

US ACCIDENTS (2016 - 2019)

COUNTRYWIDE TRAFFIC

ACCIDENT

Power BI Documentation

University of Colorado, Denver-Business School

Data Visualization

US Accidents (2016 - 2019)

Countrywide Traffic Accident

Urvish A. Patel

University of Colorado, Denver

Business School

Abstract

This document offers an analytical view on road accidents which helps the Department of Transportation, Road safety department, Hospitals & various response teams as they will be prepared for such accidents and will be knowing the exact measures to take beforehand an accident occur. Specifically, this document is a summary of findings through Power BI which tells us statistics of the dataset through various visualizations which answers questions such as which State/Zip-code/County has the highest number of accidents? At what time/day/ do accidents usually occur in the US? What are the factors which causes road accidents? Predictions of accidents in future. This document lays out everything which might be helpful for safety of travelers.

Keywords: Time Series Forecasting, R-Script, Key-influencers, Decomposition Tree, Top Segments, Predictions

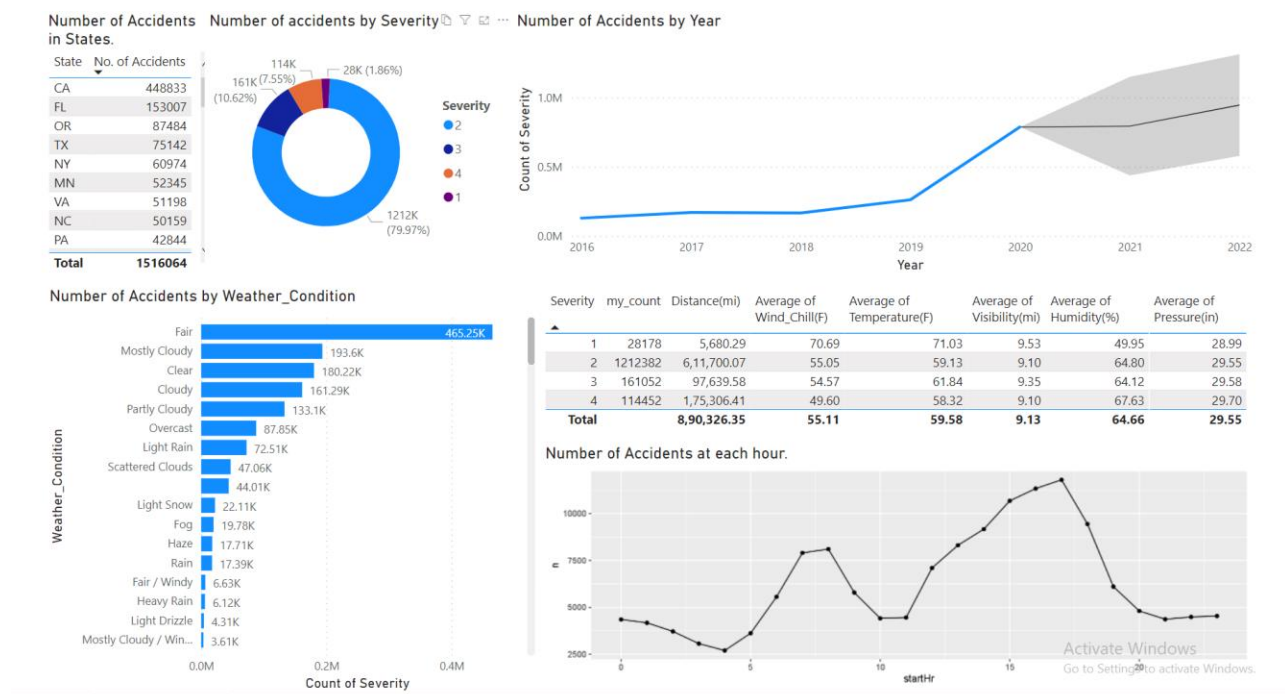
US Accidents (2016 - 2019)

