Data Information:

The leading cause of death and severe injuries in the world is the Road Accident. In India, the rate of occurrence of an accident is more than a critical limit. In 2018 alone, the country reported around 151 thousand fatalities due to road accident. It may not be stopped altogether, but can be reduced. A World Bank study has found that if India were to halve road deaths and injuries between 2014 and 2038 successfully, it could potentially add 14 per cent to its GDP per capita. This paper deals with the visualization of Road Accident data to identify the trends that are causing all the possible accident. Visualization will help to find the hidden pattern in the data to find the standard features between accident.

Dataset:

Data has been taken from **https://data.gov.in/,** which is an open government data platform of India. Data consist of 3 CSV Sheet having 5 million rows from the year 1970 to 2017. Data of Road accident in India are collected by Transport Research Wings, Ministry of Road Transport and Highway, and from the police headquarters of various Sates, UT's and million-plus cities of India. (Telangana State didn't exist for this dataset and hence empty.)

Attributes:

Original data refers to Cause of Accident, type of Vehicles (Government / Private), monthly death report, states in India, the Death count of Male, the Death count of Female, Total Death, Total Number of Road Accidents, Persons Killed and Injured, Number of Accidents per Ten Thousand Vehicles, Number of Accidents/ Lakh Population, Number of Persons Killed / Ten Thousand Vehicles, Number of Persons Killed / Lakh Population, Number of Persons Killed/ Ten Thousand Kms of Roads, Number of Persons Injured / Lakh Population, Urban and Rural distribution.

Data Exploration:

Tableau tool was used to visualize the data obtained from data.gov.in website and discovered various relations like Number of State Vs Number of Accident, the Percentage difference in total accident state wise, Population Vs Accident over the period of time and many other making the data a rich source for visualizers. For my visualization I am going to consider the different state, different cause of accident over the period of time, the number of deaths of male and female in these states, urban and rural division of state (of 2016), and percentage of male and female in a different type of accident.

Data Processing, Cleaning and Integration:

Data is needed to be pre-processed before visualization. Data size is too large, so we take the data from the Year 2006 to 2017 into consideration. Sate name in one CSV is converted to upper case, in order to make proper relation with the remaining dataset. The data type is also changed according. Taking Government and Private vehicle as a single unit. Other attributes that are taken into consideration are type of vehicle used (CAUSE), State name (STATE/UT), Year of accident (Year), the number of male and female (Male, Female), urban and rural area state wise, and Total Column. The data has also shown each vehicle to Male and Female death count from the year (2006-2017). All the NULL values were replaced by 0 in the data.

Visualisation:

In India, there are 28 States and 8 Union Territories, so knowledge about the State and UT's with a large number of accidents is essential that can help us to take some additional preventive measures to those States and UT's. Representation of these State and UT's is done using Map graph, that helps us to analyse the data as the location on an interactive map. Areas are coloured on the basis of the total number of accidents of that region. Darker colour represents a greater number of accidents while a lighter shade of the colour represents a smaller number of accidents. State name and the number of accidents is displayed when hovered over the map.

