Business Analytics: Data Visualization

Quarter: Autumn 2021

Date: 12/12/2021

Title of the project: Analysis of US Accidents (2020)

Documentation – Tableau Functionality Tutorial

Vertical industry: Mobility Industry

Business process: Safety, Administration and Preventive Measures

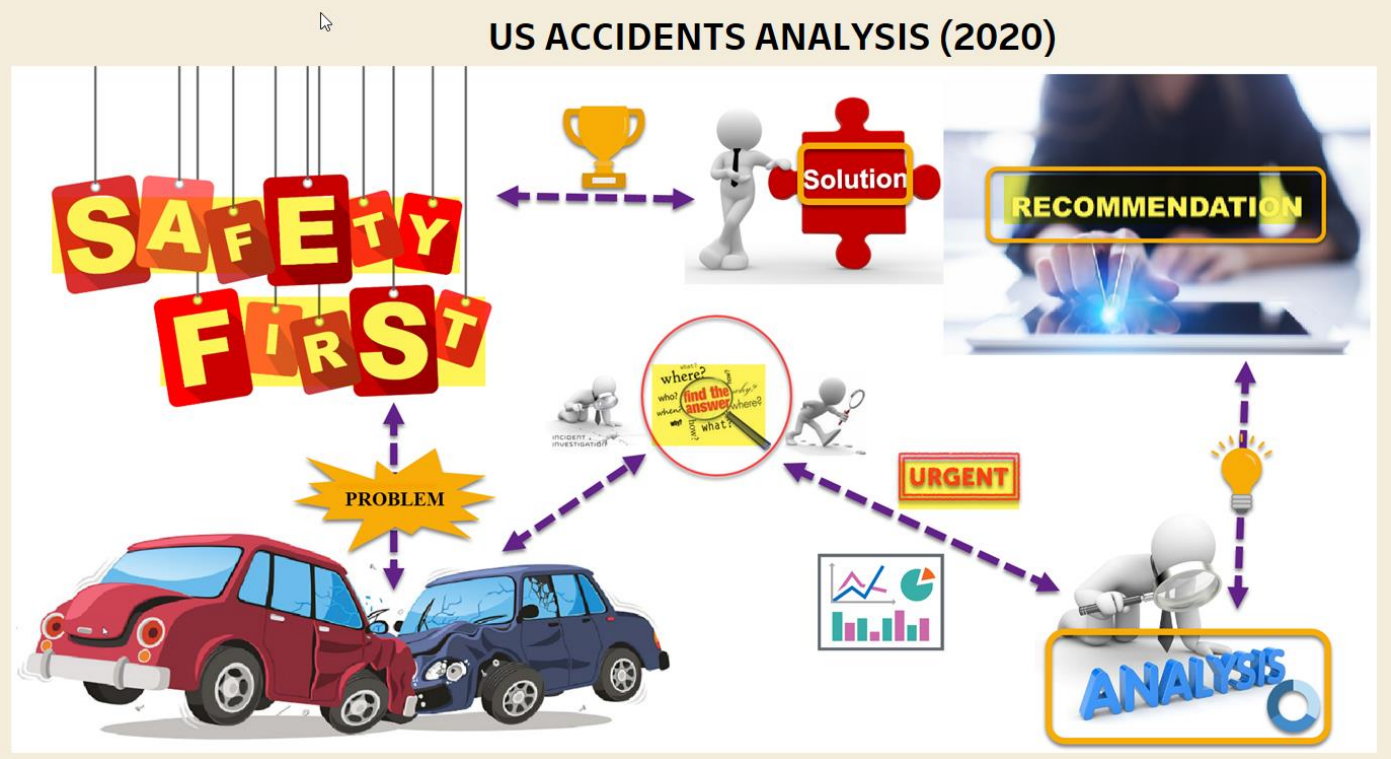
Submitted by: Shephali Jain

I certify that I have completed this assignment within the Academic Integrity guidelines presented in the UW General Catalog. Further, I certify that I do not have any knowledge of any other individual(s) violating these guidelines.

INDEX OF CONTENTS

INDEX OF TABLES

1 ANALYSIS ON US ACCIDENTS (2020)



1.1 Executive Summary

Reducing traffic accidents is an essential public safety challenge all over the world; therefore, accident analysis has been a subject of much research in recent decades. The objective of the project is to perform exploratory and diagnostic analysis of the changes in year 2020 to analyze the conterminous US accident data from 49 states (based on a data set obtained from Kaggle, US Accidents), and to inform the US government agencies possible causes of traffic accidents since Covid-19 hit, specifically for year 2020 and what could be done to reduce them.

During this time, the pandemic due to the coronavirus (SARS –CoV-2) was a threat to the health and welfare of all people residing in US. Government issued several executive orders that would help to slow the spread of the virus, and its associated health impacts now commonly known as COVID-19. The analysis include number of accidents by year, number of accidents by state, best time to travel by month, day and hour, accident-prone area in each state, factors responsible of the accidents like weather, wind flow, temperature, location, etc.

1.2 Learning Objectives

The following are the main learning objectives for this assignment file:

Purpose of below questions: Tableau Functionalities and applications

1. **Lollipop Chart using custom icons and shapes in Tableau?**
   To visually represent the data just by figures, sometimes makes the insights drawn to be fade away, to engage audience by visuals, the data can be made interesting without losing the attention to detail.

2. **Hex map functionality of Tableau ?**
   To make the interface more interactive, so that the statistical facts of the data can become handy, Hex map is very useful such as in this case when the number of states- variable has 48 unique values to be included in dashboard as a filter option.

3. **Maps:** How to depict via only visualizations plotted which zip codes are having enormous number of accidents in which state?

➔ Above stated learning objectives will be demonstrated with click through solutions for the listed scenarios above with the following 4 Types taken into consideration:

(a) Type 1: to practice- upload data, create visualizations, etc.

(b) Type 2: to practice- dashboard

(c) Type 3: to practice- storyboard

(d) Type 4: Different functionalities of Tableau (Choice Selected)

Note- Using different charts for different type of data is a niche skill which helps in delivering stories in a more succinct and impactful way. Remembering a story is easier than remembering few metrics of the dashboard.

1.2.1 Data Dictionary

There are 47 attributes along with the field description as shown in below figure.

#	Attribute	Description
1	ID	This is a unique identifier of the accident record.
2	Severity	Shows the severity of the accident, a number between 1 and 4, where 1 indicates the least impact on traffic (i.e., short delay as a result of the accident) and 4 indicates a significant impact on traffic (i.e., long delay).
3	Start_Time	Shows start time of the accident in local time zone.
4	End_Time	Shows end time of the accident in local time zone. End time here refers to when the impact of accident on traffic flow was dismissed.
5	Start_Lat	Shows latitude in GPS coordinate of the start point.
6	Start_Lng	Shows longitude in GPS coordinate of the start point.
7	End_Lat	Shows latitude in GPS coordinate of the end point.
8	End_Lng	Shows longitude in GPS coordinate of the end point.
9	Distance(mi)	The length of the road extent affected by the accident.
10	Description	Shows natural language description of the accident.
11	Number	Shows the street number in address field.
12	Street	Shows the street name in address field.
13	Side	Shows the relative side of the street (Right/Left) in address field.
14	City	Shows the city in address field.
15	County	Shows the county in address field.
16	State	Shows the state in address field.
17	Zipcode	Shows the zip code in address field.
18	Country	Shows the country in address field.
19	Timezone	Shows time zone based on the location of the accident (eastern, central, etc.).
20	Airport_Code	Denotes an airport-based weather station which is the closest one to location of the accident.
21	Weather_Timestamp	Shows the timestamp of weather observation record (in local time).
22	Temperature(F)	Shows the temperature (in Fahrenheit).
23	Wind_Chill(F)	Shows the wind chill (in Fahrenheit).
24	Humidity(%)	Shows the humidity (in percentage).
25	Pressure(in)	Shows the air pressure (in inches).
26	Visibility(mi)	Shows visibility (in miles).
27	Wind_Direction	Shows wind direction.

28	Wind_Speed(mph)	Shows wind speed (in miles per hour).
29	Precipitation(in)	Shows precipitation amount in inches if there is any.
30	Weather_Condition	Shows the weather condition (rain, snow, thunderstorm, fog, etc.)
31	Amenity	A POI annotation which indicates presence of amenity in a nearby location.
32	Bump	A POI annotation which indicates presence of speed bump or hump in a nearby location.
33	Crossing	A POI annotation which indicates presence of crossing in a nearby location.
34	Give_Way	A POI annotation which indicates presence of give_way in a nearby location.
35	Junction	A POI annotation which indicates presence of junction in a nearby location.
36	No_Exit	A POI annotation which indicates presence of no_exit in a nearby location.
37	Railway	A POI annotation which indicates presence of railway in a nearby location.
38	Roundabout	A POI annotation which indicates presence of roundabout in a nearby location.
39	Station	A POI annotation which indicates presence of station in a nearby location.
40	Stop	A POI annotation which indicates presence of stop in a nearby location.
41	Traffic_Calming	A POI annotation which indicates presence of traffic_calming in a nearby location.
42	Traffic_Signal	A POI annotation which indicates presence of traffic_signal in a nearby loction.
43	Turning_Loop	A POI annotation which indicates presence of turning_loop in a nearby location.
44	Sunrise_Sunset	Shows the period of day (i.e., day or night) based on sunrise/sunset.
45	Civil_Twilight	Shows the period of day (i.e., day or night) based on civil twilight.
46	Nautical_Twilight	Shows the period of day (i.e., day or night) based on nautical twilight.
47	Astronomical_Twilight	Shows the period of day (i.e., day or night) based on astronomical twilight.

