Leveraging Python for Advanced Traffic Crash Analysis: Insights, Patterns, and Predictive Solutions in Chicago

Abstract:

In 2024, traffic crashes across Chicago's city streets were closely examined using data from the Chicago Police Department (CPD), sourced through Data.gov. This analysis finds a detailed picture of the city's road safety challenges and opportunities for improvement. Every year, more than 900,000 crashes occur, and a striking insight is that over 700,000 of these accidents take place under clear weather conditions. This suggests that factors such as driver behavior, traffic volume, and infrastructure play a far larger role in accidents than adverse weather conditions. Breaking this down further, it means that on average, about 2,500 crashes happen each day, with the majority occurring when visibility and road conditions are optimal. This emphasizes the need to focus on improving road safety through better traffic management, enhanced driver education, and infrastructure improvements rather than solely focusing on weather-related risks.

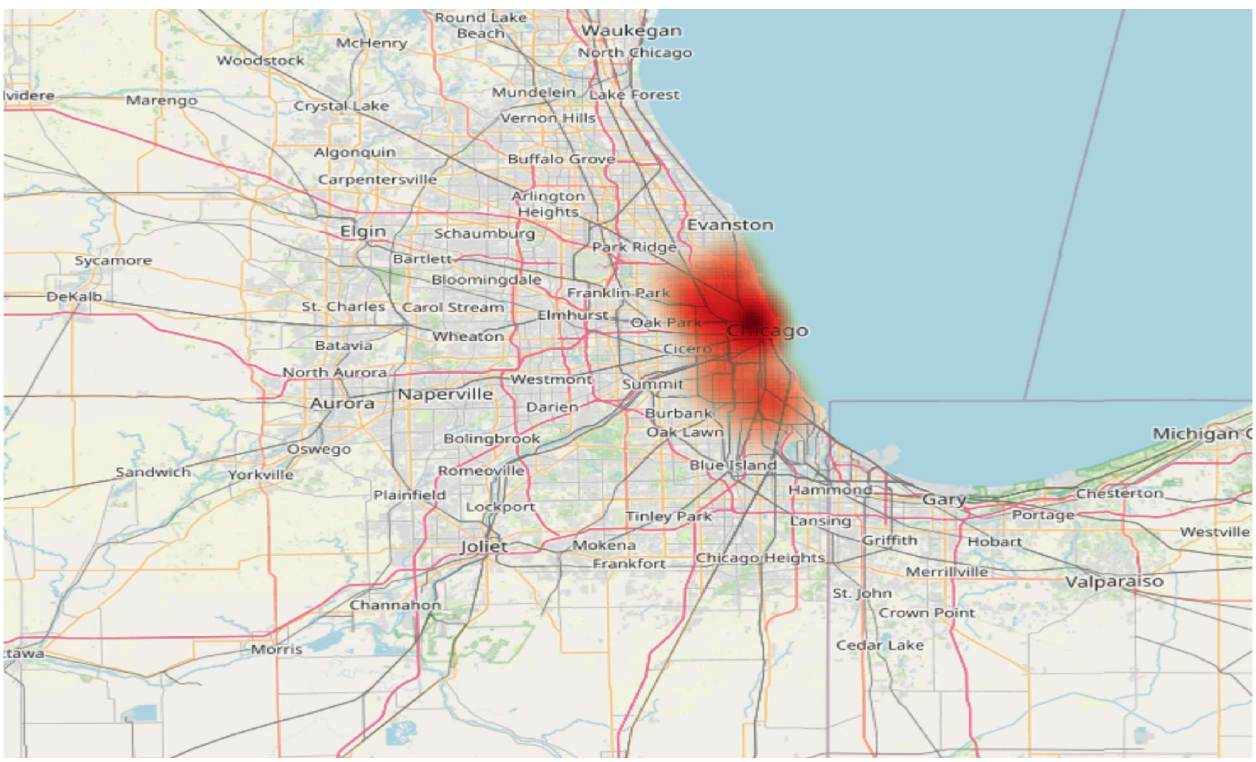

Figure 1: Heat map of Chicago traffic accident (QGIS: 2025)

Rainy weather came in second, while a significant number of crashes were recorded under "unknown" weather conditions, pointing to gaps in data reporting. Timing also stood out.

October saw the highest number of crashes, with late afternoons between 3 PM and 5 PM being the riskiest time of day.

