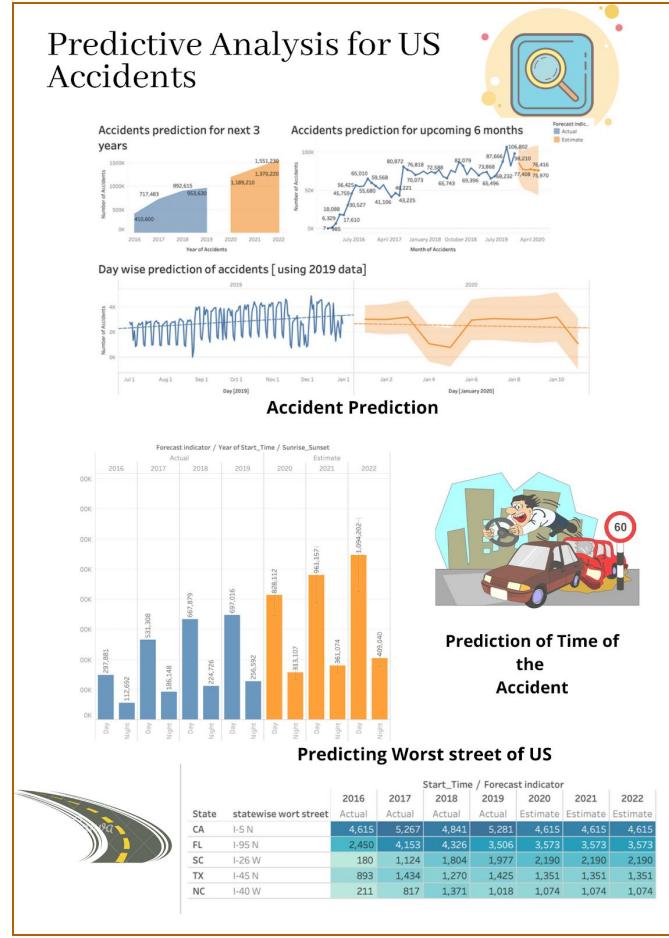
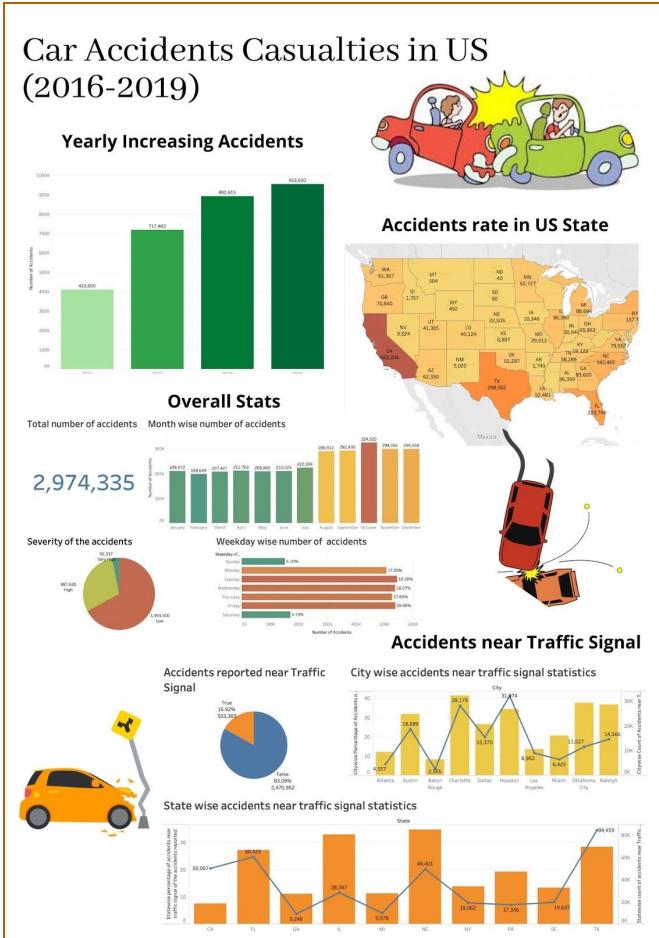


US-Accidents: A Countrywide Car Accident Dataset (2016-2019)



Shweta Gupta

Student Number - 3009898

MSc in Big Data Management and Analytics

Applied Data Analytics

Domain

A **traffic collision**, also called Car accident, occurs when a vehicle collides with another vehicle, pedestrian, animal, road debris, or other stationary obstruction, such as a tree, pole or building. Traffic collisions often result in injury, disability, death, and property damage as well as financial costs to both society and the individuals involved.

Road accidents have become very common these days. Nearly 1.25 million people die in road crashes each year, on average, 3,287 deaths a day. Moreover, 20–50 million people are injured or disabled annually. Road traffic crashes rank as the 9th leading cause of death and accounts for 2.2% of all deaths globally. Road crashes cost USD 518 billion globally, costing individual countries from 1–2% of their annual GDP. In the USA, over 37,000 people die in road crashes each year, and 2.35 million are injured or disabled. Road crashes cost the U.S. \$230.6 billion per year or an average of \$820 per person. Road crashes are the single greatest annual cause of death of healthy U.S. citizens travelling abroad.

Source of Dataset

Dataset has been taken from Kaggle - <https://www.kaggle.com/sobhanmoosavi/us-accidents>

Description of Dataset

This is a countrywide car accident dataset, which covers 49 states of the United States. The data is continuously being collected from February 2016, using several data providers, including two APIs which provide streaming traffic event data. These APIs broadcast traffic events captured by a variety of entities, such as the US and state departments of transportation, law enforcement agencies, traffic cameras, and traffic sensors within the road-networks. Currently, there are about **3.0 million** accident records in this dataset.

Attributes of Dataset

	Attribute	Description
1	ID	This is a unique identifier of the accident record.
2	Source	Indicates the source of the accident report (i.e. the API which reported the accident.).
3	TMC	A traffic accident may have a Traffic Message Channel (TMC) code which provides a more detailed description of the event.
4	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).
5	Start_Time	Shows start time of the accident in local time zone.
6	End_Time	Shows end time of the accident in local time zone.
7	Start_Lat	Shows latitude in GPS coordinate of the start point.
8	Start_Lng	Shows longitude in GPS coordinate of the start point.
9	End_Lat	Shows latitude in GPS coordinate of the end point.
10	End_Lng	Shows longitude in GPS coordinate of the end point.
11	Distance(mi)	The length of the road extent affected by the accident.
12	Description	Shows natural language description of the accident.
13	Number	Shows the street number in address field.

14	Street	Shows the street name in address field.
15	Side	Shows the relative side of the street (Right/Left) in address field.
16	City	Shows the city in address field.
17	County	Shows the county in address field.
18	State	Shows the state in address field.
19	Zipcode	Shows the zipcode in address field.
20	Country	Shows the country in address field.
21	Timezone	Shows timezone based on the location of the accident (eastern, central, etc.).
22	Airport_Code	Denotes an airport-based weather station which is the closest one to location of the accident.
23	Weather_Timestamp	Shows the time-stamp of weather observation record (in local time).
24	Temperature(F)	Shows the temperature (in Fahrenheit).
25	Wind_Chill(F)	Shows the wind chill (in Fahrenheit).
26	Humidity(%)	Shows the humidity (in percentage).
27	Pressure(in)	Shows the air pressure (in inches).
28	Visibility(mi)	Shows visibility (in miles).
29	Wind_Direction	Shows wind direction.

30	Wind_Speed(mph)	Shows wind speed (in miles per hour).
31	Precipitation(in)	Shows precipitation amount in inches, if there is any.
32	Weather_Condition	Shows the weather condition (rain, snow, thunderstorm, fog, etc.)
33	Amenity	A POI annotation which indicates presence of amenity in a nearby location.
34	Bump	A POI annotation which indicates presence of speed bump or hump in a nearby location.
35	Crossing	A POI annotation which indicates presence of crossing in a nearby location.
36	Give_Way	A POI annotation which indicates presence of give_way in a nearby location.
37	Junction	A POI annotation which indicates presence of junction in a nearby location.
38	No_Exit	A POI annotation which indicates presence of no_exit in a nearby location.
39	Railway	A POI annotation which indicates presence of railway in a nearby location.
40	Roundabout	A POI annotation which indicates presence of roundabout in a nearby location.
41	Station	A POI annotation which indicates presence of station in a nearby location.
42	Stop	A POI annotation which indicates presence of stop in a nearby location.
43	Traffic_Calming	A POI annotation which indicates presence of traffic_calming

		in a nearby location.
44	Traffic_Signal	A POI annotation which indicates presence of traffic_signal in a nearby location.
45	Turning_Loop	A POI annotation which indicates presence of turning_loop in a nearby location.
46	Sunrise_Sunset	Shows the period of day (i.e. day or night) based on sunrise/sunset.
47	Civil_Twilight	Shows the period of day (i.e. day or night) based on civil twilight.
48	Nautical_Twilight	Shows the period of day (i.e. day or night) based on nautical twilight.
49	Astronomical_Twilight	Shows the period of day (i.e. day or night) based on astronomical twilight.

Analysis Goals

1. Perform exploratory data analysis on this dataset and generate insights about car accidents in the United States.
2. Take a closer look into what factors affect the severity levels of car accidents in the United States. The severity variable here indicates the impact of the accident on traffic delay, not how severe the damage was to the vehicle.
3. Find out what are the factors affecting the car accidents in the US.

Questions

1. How is weather condition correlated with accidents?
2. What areas of the country are geographically critical in terms of accidents?
3. What is the impact of time horizons on accidents?
4. How accidents affect traffic in the US?
5. What are the worst roads or highways in the US in terms of accidents?

6. What is the ratio of accidents near the traffic signal?
7. What is the impact of the time horizon?
8. What are the geographically critical areas?

Who can utilize this Analysis

US-Accidents analysis can be used for numerous applications such as real-time accident prediction, studying accident hotspot locations, casualty analysis and extracting cause and effect rules to predict accidents, or studying the impact of precipitation or other environmental stimuli on accident occurrence.

Problem Statement

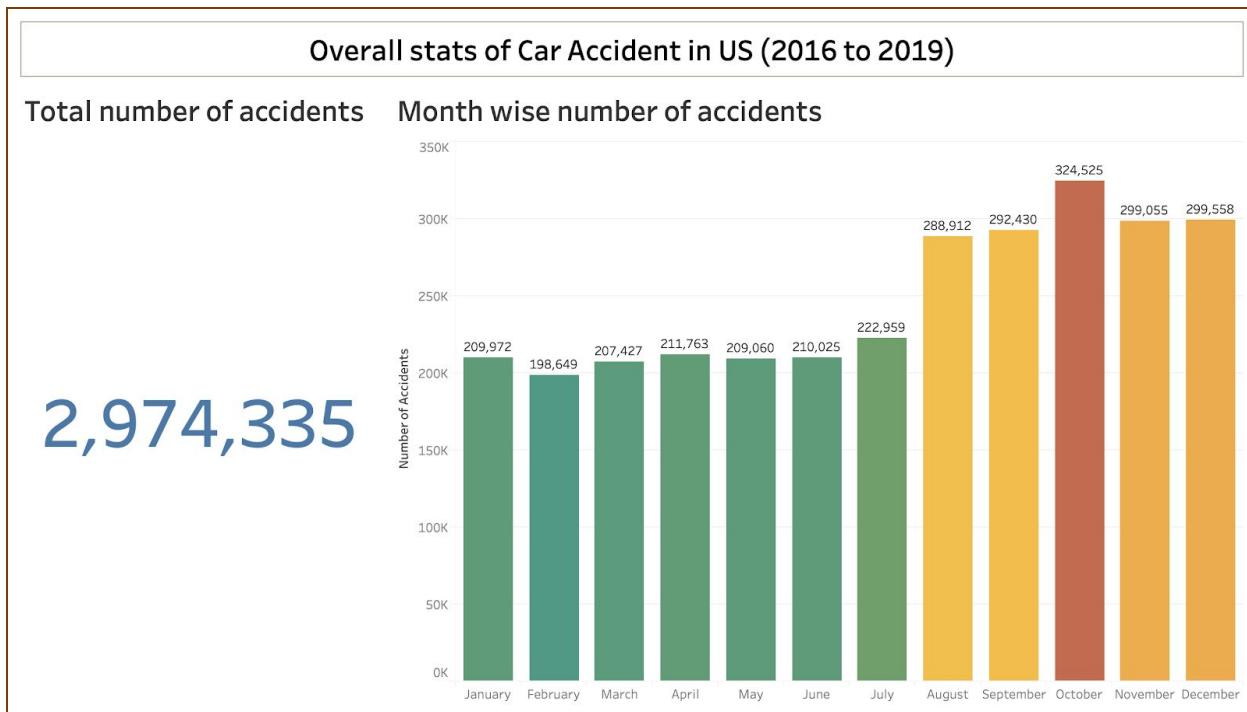
Reducing traffic accidents is an important public safety challenge around the world. The main purpose of the analysis is to find out the factors behind accidents such as weather, traffic, temperature etc. Accidents are also causing traffic in the states of the US, find out how much accidents are affecting the traffic. Accident prediction is important for optimizing public transportation, enabling safer routes, and cost-effectively improving the transportation infrastructure, all in order to make the roads safer.

Dashboards

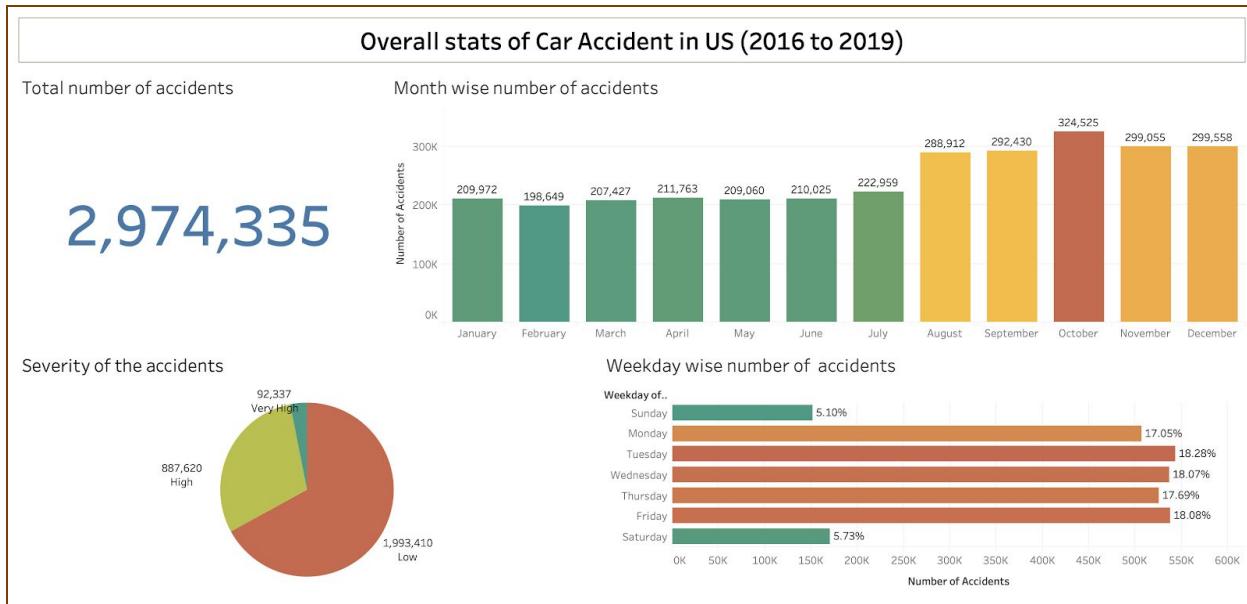
[Informative Dashboards](#)

Dashboard 1 - Overall stats of Car Accidents in the US (2016-2019)

As the Accidents are increasing year by year and the deaths because of the accidents are also increasing. So here is the very first dashboard, we will analyse the rate of an increasing number of accidents in the US from the year 2016 to 2019.