Table 1-1. Data Dictionary Table for US-Accidents Dataset

1.3 Deliverables

Link for Tableau Public Storyboard and Dashboard:

https://public.tableau.com/views/USAccidentsAnalysis2020/Storyboard?:language=en-US&publish=yes&:display_count=n&:origin=viz_share_link

1.3.1 Analysis & Visualizations

Variables	Significance
Pressure (in)	12.20%
Weather_Conditions	10.50%
Humidity (%)	10.34%
Temperature (F)	3.55%
Day	9.18%
Traffic_signal	5.67%
Hours	8.03%
Wind_Speed	3.64%
Stop	4.23%
Junction	4.09%
Month	4.79%
Crossing	3.39%
Visibility(mi)	2.31%

Table 1-2 List of Significant attributes

Above list of attributes are used in tutorial development as these are the identified KPIs for the solution to problem analysis of US Accidents increasing count over the year 2020.

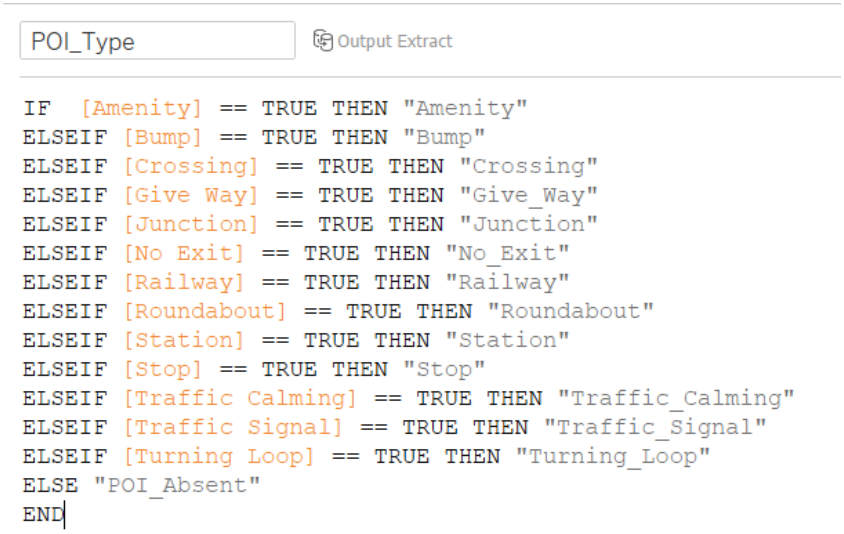
1.3.2 Lollipop Chart

Introduction: Best practices for creating a lollipop chart in Tableau.

- Sorting the lollipop chart makes it more interpretable.
- Use color and gridlines with caution to avoid clutter.
- Always label the axis clearly.
- Provide additional details on the tooltip.

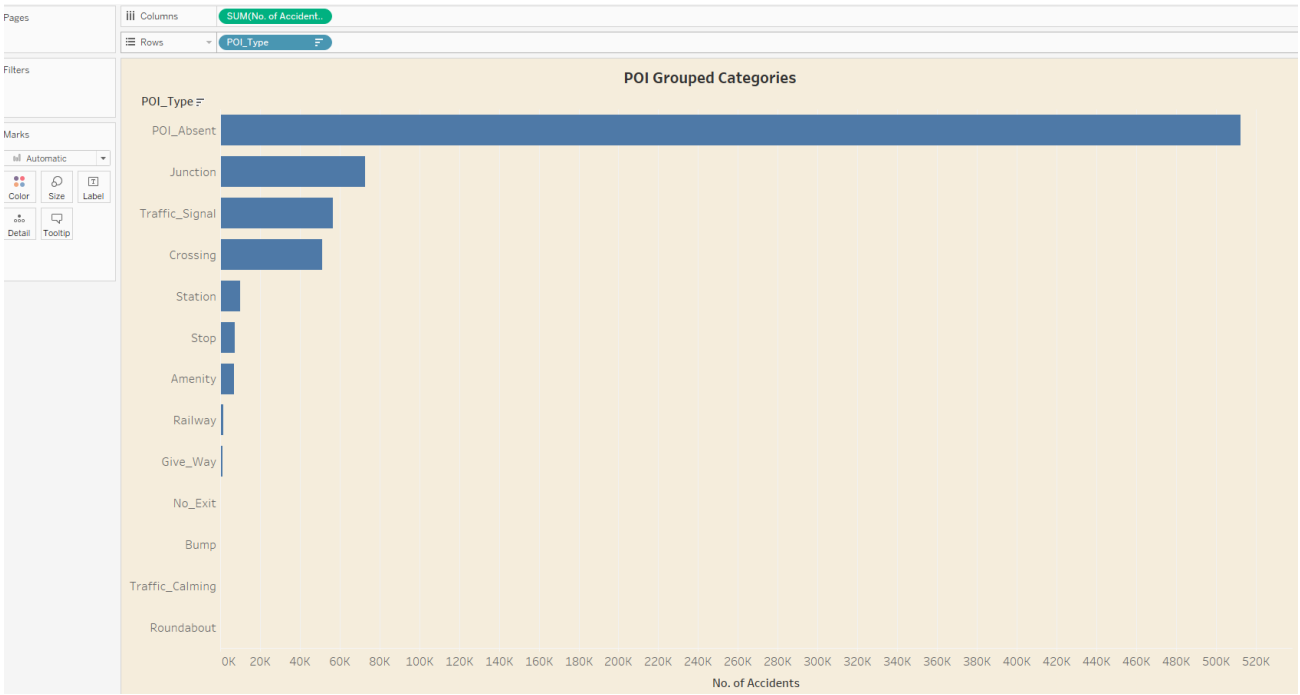
Q1. How to make Lollipop chart in Tableau using custom icons and shapes in Tableau?

The visualization represented for demonstrating Lollipop chart embeds the business scenario to represent the accident count distribution against the total number of accidents in the US.
To extract the associated findings by lollipop chart, the category types were grouped together via creating a calculated field: "POI_Type" as shown in the below screenshot.



