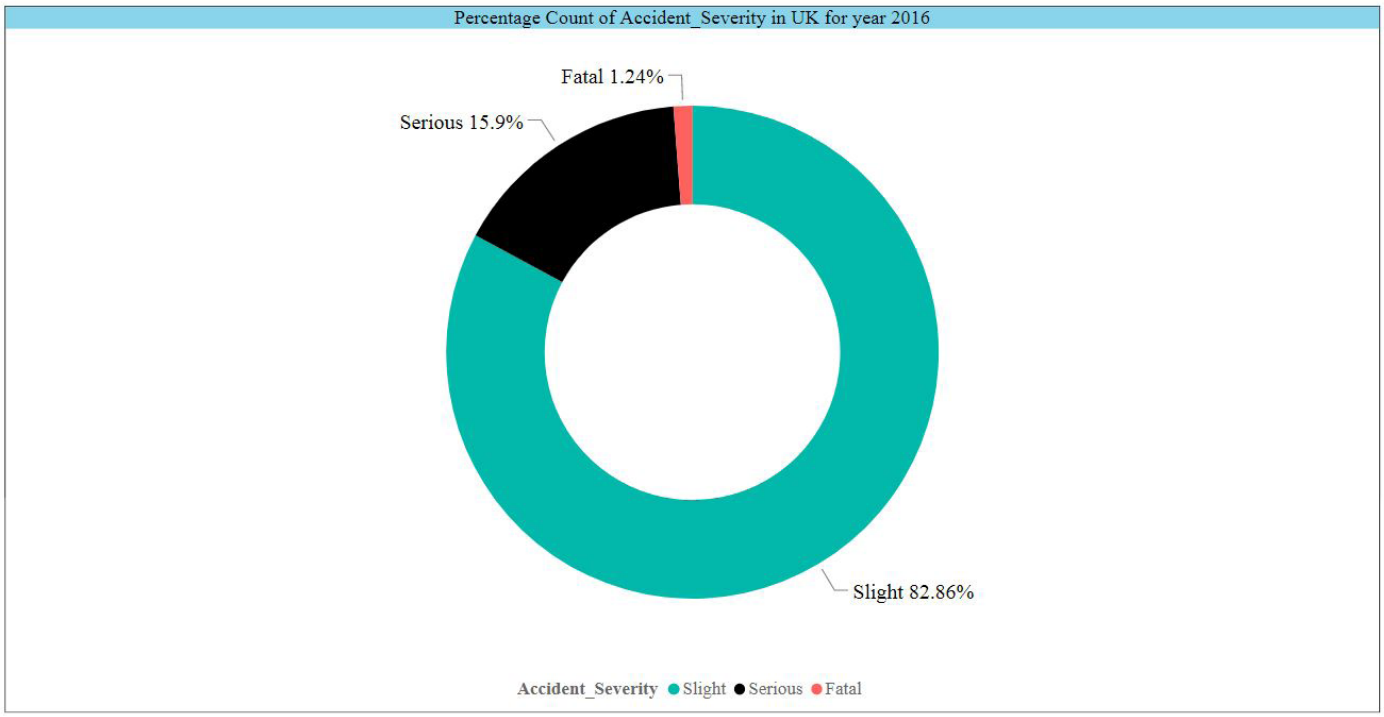

Figure 7: Percentage count of accidents severity in UK for year 2016.

Visualizations

Case 7: Occurrence of road accidents according to various circumstances in UK for year 2016.

This visualization clearly tells us about the causes of road accidents in UK for 2016. In the above graph we can see that there are 4 major causes of road accidents which are above the average line, out of which going ahead is on top. We have chosen horizontal bar plot for this because we have many instances so bar plot is most suitable for such visualizations to be more effective.

