Fatal Traffic Accidents in Chicago City

Onur Tekiner (Solo)

Summary

In my research on traffic accidents in Chicago, I discovered a crucial link between the time of day and the severity of accidents. The data revealed that accidents became significantly worse at night, particularly from 7 PM to 5 AM. This demonstrates the challenges of reduced visibility and other factors that increase the likelihood of fatal accidents during these hours. Additionally, my project explores the patterns of accidents occurring on different weekdays.

Introduction

Understanding why accidents happen and how factors like time and days play a role is what I'm exploring. For instance, as a young person, sometimes we make careless choices, like driving when we shouldn't. This behavior, seen in many people, interests me from a statistical perspective. It's not just about us; traffic accidents affect everyone. According to Federal Highway Administration There were 39,508 fatal motor vehicle crashes in the United States in 2021 in which 42,939 deaths occurred. This resulted in 12.9 deaths per 100,000 people and 1.37 deaths per 100 million miles traveled. The fatality rate per 100,000 people ranged from 5.7 in Rhode Island to 26.2 in Mississippi. The death rate per 100 million miles traveled ranged from 0.71 in Massachusetts to 2.08 in South Carolina. So traffic accidents actually involve every one of us.

I'm curious if specific days or hours relate to more severe accidents. Most people work on weekdays, and Fridays, Saturdays, and Sundays are usually when people relax and have fun. Referencing to National Institue of Health, In the US,The percent of drinkers on each day was as follows: 17.2 % on Monday, 16.0 % on Tuesday, 19.8 % on Wednesday, 21.7 % on Thursday, 51.1 % on Friday, 62.8 % on Saturday, and 28.4 % on Sunday. I want to look into are there any correlation between these days are more fatal accident or not.

Daylight is essential for us. We go to work when the sun rises and return home after it sets. Our vision isn't as good after sunset, which isn't great for drivers. The National Safety Council highlights reasons why driving at night can be riskier than during the day, like shorter days, tiredness, reduced vision, rush hour, and impaired drivers.

I'm curious about how accidents change later in the day. Does reduced vision or alcohol play a role in making accidents more severe? According to the National Institutes of Health, the percentage of accidents varies a lot by time of day. For instance, late nights and early mornings on weekends pose serious risks, especially for shift workers driving home. It's surprising that on

weekday nights from 10 PM to 1 AM, one in 13 drivers is drunk. This high number impacts how severe accidents can be.

Stats from CrashStats show that nearly half of passenger vehicle fatalities occur at night, even though only a quarter of travel happens during these hours. This shows that the fatality rate per mile traveled is about three times higher at night than during the day. Understanding these connections between time, behavior, and accidents is crucial.

Process

Data Description

I used data about traffic accidents from the City of Chicago Government's website. But it was missing data before 2017. To have a better dataset, I compared it with other sources and found that info before 2017 was also missing. So, I focused on data from 2017 to 2023, which had 362,881 rows and 46 features.

