CHICAGO CITY CITY ANALYSIS

IAN TULIENGE

OVERVIEW

The goal of this project is to analyze car accidents in Chicago City and build a classifier that can predict the primary contributory cause of these accidents. The project aims to identify specific locations or road segments with a higher frequency of accidents, understand the contributing factors associated with severe accidents, and uncover any seasonal or temporal patterns in accidents. The analysis will be conducted using three datasets obtained from the Chicago Data Portal: Traffic_Crashes_-*People, Traffic_Crashes-Vehicles, and Traffic_Crashes*-_Crashes. These datasets provide comprehensive information about the accidents, including details about the people involved, vehicles, crashes, and various factors related to the accidents.

The project will follow a structured approach to derive insights and provide recommendations for accident prevention and improving road safety

Outline

- Data Understanding
- Business understanding
- Defining the Question
- Feature selection
- Conclusion
- Recommendation
- Next step

Business understanding

The City of Chicago Vehicle Safety Board (CCVSB) is interested in reducing traffic accidents and improving road safety in Chicago City. To achieve this, they want to analyze car accidents and identify patterns and contributing factors associated with these accidents. The ultimate goal is to develop strategies and initiatives to prevent accidents and enhance safety measures in the city.

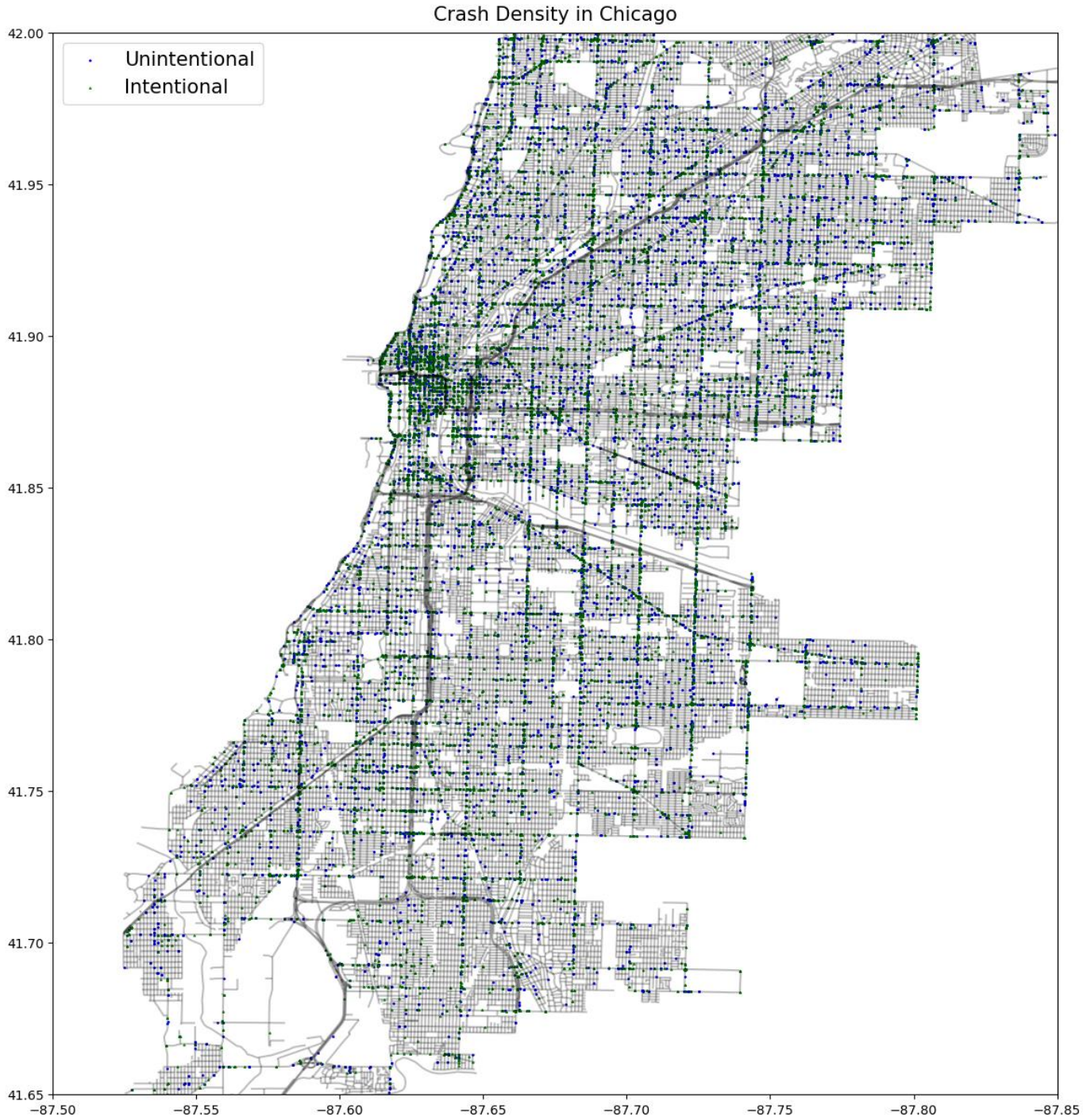
Defining the Question

Question One

Are there any specific locations or road segments in Chicago City that have a higher frequency of car accidents?

By identifying high-frequency accident locations, the CCVSB can focus their efforts on implementing targeted safety measures in those areas.

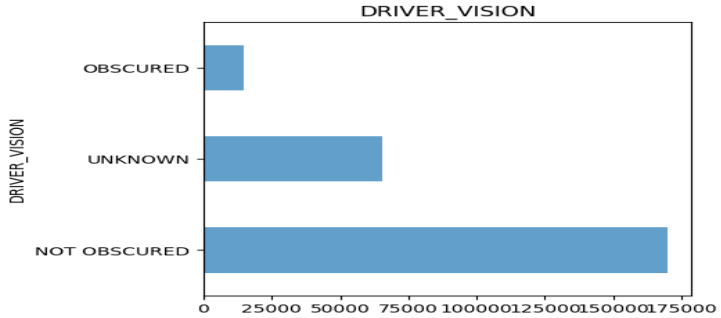
The map highlights a significant concentration of accidents in the downtown area of Chicago, indicating a higher density of incidents in that region. The predominant color in this area is green, indicating that a majority of the accidents are attributed to intentional actions or driver errors. However, it's important to note that there are also scattered blue plots throughout the map, suggesting a considerable number of accidents that occur unintentionally or present opportunities for improvement in terms of safety measures.

