Project Report:

Demographic, Substance, and Environmental Factors in Predicting Accident Severity

By: Chloe Nicola, Audric Charles, Divy Shah, David Zhao

Abstract

This study investigates the interplay between environmental, demographic, and substance-related factors in predicting traffic accident severity in the City of Tempe. Traffic accidents are a major cause of fatalities in the United States, making it imperative to identify and understand the key predictors of severe accidents to inform prevention strategies. Using a dataset of over 35,000 crash records, we analyzed variables including weather conditions, lighting, surface quality, junction presence, and driver demographics (age, gender) to assess their impact on crash outcomes. Statistical techniques such as Welch's ANOVA and hypothesis testing were employed to evaluate differences in injury severity across environmental and substance-related factors.

Our findings reveal that adverse weather conditions (e.g., fog, snow), impaired driving—particularly involving drugs—and older drivers (75+ years) are significant predictors of higher accident severity. Conversely, gender had minimal influence on crash outcomes. Additionally, certain high-traffic streets, such as Baseline Road and Rural Road, emerged as consistent hotspots for severe crashes due to factors like complex intersections and driver behaviors. These results emphasize the importance of targeted interventions, including infrastructure improvements, stricter driving regulations for high-risk demographics, and campaigns against drug-impaired driving, to mitigate accident severity and enhance traffic safety.

Contents

Motivation

Traffic accidents are a significant public safety concern, with severe economic, social, and personal implications. Each year, millions of crashes worldwide result in injuries, fatalities, and significant property damage, placing a substantial burden on healthcare systems, insurance companies, and government infrastructure. As such understanding the factors contributing to accident severity is crucial for developing effective preventative measures and improving urban planning and traffic management.

Environmental conditions such as weather, lighting, and road surface quality are among the most prominent determinants of accident outcomes. Adverse weather conditions, such as rain, snow, or fog, can also significantly affect vehicle handling, and increase stopping distances, thereby elevating the risk of severe collisions. Poorly maintained or slippery road surfaces can further increase these risks, especially in high-traffic areas or regions with frequent inclement weather. Similarly, insufficient lighting can impair a driver's ability to detect hazards, increasing the likelihood of nighttime accidents. Lastly, demographic factors such as age and gender can reveal a pattern in driving behavior and risk perception.

The motivation for this study arises from the need to identify and analyze these factors to improve traffic safety outcomes. By leveraging a comprehensive dataset and data analysis techniques, we aim to predict accident severity more accurately, helping municipalities and policymakers design targeted interventions. Additionally, understanding the interaction between environmental and demographic variables could lead to more inclusive and effective traffic safety measures, reducing the likelihood of severe accidents and saving lives.

Dataset

To explore the factors influencing traffic accident severity, we utilized a detailed crash dataset from the City of Tempe. The dataset contains over 35,000 records, capturing a wide range of environmental, demographic, situational, and substance-related attributes. Key features include:

- **Crash details**: Type of crash (e.g., sideswipe, angle), contributing factors (e.g., speeding, distractions), and severity of injuries.
- **Environmental factors**: Weather, lighting, and road surface conditions, which help determine how external conditions might impact accident severity.
- **Location information**: Coordinates, street names, distance from junctions, and junction types (e.g., intersections, driveways), allowing for an analysis of how the location of accidents may affect their outcomes.
- **Driver demographics**: Age, gender, travel direction, and any reported alcohol or drug influence, which provide insights into the role of driver characteristics in accident severity.
- **Substance influence**: Information on whether the driver was under the influence of alcohol or drugs. This factor is crucial as impaired driving is known to significantly increase the likelihood of accidents and the severity of injuries. The dataset includes variables for both **alcohol use** and **drug use** (if reported), which allow us to analyze the impact of substance impairment on crash outcomes.

This comprehensive dataset enables an in-depth analysis of the relationships between these factors and crash severity. By leveraging statistical and data analysis techniques, we aim to uncover patterns and insights that could help improve traffic safety policies, reduce accident severity, and provide a more accurate understanding of how both demographic and substance-related factors contribute to crashes.

Analyses

3.1 Demographic Factors

We aim to explore the relationship between demographic factors (such as driver age, gender, and potential substance use) and the severity of crashes, measured in terms of fatalities and the number of injuries. To achieve this, we analyze the data using graphical techniques and summarize the findings in a series of visualizations, highlighting key patterns and trends.