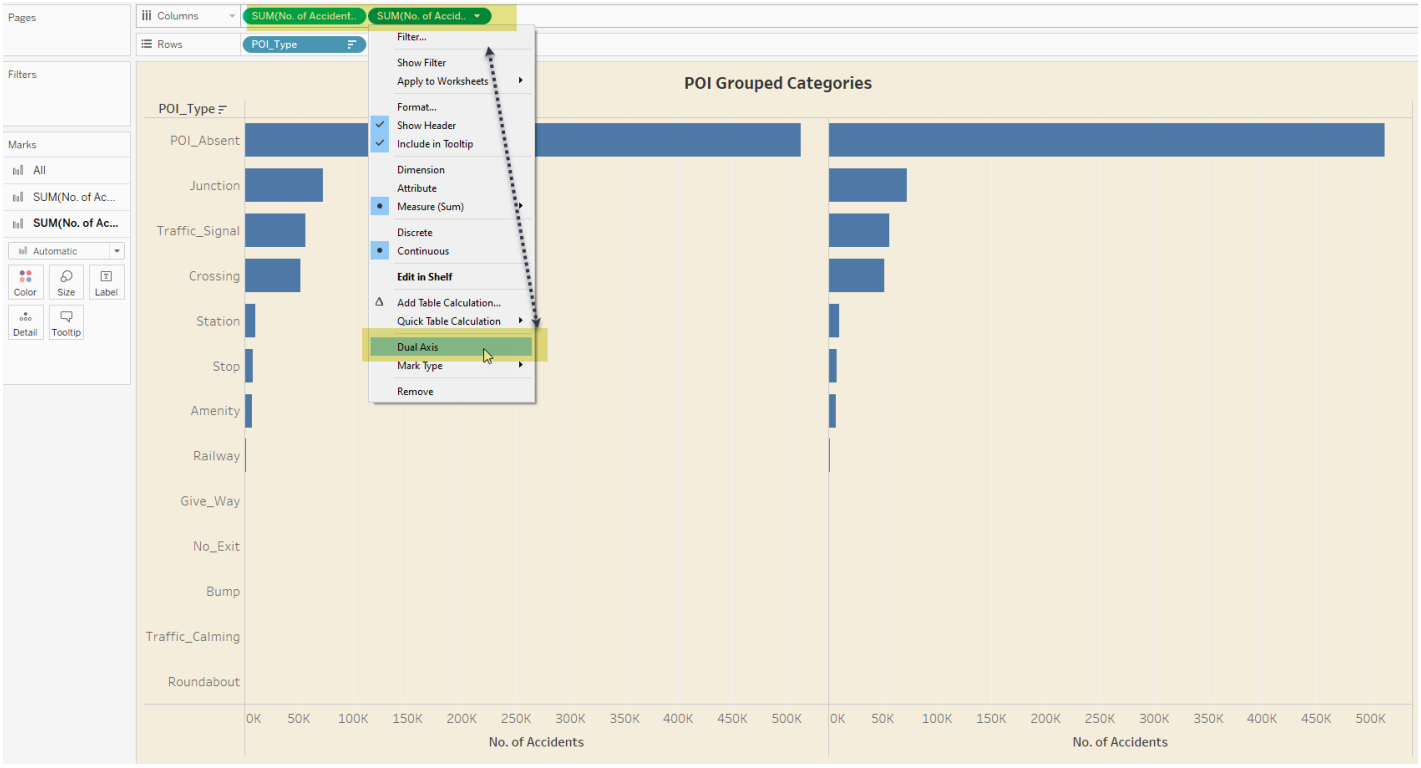
Step 1: Build a simple bar chart

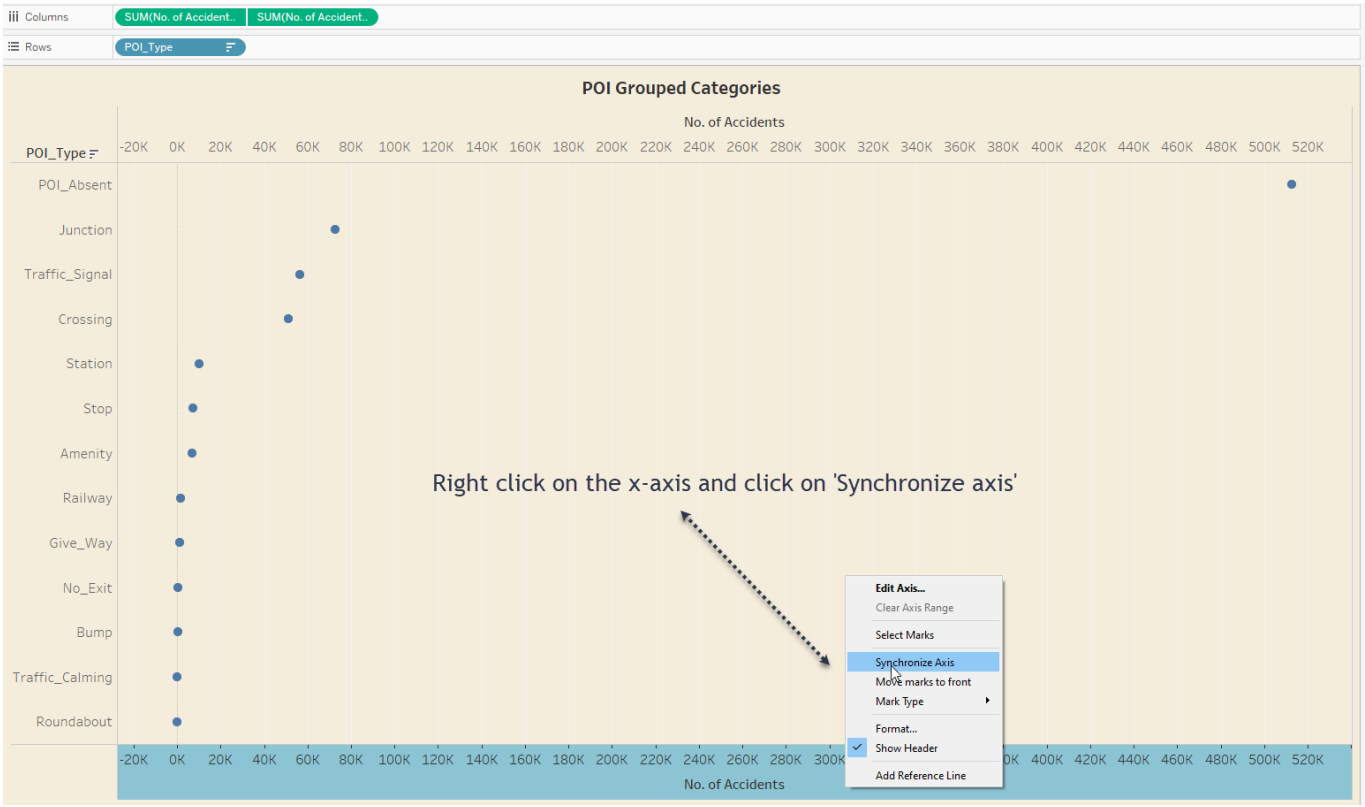
➔ Drag POI_Type into rows and no. of accidents into columns and sort POI_Type in descending order by field.
➔ Select 'Bar' under marks card.
The result should look like below:



Step 2: Create a dual bar chart (duplicate the view)

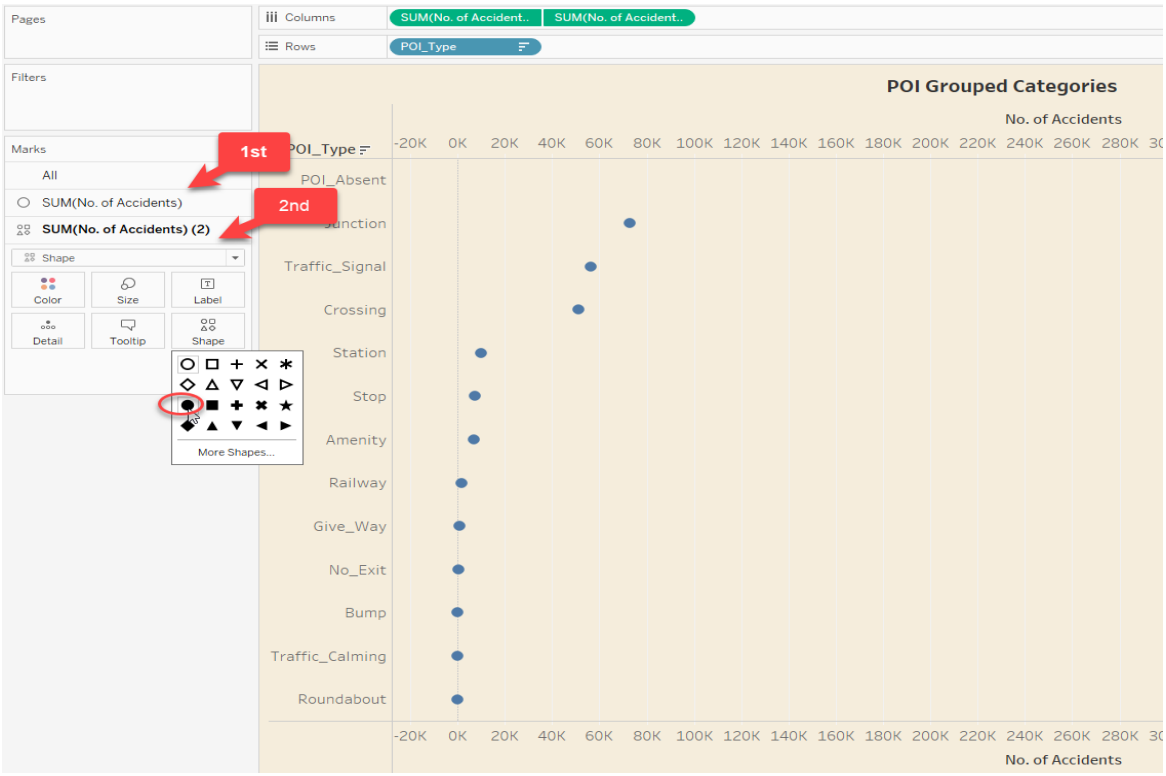
- ➔ Drag the measure 'no. of accidents' to the columns shelf next to the aggregate SUM(No. of Accidents)
- ➔ Make the cards dual and synchronize the axis.