Countrywide Traffic Accident

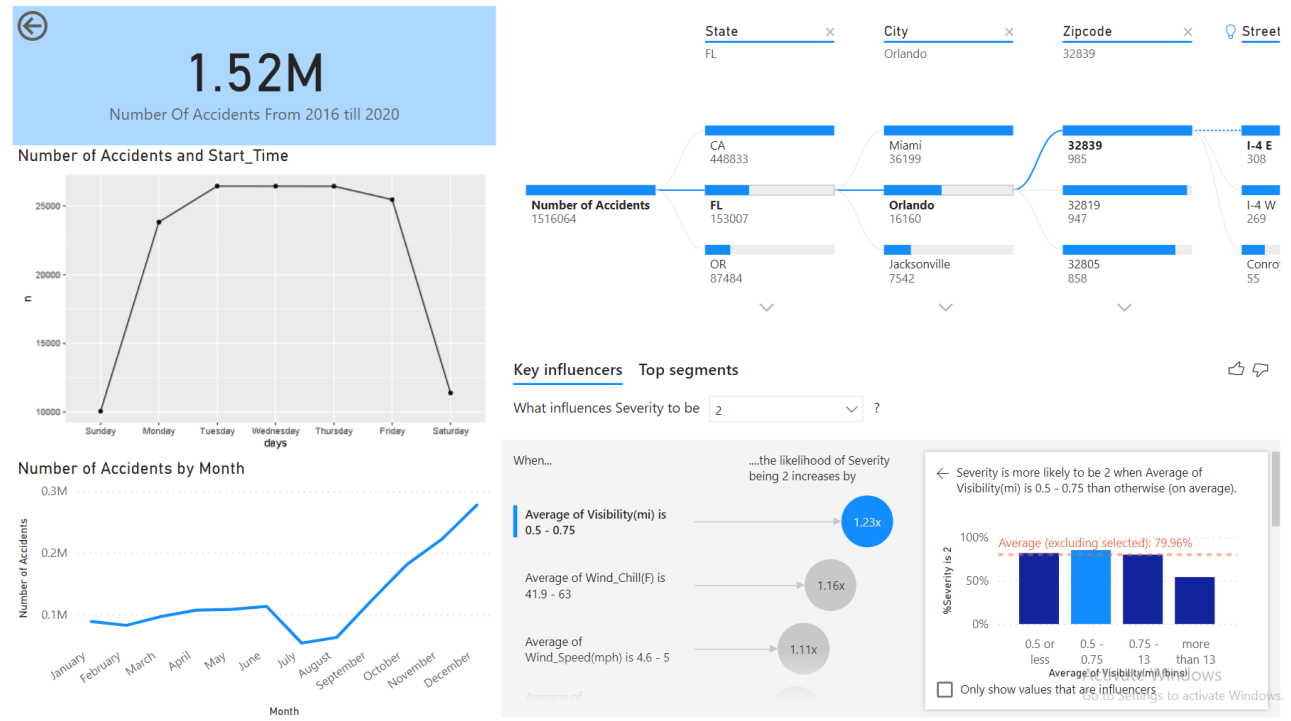
The US Accidents dataset is a data reported by the relevant police department, US and state departments of transportation, law enforcement agencies, traffic cameras, and traffic sensors within the road-networks whenever an accident occurs. The dataset used covers 49 states of the USA. The accident data are collected from February 2016 to Aug 2019. They note down the exact location of accident and the distance of traffic which is affected after an incident occurs. The very first thing which Department of Transportation wants to know is the severity of an accident and how frequently an accident occurs during the day/week/time. How can a response team work efficiently given these circumstances? Weather plays a severe role in any type of accident. This analysis discloses the facts and relations about accidents which might be directly linked to the type of weather condition during an accident.

Power BI Dashboards.

PAGE 1:



PAGE 2:



Visualizations

Visualizations which answers the following questions.

1. What is the count of accidents in each state?

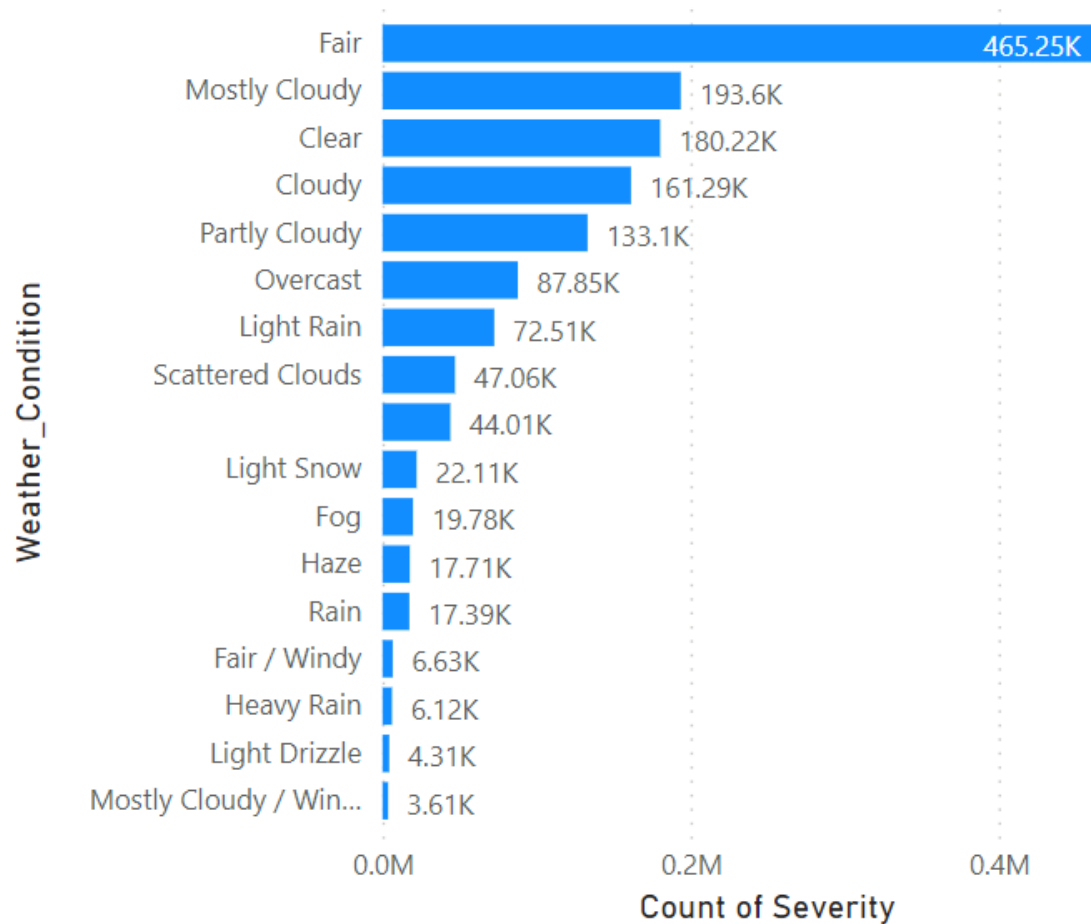
Number of Accidents in States.

