Urvish's Portfolio



Hi, I'm Urvish!

I am an International student at the University of Colorado at Denver (Business School), pursuing Master's in Business Analytics. From the first day of my undergraduate, data intrigued me. I love solving puzzles and finding patterns in large data sets.

While pursuing my master's I have worked on various languages and visualization tools such as R, Python, PowerBi, and Tableau.

Professional experience

Currently: Graduate assistant as IT LAB advisor at the University of Colorado, Denver Business School

Previously: Ten months of experience as a .Net Developer at UniqTech Solutions where majorly I used to create CRM (Client Relationship Management) applications with C# language, which included creating stored procedures & views in the database.

Projects

1. What factors influence payments for medical malpractice? (RStudio)

An insurance company wants to know more about the number of claims it pays out for medical malpractice lawsuits. The data table, taken from recently adjudicated or settled lawsuits, shows claim payment amounts, the specialty of the presiding physician, and information about the claimant or patient. Data Table: Lawsuits.xlsx

The data table contains a random sample of 118 claim payments. The variables in the data table are: Payment Amount of the payment in thousands of dollars

Severity An estimate of the severity of damage to the patient:

- 1 = emotional trauma
- 2 = insignificant damage
- 3 = minor temporary damage
- 4 = major temporary damage
- 5 = minor permanent damage
- 6 = significant permanent damage
- 7 = major permanent damage
- 8 = grave damage
- 9 = death

10 = Age of the patient in years

Private Attorney Whether the claimant was represented by a private attorney (1 = private attorney, 0 = not a private attorney)

Marital Status: Marital status of the claimant (0 = divorced, 1 = single, 2 = married, 3 = widowed, 4 = unknown)

Specialty: Specialty of the physician involved in the lawsuit Insurance: Type of medical insurance carried by the patient

Gender: Gender of the patient

Questions this analysis will carry out.

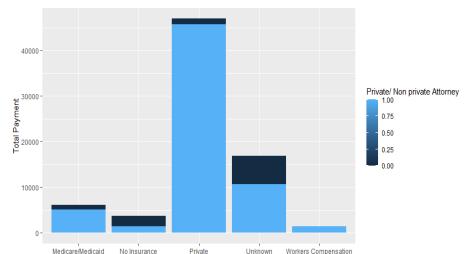
- 1) What to Report to Management? (Analysis)
- 2) What factors influence payments for medical malpractice? (Insights)
- 3) What're the recommendations?

EXECUTIVE SUMMARY

Major Findings:

- Across all the payments, the total sum of private attorney payments weighs 88.27% of the total payments. Also, payments more than \$800,000 are mostly private attorneys that's why claims have to be paid more as compared to a non-private attorney.
- Total of private insurance payments is highest by value at \$46,950,500 as compared to all other types of insurances, covering 62.53% of the total.
- Severity of 7 and 8 have the highest payments which are 38.69% of the total payments, where severity 7 has paid a maximum total of payments of \$14,606,200 and severity 8 of \$14,446,100.
- The Maximum number of lawsuits are filed by people whose age was 25 to 50 in between.
- Maximum of the payments have private medical insurance, as the total sum of private insurance is the largest with a value of \$46,950,500. And lowest was for Worker's Compensation worth \$1,387,555.
- Maximum sum of payments are paid by married ones with private insurance.
- Medical payments are more than payments for surgeries by \$21,308,100.

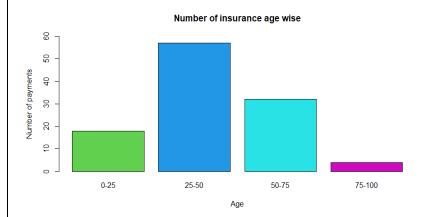
Relation between payments from various types of insurance for private/nonprivate attorney



Recommendations for Action

- Company should focus on people with an age range from 25 to 50 and improve their policy terms.
- Private attorneys are the ones who are costly, so companies should hire more experienced lawyers to fight with them in court so that they could win it and compensate less to the patients.
- Should make a policy of compensating a small amount for medical bills as they are more than surgical ones.

Should increase Insurance costs of doctors who are in family practice and OBGYN.



the Documentation Page.

Analytical Overview

- Excel data was cleaned before analyzing such as duplicates and where age was 0.
- For all graphs and data visualization, only R programming was used.
- All major findings and recommendations are based on EDA which is explained on

PROPER DOCUMENTATION OF THIS ANALYSIS CAN BE VIEWED THROUGH THE BELOW LINK ADDRESS.

• https://github.com/Urvish130/RStudio_Insurance_Project/blob/main/Documentation on%20of%20Project's%20Finding.pdf

2. US accidents analysis. (Visualizations through R, PowerBi, & Tableau) Problem Statement.

Now suppose this is the year 2021 and I am an executive who works for the Department of Transportation and with an increase in road accidents, the department head wants to know the following things which I will be working on.

