Chicago Traffic Crashes Analysis

By DATA CHAMPS

Team Members

Shivani Konyala

Yeshwanth Nellikanti

1. Why did your team pick this data set?

As a Team we went through several datasets from the suggested sites, we investigated some of the datasets like dual credit card, cyber security and few other. Eventually, the team selected Chicago Traffic Crashes, as it can be helpful in analysing the traffic crashes of Chicago as it has over 488K rows in total with 49 columns. This dataset has data of the Traffic Crashes in Chicago state from 2013 but the data of all districts is not available until 2017. So, we intended to analyse the data from 2018 as it has information of all the districts of Chicago. Each record of this dataset represents a traffic crash occurred in Chicago with necessary columns like Date and time of the crash occurred, Primary and secondary cause, posted speed, weather and others. This dataset was selected as the most recent data is available (March 2021) and has clear details of the crashes occurred. Based on this dataset we have developed some of the research questions to analyse the traffic crashes dataset.

2. What is the web link to this data set?

Chicago Traffic Crashes URL:-

https://data.cityofchicago.org/Transportation/Traffic-Crashes-Crashes/85ca-t3if

3. Anytime you are dealing with secondary data you need to assess the overall quality of your data set.

a. Does your data set have all the information that you need or are granular details missing that should have been included when the data was gathered?

Traffic crashes dataset contains all the granular details required for two of the research questions the team has picked. But for one of the research questions, Street Name column is required, and it was found missing for a record in the dataset.

The dataset contains 49 columns in total, out of these columns some of the records have blank values but that does not affect the analysis as they are not required to analyse the research questions.

b. Does this impact the creation of your research question?

The Dataset has 488K records of which the street name is missing for one traffic crash. The missing one value is negligible as it contributes almost 0% of the complete dataset.

4. Research Questions:

1. What is the traffic crash rate in Chicago from 2018 to 2020 (Annual analysis)?

2. What are the top streets that have the greatest number of crashes in Chicago from 2018 to 2021? (Street = Street No + Street Name)

3. Which day of the week most of the crashes happened in Chicago from 2018 to 2021?

Rejected Research Questions:

1. What was the Lighting Condition when most of the crashes happened?

2. During which Weather Condition the crashes happened?

3. What is the most common type of traffic reported in Chicago from 2018 to 2021?

4. What is the Prime Cause for the traffic crash in Chicago from 2018 to 2021?