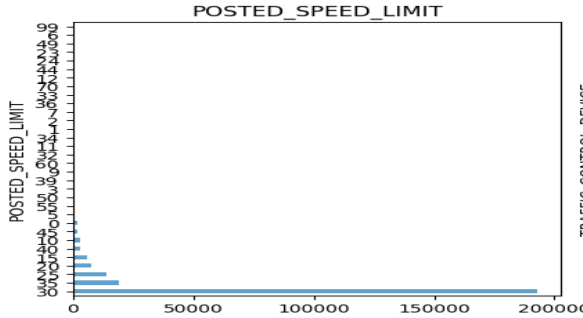


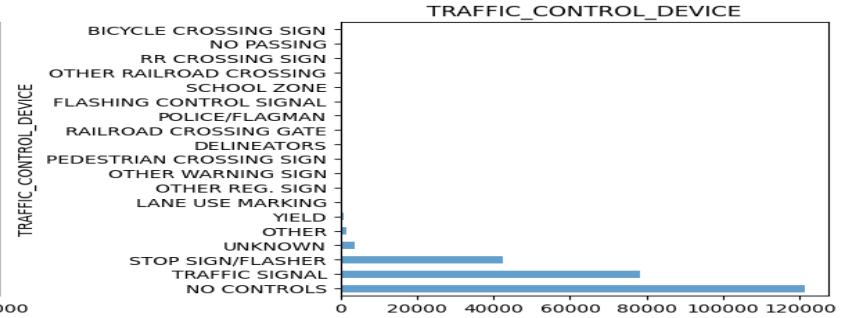
Question Two

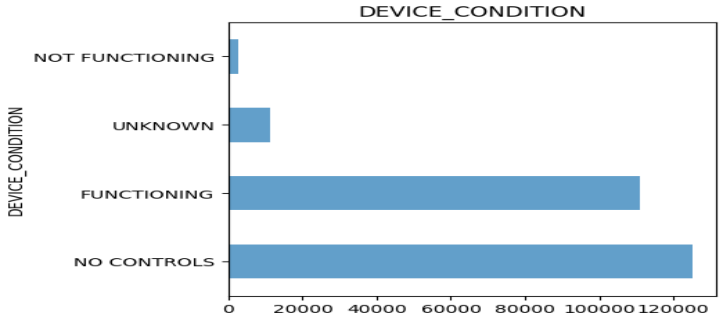
What are the contributing factors or characteristics associated with severe car accidents in Chicago City?

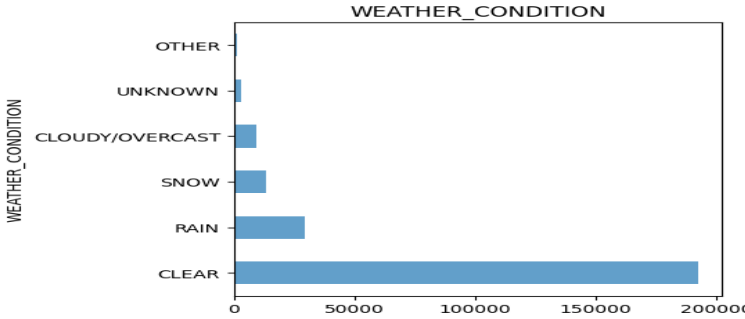
Our focus will be on accidents that were categorized as 'Unintentional' in order to delve deeper into the underlying causes and identify potential areas for improvement. By narrowing our analysis to these specific incidents, we can gain valuable insights into the root causes of unintentional accidents and uncover opportunities for enhancing safety measures and preventing similar occurrences in the future.