- 1. Which state has the highest number of accidents?
- 2. Which Zipcode/County has the highest number of accidents?
- 3. At what time do accidents usually occur in the US?
- 4. At what day of the week do accidents occur most?
- 5. What are the Factors Affecting Accident Severity?
- 6. Predict the location of the accident.
- 7. Predict the accident-prone zone in each state.
- 8. Generating the EDA for the data.
- 9. Predict the number of accidents that might occur in the next few years.
- 10. Getting possible trends in the Number of accidents. (Like every March of each year the highest number of accidents).

The dataset is from Kaggle.com (https://www.kaggle.com/sobhanmoosavi/us-accidents).

This Analysis will be useful to the Road safety department, hospitals, and response teams as they will be prepared for such accidents and will be knowing the measures to take beforehand the accident occurs in

an accident-prone zone (after analysis). This will tell us what steps must be taken to avoid such accidents. Will be recommending some actions to take when a similar condition arrives.

Methodologies I used:

- 1. Regression models (linear/ multilinear/ logistic)
- 2. Time series forecasting.
- 3. Various libraries for less coding in RStudio.
- 4. Dashboard with great UI.
- 5. Exploratory Data Analysis.
- 6. Geospatial visualization.
- 7. Some Calculated Fields.
- 8. Time series Visualization.

Tools I used:

- 1. RStudio
- 2. PowerBi
- 3. Tableau
- 4. PowerPoint

In RStudio:

I used this for Exploratory Data Analysis and to build some regression models. Will also try to predict values (Number of Accidents in Future) in RStudio, while training and testing my dataset. I believe RStudio is a great tool for Statistical Analysis so there will be some statistics involved, like getting to know the normality of columns, mean, median, mode, etc.

In Tableau/PowerBi:

I used Tableau & PowerBi to make visualizations, Presenting to my boss, and uploading to the department of transportation website for the citizens, so that they can evaluate and take precautions beforehand if something similar happens to them while on road.

In PowerPoint:

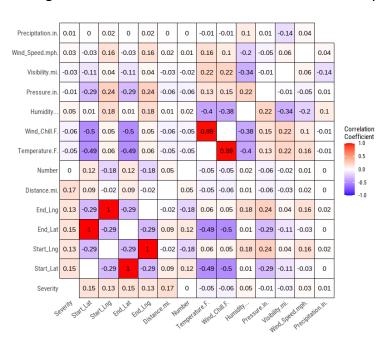
I will be presenting all my findings to my boss so that the accident department could be alerted of these findings. And make several measures and plan to reduce these accidents in the future. This will include my suggestions and views for the same.

Abstract

The US Accidents dataset is 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 the accident and the distance of traffic which is affected after an incident occurs. The very first thing which the 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 the 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 that might be directly linked to the type of weather condition during an accident.

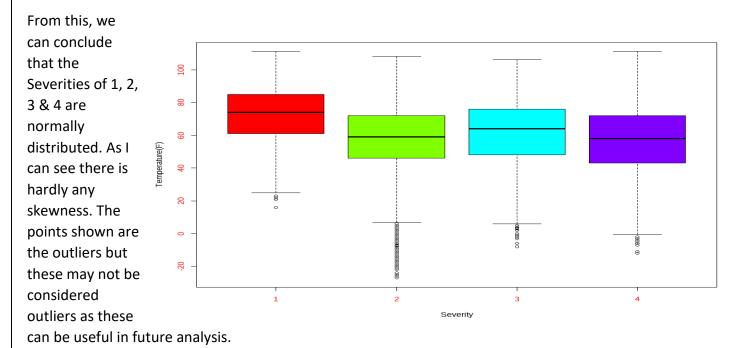
Understanding and cleaning Data in RStudio

The very basic steps of EDA analysis are to first approach the data and analyze categorical, numerical, & both at the same time. After loading the dataset into R, all the values with "NA" is removed. As this may become outliers in future analysis. After removing NA from the original dataset which consisted of 1516064 rows shorted to 335552 rows. Now, there are instances where your data rows are repeated, the reason might be a machine error. So we check if we have any repeated rows. Understanding the summary



of the data frame helps us to know which columns are of what data type. Describing the function of R gave me the statistics of each column. It shows mean, standard deviation, skewness, Kurtosis & Inter Quartile Range. R has this amazing function by which we can generate reports. After doing all the things mentioned above, I generated a normality report, which shows the normality plot and along with the plots if we transform the columns into a log or squared value. The correlation matrix is the best way to find out which columns are dependent in pairs.

Now let's see if each severity is having a normal distribution or not. To prove this I created a Box Plot which is shown below.

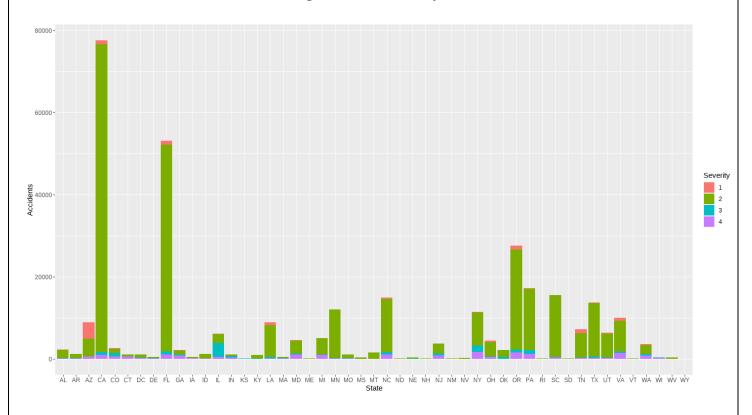


After plotting a correlation matrix I converted Boolean data types (True, False) to 1 and 0 so that I could build a Logistic regression model for knowing the factors which cause accidents.