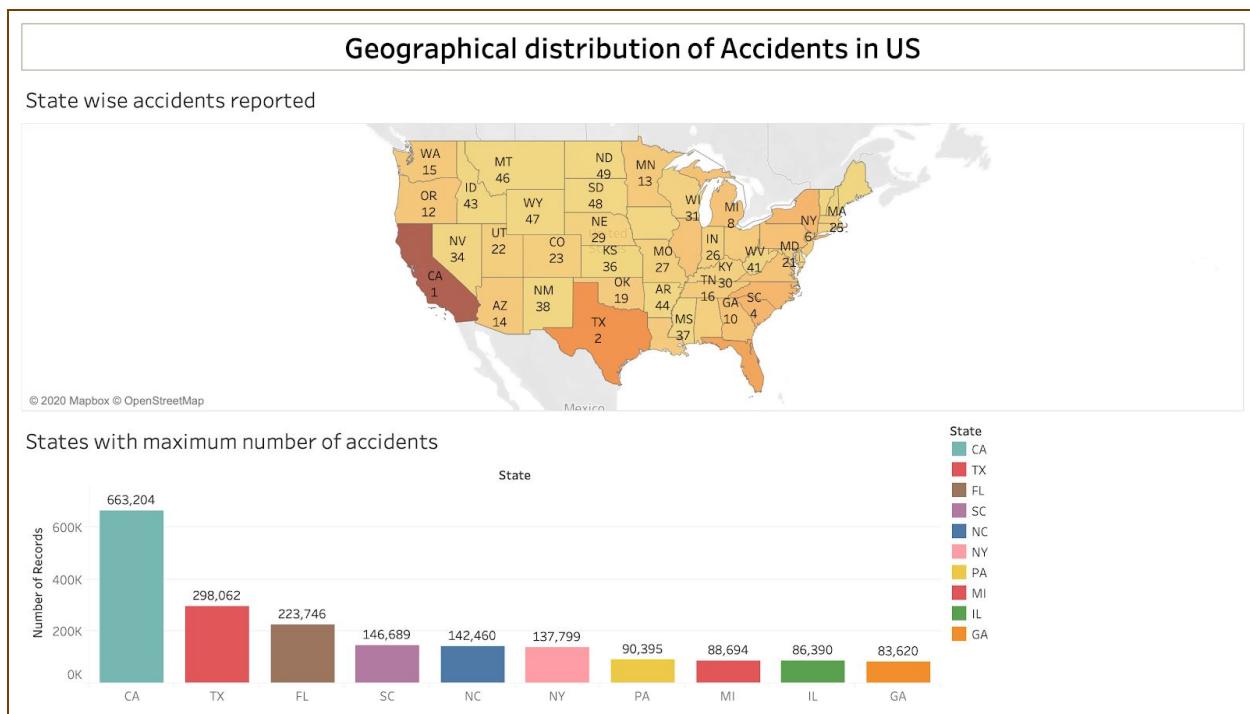
Here we can see the total number of accidents in the last few years is 2,97335 which is such a high rate of accidents in the US. If we see these numbers monthly, every month there are more than 198649 number of accidents reported. In the month of October, the number of accidents reported are higher than the other months. We will go in more detail to analyse the number of accidents in the US.



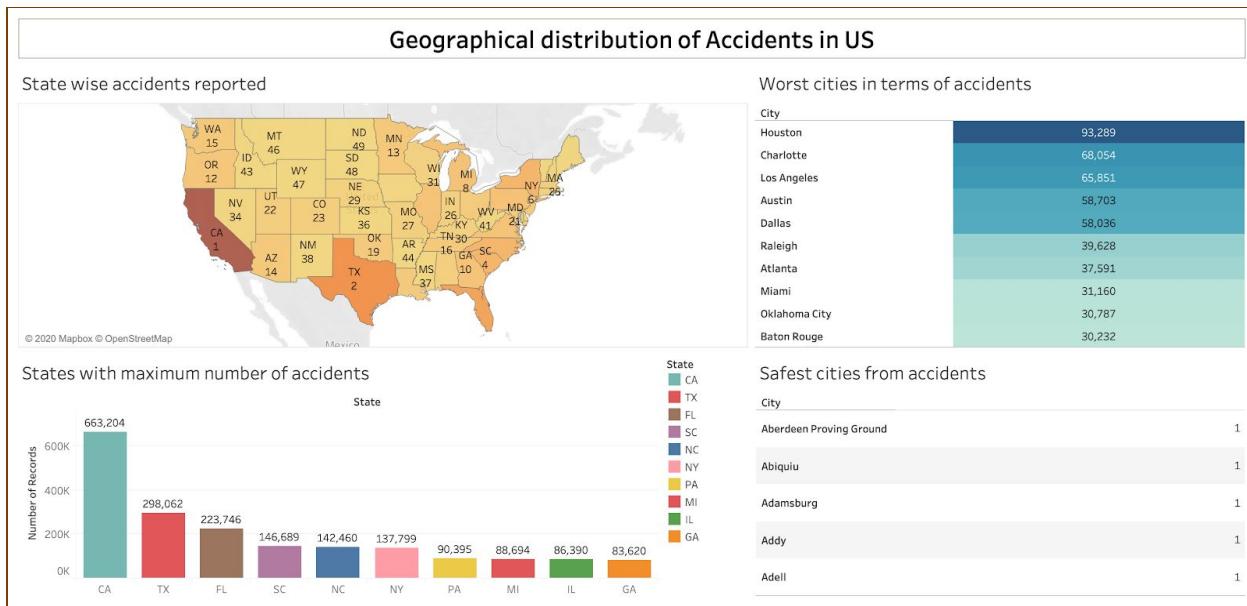
As we can see in the chart the severity of the accidents of most of the accidents are low but there are large numbers of accidents which have high severity, this is where the number of deaths due to accidents increases. On the other hand, there are very less cases which have high severity. If we see the increasing number of accidents according to the weekdays Tuesdays have the more number of accidents reported compared to Sunday and Saturday (weekends). So we can summarize that weekdays have the maximum number of accidents.

Dashboard 2 - Geographical distribution of Accidents in the US (2016-2019)

Looking at the geographical distribution of the accidents we found that most of the states of the US are having a high number of accidents but the popular states have the maximum number of accidents such as California is ranked first in terms of accidents after California, Texas is on the second rank in terms of accident and Florida is on the third. So here we can find that the more popular the state is, the more accidents and casualties it has.

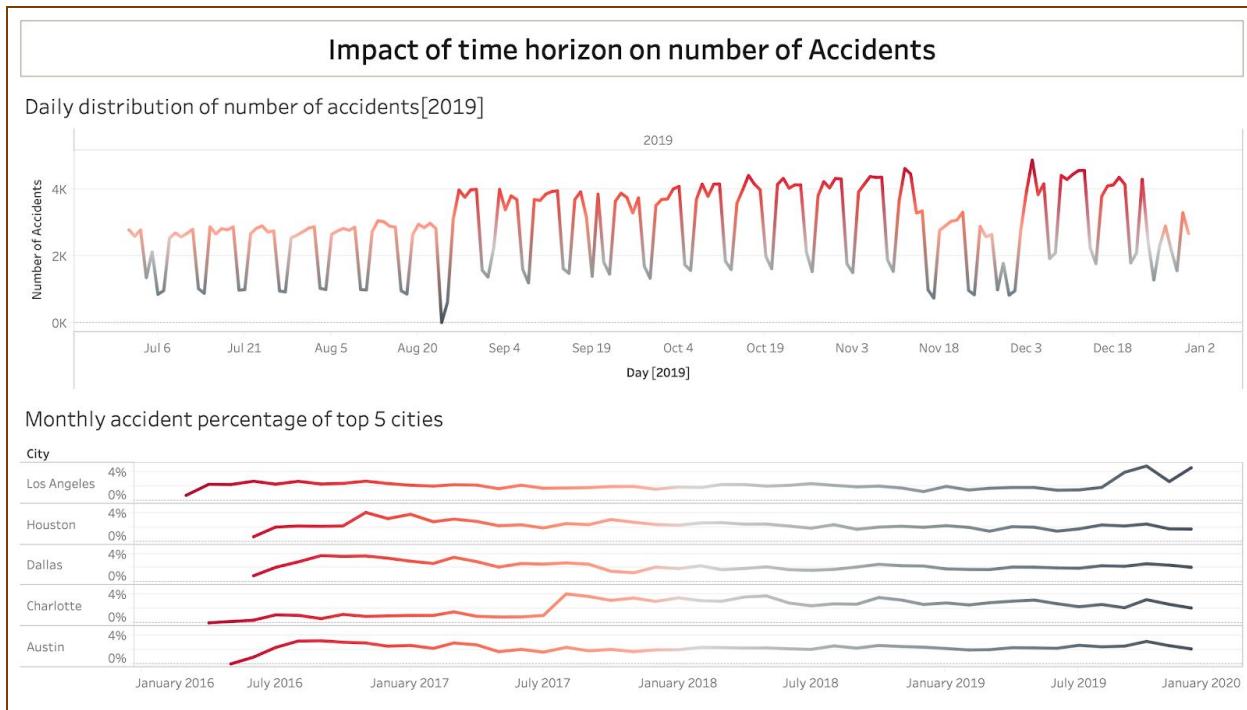


The safest cities in the US in terms of accidents are found after this analysis and the worst cities are also found which have a high rate of accidents in the US. Houston city has the maximum number of accidents, Los Angeles and Charlotte are also on a high rate of accidents. With the help of this analysis, the red zone areas of the US can be found for the accidents and the necessary action towards can be taken for these areas.



Dashboard 3 - Impact of time horizon on number of Accidents in the US (2016-2019)

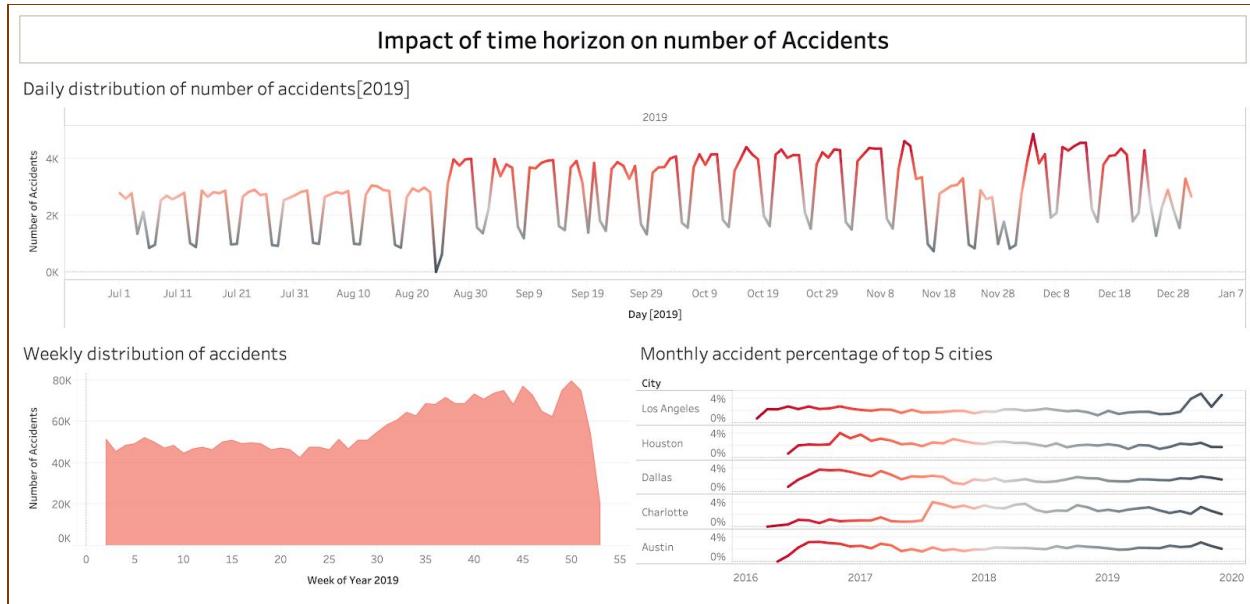
We are analysing here over the time how the number of accidents are increasing or decreasing. This Dashboard is the analysis of accidents over the period of time.



The number of accidents are drastically increasing day by day. As we are analysing the 2019's

data for the accident the graph depicts that the number of accidents are not decreasing at all they are keep on growing after some period of time.

There are top 5 cities with the highest percentage of accidents over the year. Los Angeles had a number of accidents in the year of 2019 which was very low in 2016. Whereas other cities are quite stable in terms of accidents over the year.



Weekly distribution of the accidents tells that the number of accidents are drastically increased after the 30th week of the year. And the accidents have reached near 80K in the week of 52.

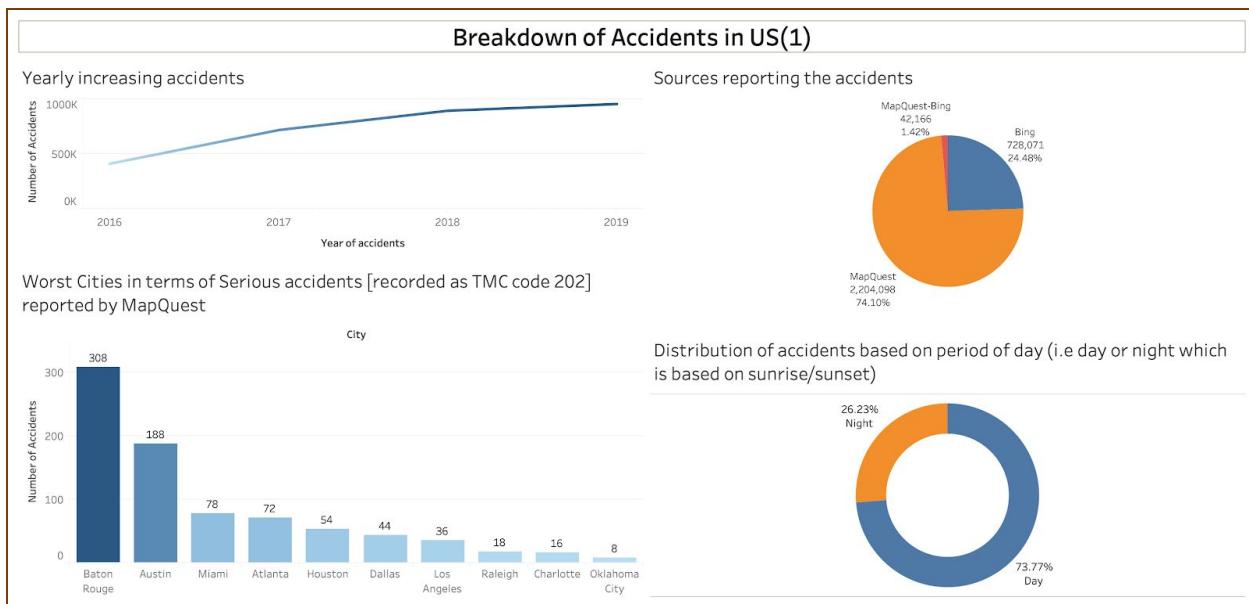
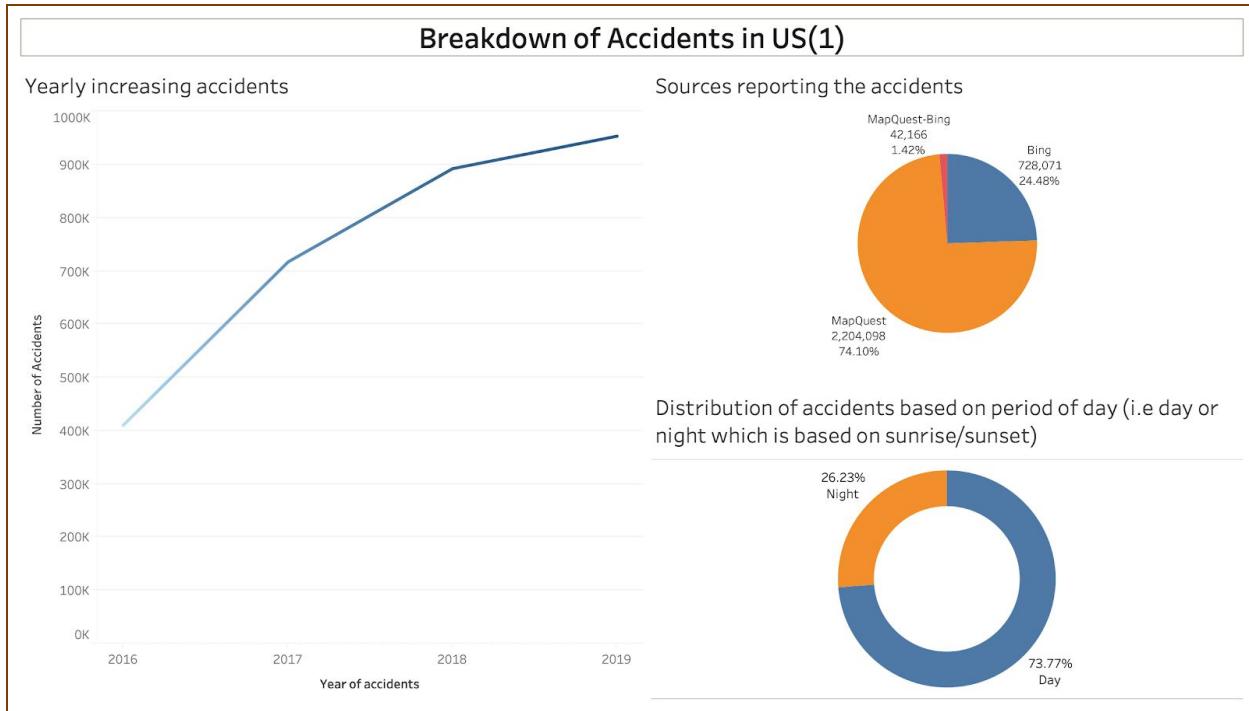
The necessary action can be taken to improve the accidents in the city of Los Angeles and the weekly increasing number of accidents.

