Charting the Route to Safety: Visualizing Chicago's Road Crash Data

Vaishnavi Nanduri , **Charishma Tamarana** ,
Sushma Tacholi Kudai , and **Laxmi Banas**i.

Abstract—**This project aims to perform a thorough analysis of car accidents in Chicago using a wide-ranging dataset covering different aspects related to accidents, injuries, and contributing factors. Through meticulous examination of data and visualization methods, the analysis aims to uncover significant patterns and connections that can help improve road safety throughout the city. Employing a variety of visual representations, the project aims to reveal valuable insights. The findings from this study can inform the development of targeted measures, policy decisions, and educational initiatives aimed at enhancing road safety, reducing the human and financial impact of accidents, and ultimately fostering a safer and more sustainable transportation environment for both residents and visitors of the city. Our project includes an analysis of crash occurrences over time, injury severity distributions, roadway surface conditions, Investigation into top contributing factors to crashes, weather conditions and crash types, analysis of primary and secondary contributions to crashes, and damage amounts.**

I. Introduction

Chicago, the vibrant heart of the Midwest, is a bustling metropolis that attracts millions of residents and visitors alike. With its rich cultural tapestry, architectural wonders, and diverse communities, the city offers an endless array of opportunities and experiences. However, amidst this urban dynamism lies a pressing challenge – the prevalent issue of vehicular crashes and the consequent injuries and fatalities that impact the lives of countless individuals.

Vehicular crashes in Chicago not only result in tragic loss of life and personal suffering but also impose substantial economic burdens on the city and its inhabitants. These incidents disrupt the flow of traffic, cause property damage, and strain emergency response resources, ultimately affecting the overall quality of life and economic vitality of the city.

By leveraging a comprehensive dataset encompassing a multitude of variables related to crashes, this project seeks to uncover invaluable insights that can inform targeted interventions and policy decisions to enhance road safety across the city.

II. Motivation

This project is fueled by the urgent necessity to address the significant challenge of road safety in Chicago. As a bustling city with intricate road networks and diverse populations, Chicago faces profound repercussions when crashes occur. These incidents not only harm those directly involved but also pose extensive societal, economic, and public health challenges.

A. *Impact on Physical and Emotional Well-being*

When individuals suffer injuries or fatalities in crashes, it inflicts immense pain on them, their families, and their friends. Moreover, crashes strain hospitals and emergency services, making it harder for others to access medical assistance promptly. Additionally, crashes diminish people's trust in the safety of roads, affecting their sense of belonging within their communities.

B. *Pressure on Healthcare Systems*

The occurrence of crashes places immense pressure on hospitals and emergency services, impeding timely access to medical care for those in need. This strain exacerbates existing healthcare challenges and undermines community trust in road safety measures.

C. *Disruption of Community Harmony*

Crashes undermine people's confidence in road safety and erode their sense of community belonging. Furthermore, crashes incur significant financial costs, not only in terms of medical expenses and vehicle repairs but also due to their adverse effects on businesses and workforce productivity.

D. *Economic Implications*

Crashes impose substantial economic burdens, encompassing medical bills, vehicle repairs, and losses in productivity. These financial strains impact individuals, businesses, and the city's budget, hindering economic growth and prosperity.

E. *Commitment to Public Safety*

Our commitment is to prioritize public safety by understanding the root causes of crashes and implementing measures to prevent them. By fostering safer roads and promoting responsible driving behavior, we aim to protect the well-being of all road users.

F. *Empowering Insights Through Data*

We endeavor to leverage data-driven insights to enhance road safety in Chicago. By analyzing crash data and identifying key factors contributing to crashes, we can formulate targeted interventions to mitigate risks and improve safety outcomes. Our goal is to empower stakeholders with the knowledge and resources needed to create safer roadways and communities.

III. Problem Statement

Traffic collisions pose a multifaceted challenge in urban settings, jeopardizing public safety, straining emergency services, and incurring significant economic burdens. Within Chicago, a sprawling metropolitan area characterized by intricate transportation networks and diverse demographics, the prevalence of traffic accidents underscores the imperative of addressing this issue comprehensively. The problem statement emphasizes the necessity of grasping the underlying causes and dynamics of traffic collisions in Chicago to effectively mitigate their repercussions and bolster road safety.