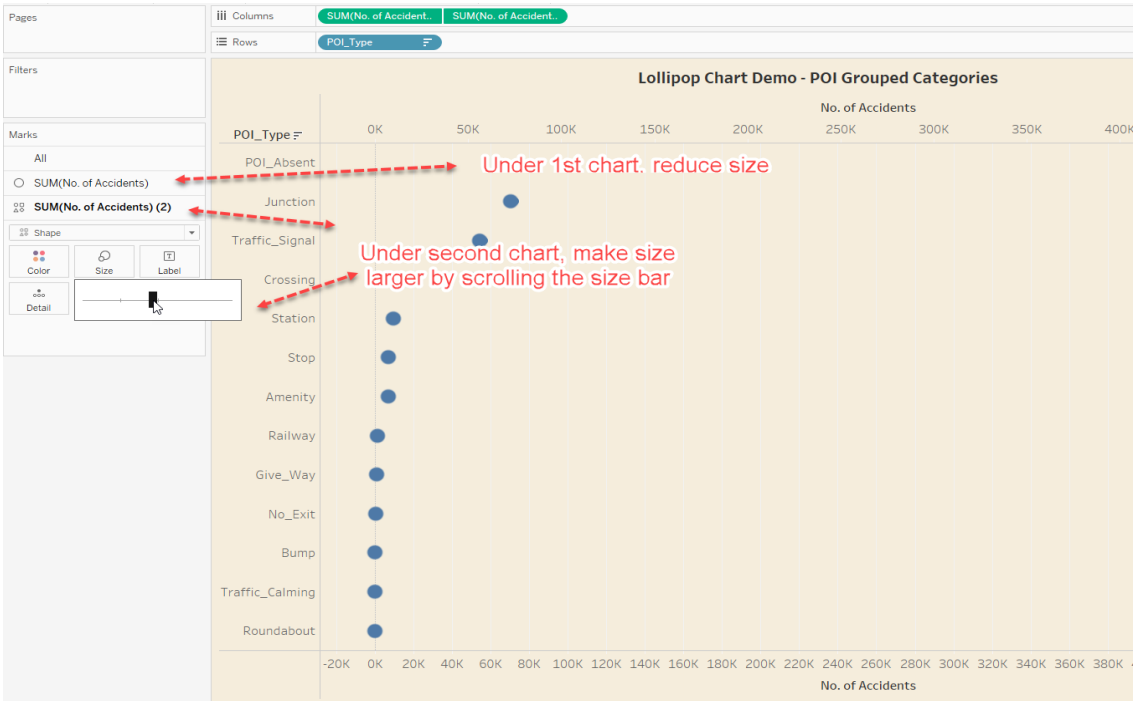
Step 3: Change the second bar char to a circle

- Select the second bar chart and change it to '**Shape**' under marks card.
- Choose circle on the '**Shape**' tab.

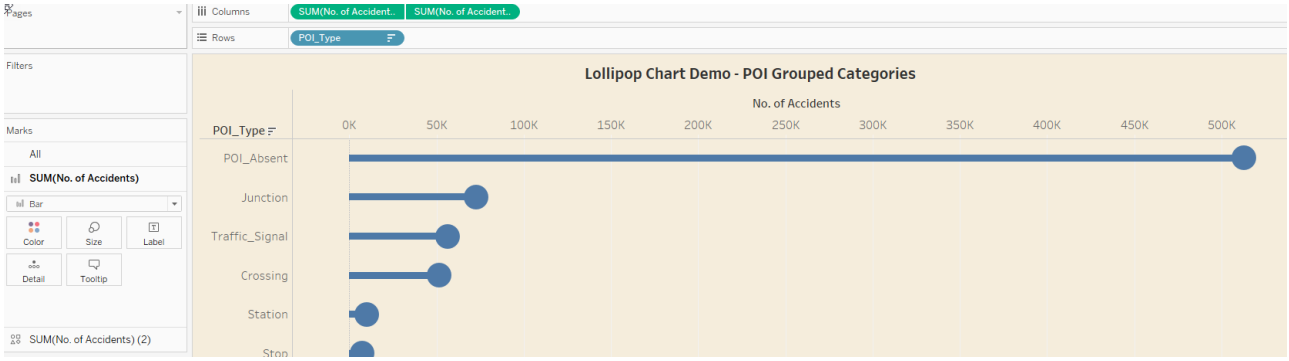


Step 4: Adjust the two charts as follows:

- Make the bar chart (1st chart) thinner and the circles (2nd chart) larger by adjusting their respective sizes under **Size** tab.

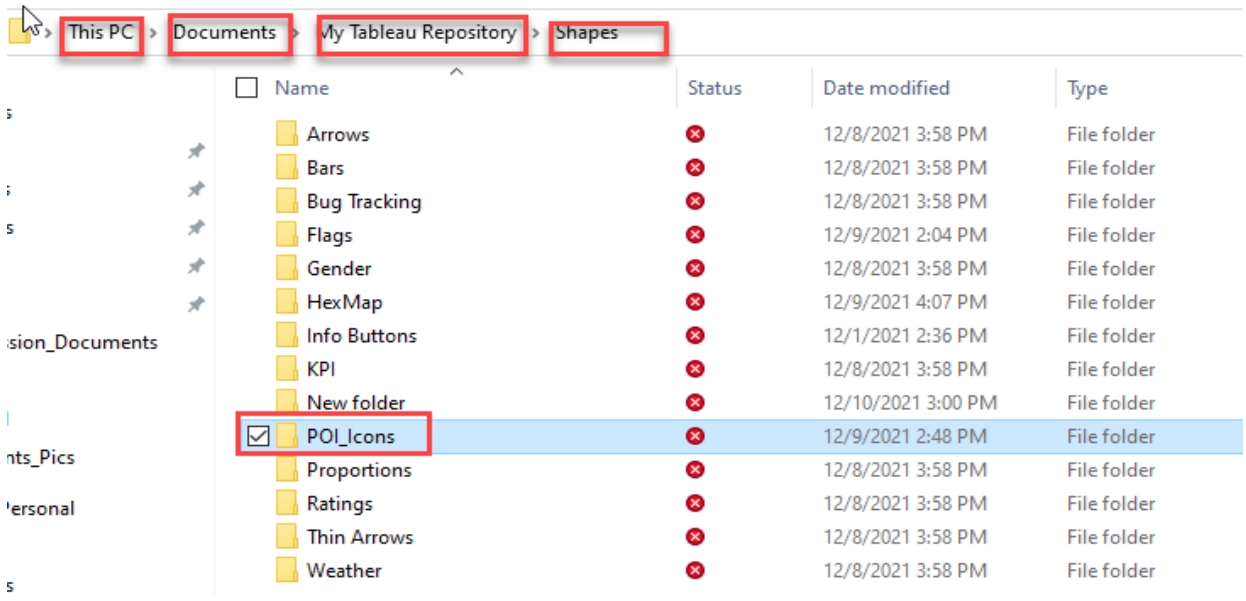


Executing this gives us Lollipop chart below:

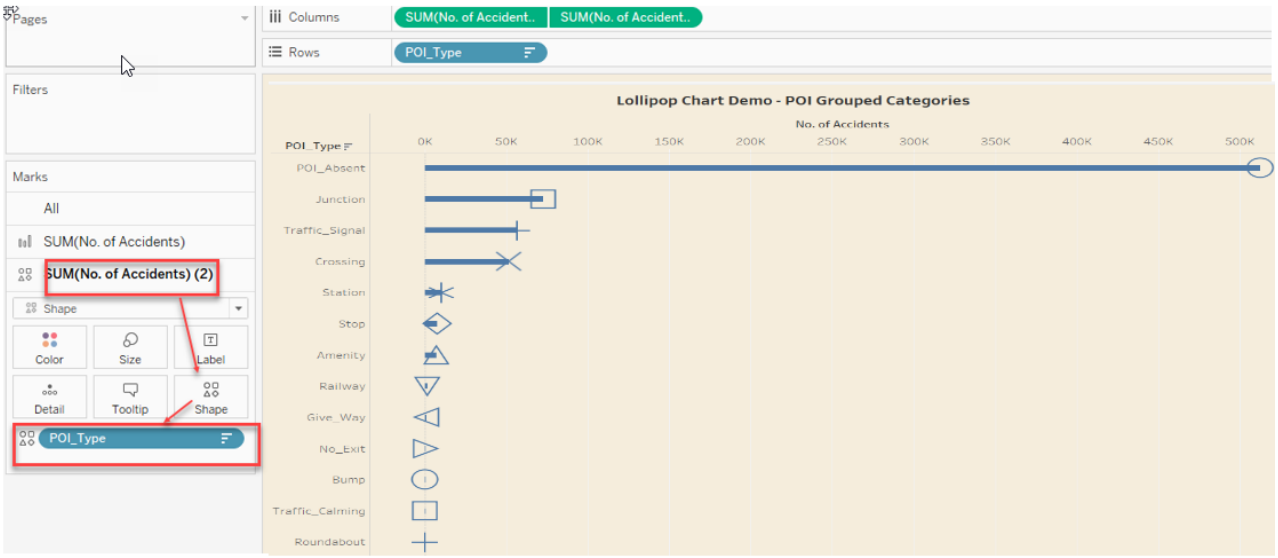


Step 5: Download the shapes which are related to the data (POI_Types) shown in the bar chart axis. Icons can be downloaded via- flaticon.com