These patterns emphasize the importance of addressing specific time periods and seasonal factors to reduce accident risks. When it came to crash types, collisions with parked vehicles were alarmingly common over 200,000 incident followed by rear-end crashes. While pedestrian and cyclist-related crashes were less frequent, they posed higher risks to vulnerable road users, underscoring the need for focused safety measures for these groups. The data also highlighted issues with traffic control. More than 500,000 crashes occurred in areas without any traffic controls, but even intersections with functioning signals and stop signs weren't immune, as driver non-compliance and design flaws contributed to accidents. Injury severity analysis revealed that most crashes caused non-incapacitating injuries, but more serious injuries and fatalities were concentrated in areas with insufficient safety measures. By using Python to analyze and visualize this data, the study provides actionable insights to make Chicago's streets safer. It calls for enhanced traffic controls, improved public awareness, and focused interventions during high-risk times and locations. These steps can help reduce both the frequency and severity of crashes across the city.

Introduction

Traffic accidents have become a pressing issue for cities around the globe, and Chicago is no exception. As one of the largest and busiest cities in the United States, Chicago has long grappled with the challenges posed by traffic-related incidents. (Wenzhao Zhang, 2020) In 2024, the city faced a staggering 900,000 traffic crashes, highlighting the complex mix of urban density, driver behavior, and infrastructure issues contributing to this high rate of accidents. This article delves into Chicago's 2024 traffic crash data, analyzing key factors such as road conditions, weather, traffic control measures, and crash types. By studying these patterns, we aim to uncover meaningful trends and identify areas where targeted interventions could significantly enhance road safety. The data, sourced from the city's official crash reports via Data.gov, sheds light on the urgent need for action. (Portal, 2024) Whether it's improving traffic controls, addressing infrastructure design, or raising driver awareness, this analysis seeks to guide efforts to make Chicago's streets safer for everyone. (Xu, Gao, Zuo, & Ozbay, 2024)

Data_Source_(*https://data.cityofchicago.org/api/views/85cat3if/rows.csv?accessType=DOWNLOAD)*

Crashes due to weather condition:

Figure 2 reveals how different weather conditions influence traffic crashes in Chicago. Surprisingly, clear weather accounts for the majority of accidents, with over 700,000 incidents recorded. While we often associate bad weather with higher crash risks, this data suggests otherwise. Factors like heavy traffic, overconfidence, or higher speeds on clear days may play a significant role in these numbers. It seems that clear skies, often thought to be ideal for driving, come with their own set of challenges tied to driver behavior and systemic issues.

Rainy weather follows as the second-most common condition for crashes, with "Unknown" weather not far behind. The large number of crashes attributed to "Unknown" conditions points to potential gaps in data collection, making it harder to fully understand weather-related risks. Snow, cloudy skies, and freezing rain or drizzle also contribute significantly, highlighting the importance of extra caution during poor visibility or slippery road conditions.