State	No. of Accidents
CA	448833
FL	153007
OR	87484
TX	75142
NY	60974
MN	52345
VA	51198
NC	50159
PA	42844
Total	1516064

This table represents the number of accidents happened in each state from highest to lowest. Sometimes a simple table is enough for giving an insight. After all the understanding of data is totally dependent on these numbers.

2. In which Weather Condition do accidents occur most?

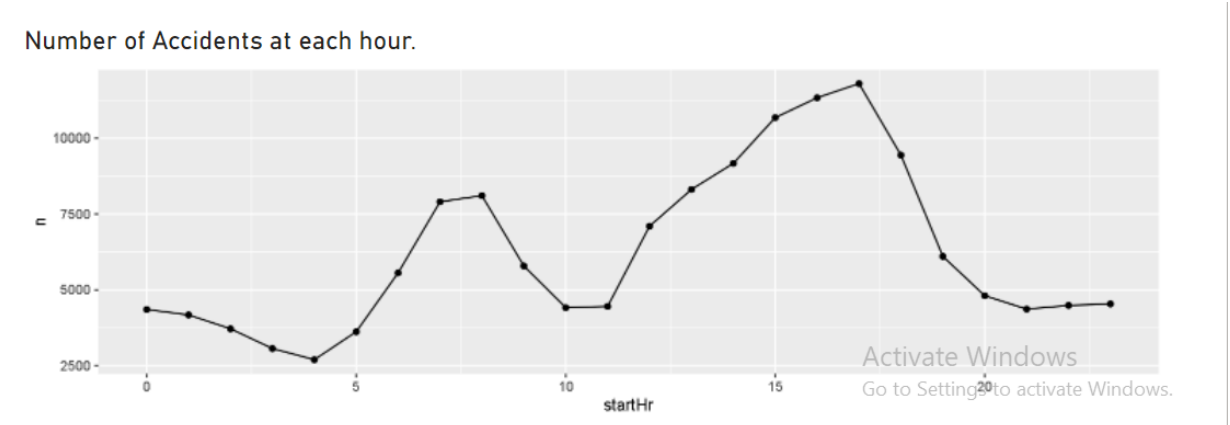
Number of Accidents by Weather_Condition



This is a bar graph which represents that most of the accidents occur when the weather is fair.

Apart from that, accidents also occur when the weather is mostly cloudy, windy and having a light rain. This is a very simple way to represent quantity of a categorical columns of a dataset.

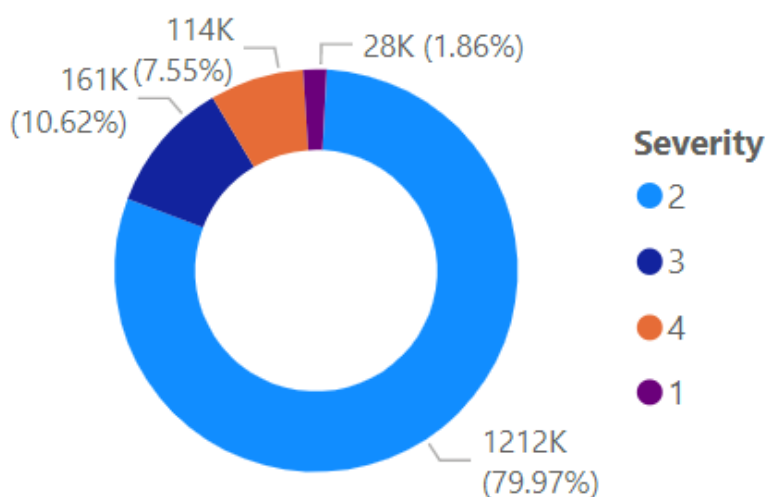
3. Usually at what hour most accidents occur in US?



According to this plot we can conclude that most accidents occur between 15 hrs. To 17 hrs. This visualization is generated through R script. Now I had to use R script because it was the easy way and as power BI did not provided me ease to build this particular visualization. The column of Start_time which had both date and time did not generated time hierarchy, but tableau does this with ease.

