Analyzing Chicago Traffic Accident and Casualty Trend Using Tableau

CIS 5270 Professor Shilpa Balan

Yangyang Jia (yjia12@calstatela.edu)

Department of Information Systems, California State University Los Angeles

TABLE of CONTENTS

1. Dataset URL

Traffic Crashes – People:

https://data.cityofchicago.org/dataset/Traffic-Crashes-People-Dashboard/7fud-yfx4

This dataset contains information about people involved in a crash and if any injuries were sustained from 2015 to 2020. Some people involved in a crash may not have been an occupant in a motor vehicle, but may have been a pedestrian, bicyclist, or using another non-motor vehicle mode of transportation.

Traffic Crashes – Crashes:

https://data.cityofchicago.org/Transportation/Traffic-Crashes-Crashes-Dashboard/8tdq-a5dp

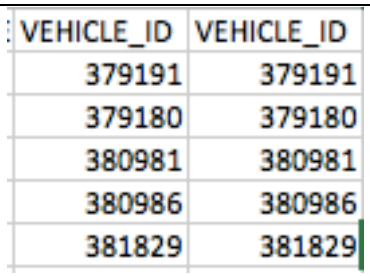
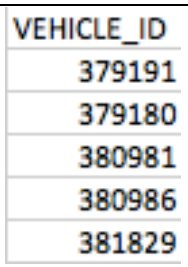
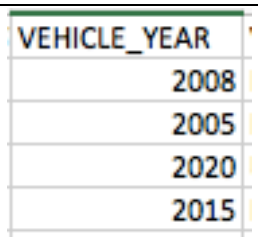
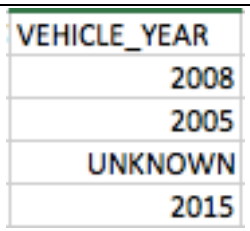
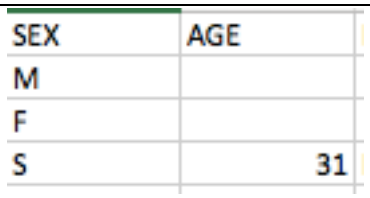
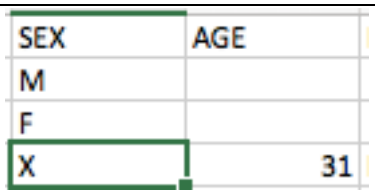
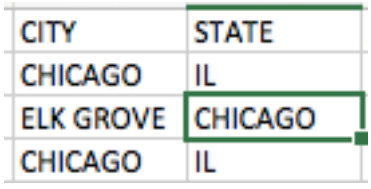
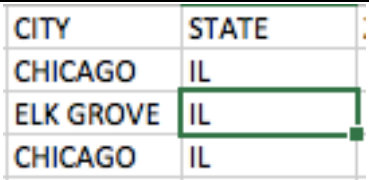
This dataset shows information about each traffic crash on city streets within the City of Chicago limits and under the jurisdiction of Chicago Police Department (CPD) from 2015 to 2020. Many of the crash parameters, including street condition data, weather condition, and posted speed limits, are recorded by the reporting officer based on best available information at the time.

2. Initial Questions

Before exploring the data, I created a list of questions that I want to address:

1) How do the different types of person, gender and age related to the crash?
2) Is there a relationship between the time of day and the day of week to the crash?
3) What's the most contributing factors of the accident?
4) How does location influence the crashes?
5) How does past trend forecast the future?