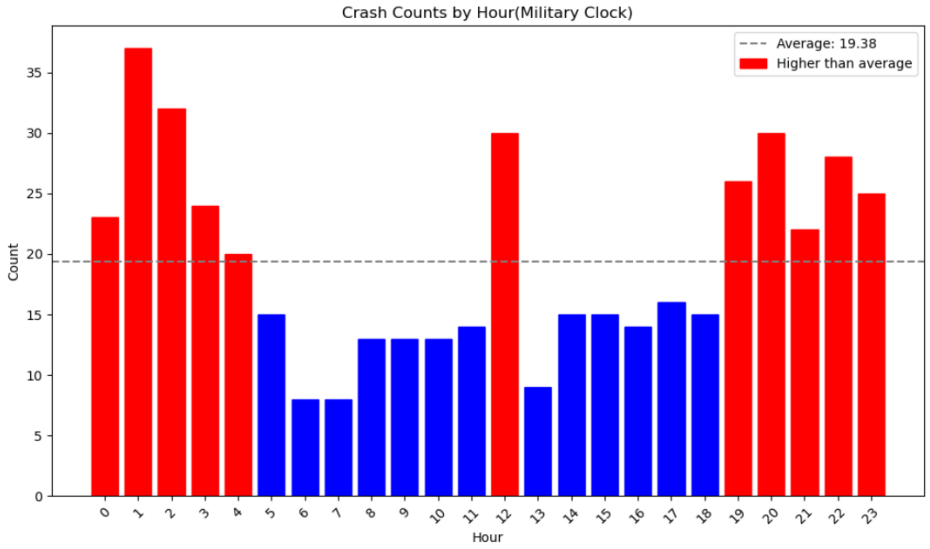
Data Preparation

- I got rid of columns mostly filled with nothing.
- For columns with a few empty spots, I didn't want to lose data, so I filled with the most common value for categories variables.
- There were empty spots in the latitude and longitude columns, so I filled those with the average location.

Data Analysis

My initial attempt involved examining the "Hit and Run" column. Most values were either null or marked as "yes" or "no." I explored if there was any connection between injury rates and the missing data in the "Hit and Run" column. For instance, if the majority of entries were 'No,' I thought the missing values might actually mean 'Yes' due to potential data entry errors. However, after investigating correlations with injury rates, value counts, and other factors, I couldn't identify any significant pattern. Consequently, I decided to remove the 'Hit and Run' column from the analysis.

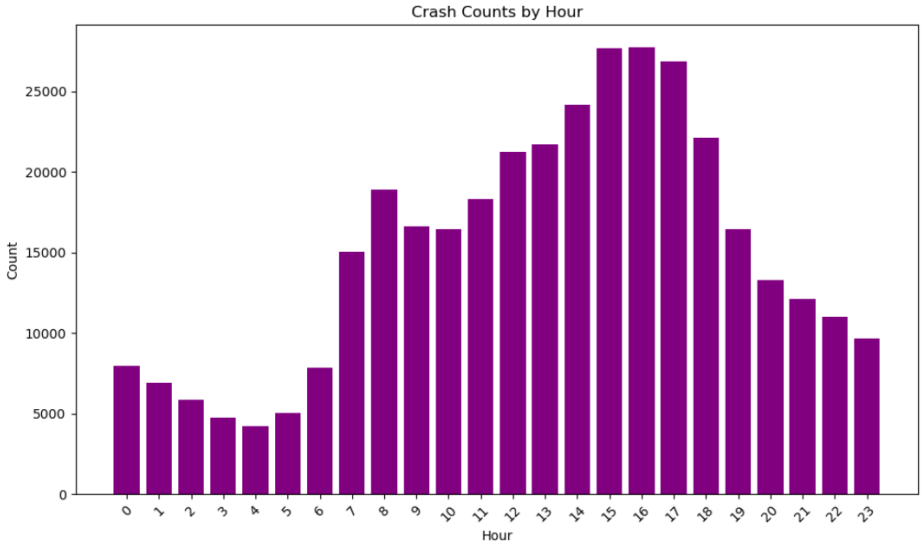
One of my goals was to compare fatal accidents during the day and night. I made graphs showing the count of fatal accidents by the time of day. The graph uses a 24-hour timeline where 1 PM becomes 13:00, 2 PM becomes 14:00, and so on. Here is the graph;



The red bars show a higher number of fatal crashes than average. We can observe that from 6 PM until 5 AM, there are more dangerous traffic accidents. This aspect points out a shortcoming in my analysis. I didn't check the distribution of the dataset. If the accident rate is higher during these times, it's understandable that fatal crashes would be more frequent during these hours compared to daytime.

