# Road Accident Analysis in USA

Aswini Pusuluri

Department of Applied Sciences, San Jose State University

DATA 230: Data Visualization

Prof. Andrew Bond

May 17, 2022

Table of Contents

## Abstract

Road Safety has become a crucial factor in determining the day-to-day lives of every individual. Improvement in living standards along with social and economic statuses has introduced new trends in the motor vehicle sector, thus resulting in an increased number of vehicles on the road each year. Road accidents have become a major unnatural global problem claiming staggering numbers of fatalities every year. Despite the remote work options and reduced traffic trends in major cities in the last couple of years, the traffic mortality rate is still on the rise. The death estimates from motor vehicle crashes in the first half of 2021 have risen by 18.4% over the preceding year with an estimate of 20160 deaths. This is a shocking revelation to the concerned authorities as road accidents claim a fair share of the economy in the form of shattered lives, disabled persons, cost for the treatment of injuries and repair costs to the roads and vehicles involved. There are also intangible effects that prevail from road accidents that might not reflect in any of the statistics, such as the effect on the mental and physical wellbeing of the individuals involved in accidents, long term, or permanent disabilities, etc. As per WHO statistics 1.35 million succumb to road accidents every year globally of which 38,000 deaths occur in the United States involving a total of 280 million vehicles and 227.5 million drivers in the US. The impact on the US economy is estimated as high as $800 billion each year. Although there are numerous causes for road accidents, most of them would broadly fall under the categories which are humanly controllable and can be avoided by exercising caution and necessary safety measures in place. The most common causes were the behavior of the drivers, weather, age, and time.

The main aim of this project is to perform an analysis on road accidents in the United States caused by major factors such as climatic conditions, location, traffic flow interruptions,

time period (day or night), and determine which one has a significant effect on the economy and lives of people involved. This analysis would further help in identifying the key factors that are responsible for the cause of accidents and suggest preventive and precautionary measures to curb road accidents.

## Objectives

The objective of this project is to analyze the accident data of USA and draw insights about what factor plays crucial role in increasing the number of accidents every year. To achieve these three dashboards has been created. The first dashboard is to understand the severity analysis of accidents. Second dashboard provides insights about accident trends over time and the third dashboard provides insights about the conditions and factors causing Accidents.

**The following are the visualizations proposed as part of this analysis:**

- States with Highest number of accidents
- Analysis of crash severity levels
- Distribution of accidents across US time zones
- Distribution of severity across US time zones
- Top accident-prone cities
- Major accident-prone zones
- Overview of total number of accidents across the cities of US
- Most accident-prone day of the week
- Accidents at different times of the day
- Trendline of accidents in 2021
- Severity of accidents recorded by temperature
- Accidents in different weather conditions
- Analysis on which side of the road most accidents occur
- Accidents based on density of population
- Major highways on which most accidents happened

## Dataset Description

This analysis will be performed using a public dataset from Kaggle which has road accidents details from 49 states of the United States. This data was collected from multiple APIs that provide continuous streaming data from 2016 to 2020. These APIs transmit traffic events collected by a range of institutions inside the road networks, including US transportation authorities, traffic cameras and sensors, and law enforcement agencies. This dataset reportedly contains around 2.8 million accident records. The dataset used for this project has 47 attributes with the description as listed below

**Figure 1**

*Sample of US Accidents Dataset*

| ID | Severity | Start_Time | End_Time | Start_Lat | Start_Lng | End_Lat | End_Lng | Distance(r | Descriptio | Number | Street | Side | City | County | State | Zipcode | Coun |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| A-1 | 3 | 2/8/2016 0:37 | 2/8/2016 6:37 | 40.10891 | -83.0929 | 40.11206 | -83.0319 | 3.23 | Between Sawmill Rd, | | Outerbelt | R | Dublin | Franklin | OH | 43017 | US |
| A-2 | 2 | 2/8/2016 5:56 | 2/8/2016 11:56 | 39.86542 | -84.0628 | 39.86501 | -84.0487 | 0.747 | At OH-4/OH-235/Exit | | I-70 E | R | Dayton | Montgom | OH | 45424 | US |
| A-3 | 2 | 2/8/2016 6:15 | 2/8/2016 12:15 | 39.10266 | -84.5247 | 39.10209 | -84.524 | 0.055 | At I-71/US-50/Exit 1 | | I-75 S | R | Cincinnati | Hamilton | OH | 45203 | US |
| A-4 | 2 | 2/8/2016 6:51 | 2/8/2016 12:51 | 41.06213 | -81.5378 | 41.06217 | -81.5355 | 0.123 | At Dart Ave/Exit 21 - | | I-77 N | R | Akron | Summit | OH | 44311 | US |
| A-5 | 3 | 2/8/2016 7:53 | 2/8/2016 13:53 | 39.17239 | -84.4928 | 39.17048 | -84.5018 | 0.5 | At Mitchell Ave/Exit ( | | I-75 S | R | Cincinnati | Hamilton | OH | 45217 | US |
| A-6 | 2 | 2/8/2016 8:16 | 2/8/2016 14:16 | 39.06324 | -84.0324 | 39.06731 | -84.0585 | 1.427 | At Dela Palma Rd - A | | State Rout | R | Williamsb | Clermont | OH | 45176 | US |
| A-7 | 2 | 2/8/2016 8:15 | 2/8/2016 14:15 | 39.77565 | -84.186 | 39.77275 | -84.1881 | 0.227 | At OH-4/Exit 54 - Acc | | I-75 S | R | Dayton | Montgom | OH | 45404 | US |
| A-8 | 2 | 2/8/2016 11:51 | 2/8/2016 17:51 | 41.37531 | -81.8202 | 41.36786 | -81.8217 | 0.521 | At Bagley Rd/Exit 23 | | I-71 S | R | Cleveland | Cuyahoga | OH | 44130 | US |
| A-9 | 2 | 2/8/2016 14:19 | 2/8/2016 20:19 | 40.70225 | -84.0759 | 40.69911 | -84.0843 | 0.491 | At OH-65/Exit 122 - A | | E Hanthor | R | Lima | Allen | OH | 45806 | US |
| A-10 | 2 | 2/8/2016 15:16 | 2/8/2016 21:16 | 40.10931 | -82.9685 | 40.11078 | -82.984 | 0.826 | At I-71/Exit 26 - Accic | | Outerbelt | R | Westervill | Franklin | OH | 43081 | US |
| A-11 | 2 | 2/8/2016 15:43 | 2/8/2016 21:43 | 39.19288 | -84.4772 | 39.19615 | -84.4734 | 0.307 | At OH-4/Paddock Rd | | I-75 N | R | Cincinnati | Hamilton | OH | 45216 | US |
| A-12 | 2 | 2/8/2016 16:50 | 2/8/2016 22:50 | 39.13877 | -84.5339 | 39.13977 | -84.5343 | 0.072 | At US-52/Hopple St/I | | Hopple St | R | Cincinnati | Hamilton | OH | 45225 | US |
| A-13 | 2 | 2/8/2016 17:27 | 2/8/2016 23:27 | 41.4739 | -81.7042 | 41.47388 | -81.7056 | 0.07 | At US-42/Exit 170 - A | | I-90 E | R | Cleveland | Cuyahoga | OH | 44113 | US |
| A-14 | 2 | 2/8/2016 17:30 | 2/8/2016 23:30 | 39.58224 | -83.6778 | 39.60301 | -83.6373 | 2.59 | Between OH-72/Exit | | I-71 N | R | Jamestow | Greene | OH | 45335 | US |
| A-15 | 3 | 2/8/2016 18:11 | 2/9/2016 0:11 | 40.15179 | -81.3126 | 40.15175 | -81.3127 | 0.004 | At Shipley | 48999 | Titus Rd | R | Freeport | Guernsey | OH | 43973 | US |
| A-16 | 3 | 2/8/2016 18:11 | 2/9/2016 0:11 | 40.15175 | -81.3127 | 40.15179 | -81.3126 | 0.004 | At Titus Rc | 22549 | Cadiz Rd | L | Freeport | Harrison | OH | 43973-862 | US |
| A-17 | 3 | 2/8/2016 19:47 | 2/9/2016 1:47 | 39.97241 | -82.847 | 39.98529 | -82.8567 | 1.028 | At OH-16/Broad St/E | | Outerbelt | R | Columbus | Franklin | OH | 43213 | US |
| A-18 | 2 | 2/8/2016 19:47 | 2/9/2016 1:47 | 39.9838 | -82.8566 | 39.9725 | -82.8475 | 0.918 | At I-270 - Accident. | | Outerbelt | R | Columbus | Franklin | OH | 43213 | US |

**ID**: This is used for the unique identification of accident records.

**Severity**: Shows the accident's severity as a number between 1 and 4, with 1 as low severity and 4 as high severity

**Start_Time**: Start time of the accident

**End_Time**: End time of the accident with all the traffic cleared

**Start_Lat & Start_Lng:** Accident starting point GPS coordinates

**End_Lat & End_Lng**: Accident ending point GPS coordinates

**Distance(mi):** The accident's impact on the entire path

**Description:** Description of the accident

**Number:** Street number where the accident took place

**Street:** Street name where the accident took place

**Side:** Gives information regarding the side of the road the accident took place

**City, County, State, Zipcode & Country:** This attribute shows the entire address of the incident

**Timezone:** Timezone depending on the accident's location.

**Airport_Code:** The nearest weather station to the accident site is an airport-based weather station.

**Weather_Timestamp:** Timestamp when the weather details are noted

**Temperature(F):** Temperature at the time of the accident

**Wind_Chill(F):** Windchill at the time of the accident

**Humidity (%):** Humidity during the time of the accident

**Pressure(in):** Air pressure during the time of the accident

**Visibility(mi):** Visibility during the time of the accident

**Wind_Direction & Wind_Speed(mph):** Direction and speed of wind at the time of the accident

**Precipitation(in):** Precipitation measured in inches

**Weather_Condition:** Condition of weather for example snow, rain, etc

**Amenity:** Shows the presence of a local amenity using POI annotation

**Bump:** Indicates the existence of a speed bump or hump in the vicinity using POI annotation

**Crossing:** Shows the existence of a crossing in the immediate vicinity.

**Give_Way:** Shows the existence of a Give_Way in the immediate vicinity.

**Junction:** Shows the existence of a Junction in the immediate vicinity.

**No_Exit:** Shows the existence of a No_Exit in the immediate vicinity.

**Railway:** Shows the existence of a Railway in the immediate vicinity.

**Roundabout:** Shows the existence of a Roundabout in the immediate vicinity.

**Station:** Shows the existence of a Station in the immediate vicinity.

**Stop:** Shows the existence of a Stop in the immediate vicinity.

**Traffic_Calming:** Shows the existence of a Traffic_Calming in the immediate vicinity.

**Traffic_Signal:** Shows the existence of a Traffic_Signal in the immediate vicinity.

**Turning_Loop:** Shows the existence of a Turning_Loop in the immediate vicinity.

**Sunrise_Sunset:** The time of day or night is displayed based on the Sunrise_Sunset.

**Civil_Twilight:** The time of day or night is displayed based on theCivil_Twilight.

**Nautical_Twilight:** The time of day or night is displayed based on the Nautical_Twilight.

**Astronomical_Twilight:** The time of day or night is displayed based on Astronomical_Twilight.

Population data of each state in US is collected from UScensus.com. This dataset has more than 60 columns containing population estimates from 2010 to 2021 and the birth and death count for those years. The population estimates of all states from 2016 to 2021 is being used in this project.