A. *Threats to Public Safety*

Traffic collisions pose immediate hazards to public safety, resulting in physical harm, loss of life, and psychological distress for victims and their families. Furthermore, the aftermath of collisions often leads to traffic congestion, hindering emergency response operations and heightening risks for subsequent incidents.

B. *Pressure on Emergency Response Systems*

Traffic collisions strain emergency response systems, including ambulance services, hospitals, and law enforcement agencies. The surge in collision-related injuries necessitates prompt and efficient allocation of resources to deliver timely medical assistance and ensure public safety. Additionally, diverting resources to manage collision incidents can detract from other emergency response endeavors, potentially compromising overall readiness and effectiveness.

C. *Economic Impacts*

In addition to the human toll, traffic collisions impose substantial economic costs on society, encompassing medical expenditures, property damage, productivity losses, and legal expenses. Moreover, indirect costs associated with traffic congestion, vehicle delays, and reduced mobility exacerbate the economic burden on individuals, businesses, and the wider community.

D. *Holistic Understanding*

The problem statement underscores the importance of gaining a comprehensive understanding of traffic collisions in Chicago. This involves scrutinizing diverse datasets containing collision-related data, including but not limited to collision characteristics, contributing factors, environmental circumstances, and socioeconomic variables. Through thorough analysis of these datasets, we aim to discern patterns, trends, and risk factors that can inform evidence-based strategies for prevention, mitigation, and response.

E. *Unveiling Complex Dynamics*

Traffic collisions arise from a complex interplay of factors, including road design, traffic patterns, driver conduct, vehicle attributes, and environmental elements. Understanding the interactions among these variables and their impact on collision occurrences and severity outcomes is vital for devising effective interventions. By unraveling these intricate dynamics, we can identify intervention opportunities and allocate resources to address the most pertinent risk factors.

F. *Evidence-Driven Approaches*

Ultimately, the overarching goal of comprehensively understanding traffic collisions in Chicago is to inform evidence-driven approaches for prevention, mitigation, and response. This entails not only pinpointing high-risk zones and populations but also crafting targeted interventions tailored to address specific risk factors. By leveraging data-driven insights, we can optimize resource allocation, enhance public awareness, and implement infrastructure enhancements to cultivate safer roadways and diminish the frequency and severity of traffic collisions in Chicago

IV. Objective

The primary objective of this endeavor is to utilize advanced data analysis and visualization methods to glean practical insights into the intricacies of traffic crashes in Chicago. Our project aims revolve around the following core areas:

A. *Temporal Analysis*

Objective: Identify temporal patterns and trends in crash occurrences to pinpoint high-risk periods and locations.

Approach: Employ time-series analysis techniques to scrutinize variations in crash frequencies across different time intervals (e.g., hourly, daily, monthly).

Outcome: By detecting temporal patterns, we aim to inform targeted interventions and resource allocation strategies to address high-risk periods and mitigate crash occurrences.

B. *Contributing Factors Identification*

Objective: Investigate the multifaceted factors contributing to crash occurrences, including environmental, infrastructural, vehicular, and human factors.

Approach: Conduct thorough data analysis to explore the relationships between various factors and crash occurrences, examining variables such as weather conditions, road infrastructure, vehicle characteristics, and driver behavior.

Outcome: By identifying contributing factors, we aim to develop a nuanced understanding of the root causes of crashes and inform evidence-based interventions targeting specific risk factors.

C. *Severity Assessment*

Objective: Analyze the severity outcomes of traffic crashes, including injuries and fatalities, to discern patterns and risk factors associated with adverse outcomes.

Approach: Utilize statistical analysis techniques to evaluate the distribution and severity of injuries and fatalities resulting from crashes, examining variables such as crash type, vehicle speed, and safety equipment usage.