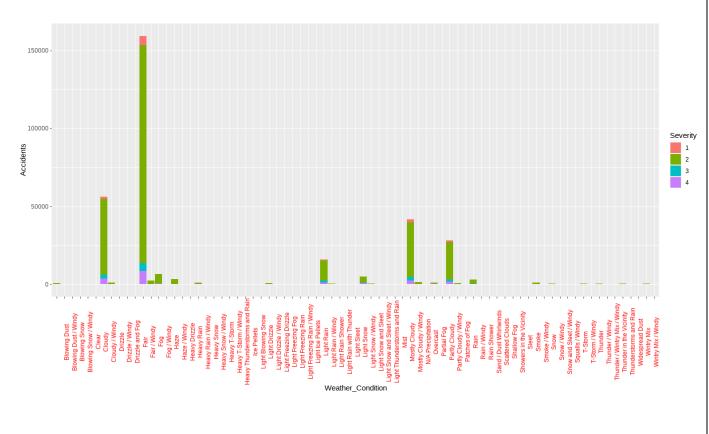
Common Visualizations.

After going through the cleaning part I created some visualizations which answer the following questions.

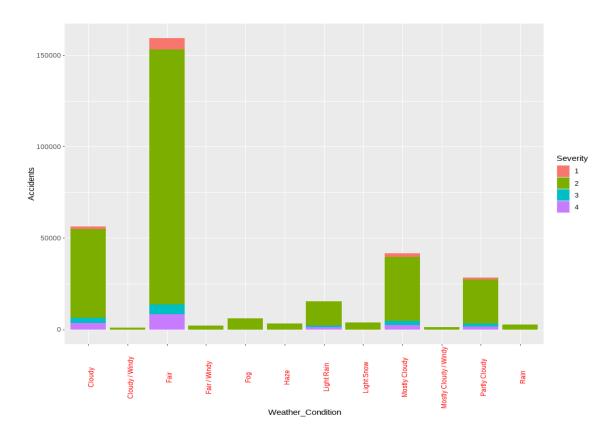
1. What is the count of accidents along with their severity in each state?



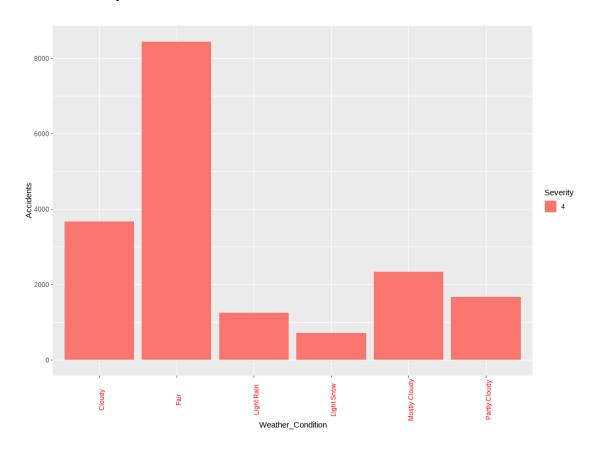
2. In which Weather Condition do accidents occur most?



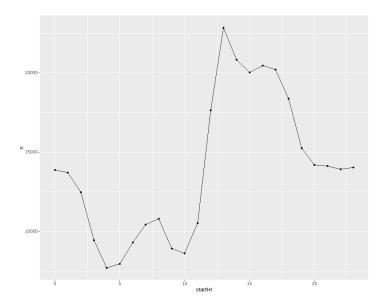
3. In which Weather Condition do accidents occur most where accidents are greater than 1000?



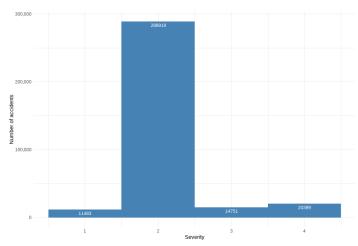
4. In which Weather Condition do accidents occur most where accidents are greater than 1000 and have a severity of 4?



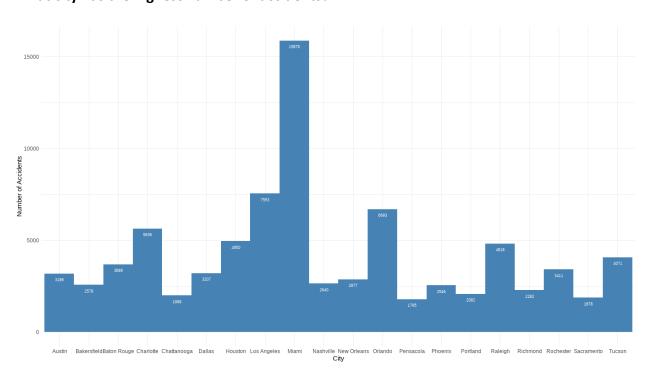
5. Usually at what hour do most accidents occur in the US?