**Figure 2**

*Sample of population dataset*

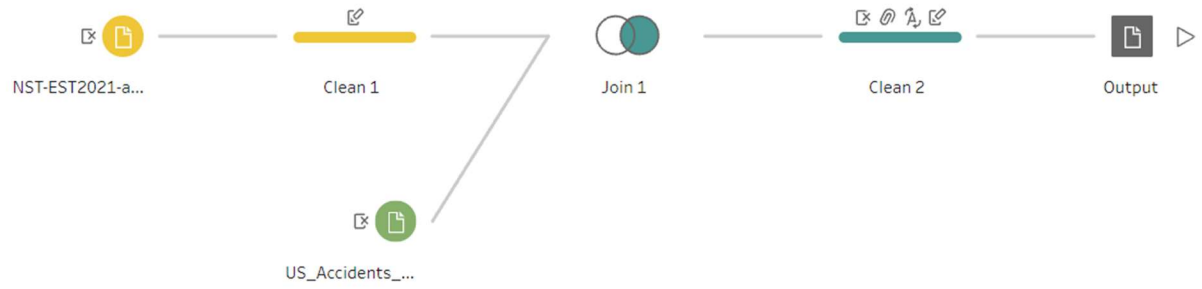
| SUMLEV | REGION | DIVISION | STATE | NAME | CENSUS2010POP | ESTIMATESBASE2010 | POPESTIMATE2010 | POPESTIMATE2011 | POPESTIMATE2012 | POPESTIMATE2013 | POPESTIM | POPES |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 10 | 0 | 0 | 0 | United States | 308745538 | 308758105 | 309321666 | 311556874 | 313830990 | 315993715 | 3.18E+08 | 3.21E |
| 20 | 1 | 0 | 0 | Northeast Region | 55317240 | 55318443 | 55380134 | 55604223 | 55775216 | 55901806 | 56006011 | 56034 |
| 20 | 2 | 0 | 0 | Midwest Region | 66927001 | 66929725 | 66974416 | 67157800 | 67336743 | 67560379 | 67745167 | 67860 |
| 20 | 3 | 0 | 0 | South Region | 114555744 | 114563030 | 114866680 | 116006522 | 117241208 | 118364400 | 1.2E+08 | 1.21E |
| 20 | 4 | 0 | 0 | West Region | 71945553 | 71946907 | 72100436 | 72788329 | 73477823 | 74167130 | 74925793 | 75742 |
| 40 | 3 | 6 | 1 | Alabama | 4779736 | 4780125 | 4785437 | 4799069 | 4815588 | 4830081 | 4841799 | 4852 |
| 40 | 4 | 9 | 2 | Alaska | 710231 | 710249 | 713910 | 722128 | 730443 | 737068 | 736283 | 737 |
| 40 | 4 | 8 | 4 | Arizona | 6392017 | 6392288 | 6407172 | 6472643 | 6554978 | 6632764 | 6730413 | 6829 |
| 40 | 3 | 7 | 5 | Arkansas | 2915918 | 2916031 | 2921964 | 2940667 | 2952164 | 2959400 | 2967392 | 2978 |
| 40 | 4 | 9 | 6 | California | 37253956 | 37254519 | 37319502 | 37638369 | 37948800 | 38260787 | 38596972 | 38918 |
| 40 | 4 | 8 | 8 | Colorado | 5029196 | 5029319 | 5047349 | 5121108 | 5192647 | 5269035 | 5350101 | 5450 |
| 40 | 1 | 1 | 9 | Connecticut | 3574097 | 3574147 | 3579114 | 3588283 | 3594547 | 3594841 | 3594524 | 3587 |
| 40 | 3 | 5 | 10 | Delaware | 897934 | 897937 | 899593 | 907381 | 915179 | 923576 | 932487 | 941 |
| 40 | 3 | 5 | 11 | District of Columbia | 601723 | 601767 | 605226 | 619800 | 634924 | 650581 | 662328 | 675 |
| 40 | 3 | 5 | 12 | Florida | 18801310 | 18804564 | 18845537 | 19053227 | 19297822 | 19545631 | 19845011 | 20209 |

## Data Processing

Several Traffic APIs were used to acquire the US accidents dataset due to which many duplicate and null records were introduced. In order to clean and process the datasets tableau data prep has been used. Initially load both the datasets into tableau and remove the unwanted columns then create a clean step to rename the common column name to merge both the population and accident data. Left join both the datasets using common state column. After joining both the datasets there are some redundant columns to remove them create a clean step. After cleaning the dataset change the datatype of state, city and zip code columns to geolocation and add output step to the dataflow. Run the output flow to generate an output CSV which can be further used for creating visualizations.

**Figure 3**

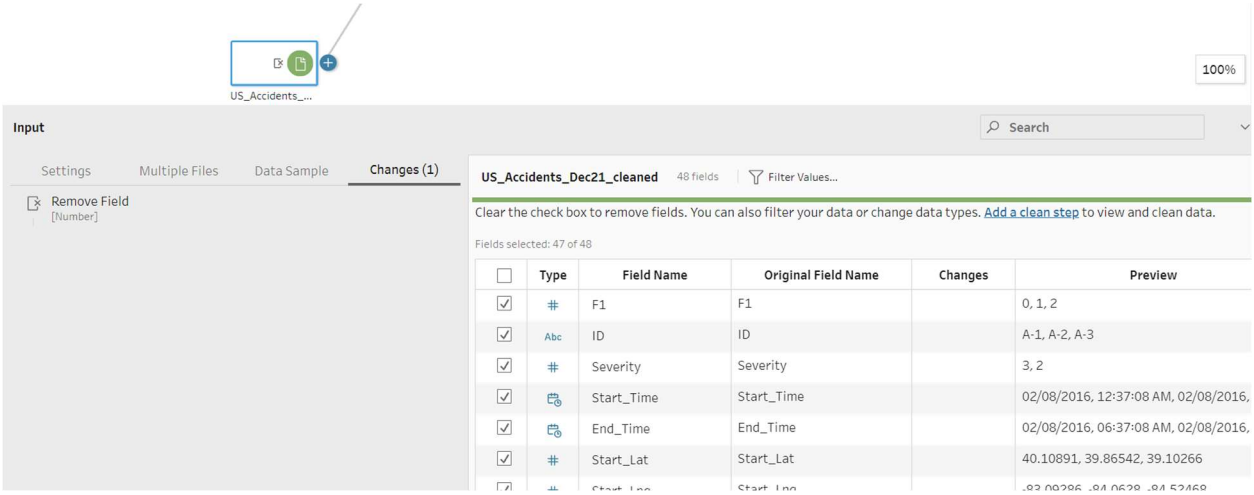
*Data flow for preprocessing US accidents and population dataset*



**Dataset1:** US Population dataset

Column number is removed it adds no meaning in analysis of the data

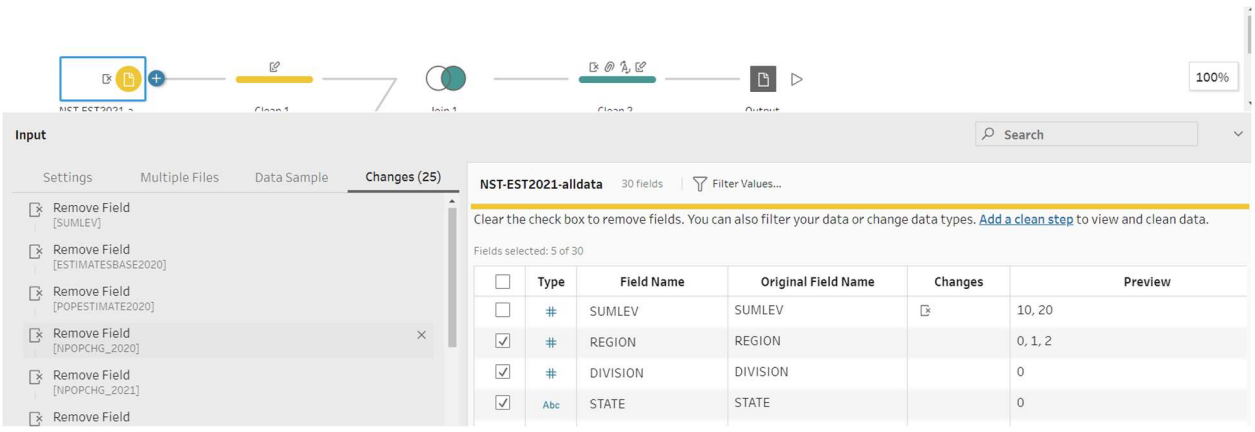
**Figure 4**




**Dataset2**: US population dataset

The population dataset has information related to number of deaths and number of births for each state in united states. That data is irrelevant to this project hence removing the unwanted columns and taking the population estimates data from 2016 to 2021

**Figure 5**



## Figure 6

*Sample data after data processing*

| Start_Lat | Start_Lng | End_Lat | End_Lng | Distance(r | Descriptio | Street | Side | City | County | State | Zipcode | Country | Timezone | Airport_C | Weather_Timestamp | Temperati | Wind_Chi |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 40.10891 | -83.0929 | 40.11206 | -83.0319 | 3.23 | Between S | Outerbelt | R | Dublin | Franklin | OH | 43017 | US | US/Easter | KOSU | 2/8/2016 0:53 | 42.1 | 36.1 |
| 39.86542 | -84.0628 | 39.86501 | -84.0487 | 0.747 | At OH-4/C | I-70 E | R | Dayton | Montgom | OH | 45424 | US | US/Easter | KFFO | 2/8/2016 5:58 | 36.9 | 59.65823 |
| 39.10266 | -84.5247 | 39.10209 | -84.524 | 0.055 | At I-71/US | I-75 S | R | Cincinnati | Hamilton | OH | 45203 | US | US/Easter | KLUK | 2/8/2016 5:53 | 36 | 59.65823 |
| 41.06213 | -81.5378 | 41.06217 | -81.5355 | 0.123 | At Dart Av | I-77 N | R | Akron | Summit | OH | 44311 | US | US/Easter | KAKR | 2/8/2016 6:54 | 39 | 59.65823 |
| 39.17239 | -84.4928 | 39.17048 | -84.5018 | 0.5 | At Mitchel | I-75 S | R | Cincinnati | Hamilton | OH | 45217 | US | US/Easter | KLUK | 2/8/2016 7:53 | 37 | 29.8 |
| 39.06324 | -84.0324 | 39.06731 | -84.0585 | 1.427 | At Dela Pa | State Rout | R | Williamsb | Clermont | OH | 45176 | US | US/Easter | KI69 | 2/8/2016 8:16 | 35.6 | 29.2 |
| 39.77565 | -84.186 | 39.77275 | -84.1881 | 0.227 | At OH-4/E | I-75 S | R | Dayton | Montgom | OH | 45404 | US | US/Easter | KFFO | 2/8/2016 8:18 | 33.8 | 59.65823 |
| 41.37531 | -81.8202 | 41.36786 | -81.8217 | 0.521 | At Bagley | I-71 S | R | Cleveland | Cuyahoga | OH | 44130 | US | US/Easter | KCLE | 2/8/2016 11:51 | 33.1 | 30 |
| 40.70225 | -84.0759 | 40.69911 | -84.0843 | 0.491 | At OH-65/ | E Hanthor | R | Lima | Allen | OH | 45806 | US | US/Easter | KAOH | 2/8/2016 13:53 | 39 | 31.8 |
| 40.10931 | -82.9685 | 40.11078 | -82.984 | 0.826 | At I-71/Ex | Outerbelt | R | Westervill | Franklin | OH | 43081 | US | US/Easter | KCMH | 2/8/2016 15:12 | 32 | 28.7 |
| 39.19288 | -84.4772 | 39.19615 | -84.4734 | 0.307 | At OH-4/P | I-75 N | R | Cincinnati | Hamilton | OH | 45216 | US | US/Easter | KLUK | 2/8/2016 15:47 | 33.8 | 29.6 |
| 39.13877 | -84.5339 | 39.13977 | -84.5343 | 0.072 | At US-52/I | Hopple St | R | Cincinnati | Hamilton | OH | 45225 | US | US/Easter | KLUK | 2/8/2016 16:53 | 35.1 | 32.2 |
| 41.4739 | -81.7042 | 41.47388 | -81.7056 | 0.07 | At US-42/I | I-90 E | R | Cleveland | Cuyahoga | OH | 44113 | US | US/Easter | KBKL | 2/8/2016 17:36 | 33.1 | 24.4 |
| 39.58224 | -83.6778 | 39.60301 | -83.6373 | 2.59 | Between C | I-71 N | R | Jamestow | Greene | OH | 45335 | US | US/Easter | KSGH | 2/8/2016 17:28 | 33.8 | 28.6 |
| 40.15179 | -81.3126 | 40.15175 | -81.3127 | 0.004 | At Shipley | Titus Rd | R | Freeport | Guernsey | OH | 43973 | US | US/Easter | KPHD | 2/8/2016 18:23 | 33.1 | 59.65823 |
| 40.15175 | -81.3127 | 40.15179 | -81.3126 | 0.004 | At Titus Rc | Cadiz Rd | L | Freeport | Harrison | OH | 43973-862 | US | US/Easter | KPHD | 2/8/2016 18:23 | 33.1 | 59.65823 |

## Challenges Faced

US accident dataset is huge with 2.8 million records to load the data and perform each clean step is a time taking process.

## Deliverables

The deliverable of this project includes three dashboards that provides insights about severity analysis of accidents, Accident trends over time, and Conditions causing the accidents.

## Scope of work

The scope of this project covers only 49 states in US. Hence the data for three states Alabama, Alaska and Connecticut were missing this project does not provide insights about the missing states.