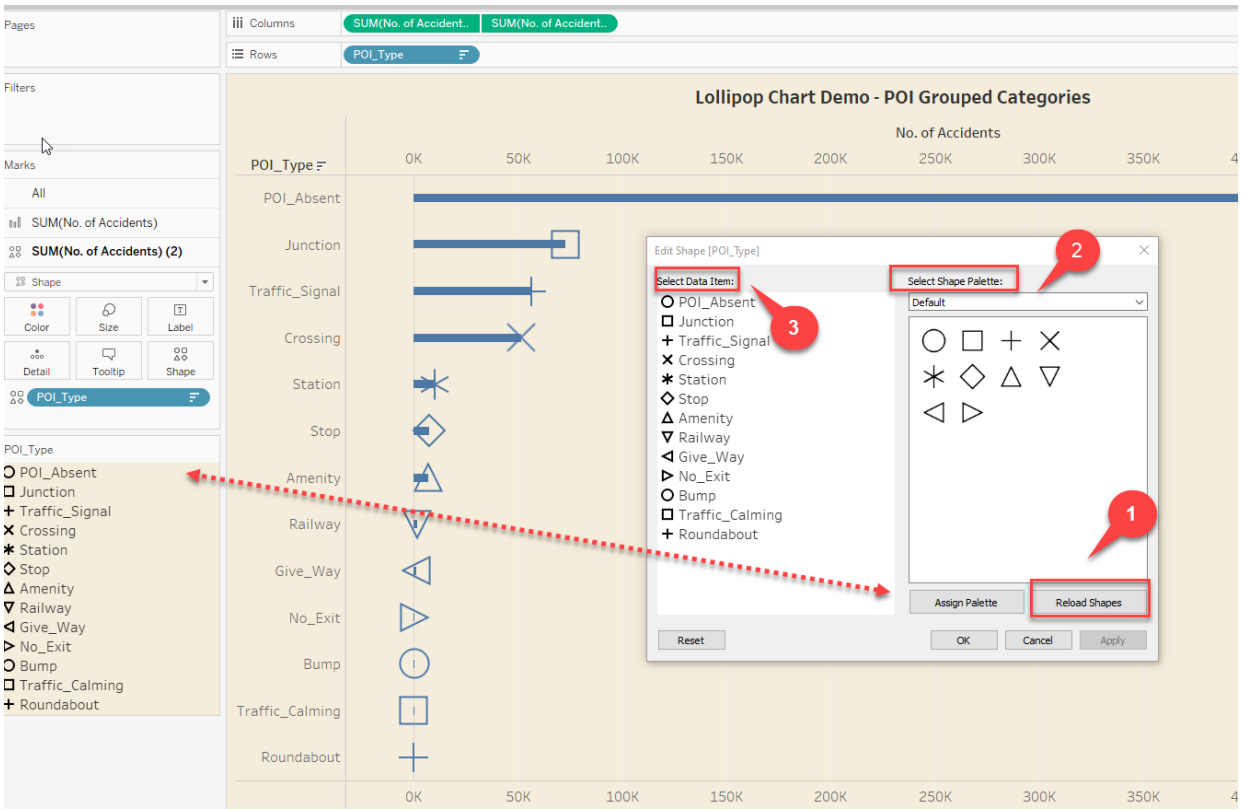
Step 6: Load icons downloaded to "Shapes" in Tableau repository creating a new folder for 'POI_Types'.



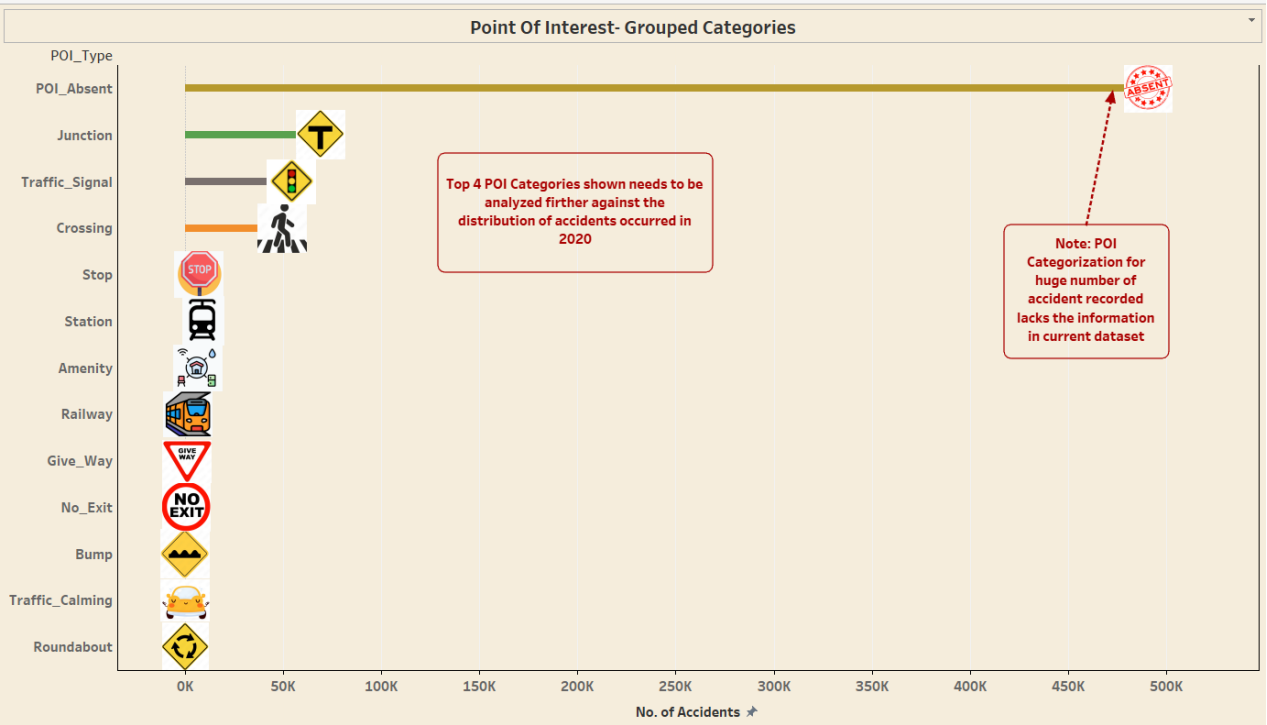
Step 7: Add POI_Type into 2nd marks card Shape pallet as shown below to use flag shapes in our viz.



From the legends displayed in right corner, change shape icons by reloading shapes into the shape palette:



Step 8: Match the icons downloaded from selected palette and assign to their corresponding POI_Type , executing this will give the following results representing a lollipop chart with all the grouped POI_Types.



Insights drawn:

- ➔ Among all the 13 POI types, the topmost significant attributes for analysis purpose are determined which are Junction, Traffic Signal. Crossing and Stop sign.
- ➔ The findings are carried out in further stages of analysis to view how the incident counts are increasing for accidents happened in US around these 4 POIS_Types.

1.3.3 Hex Map

How to make Hex map in Tableau ?
Hex maps are great way of visually representing filter values more than 20. As in this scenario, demo will be carried out to answer the business question:

How are 4 POI_types affecting the accident counts per state?

As in the dataset, there are 49 states to be precise for filtering across POI_Types, let's start!
Prerequisites:

1. Data source (US Accident dataset)
2. Hex Map Data Source (csv file attached below)
3. Save the hexagon image in the shape folder of tableau repository.

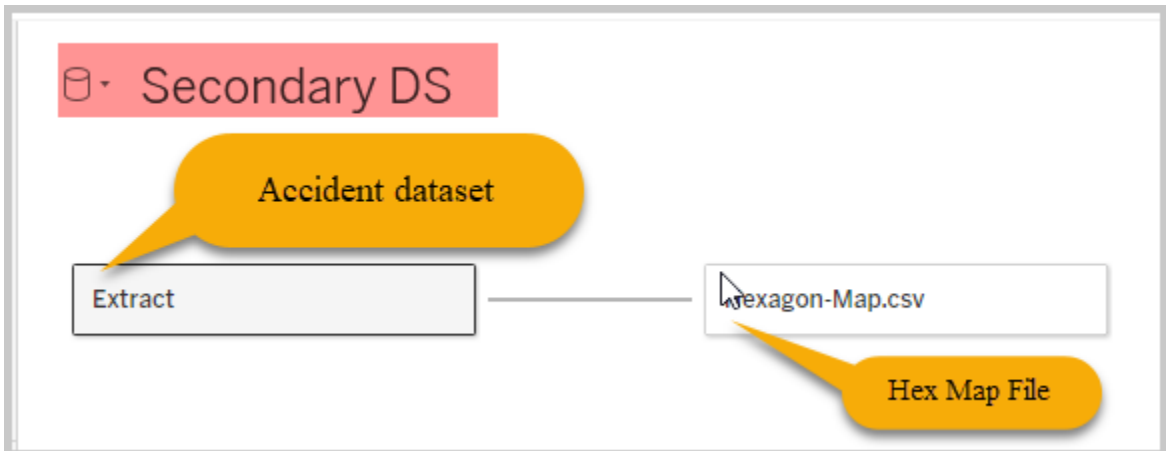
hexagon.png

Hexagon-Map.csv

Note: *This will contain the X and Y locations for each individual state.*

Steps to be follow along to create Hex Map are defined below:

S1. To avoid the mismatch of applicable filters for this hex map, it's better to create a separate data source and create a relationship for the Hexagon-Map.csv. The secondary data source is shown below.