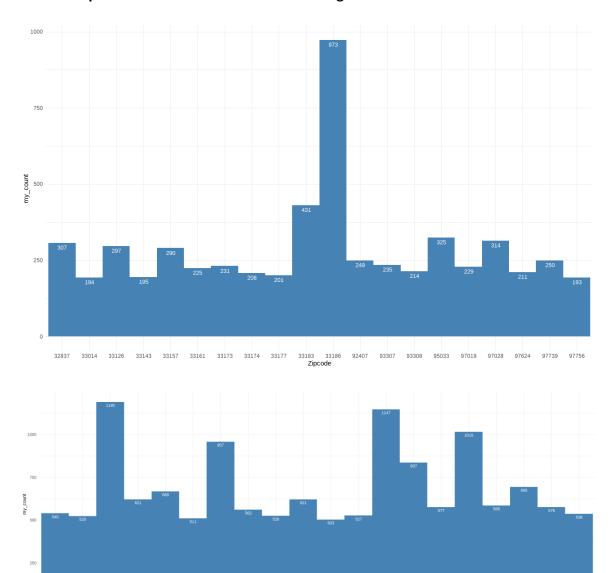
6. What type of severity accidents mostly occur?



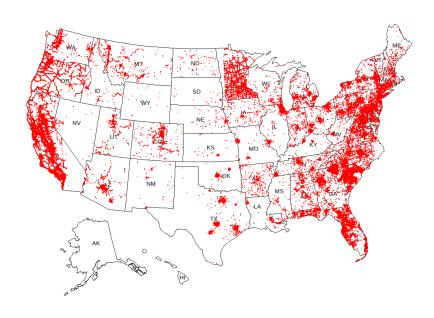
7. What city has the highest number of accidents?



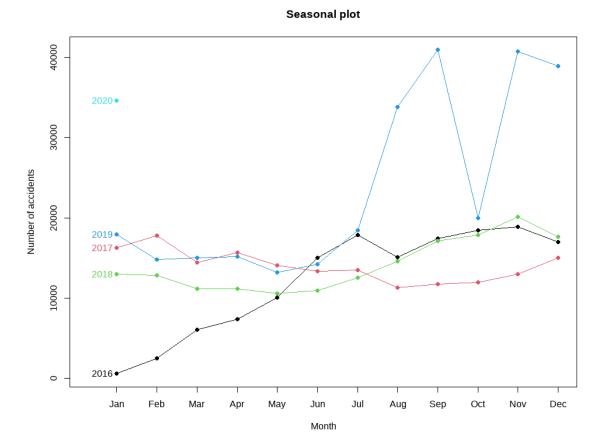
8. Which Zip code and street in the US has the highest number of accidents?



9. General view on the number of accidents in the US using the map through R code.



10. What is the comparison between different years concerning the count of accidents?



Logistic Regression Model.

From this model, we will be able to identify the actual causes of road accidents. The sole objective of this model is to find out factors that affect the severity of road accidents.

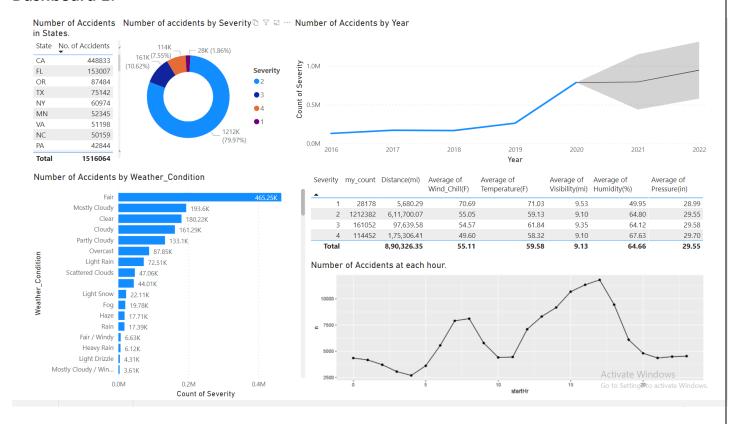
```
> summary(best_logistic_model)
call:
glm(formula = as.factor(Severity) ~ Temperature.F. + Civil_Twilight +
    Nautical_Twilight + Astronomical_Twilight + Humidity... +
Pressure.in. + Wind_Speed.mph. + Precipitation.in., family = binomial,
    data = df
Deviance Residuals:
    Min
              1Q
                    Median
                                  3Q
                                          мах
          0.0983
-3.5991
                    0.1780
                             0.3018
                                       1.0246
Coefficients:
                         Estimate Std. Error z value Pr(>|z|)
                        1.1583568 0.2064055
                                                          2e-08 ***
(Intercept)
                                                 5.612
                       -0.0328198
                                    0.0007383
                                               -44.454
                                                          2e-16 ***
Temperature.F.
                                               -3.221 0.001279 **
civil_Twilight
                       -0.1880180
                                    0.0583786
                                                -3.345 0.000822 ***
Nautical_Twilight
                       -0.2986898
                                    0.0892893
Astronomical_Twilight -0.9932627
                                    0.0782458
                                               -12.694
                                                        < 2e-16 ***
Humidity.
                        0.0186887
                                    0.0005022
                                                37.215
                                                        < 2e-16 ***
Pressure.in.
                        0.1526831
                                    0.0073554
                                                20.758
Wind_Speed.mph.
                        0.0062393
                                    0.0018171
                                                 3.434 0.000595 ***
                                                -3.472 0.000517 ***
Precipitation.in.
                       -0.4012353
                                    0.1155659
Signif. codes: 0 '*** 0.001 '** 0.01 '* 0.05 '.' 0.1 ' ' 1
(Dispersion parameter for binomial family taken to be 1)
    Null deviance: 100077
                            on 335551 degrees of freedom
Residual deviance: 85670 on 335543 degrees of freedom
Number of Fisher Scoring iterations: 8
```