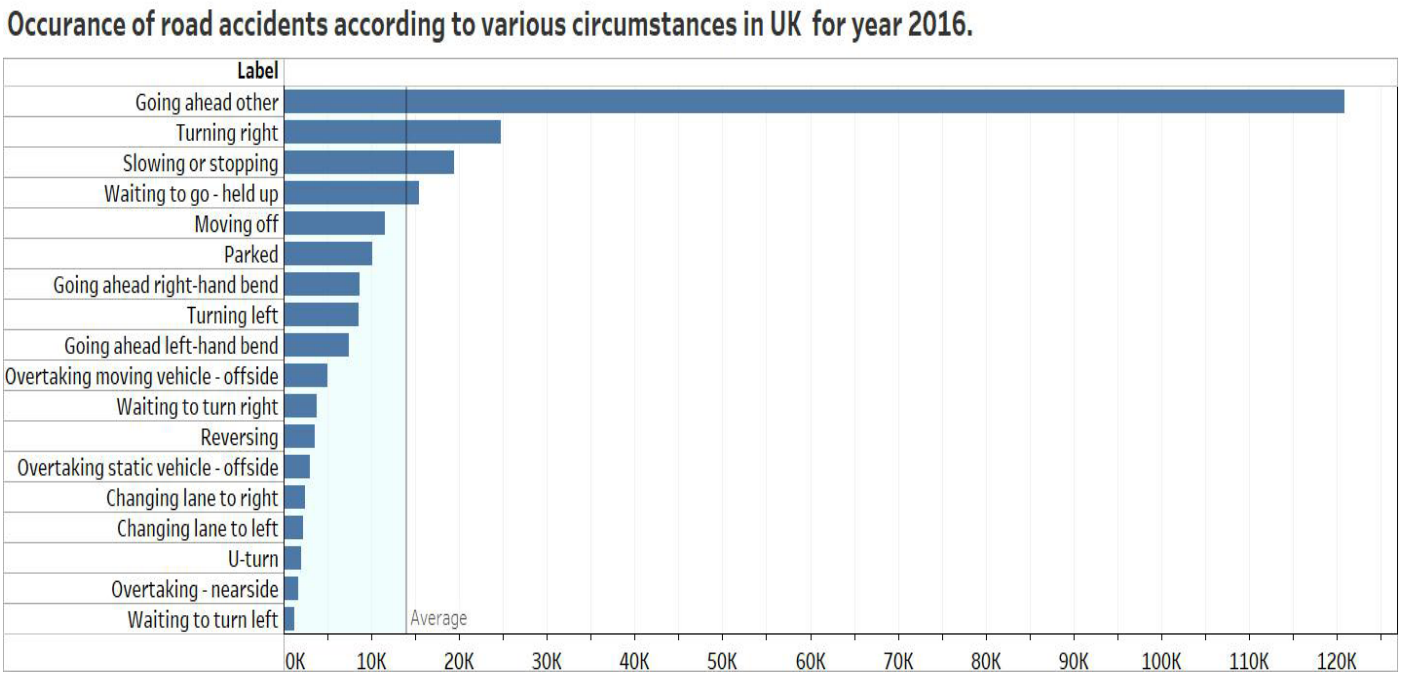

Figure 8: Occurrence of road accidents according to various circumstances in UK for year 2016.

Visualizations

Case 8: Number of accidents according to vehicle fuel type in UK for year 2016.

The visualization above tells us about the number of vehicle accidents according to their fuel type. So, the higher number of accidents are highlighted with the darker shade of blue and least number is with lighter shade of blue.

Here, we have also mentioned the total number of accidents in UK for year 2016. Heavy oils and petrol type have more number of accidents compared to other and LPG fuel type has least.

Number of accidents according to vehicle fuel type in UK for year 2016.

Label	
Electric	549
Electric diesel	696
Gas	150
Gas/Bi-fuel	1,113
Heavy oil	1,63,810
Hybrid electric	19,376
Petrol	1,09,444
Petrol/Gas (LPG)	228
Grand Total	2,95,366

Code

150 1,63,810

Figure 9: Number of accidents according to vehicle fuel type in UK for year 2016.