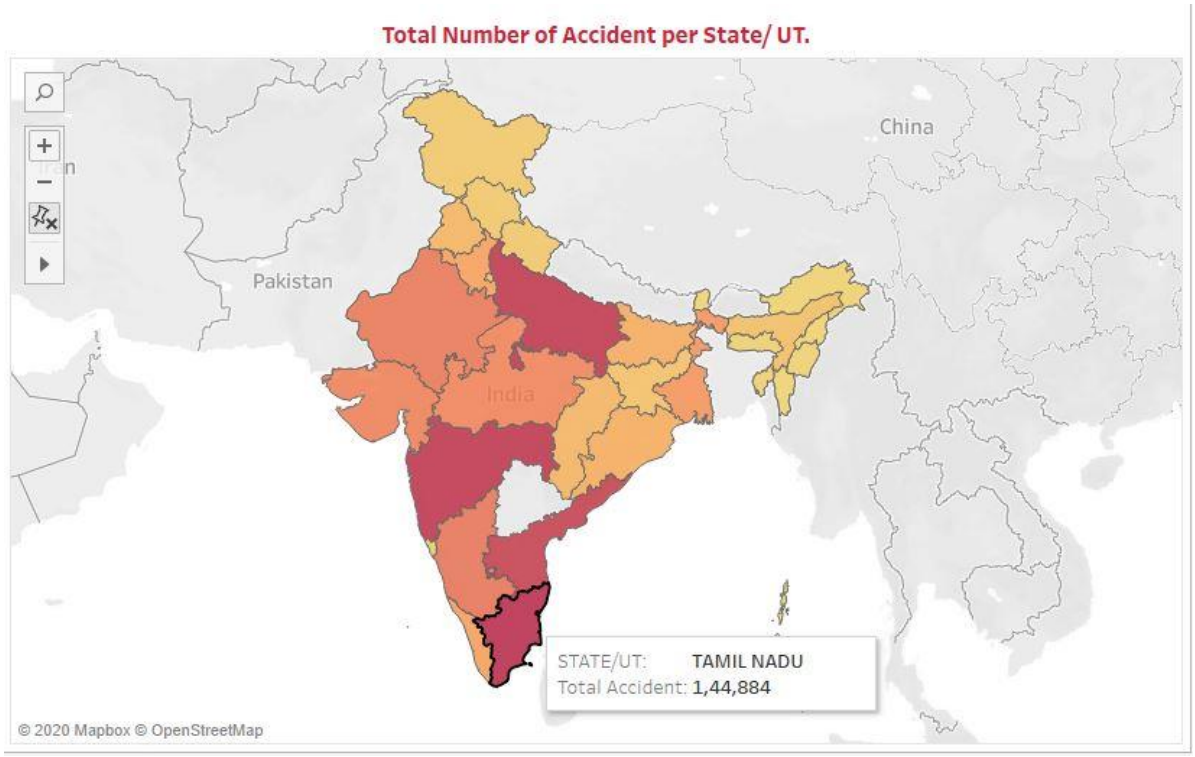

Fig 1: Map Graph

By analysing the Map Graph, we can see that Tamil Nadu a state in the southern part of India has the largest number of accident cases. To get more insight into this analysis, I have created one trend graph. This graph shows the percentage difference (created by using the calculated field in Tableau) of the total number of accidents over the years of graphs from the immediate previous year for particular State and UT. It gives the overall information about the number of accidents in a particular state on a yearly basis and helps in identify trends of the accident when compared to the previous year. Animated Trend Line Chart is used to visualise this trend graph from the year 2006 to 2017. After analysing trend graph for Tamil Nadu, we can see that the highest number of road accident was in the year 2010 after which there was more

than 60% drop in the accident for the year 2011. From 2012 till 2107 no major dip in trend.