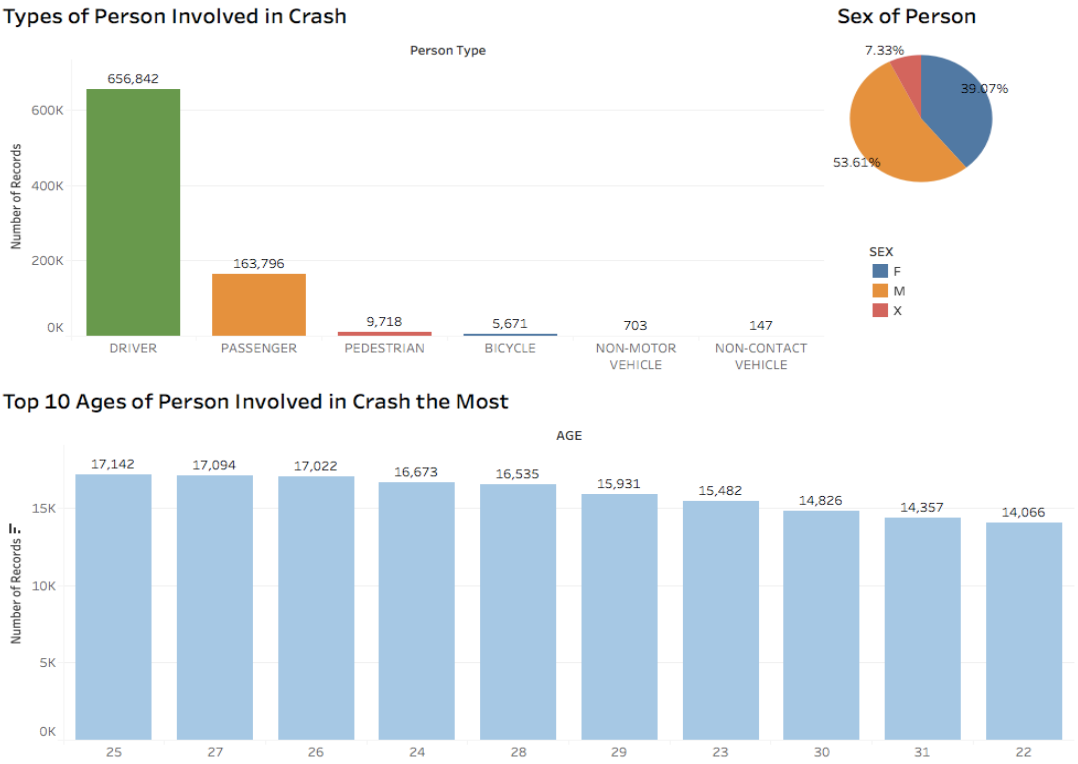
3. Data Cleaning

Score / Problem	Dirty Data	Cleaned Data / Remarks
1. Duplicated Record	 Vehicle_ID is duplicated.	 Delete one of the duplicated columns.
2. Contradicting Records	 2020 is contradicting the records.	 Correct the record to be unknown
3. Illegal values	 “S” doesn’t represent anything for SEX	 Changed “S” to “X” as unknown value.
4. Misfielded values	 “Chicago” as a city name shouldn’t show on the STATE filed.	 Corrected “Chicago” to “IL” as Chicago is in IL State

5. Embedded values	VEHICLE_ID 10 08/04/2015 12:40:00 PM 96 07/31/2015 05:50:00 PM 954 09/02/2015 11:45:00 AM It combined Vehicle_ID and Crashes Time to one Column	VEHICLE_ID | CRASH_DATE 10 | 08/04/2015 12:40:00 PM 96 | 07/31/2015 05:50:00 PM 954 | 09/02/2015 11:45:00 AM 9561 | 10/31/2015 09:30:00 PM Separated two values to two different columns.

4. Data Visualizations & Explanations

I How different types of person, gender and age related to the crash from 2015 to 2020?