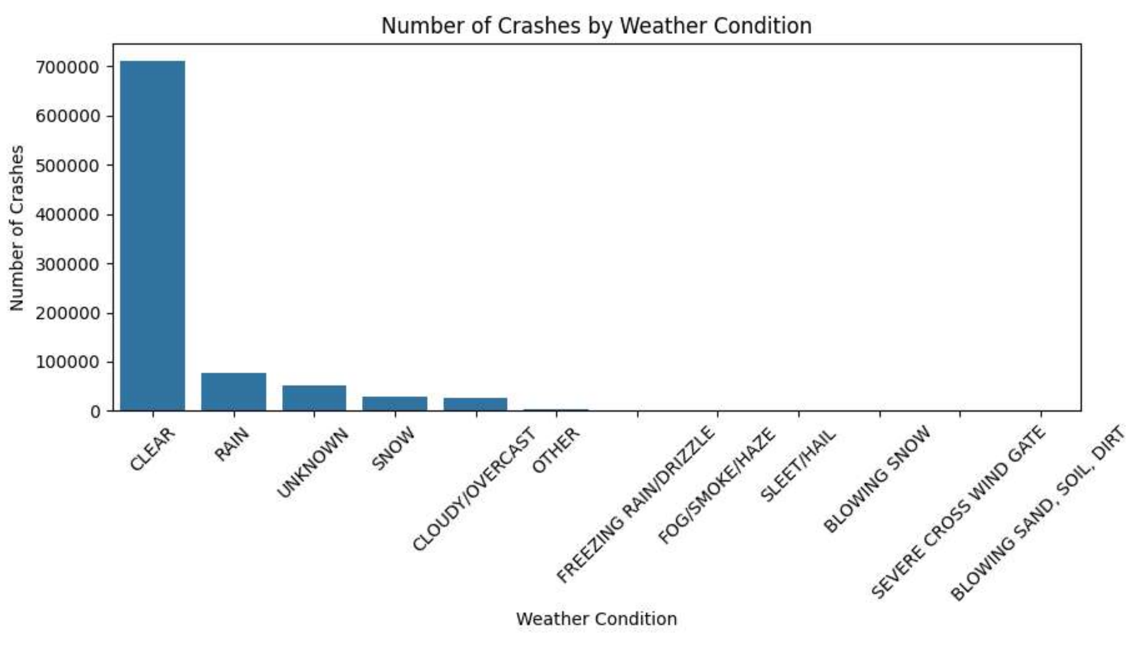

Figure 2: Number of Crashes by weather Condition

Less common weather events like fog, sleet, blowing snow, and strong crosswinds contribute far fewer crashes, reflecting their rarity in Chicago's climate. However, these situations still pose dangers and underscore the importance of being prepared for unusual but potentially hazardous conditions. The findings emphasize the need for targeted driver education and awareness efforts. For clear weather, this means addressing issues like speeding and complacency. Additionally, improving how weather conditions are reported in crash records is crucial for designing data-driven strategies to enhance road safety in all conditions.

Crash types Toe due to Injury vs Drive away:

The total injuries caused by different types of crashes in Chicago, offering valuable insights into the city's traffic safety challenges. The data reveals that the vast majority of injuries occur in crashes categorized as "Injury and/or Tow Due to Crash." These are typically severe collisions that require vehicles to be towed from the scene, underscoring the serious public health impact of high-intensity crashes. Such incidents also place a heavy burden on emergency services, from first responders to medical facilities.

On the other hand, crashes labeled as "No Injury / Drive Away" contribute only a tiny fraction to the overall injury count. These minor incidents, where drivers can leave the scene without assistance, suggest the effectiveness of modern vehicle safety features or the lower severity of these collisions.

The stark contrast between these two categories highlights the need for focused efforts to prevent serious crashes. Strategies like stricter enforcement of speed limits, smarter road designs, and increased public awareness of risky driving behaviors could play a key role in reducing the frequency and severity of high-impact collisions. By addressing the root causes of these severe incidents, we can take meaningful steps toward improving safety on Chicago’s roads.

Crash Data in month and hour basis

The figure 3 illustrates how traffic crashes in Chicago vary throughout the year, showing clear seasonal patterns. Crashes begin at a moderate level in January and February, gradually climbing as the weather warms in spring and summer. The numbers reach their highest point in October, likely reflecting increased travel and activity during this time of year. Interestingly, there’s a slight dip in crashes during November and December. This decrease could be tied to reduced travel during the colder months or changes in driving habits around the holidays. These trends highlight how factors like seasonal activities, weather, and traffic volume can influence the frequency of crashes, offering insights into when and why accidents are more likely to happen.