4. What type of severity accidents mostly occur?

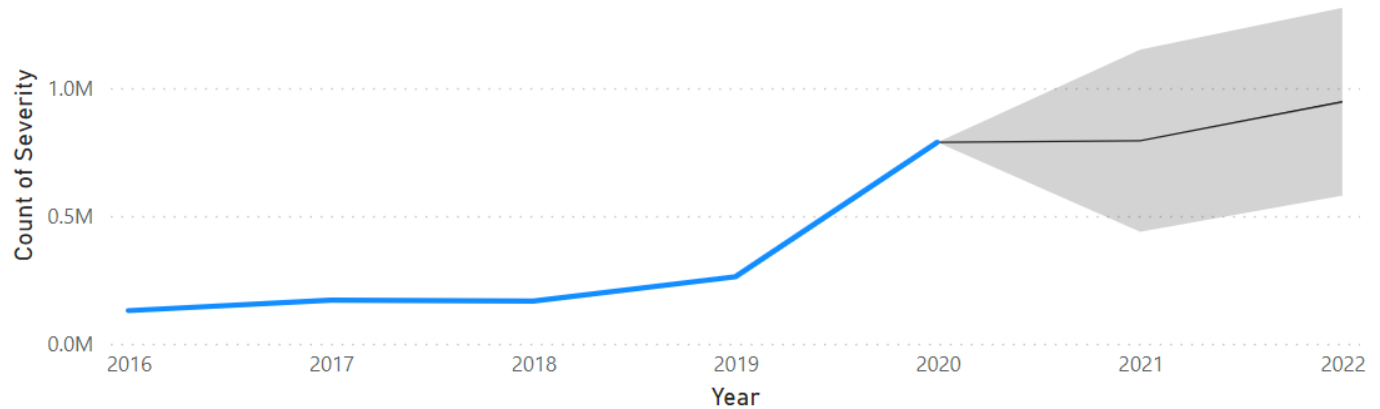
Number of accidents by Severity



This is a pie chart which represents the number of accidents by Severity. We can conclude from this visual that, Severity of type 2 has the highest number of followed by severity 4, 3, & 1.

5. What are the predictions for the future? What number of accidents we can expect?

Number of Accidents by Year



This is a time series forecasting graph. From this we can say that the expected number of accidents in future is likely to increase according to the trend. This prediction is generated not by year but by considering month wise.

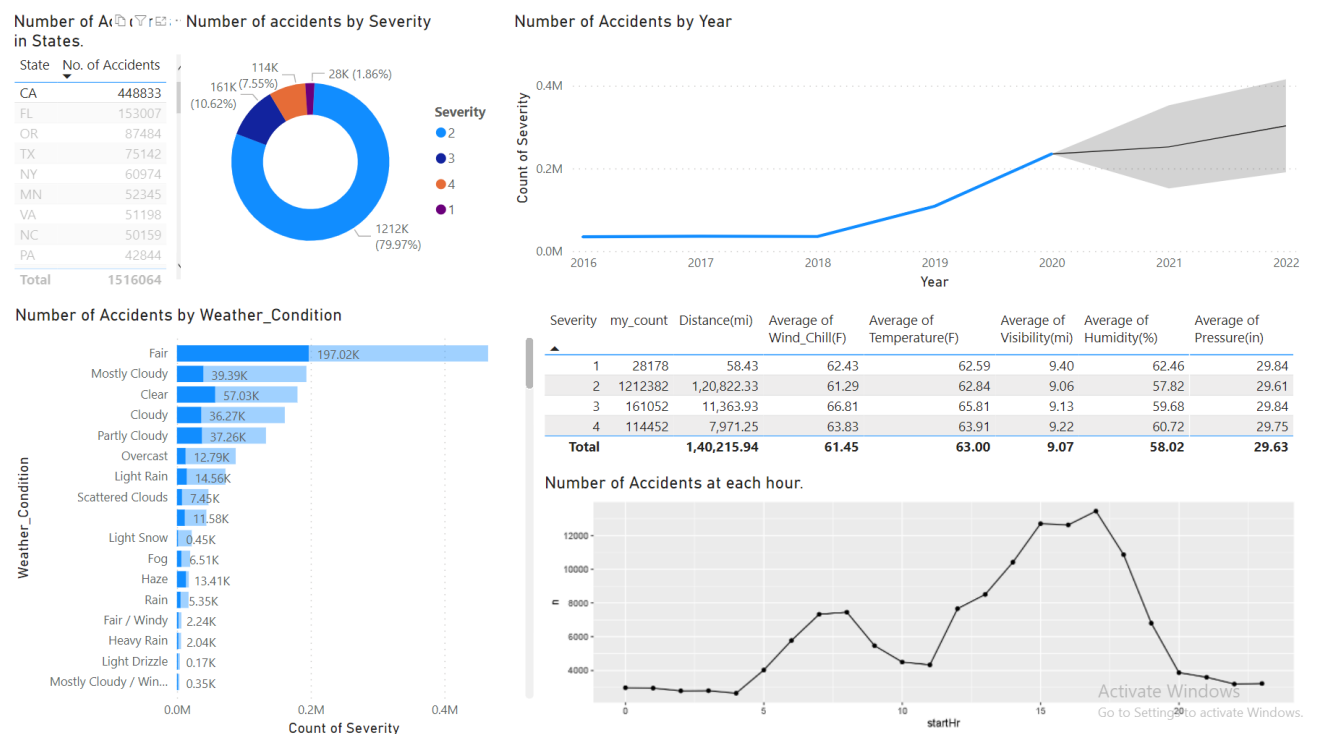
6. What are the number of accidents happened according to severity while considering the average of temperate, humidity, pressure, Visibility, and wind chill?