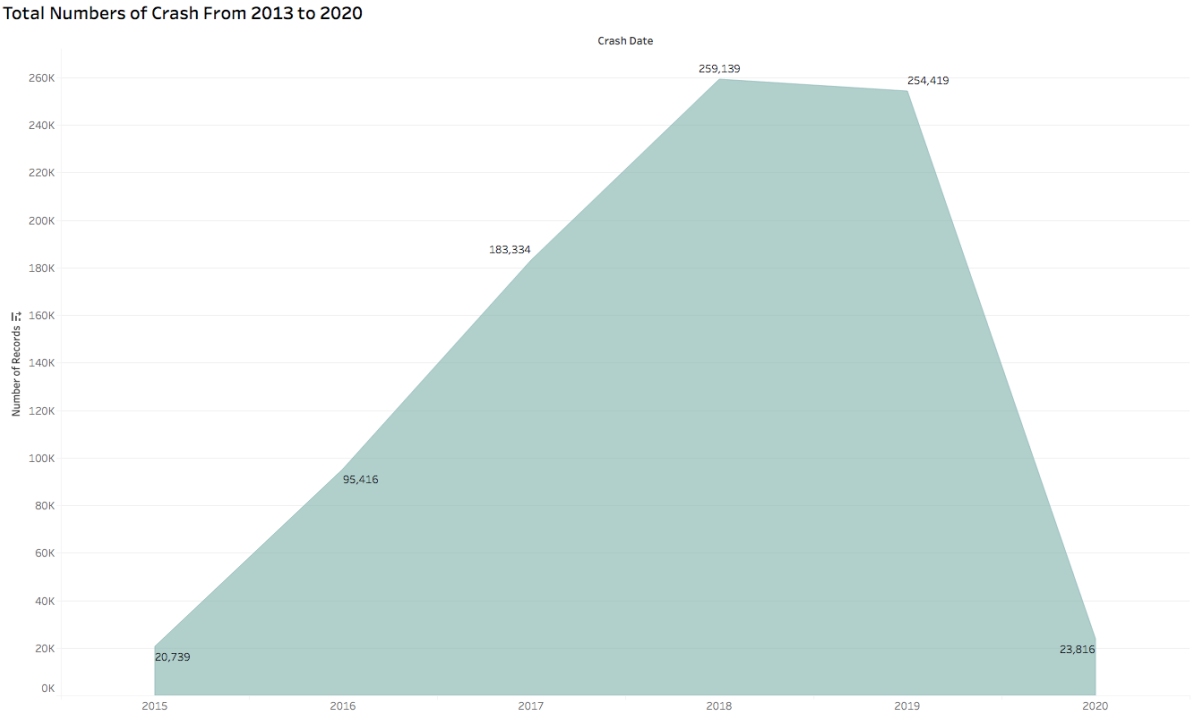
[Fields used: person type, sex, age, number of records]

The above chart tables show the following questions for each year from 2015 – 2020.

- Which types of people are most likely to be involved in a crash?
- What are the most 10 age group of people have the highest rate of accident?
- What's the number difference between male and female in a crash from 2015 to 2020?

This visualization helps us know that the driver and passenger are the top 2 highest type of person that involved in a crash, which reaches to 655,842 and 163,796 in total from 2015 to 2020. Besides, age of 25, 27 and 26 are the top 3 highest total numbers of age group who had crashed from 2015 to 2020, on the contract, the people who above 80 years old is the lowest age group. Moreover, male reaches to 53.61% of people that had crashed, which is 14.6% higher than the female.

II What is the timeline of total amount of crashes from 2015 to 2020?