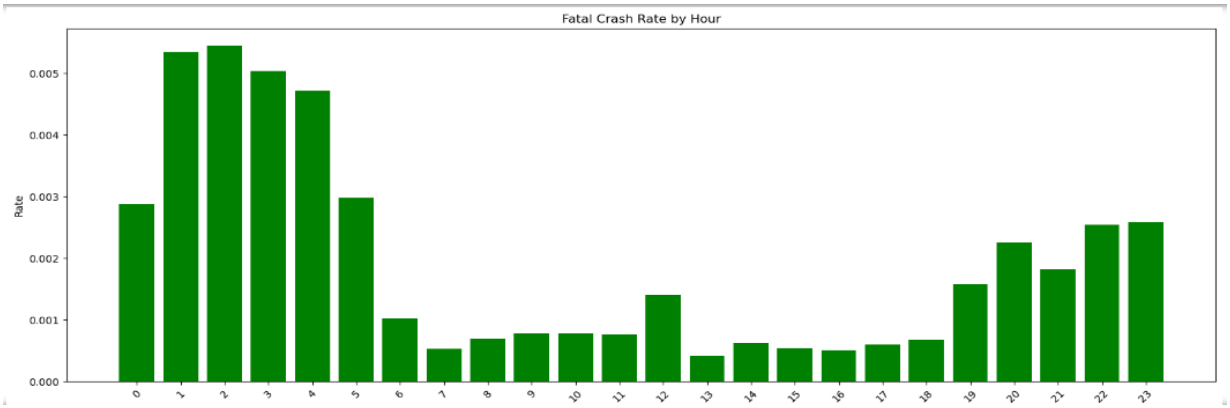
Next, I aim to explore the distribution of non-fatal traffic accidents to ensure I'm interpreting the data accurately.

Let's examine the traffic crashes that aren't fatal:



This is getting interesting because the graphs indicate that the data distribution leans more towards daytime accidents than nighttime. These results reveal something about the severity of

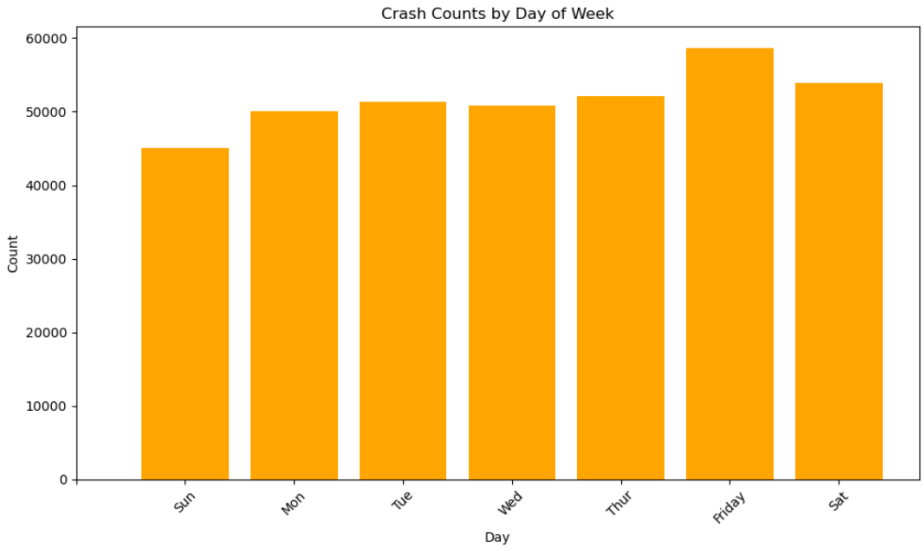
nighttime crashes. Let's take another step forward and calculate the ratio of fatal accidents by dividing the total number of fatal accidents by total accidents.



The graph distinctly illustrates a significantly higher ratio of fatal accidents occurring during the nighttime, specifically from 7 PM to 5 AM, compared to daytime.

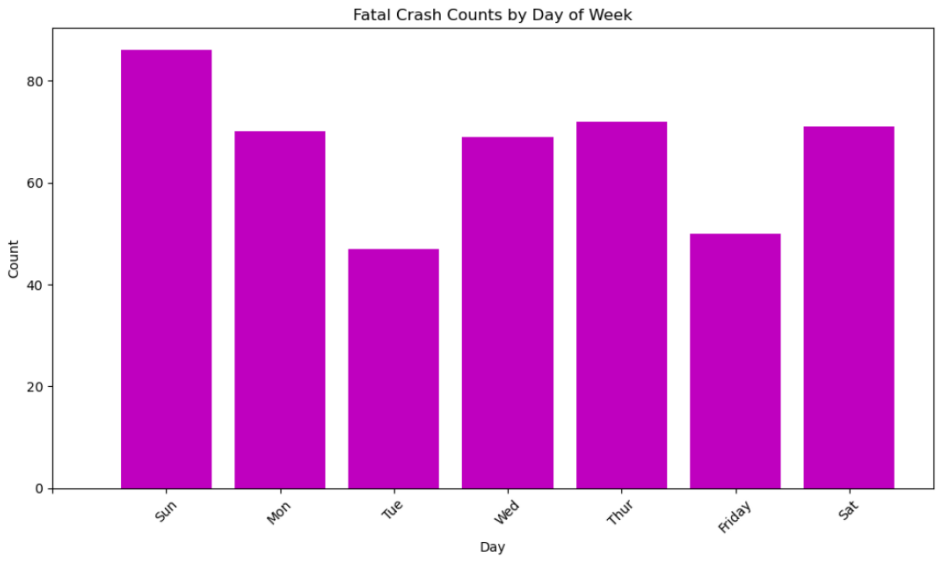
As for another objective, I'm exploring how weekdays correlate with fatal accidents. I hypothesize that Fridays and Saturdays might pose higher risks for driving. According to the National Safety Council, fatal car crashes were more frequent on weekends in 2022, with a peak on Saturdays. Conversely, nonfatal crashes tended to be more common on weekdays, peaking on Fridays. I aim to investigate if the Chicago dataset reflects a similar pattern observed by the National Safety Council.