S2. Our dataset contains all the needed dimensions and measure now.

➔ Drag 'Severity' dimension to the color marks card.

➔ From hexagon-map table:

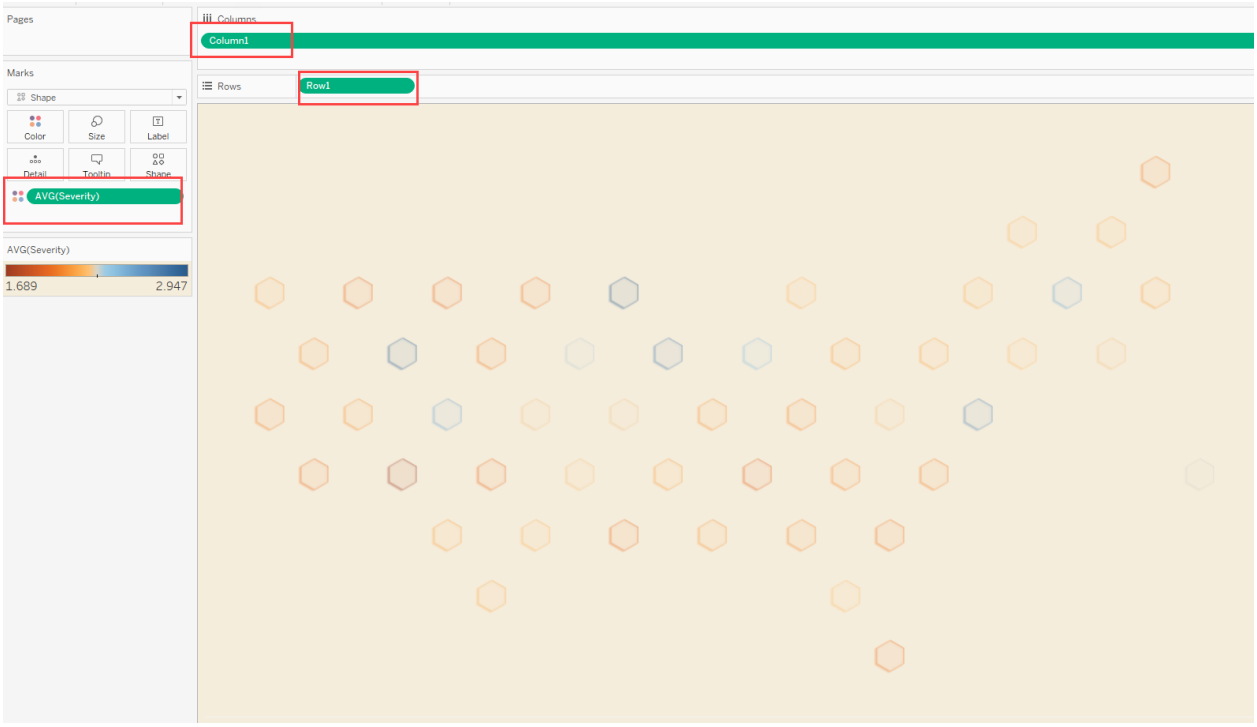
- Drag **Column** onto the **Columns Shelf**; right-click on this Pill and convert it to a Dimension
- Drag **Row** onto the **Rows Shelf**; right-click on this Pill and convert it to a Dimension
- Drag **Code** onto the **Label Mark**

Optional: If the reader is using 2 separate data sources without establishing relationship among both the files, then, on the Hexagon-Map user should see a link next to State, click on the 'Link' icon to set this as a blend condition.

S3: Change Mark Type to shape

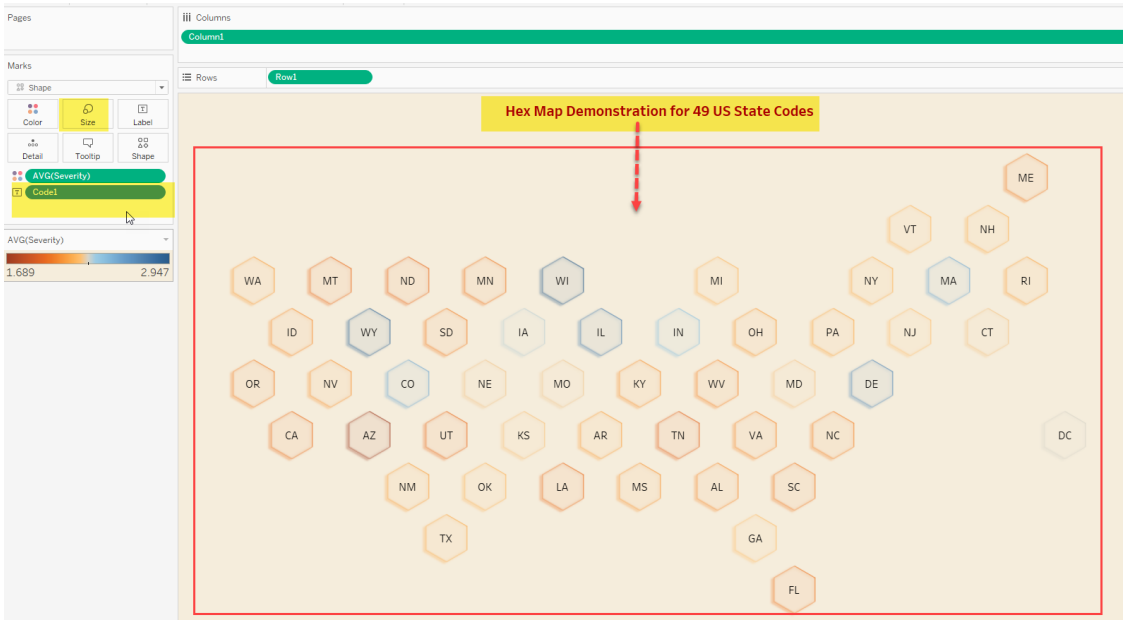
➔ Click on the **Shape Mark** and select **More Shape** and apply "Hexagon" shape image which we saved.

The sheet should appear as below screenshot:

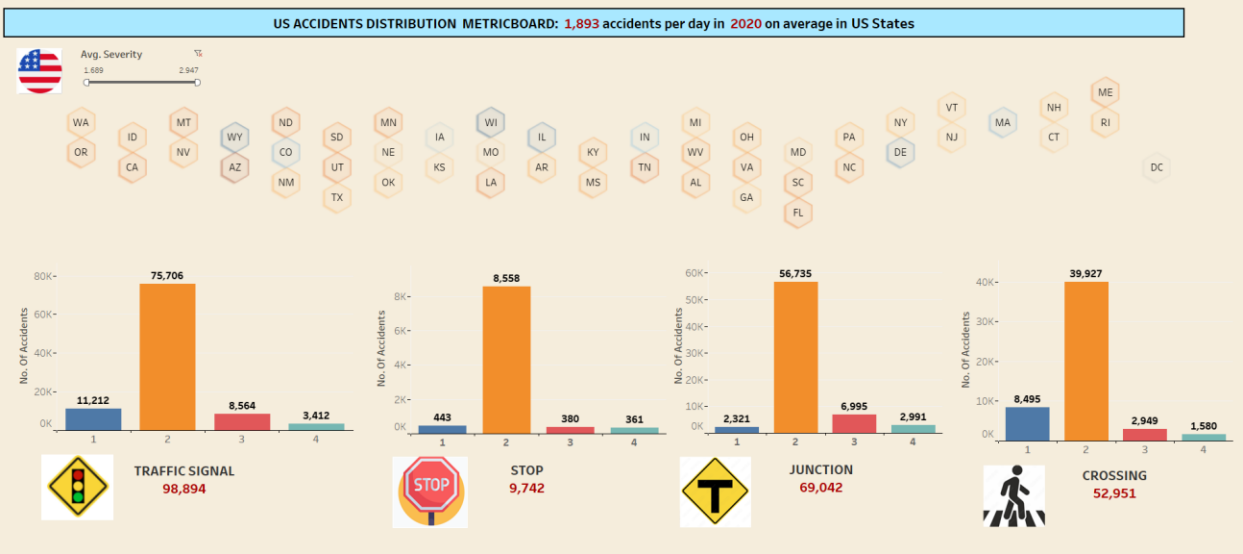


S4: Choose appropriate colors asper the choice and add ‘Code’ from Hexagon-Map data source to Label. Adjust the shape to avoid overlapping of Hexagon tiles shown over the sheet.