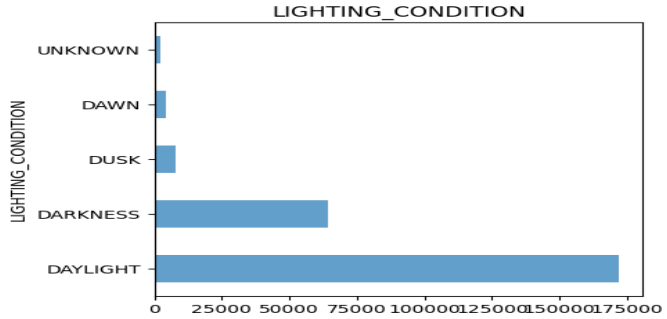


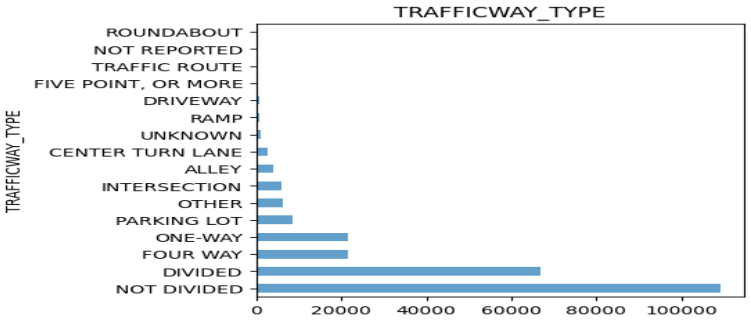


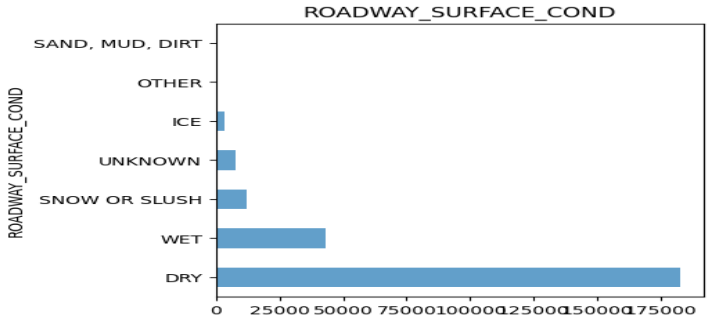


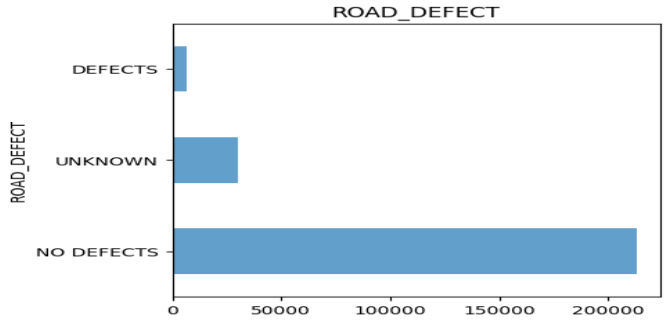












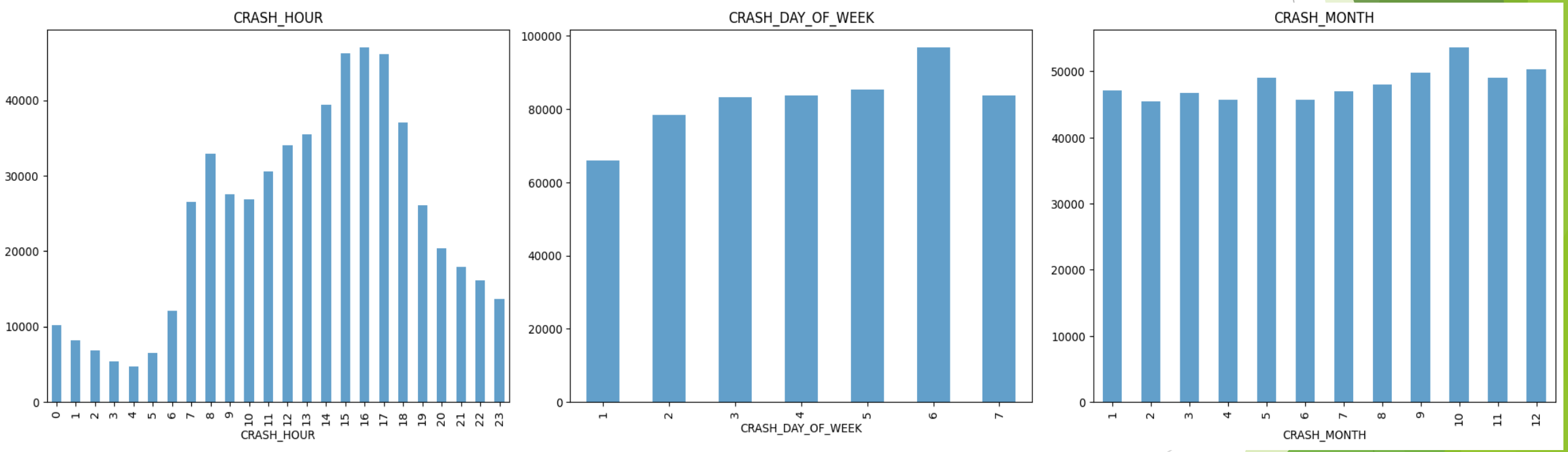
From the above plot we can draw the following conclusions:

- Upon analyzing the contributing factors associated with control failures in unintentional accidents, it is evident that a majority of the accidents occurred when the driver's vision was not obscured. Furthermore, it is notable that these accidents occurred while the drivers were adhering to the posted speed limit, typically set at 30 mph. These findings suggest that factors other than vision or speed might be contributing to control failures in these accidents.
- An important finding from the analysis is that the absence of traffic control devices has been the primary contributing factor to the number of accidents in Chicago. This suggests that increasing the presence of traffic control devices throughout the city could potentially reduce the occurrence of unintentional accidents. This finding is further supported by the Device Condition plot, which indicates a higher count of accidents when there are no traffic control devices in place. Implementing and improving traffic control measures can therefore be an effective strategy to mitigate control failures and enhance road safety in Chicago.
- The analysis indicates that weather condition and lighting condition have relatively minimal impact on the occurrence of accidents. These factors do not show a strong correlation with the number of accidents in Chicago.
- Significant number of accidents occur on roads categorized as "Not Divided" in terms of trafficway type. This suggests that implementing road division measures, such as adding medians or physical barriers, can potentially mitigate the occurrence of accidents. Dividing the roads can enhance traffic management, separate opposing flows of traffic, and reduce the likelihood of collisions, thereby contributing to improved road safety.
- The analysis indicates that the roadway surface condition and road defects have a relatively minimal impact on the occurrence of these accidents. It suggests that the condition of the road surface, such as potholes or uneven pavement, and the presence of road defects, such as cracks or debris, may not be significant contributors to the unintentional accidents in Chicago.

Question Three

Are there any seasonal or temporal patterns in car accidents in Chicago City?

Analyzing seasonal and temporal patterns can provide insights into when accidents are more likely to occur. This information can be used to implement specific measures during high-risk periods.



From the above plots we can come to conclusion:

- The analysis of the crash time data reveals that a significant number of accidents in Chicago occur between the hours of 14 to 18, which coincides with the peak rush hour traffic. This suggests that the high volume of vehicles during these hours contributes to the increased accident rate. Considering the concentration of accidents in the downtown area during this time frame, it becomes apparent that better traffic management strategies are needed.

To address this issue, it is recommended that the city implements additional measures to facilitate traffic flow and reduce congestion in the downtown area during these peak hours. This can include deploying more traffic management personnel or implementing intelligent transportation systems to optimize traffic signal timings and improve the coordination of traffic movements. By enhancing traffic management during rush hours, the city can mitigate the number of accidents and improve overall road safety in the downtown area.