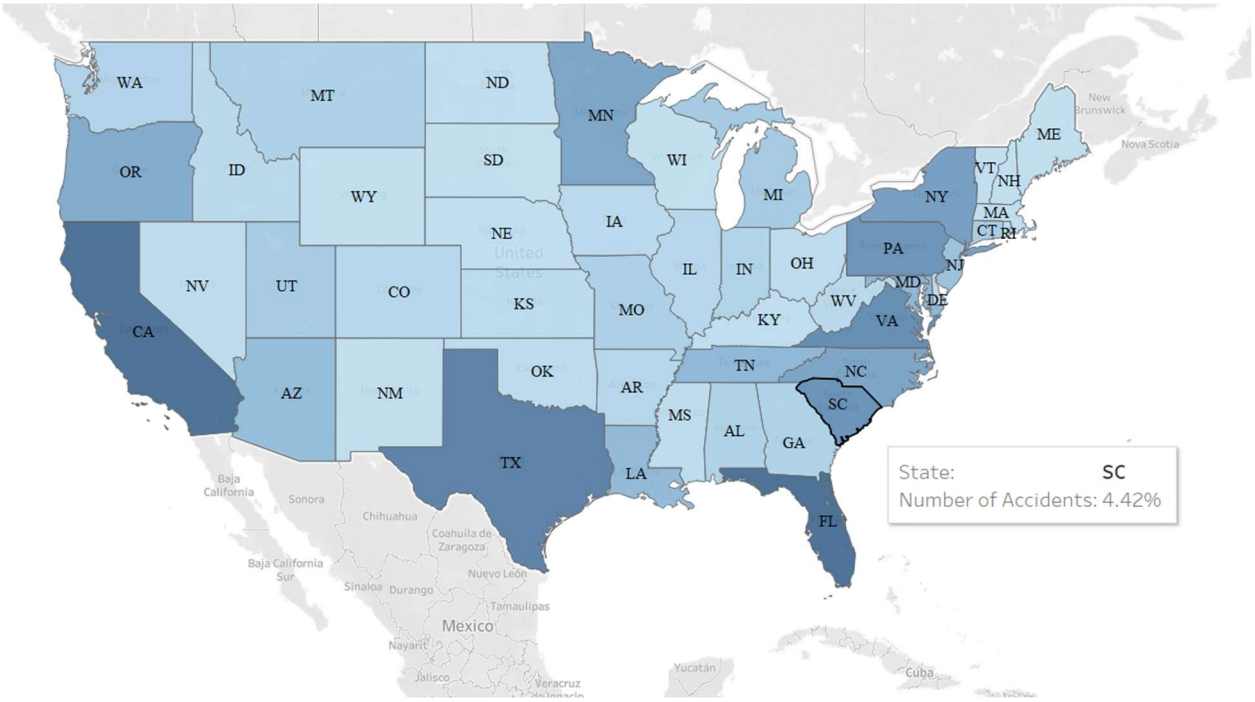
## Visualization Patterns and Diagrams

### Regional Map

As we all know, the foundation of a geographical research is the region, which is a unit of space defined by a shared attribute. Maps illustrating elements of particular interest can be created using the regional maps, and data from several areas can be compared to help detect trends, and patterns that provide useful insights about a particular problem.

In this project regional map is used for comparing the number of accidents recorded for each state from the years of 2016 to 2021. The color range in regional maps helps us in identifying the states with highest number of accidents.

**Figure 7**



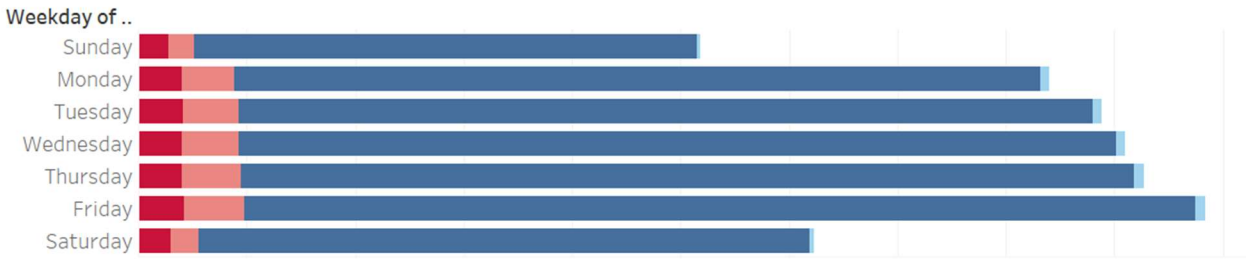
## Bar Chart

The conventional Bar Chart uses horizontal or vertical bars to depict numerical correlations across classes. Each bar's length correlates to the value of that bar. A bar claims ownership of each value. A bar chart is used when you wish to display a distribution of data points or compare metrics between distinct groupings of data.

## Horizontal Bar Chart

In horizontal bar chart category values are represented along y axis and continuous data points are represented along the x axis

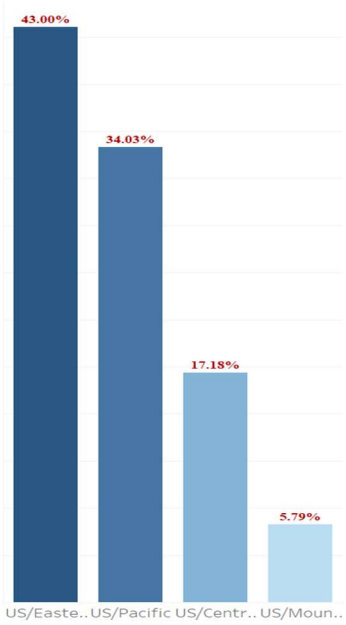
**Figure 8**




## Vertical Bar Chart

For Vertical bar chart continuous data points are represented along y axis and category values are represented along the x axis

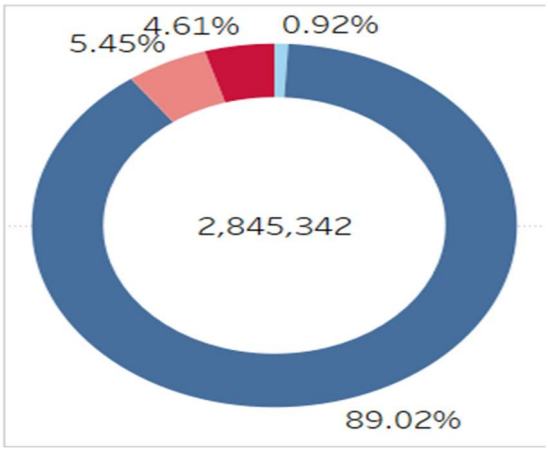
**Figure 9**




**Donut Chart**

Donut chart is often used to depict numerical and categorical proportions, in this each component signifies the percentage of each class or category. It is also used to quickly comparing a few categories and how they connect to the overall.

**Figure 10**

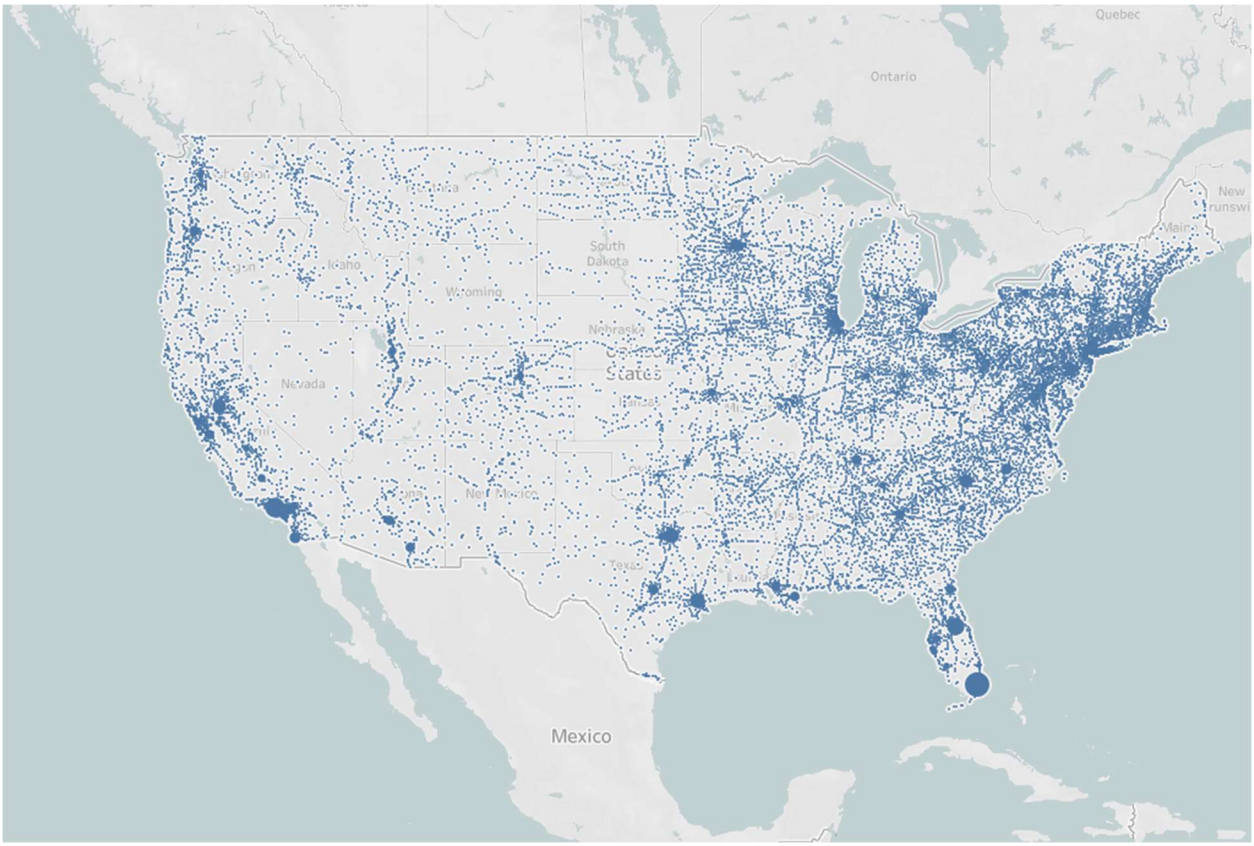


**Dot Maps**

Dot maps are a quick and easy technique to visualize density variations in regional distributions over a terrain. dot map is a useful way to depict actual numbers. Each dot has a value, and the number of dots can be distributed throughout a region according to their geographical distribution.

For this project dot map is used to provide an overview of number of accidents for each city

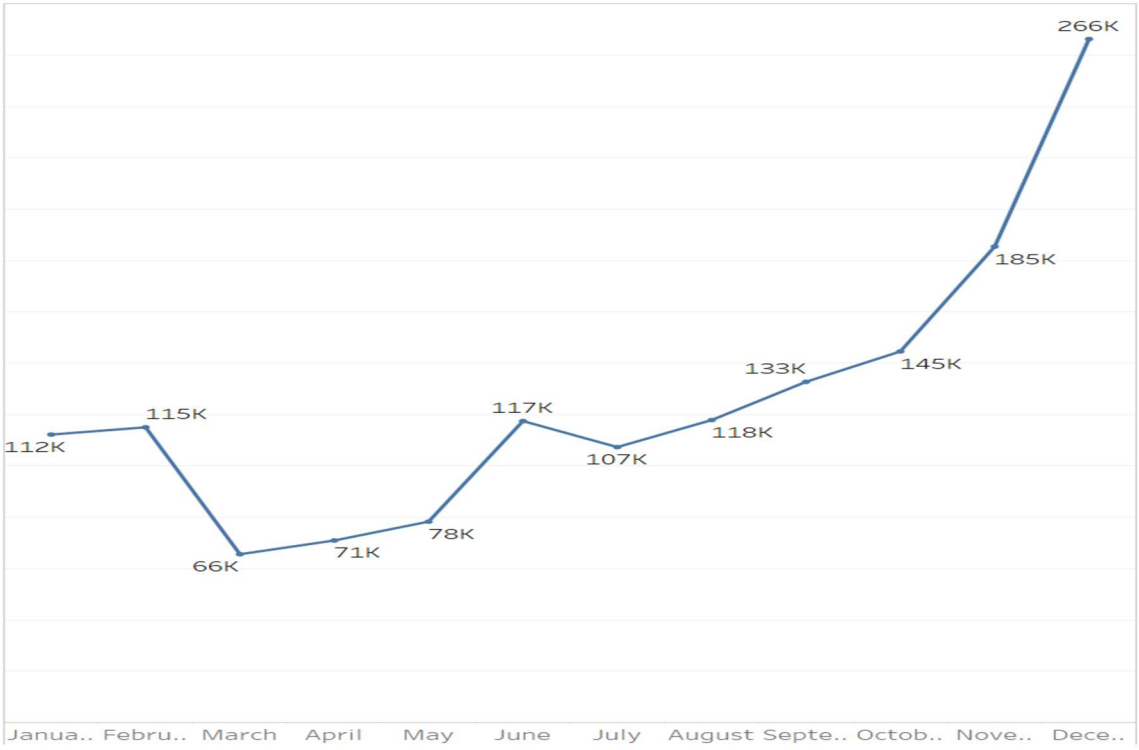
**Figure 11**



## Line Graph

Line graph is used to provide insights about data constantly changing throughout time. When there are modest changes, it is best to use a line graph. In this project line graph is used for visualizing the trend of accidents over the year of 2021.