Upon using and trying many models I found one bet model with a good P-value. Following is the image of the Logistic Regression model.

From this we can conclude that the Severity of an accident mostly depends on Temperature, Civil Twilight (Beginning in the morning or ending in the evening), Humidity, Pressure, Wind speed, and precipitation. So it is most like that an accident might occur if these conditions are fulfilled.

Power BI Dashboards.

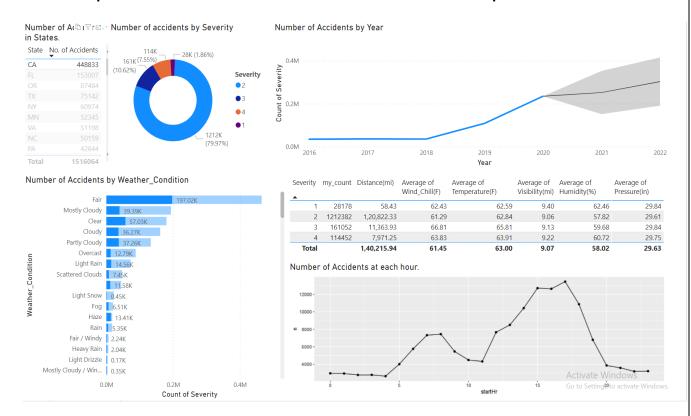
Dashboard 1:



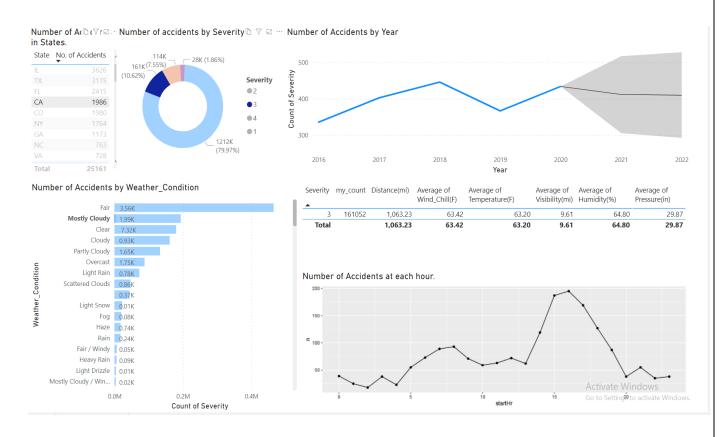
This Dashboard answers the following questions.

- 1. What is the count of accidents in each state?
- 2. In which Weather Condition do accidents occur most?
- 3. Usually at what hour do most accidents occur in the US?
- 4. What type of severity accidents mostly occur?
- 5. What are the predictions for the future? What number of accidents we can expect?
- 6. What is the number of accidents that happened according to severity while considering the average temperate, humidity, pressure, Visibility, and wind chill?

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



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?



Dashboard 2:



This Dashboard answers the following questions.

- 1. What is the number of accidents according to the state, city, zip code, and street?
- 2. During which Week do accidents occur most?
- 3. On what weather factor does the severity of the accident depends?
- 4. In what month do most accidents occur?
- 5. What is the Total number of accidents for the dataset?
- 6. 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?

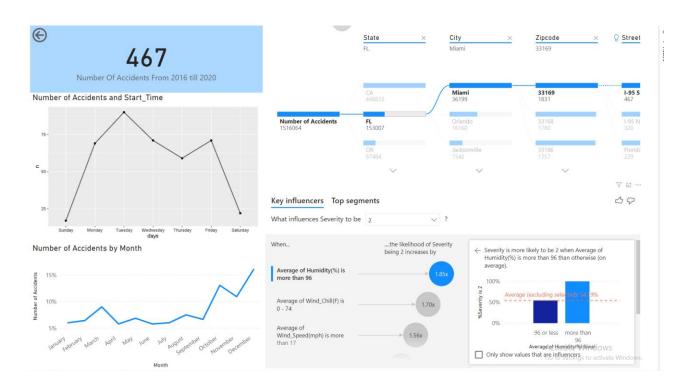
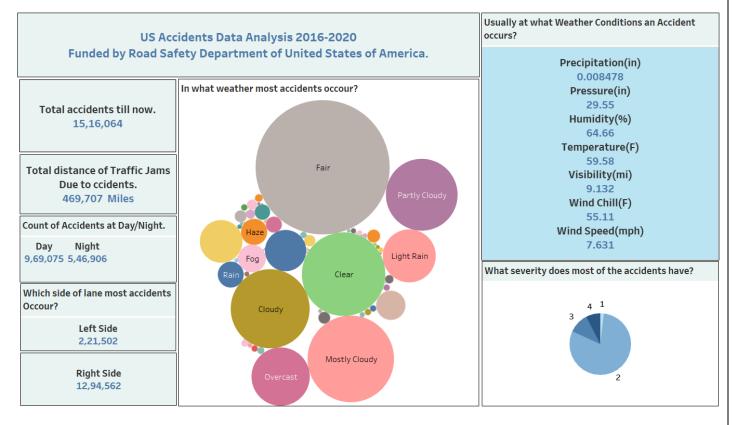