Let's examine the distribution of the dataset for further insights.



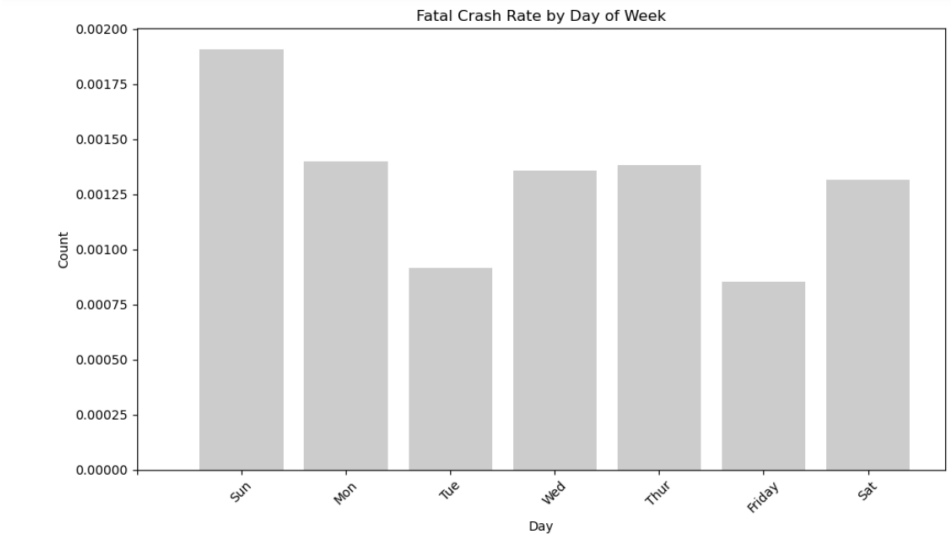
Here, we observe that Friday records the highest rate of traffic accidents compared to other weekdays, aligning with the observation made by the National Safety Council.

Now, let's analyze the total number of fatal accidents categorized by each weekday.



It shows that Sunday had the highest number of fatal accidents in Chicago. This finding differs significantly from the National Safety Council's data, but it's within acceptable bounds.

Now, let's examine the fatal accident rate, which is calculated by dividing the total number of fatal accidents by total accidents.