The distribution of injuries by gender is visualized in **Figure 3.1.1**, which includes histograms and a boxplot. The histograms show that the majority of crashes result in zero injuries for both male and female drivers, indicating that most accidents are minor in severity. However, male drivers exhibit a slightly higher count of injuries compared to female drivers, which could reflect differences in driving behavior, exposure to risk, or situational factors. The boxplot illustrates similar distributions for both genders, with most injuries clustering around zero. However, the presence of outliers for both genders indicates that some crashes result in disproportionately high numbers of injuries, potentially involving high-speed impacts or multi-vehicle collisions. These observations suggest that while gender may not be a primary predictor of crash severity, further statistical testing could uncover significant differences.

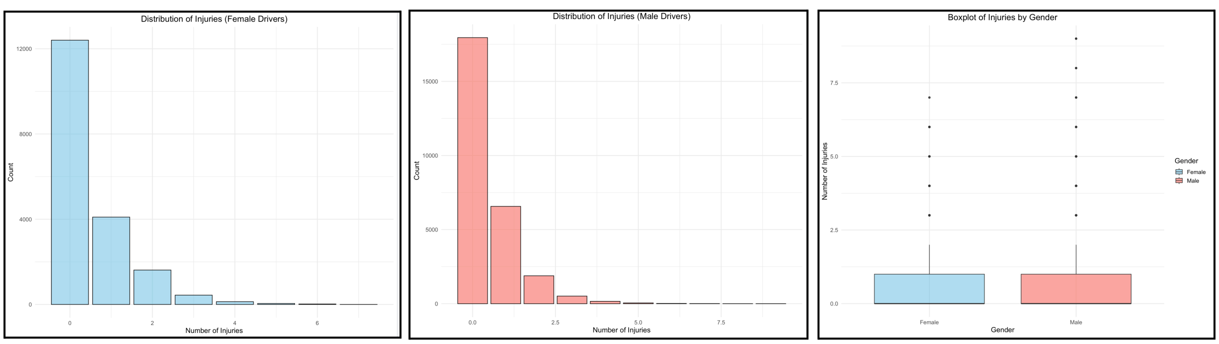

Figure 3.1.1: **Histograms and boxplot for number of accident injuries by gender**

The relationship between driver age and the average number of fatalities is illustrated in **Figure 3.1.2** using a scatterplot. The data reveals that fatalities remain consistently low across most age groups, with minor fluctuations. However, a noticeable increase in average fatalities is observed among older drivers, particularly those aged 75 and above. This trend may be attributed to factors such as slower reaction times, increased physical frailty, and reduced resilience during severe crashes. The smoothed trend line in Figure 3.1.2 highlights a peak in average fatalities among the elderly age group, followed by a slight decline at very high ages, which may reflect fewer older drivers on the road. Conversely, younger drivers (under 25) exhibit consistently low fatality averages, possibly due to modern vehicle safety technologies or differences in crash dynamics. These findings suggest that targeted interventions, such as stricter testing for older drivers and advanced vehicle technologies (e.g., collision-avoidance systems), could help reduce fatalities in high-risk age groups.

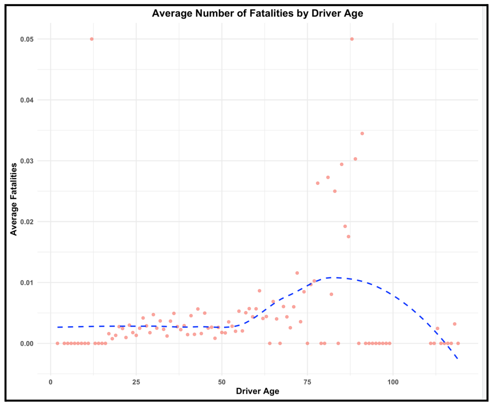

Figure 3.1.2: **Scatterplot of average fatalities by driver age**

From these analyses, we observe that age has a clearer relationship with crash severity than gender. While the differences in injuries by gender are subtle, the increasing average fatalities among older drivers highlight a critical risk group. These insights underscore the need to consider demographic factors in road safety interventions.

3.2 Substance Influence

In this section, we examine the relationship between influence factors, including drug use, alcohol use, and no apparent influence, and the frequency of accidents. Understanding how these influence categories affect crash rates is essential for identifying patterns and implementing effective interventions to reduce risk. **Figure 3.2.1** presents a bar chart that compares the total number of accidents across the three influence categories.