Tableau Dashboards.

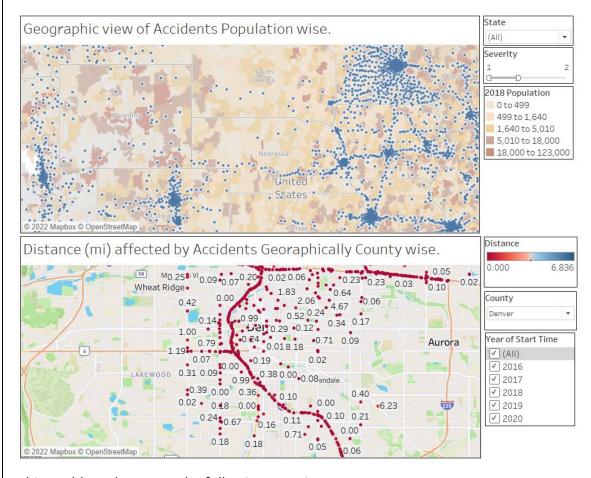
Dashboard 1:



The very first dashboard gives you some insights and an overview of what this analysis is all about. It answers the following questions.

- 1. What is the total number of accidents till now?
- 2. What is the total distance in miles affected by the accidents happening on road?
- 3. What is the number of accidents happening in daylight and nightlight?
- 4. What is the number of accidents happening on the left & right sides of a lane?
- 5. In what weather do most accidents occur?
- 6. Usually what Weather Conditions does an Accident occur?
- 7. What type of severity accidents mostly occur?

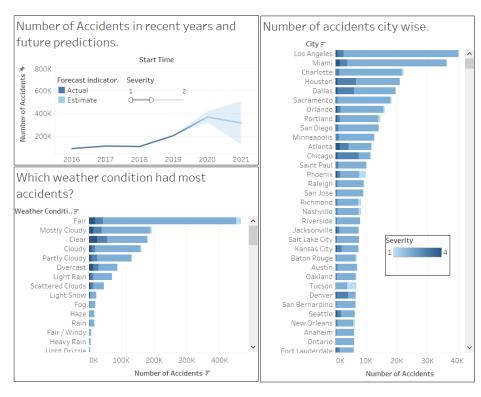
Dashboard 2:



This Dashboard answers the following questions.

- 1. What is the overview? In which region does the most number of accidents occur in the USA?
- 2. What is the Distance (mi) affected by Accidents Geographically?

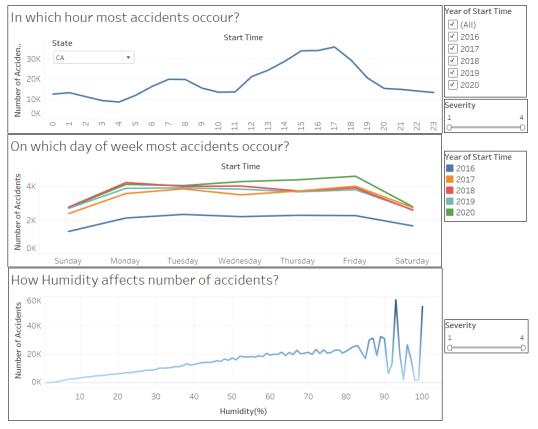
Dashboard 3:



This Dashboard answers the following questions

- 1. What are the predictions for the future? What number of accidents we can expect?
- 2. In which Weather
 Condition do accidents occur
 most?
- 3. What is the count of accidents in each city?

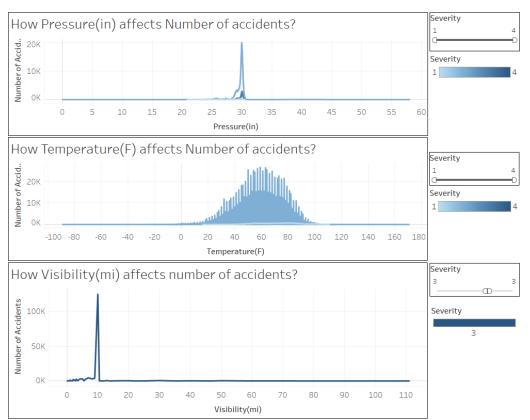
Dashboard 4:



This Dashboard answers the following questions

- 1. Usually at what hour do most accidents occur in the US?
- On which day of the week do most accidents occur compared to each year?
 How the weather conditions like humidity affect the number of accidents?