Severity	my_count	Distance(mi)	Average of Wind_Chill(F)	Average of Temperature(F)	Average of Visibility(mi)	Average of Humidity(%)	Average of Pressure(in)
1	28178	5,680.29	70.69	71.03	9.53	49.95	28.99
2	1212382	6,11,700.07	55.05	59.13	9.10	64.80	29.55
3	161052	97,639.58	54.57	61.84	9.35	64.12	29.58
4	114452	1,75,306.41	49.60	58.32	9.10	67.63	29.70
Total		8,90,326.35	55.11	59.58	9.13	64.66	29.55

From this we can conclude that most accidents occur with severity of 2 with average of respective weather statistics. This also gives us information on how much the total distance of traffic is affected with their respective severity number. A total of 890326.35 Miles of road with traffic is affected due to all the accidents.

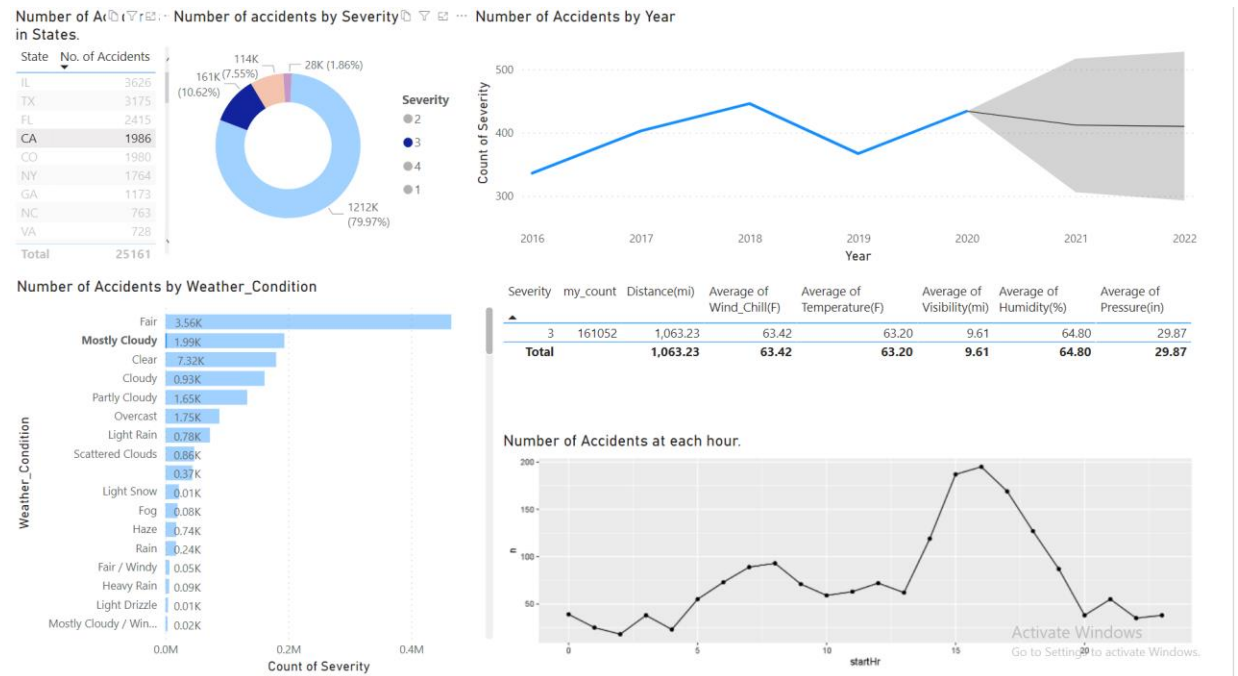
Now, these same visualizations can also be made in R as I did before. But the usage of tools like power BI is it makes everything easy. Now power Bi gives option to select and filter data. This can help us to answer the following questions.

7. In which weather condition accidents occur most given that a particular state with highest accident count & what would be the average of weather statistics according to the Severity? What will be the prediction of number of accidents in future for that particular state?