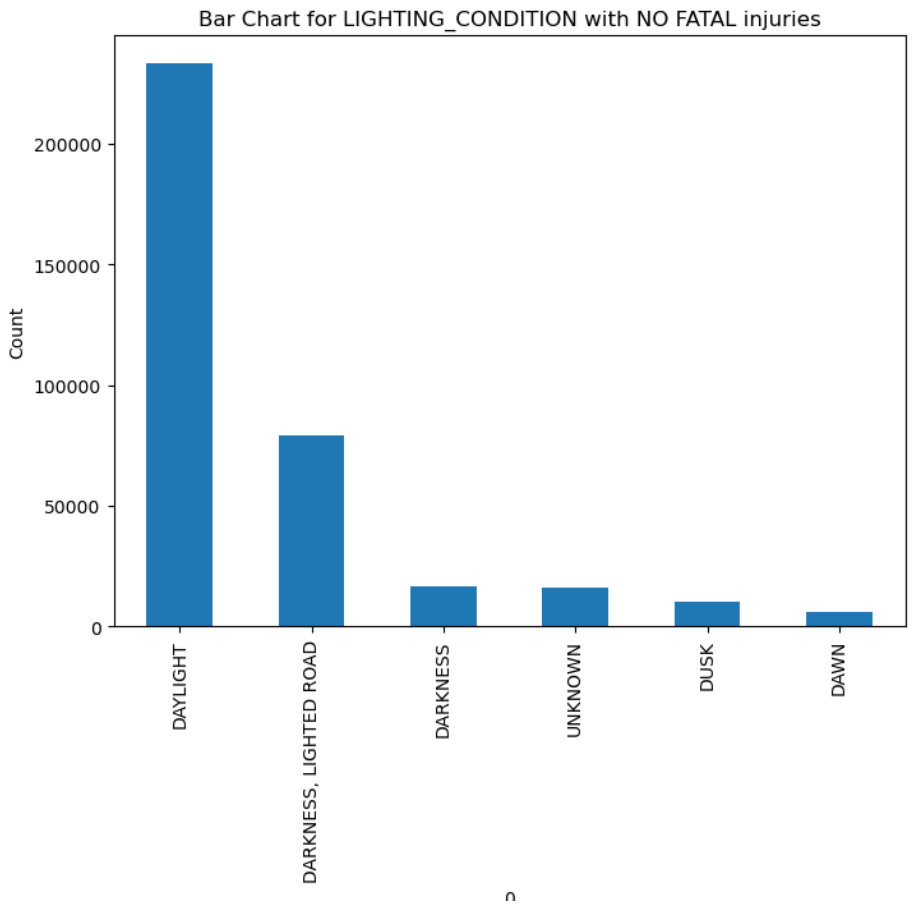
The data shows a similar trend, with Sunday being the day with the highest number of fatal crashes.

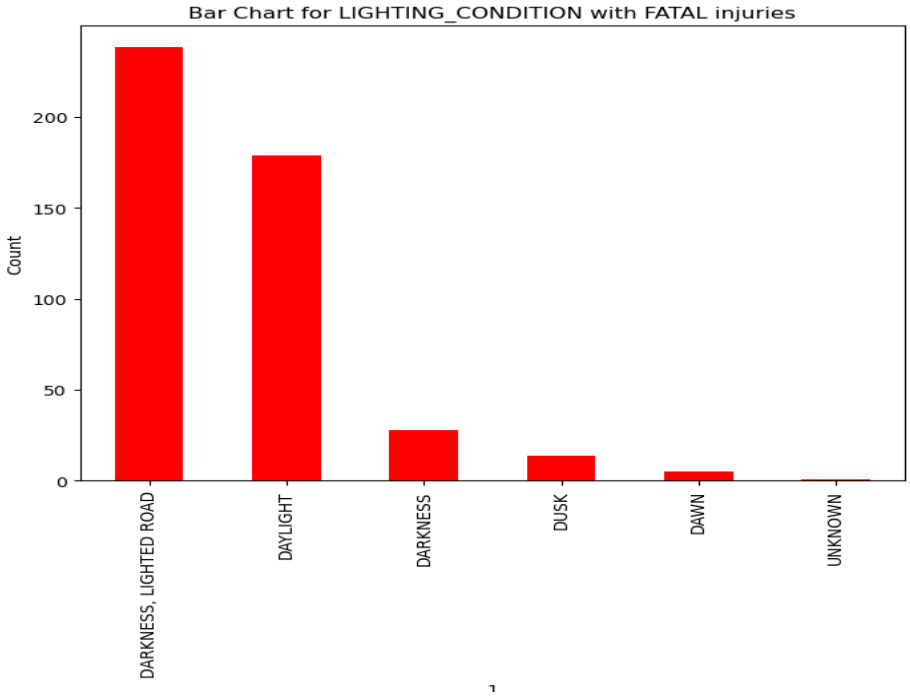
There could be several reasons for this difference from the National Safety Council's data. One potential reason might be the relatively low quantity of fatal crash data available for Chicago. If we had a larger dataset regarding fatal crashes, the results might align more closely with general averages. Another possible explanation is that different regions may exhibit different patterns. The National Safety Council's data represents the entire US, while my analysis focuses solely on Chicago.