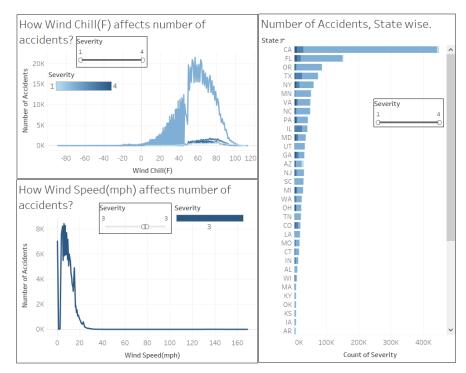
Dashboard 5:



This Dashboard answers the following question:

How the weather conditions like pressure, temperature, visibility, wind chill, & wind speed affect the number of accidents?

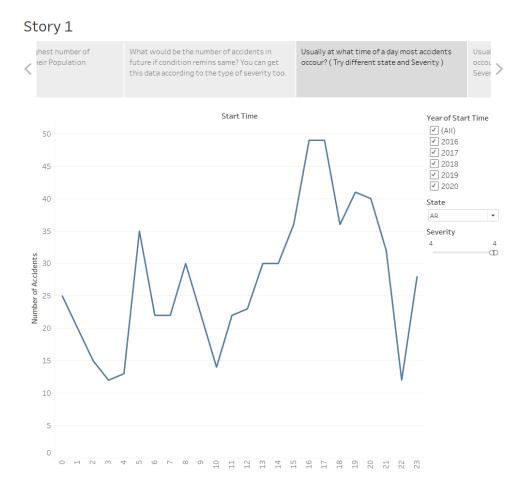
Dashboard 6:



This Dashboard answers the following question:

- 1. What is the count of accidents in each state?
- 2. How the weather conditions like wind chill, & wind speed affect the number of accidents?

Now Tableau has several features and one of them is the story. I have created a story that shows the same as the dashboard but it makes it easy for the user to understand. Below is just **one example**, you can have a look at the tableau file for more insights.



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 occur.
- Warning signs about speed limits are to be put on the accident-prone streets
- The state with the 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 for vehicles to have a first aid kit should be passed.
- Online surveillance for a prompt response from emergency services should be implemented.
- To have enough response teams to rescue in accident-prone locations.

PROPER DOCUMENTATION OF THIS ANALYSIS CAN BE VIEWED THROUGH THE BELOW LINK ADDRESSES.

1. RStudio code & Documentation https://github.com/Urvish130/Visualization-through-RStudio

A DIFFERENT VERSION OF US ACCIDENT ANALYSIS IN RSTUDIO CAN BE FOUND THROUGH THIS LINK.

https://github.com/Urvish130/RStudio_US_ACCIDENT_Project-EDA

2. PowerBi Documentation

https://github.com/Urvish130/Visualization-through-PowerBI

PowerBi File

https://drive.google.com/file/d/1uN99CxN6R1Ky-RHRdIYZI3Q5 wAM5ZMw/view?usp=sharing

3. Tableau Documentation

https://github.com/Urvish130/Visualization-through-Tableau

Tableau Dashboard

https://public.tableau.com/app/profile/urvish.patel/viz/FinalProject_16501460218 280/Dashboard1

4. Project's Final Report and Presentation.

https://github.com/Urvish130/Visualizations-Presentations-and-Reports

3. Python analysis on COVID-19 Vaccine drive & Sentiments of people on Vaccinations through Twitter Data

Suppose I am an analyst at WHO and my boss wants to know about the progress each country is having on COVID vaccination drives. He specifically wants answers to the following questions.

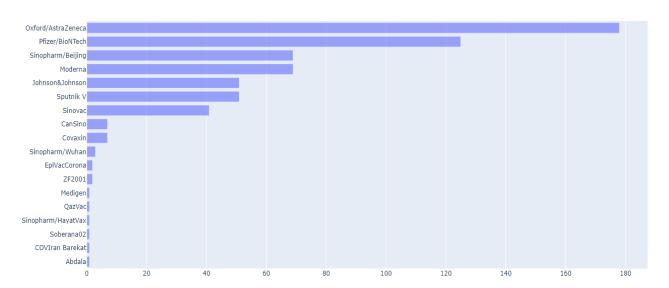
- Which country's vaccine drive was successful?
- What percent population of a certain country got vaccinated?
- What vaccination schemes (combination of vaccines) are used and in which countries?
- What manufacturer had produced the most amount of vaccine and sold it to what country?
- What was the relationship between vaccination evolution and sentiment toward vaccinations?
- What are the factors that influence vaccinations?
- What vaccination schemes (combination of vaccines) are used and in which countries?
- What country has vaccinated more people?
- What country has immunized the largest percentage of its population?
- How to progress vaccination programs around the World (or in a specific country) is received by the public, as reflected in the tweets about all vaccines.