The analysis reveals that the majority of accidents are linked to drivers with no apparent influence, as reflected in the significantly tall bar for this category. This aligns with expectations, as most drivers are not impaired by substances. However, drug use-related accidents still represent a notable proportion, followed by alcohol use-related accidents. While alcohol-related crashes account for the smallest share, their significance cannot be overlooked, given the well-established dangers of impaired driving.

This visualization emphasizes that both drug and alcohol use are contributing factors to accidents, although less frequently than crashes involving no apparent influence. Interestingly, the relatively higher frequency of accidents involving drug use compared to alcohol use suggests that drugs may play a more significant role in crash outcomes than previously thought. Future research could delve into the severity of crashes within these influence categories and explore whether targeted interventions for impaired driving can help mitigate these numbers.

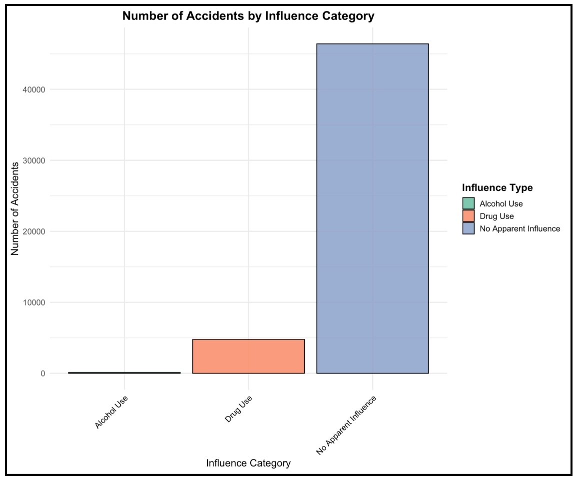

Figure 3.2.1: **Bar Chart of Accidents by Influence Category**

Further exploration into the severity of accidents reveals a strong correlation between substance use (specifically drug use and alcohol use) and higher numbers of fatalities and injuries. **Figure 3.2.2** highlights that accidents involving drug use or alcohol use tend to have significantly more fatalities and injuries compared to accidents where there is no apparent use.

This finding underscores the heightened risks associated with impaired driving. Drug use and alcohol use not only increase the likelihood of accidents but also lead to more severe outcomes. The data suggests that while accidents involving no apparent influence are the most common, those involving substance use often result in more significant harm. This highlights the need for focused efforts to reduce impaired driving, as even a small proportion of such accidents could lead to disproportionately severe outcomes.

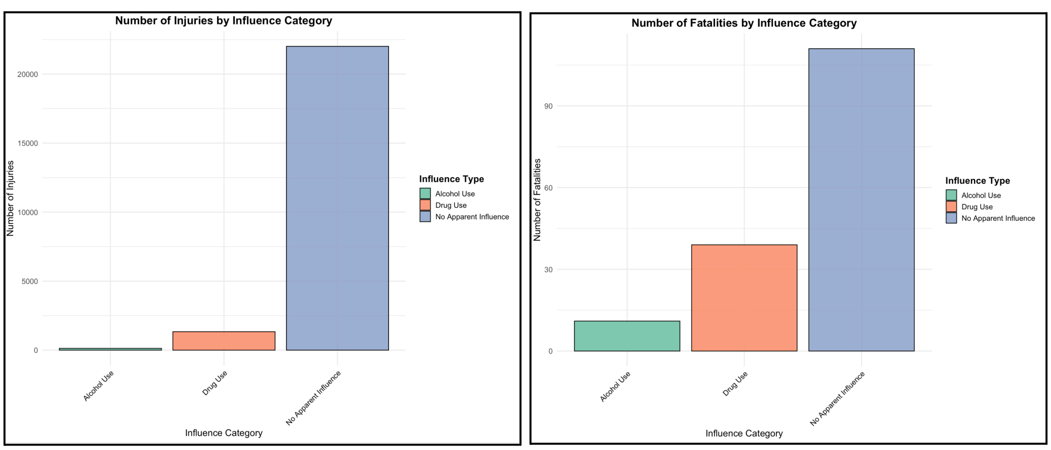

Figure 3.2.2: **Fatalities and Injuries by Influence Category**

From these analyses, we observe that drug use has a more significant impact on crash severity than alcohol use. While alcohol-related crashes are less frequent, those involving drugs result in more severe outcomes, including higher numbers of fatalities and injuries. These findings highlight the need for targeted interventions to address drug-impaired driving and reduce its contribution to traffic-related harm.