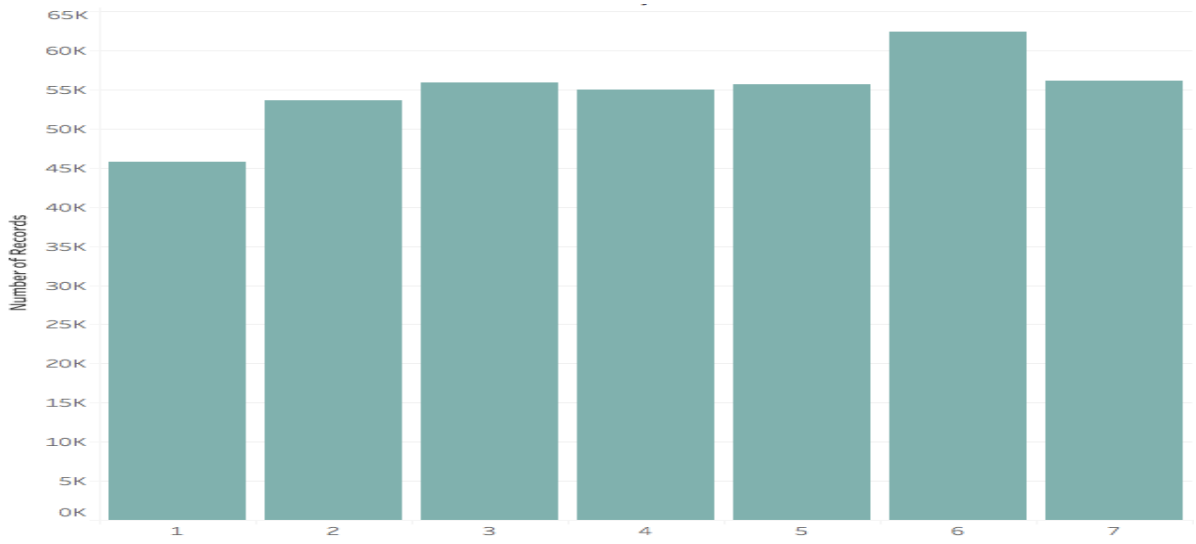


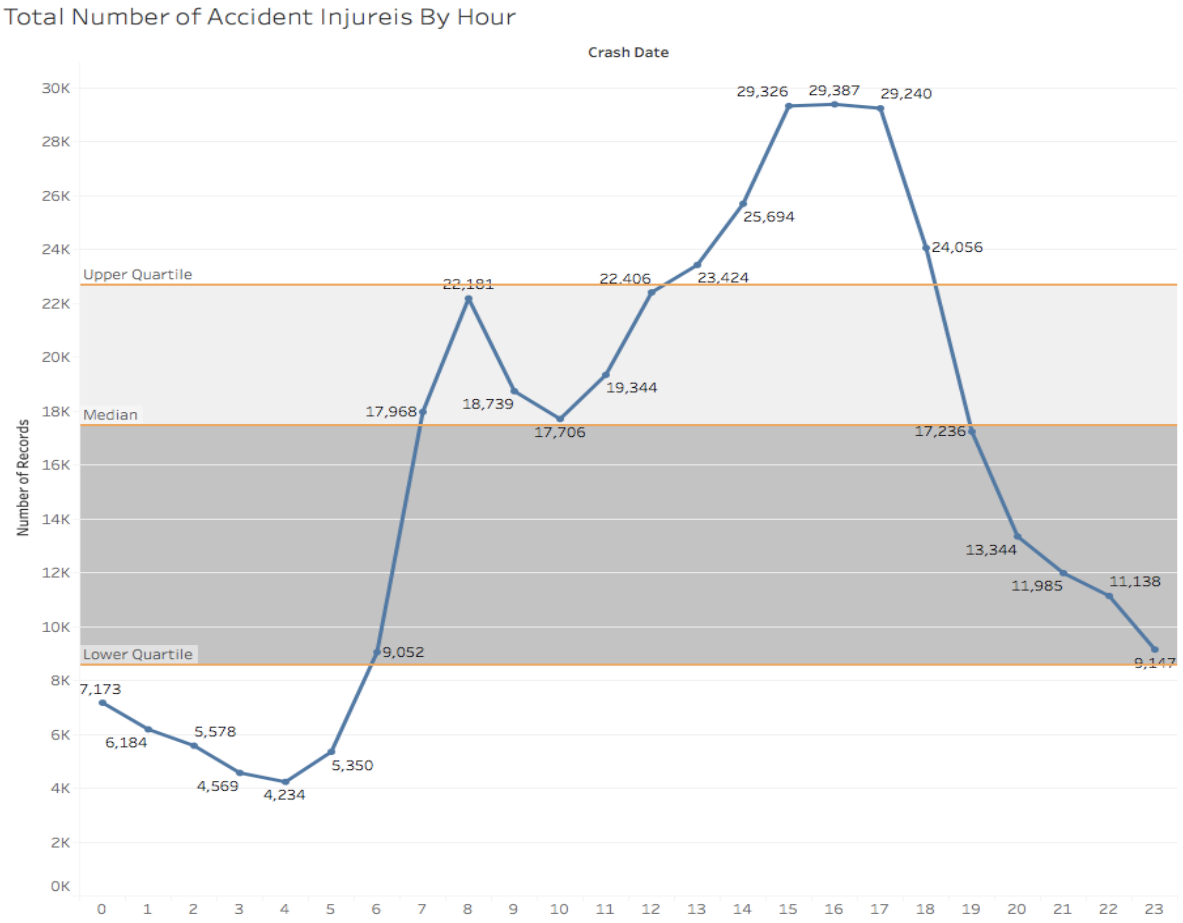
[Field used: crash dates (year), number of records]

The above analysis shows the timeline for the amount total of people that had crashed from 2015 to 2020. As a result, the year of 2018 has the highest rate of total amount of crashes, which reaches to 259,139, the year of 2019 comes the next, which is 254,429. Surprising, the total amount of crashes was growing up from 2015 to 2018, but it was decreased by 4,710 in 2019. So, we could explore deeper about what's the reason for that fall, with the reasonable

study we could expect the total amount of crashes may decrease in 2020 as well based on the trend.