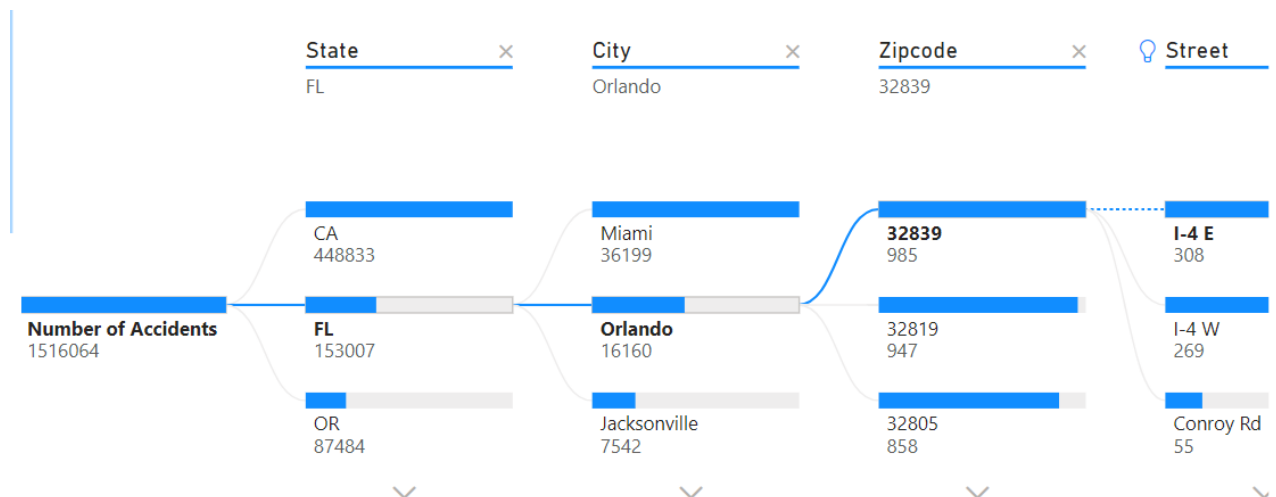
We can see that State CA is selected and all the values in visualizations are changed according to the selected data.

8. Now, What if I wanted to know the Number of accidents at each hour for severity of 3 where the weather is Mostly Cloudy? And what would be the prediction for future years?



We can see that for the following conditions how our visualizations changed. We can conclude that for the same conditions the number of accidents will decrease in near future.

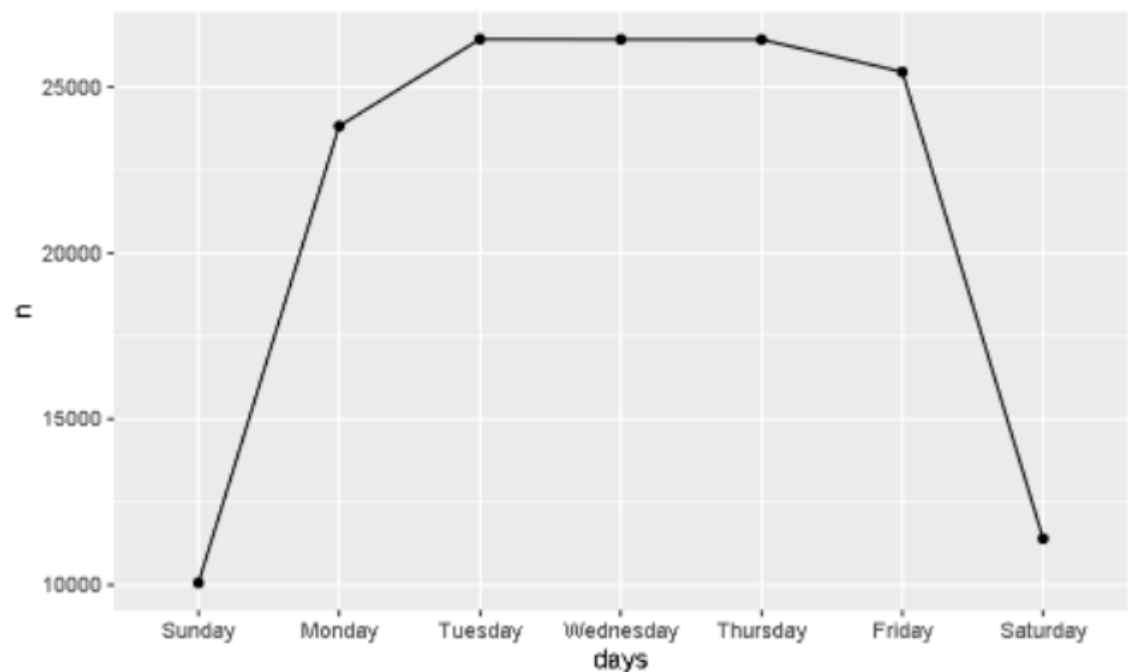
9. What is the number of accidents according to the state, city, zip code and street?



This is a decomposition tree. This represents the number of accidents in a hierarchical order. Suppose, you want to know the state with highest number of accidents, just click on Number of Accidents, then you want to know which city from that state has highest accidents. You click on the state CA, you get sorted cities and so on till you want to know which street has highest accidents.