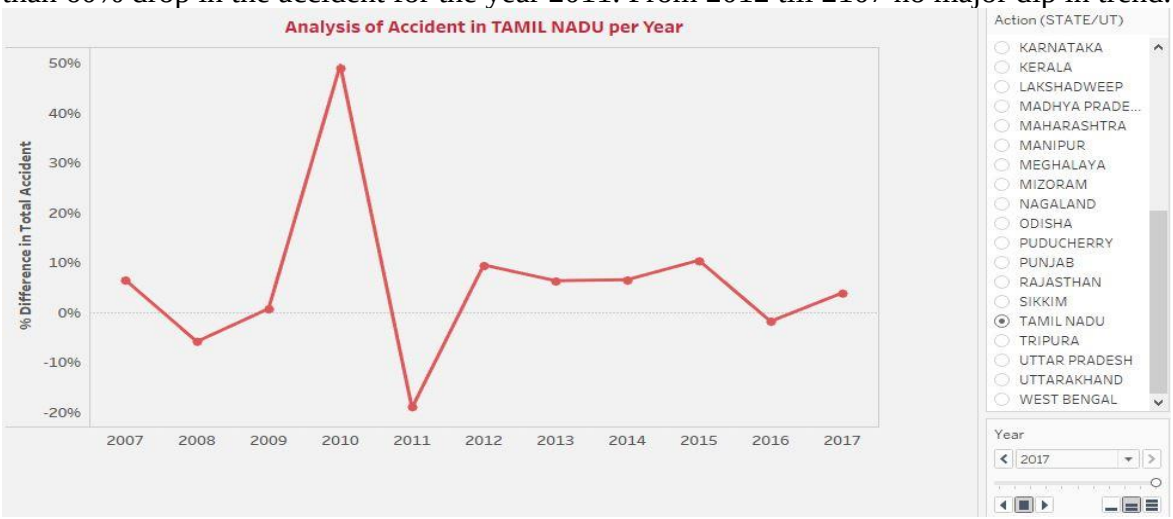

Fig 2: Animated Trend Graph

In Addition to the year with the highest number of accidents, the cause of this accident is needed to be analysed. I am splitting the cause of death on the bases of the percentage of male and female death. In order to display two data side by side, Butterfly Chart also knows as Tornado Chart is used. For the creation of butterfly chart, Two zero measures are created, which is then added to the percentage of male and female in different causes of accident respectively using dual axis function. From the butterfly chart, we can conclude that Truck, two-wheeler, and Buses are the primary cause of an accident.

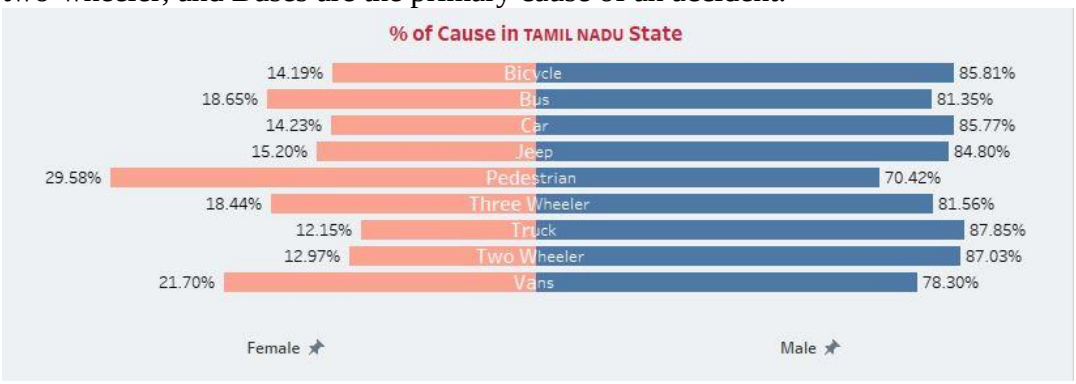

Fig 3: Tornado Graph

Below bar graph and area graph gives information about the number of males and females death due to different types of vehicles. Bar graph represents the number of males whereas the area graph is used for female. In the case of Tamil Nadu cause of death for male was

Truck while for women it was Bus.