Another objective I pursued was understanding the causes behind fatal traffic accidents. I aimed to illustrate the potential reasons for fatal accidents using a bar chart. During this process, I had to adjust my visualization methods to ensure comparability between fatal and non-fatal accidents. I realized that by visualizing only the causes of fatal accidents, it became challenging to compare them with non-fatal accidents and identify patterns.

I carefully analyzed various potential variables to uncover patterns related to fatal accidents. I discovered some valuable insights. The red bars represent fatal traffic accidents, while the blue bars depict non-fatal accidents.

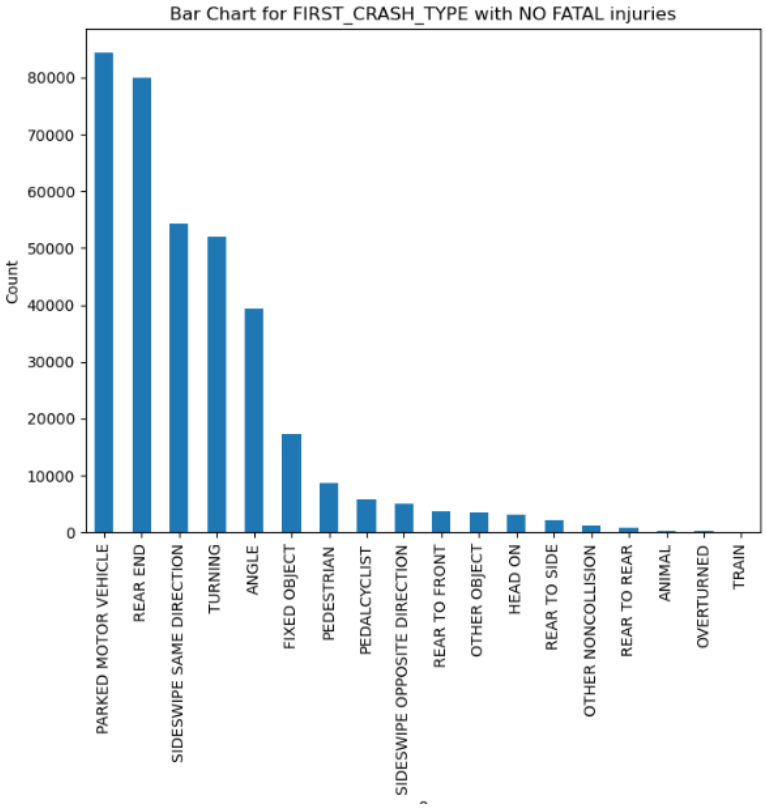
Let's delve into one of these variables, starting with the lighting conditions of the road when accidents occur.