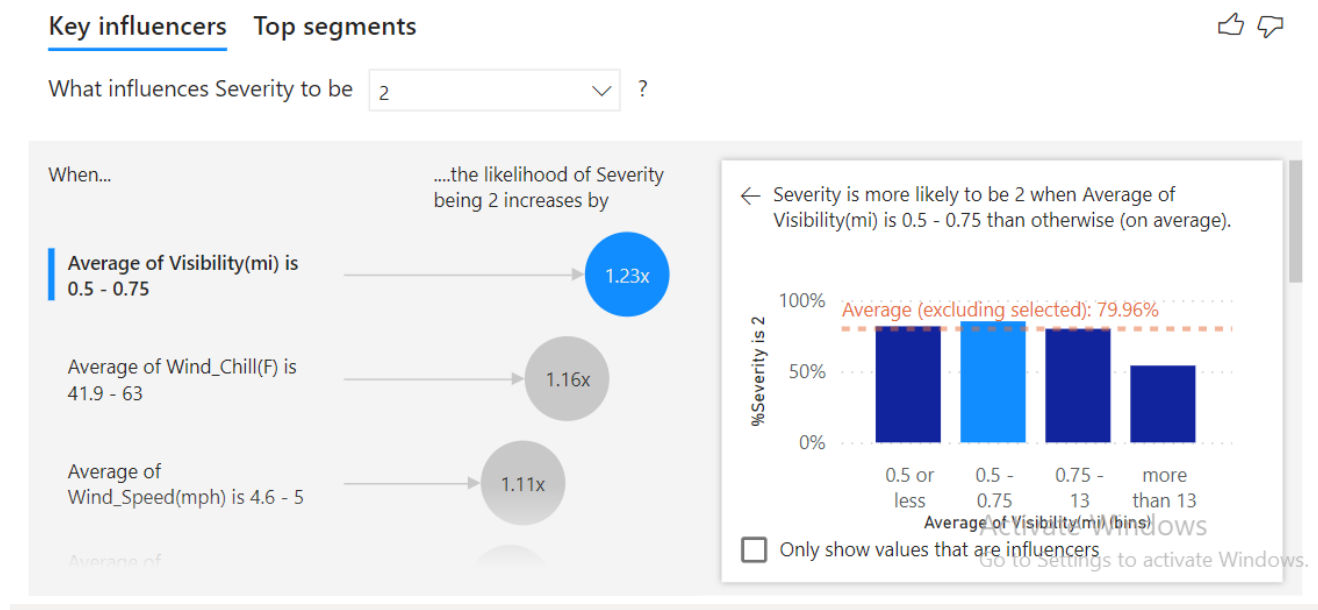
10. During which Week day accidents occur most?

Number of Accidents and Start_Time



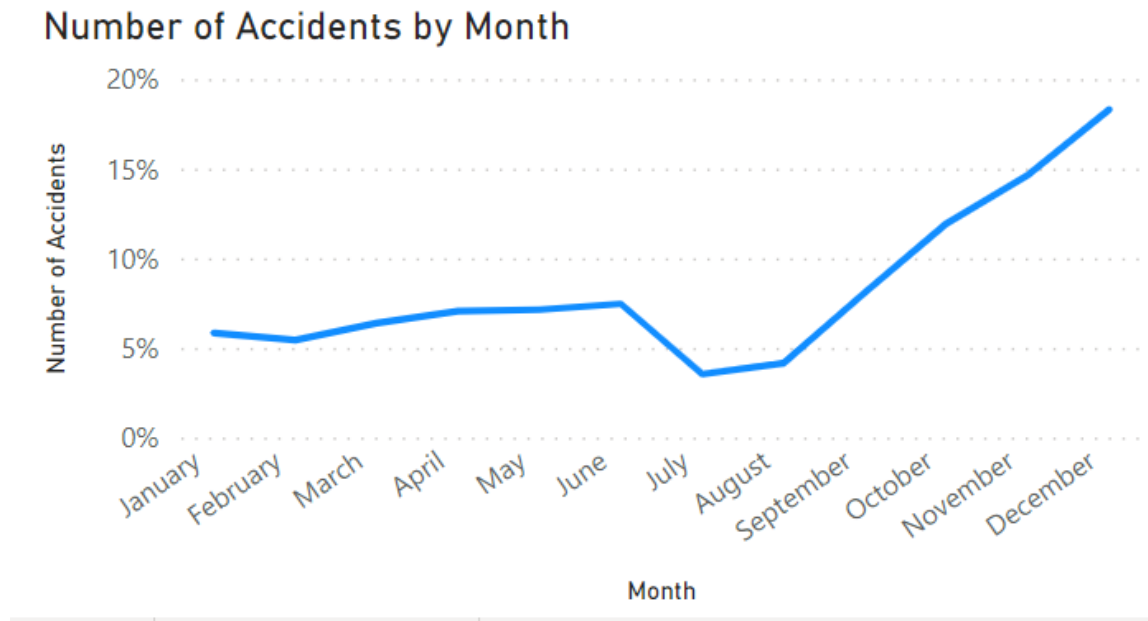
This is also an R script. This tells us that there are less number of accidents on weekends as compared to the regular weekdays.

11. On what weather factor does the severity of accident depends on?



This visualization is known as Key influencers. The key influencers' visual helps us to understand the factors that drive a metric we're interested in. It analyzes our data, ranks the factors that matter, and displays them as key influencers. For example, the above image shows what influences the Accident with severity of 2? The answer to that is. When average of visibility is between to 0.5 and 0.75 the likelihood of having accident with severity 2 increases by 1.23 X. The same type of interpretation can be done for all the influencers mentioned in the visual box.