Dashboard 4 - Breakdown of Accidents in the US(1) (2016-2019)

This dashboard is the depth analysis of accidents in the US. The breakdown of the aspects of accidents in the US. The first chart in the Dashboard displays that the number of accidents are increasing each and every year. From 2016 to 2019 the number of accidents has increased from 500k to 1000k. The sources which are reporting the accidents MapQuest and Bing and MApQuest-Bing most of the accidents are being reported by the Mapquest. Very few accidents are being reported by MapQuest-Bing.

Another chart tells the time of the day when the accident happened. Most of the accidents are reported in the daytime. The reason might be the traffic which is causing accidents. The

number of accidents at night are very less compared to the daytime.

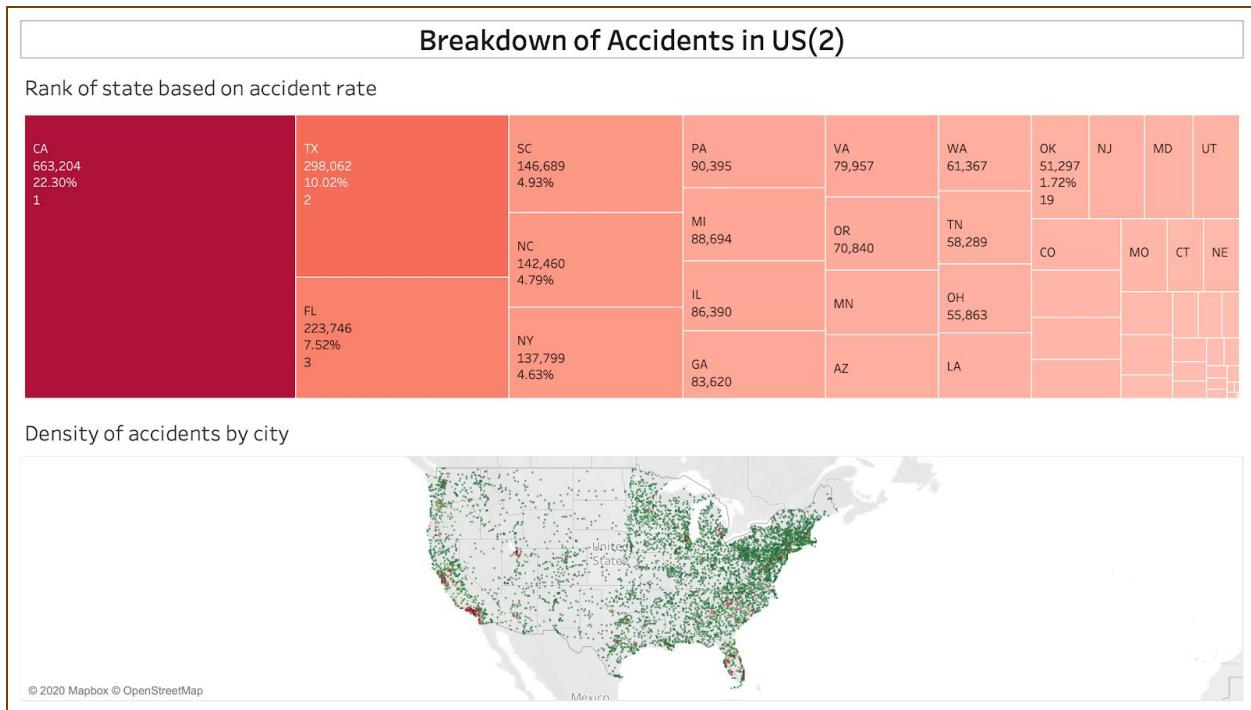


This dashboard also tells the worst cities in terms of serious accidents reported. Baton Rouge has the most serious cases of accidents in the US. Austin is on the second number.

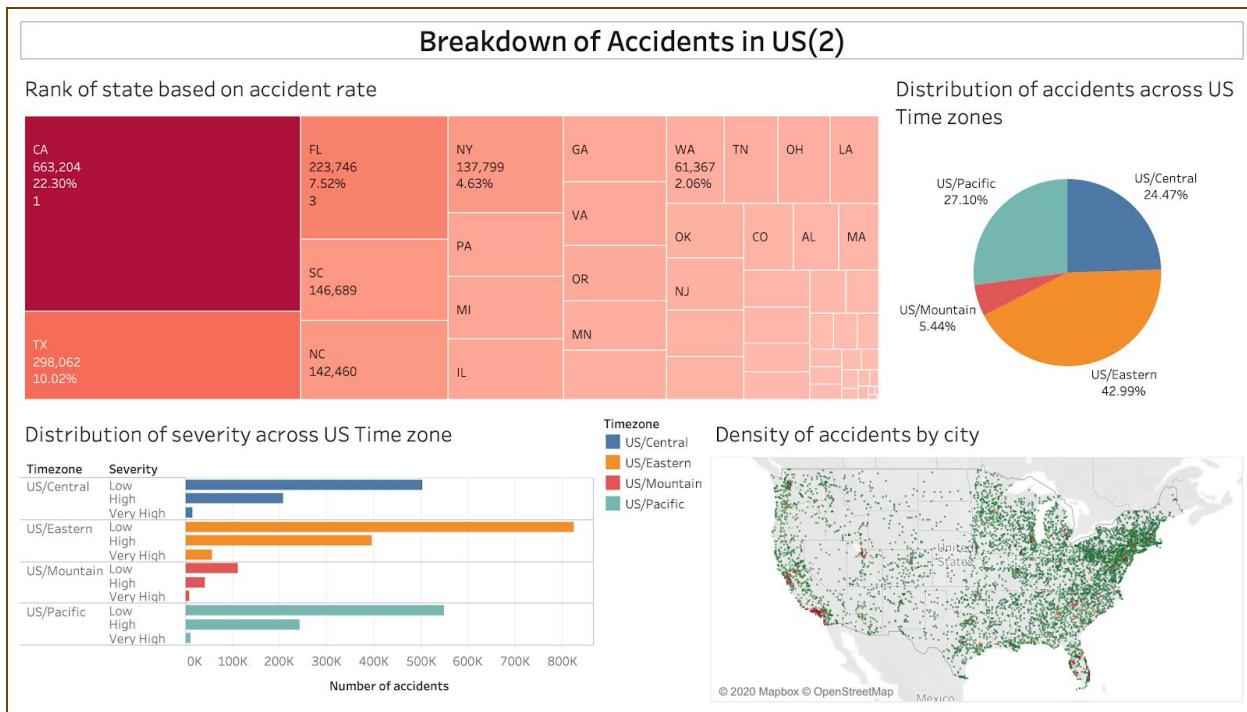
So, here as we know the worst city for serious accidents and the time of the accidents, the actions can be taken towards the city of Baton Rouge and Austin.

Dashboard 5 - Breakdown of Accidents in the US(2) (2016-2019)

This is the continuation of dashboard 4 for breaking down the accidents that happened in the US from the year 2016 to 2019.

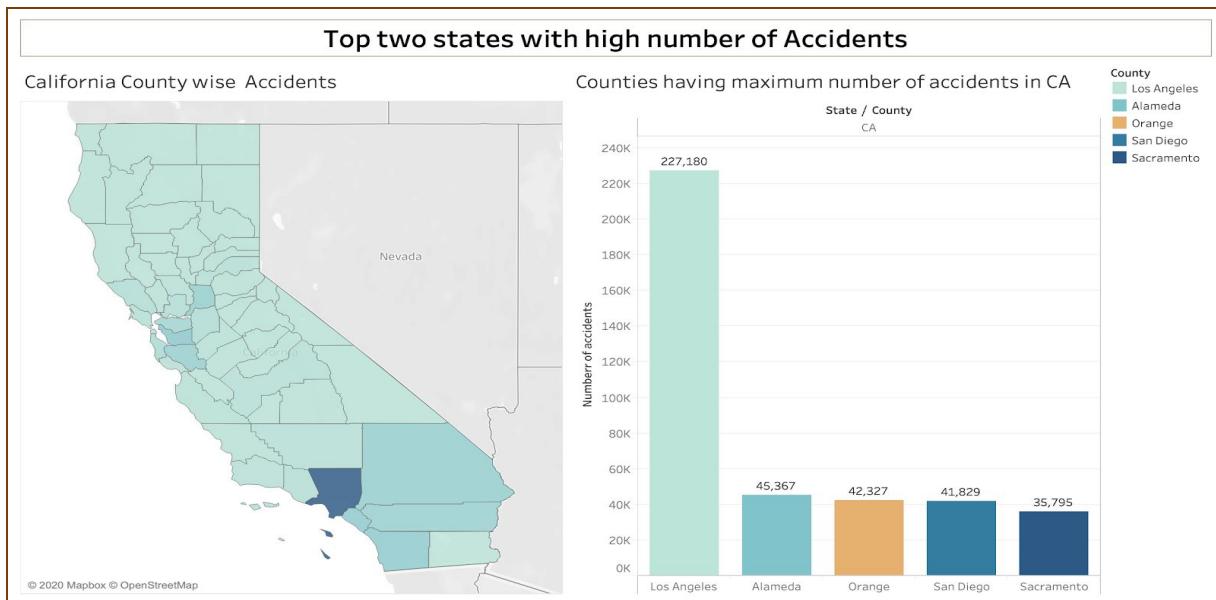


The TreeMap analysis of accidents according to the states tell that California has 22.30% of the accidents from 2016 to 2019 and has scored the first rank in terms of the accidents. This state should be the red zoned state for the accidents happening in the US. TX and FL also have 10.2% and 7.52% of accidents. Other states have less than 5% of the overall accidents.



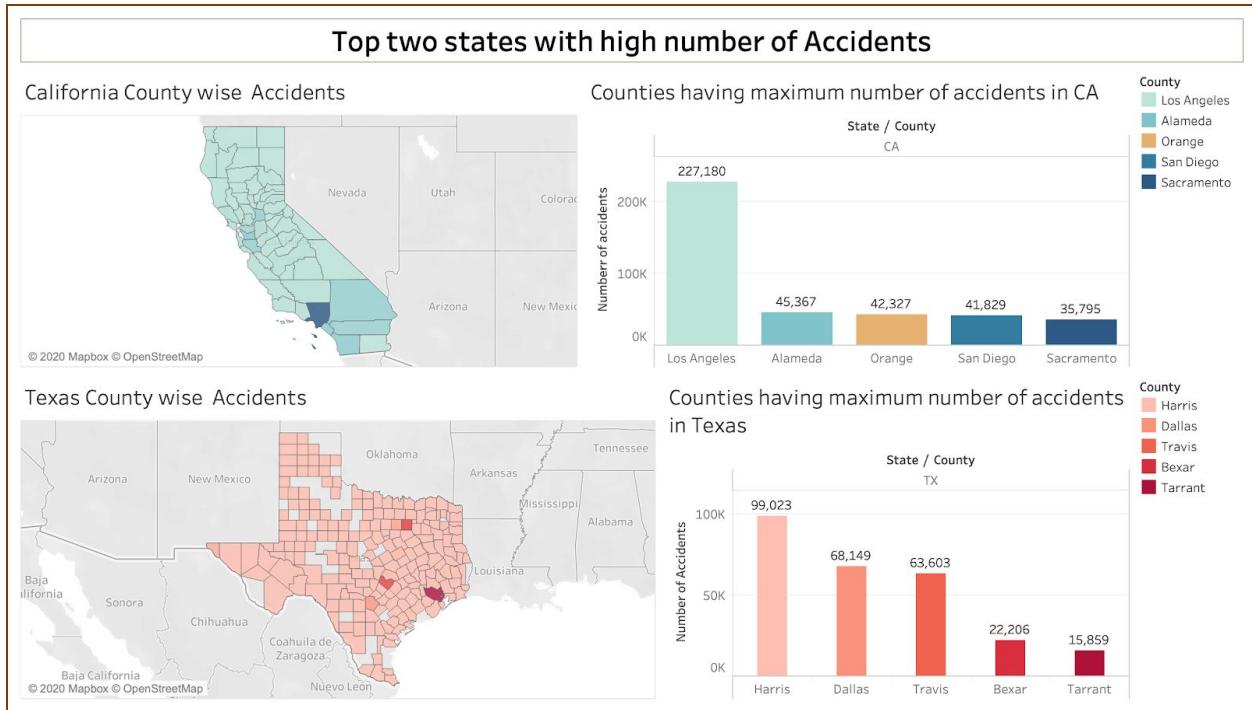
US Eastern has the maximum number of accidents and the severity of the accidents in US Eastern is also high when compared to others. US Mountain has a lesser number of accidents but the number of high severity accidents are similar to US Central. All the red marked cities have more number of accidents in the density graph.

Dashboard 6 - Top two states with the high number of Accidents in the US (2016-2019)



In this Dashboard, we will be analysing the top 2 states of the US which are having the maximum number of accidents. As we have seen till now California is on the top for the reported number of accidents. So now here we will be checking the counties which have the highest number of accidents in California. Los Angeles is the county which has a large amount of

accidents.