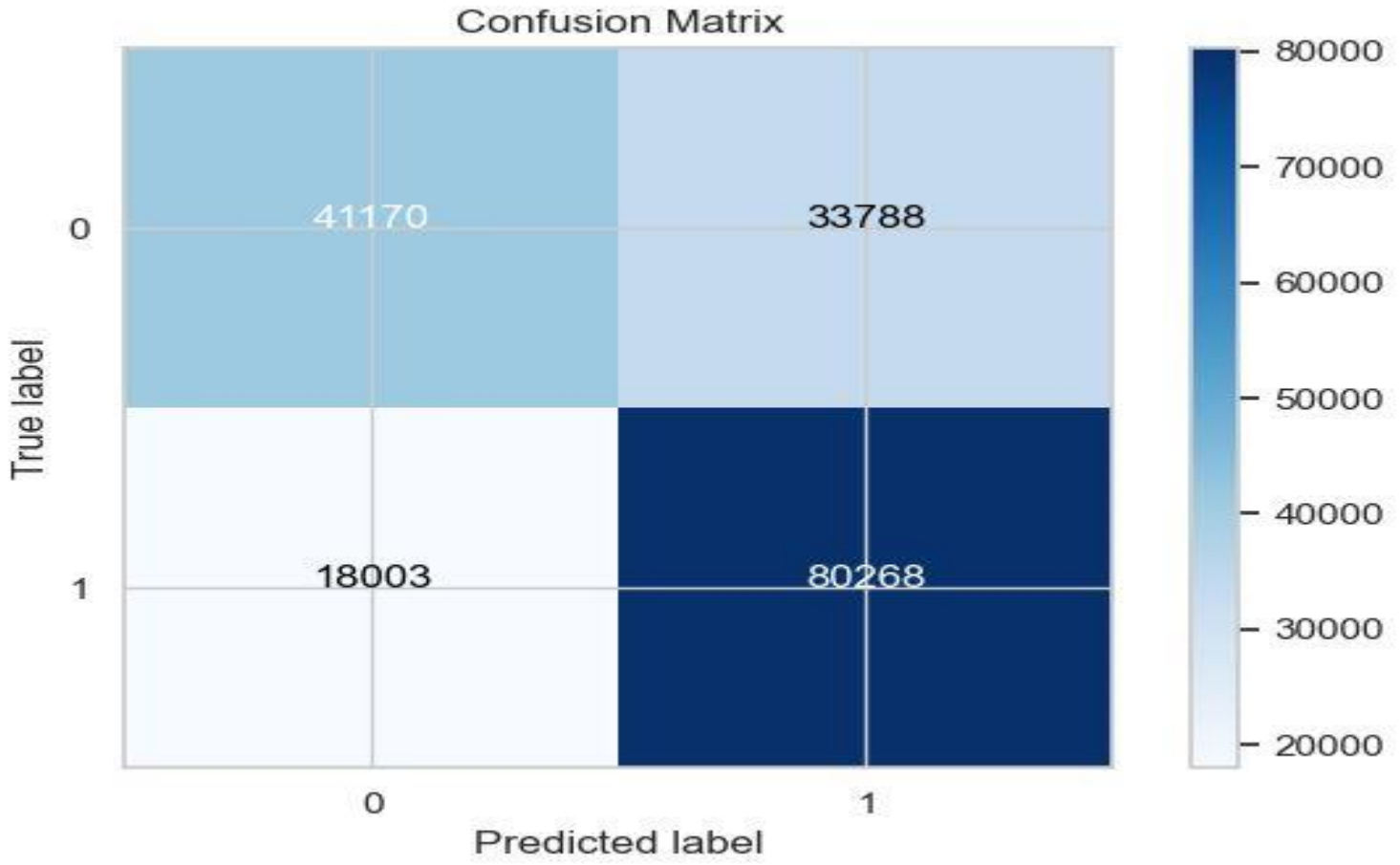
- The analysis of the crash data by day of the week indicates that there is a slightly higher number of accidents during the weekends compared to other days. However, the difference in accident frequency between weekdays and weekends is not substantial. Therefore, it can be concluded that the crash hour plays a more significant role in determining accident occurrence than the specific day of the week.
- Analyzing the crash data by month reveals some interesting patterns. The number of car accidents in Chicago tends to be higher during the summer months, particularly in June, July, August and September. This can be attributed to various factors such as increased travel and tourism, more outdoor activities, and potentially more congested roads during the summer season.

However, it is important to note that while there may be higher accident rates during certain months, the difference in accident frequency between months is not significant enough to warrant major adjustments in road safety strategies based solely on the crash month.

Question Four

Can we build a classification model to predict the primary contributory cause of car accidents?