12. In what month most accidents occur?



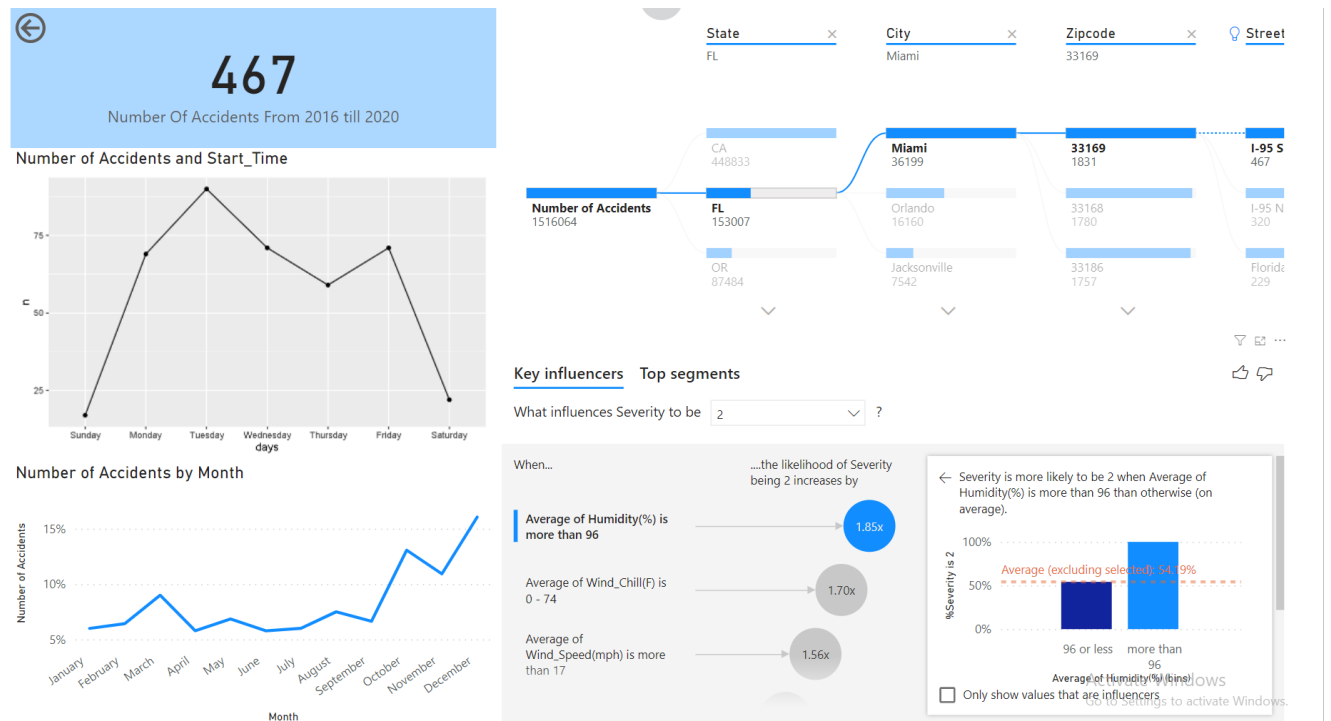
This is a line chart which shows most of the accidents occur in the month of October, November, & December.

13. What is the Total number of accidents for the dataset?



This is a card which hold the numerical value. In this case it shows the number of accidents.

14. In this too we can filter the data if we select something for example, if I wanted to know what weather factors affect the severity of 2 for the highest number of accidents on the street of Florida State, and at what month and day most accidents occur, for the same state?



So according to the selected decomposition tree of Florida State, until 2020 there were 467 accidents and most accident occurred on Tuesdays, and in the month of December. Now the above image also shows what affects the accident severity of 2. When humidity is more than 96, the chances of having accident with severity 2 increases by 1.86x.

Recommendations.

After analyzing and visualizing the data I will recommend the following things which may reduce the number of road accidents.

- The response teams and hospitals must be given special provisions in the hours in which most accidents are occurred.
- Warning signs about speed limits are to be put in the accident prone streets
- The state with highest accidents must be provided with better resources and budget plans to avoid accidents and rescue the victims.
- Warnings are to be put depending on the weather conditions which cause accidents.
- A mandate of vehicles to have first aid kit should be passed.
- Online surveillance for prompt response from emergency services should be implemented.
- To have enough response teams to rescue in accident prone locations

References

<https://zebrabi.com/top-power-bi-dashboard-tips-and-tricks/>

<https://databear.com/3-hidden-tricks-in-power-bi/>

<https://docs.microsoft.com/en-us/power-bi/create-reports/desktop-tips-and-tricks-for-creating-reports>

<https://docs.microsoft.com/en-us/power-bi/visuals/power-bi-visualization-influencers>

<https://www.youtube.com/watch?v=i1kCHZEhnEY>

<https://www.youtube.com/watch?v=DWjbd7NPUCM>

<https://powerbi.microsoft.com/en-in/partner-showcase/bizone-traffic-accident-analytics/>

<https://www.averanalytics.com/accident-dashboard>