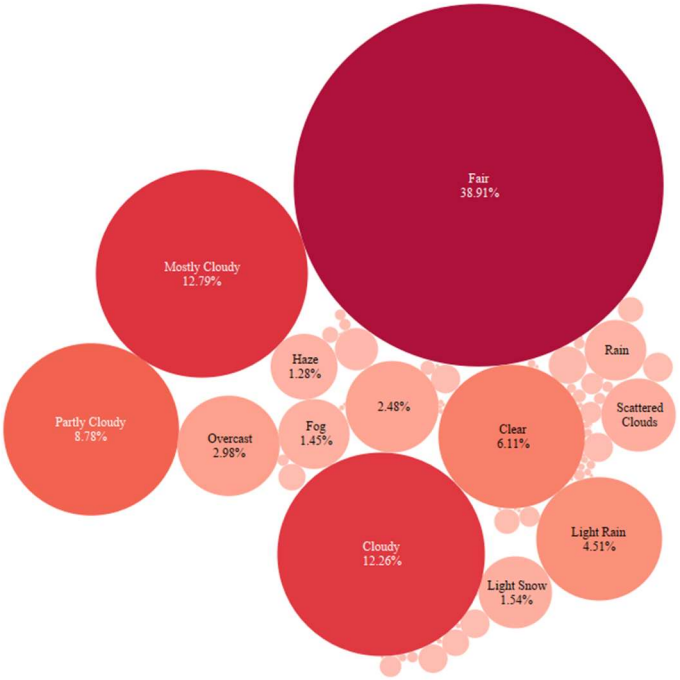
**Figure 12**




## Packed Bubble Chart

To represent information in a group of circles, use packed bubble charts. Each bubble is defined by dimensions, while individual circles are described by measures. For this project each circle in bubble chart represents different weather conditions.

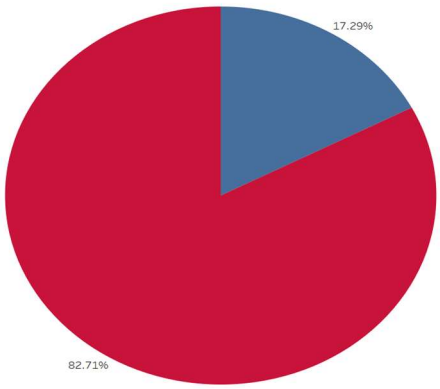
**Figure 13**




## Pie Chart

Pie chart is used to represent what proportion of each categorical value in the whole. Each proportion is represented in the form of a slice in a circle. For this project pie chart is used to represent the lanes on which more accidents happened.

**Figure 14**



## Use Cases

The analysis for this project is performed by using different use cases. This use cases are presented in the form of three dashboards.

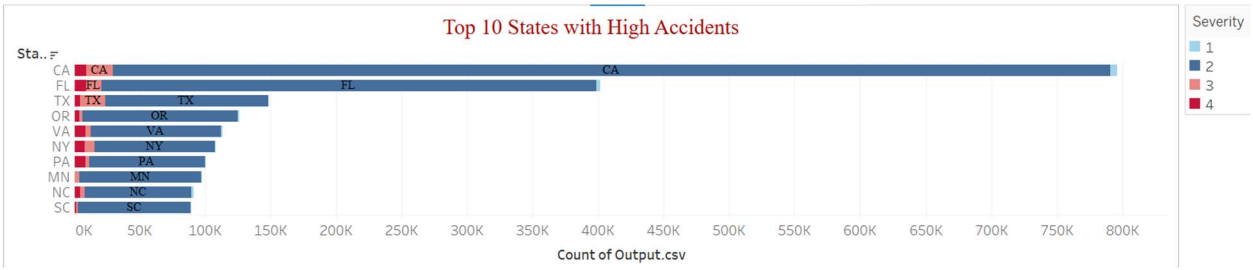
Below are the use cases represented in the dashboards of this project

1. Top 10 states with highest number of accidents over the years of 2016 to 2021
2. Total number of accidents for the year of 2021
3. Percentage of crash severity levels
4. Distribution of accidents across US time zones
5. What are top accident-prone cities
6. What are the major accident-prone zones?
7. Total number of accidents across US for each city from 2016 to 2021
8. On which day most number of accidents are recorded
9. Accidents at different times of the day
10. What is the trend line of accidents in 2021?
11. Number of accidents recorded at different temperatures
12. At what weather conditions majority of accidents are recorded
13. At what side of the road major number of accidents recorded
14. Accidents based on density of population
15. What are the dangerous highways where the greatest number of accidents are recorded?

**Use Case 1: Top 10 states with highest number of accidents over the years of 2016 to 2021**

***Action:*** A bar chart is used to provide information about the top ten states in USA with highest number of accidents. Along with the number of accidents in each state severity of accidents is also represented in this chart.

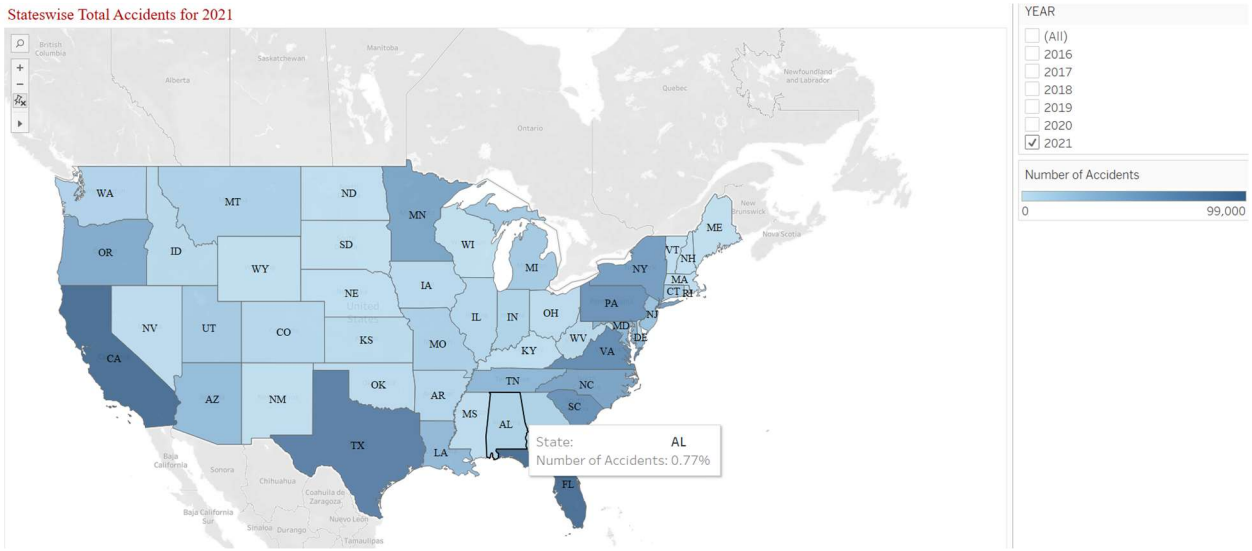
**Figure 15**




**Use Case 2: Total number of accidents for the year of 2021**

***Action:*** This map chart is used to represent the number of accidents happened in 2021 for each state. The color is proportional to the number of accidents. As the number of accidents increases the color gets darker. From this chart we can see that California is having highest number of accidents with 25.72%. Florida is having 2nd highest number of accidents with 18.42 percent. Followed by Florida Texas is having more accidents with5.53%. South Dakota is having least number of accidents.

## Figure 16




**Use Case 4: Percentage of crash severity levels**

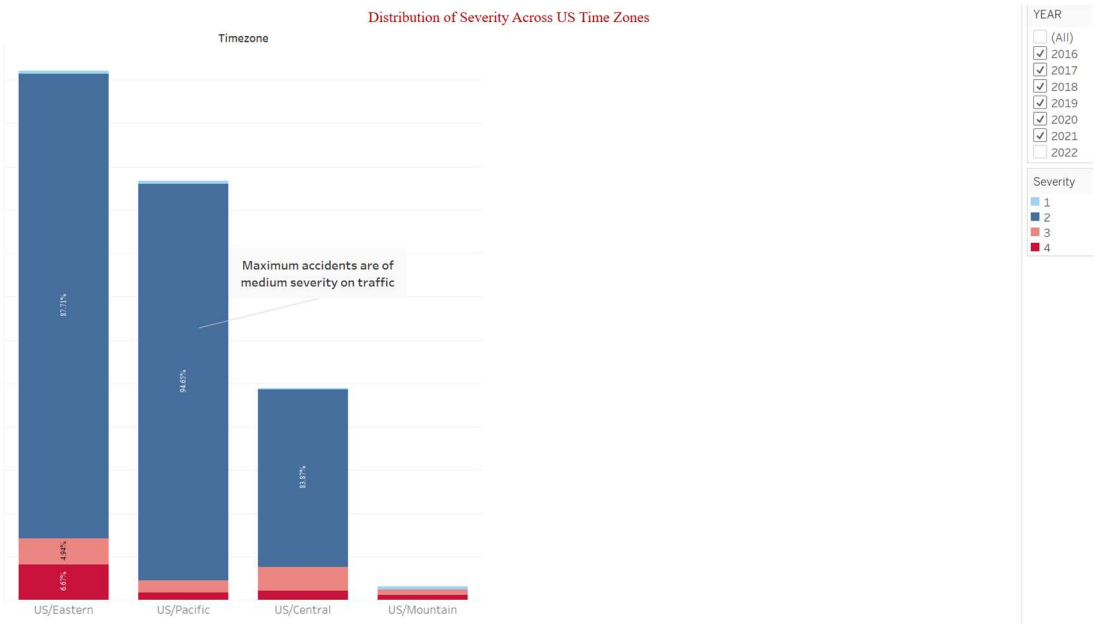
***Action:*** The accidents are classified into four types of severities higher the number of severities more the impact is. Severity 1 indicates the least impact on traffic that is shortest delays as the result of accidents. Severity 4 is the significant impact on traffic that is long delays as the result of accidents. Unfortunately, very small portion of accidents are of Severity 1 and almost 90% of accidents happening are of Severity 2.

## Figure 17

**Use Case 4: Distribution of accidents across US time zones**

***Action:*** This chart is used to represent the distribution of severity across us time zones. This distribution shows highest proportion of accidents happens at eastern time zone followed by pacific time zone and mountain time zone is having least proportion of accidents compared to other time zones. In eastern time zone almost 88% of accidents are of Severity 2

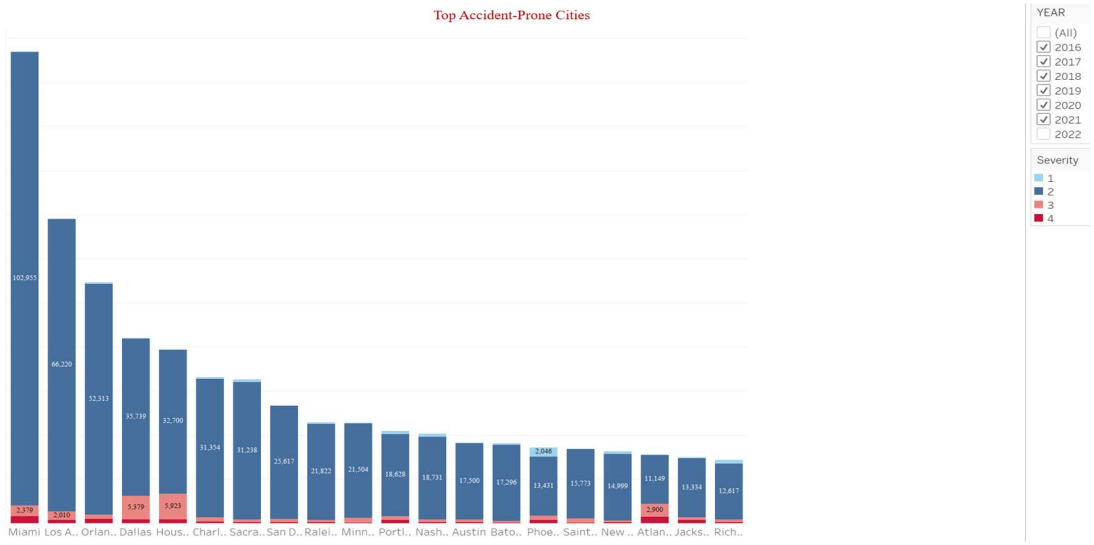
**Figure 18**




**Use Case 5: What are top accident-prone cities**

***Action:*** This bar chart is used to get insights about the city with highest number of accidents. Miami is the city with highest number of accidents across US followed by Los Angeles. In Los Angeles 103k accidents happened from 2016 to 2021 are of Severity 2 type.