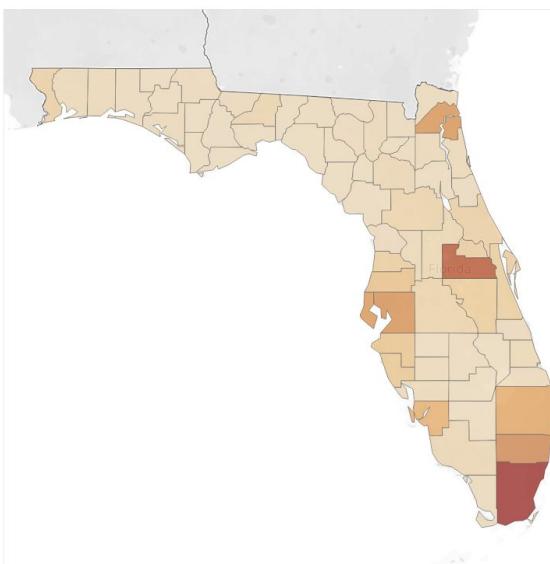
The second topmost state is Texas in terms of accidents and in Texas Harris is the county which has a high rate of accidents. Dallas and Travis also have a large number of accidents reported.

Dashboard 7 - Top Third and Fourth states with the high number of Accidents in the US (2016-2019)

In this Dashboard, we will be analysing the top third and fourth states of the US which are having the maximum number of accidents. As we have seen till now Florida is on the top third for the reported number of accidents. So now here we will be checking the counties which have the highest number of accidents in Florida. Miami is the county which has a large number of accidents over the year 2016 to 2019 After that Orange and Broward has a large number of accidents.

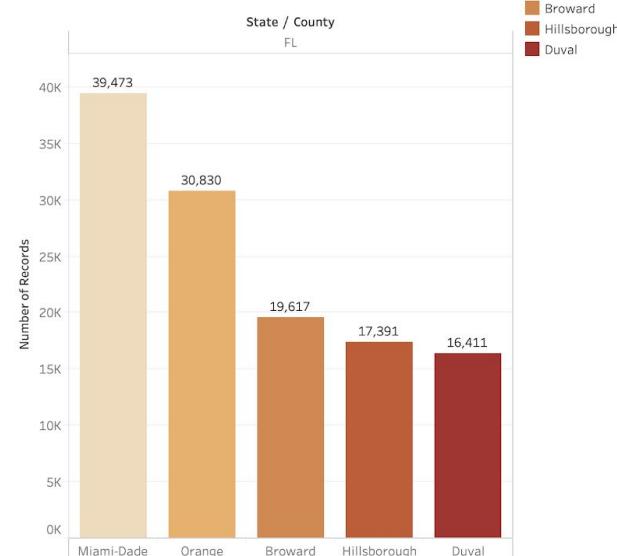
Top third and fourth states with high number of Accidents

Florida County wise Accidents



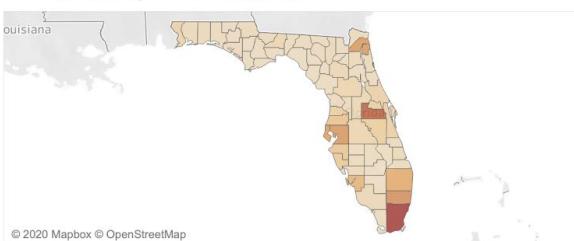
© 2020 Mapbox © OpenStreetMap

Counties having maximum number of accidents in Florida

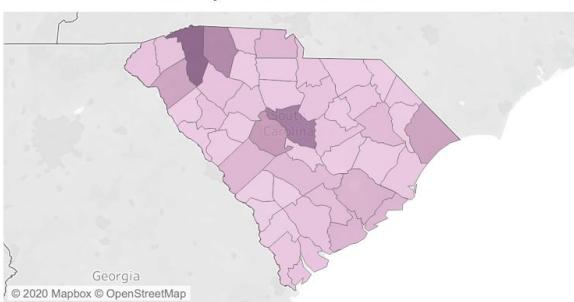


Top third and fourth states with high number of Accidents

Florida County wise Accidents

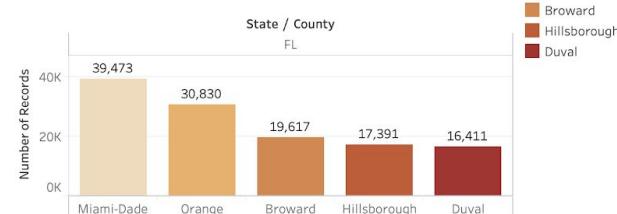


South Carolina County wise Accidents

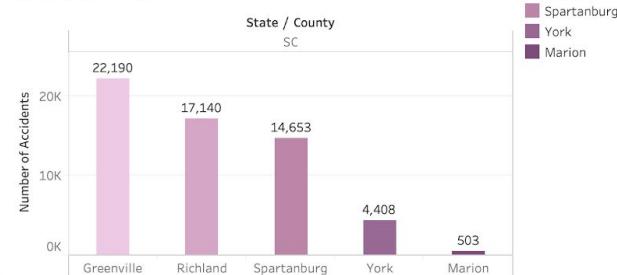


© 2020 Mapbox © OpenStreetMap

Counties having maximum number of accidents in Florida



Counties having maximum number of accidents in South Carolina

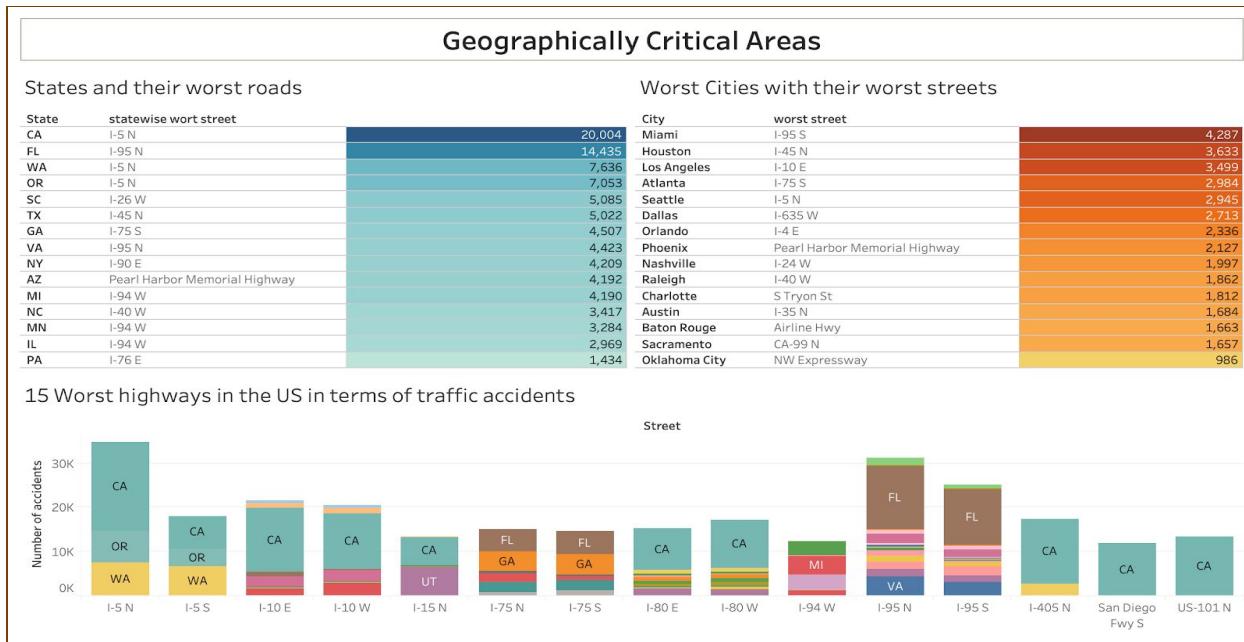


The fourth highest state in terms of accidents is South Carolina and the county which has the maximum number of accidents in South Carolina is Greenville with the number of accidents 22190.

Dashboard 8 - Geographically Critical Areas of in the US (2016-2019)

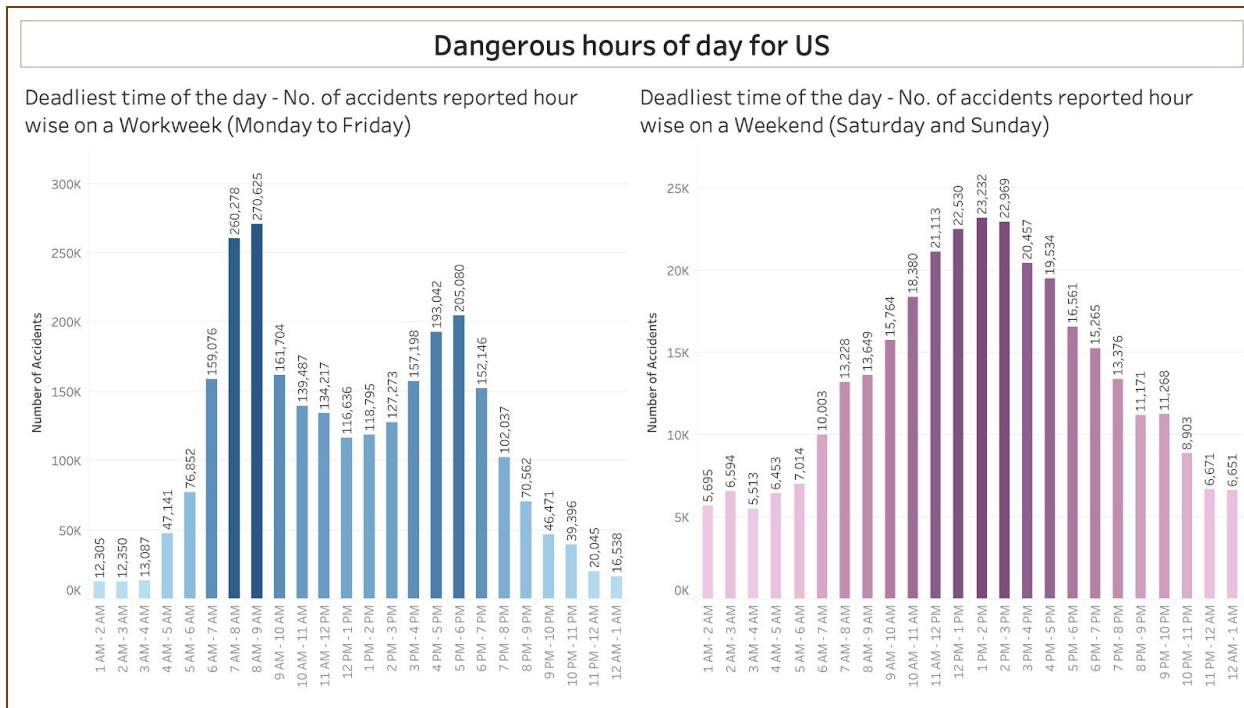
Geographically Critical Areas			
States and their worst roads		Worst Cities with their worst streets	
State	statewise worst street	City	worst street
CA	I-5 N	Miami	I-95 S
FL	I-95 N	Houston	I-45 N
WA	I-5 N	Los Angeles	I-10 E
OR	I-5 N	Atlanta	I-75 S
SC	I-26 W	Seattle	I-5 N
TX	I-45 N	Dallas	I-635 W
GA	I-75 S	Orlando	I-4 E
VA	I-95 N	Phoenix	Pearl Harbor Memorial Highway
NY	I-90 E	Nashville	I-24 W
AZ	Pearl Harbor Memorial Highway	Raleigh	I-40 W
MI	I-94 W	Charlotte	S Tryon St
NC	I-40 W	Austin	I-35 N
MIN	I-94 W	Baton Rouge	Airline Hwy
IL	I-94 W	Sacramento	CA-99 N
PA	I-76 E	Oklahoma City	NW Expressway

This Analysis tells the critical areas of the US in terms of accidents that happened in 2016 to 2019. Here is the list of top dangerous states and top dangerous cities with the highest number of accidents reported. State CA is the most critical state and City Miami is the most critical city in terms of accidents.



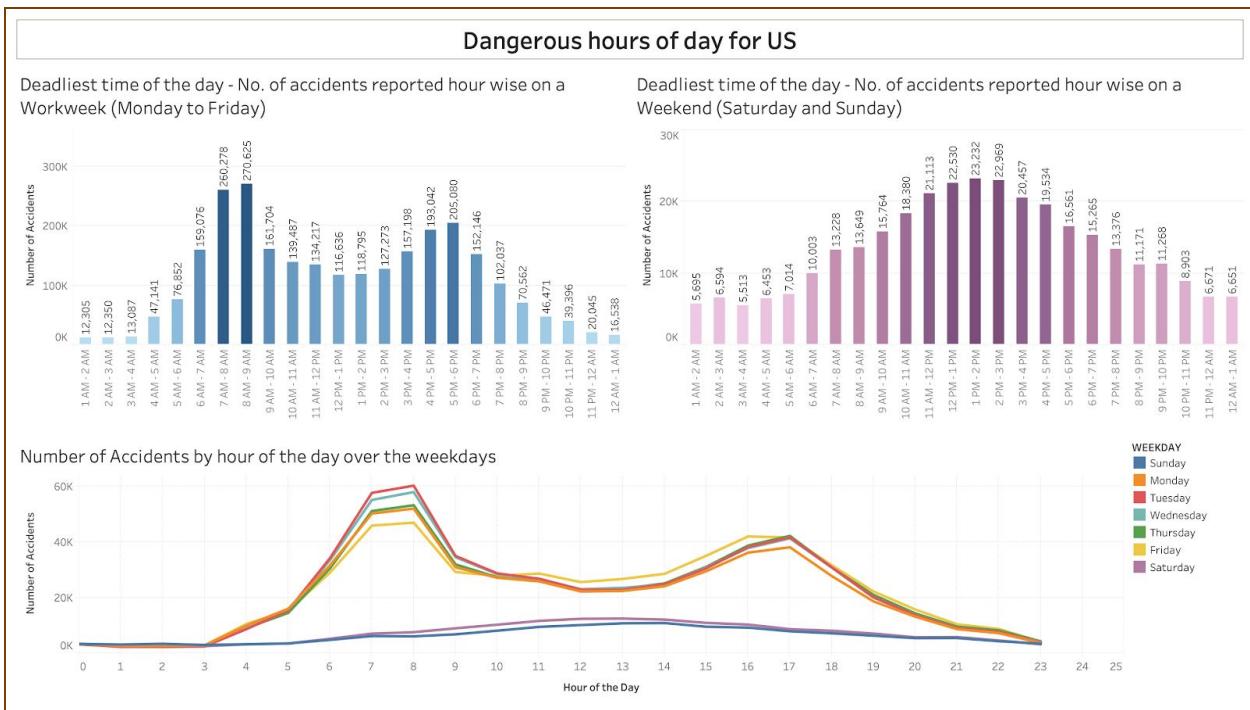
Here is the list of the worst highways in the US which are dangerous for accidents. The maximum number of accidents that happen in the US is on the highway I-5 N. After that I-95 S and I-405 N have a high number of accidents. These highways in the US can be more secure and might require more traffic signals to improve the accident rate.

Dashboard 9 - Dangerous hours of day for US (2016-2019)



Here we are finding the dangerous hours of the day for the US in terms of accidents. The hour of the day when the accident rate is high. As we can see from the graph the accident rate is way higher on weekdays than the weekends. And if we look at the hour, The maximum number of accidents are happening in the morning at 7 to 8 am and 9 to 10 am which is the timing for maximum of the office and schools and etc therefore it is clear that the traffic rate is also high and this is the peak time of the day so the number of accidents are high.