Outcome: By assessing crash severity, we aim to identify vulnerable populations and high-risk scenarios, guiding efforts to enhance emergency response protocols, medical care provision, and post-crash support services.

D. Correlation Exploration

Objective: Examine interrelationships between various variables such as weather conditions, road conditions, driver behavior, and crash severity to uncover causal pathways and inform targeted interventions.

Approach: Conduct correlation analysis and data visualization techniques to explore the associations between different variables, generating correlation matrices, scatter plots, and heatmaps.

Outcome: By uncovering correlations, we aim to elucidate the complex interactions between different factors influencing crash occurrences and severity outcomes. This knowledge can inform the development of predictive models, early warning systems, and targeted interventions to mitigate crash risks and improve road safety.

E. Dynamic Dashboards Creation in Power BI

Objective: Develop dynamic dashboards in Power BI to provide stakeholders with intuitive access to key findings and trends in real-time.

Approach: Utilize Power BI's interactive visualization capabilities to create dynamic dashboards enabling users to explore crash-related data from various perspectives, including filters, slicers, and drill-down functionality.

Outcome: By crafting dynamic dashboards in Power BI, we aim to enhance stakeholder engagement and decision-making by offering real-time access to actionable insights. These dashboards facilitate data-driven decision-making processes, enabling stakeholders to identify trends, monitor key metrics, and respond proactively to emerging challenges in road safety management.

Through these objectives, our aim is to leverage advanced data analysis techniques, including dynamic dashboard creation in Power BI, to gain deeper insights into the dynamics of traffic crashes in Chicago. We endeavor to empower stakeholders with the knowledge and tools needed to implement targeted interventions, improve road safety outcomes, and ultimately save lives.

V. LITERATURE REVIEW

This comprehensive literature review delves into various dimensions of traffic crashes and road safety measures, drawing insights from a range of research studies. Bačkalić's (2013) temporal analysis of traffic accidents in the Province of Vojvodina underscores the importance of understanding temporal patterns to pinpoint critical periods for decision-making in road safety management. By examining accident data over time, the study aims to identify recurring patterns and trends, enabling authorities to allocate resources effectively and implement targeted interventions during high-risk periods. Shahi's (2022) spatial analysis of road traffic crashes in Rotterdam offers valuable insights into the identification of hotspots and road users' perceptions of safety, vital for crafting effective road safety plans. Utilizing spatial analysis techniques, the research not only identifies areas with high crash rates but also explores the underlying factors contributing to these patterns, facilitating the development of tailored safety interventions and infrastructure improvements. Additionally, Garcia et al.'s (Year) study on weather impacts on traffic crashes in Chicago illuminates the influence of adverse weather conditions on crash occurrences, emphasizing the need for weather-responsive traffic management strategies. By analyzing crash data in relation to weather variables such as rain, snow, and temperature, the research provides valuable insights into the complex interactions between weather conditions and road safety, informing the development of proactive measures to mitigate weather-related risks. Naik et al.'s (2016) investigation into weather impacts on single-vehicle truck crash injury severity provides further depth by highlighting the correlation between detailed weather data and crash-related injury severity. By integrating comprehensive weather data with crash and roadway data, the study identifies how factors such as wind speed, rain, humidity, and air temperature influence the severity of truck crashes, offering valuable insights for enhancing truck safety measures in different weather conditions. Furthermore, Martinez et al.'s (Year) analysis of pedestrian and cyclist crashes in Chicago sheds light on the vulnerability of non-motorized road users and underscores the importance of implementing safety measures to protect them. Through an examination of crash data involving pedestrians and cyclists, the study identifies risk factors and trends specific to these road users, guiding the development of targeted interventions such as improved infrastructure, traffic calming measures, and public awareness campaigns. Lastly, our project focuses on analyzing Chicago crash data to identify contributing factors and spatial-temporal patterns, providing valuable insights for policymakers and stakeholders to enhance road safety measures in the city. By leveraging advanced analytical techniques and machine learning algorithms, our research aims to uncover hidden patterns and correlations within crash data, informing evidence-based strategies for reducing crash occurrences and improving road safety outcomes. Collectively, these studies contribute to a holistic understanding of traffic crash dynamics and offer actionable insights for improving road safety outcomes in urban environments.