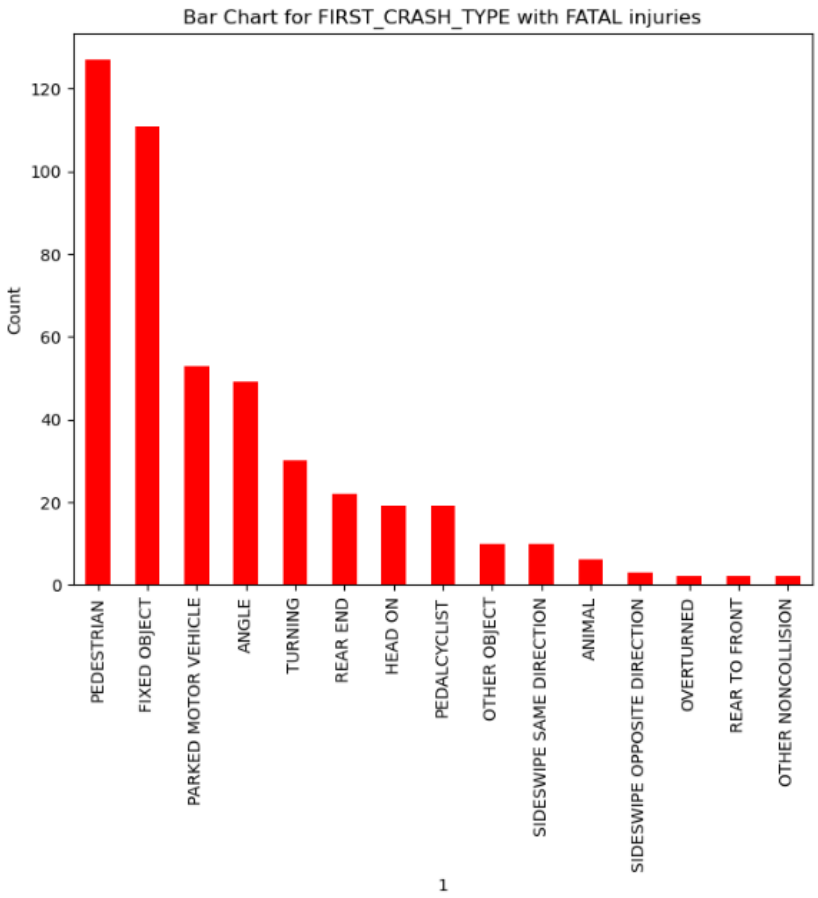




It's evident that nighttime is associated with fatal accidents, which aligns with the earlier findings in our analysis.

Moving on to another variable, let's explore the primary type of crash and compare the differences between fatal and non-fatal crashes.


Bar Chart for FIRST_CRASH_TYPE with NO FATAL injuries
Count
80000
70000
60000
50000
40000
30000
20000
10000
0
PARKED MOTOR VEHICLE
REAR END
SIDESWIPE SAME DIRECTION
TURNING
ANGLE
FIXED OBJECT
PEDESTRIAN
PEDALCYCLIST
SIDESWIPE OPPOSITE DIRECTION
REAR TO FRONT
OTHER OBJECT
HEAD ON
REAR TO SIDE
OTHER NONCOLLISION
REAR TO REAR
ANIMAL
OVERTURNED
TRAIN



From these two graphs, it's apparent that fatal crashes are much more commonly associated with pedestrians than non-fatal crashes.

After exploring the patterns between fatal and non-fatal accidents based on categorical variables, I applied One-hot encoding to convert my categorical variables into numerical columns. This encoding increased the number of columns to 260. Due to the high number of dimensions, I aimed to use feature selection techniques to reduce the dataset's dimensionality.

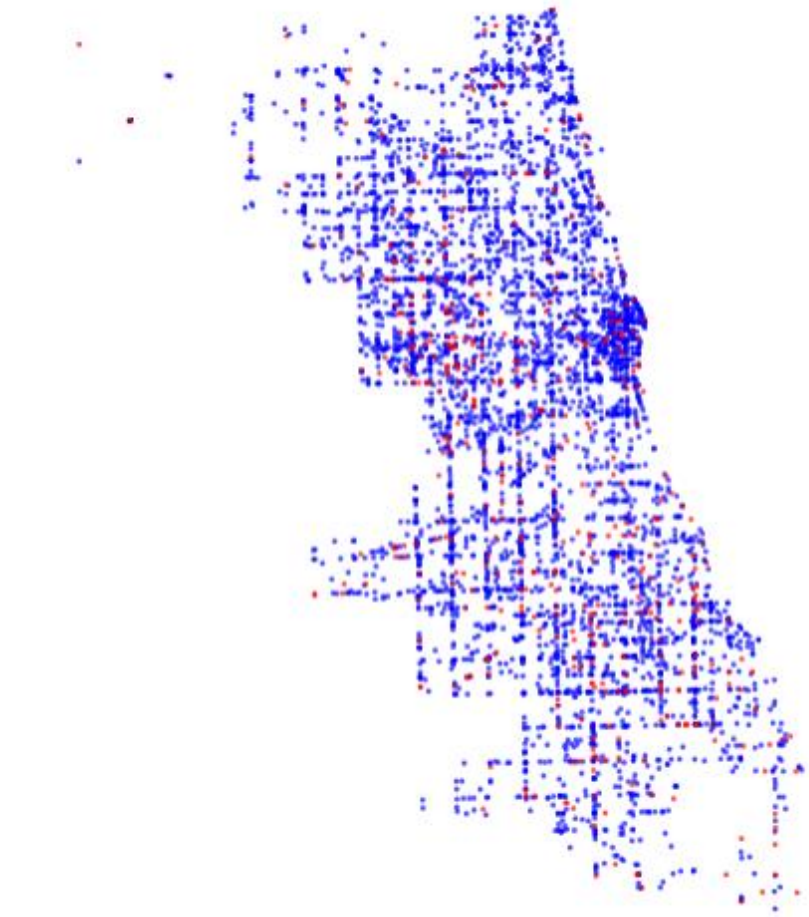
I conducted correlation tests and identified the top 10 important features. Here is the list of columns:

```
In [114]: top_features = correlation_with_target.iloc[1:].head(10).index.tolist()
          top_features

Out[114]: ['INJURIES_FATAL',
           'INJURIES_TOTAL',
           'MOST_SEVERE_INJURY_NO INDICATION OF INJURY',
           'FIRST_CRASH_TYPE_PEDESTRIAN',
           'CRASH_TYPE_NO INJURY / DRIVE AWAY',
           'CRASH_TYPE_INJURY AND / OR TOW DUE TO CRASH',
           'INJURIES_INCAPACITATING',
           'PRIM_CONTRIBUTORY_CAUSE_PHYSICAL CONDITION OF DRIVER',
           'INJURIES_NO_INDICATION',
           'FIRST_CRASH_TYPE_FIXED OBJECT']
```

