**IFT 598 Data Visualization & Reporting for IT**

Hima sree Chalasani

Nikhil Jain Sanghami Padmaprabha

Venkata Bala Nagendra Srirangam

Project - Phase II Decision Making

Asmaa Elbadrawy

April 16th, 2023

**Section 1: Choosing Visualization tools**

For this project, we will use Tableau as our primary tools. This is because Tableau is known for turning data from almost any system into useful insights quickly and easily. The drag-and-drop method is all that is required to use Tableau. In addition, their world-wide data community, top-notch training resources, and support services offer unparalleled assistance. With Tableau, we can effectively deal with huge volumes of information and accomplish phenomenal outcomes in minutes by utilizing different strategies to break down the information.

We prefer Tableau over other tools because it is easy to use and doesn't require any coding or technical knowledge, making it easy for new users to use. By integrating common fields, Tableau can quickly incorporate new data sources. To get valuable insights, users can easily create multiple charts and switch between different visualization models to find the one that best conveys their message.

R can be used to run more complex analytical models, and the results can be imported into Tableau to create flexible and easy-to-understand visualizations. Tableau also supports some basic calculations. Businesses are able to derive valuable insights from their data thanks to the visual analytics that are provided by this combination of R and Tableau. Scene's visual investigation interface works on information examination and association, making it simpler to reveal inconspicuous examples inside huge datasets.

Businesses are able to maintain their competitiveness in their respective industries by analyzing data for key insights. As data volume and velocity increase, businesses attempt to integrate diverse and complex data. These companies can use Tableau to quickly and easily extract hidden insights from their data.

**Section II: Explanation of Required Data Pre-processing**

Below is the initial dataset with all the columns:

A screenshot of a computer

Description automatically generated with medium confidence

A screenshot of a computer

Description automatically generated with medium confidence

A screenshot of a computer

Description automatically generated with medium confidence

1. Extracted Start time hours as a separate column in Tableau so that it will be easy to visualize the hour-wise data.
2. Excluded certain weather conditions as much data was not available for certain weather conditions and grouped the common weather conditions together to avoid too much data.
3. Excluded the accident IDs column as it was not serving any purpose in our visualizations.
4. Removed the data column named precipitation as this info was not mentioned for all accidents.
5. End time excluded.
6. Removed data column named precipitation – as this info was not mentioned for all accidents.
7. Calculated the total time of an accident from start time and end time data provided.

**Section III: List of Final Set of Questions**

1. What is the distribution of accidents across different states in the US? Which states have the highest and lowest numbers of accidents?
2. What is the distribution of accidents across different days of the week and the severity trends?
3. What is the distribution of accidents across different times of day?
4. How do weather conditions and other factors (such as road surface conditions or visibility) impact the likelihood of accidents occurring throughout the year?
5. What is the average time taken to clear an accident at different locations?
6. How does the frequency and severity of accidents vary across different traffic signs, such as bumps, crossing, junctions?
7. What is the year-wise severity trend of the accidents?
8. How does the twilight time affect the accident severity?
9. How does the severity of the accident affect the traffic congestion (distance of the road affected due to accident)?
10. How does the visibility in miles affect severity and number of accidents?

**Section IV: Dashboard Plots**

Graphical user interface

Description automatically generated with low confidence

Graphical user interface, application, website, map

Description automatically generated

The above plot shows the number of accidents that have occurred in each state. This will help transportation companies, government agencies and insurance companies to take calculated decisions. The pre-attentive attributes used in the above plot are:

Color,

Graphical user interface, application

Description automatically generated

The above plot shows the number of accidents that have occurred on each day of the week. This plot shows relationship between days of the week and the number of accidents.

The pre-attentive attributes in the above plot are:

Orientation

Chart, bubble chart

Description automatically generated

The above plot shows the relationship between the number of accidents and the time of the day. This will help us figure out if the accidents are occurring more at night or day.

The pre-attentive attributes used in the above plot are:

Color

A piece of paper with writing on it

Description automatically generated with medium confidence

The above plot shows the average time taken to clear the accident in each state.

The pre-attentive attributes used are color.

Graphical user interface

Description automatically generated with medium confidence

The above plot shows the relationship between weather conditions and the number of accidents. This will help us see if there is any relationship between the weather condition and number of accidents.

The pre attentive attributes are:

Length

Graphical user interface, chart

Description automatically generated

The above graph shows the number of accidents that have occurred at each traffic sign.

The pre-attentive attributes used are:

Length

Chart, line chart

Description automatically generated

The above plot shows relationship between the number of accidents that have occurred across different years. The pre-attentive attributes used are:

OrientationGraphical user interface, text, application

Description automatically generated

The above plot shows the number of accidents that have occurred during different twilight times. The pre-attentive attributes used are:

Color

Graphical user interface, application, table

Description automatically generated

The above plot related the severity of accidents with the distance of road affected with the accident. This will help us know how much the severity of accident will affect the traffic congestion.

The pre-attentive attributes used are:

Graphical user interface, application, table, Excel

Description automatically generated

The above plot shows how the visibility in miles will affect the number of accidents.

The pre-attentive attributes used here are:

Orientation

**Section V: Dashboard Interactivity**

1. Severity level control from the ‘Severity’ data attribute which contains a dimension value of the severity to interact with the traffic congestion visualization, effect of weather condition visualization and accidents in different twilight visualization.
2. Year control from the ‘Start Time’ data attribute which contains only years to interact with the severity visualization.

**References**

**Mural -** <https://app.mural.co/t/greeshmasworkspace0156/m/greeshmasworkspace0156/1680555917958/0960ed43596c4be3fb8a716bdd2332f73c5e5a3d?sender=ufd0807b59c2de6a557262362>

**Dataset source –** [**https://www.kaggle.com/datasets/sobhanmoosavi/us-accidents**](Dataset%20Link)

Moosavi, Sobhan, Mohammad Hossein Samavatian, Srinivasan Parthasarathy, and Rajiv Ramnath. *“A Countrywide Traffic Accident Dataset.”*, arXiv preprint arXiv:1906.05409 (2019).

Moosavi, Sobhan, Mohammad Hossein Samavatian, Srinivasan Parthasarathy, Radu Teodorescu, and Rajiv Ramnath. *“Accident Risk Prediction based on Heterogeneous Sparse Data: New Dataset and Insights.”* In proceedings of the 27th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, ACM, 2019