Final output should look like below:



Executing this with global filters and actions for answering the business question will result into the final dashboard shown below:

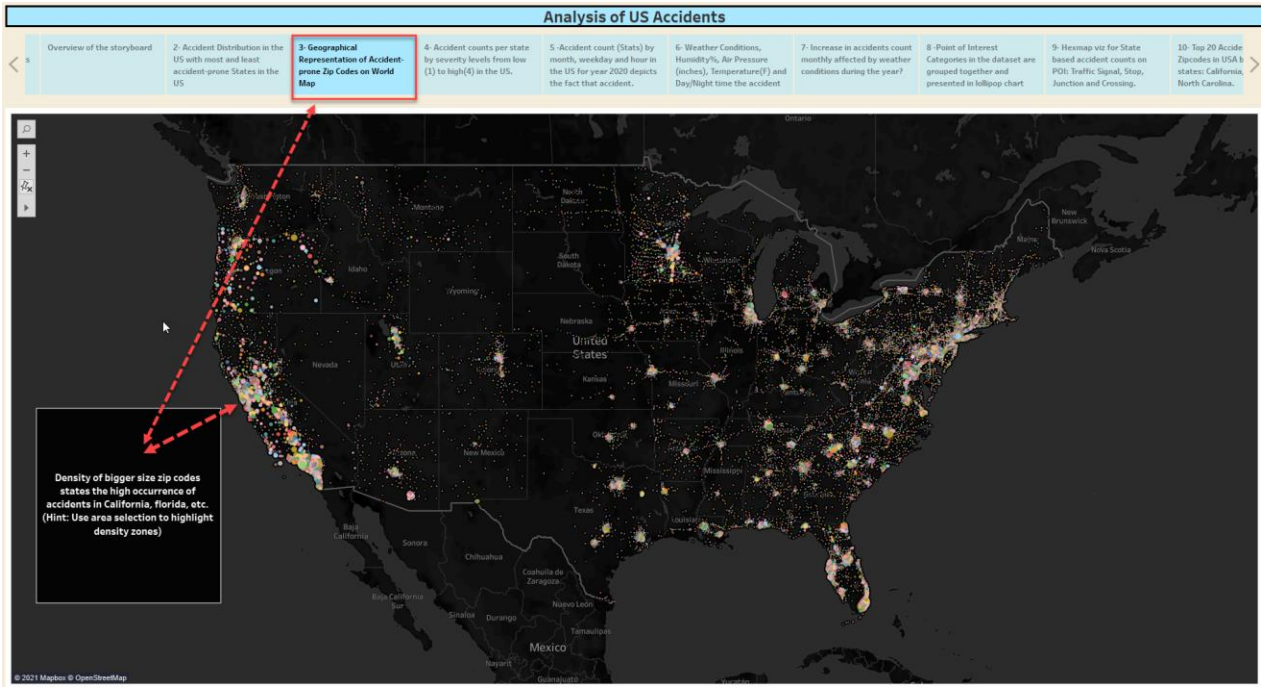


Dashboard Click through Steps:

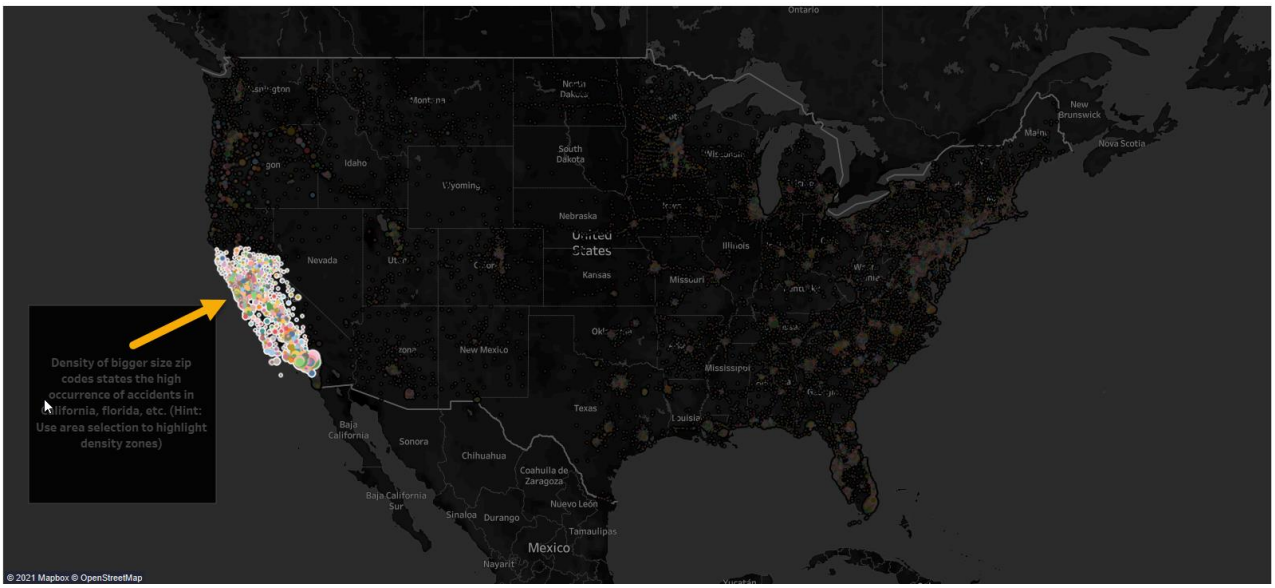
1. Select the state by clicking on the hexagon tiles for specific state you wish to analyze the metrics for.
2. Adjust the severity (optional) by using the slider shown in the top left most corner of the dashboard.
3. Execution of this exploration updates the figures on the bar chart for POI Types.
4. As observed, highest counts of accidents are from severity – 2 causing short delay in the subsequent traffic causing congestions on the road.
5. By default, the stats shown in the dashboard are applicable for overall 49 states.
6. On average, number of accidents occurring per day are updated on the dashboard title and changes are updated based on the state selection criteria chosen.
7. Traffic signal and Junction are having the highest count of accidents for year 2020, aggregated states figures.

1.3.4 MAPS used for demographic representations

Demonstration: How to depict via only visualizations plotted which zip codes are having enormous number of accidents in which state?



Using the dark map theme created a mesmerizing and impactful visualization of zip code-prone accident zones in the US map. The caption is part of the Area annotation and selected area will be highlighted containing only those zip codes in the map, the viz should look something like this:



Following are the steps to follow:

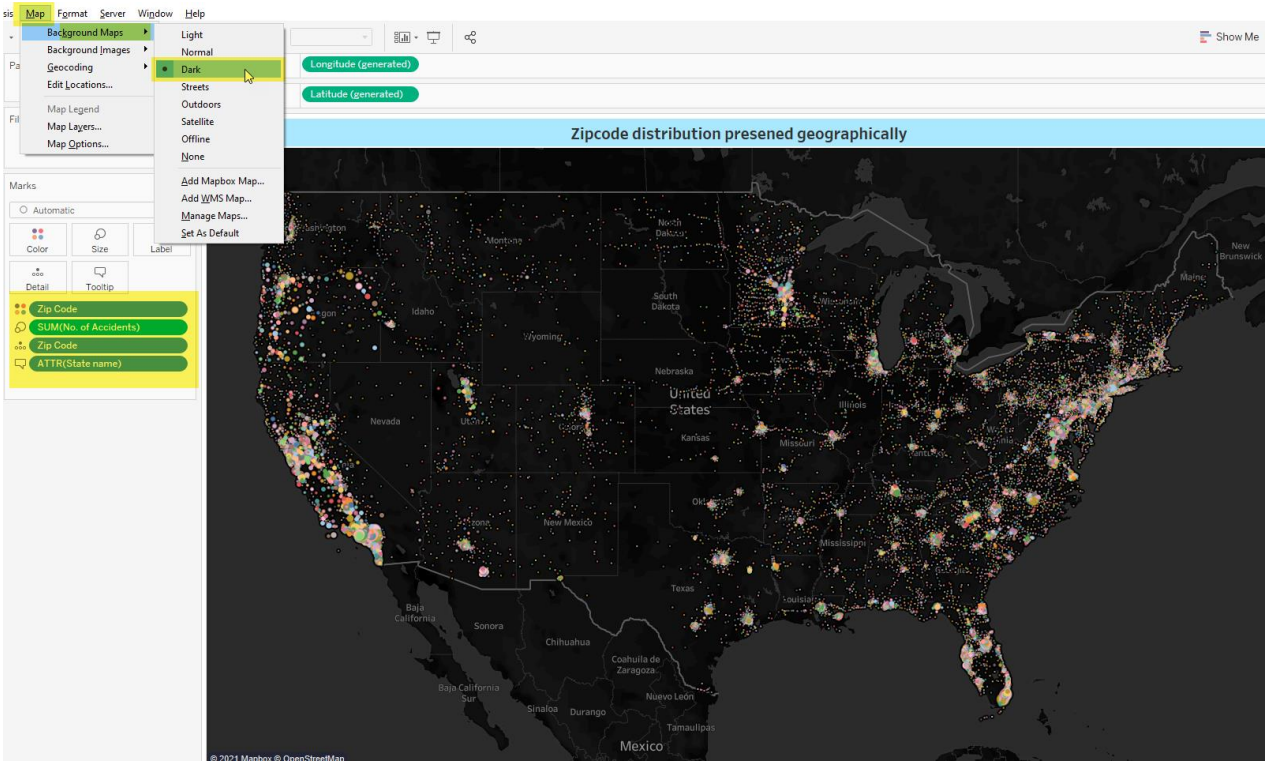
1. Double click on State and the row and column shelve will be auto populated with latitude and longitude generated.