REFERENCES

[1] Naik, Bhaven Tung, Li-Wei Zhao, Shanshan Khattak, Aemal. (2016). Weather impacts on single-vehicle truck crash injury severity. Journal of Safety Research. 58. 10.1016/j.jsr.2016.06.005.

[2] Bačkalić, Svetlana. (2013). Temporal analysis of the traffic accidents occurrence in province of Vojvodina. Transport Problems. 8. 87-93.

[3] Hale, A.R. Guldenmund, Frank Loenhout, P.L.C.H. Oh, J.. (2010). Evaluating safety management and culture interventions to improve safety: Effective intervention strategies. Safety Science. 48. 1026-1035. 10.1016/j.ssci.2009.05.006.

VI. DATAFLOW DIAGRAM

Fig. 1. Dataflow Diagram

VII. Data Description

Several datasets are used in the project ”Charting the Route to Safety: Visualizing Chicago’s Road Crash Data” to provide a comprehensive analysis of road crashes in Chicago. The primary dataset, crash.csv, contains essential information about every crash, such as the Crash Type, which categorizes the type of incident (such as a rollover or collision), and the Crash Date and Crash Time, which show when the crash happened. The crash’s impact level, which ranges from minor to fatal, is detailed in the Severity column, giving important context for understanding how serious the crash is.

In addition, crash_weather.csv provides weather-related information at the time of the crash. Columns such as Weather Condition, which describes weather types like clear, rainy, or snowy, Visibility, which indicates view clarity (like clear, foggy), and Temperature, which displays the temperature at the time of the crash, are included. With Latitude and Longitude columns providing the precise crash location and Location Description providing a textual description of the crash site, such as an intersection or highway, the crash_location.csv dataset adds a spatial layer to the analysis.

With respect to the human impact, the crash_injury.csv dataset provides information on the quantity and seriousness of injuries. The total number of injuries is shown in the Injuries Total column, and the counts of fatal, serious, and minor injuries are indicated in the Injuries Fatal, Injuries Serious, and Injuries Minor columns, respectively. Last but not least, the crash_cause.csv file lists the variables that contribute to crashes. The Primary Cause column indicates the primary cause of the crash, such as speeding or distracted driving, while the Secondary Cause column lists additional variables. The types of vehicles involved, such as cars, trucks, or motorcycles, are listed in the Vehicle Type column. When combined, these datasets provide a thorough perspective that makes it possible to analyze crash dynamics in great detail and aids in the creation of successful road safety initiatives.

VIII. Data cleaning and Preprocessing

Ensuring data accuracy and reliability is paramount in any data analysis endeavor, and data cleaning and preprocessing play a pivotal role in achieving this goal. For this project, multiple measures were implemented to clean and prepare the dataset for analysis.

One crucial step involved addressing null or missing values, which can arise from various factors such as data entry errors or incomplete information. Depending on the nature of the data and the specific features, different techniques were utilized to handle these null values. In our instance, the null values were removed as their presence did not significantly impact the analysis.

Another important task was the identification and removal of unnecessary features that did not contribute to the project’s objectives. Retaining irrelevant features not only increases computational complexity but can also introduce noise, potentially leading to erroneous or misleading results. By carefully evaluating each feature’s relevance and eliminating those that did not contribute to the analysis, the dataset was streamlined, making it more manageable and focused.

To enhance organization and facilitate more targeted analysis, the dataset was divided into different themes based on shared characteristics or attributes. This division enabled a more focused approach, allowing for deeper insights to be drawn from each thematic subset while maintaining the overall integrity of the data.

Furthermore, duplicate values, which can arise due to data entry errors or other factors, were identified and handled appropriately. Duplicate records can skew the analysis and lead to inaccurate results.

Through the implementation of these data cleaning and preprocessing steps, the dataset underwent a transformation, resulting in a more reliable and accurate representation of the underlying information. This refined dataset was better suited for meaningful analysis and insight generation, ensuring the validity and robustness of the project’s findings.