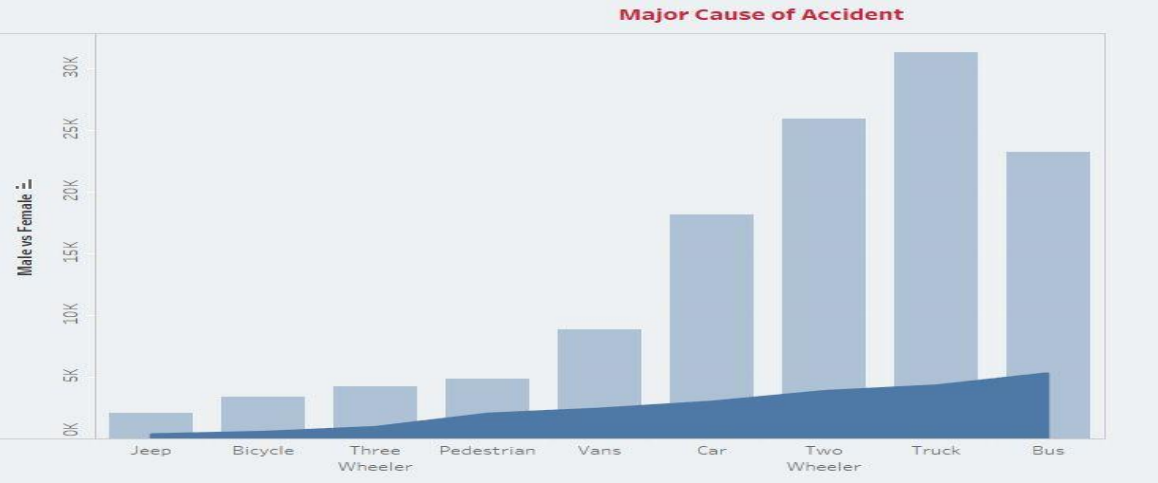

Fig 4: Overlapping Graph

Packet Bubble chart helps in visualizing the Cause of death. Larger the bubble and lighter the colour represents a greater number of accidents by that medium. In the case of Tamil Nadu, it can be inferred that 35,628 accident was reported by Truck whereas accident occurred by jeep is comparatively small in number.

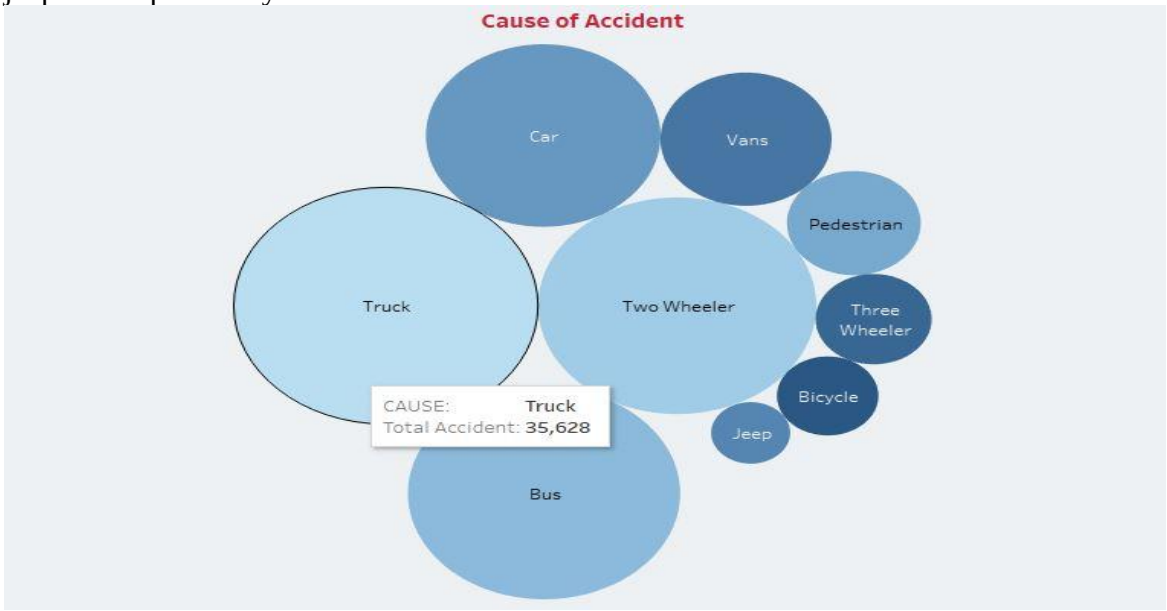

Fig 5: Packet Bubble Chart

I have also analysed the Rural and Urban distribution of Tamil Nadu using a Stacked bar for the year 2016; it shows that number of accidents in a rural area is more as compared to the urban area. This in all tells us that in order to reduce the number of accidents in Tamil Nadu, more measurement should be taken in rural areas of Tamil Nadu.