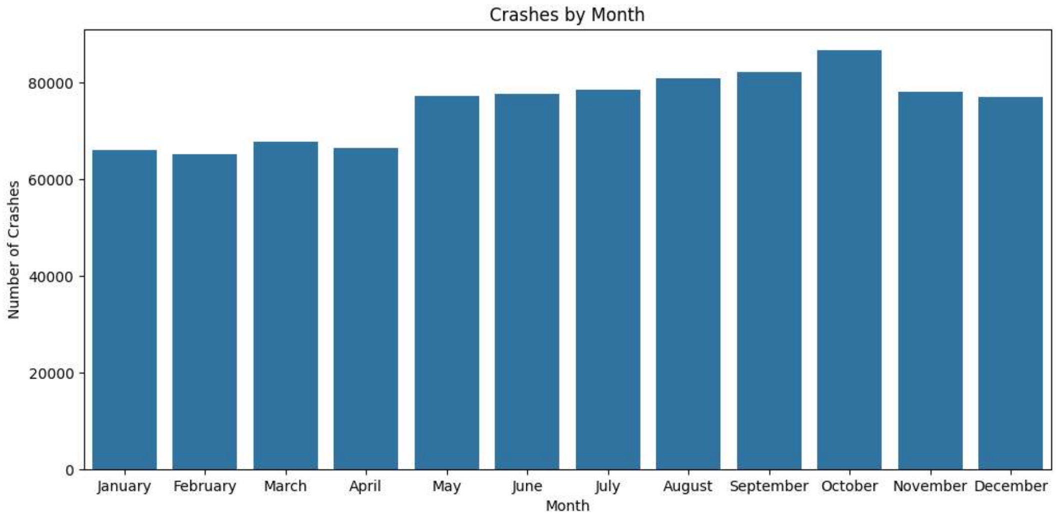

Figure 3: Crashes by Month

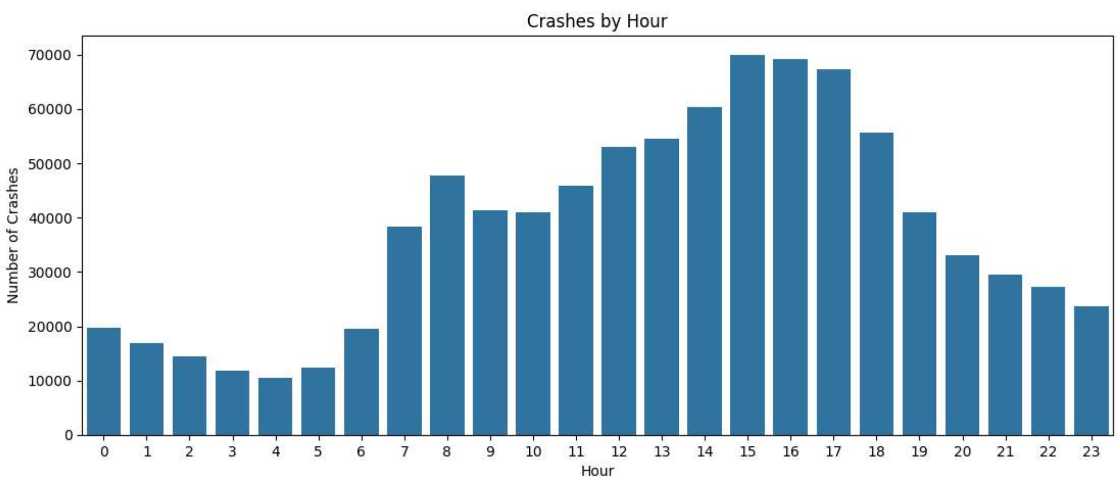

Figure 4: Crashes By Hour

The figure 4 highlights how traffic crashes in Chicago vary by the hour, with noticeable spikes in the late afternoon, especially between 3 PM and 5 PM. These times coincide with peak commuting hours, when roads are busiest, and factors like driver fatigue, stress, or a sense of urgency can increase the risk of accidents. In contrast, the early morning hours, from midnight to

5 AM, see the fewest crashes, likely due to lighter traffic and fewer vehicles on the road during these quieter hours.

When comparing this data with monthly trends, it's clear that traffic crashes are shaped by both seasonal and daily patterns. Monthly variations suggest the influence of factors like weather, daylight, and seasonal activities, while hourly patterns emphasize the impact of commuting behaviors and traffic flow. Together, these findings highlight opportunities for more targeted safety measures.

Efforts to improve traffic management and raise public awareness during high-risk times such as the busy afternoon commute or peak crash months like October could significantly reduce accident rates. Additionally, aligning road safety strategies with both seasonal and hourly trends can help create a safer driving environment for everyone throughout the year.

Types of Crash:

This figure 5 paints a clear picture of the most frequent types of crashes in Chicago, with "Parked Motor Vehicle" incidents leading the way at over 200,000 cases. These crashes highlight a widespread issue with stationary vehicles being struck, which could stem from poor parking practices, limited space, or driver inattention. Following closely are "Rear-End" crashes, often caused by behaviors like tailgating, distracted driving, or abrupt braking. These common crash types emphasize the importance of safe driving habits and better traffic flow management.