IX. Visualizations

A. *Time Series Plot showing Number of crashes over time*

A time series plot illustrates the progression of data points gathered over time, with the horizontal axis denoting time and the vertical axis representing the measured variable’s values. Specifically, this plot visualizes the frequency of crashes over time. Each plotted point corresponds to a particular time frame, typically months, with its vertical position indicating the number of crashes during that period. By connecting these points using lines or markers, the plot offers a visual depiction of how crash occurrences have evolved over time. Prior to 2015, there were no recorded crashes, suggesting either a lack of data for that year or an exceptionally safe period with zero incidents. Starting from 2016, there was a consistent uptrend in crash frequency, interrupted by a minor decrease in early 2017 followed by subsequent increases in the ensuing months. The peak in crash occurrences was observed in 2018 and 2019, with a notable decline in 2020 likely attributable to factors such as travel restrictions imposed due to the COVID-19 pandemic. Both 2022 and 2023 displayed a similar cyclic pattern of crash frequency, characterized by alternating increases and decreases.

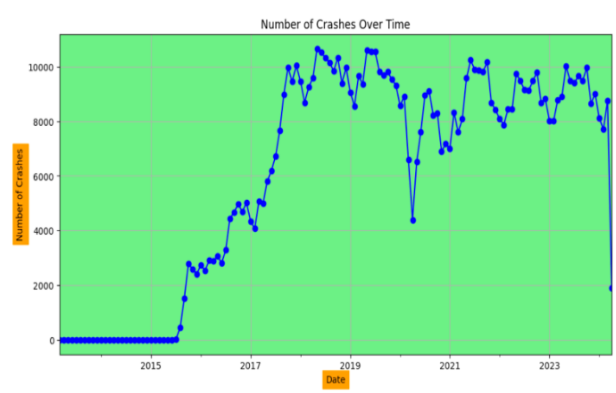

Fig. 2. Time Series Plot showing Number of crashes over time

In essence, the time series plot facilitates the examination of trends, patterns, and fluctuations in crash frequency across the specified timeframe, thereby aiding in the analysis and comprehension of the underlying dataset.

B. Pair Plot showing Pairwise relationships between total injuries , fatal injuries and posted speed limit

The pair plot, consisting of a grid of subplots, allows for comprehensive visualization of relationships between multiple variables simultaneously. Each subplot within the grid represents a unique combination of variables, facilitating exploration of their interdependencies. Along the diagonal of the grid, histograms are displayed to illustrate the distribution of individual variables, aiding in the understanding of their data distribution characteristics.

In the scatter plots generated for pairwise comparisons, each point represents a data point from the DataFrame, with one variable plotted on the x-axis and the other on the y-axis. Through examination of these scatter plots, patterns such as linear trends, clusters, or outliers can be identified, providing valuable insights into the relationships between the variables.