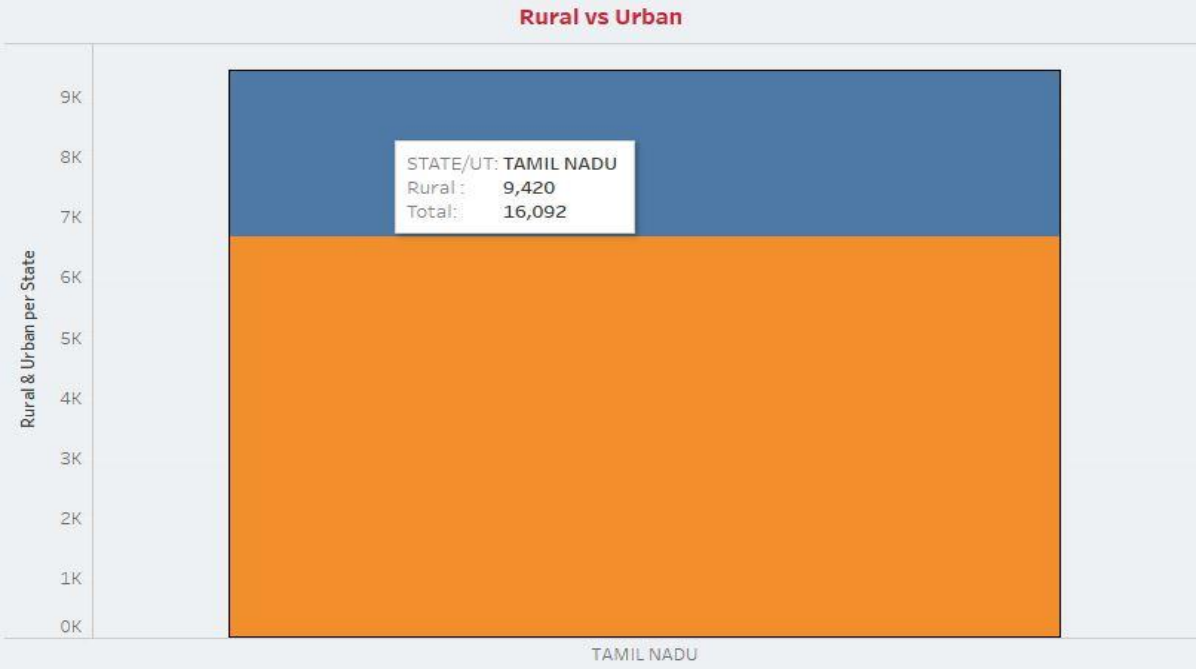

Fig 6: Stacked Graph

Tableau Tool Used for Visualization:

Tableau is used for visualization as it supports complex computations and various data exploration technique to get the hidden insight which cannot be derived by just looking at the excel sheet.

Conclusion:

India is the 2nd largest populated country in the world. In support of proper connectivity between different regions with in the country, India has vast road networks. However, Road safety is a national concern. With the available data on a road accident from the Indian government website, I try to analyze different dimensions and magnitude of the road accident.

Visualization of this data gives information about the vehicle, which is causing more fatalities in a particular state and at the country level also. Drop in the percentage of accident from previous year give an idea about measures and precaution that have been taken by the government, road authorities at state and national level towards road safety. Map chart can help us to define a particular state which needs some immediate action in order to reduce the accident count quickly and effectively. For making more precision decision one can look for the urban and rural graph to dig more into a particular sate and find which region is more affected and need more attention. This visualization might create awareness and can help government policymaker, citizen, and civil society organization working in the road safety department. Though the visualization provided us with the right amount of information regarding accidents and their causes, we realized that data would be never to be enough to make a firm decision. If more data, like the time when the accident happens, state of the driver (drunk or not) are available, more test could be performed thus more analysis could be made from the data.

References:

1. Nishtha Saluja (November 19, 2019) Road accidents claimed over 1.5 lakh lives in 2018, over-speeding major killer (Online)
   Available :https://economictimes.indiatimes.com/news/politics-and-nation/road-accidents-claimed-over-1-5-lakh-lives-in-2018-over-speeding-major-killer/articleshow/72127418.cms?utm_source=contentofinterest&utm_medium=text&utm_campaign=cppst
2. Amit (MARCH 6, 2016) Analysis of Road Accidents – India (2008-12) [online]
   Available at: https://amitman3.wordpress.com/2016/03/06/analysis-of-road- accidents-india-2008-12/
3. Bindu ((MAY 7, 2019) Road accidents still cause most deaths: WHO report [online]
   Available at: https://www.thehindu.com/news/national/road-accidents-still-cause-most-deaths-who-report/article27061224.ece
4. Prachi (March 28, 2017) Overview of Road Accidents in India [online] Available at: http://prsindia.org/policy/vital-stats/overview-road-accidents-india
5. Shraddha (July 28, 2015) This Is How Pink Came To Be A 'Girly' Colour [online] Available at: https://www.scoopwhoop.com/inothernews/how-pink-became-the-girly-colour/