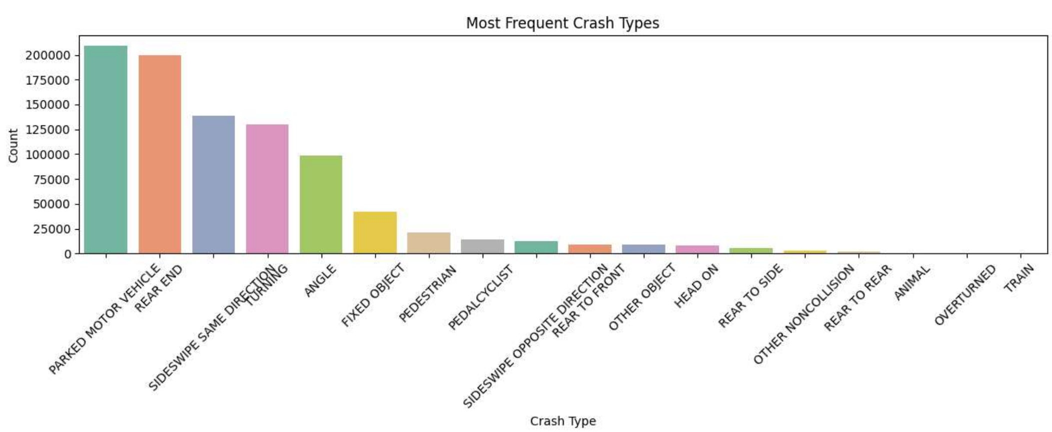

Figure 5: Most frequent Crash Types

Other notable crash types include "Sideswipe Same Direction," "Turning," and "Angle" collisions, which frequently occur during lane changes, at intersections, or when making improper turns. While less common, crashes involving "Fixed Objects," "Pedestrians," and "Pedal Cyclists" point to risks for vulnerable road users and challenges with stationary hazards. Rare incidents like "Train" or "Overturned" crashes are almost negligible but remind us of the need for preparedness in all scenarios. Together, these patterns highlight the need for targeted safety measures, such as stricter parking enforcement, improved infrastructure at intersections, and public awareness campaigns focused on reducing risky driving behaviors.

Traffic Control Devices:

Figure 6 highlights how traffic crashes in Chicago are distributed based on the presence of traffic control devices, revealing some striking patterns. The majority of crashes over 500,000 occur in areas without any traffic controls. This underscores the risks associated with uncontrolled intersections or road segments, where the lack of clear guidance can lead to confusion and a higher likelihood of collisions. It’s a reminder of how crucial well-placed traffic control measures are in maintaining order and safety on the roads.