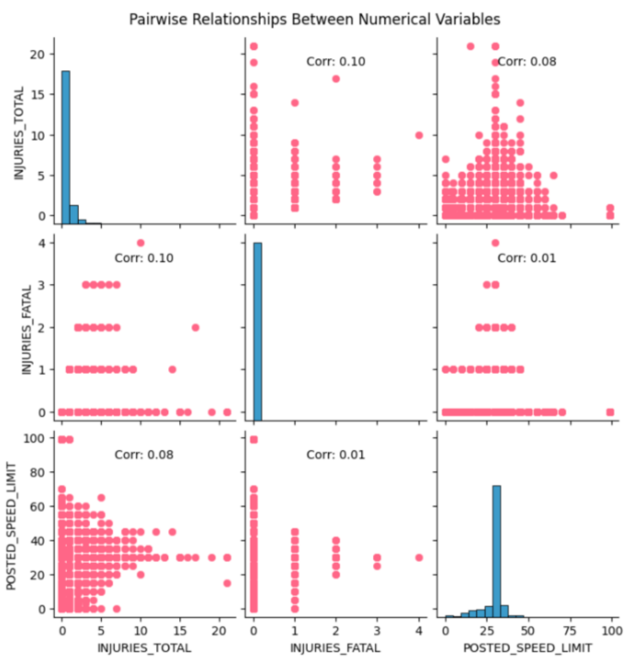

Fig. 3. Time Series Plot showing Number of crashes over time

The correlation coefficients calculated for each pair of variables further enhance our understanding. For instance, the comparison between Injuries_Total and Injuries_Fatal reveals a mild positive correlation (correlation coefficient: 0.10), suggesting that as the total number of injuries increases, there's a slight tendency for the number of fatal injuries to rise as well. Similarly, the comparison between Injuries_Total and Posted_Speed_Limit yields a correlation coefficient of 0.08, indicating a subtle inclination for the total number of injuries to increase with higher posted speed limits, albeit with relatively weak correlation.

When analyzing Injuries_Fatal versus Injuries_Total, the same correlation value as the first comparison reaffirms the observed mild positive correlation between these variables. Conversely, the comparison between Injuries_Fatal and Posted_Speed_Limit results in a correlation coefficient of 0.01, indicating an almost negligible positive correlation, suggesting an almost nonexistent linear relationship between fatal injuries and the posted speed limit.

Similarly, the correlation coefficient of 0.08 observed in the comparison between Posted_Speed_Limit and Injuries_Total reaffirms the mild positive correlation between these variables. Lastly, the examination of Posted_Speed_Limit versus Injuries_Fatal corresponds with the fourth comparison, reaffirming the extremely faint positive correlation between the number of fatal injuries and the posted speed limit.

In summary, these correlation coefficients provide valuable insights into the relationships between variables, aligning with the observed patterns in the scatter plots and contributing to a comprehensive understanding of the dataset.

C. Correlation matrix as a heatmap to explore relationships between numerical variables

- The Crash_ID exhibits a minimal negative correlation with both CRASH_HOUR and CRASH_DAY_OF_WEEK, indicating a weak linear relationship.
- There is a slight positive correlation between CRASH_HOUR and CRASH_DAY_OF_WEEK, suggesting that as the hour of the crash increases, there is a tendency for it to happen slightly more often on certain days of the week.
- CRASH_MONTH shows a very weak positive correlation with CRASH_DAY_OF_WEEK, hinting at a minor inclination for specific days of the week to experience more crashes during certain months.
- POSTED_SPEED_LIMIT demonstrates a weak positive correlation with CRASH_HOUR, implying a slight rise in crash occurrences with higher posted speed limits.
- There is a moderate positive correlation between INJURIES_TOTAL and POSTED_SPEED_LIMIT, indicating a noticeable increase in total injuries as the posted speed limit goes up.
- INJURIES_FATAL has a weak positive correlation with INJURIES_TOTAL, suggesting a minor tendency for the number of fatal injuries to increase as the total number of injuries rises.
- INJURIES_INCAPACITATING displays a moderate positive correlation with INJURIES_TOTAL, implying a significant increase in incapacitating injuries with an increase in total injuries.
- INJURIES_NON_INCAPACITATING shows a strong positive correlation with INJURIES_TOTAL, indicating a notable increase in non-incapacitating injuries as the total number of injuries rises.
- STREET_NO exhibits a very weak positive correlation with INJURIES_TOTAL, indicating a minimal linear relationship between them.

- The LONGITUDE and LATITUDE variables have a strong negative correlation, suggesting that as one increases, the other decreases, implying a spatial relationship.

Understanding Heatmaps:

Color Range: Heatmaps use colors to show how things are connected. Warm colors like red and orange mean things go together well, while cool colors like blue mean they don't. Darker colors mean a stronger connection.