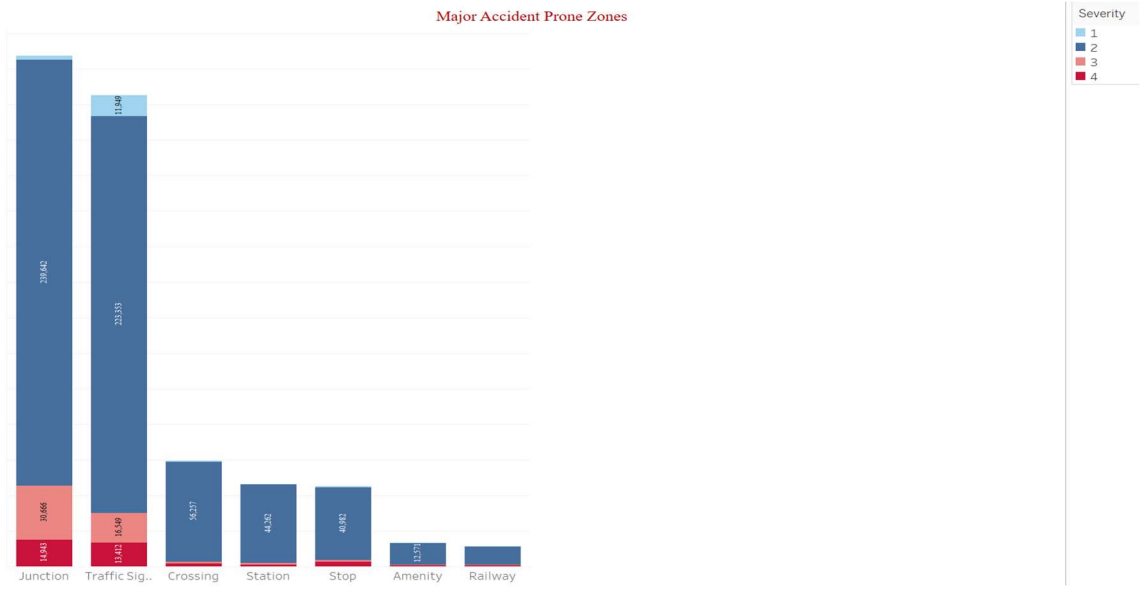
**Figure 19**




**Use Case 6: What are the major accident-prone zones?**

***Action:*** This bar chart is used to represent the major accident hotspot zones. From this we can see that highest proportion of accidents are happening at Junctions, traffic signals followed by crossings

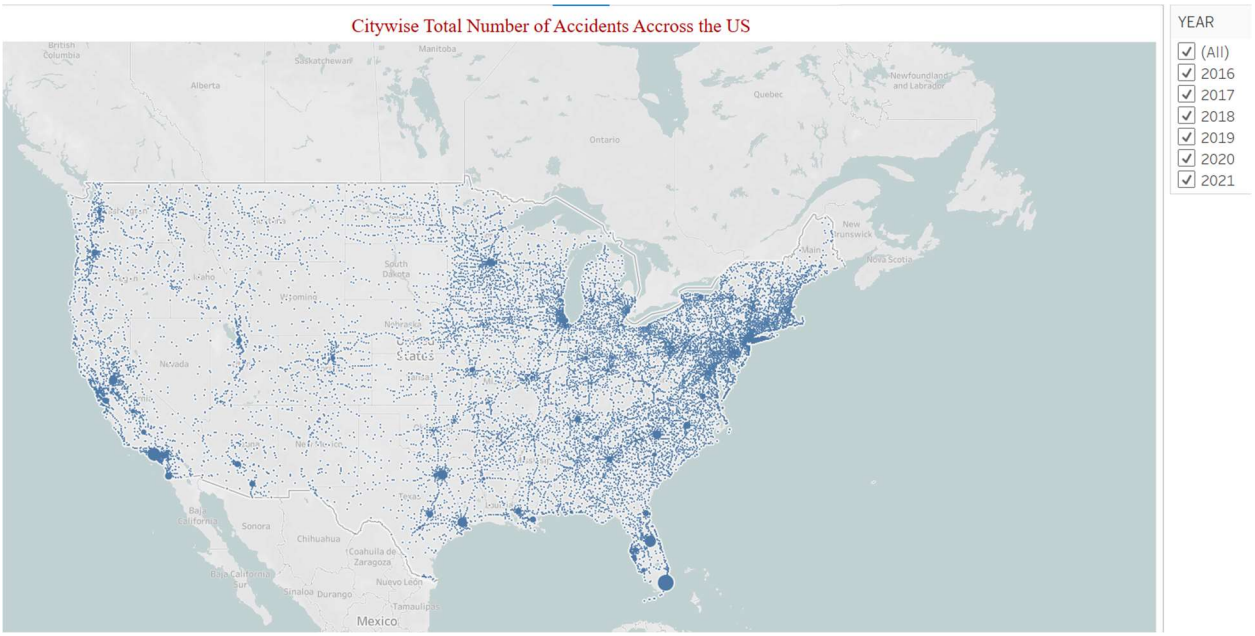
**Figure 20**



**Use Case 7: Total number of accidents across US for each city from 2016 to 2021**

***Action:*** This dot map is used to represent the number of accidents happened for each city. The size of dot is proportional is the number of accidents as the number of accidents increases the size of dot increases. In Florida Miami is having the highest proportion of accidents and in California Los Angeles is having highest number of accidents.

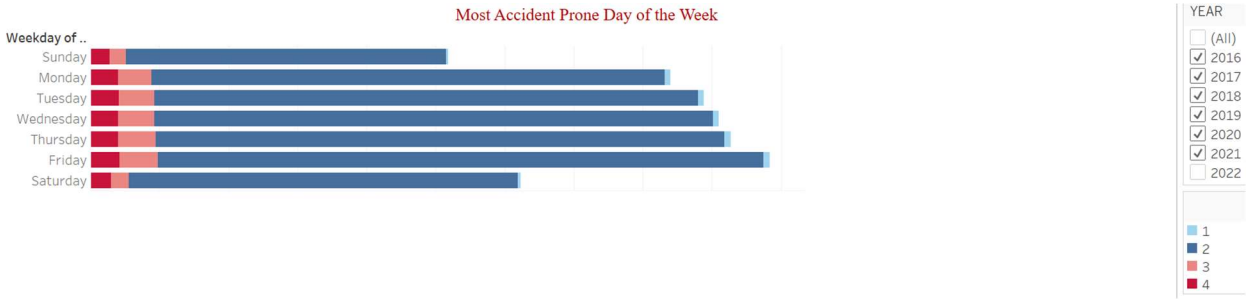
**Figure 21**




**Use Case 8: On which day most number of accidents are recorded**

***Action:*** This bar chart is used to provide insights about on what day of a week the most number of accidents are recorded. From this chart we can see that number of accedents recorded at weekdays are higher than that of weekends. The possible reason could be that most people commute to work during weekdays. Friday of the week is recorded with highest number of accidents than the other days of the week.

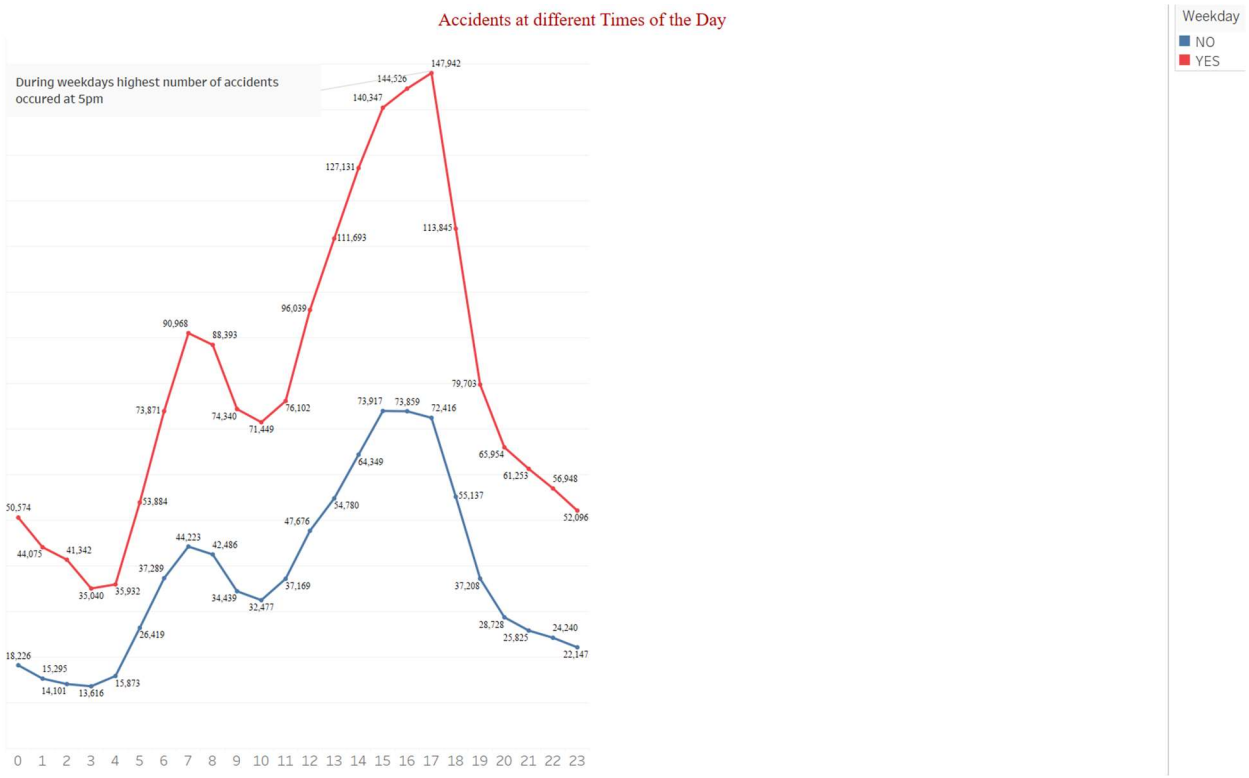
## Figure 22




### Use Case 9: Accidents at different times of the day

***Action:*** This visualization is used to give information about at what time of the day most number of accidents are recorded. From this it is observed that majority of the accidents are recorded between 6 to 9 AM and 3 to 6 PM in both weekdays and weekends.