Visualizations

Case 9: Gender wise count of car accidents in UK for year 2016.

The visualization above divides the number of total accidents into two that is males and females. Here, we can see that in UK more number of accidents occur due to female drivers thanmale drivers.

We have also mentioned the grand total. females are denoted by pink color and males are denoted by blue color and red signifies the total number of accidents.

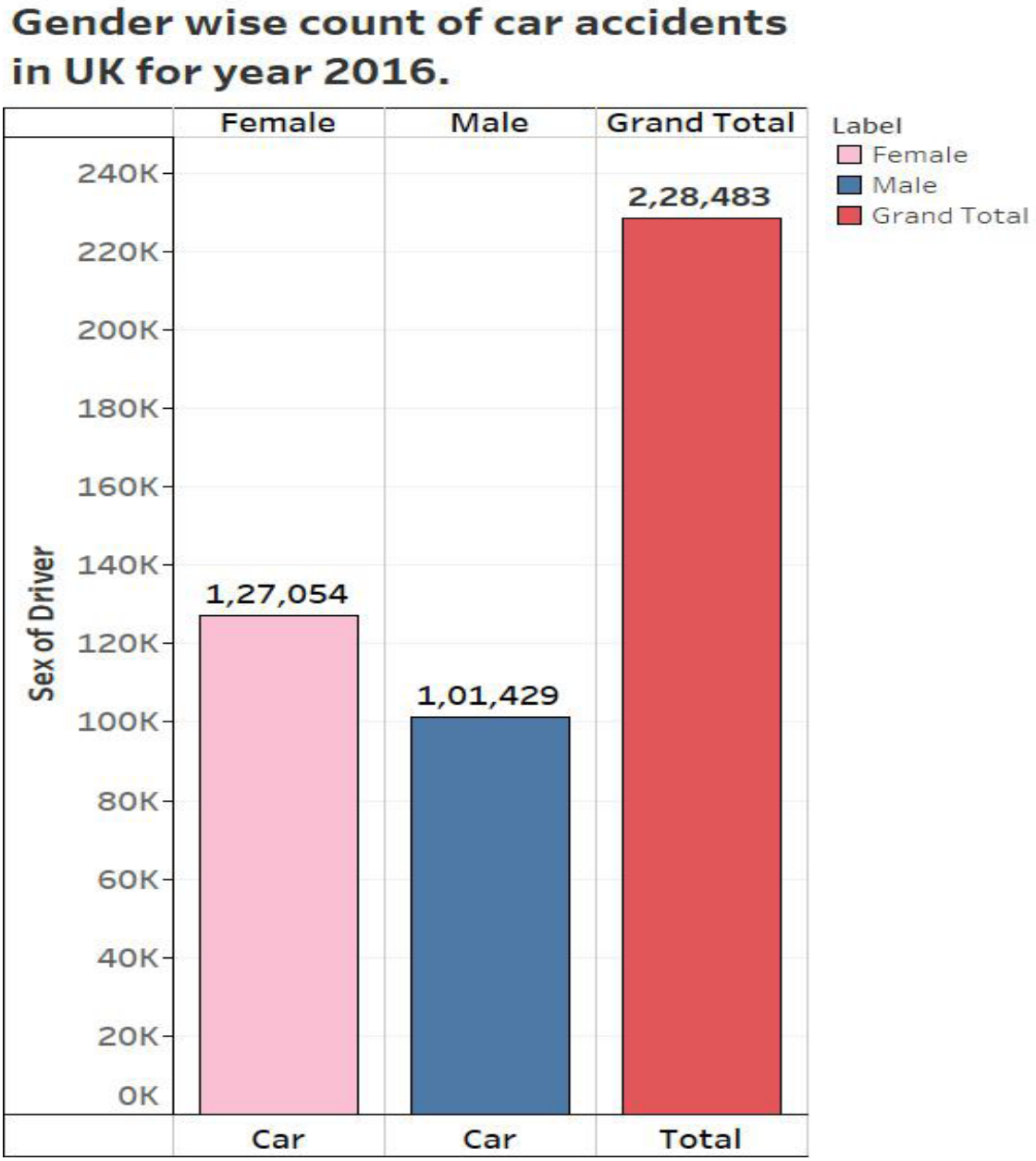

Figure 10: Gender wise count of car accidents in UK for year 2016.