III What is the amount of total number of crashes by day and by hour from 2015 - 2020?





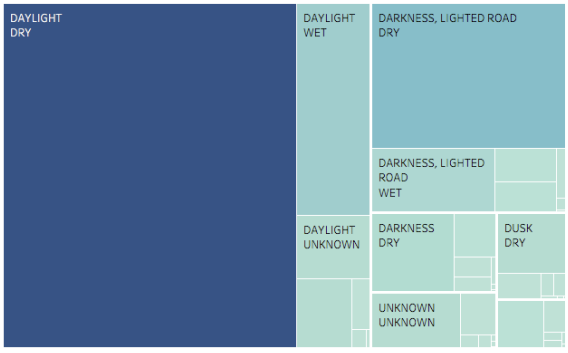
[Field used: crash date (hour), crash date (day), number of records

Tool used: Lines (Continuous), Reference Line - distribution, 4 tiles of quartiles]

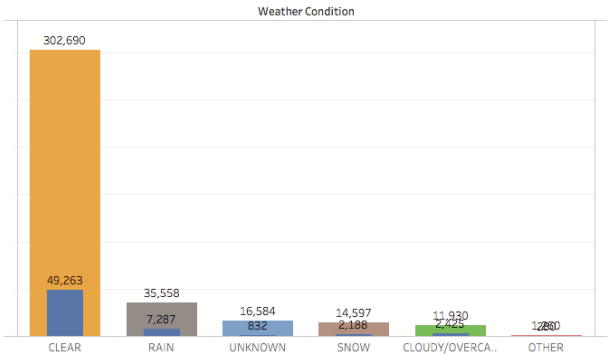
The first bar chart shows how specific day affects the crash, “Sunday = 1”, “Monday = 2”, etc, Interestingly, It shows less accidents happened on Sunday, more accidents happened on Friday and Saturday. The second analysis using a Line chart shows the count of total number of crashes that have happened by hour from 2015 - 2020. By adding the reference line (Distribution, 4 tiles of quartiles) on the hour chart, it shows the lower quartile of total amount of crashes that happened between 0am and 6am. The upper quartile of total amount of crashes are happened between 12pm to 18pm, which has above 75% of total amount of crashes. Besides, the median number lands in 6am to 12pm and 18pm to 23pm.

IV What are the other contributing factors to the crash?

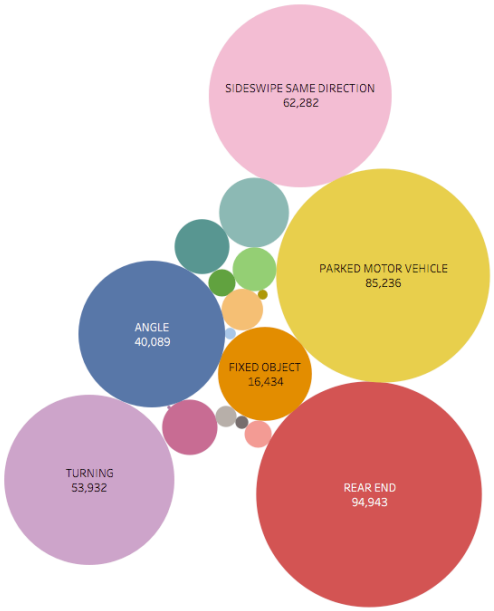
Types of Road Condition That Casue Crash