The important features include the number of injuries, the presence of fatalities, pedestrian involvement in the crash, the physical condition of the driver, and more.

Lastly, the dataset contains longitude and latitude columns. Due to Altair's visualization limit of 5000, I sampled 5000 non-fatal accidents. These are represented by blue dots on the map. Fatal accidents are indicated by red dots on the map. Here is the Altair map visualization of Chicago.



The map didn't show any significant patterns for fatal or non-fatal traffic crashes as I expected. Even if there were some patterns in this sample of non-fatal accidents, the maximum sample size for Altair, which is 5000, isn't enough to prove all the data patterns are accurate since there are more than 300 thousand data entries.

Discussion

The number of fatal accidents is too low to identify overall patterns.

What could be considered as fatal or not? What are the parameters for labeling accidents as fatal or non-fatal?

Conclusion

I can conclude that there is strong evidence that traffic crashes after sunset are more dangerous than daytime crashes. Factors such as reduced visibility and possible alcohol consumption during these hours might contribute to this increased risk. However, it's essential to explore this further with more data and detailed analysis to understand the complete picture.

I've learned that interpreting data involves more than just looking at the numbers. It requires considering various factors, cleaning the data for accuracy, and understanding its context. This understanding is critical for effective data visualization, ensuring that our insights are meaningful and accurate.

Future Work

In future investigations, it's crucial to expand the dataset to get a clearer view of the patterns and causes behind fatal crashes. This might involve exploring how different factors, such as alcohol consumption, weather conditions, or road infrastructure, contribute to these accidents. Such a comprehensive analysis could provide more robust insights into preventing fatal accidents.

References

Federal Highway Administration Fatal Accidents Rate in the US

https://www.iihs.org/topics/fatality-statistics/detail/state-by-state#:~:text=Posted%20May%202023.-,Fatal%20crash%20totals,Island%20to%2026.2%20in%20Mississippi.

National Institue of Health Day's Stats

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5898433/#:~:text=The%20percent%20of%20drinkers%20on,%2C%20and%2028.4%20%25%20on%20Sunday.

National Safety Council

https://www.nsc.org/road/safety-topics/driving-at-night#:~:text=Shorter%20days%2C%20fatigue%2C%20compromised%20night,than%20driving%20during%20the%20daytime.

National Institue of Health Day's Drunk Drivers Stats

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3400188/#:~:text=The%20proportions%20also%20vary%20greatly,drunk%20(BAL%E2%89%A50.08%25).

CrashStats

https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/810637#:~:text=Nationwide%20almost%20half%20(49%25),night%20than%20during%20the%20day.

Bytebloc

https://www.bytebloc.com/Help/Content/24_Hour_Time_Format.htm#:~:text=Many%20places%20in%20the%20world,%3A00%2C%20and%20so%20on.