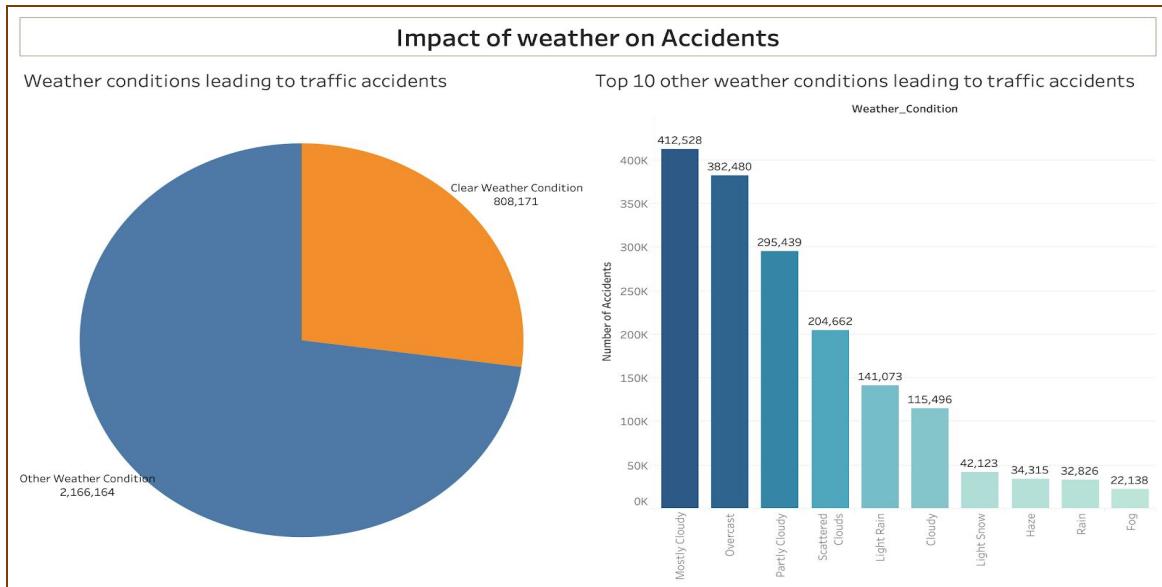
Whereas on weekends 1pm to 2pm is the time when the maximum number of accidents are happening. 11 am to 3pm is the alert time for accidents on weekends.



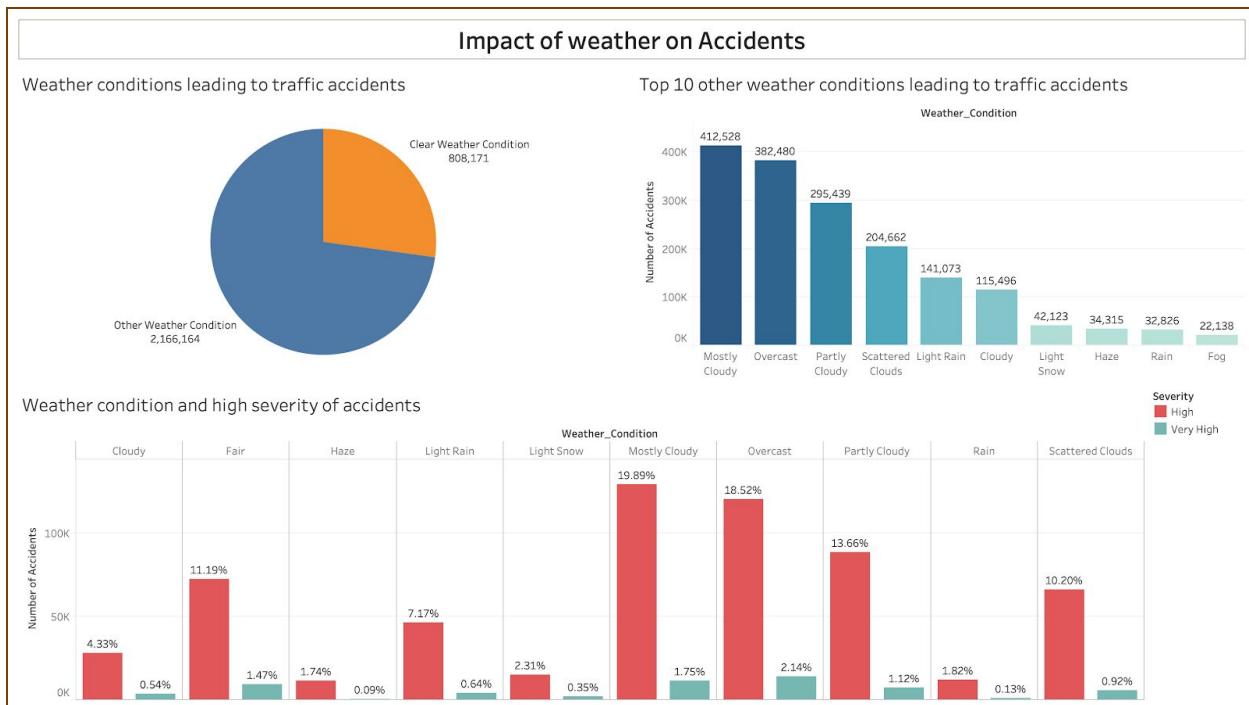
By seeing the graph of Number of accidents by hour of the day we can say that the 7th and 8th hour of the day have the maximum number of accidents.

The danger time can be given to the public so that people can avoid travelling at the red alter time of the day in order to prevent accidents.

Dashboard 10 - Impact of weather on Accidents in US (2016-2019)



How weather is affecting the number of accidents in the US? This Dashboard is the answer for the question. AS you can see that most of the accidents are happening in other weather conditions. When the weather is clear the number of accidents are less. If we go in more depth the worst weather for accidents is cloudy and after that Overcast.

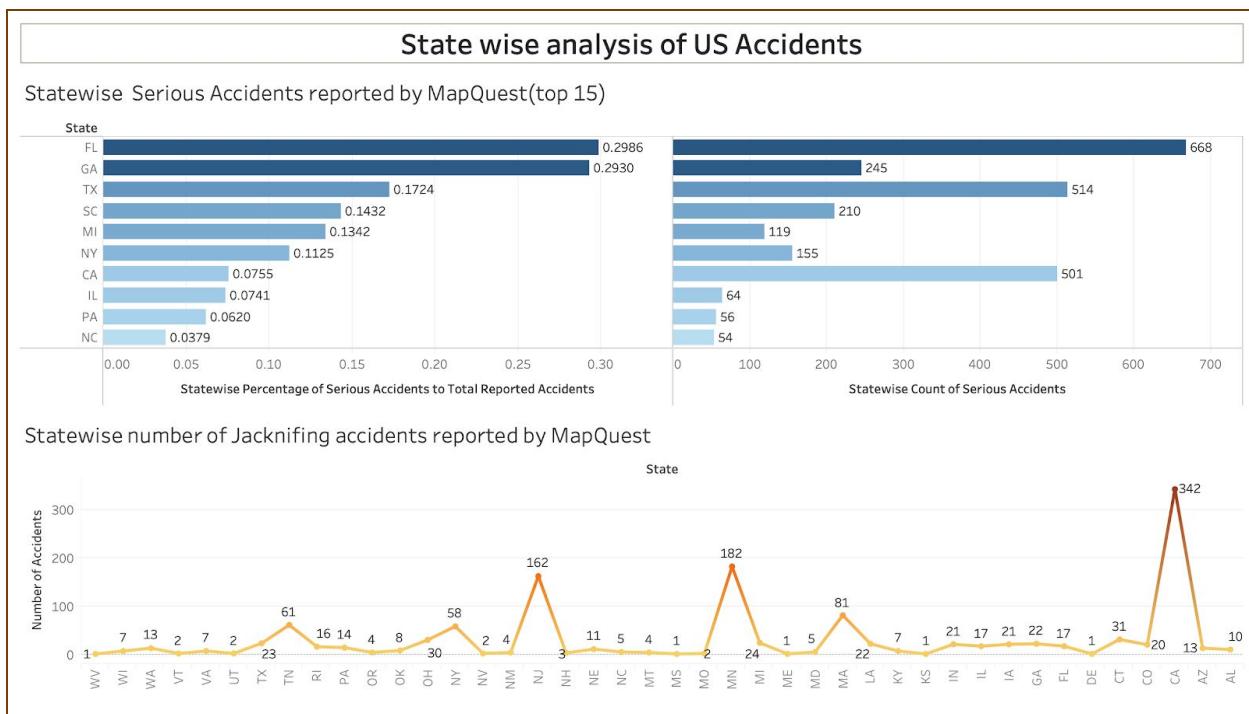


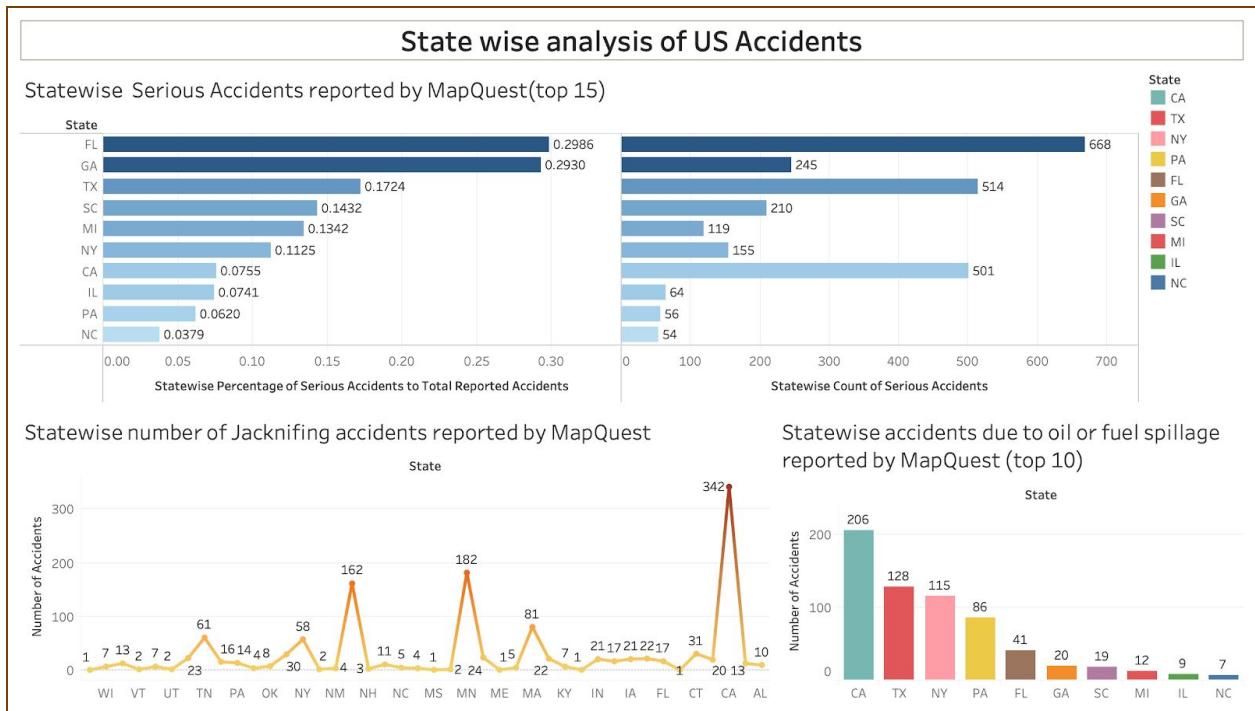
What is the severity of the accidents according to the weather? So if you see the last chart which tells if the weather is Mostly Cloudy then the severity of accidents are high for maximum of the accidents and same for overcast.

So an alert can be given to the public when the weather is cloudy or overcast to prevent the accidents.

Dashboard 11 - Impact of weather on Accidents in US (2016-2019)

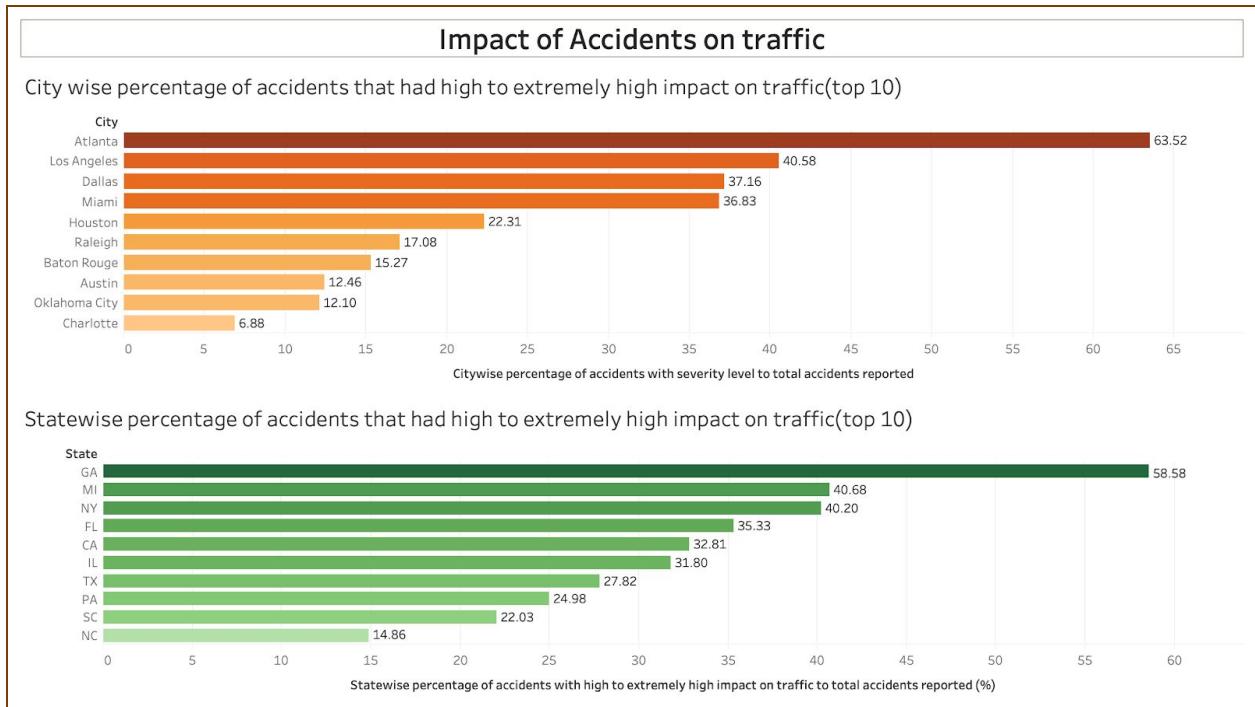
If we check the number of accidents according to each state most of the serious accidents reported mapQuest are in FL the count of the accidents are 668 and the percentage is 0.2986. GA is also almost equal to the FL in terms of the number of serious accidents reported by mapQuest.

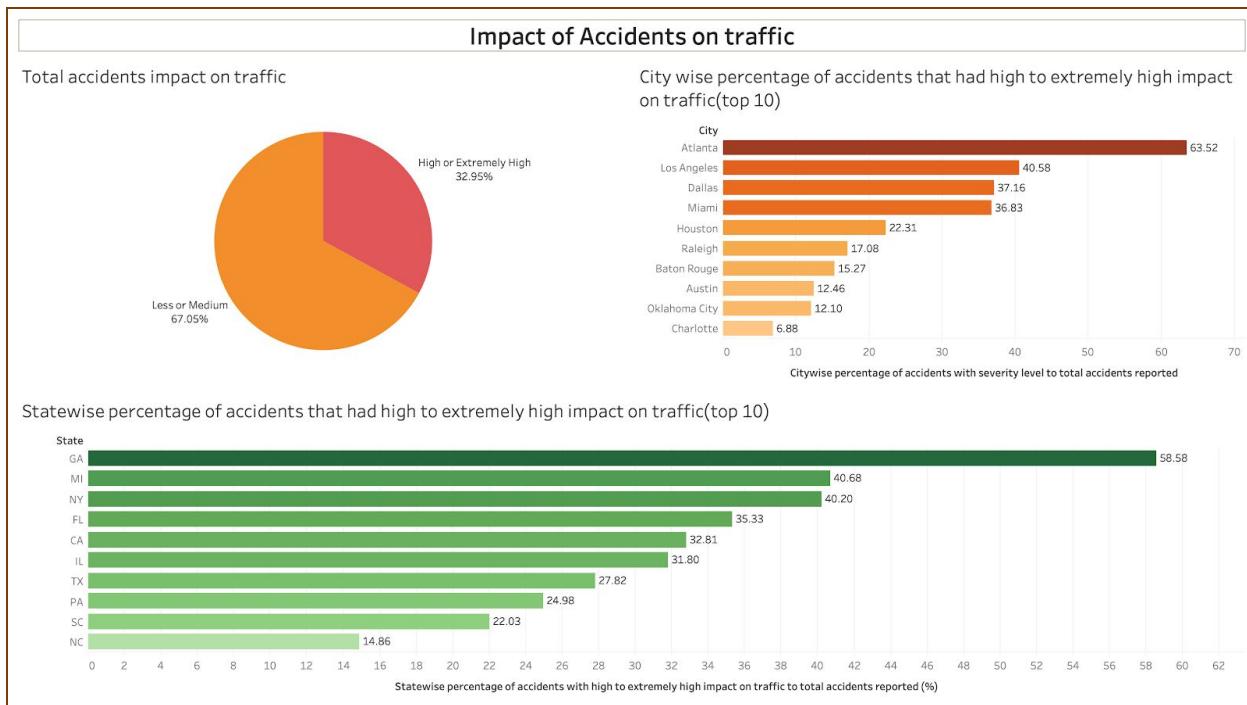




We can see that the Number of jackknifing accidents are very high in the state of CA and numbers of accidents due to oil or fuel spillage are also in CA.