Types of Weather Condition That Casue Crash and Injuries Fatal



Tyeps of First Crash



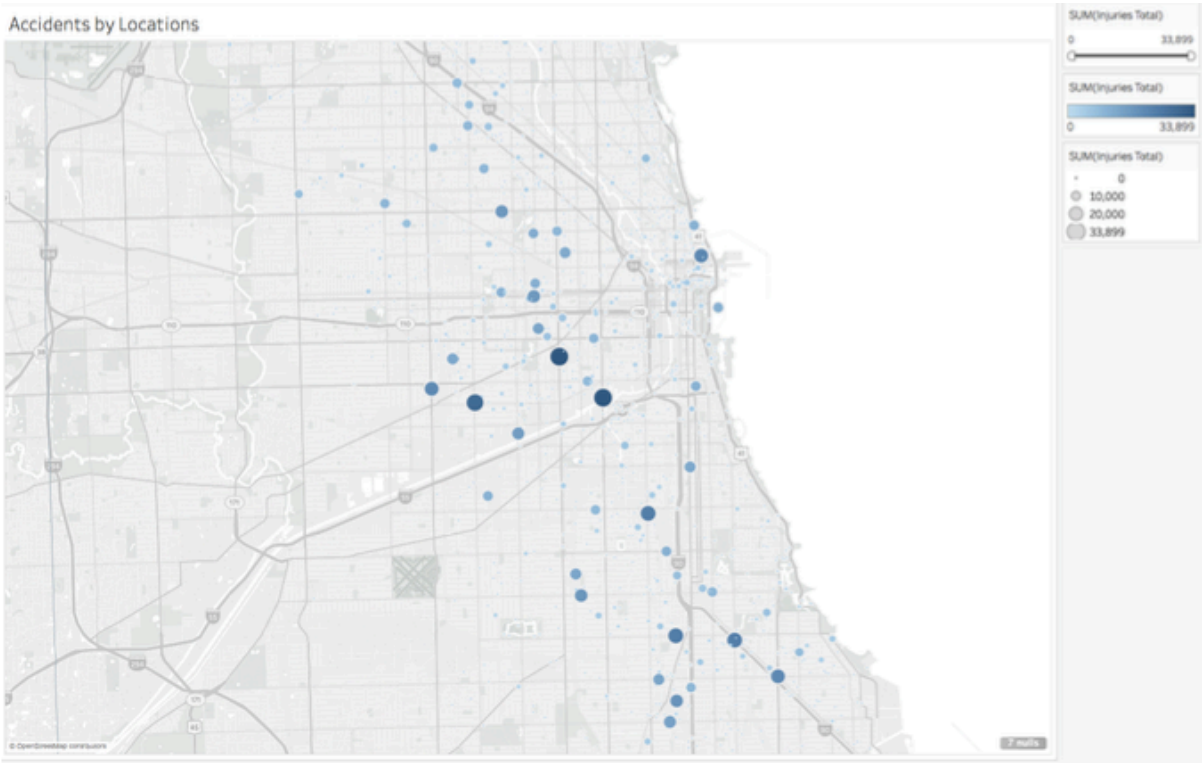
[Fields used: road condition, first crash, weather condition, number of records, injuries total
Tools used: Treemap, packed bubbles, dual axis, synchronize axis]

The treemap shows that the different road conditions which cause the crash from 2015 – 2020. The darker blue color shows the greater amount of total number of crashes. This helps us to know that daylight & dry, darkness & lighted dry road and daylight & wet road conditions has higher possibility of accident crash.

This packed bubbles chart shows that the highest possibilities of first crash type that cause a crash. The bigger size of circle, the greater of the total number of crash happened from 2015 to 2020. As a result, the Crash happened to the "REAR END" has the highest amount of number which reaches to 94,943, and "PARKED MOTOR VEHICLE" comes next, which is 85,236.

The bar char combined number of records and injuries total in a synchronize axis. It shows how the different weather conditions effect on the total number of crashes and the total number of injuries. As a result, with "CLEAR" weather condition has the highest rate of crashes number and injuries number, which are 302,690 and 49,263, "RAIN" comes next as the second highest of number of crashes. The interesting founding is even "CLOUDY/OVERCAST" weather condition has low total number of crashes, but it has highest possibility of injuries among all other weather condition, which reaches to 20.3%, while, "CLEAR" weather condition has 16% possibility that people may get injuried.

V How does the location influence the crash?