3.3 Environmental Factors

We now examine the relationship between environmental conditions (natural light, weather, and road surface conditions) and the severity of the crash (i.e. fatalities and/or number injured). We can first visualize this data using graphs which compare the different environmental conditions against the severity of the crash.

Figure 3.3.1 shows the boxplot comparing injuries by weather condition. The data reveals that the majority of accidents occur under clear weather conditions, which is expected as clear weather is the most common. However, the severity of injuries tends to be slightly higher in conditions such as fog and snow. For example, snow conditions exhibit a higher average number of injuries, along with a broader spread of data (indicating more variation in injury severity). The outliers for conditions such as "Heavy Smoke" and "Rain" also suggest that some accidents in these weather conditions may result in particularly severe injuries or fatalities.

From this, it can be inferred that, while clear weather is the most common, adverse weather conditions such as snow and fog contribute to a higher severity of injuries. This could suggest that driving behavior or road conditions deteriorate under these adverse weather conditions, increasing the risk of more severe accidents.

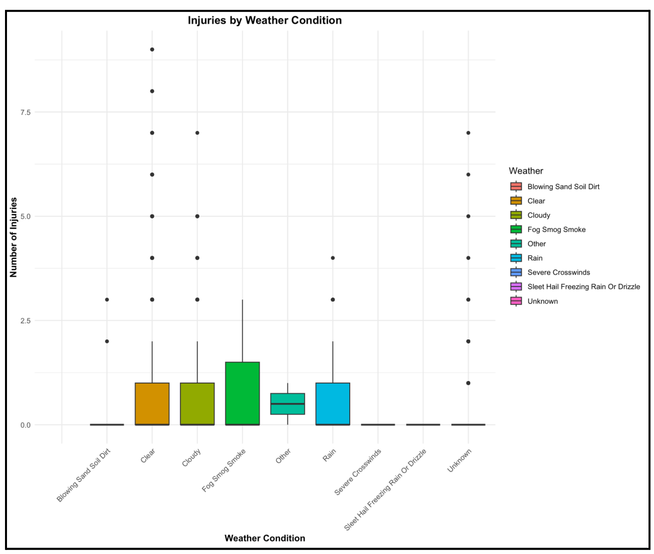

Figure 3.3.1: **Fatalities by Weather Condition**

Figure 3.3.2 shows the bar charts comparing injuries by road surface condition and light condition. The analysis of road surface conditions indicates that "Wet" and "Icy" conditions lead to the highest average number of injuries. This aligns with common understanding, as wet and icy roads significantly reduce tire traction, making it more difficult for drivers to control their vehicles, particularly at higher speeds. On the other hand, dry road conditions have the lowest injury levels, indicating that proper maintenance and dry conditions reduce the severity of accidents.

In terms of light conditions, the bar chart shows that "Daylight" conditions have the lowest average number of injuries, while "Dark Lighted" and "Dark No Light" conditions show slightly higher average injuries. This suggests that reduced visibility during nighttime or in poorly lit areas significantly impacts driving safety, increasing the chances of accidents, which often lead to more severe injuries.

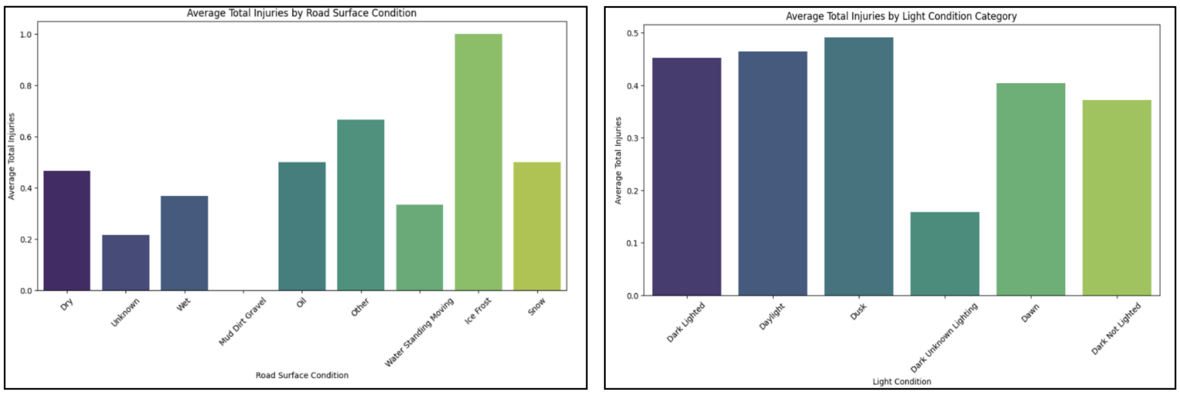

Figure 3.3.2: **Average number of injuries by road surface condition and light condition**