Dashboard 12 - Impact of Accidents on Traffic of US (2016-2019)

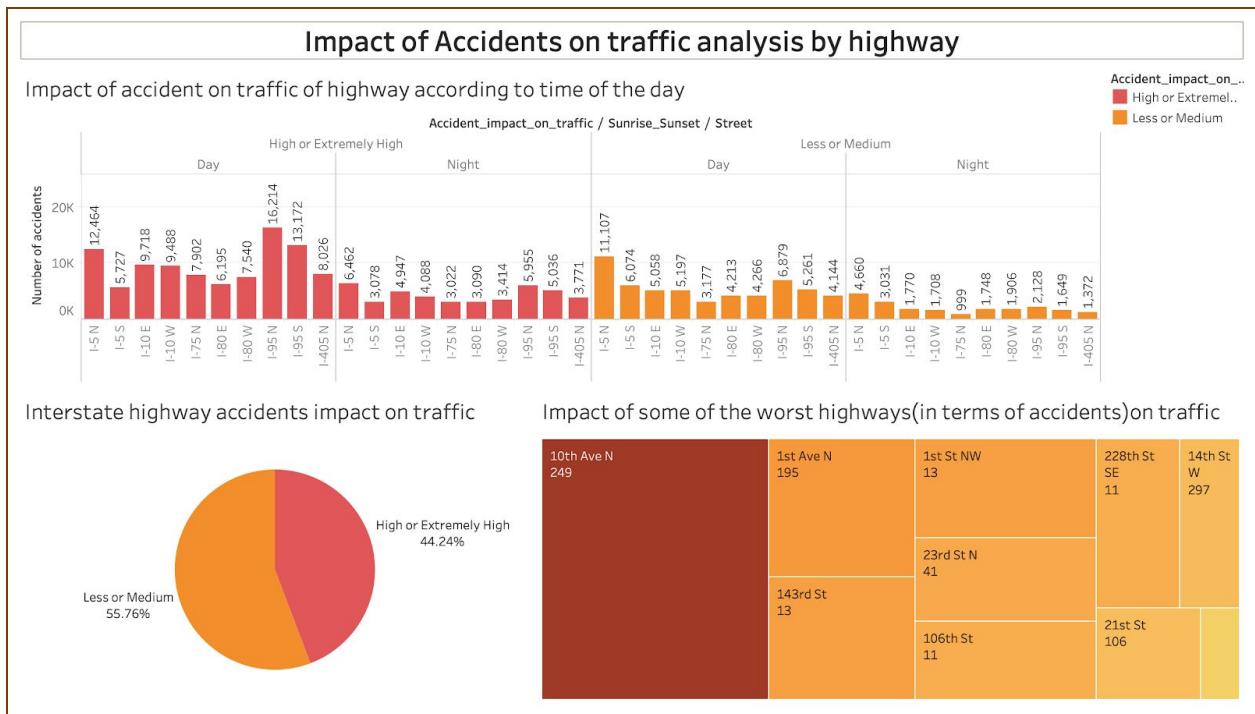
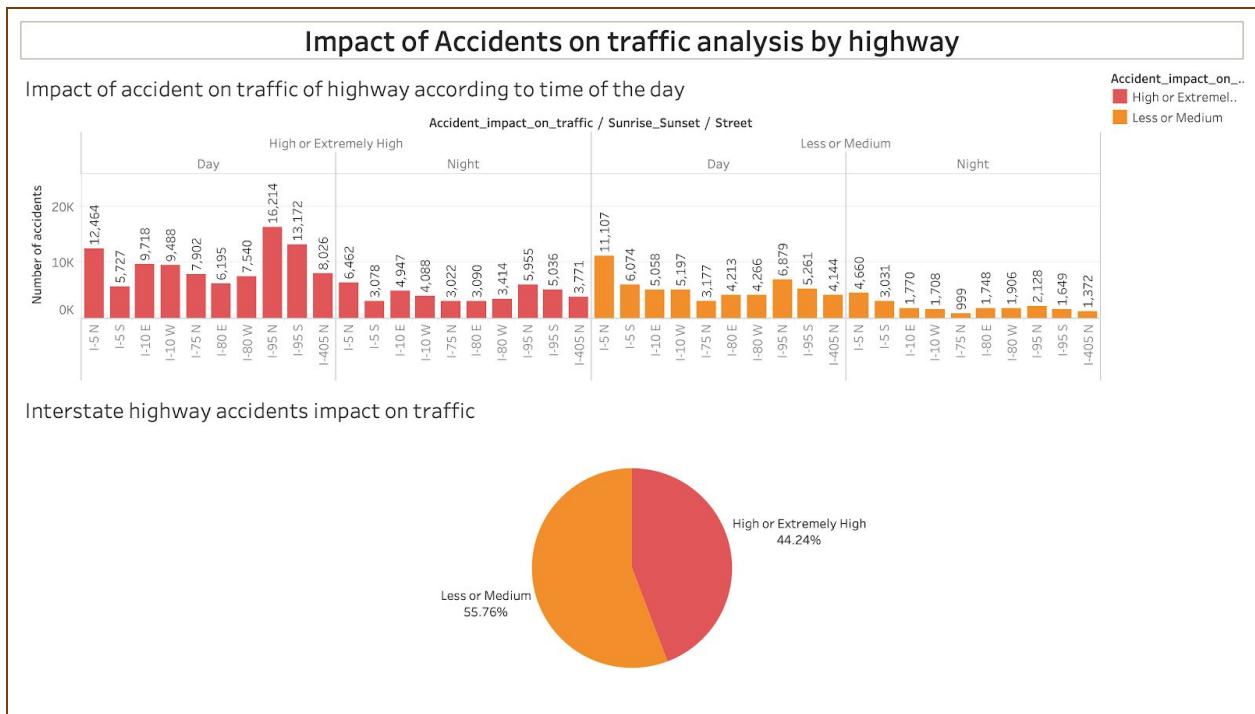




There are 32.95% of accidents which have high impact on traffic of the US and 67.05% of the Accidents have less or medium impact on traffic. If we look at city wise traffic impact by accidents, Atlanta has the most of the traffic (63.52%) caused by the accidents. If we look at the State wise GA has the most of the traffic(58.58%) caused by the accidents. The city Los Angeles has 40.58% of their traffic caused by the accidents. So in order to decrease the traffic we have to decrease the number of accidents in each state and city.

Dashboard 13 - Impact of Accidents on Traffic by Highway's of US (2016-2019)

By looking at the chart Impact of the accident on traffic of highway according to time of the day we can say that most of the high impact accidents are happening in day time and the worst street is I-95 N. Majority of accidents on Interstate highways cause high or extreme high impact on traffic.

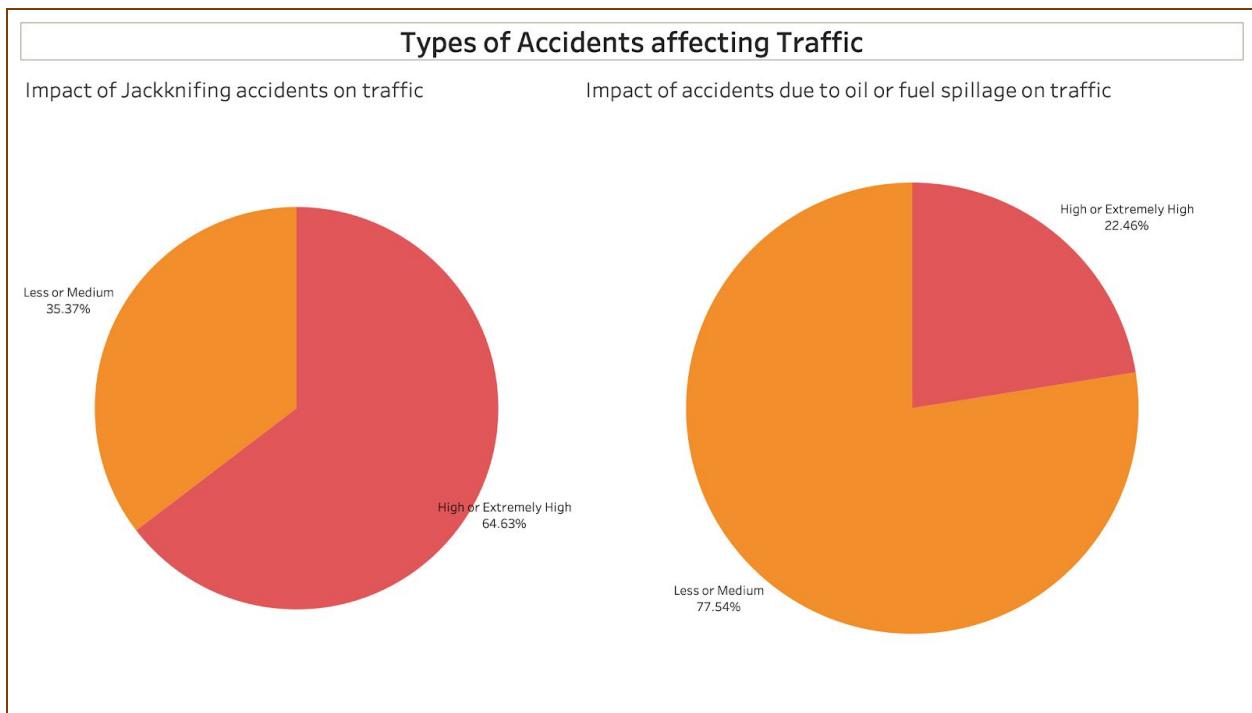


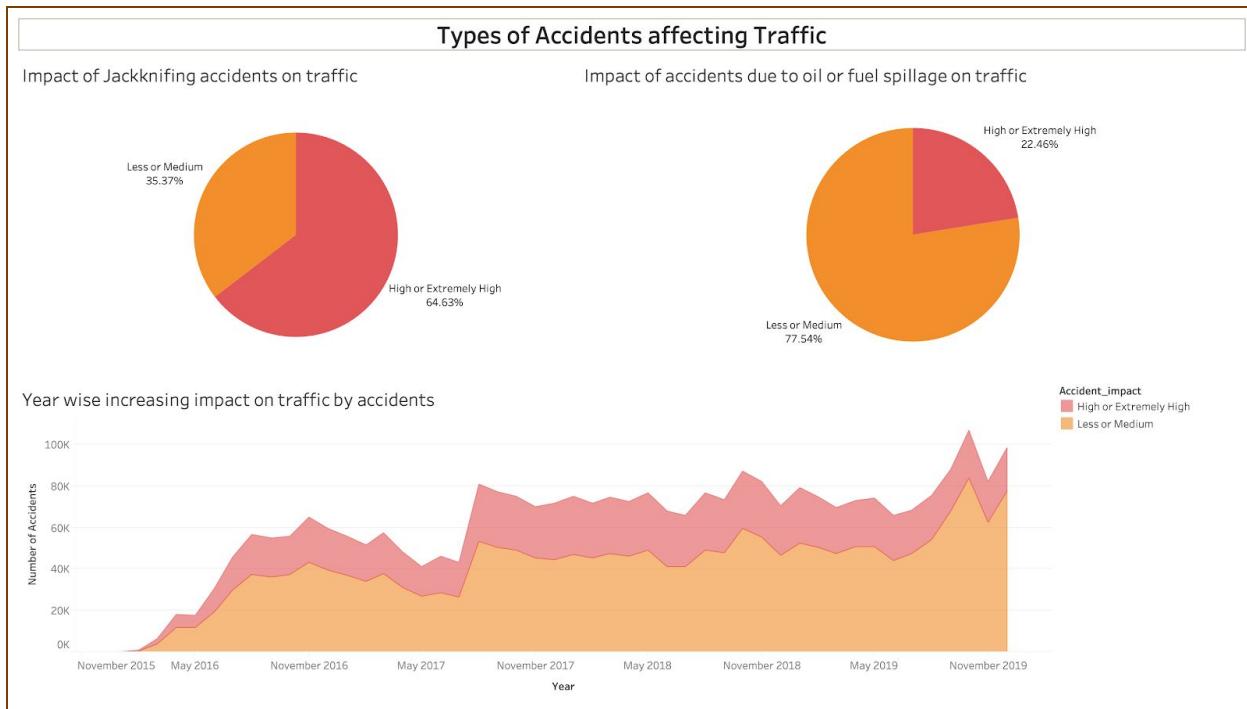
Highway 10 Ave N has the worst impact on the traffic and 1st Ave N, 143rd st as well. We can summarize that highways in the day time are more dangerous and Interstate highways have more number of high impact accidents.

In order to prevent accidents the number of vehicles can be restricted on these highways in day time.

Dashboard 14 - Types of Accidents Affecting Traffic in US (2016-2019)

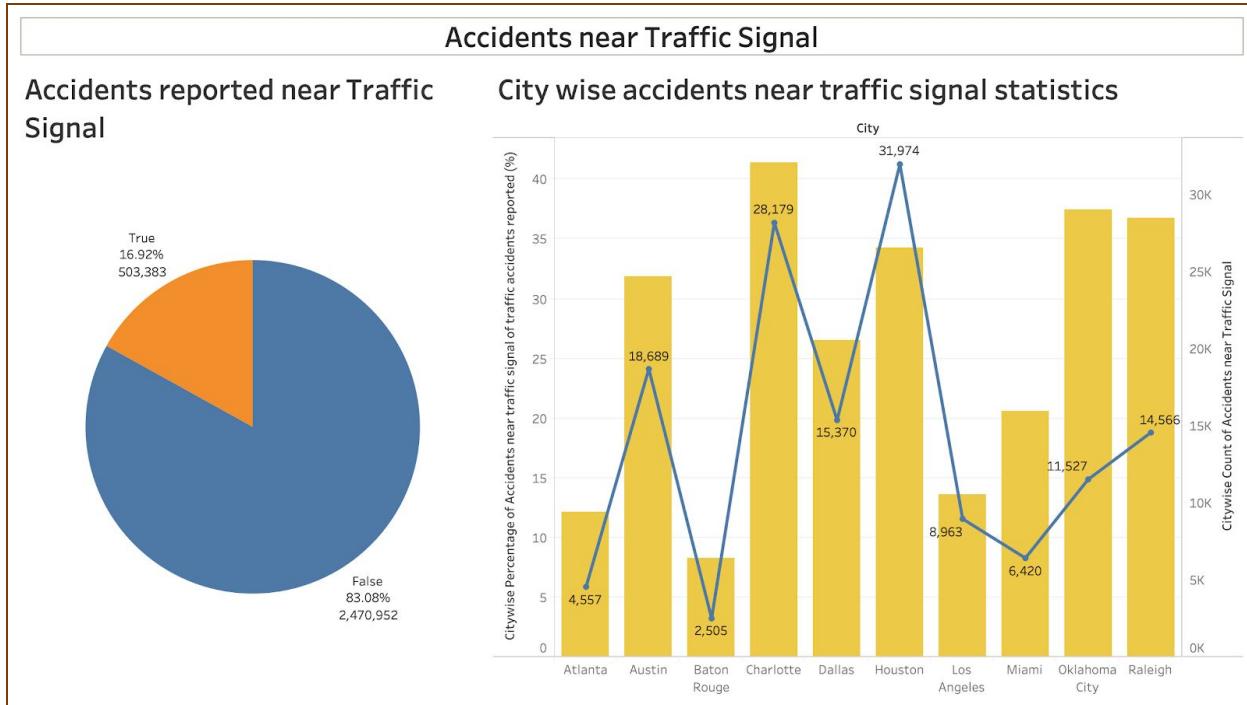
The jackknifing accidents impact the extremely high on the traffic of the US compared to impact of accidents due to oil or fuel spillage. 64.63% of Jackknifing Accidents affects the extremely high on traffic of US whereas 77.54% of accidents due to oil or fuel spillage impacts the low on the traffic only 22.46% are impacting high.