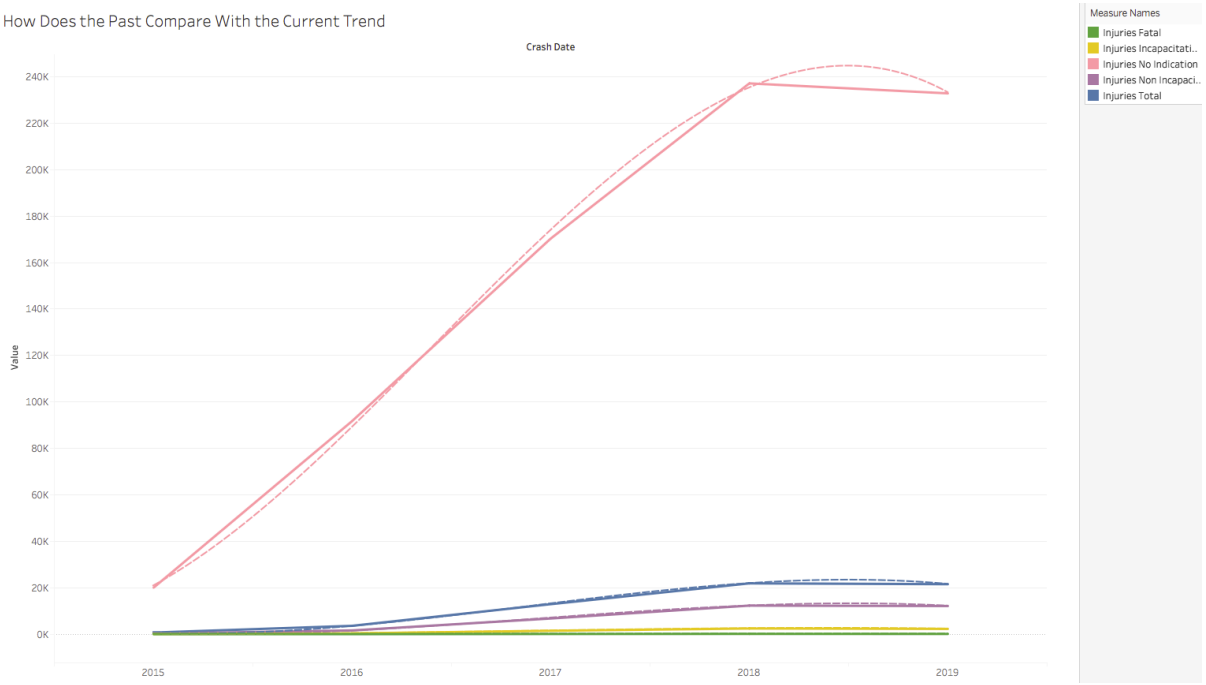


[Fields used: location, number of records

Tools used: symbol map]

The above analysis using a symbol map table shows that the different locations that have the different count of total number of the accident injures from 2015 – 2020. From the table, we could tell, the darker blue and bigger size of circle has greater number of accident injuries. The location (Latitude: 41.8607, Longitude: -87.686) has the highest count of total number of accident injuries which reaches to 33899. While, Latitude: 41.8467, Longitude: -87.666 and Latitude: 41.8450, Longitude: -87.725 reaches to top 2 and 3 highest count of total number of accident injuries. This map simple shows us that the top 3 highest count of total number of accident injuries happen really close to each other, which are within 5 miles to each location.

VI How does the past compare with the current trend?