Visualizations

Case 10: Percentage of total number of accidents at various vehicle speeds in MPH, in UK for year 2016.

This visualization tells us at which speed more number of accidents likely occurs. Thus we have implemented line chart with at different speed levels in MPH. The line chart shows that at the speed of 30 MPH there are more than 60 % occurrence of accidents.

We can assume that at 30 MPH people are less caring therefore might end up having an accident. At 50 MPH it may be safer high speed so people are comfortable and carful. At speed of 60 MPH and above it sometimes becomes out of control so there is possibility for accidents to occur.

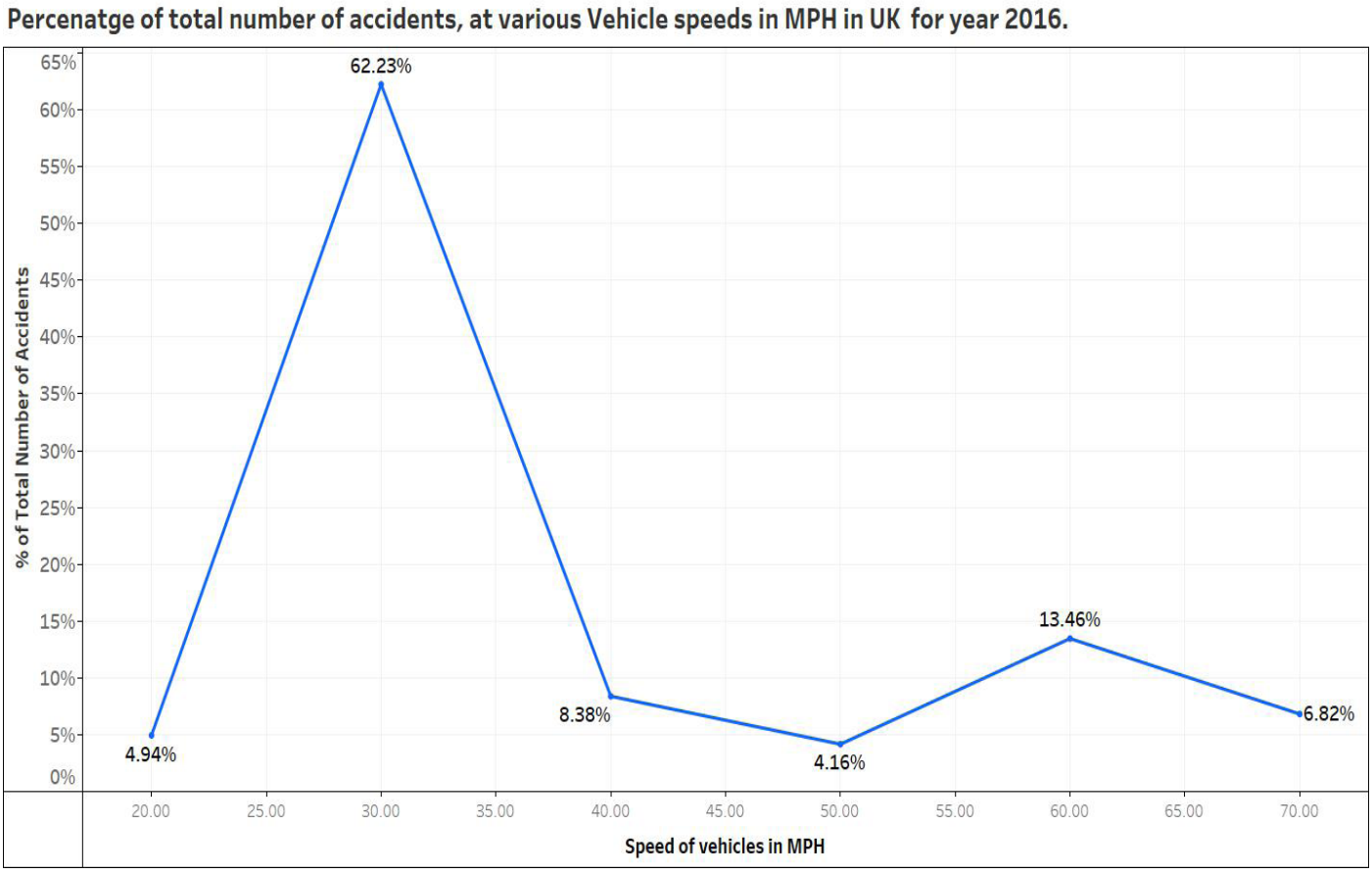

Figure 11: Percentage of total number of accidents at various vehicle speeds in MPH, in UK for year 2016.

Conclusion

Thus, we have implemented various visualizations on road safety and accidents for New York and UK. In this project we have included various number of graphs and plots and maps, implemented them on tools such as tableau and Power BI. These visualizations analyzes complex data and converts them into easily readable and understandable format.

References

Anderson, T. K. (2009) 'Kernel density estimation and K-means clustering to profile road accident hotspots', Accident Analysis and Prevention, 41(3), pp. 359–364.

Friendly, M. (2009) 'Milestones in the history of thematic cartography , statistical graphics , and data visualization', Engineering, 9, p. 2008.

https://opendata.socrata.com/w/gzw2-wb9x/y34g-bnf3?cur=--VLEbFtbcM&from=root

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

Appendix

Figure 1: Dataset from socrata open data website.

Figure 2: Number of accidents in New York according to the day of week for year 2015.
Graph used: Bar plot is used at it gives clear idea of number of accidents for each day of week.
Tool used: Tableau

Figure 3: Percentage of accident occurrences in Monroe county (New York) according to weather conditions in 2015.