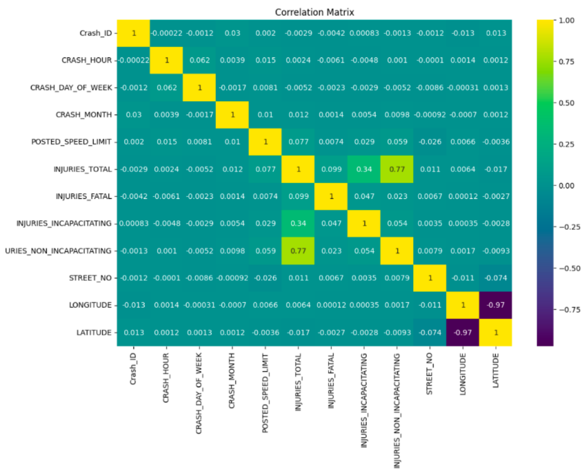

Fig. 4. Correlation matrix as a heatmap to explore relationships between numerical variables

Diagonal Line: There's a line from the top-left to the bottom-right of the heatmap. It shows how each thing compares to itself, always with a perfect match.

Other Squares: The rest of the squares show how two different things are connected. Each square has a number that tells us how strong the connection is and which way it goes.

What It Shows: By looking at the colors and numbers in the squares, we can quickly see which things are closely linked, whether in a good or bad way, and which ones aren't linked much at all.

D. *Horizontal bar plot for top Contributing Factors to crashes*

The predominant cause of crashes, "Unable to Determine," tops the list with around 300,000 instances. Additionally, "Failing to Yield Right of Way" and "Following Too Closely" are prevalent, with estimated occurrences ranging between 50,000 and 100,000 each.

Incidents categorized as "Not Applicable" fall within the range of 40,000 to 50,000 occurrences. "Improper Overtaking/Passing" is responsible for approximately 40,000 incidents.

Further down the list, "Failing to Reduce Speed to Avoid Crash" and "Improper Backing" each have less than 50,000 occurrences. Similarly, "Improper Lane Usage" and "Improper Turning/No Signal" hover around 25,000 to 30,000 instances. Reports of issues related to "Driving Skills" total roughly 25,000 times.

These findings shed light on various aspects:

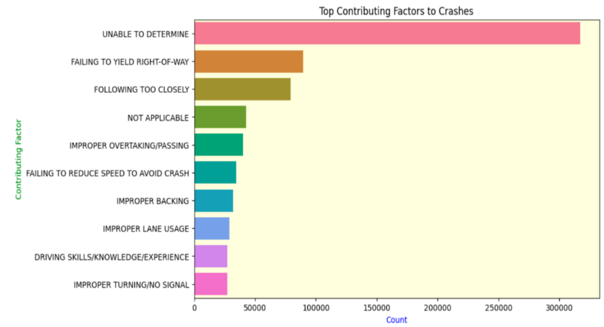

Fig. 5. Horizontal bar plot for top Contributing Factors to crashes

- **Common Causes:** "Unable to Determine," "Failing to Yield Right of Way," and "Following Too Closely" highlight areas for enhancing driving conduct and enforcing traffic rules.
- **Safety Concerns:** Instances such as "Not Applicable" and "Failing to Reduce Speed to Avoid Crash" raise concerns about safety lapses, necessitating proactive measures to prevent accidents.
- **Specific Driving Errors:** Errors such as "Improper Overtaking/Passing," "Improper Backing," "Improper Lane Usage," and "Improper Turning/No Signal" pinpoint specific driving behaviors contributing to accidents, indicating the need for educational initiatives and enforcement actions.
- **Frequency Analysis:** The distribution of incidents across different causes offers insights into where resources and efforts should be channeled for targeted interventions and safety campaigns.