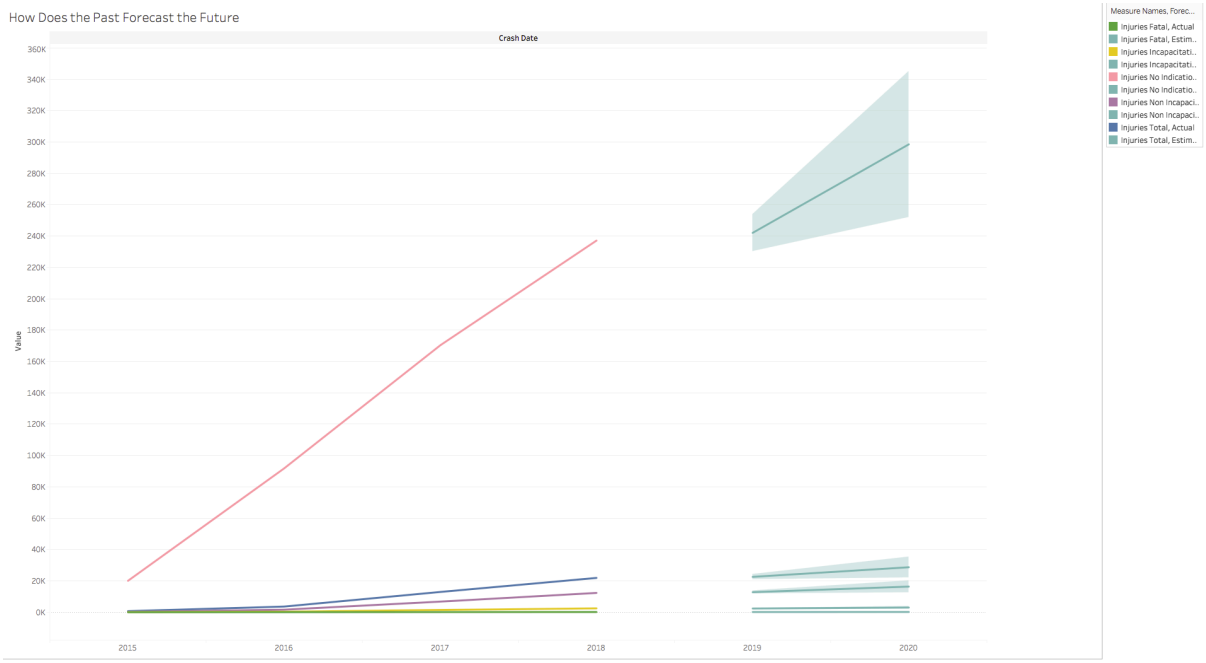


[Fields used: injuries fatal, injuries incapacitating, injuries no indication, injuries non incapacitating, injuries total

Tools used: Line chart, Trend – Polynomial, degrees: 3]

This trend chart shows how the trends of each numbers of injuries fatal, injuries incapacitating, injuries no indication, injuries non incapacitating, injuries total from 2015 to 2019. Overall, from 2015, the number of reported crashes has been trending up for all the severity injury types. However, there has been a slight decrease starting from 2018. Fortunately, injuries no indication has the highest rate of total numbers of crashes from all the time than any other injuries types.

VII How does the past forecast the future?



[Fields used: injuries fatal, injuries incapacitating, injuries no indication, injuries non incapacitating, injuries total

Tools used: Line chart, forecast]

This forecast analysis shows the prediction to the crash possibility in 2020 based on the crashes happened from 2015 to 2019. It includes all the prediction for injuries fatal, injuries incapacitating, injuries no indication, injuries non incapacitating, injuries total. From the forecast, we could tell the number of crashes would still be higher than the past 5 years in 2020.

5. Final Thoughts

In conclusion. Given the massive traffic accident dataset in Chicago, accident occurrence analysis is known to identify the main factors that contribute to crash type, crash location and severity. By analyzing the traffic accident data from 2015 to 2020 in Chicago, USA, I came up with this research paper of accident occurrence analysis and visualization method in both spatial and temporal dimensions, in order to predict when and where an accident with a certain crash type will happen sequentially by whom. But there is still have room for me to improve with the process of researching deeper and deeper.

References:

1. F.Ahmed Malik, “Road Accidents and Prevention” (2017.04)

2. E.Laiza King, “Top 15 Causes of Car Accidents and How You Can Prevent Them” (2016.08)

3.O.Maureen Donnell, “Exploring NYC Vehicle Crash Data in Tableau” (2015. 08)
https://interworks.com/blog/modonnell/2015/08/26/exploring-nyc-vehicle-crash-data-tableau/

4. L.Fan Xiao, “Context-Aware Big Data Analytics and Visualization for City-Wide Traffic Accidents”

5. Jason Chen, “Visualizing NYC Traffic Accidents Before and After Vision Zero” (2017.02)