Graph used: Pie chart is used to implement as total accidents sum up to 100%
Tool used: Tableau

Figure 4: Percentage of fatalities to total number of car accidents in eight counties of New York for year 2015.
Graph used: Bubble chart is used as we can differentiate percentage of fatalities using the size and colour of the chart.
Tool used: Tableau

Appendix

Figure 5: Percentage of serious injuries to sum of total accidents in nine counties of New York respective to days of week.
Graph used: Tree maps is used to implement this visualization as in tree map we can see the highest number of serious injuries on top with the detail of percentage and day of week.
Tool used: Tableau

Figure 6: Number of road accidents in New York from year 2006-2015.
Graph used: Line graph with area covered is used to visualize the number of accidents; line chart properly shows the trend of accidents.
Tool used: Tableau

Figure 7: Percentage count of accidents severity in UK for year 2016.
Graph used: Donut chart is used as total distribution of accidents is 100% so donut chart is efficient for this visualization.
Tool used: Power BI

Figure 8: Occurrence of road accidents according to various circumstances in UK for year 2016.
Graph used: Horizontal bar chart is used as there are more number of instances.
Tool used: Tableau

Figure 9: Number of accidents according to vehicle fuel type in UK for year 2016.
Graph used: Text tables is used for highlithing the maximum number of accidnents by the fuel type of the vehicles.
Tool used: Tableau

Figure 10: Gender wise count of car accidents in UK for year 2016.
Graph used: Bar chart is used to signify male and female's accidents.
Tool used: Tableau

Figure 11: Percentage of total number of accidents at various vehicle speeds in MPH, in UK for year 2016.
Graph used: Line graph is efficient for visualizing percentage of accidents according to different speed in MPH.
Tool used: Tableau