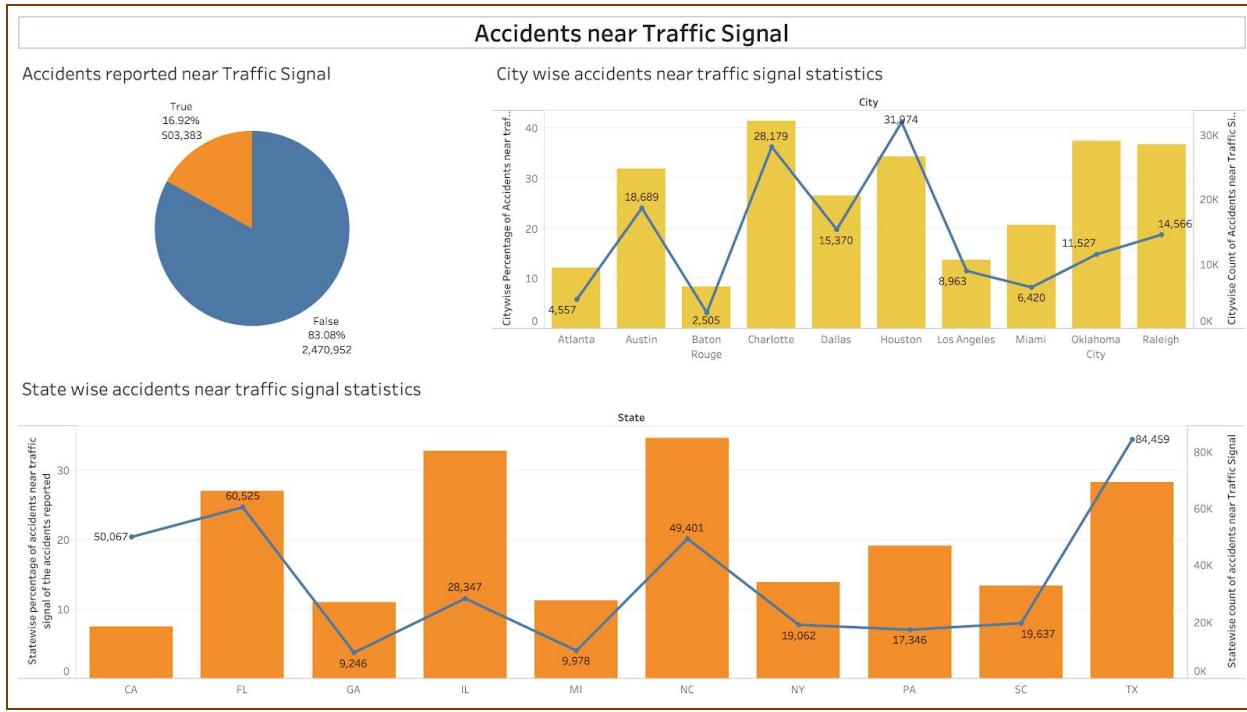


Every year traffic is getting increased by the number of accidents. In the year of 2017 the impact of accidents on traffic was very high whereas in 2016 the impact on traffic was very less.

Dashboard 15 - Accidents near Traffic Signals in US (2016-2019)



This Dashboard states that only 16.92% of the accidents happened near traffic signals in the US. which means traffic signals are helpful in reducing the number of accidents.



The place which has a high number of accidents requires a traffic signal system. If we look at the city wise and state wise accident near traffic signal chart, The Charlotte city has the highest percentage of accidents near traffic signals and the number of accidents near traffic signals are high in the city Houston. NC state has the highest percentage of accidents near traffic signals TX has the highest number of accidents near traffic signals.

Insights Gained

1. Yearly number of accidents are increasing.
2. Month of october has more number of accidents.
3. Weekdays have more accidents than the weekends.
4. Got the worst and best cities and states in terms of accidents.
5. MapQuest is reporting the maximum number of accidents.
6. Most of the accidents are happening in day time.
7. US/Eastern has a maximum number of accidents, US/central and US/Pacific have high severity accidents.

8. Got worse highways and worse streets in terms of accidents.
9. Got the deadliest time of the day in terms of accidents.
10. Most of the accidents are happening when the weather is not clear.
11. Most of the accidents are making an impact on traffic.
12. Jackknifing accidents are making more impact on traffic than other accidents.
13. Very few accidents are happening near traffic signals.

This Includes all the answers of the question raised before the analysis.

Predictive Dashboards

Variables Used for Predictive Analysis

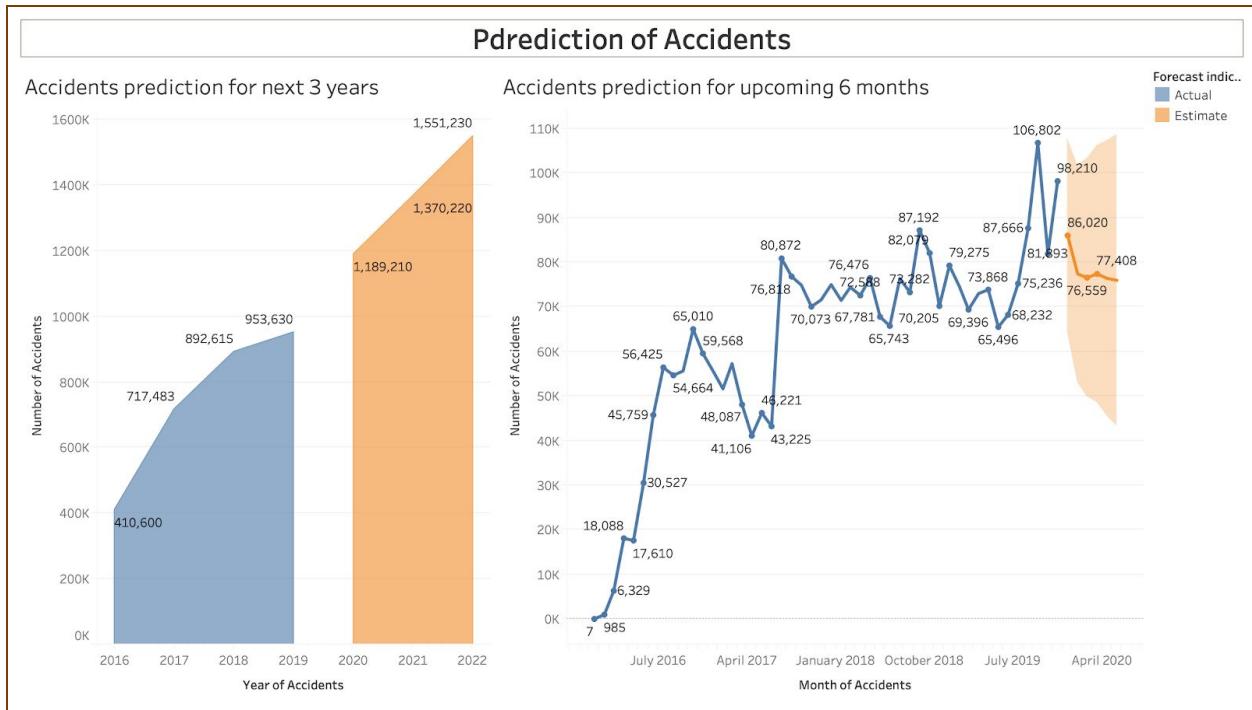
As I am doing the prediction in tableau the time series is must so -

- Date of Accidents
- Number of records

These are mandatory fields apart from this the list of attributes I have used is bellow -

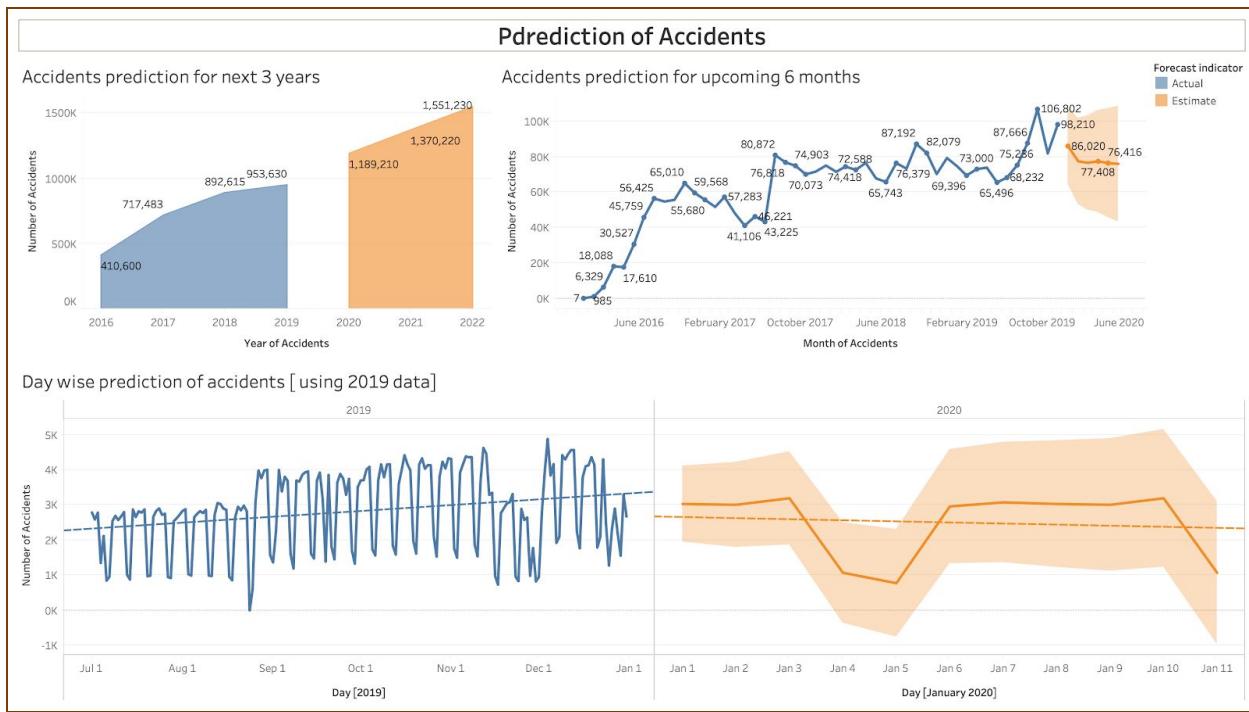
- Timezone of accident
- Time of the day (sunrise - sunset) of accident
- Severity
- City, State

Dashboard 16 - Prediction of Accidents in US



This Dashboard represents the prediction of accidents for upcoming years and months. The first chart is predicting the number of accidents for the upcoming year 2020, 2021 and 2022 using accidents data for 2016 to 2019. According to the prediction the number of accidents are going to be increased in the upcoming years 2022 will have the highest number of accidents till now.

Another graph predicts the number of accidents for the upcoming 6 months. According to the prediction for the upcoming 6 months the number of accidents are going to be decreased.

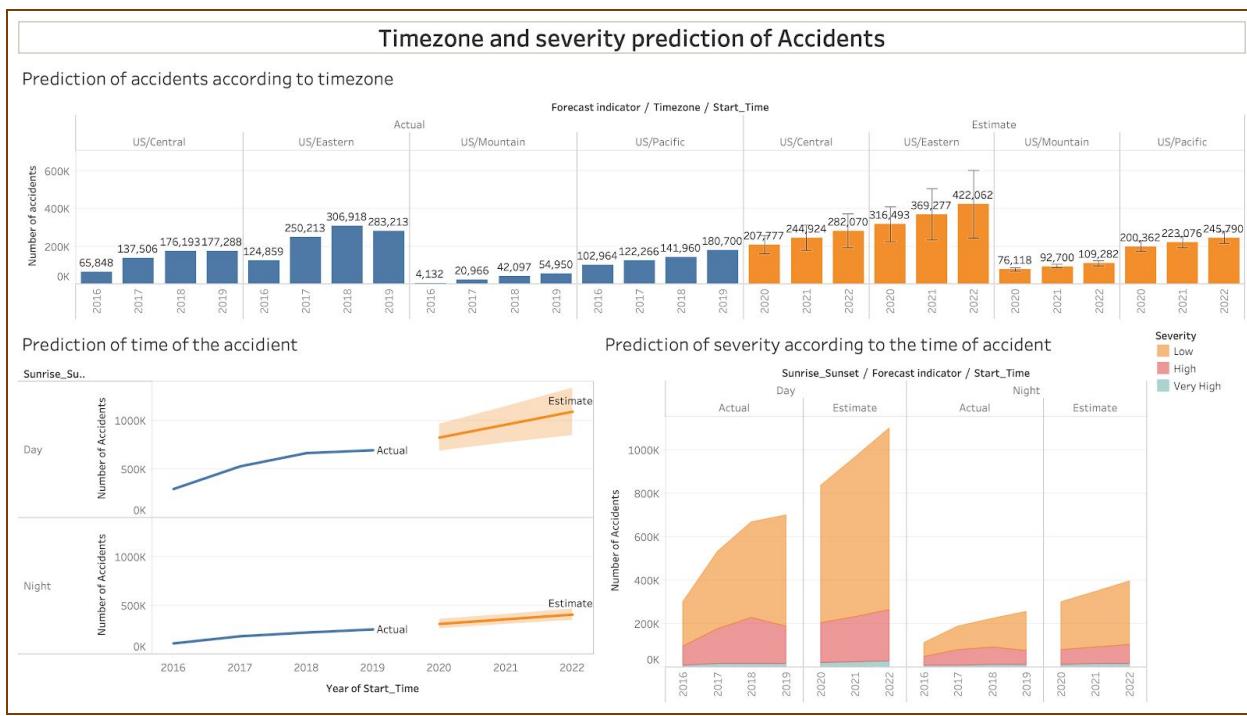
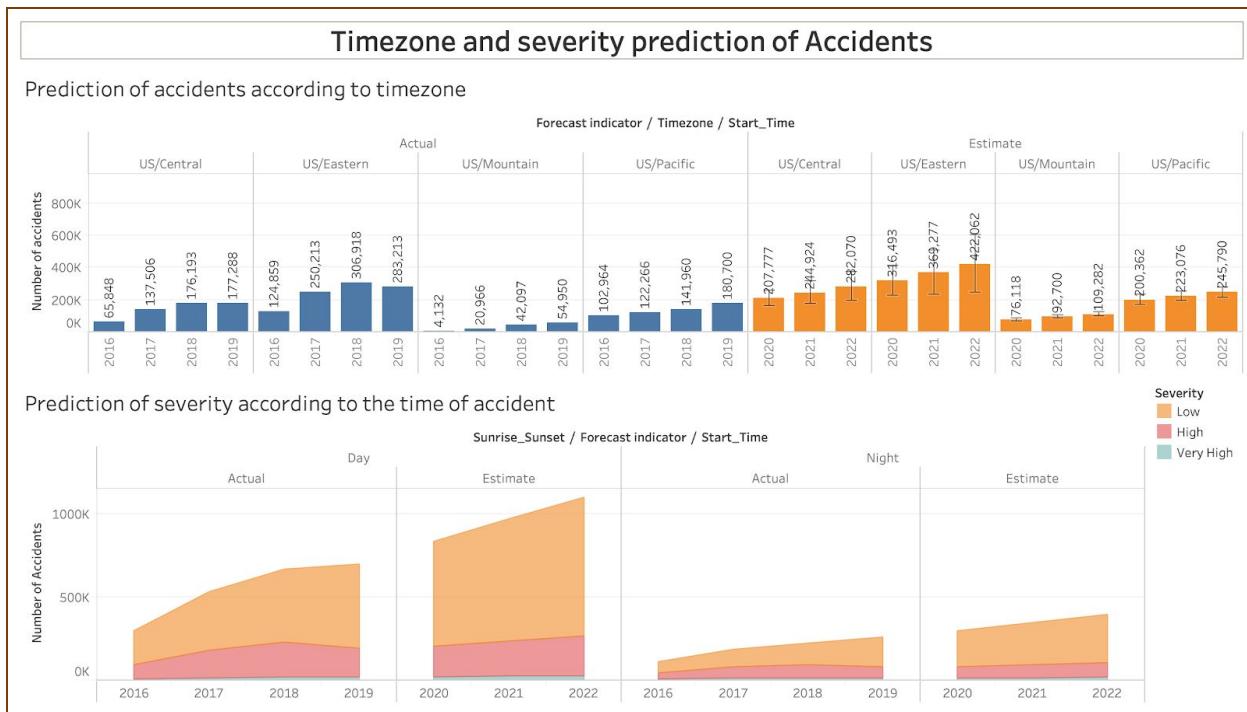