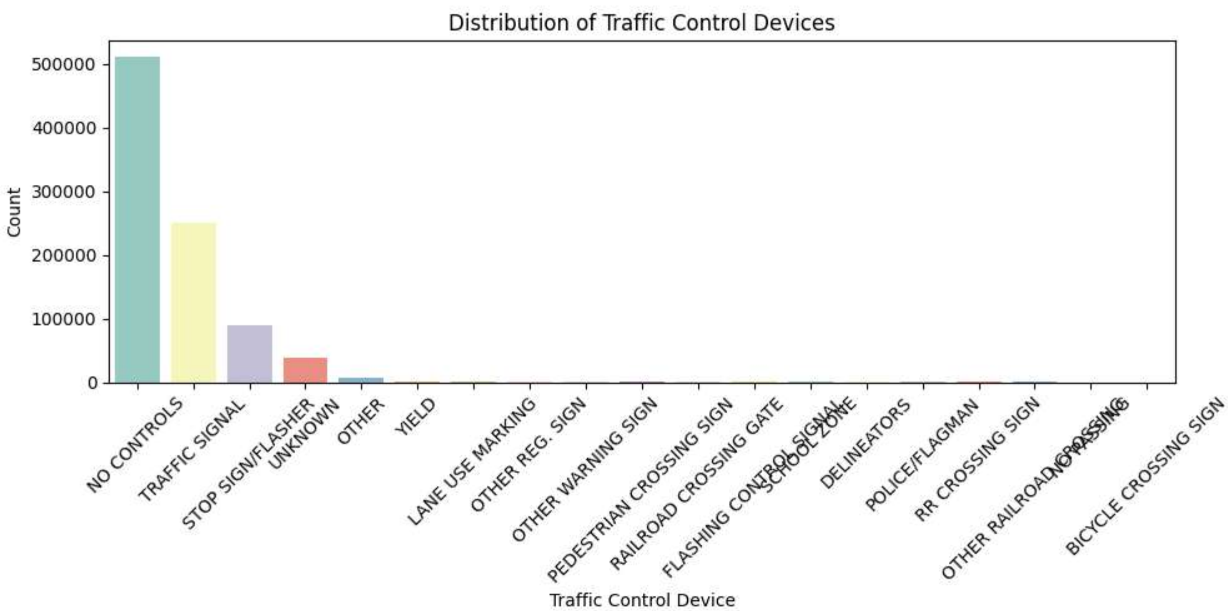

Figure 6: Distribution of Traffic Control Devices

Crashes at locations with traffic signals are also significant, pointing to issues like signal violations, distracted driving, or poorly timed lights. Stop signs and flashers account for a smaller but still notable number of incidents, likely resulting from drivers failing to yield or misjudging traffic gaps. Less common crash sites include areas with "Yield" signs, railroad crossing gates, and pedestrian crossings, suggesting these locations are generally safer, possibly due to lower traffic volumes or better compliance with the controls. These findings emphasize the need to strategically implement traffic control devices and address driver behavior, especially in uncontrolled areas, to create a safer traffic environment for all.

Crash severity breakdown by Traffic Control Device:

This chart examines the severity of crashes in Chicago based on the type of traffic control devices present at the crash site, shedding light on how these factors impact safety. Unsurprisingly, the largest share of crashes occurs in areas without any traffic controls. While most of these incidents are classified as "No Indication of Injury," a considerable number result in non-incapacitating injuries, with a smaller portion leading to severe outcomes like incapacitating injuries or fatalities. This underscores the urgent need to install traffic control devices in high-risk areas to reduce both the frequency and severity of crashes.

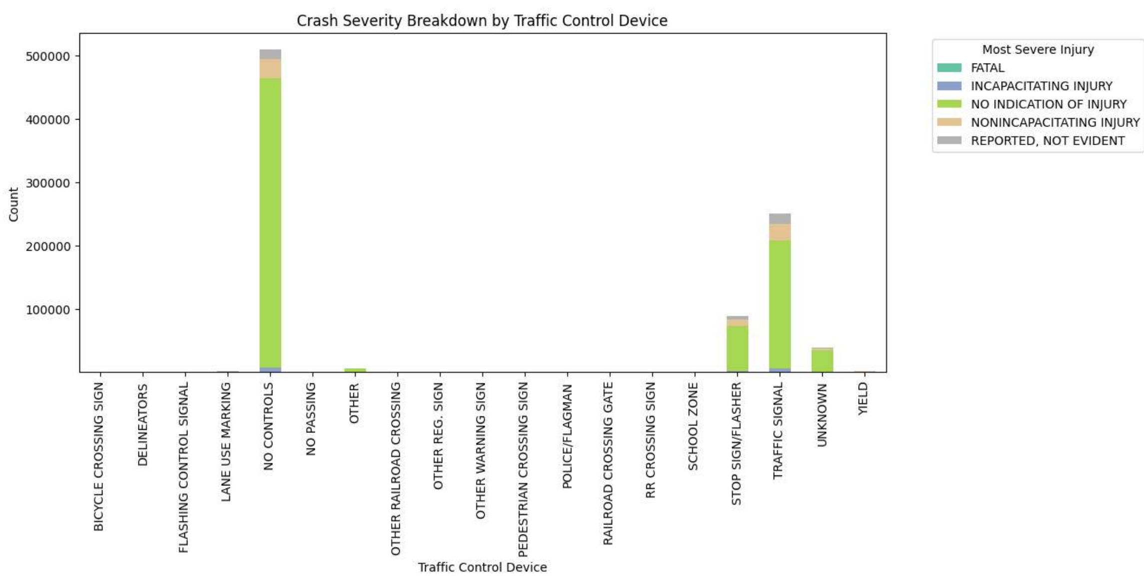

Figure 7: Crash Severity Breakdown by Traffic Control Device

Interestingly, crashes at locations with traffic signals, stop signs, or flashing lights show a similar severity pattern, with most resulting in no injuries or minor injuries. However, areas with less common controls, like pedestrian crossing signs or railroad crossing gates, experience fewer crashes overall but tend to include a higher proportion of severe injuries or fatalities. This highlights the importance of ensuring these intersections are clearly marked, well-maintained, and strictly enforced to protect vulnerable road users. Ultimately, the data emphasizes that while traffic control devices play a crucial role in reducing crash risks, their strategic placement, regular maintenance, and proper enforcement are essential to minimizing the impact of accidents.