2. Add zip codes into colour marks card and drop on to Detail section of the marks card as well.

3. Add measure “No. of accidents” on to size section on the marks card.

4. Drag and drop State name in the tool tip so that hovering should show the details of the states including the number of accidents, zip code and state name.

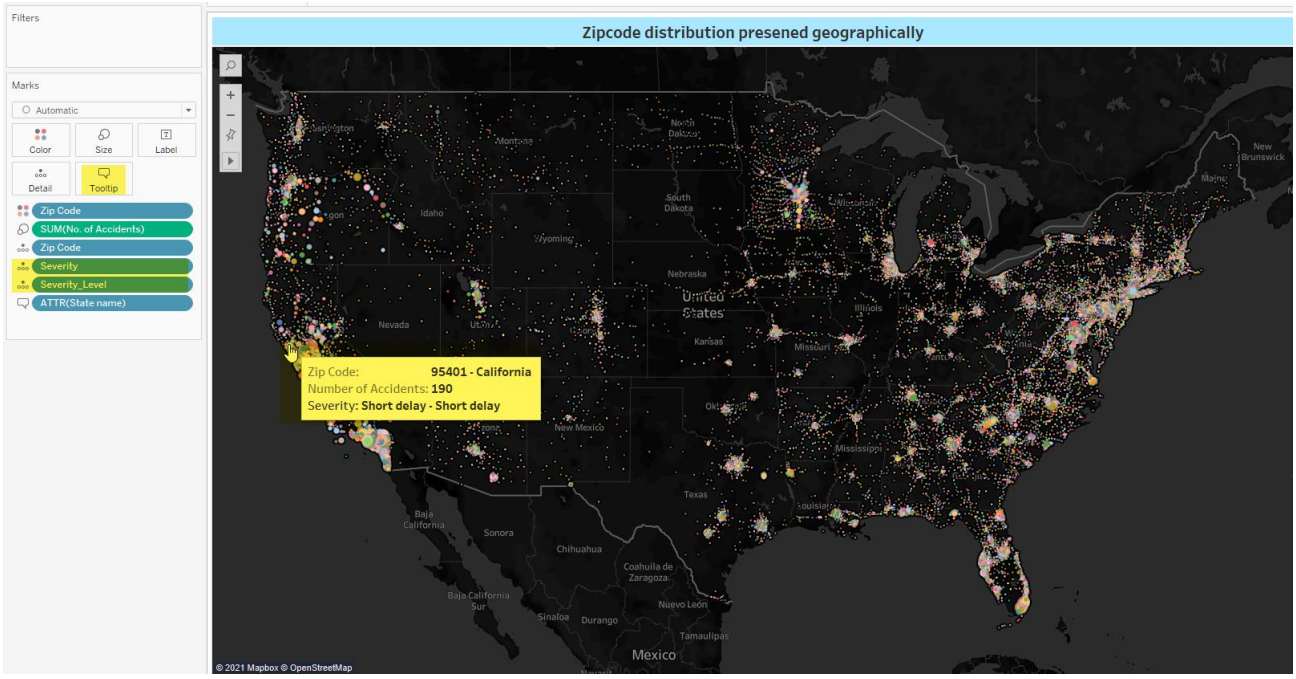
5. Last but not the least, change the map settings by choosing background map to be “Dark”.



➔ **How to read the map and make it more informative?**

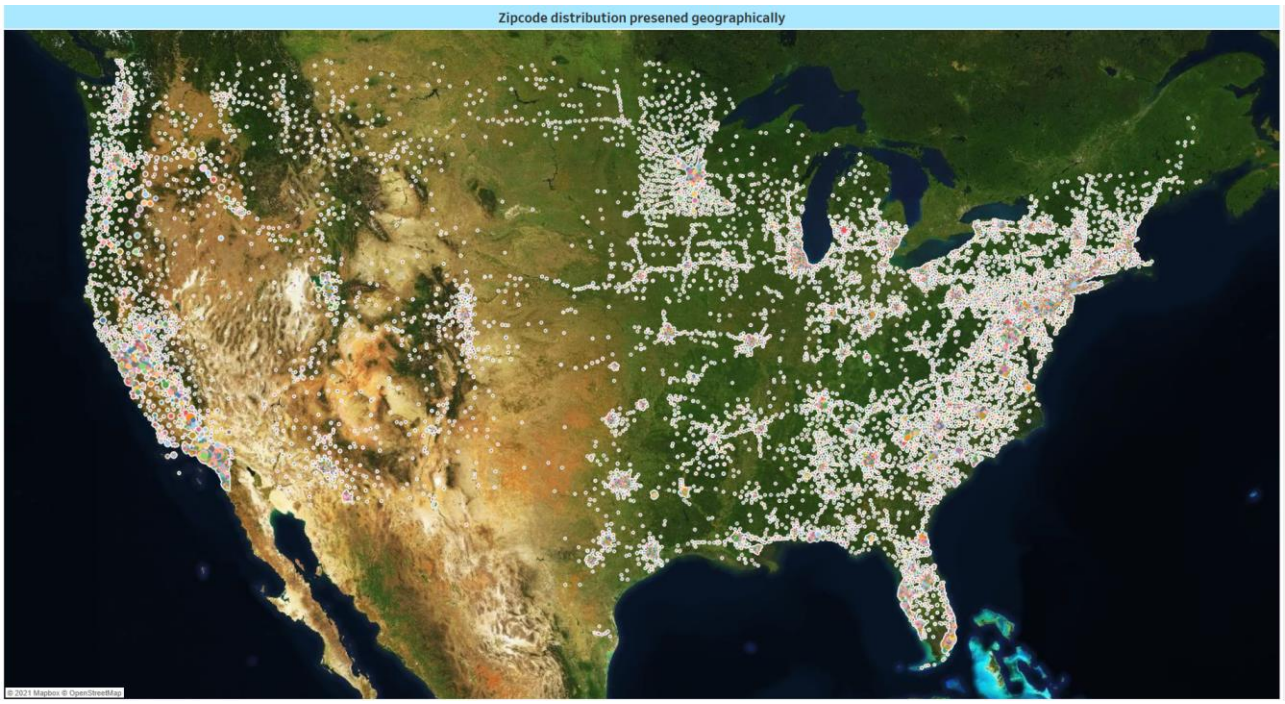
Hovering on different zip codes will tell the number of accidents happening by severity if below settings are made to the tool tip as well by adding severity and severity-level description on to the details card.

This is a top-notch visualization to quickly check out which are the states with highest density of zip code involved in accidents as size of the zip code points are bigger if there are enormous number of accidents happening over there.



Variation

There are other map backgrounds also which give a look of street view, satellite view, offline view etc. Below image is demonstration of use of **Satellite maps** with same applied settings for dimensions and measures.



1.4 Limitations

Due to information constraint, part of the original dataset was removed which consisted of 3million records in the dataset for US Car Accidents (2019). After removing the sensitive information and attributes, updated dataset "US Accidents (Updated 2020)" was created and shared on Kaggle for academic and research purposes.

**Disclaimer: It is a valid consideration to keep in mind that it is possible to find false trends because the data could actually be missing information on some crashes that did not end up being reported or recorded.*

1.4.1 Recommended Citations/Acknowledgements

- Moosavi, Sobhan, Mohammad Hossein Samavatian, Srinivasan Parthasarathy, and Rajiv Ramnath. "A Countrywide Traffic Accident Dataset.", 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.

1.4.2 Usage Policy and Legal Disclaimer

This dataset is being distributed only for Research purposes, under Creative Commons Attribution-Noncommercial-Share Alike license (CC BY-NC-SA 4.0). By downloading this dataset, user agrees to use this data only for non-commercial, research, or academic applications.

1.5 References

1. Over 100 Car Accident Statistics for 2020 | U.S. and Global (safer-america.com)
2. Accident Risk Prediction based on Heterogeneous Sparse Data: New Dataset and Insights (arxiv.org)
3. Road Safety Facts — Association for Safe International Road Travel (asirt.org)
4. NHTSA | National Highway Traffic Safety Administration
5. Causes and Solutions of Car Accidents - 1117 Words | Report Example (ivypanda.com)
6. Contiguous United States - Wikipedia
7. https://scholarworks.lib.csusb.edu/etd/979
8. [1906.05409] A Countrywide Traffic Accident Dataset (arxiv.org)
9. US-Accidents Dataset | Papers With Code
10. HiRes2-1084x1035.jpg (1084×1035) (sagoomotors.co.ke)
11. 7340cbb107d570f0260bc23746c47925.jpg (5198×4022) (pinimg.com)
12. How to Make a Lollipop Chart in Tableau (rigordatasolutions.com)
13. Najjar, A., Kaneko, S., & Miyanaga, Y. (2017). Combining Satellite Imagery and Open Data to Map Road Safety.
14. Anderson D, Anderson J.(2004) Electric and hybrid cars: A history. McFarland, Jefferson