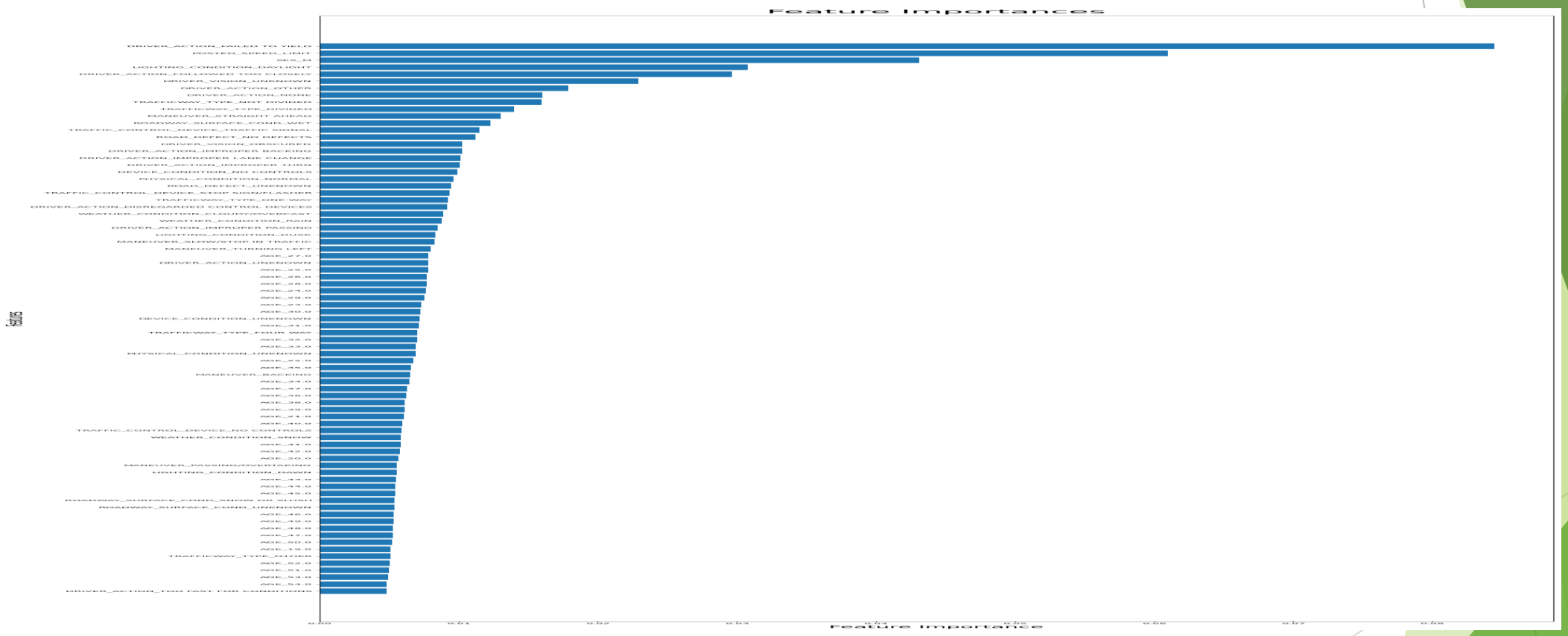
Developing a classification model will enable the CCVSB to categorize accidents into different causes, allowing for a deeper understanding of the factors contributing to each type of accident. This knowledge can inform targeted strategies for prevention. We will answer the question with the best classification model



Based on the provided results, the XG Boost model achieved the highest accuracy of 0.70, outperforming the other models (Logistic Regression, Decision Tree, and Random Forest) in terms of accuracy. The XG Boost model also had relatively higher precision scores, F1-scores for both classes (0 and 1), and AUC compared to the other models.

Therefore, based on the evaluation metrics, the XG Boost model is considered the best model for the given task.

Feature Selection



- The feature "DRIVER_ACTION_FAILED_TO_YIELD" stands out as the most important feature, followed by "POSTED_SPEED_LIMIT" and "SEX_M". These features play a crucial role in predicting the type of accidents associated with specific driver actions, whether they are classified as "INTENTIONAL" or "UNINTENTIONAL".

- * The inclusion of features such as "TRAFFIC_CONTROL_DEVICE", "WEATHER_CONDITION", "TRAFFICWAY_TYPE", and "ROADWAY_SURFACE_COND" in the list of important features indicates that these characteristics are strong predictors of accidents. These features carry significant weight in determining the occurrence and severity of accidents.