Conclusion:

In conclusion, the analysis of Chicago's 2024 traffic crash data provides critical insights that can inform targeted interventions to enhance urban road safety. Contrary to conventional expectations, clear weather, often deemed optimal for driving, was associated with the highest frequency of crashes, underscoring the complex relationship between driver behavior, systemic factors like speeding, and environmental conditions. Temporal patterns further emphasize specific windows of heightened risk, particularly in October and during late afternoon rush hours (3 PM to 5 PM), which highlight key opportunities for intervention.

The use of advanced analytical tools, including Python for data processing and visualization and QGIS for spatial analysis, proved instrumental in identifying high-risk areas and specific crash types. The predominance of rear-end collisions and incidents involving parked vehicles suggests an urgent need for public education on safe following distances and targeted urban planning initiatives to mitigate these risks. Moreover, the analysis revealed that intersections without traffic controls are disproportionately dangerous, even as areas with existing traffic signals and stop signs still necessitate improved compliance and more intelligent design strategies.

To effectively address these challenges, a multifaceted approach is required, incorporating not only enforcement and driver education but also strategic infrastructure improvements that prioritize the safety of vulnerable road users and mitigate risks in high-traffic zones. These findings highlight the potential for evidence-driven policies to foster a safer urban transportation environment, contributing to the broader goal of reducing traffic-related injuries and fatalities.

Python Code:

```python
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

```python
dataFrame = pd.read_csv("C:\\Users\\kiran\\Desktop\\data\\crashdata.csv")
print("Our DataFrame....\n",dataFrame)
df = pd.read_csv("C:\\Users\\kiran\\Desktop\\data\\crashdata.csv")
## Display check the first five rows of the data to understand its structure
print(df.head(5))
# Check for missing values
print(df.isnull().sum())
# To get overall information
print(df.info())
# Check for missing values in each column
print(df.isnull().sum())
# Drop rows where 'CRASH_DATE' is missing (if this is critical to the analysis)
df.dropna(subset=['CRASH_DATE'], inplace=True)
# You can fill missing values in other columns, for example, filling with 'Unknown'
df['WEATHER_CONDITION'].fillna('Unknown', inplace=True)

# Convert 'CRASH_DATE' to datetime format for easier analysis
df['CRASH_DATE'] = pd.to_datetime(df['CRASH_DATE'], errors='coerce')



# Count crashes by weather condition
weather_counts = df['WEATHER_CONDITION'].value_counts()
# Display the result of weather counts
print(weather_counts)
plt.figure(figsize=(12,6))
sns.countplot(data=df, x='WEATHER_CONDITION',
order=df['WEATHER_CONDITION'].value_counts().index)
plt.xticks(rotation=45)
plt.title('Number of Crashes by Weather Condition')
plt.xlabel('Weather Condition')
plt.ylabel('Number of Crashes')
```

```
plt.show()

# Sum the total injuries by crash type
injuries_by_crash_type = df.groupby('CRASH_TYPE')['INJURIES_TOTAL'].sum()
# Display the result
print(injuries_by_crash_type)
plt.figure(figsize=(12,6))
sns.barplot(x=injuries_by_crash_type.index, y=injuries_by_crash_type.values)
plt.xticks(rotation=45)
plt.title('Total Injuries by Crash Type')
plt.ylabel('Total Injuries')
plt.show()

# Extract the month from the 'CRASH_DATE'
df['CRASH_DATE'] = pd.to_datetime(df['CRASH_DATE'], format='%m/%d/%Y %H:%M')
df['CRASH_MONTH'] = df['CRASH_DATE'].dt.month
crashes_by_month = df['CRASH_MONTH'].value_counts().sort_index()
print(crashes_by_month)
plt.figure(figsize=(12,6))
sns.barplot(x=crashes_by_month.index, y=crashes_by_month.values)
plt.title('Crashes by Month')
plt.ylabel('Number of Crashes')
plt.show()

import calendar
#These data are extracted from the avove crash_month
data = {
    'CRASH_MONTH': [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12],
    'Counts': [66068, 65284, 67812, 66417, 77268, 77697, 78570, 80823, 82228, 86681, 78184,
77000]
}
```