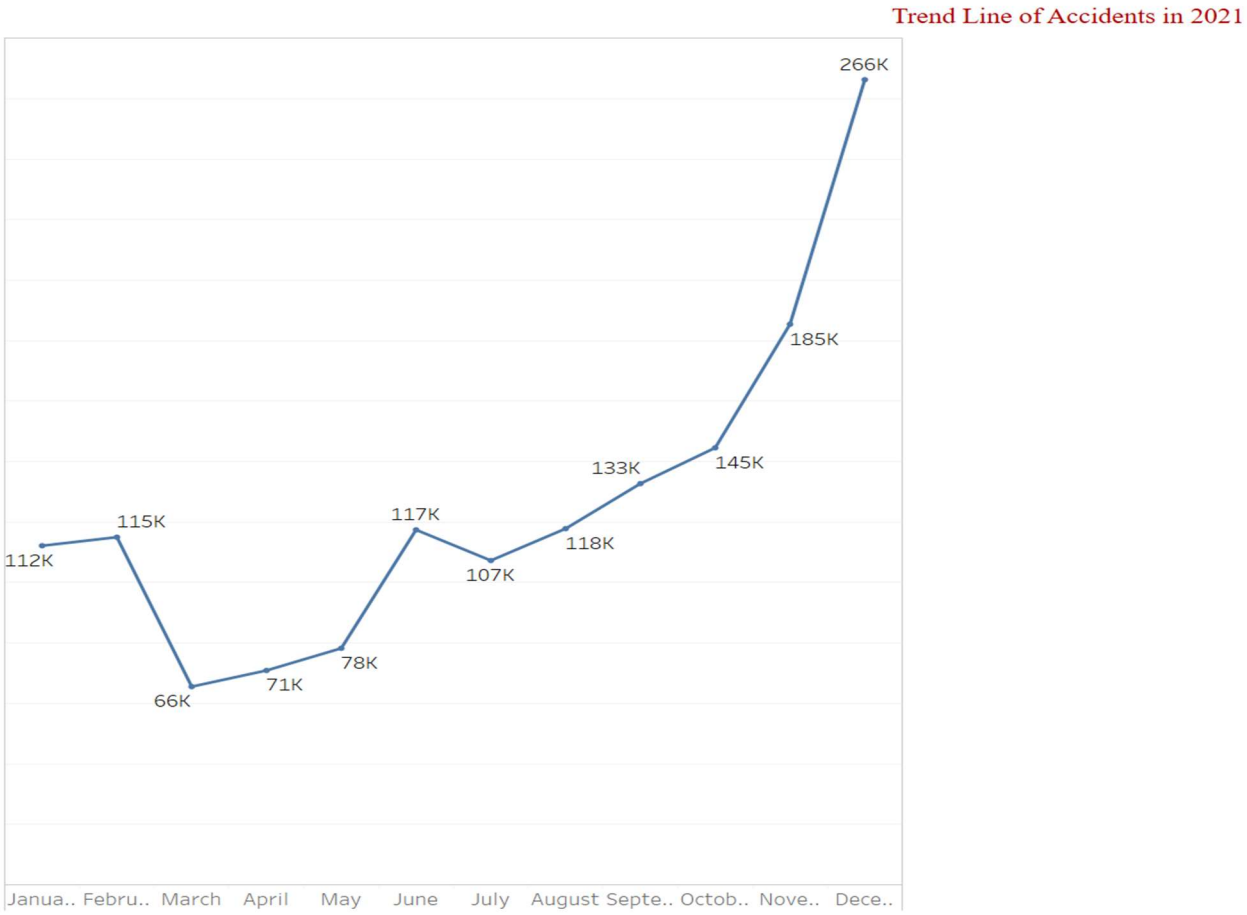
## Figure 23



**Use Case 10: What is the trend line of accidents in 2021?**

***Action:*** As you can see from the trend line of accidents for the year 2021, the number of accidents were little over 100k during the months of January and February. Then there is significant decline in the number of accidents over the next three months. The accident numbers rose gradually from the month of July, and we can observe that the trend line had a steep rise in the number of accidents from October to December. Holistically we can observe that the number of accidents towards end of the year got more than doubled to what they were in the start of the year.

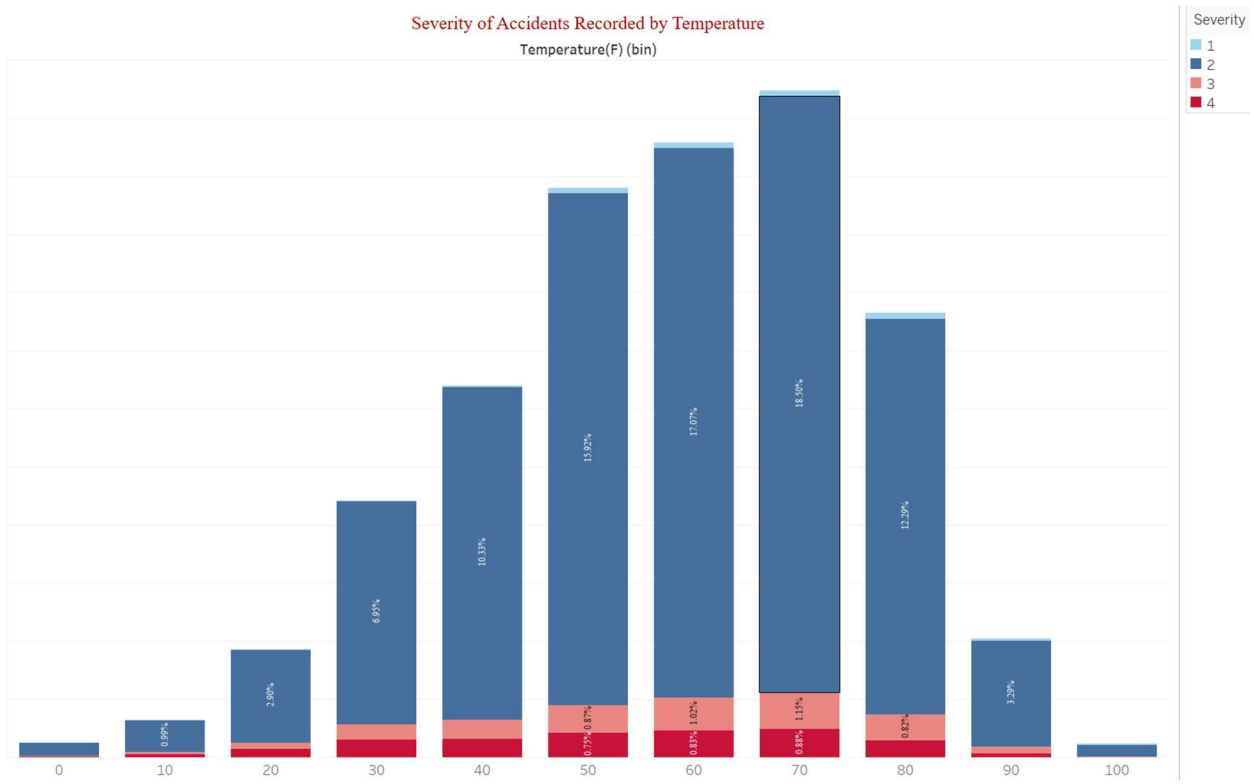
**Figure 24**



**Use Case 11: Number of accidents recorded at different temperatures**

***Action:*** This bar chart is used to provide insights about at what temperatures a greater number of accidents are recorded. From this it is observed that most accidents are recorded when the weather is pleasant that is between 50 to 80 F.

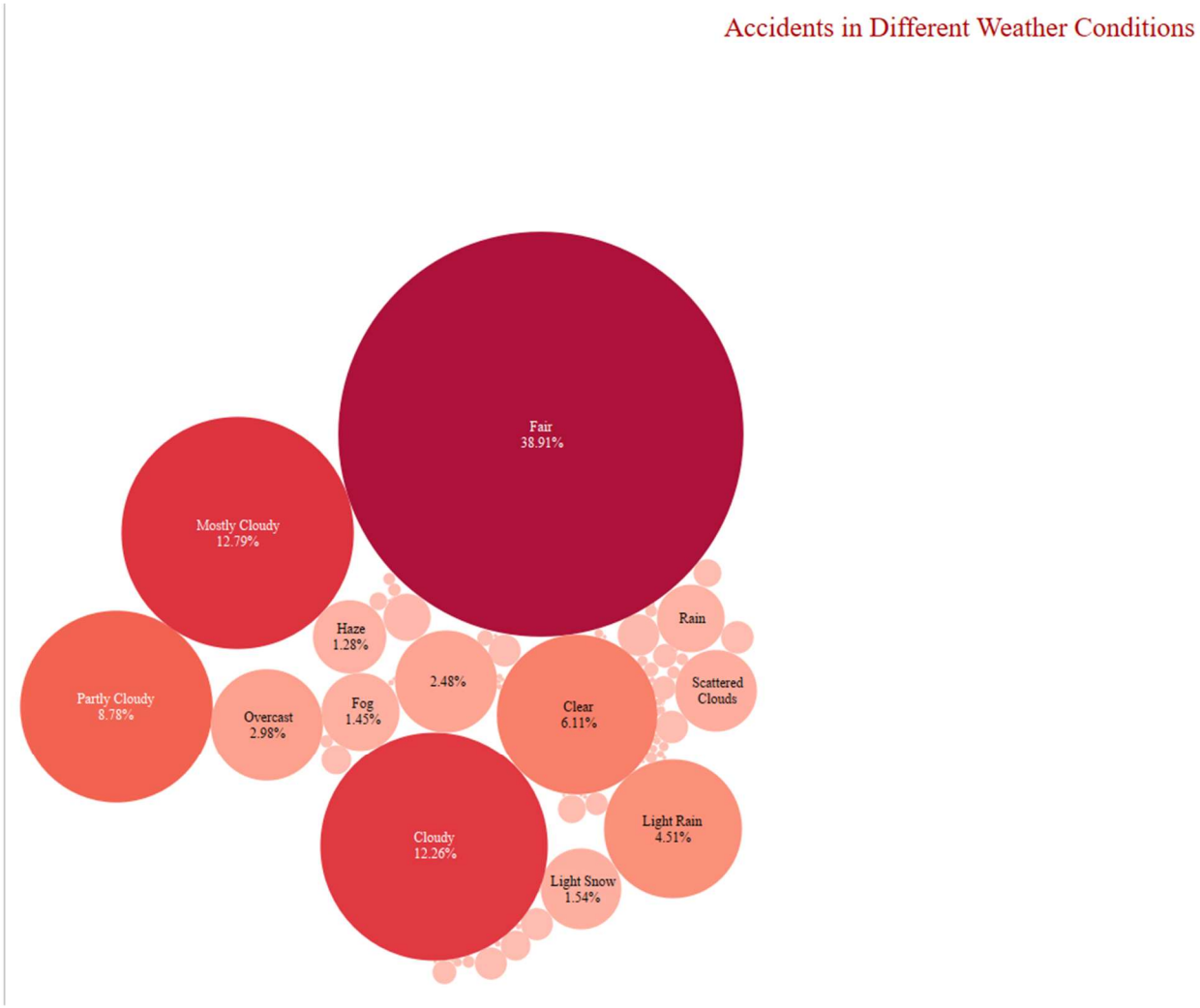
**Figure 25**




**Use Case 12: At what weather conditions majority of accidents are recorded**

***Action:*** From this visualization it can be seen that highest proportion of accidents occur in Fair weather conditions. Followed by Fair most accidents are recorded at cloudy conditions. Hear the size and color of each bubble is proportional to the number of accidents as the number of accidents increases the color and size of the bubble increases.

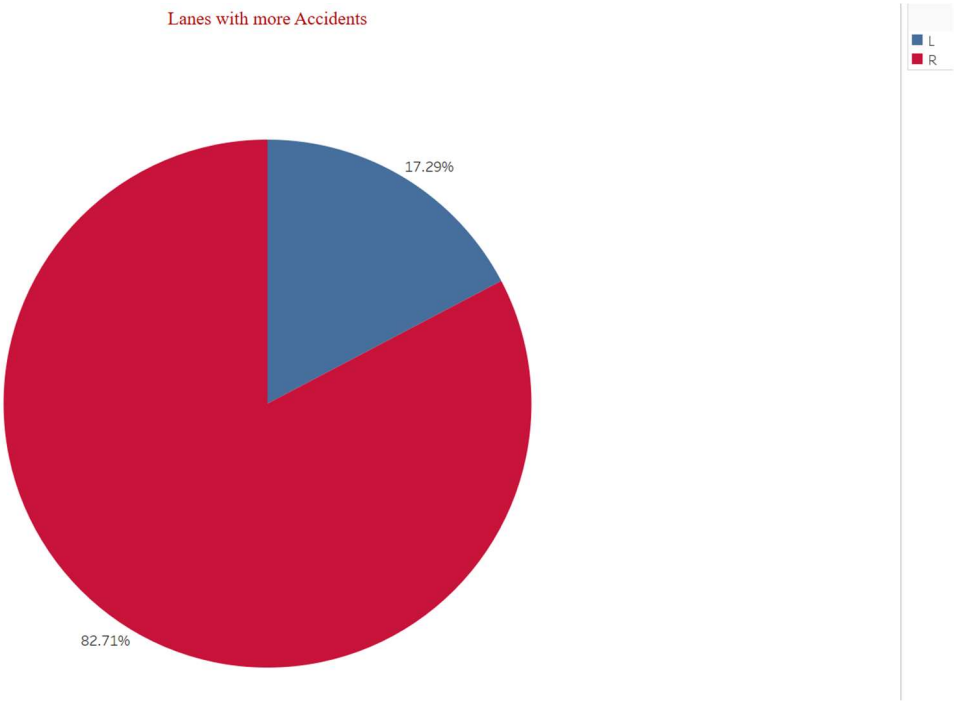
**Figure 26**




**Use Case 13: At what side of the road major number of accidents recorded**

***Action:*** From the Lanes with more Accidents Pie Chart, we can observe that: Right most lanes have major share of accidents with 82.71%. The probable reason could be the right lanes are busier than the left lanes. Left lanes are generally intended for steady ongoing traffic, whereas right lanes deal with exit and entry ramps which involves more vehicle maneuvers like merging the lanes and matching the speed of ongoing traffic.