Together, these findings highlight that both road surface and light conditions play a critical role in accident severity. Wet, icy, and dark road conditions contribute to higher injury levels, while dry roads and daylight conditions reduce the likelihood and severity of accidents. This underscores the importance of road maintenance, proper lighting, and visibility improvements to reduce injury severity in crashes.

We answer the question of if weather is a fair predictor of severity of the crash. We will start by numerically summarizing the data by computing the sample means and variances of total injuries for the different weather categories.

Table : Sample Mean and Variance of Total Injuries by Weather Category

Weather	mean	var
Blowing Sand Soil Dirt	0.500000	1.166667
Clear	0.462060	0.656531
Cloudy	0.475415	0.691136
Fog Smog Smoke	0.727273	1.218182
Other	0.500000	0.500000
Rain	0.371915	0.446740
Severe Crosswinds	0.000000	0.000000
Sleet Hail Freezing Rain Or Drizzle	0.000000	0.000000
Unknown	0.296296	0.469235

Figure 3.3.3: **Sample mean and variance of total injuries by weather category**

With the sample means and variances calculated, we can use an ANOVA (Analysis of Variance) test in order to test if the mean of the Total Injuries differs across the different weather conditions. We set our testing hypotheses like so:

H_0: $\mu_1 = \mu_2 = \mu_3 ... = \mu_i$ for all weather conditions *i*.

H_1: $\exists \mu_i \neq \mu_k$ for some weather condition *k*.

Before running the ANOVA test, we must note that the sample populations are not spread evenly, as certain categories of weather are more common than others. Because of this, we must run a modified ANOVA test: Welch's ANOVA.

After running the ANOVA test we get the result of:

F-Statistic: $F(8,\ 21.97)\ =\ 1704.500$, P-Value: $p\ =\ 7.26\ \cdot\ 10^{-29}$

Since our P-Value is very small, we can safely reject the null hypothesis. This means that a certain weather condition is resulting in either a higher or lower mean compared to others.

3.4 Common Street Accidents

Additionally, we shall also examine the connection between crash locations and the frequency of collisions, focusing on identifying streets with the highest crash counts and the types of collisions most commonly occurring there. By analyzing crash data, including information about street names, cross-streets, and collision types (e.g., rear-end, sideswipe), we provide a comprehensive understanding of where and how crashes frequently occur. This analysis utilizes both numerical aggregation and visual techniques to highlight patterns and trends.

The distribution of crash counts by street is visualized in **Figure 3.4.1**, which includes a bar graph on the most common street crash occurrences. The bar graph reveals that streets like **Baseline Rd** and **Rural Rd** experience the highest number of crashes, with **Rear-End** collisions emerging as the most frequent type across these streets. These findings suggest that high traffic volumes, complex intersections, and driver behaviors (e.g., tailgating and distractions) are also significant contributors.

These observations provide valuable insights into traffic safety, emphasizing the importance of location-specific measures to mitigate risks and reduce collision frequencies on these high-crash streets. Further exploration of environmental factors, such as lighting and road surface conditions, could deepen our understanding of why certain streets are more prone to accidents.