```python
df = pd.DataFrame(data)
df['Month_Name'] = df['CRASH_MONTH'].apply(lambda x: calendar.month_name[x])
print(df[['Month_Name', 'Counts']])
month_names = [calendar.month_name[i] for i in crashes_by_month.index]
plt.figure(figsize=(12,6))
sns.barplot(x=month_names, y=crashes_by_month.values)
plt.title('Crashes by Month')
plt.xlabel('Month')
plt.ylabel('Number of Crashes')
plt.show()


# To determine the 'CRASH_Hour' with datetime format
dataFrame = pd.read_csv("C:\\Users\\kiran\\Desktop\\data\\crashdata.csv")
print("Our DataFrame....\n",dataFrame)
df = pd.read_csv("C:\\Users\\kiran\\Desktop\\data\\crashdata.csv")
# Extract the Hour from the 'CRASH_DATE'
df['Hour'] = pd.to_datetime(df['Hour'], format='%H:%M')
df['Hour'] = df['Hour'].dt.hour
crashes_by_hour = df['Hour'].value_counts().sort_index()
print(crashes_by_hour)
plt.figure(figsize=(24,6))
sns.barplot(x=crashes_by_hour.index, y=crashes_by_hour.values)
plt.title('Crashes by Hour')
plt.ylabel('Number of Crashes')
plt.show()


# To determine the crashes Type
dataFrame = pd.read_csv("C:\\Users\\kiran\\Desktop\\data\\crashdata.csv")
```

```
print("Missing values in 'FIRST_CRASH_TYPE':",
dataFrame['FIRST_CRASH_TYPE'].isnull().sum())
dataFrame['FIRST_CRASH_TYPE'] = dataFrame['FIRST_CRASH_TYPE'].fillna('Unknown')
crash_types = dataFrame['FIRST_CRASH_TYPE'].value_counts()
print("\nMost Frequent Crash Types:\n", crash_types)
plt.figure(figsize=(20, 6))
sns.barplot(x=crash_types.index, y=crash_types.values, palette='Set2')
plt.title('Most Frequent Crash Types')
plt.xlabel('Crash Type')
plt.ylabel('Count')
plt.xticks(rotation=45)
plt.show()

# Traffic Control Signals Crash
dataFrame = pd.read_csv("C:\\Users\\kiran\\Desktop\\data\\crashdata.csv")
print("Missing values in 'TRAFFIC_CONTROL_DEVICE':",
dataFrame['TRAFFIC_CONTROL_DEVICE'].isnull().sum())
dataFrame['TRAFFIC_CONTROL_DEVICE'] =
dataFrame['TRAFFIC_CONTROL_DEVICE'].fillna('Unknown')
device_counts = dataFrame['TRAFFIC_CONTROL_DEVICE'].value_counts()
print("\nTraffic Control Device Counts:\n", device_counts)
plt.figure(figsize=(10, 6))
sns.barplot(x=device_counts.index, y=device_counts.values, palette='Set3')
plt.title('Distribution of Traffic Control Devices')
plt.xlabel('Traffic Control Device')
plt.ylabel('Count')
plt.xticks(rotation=45)
plt.show()

#Severity Breakdown by Traffic Control Devices
```

```
severity_by_device = dataFrame.groupby('TRAFFIC_CONTROL_DEVICE')['MOST_SEVERE_INJURY'].value_counts().unstack()
print("\nCrash Severity Breakdown by Traffic Control Device:\n", severity_by_device)
plt.figure(figsize=(12, 8))
severity_by_device.plot(kind='bar', stacked=True, colormap='Set2')
plt.title('Crash Severity Breakdown by Traffic Control Device')
plt.xlabel('Traffic Control Device')
plt.ylabel('Count')
plt.xticks(rotation=45)
plt.legend(title='Most Severe Injury', bbox_to_anchor=(1.05, 1), loc='upper left')
plt.tight_layout()
plt.show()
```

References:

Portal, C. D. (2024). *Traffic Crashes* . Chicago : Chicago Police Department.

Wenzhao Zhang, S. Z. (2020). Chicago Traffic Collision Data Analysis Based on Multi-Component Analysis and Exploratory Data Analysis. *CICTP 2020* , 4684 - 4696.

Xu, C., Gao, J., Zuo, F., & Ozbay, K. (2024). Estimating Urban Traffic Safety and Analyzing Spatial Patterns through the Integration of City-Wide Near-Miss Data: A New York City Case Study. *MDPI* , 14(14), 6378.