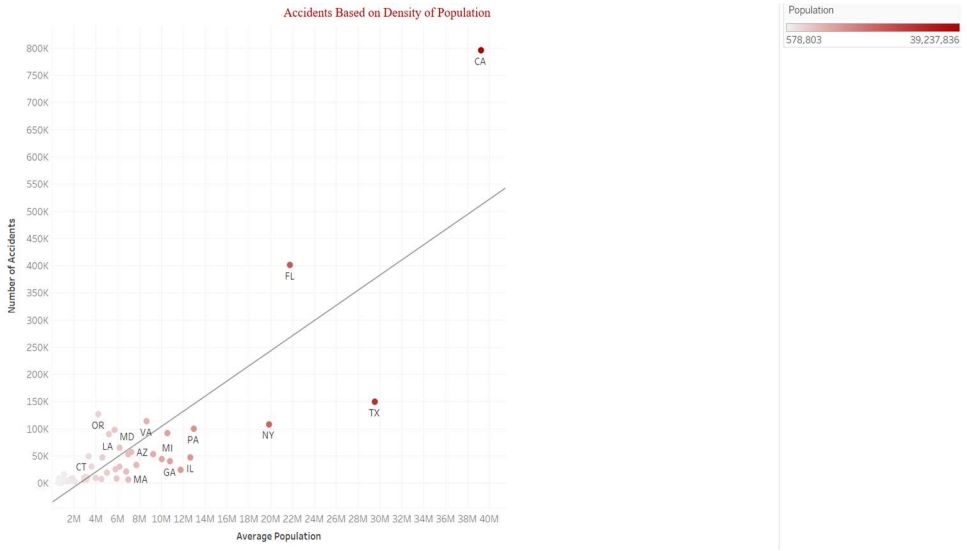
**Figure 27**




**Use Case 14: Accidents based on density of population**

***Action:*** From the Accident based on Population Visualization, it can be observed that: The accidents had a linear increase in the number with increased population. The main reason is very much evident that as the population increases, the number of vehicles on the road increases which results in more traffic and congested roads leading to rise in the scope of accidents.

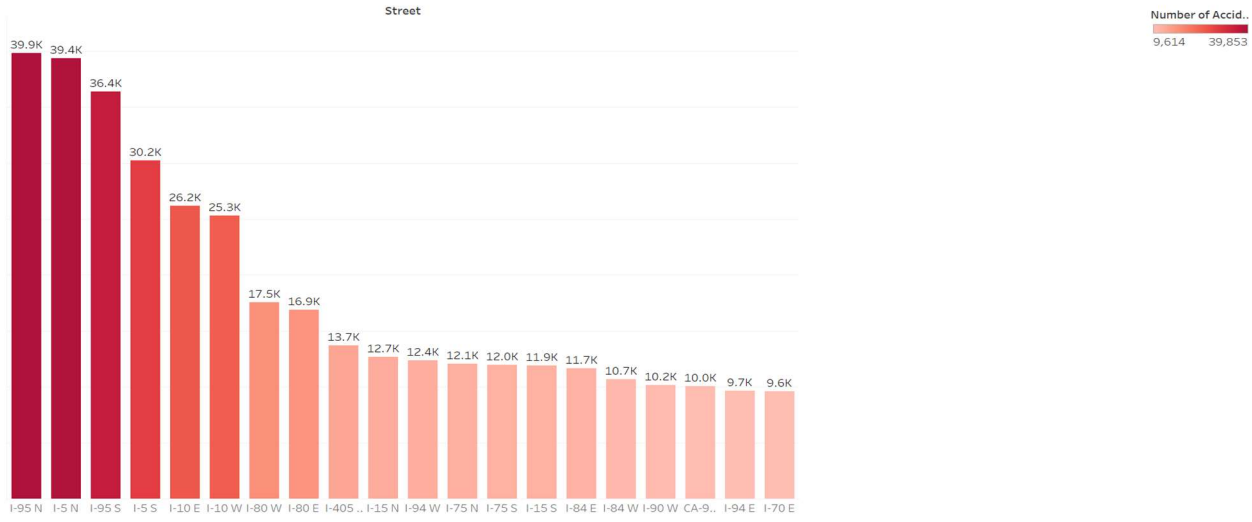
## Figure 28




**Use Case 15: What are the dangerous highways where the greatest number of accidents are recorded?**

***Action:*** This bar chart is used to represent on what Highway majority of accidents happened. Form this chart it is observed that I-95 North and south is recorded a greater number of accidents.

## Figure 29

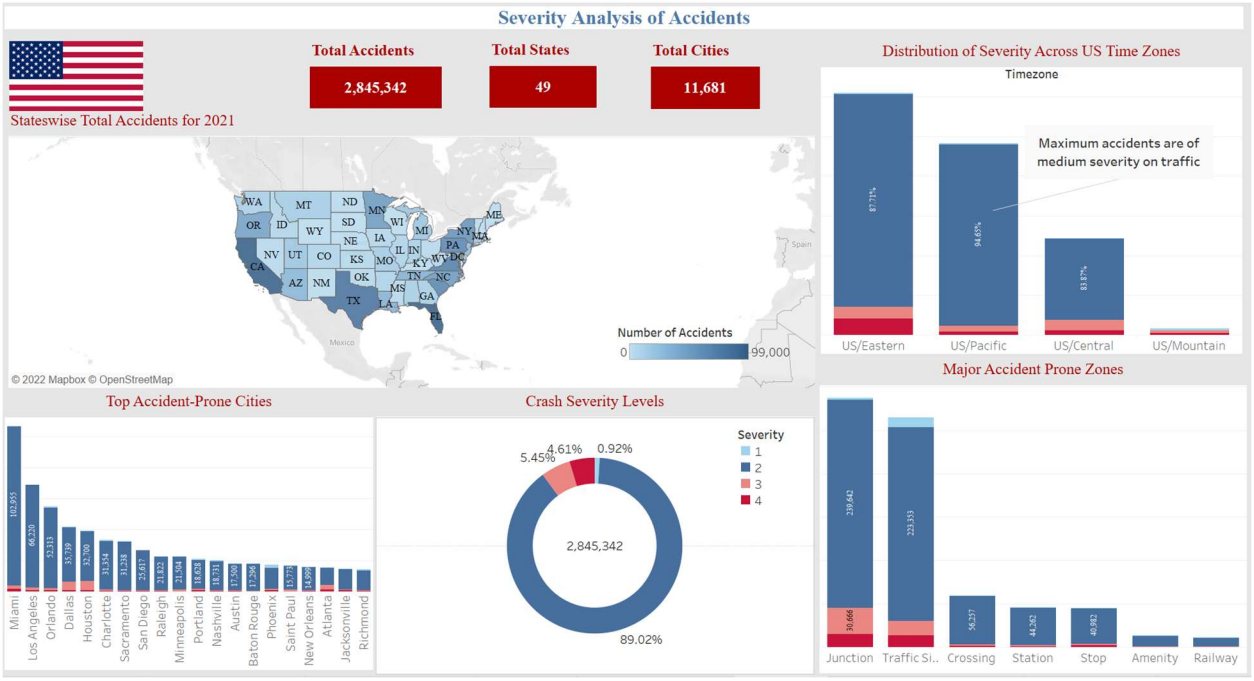


# Test Cases

## Dashboard 1: Severity Analysis of Accidents

This is an interactive dashboard used to provide information about the severity analysis of Accidents. To get severity information of each state you can just click, and all the other charts will populate information related to that state.

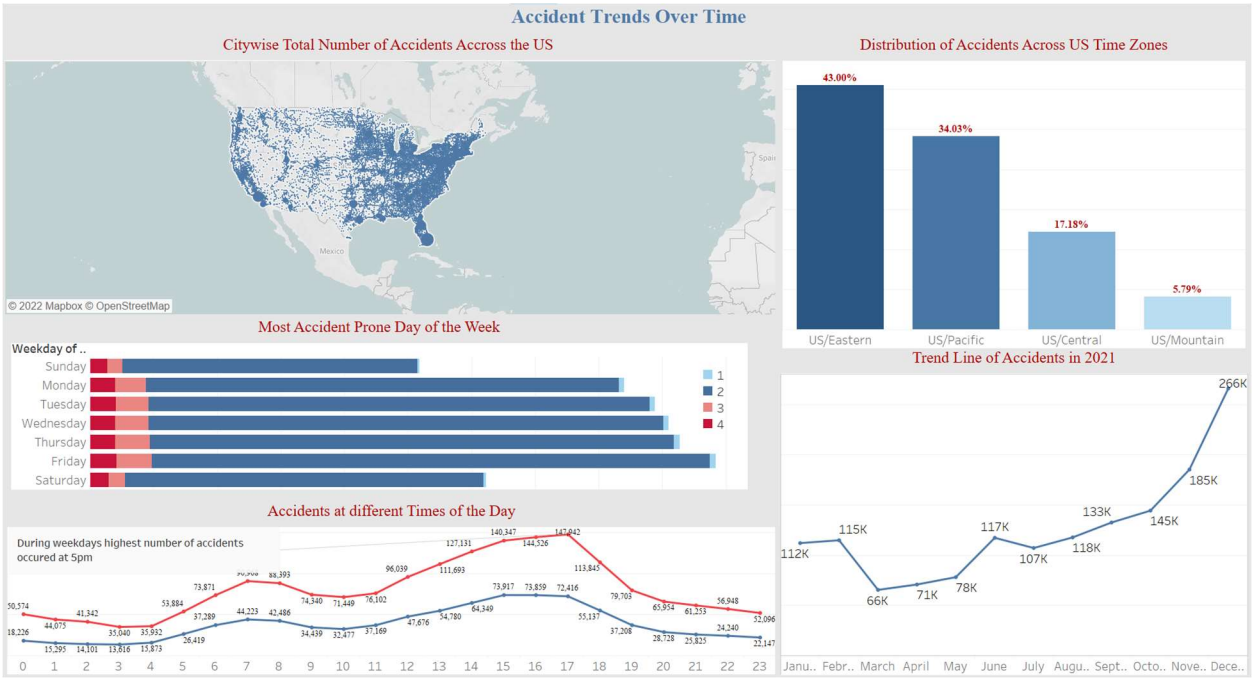
**Figure 30**




## Dashboard 2: Accident Trends Over Time

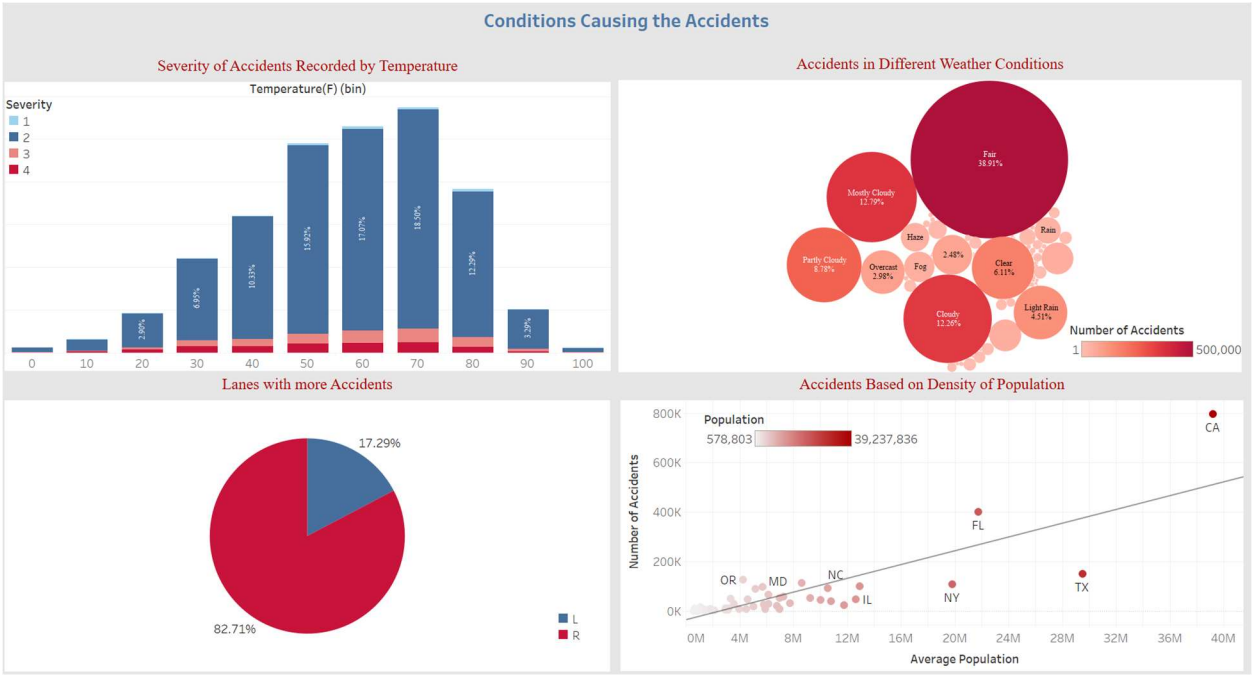
This is an interactive dashboard which is used to provide insights about Accident trends over time in united states from 2016 to 2021.