EXECUTIVE SUMMARY

Major Findings:

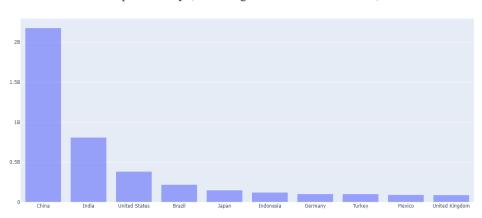
1. According to the data and graphs, we can conclude that oxford/ AstraZeneca had manufactured and distributed the most number of vaccines.

Vaccines laboratory distribution by countries

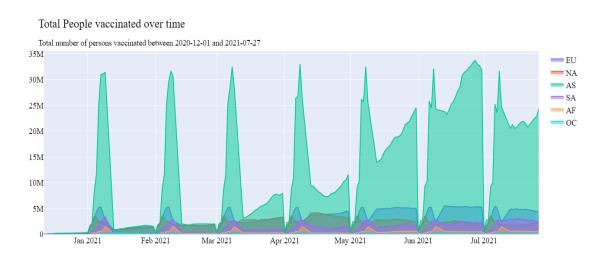


2. China is the country that has the highest number of populations vaccinated.

Total Vaccinations per country (Including First and Second Dose).



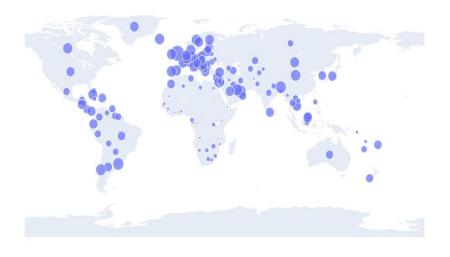
3. Asia had the best vaccination drive among all other continents.



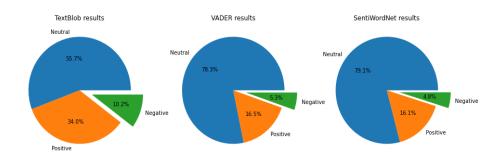
4. Great Britain is the country that has vaccinated the largest percentage of people from its population.

Vaccination ratio by country

((Vaccination/Population)*100)



5. According to different sentiment analysis methods mostly the tweets from the public were neutral. After that most tweets were positive about vaccination drives.



6. I saw that population is inversely proportional to vaccine drive from the regression analysis as the r coefficient is negative between population and daily vaccinations.

Analytical Overview

- Excel data was cleaned before analyzing such as duplicate rows and the non-essential column was removed from the databases.
- For all graphs and data visualization, I used different methods to showcase my knowledge of python.
- All major findings and recommendations are based on EDA which is explained in the Documentation Page.

PROPER DOCUMENTATION OF THIS ANALYSIS CAN BE VIEWED THROUGH THE BELOW LINK ADDRESSES.

Python code, Documentation & Presentation

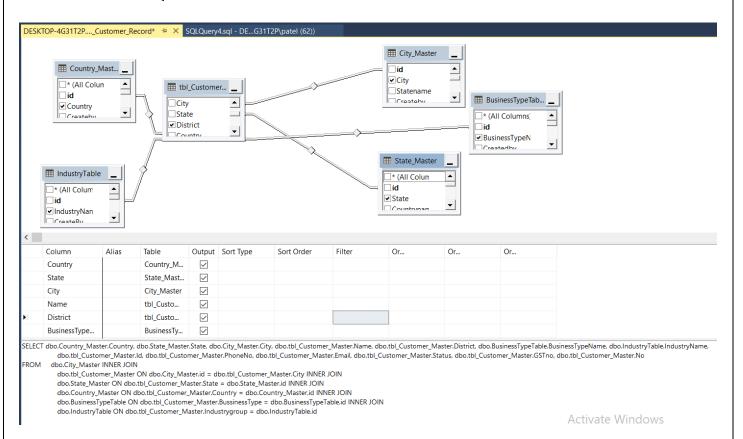
https://github.com/Urvish130/Python_Covid19_Project

4. Previous Projects that included SQL

As a .Net Developer, I used to create stored procedures that were connected to the asp .net framework web application. Below are some examples of what I used to create.

1. Stored Procedure

2. View in MS SQL



Majorly all Stored Procedures I created had inner joins. I found MS SQL more user-friendly as I was able to create inner joins quickly in views and then stored them as procedures.

THE WHOLE CRM PROJECT WITH SQL DATABASE SCRIPT FILE CAN BE VIEWED THROUGH THE BELOW LINK ADDRESSES.

https://github.com/Urvish130/CRM-Client-1

There are some projects which were on MS SQL workbench, which I don't have any access to now due to the company's policy. That was based on the .NET core which had MVC (MODEL VIEW CONTROLLER)

Structure. That had some services which were about storing data from online applications to live databases using JavaScript and converting the data into JSON format.

5. Websites

I also created/designed some websites which are still online. Below are the links for the same and also the source code on GitHub.

WEBSITES SOURCE CODE https://github.com/Urvish130/Websites

R2R website: https://www.r2rgroup.com/

SESPL website: https://sespl-ind.com/

VINMART: https://vinmartgroup.com/