- * The data suggests that drivers are at a higher risk of accidents during their mid-20s, particularly at the age of 25. As drivers progress into older age groups, the likelihood of accidents gradually decreases. This pattern implies that drivers in their mid-20s may exhibit certain characteristics or behaviors that contribute to a higher accident risk compared to other age groups.

Conclusion

- Downtown Chicago has a high concentration of accidents, primarily caused by intentional actions or driver errors. However, there are scattered incidents of unintentional accidents, indicating the need for safety improvements.
- Control failures in unintentional accidents are not significantly influenced by vision or speed. Other factors may contribute to these accidents and require further investigation.
- The absence of traffic control devices is a significant contributing factor to accidents in Chicago. Increasing their presence can help reduce unintentional accidents.
- Weather and lighting conditions have minimal impact on accident occurrence in Chicago.
- Accidents are common on non-divided roads, suggesting the need for road division measures to improve traffic management and safety.
- Road surface condition and defects have a minimal impact on unintentional accidents.
- Rush hour traffic, particularly between 14-18 hours, contributes to a higher number of accidents in the downtown area. Better traffic management strategies are needed during these peak hours.
- Weekend days show a slightly higher number of accidents compared to weekdays, but crash hour plays a more significant role in determining accident occurrence.
- Summer months have a higher number of accidents, potentially due to increased travel and outdoor activities. However, adjustments in road safety strategies based solely on the crash month may not be necessary.

Recommendations

- Increase Traffic Control Measures: Install additional traffic control devices, such as traffic lights, stop signs, and speed limit signs, particularly in areas with a high concentration of accidents. Ensure that existing devices are well-maintained and functioning properly.
- Enhance Road Infrastructure: Implement road division measures, such as adding medians or physical barriers, to separate opposing flows of traffic and reduce the likelihood of collisions. Improve road surfaces to minimize hazards like potholes or uneven pavement.
- Improve Traffic Management: Implement intelligent transportation systems and optimize traffic signal timings to facilitate traffic flow and reduce congestion, especially during peak rush hour periods. Consider deploying additional traffic management personnel to ensure efficient traffic management.
- Driver Education and Awareness: Conduct targeted educational campaigns to raise awareness about safe driving practices, including the importance of attentiveness, obeying traffic laws, and maintaining a safe speed. Emphasize the risks associated with intentional actions, such as reckless driving or aggressive behavior.
- Collaborate with Law Enforcement: Strengthen collaboration between the City of Chicago Vehicle Safety Board, law enforcement agencies, and other relevant stakeholders to enforce traffic laws effectively and deter dangerous driving behaviors.
- Continuous Monitoring and Evaluation: Establish a robust system to collect and analyze data on car accidents continuously. Regularly evaluate the effectiveness of implemented measures and adjust strategies based on evolving trends and patterns in accidents.

Next Step

- Detailed Analysis of Contributory Factors: Conduct a deeper analysis of the contributing factors associated with intentional and unintentional accidents. Explore specific driver actions, physical conditions, maneuver types, and other variables to identify patterns and prioritize targeted interventions.
- Advanced Modeling Techniques: Consider implementing advanced modeling techniques, such as ensemble methods, deep learning, or anomaly detection algorithms, to improve the accuracy and robustness of the classification model for predicting primary contributory causes of accidents.
- Geographic Hotspot Analysis: Perform a spatial analysis to identify specific locations or road segments with a higher frequency of accidents. Use techniques such as hotspot analysis or spatial clustering to pinpoint areas requiring focused attention for safety improvements.
- Long-Term Data Analysis: Expand the analysis period beyond the available data to identify long-term trends and changes in accident patterns. This can help in identifying the effectiveness of implemented measures over time and detecting emerging issues or new risk factors.
- Collaboration with Insurance Companies: Collaborate with insurance companies to access additional data, such as accident severity, insurance claims, and vehicle information. This data can provide further insights into the factors associated with severe accidents and help develop targeted strategies for prevention.


Thank you
PIRELLI
SILVERSTONE