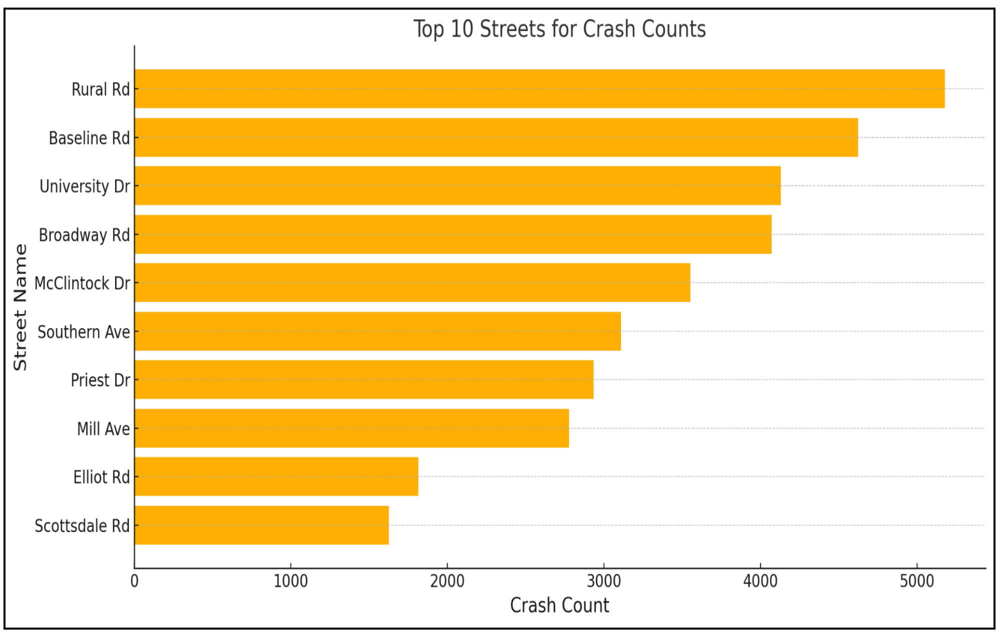

Figure 3.4.1: **Common Streets for Crash Counts**

Conclusion

4.1 Discussion

Our analyses highlight several key predictors of traffic accident severity and provide actionable insights for traffic safety measures:

1. **Substance Influence:**
   Accidents involving drug use were found to contribute significantly to fatalities and injuries compared to alcohol-related crashes, despite the latter's well-documented impact on road safety. This underscores the growing importance of addressing drug-impaired driving, which appears underrepresented in traditional safety campaigns and interventions.
2. **Environmental Factors:**
   Adverse weather conditions, such as snow and fog, significantly increase the severity of injuries, likely due to impaired visibility and reduced vehicle control. Poor surface conditions, including wet or icy roads, further exacerbate crash risks. Conversely, clear weather and dry roads were associated with lower injury severity, emphasizing the role of proper infrastructure maintenance and driver education about safe driving in inclement weather.
3. **Demographic Factors:**
   Older drivers (75+ years) exhibited a notable increase in fatalities, suggesting the need for targeted interventions, such as more frequent driver assessments, advanced safety technologies like collision-avoidance systems, or stricter licensing criteria for high-risk age groups. Gender differences were minimal and not strongly correlated with crash severity, though subtle behavioral patterns may warrant further exploration.
4. **Geographic Factors:**
   Streets such as Baseline Road, Rural Road, and University Drive were identified as high-risk locations for severe accidents, characterized by high traffic volumes, complex intersections, and variable speed limits. Tailored interventions, such as improved lighting, signage, and traffic flow designs, could help mitigate crash risks in these areas.

Together, these insights reinforce the importance of integrating demographic, environmental, and geographic considerations into traffic safety policies.

4.2 Summary of Findings

- **Substance Influence:** Drug-related crashes lead to higher fatalities and injuries than alcohol-related crashes, indicating a need for focused interventions targeting drug-impaired driving.

- **Environmental Factors:** Adverse weather (e.g., snow, fog) and poor road surfaces (e.g., wet, icy) significantly elevate crash severity, while clear weather and dry conditions correlate with fewer severe outcomes.
- **Demographic Factors:** Older drivers (75+) are disproportionately involved in severe accidents, highlighting the need for age-specific safety measures, whereas gender exhibited minimal influence on accident outcomes.
- **Geographic Factors:** High-traffic streets like Baseline Road and Rural Road are frequent sites for severe crashes due to traffic patterns and intersection complexity, warranting location-specific safety enhancements.

4.3 Future Work

To build on the findings of this study, future research should:

1. **Granular Analysis of Substance Use:** Investigate specific drug types involved in accidents to better understand their impact on crash severity.
2. **Detailed Demographic Modeling:** Explore underlying factors such as reaction time and physical frailty in older drivers to design targeted interventions.
3. **Machine Learning Approaches:** Use advanced modeling techniques (e.g., random forests, neural networks) to better account for complex, nonlinear relationships between factors.
4. **Policy and Interventions:** Develop and test campaigns addressing drug-impaired driving, tailored traffic management systems for high-risk streets, and enhanced infrastructure to mitigate risks during adverse weather conditions.