## Figure 31




## Dashboard 3: Conditions Causing the Accidents

## Figure 32

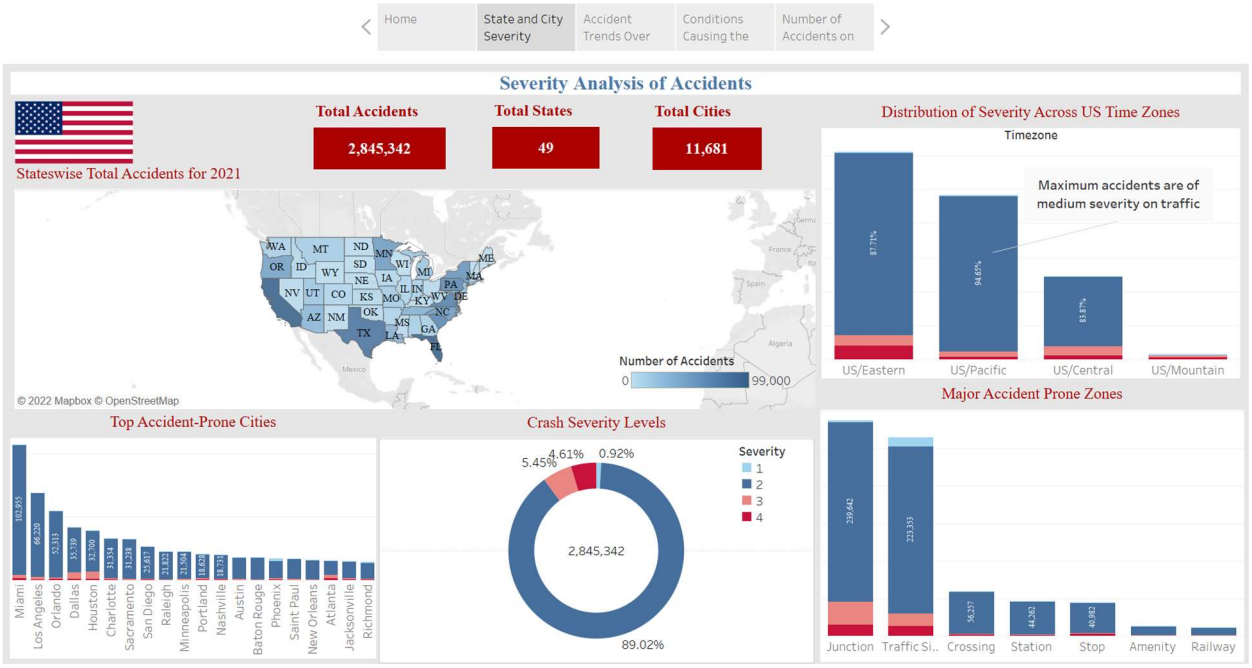


# Story Screenshots

## Figure 33

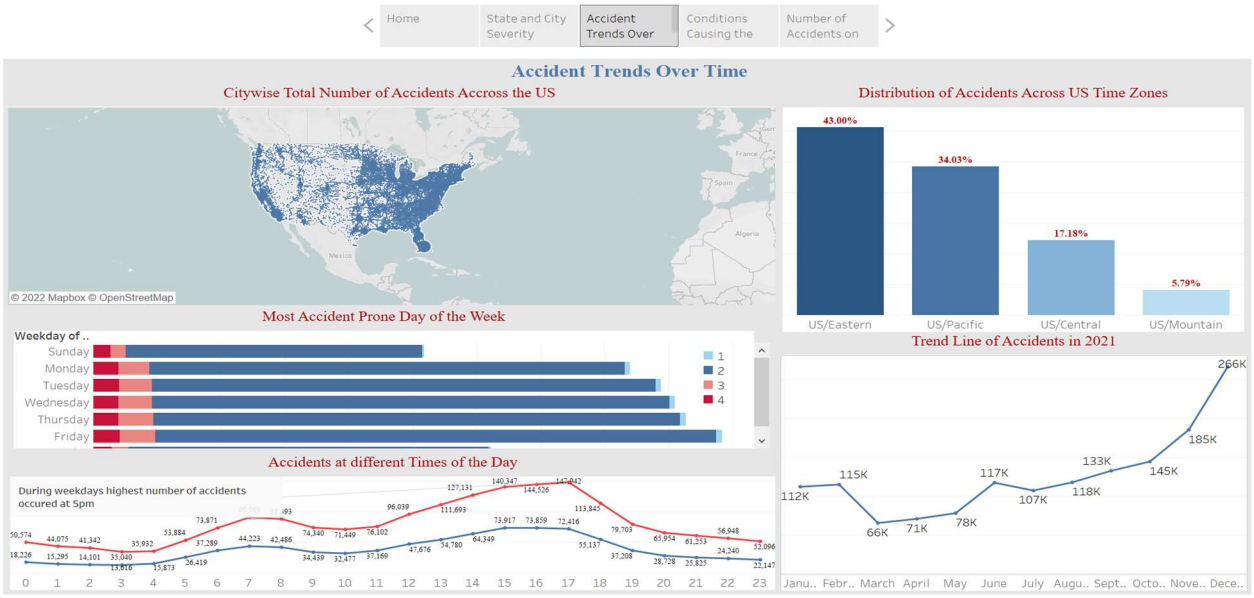

## Severity Analysis Dashboard

## Figure 34



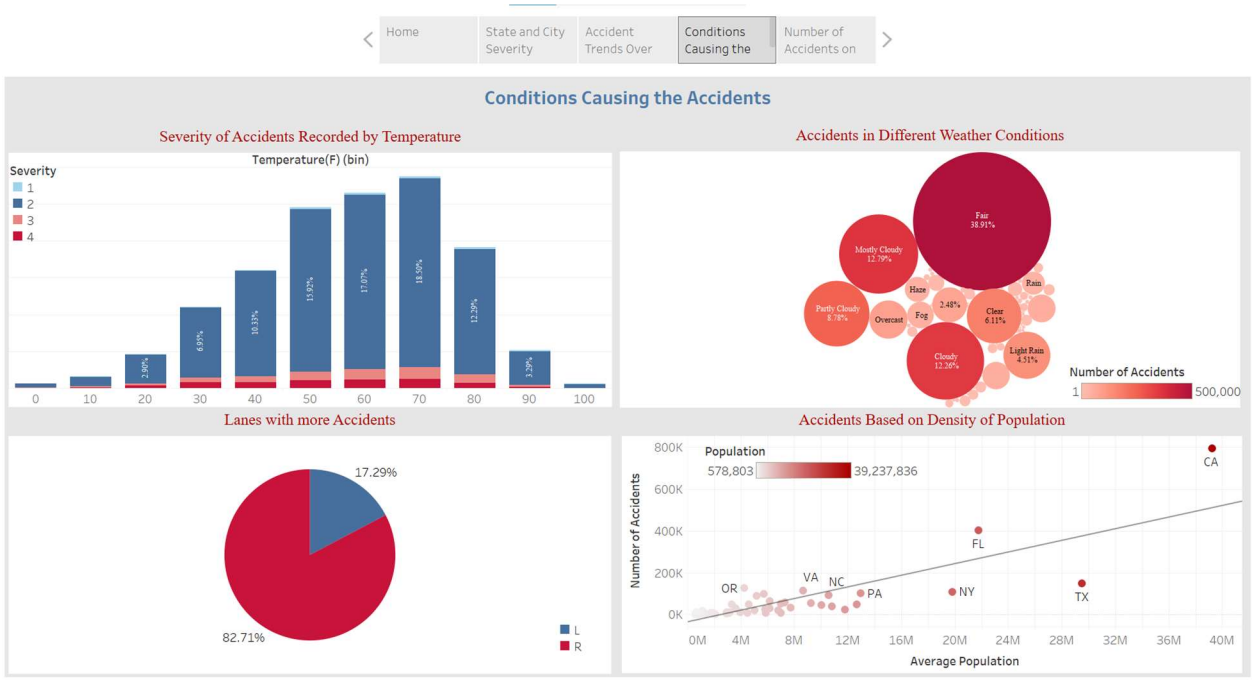
## Accident Trends Dashboard

### Figure 35



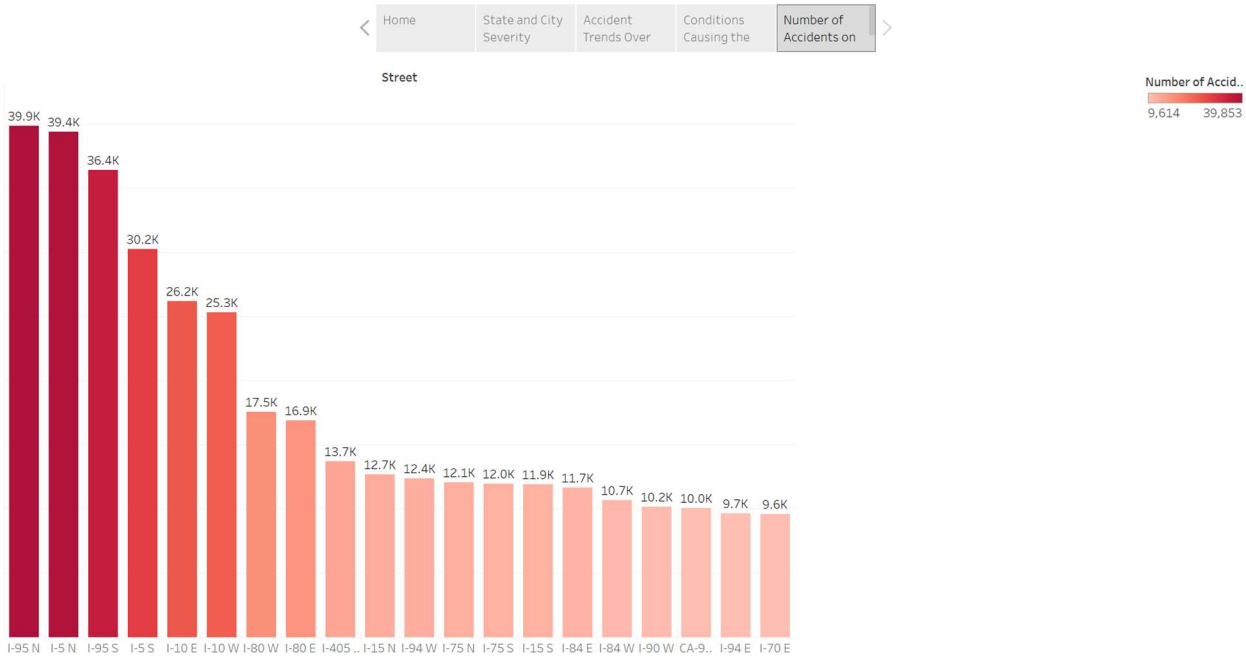

## Factors Causing Accidents

### Figure 36



## Accident analysis on Highway Dashboard

### Figure 37



## Summary

After performing the analysis in USA road accident data some of the recommendations are. Measures should be taken to reduce the accidents during peak times recognized during the day, especially on weekdays. Alerts should be provided to drivers when the weather is cloudy or overcast. Strict speed limits should be implemented in the recognized hotspot zones. Congested roads should be avoided

## Future Work

In future this project can be extended by collecting and analyzing accident data from all over the world in addition to the United States.

## References

*US Accidents (2016 - 2021)*. (2022, March 12). Kaggle. https://www.kaggle.com/datasets/sobhanmoosavi/us-accidents

US Census Bureau. (2021, December 20). *State Population Totals and Components of Change: 2020–2021*. Census.Gov. https://www.census.gov/data/datasets/time-series/demo/popest/2020s-state-total.html