The Day wise prediction of accidents chart depicts that the number of accidents for upcoming days jan1 to jan11 will reduce as the trend line is also going slightly down. On jan 11 there are very few accidents predicted. The day wise prediction is being done using the data from the year 2019.

Dashboard 17 - Timezone and severity prediction of Accidents in US

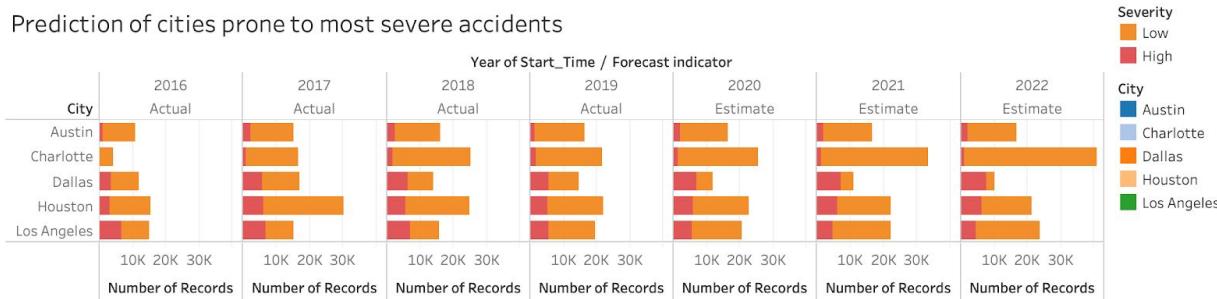
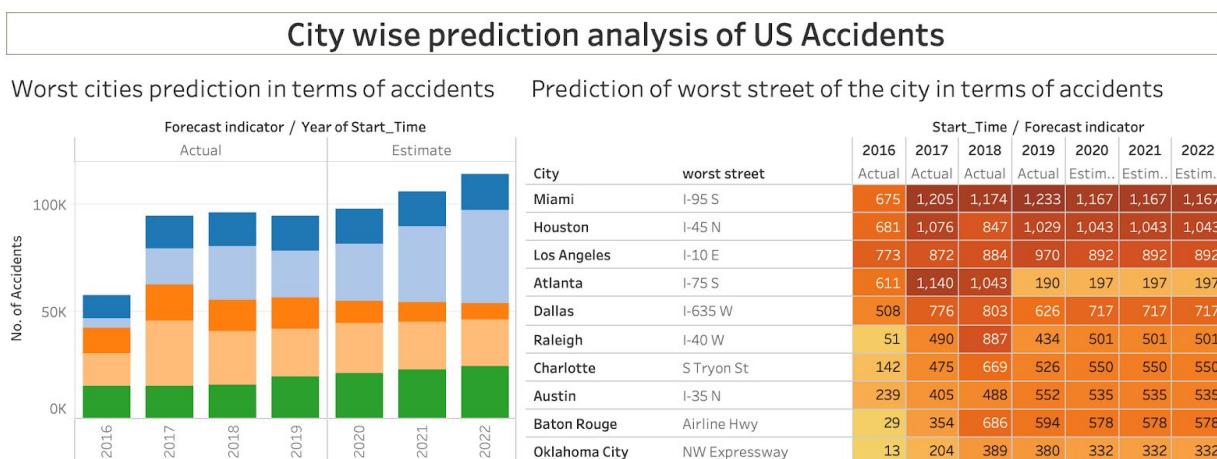
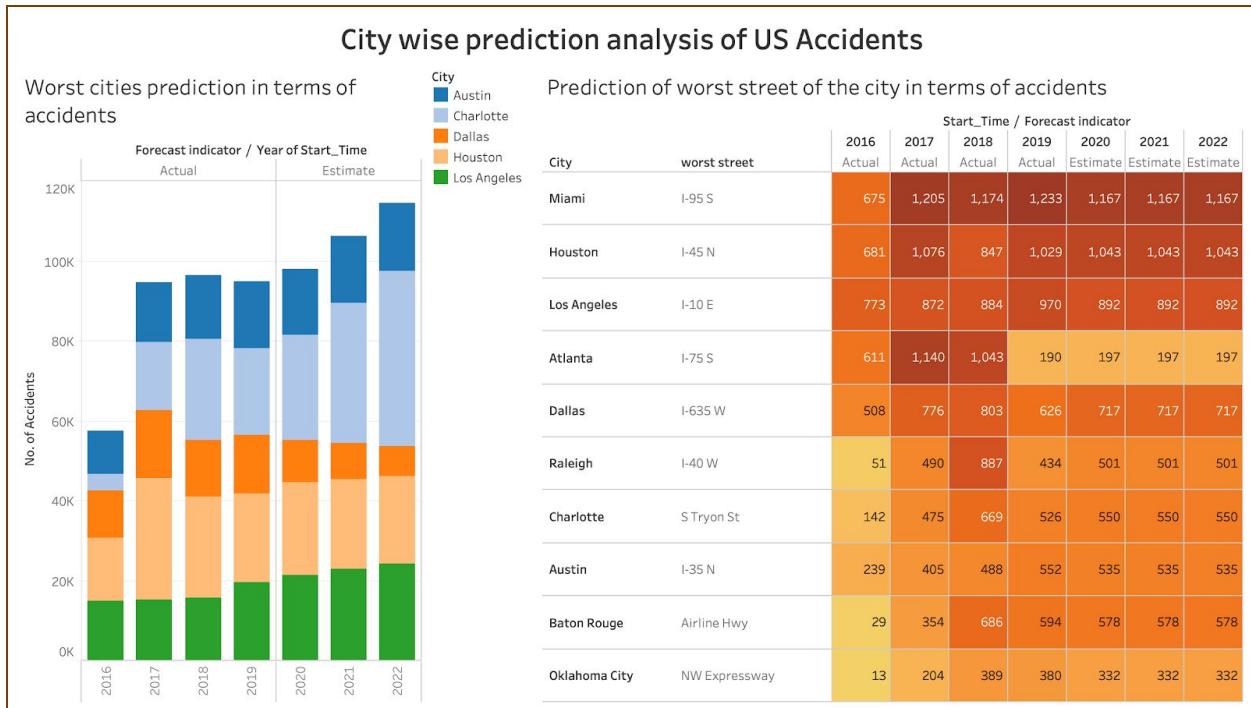
In the first chart of this Dashboard we are predicting the number of accidents for each timezone of the US. As we can see the chart we had the highest number of accidents in US/Eastern and our chart is also predicting the highest number of accidents in the year of 2020, 2021, 2022 in US/Eastern.there are very less number of accidents in US/Mountain for upcoming year.

The other chart predicts the severity of the accident according to the time of the accident. There are very less high severity accidents predicted for night time whereas for the day time the number of accidents for upcoming years with the high severity are more compared to night time.



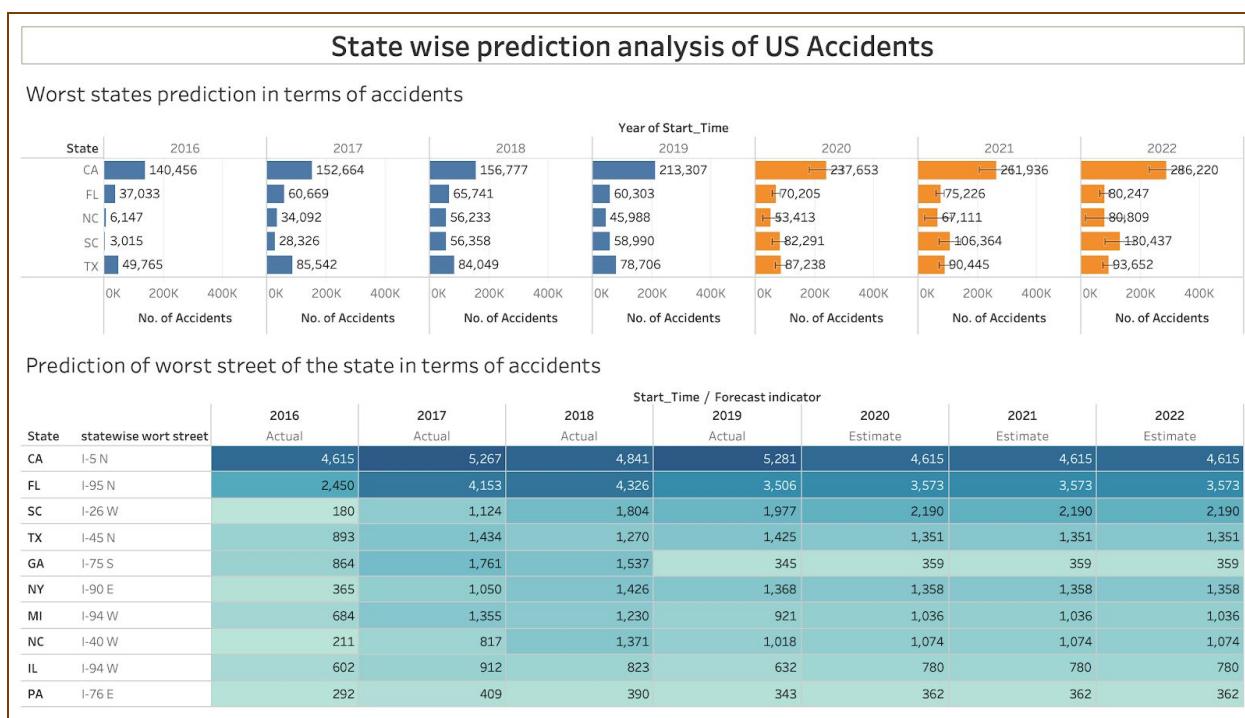
The prediction of time of the accidents tells that there will be more accidents in the day time (more than 1000k).

Dashboard 18 - City wise prediction analysis of US Accidents

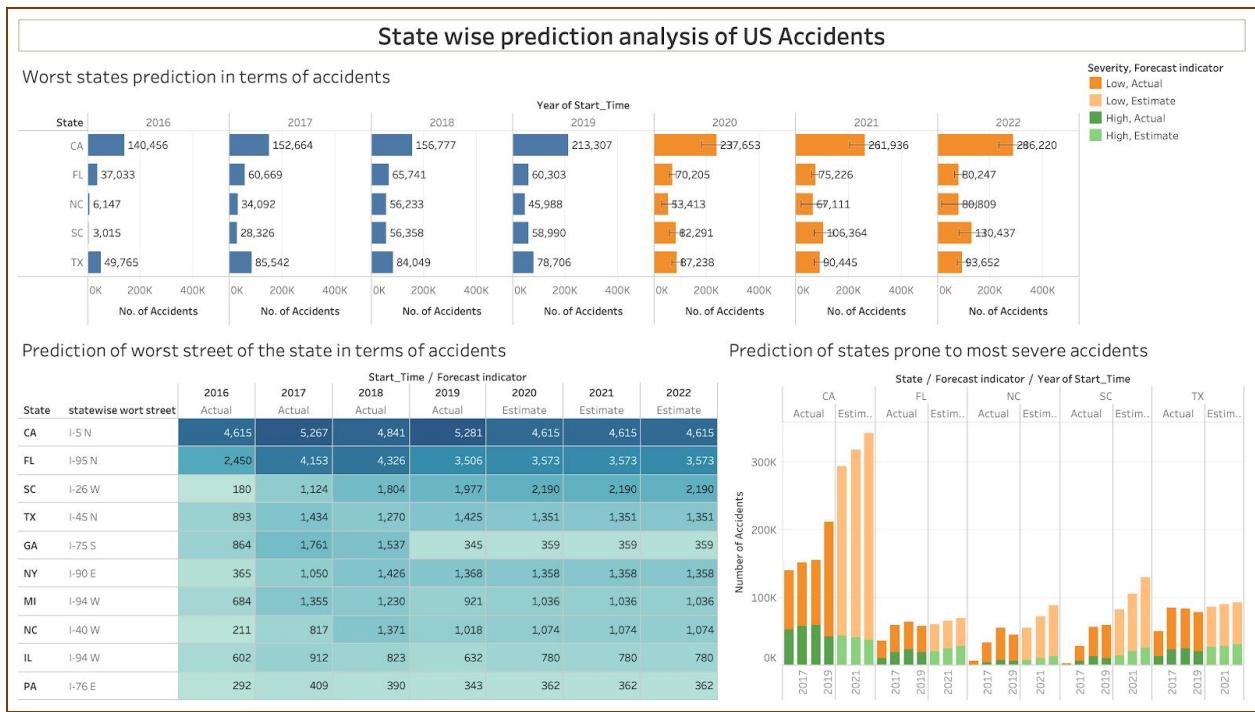


Here the prediction of the worst cities in terms of accidents chart depicts that Charlotte will have the maximum number of accidents in the upcoming year 2020, 2021 and 2022. After Charlotte Los Angeles and Houston will be the worst cities for the upcoming year. The worst street will be I-19 S of city miami with the highest number of accidents 1167 for the upcoming 3 years. And the most high severity accidents are predicted in the city Dallas for the Upcoming 3 years. And most Low severity accidents are predicted in the city Charlotte.

Dashboard 19 - State wise prediction analysis of US Accidents



State wise prediction analysis states that the CA will have the most number of accidents in the upcoming year of 2020, 2021 and 2022. And the highest number of accidents will be 286220 in 2022. The worst street of the state will be I5-N from the state CA for the upcoming years and the number of accidents on that street will be 4615. I-95 N from state FL will also be the worst street with the number of accidents 3573.



The most severe accidents will also happen in the state of CA for the upcoming years. And the low severity accidents will also be in CA 303881 number of low severity accidents will be happening in the year of 2022. So CA is the most dangerous state in terms of accidents.