E. *Injury Severity Distribution*

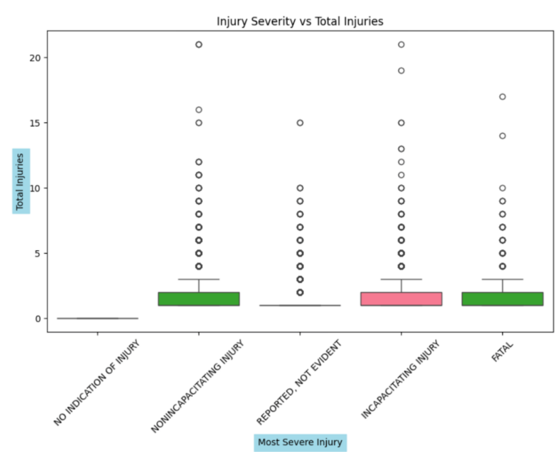

Fig. 6. Injury Severity Distribution

The box plot illustrates the distribution of total injuries across different categories of the most severe injury in a crash. The x-axis lists the various levels of injury severity: no indication of injury, non-incapacitating injury, reported but not evident, incapacitating injury, and fatal injury. The y-axis represents the total number of injuries. For each category of injury severity, the box plot shows the median (the line inside the box), the first and third quartiles (the bottom and top of the box, respectively), and the minimum and maximum values (the whiskers extending from the box). The categories like Incapacitating, non-incapacitating and fatal tend to have more severe and involve higher numbers of injuries. There are a few outliers (dots) representing crashes with an unusually high number of total injuries.

F. Roadway Surface Condition Distribution using Violin Plot

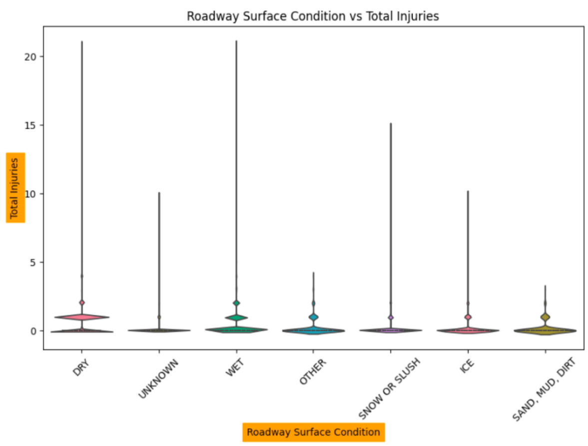

Fig. 7. Roadway Surface Condition Distribution using Violin Plot

Violin plots effectively display the full distribution of data for each category, not just summary statistics like mean or median. The width of each "violin" corresponds to the density of data points at that value, giving a clear picture of how the data is spread out within each category.

This plot reveals that "Dry" road conditions tend to have the highest number of associated injuries, followed by "Wet" and "snow" conditions.

G. Stacked Bar plot for Crash type contribution across different weather conditions

The stacked bar chart illustrates the distribution of two crash types - "NO INJURY / DRIVE AWAY" and "INJURY AND / OR TOW DUE TO CRASH" - across the top 5 weather conditions. "CLEAR" weather had the highest number of crashes for both crash types, with "NO INJURY / DRIVE AWAY" crashes being more common than "INJURY AND / OR TOW DUE TO CRASH".

"RAIN" had the second-highest number of crashes, with a similar distribution between the two crash types as seen in "CLEAR" weather. "UNKNOWN", "SNOW", and "CLOUDY/OVERCAST" weather conditions have significantly fewer crashes. Across all weather conditions, "NO INJURY / DRIVE AWAY" crashes are more frequent than "INJURY AND / OR TOW DUE TO CRASH", with the proportion between the two crash types remaining relatively consistent.

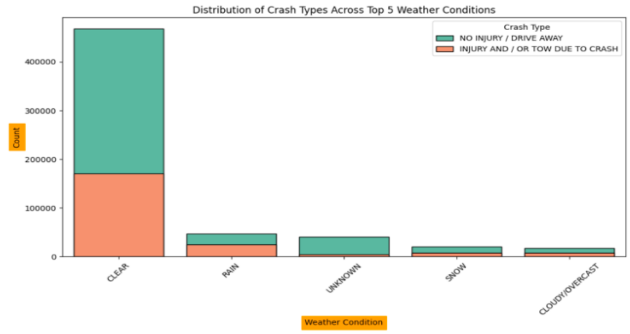

Fig. 8. Stacked Bar plot for Crash type contribution across different weather conditions

The 12-year trend in fatal injuries from traffic crashes between 2013 and 2024 is depicted in the line graph. At first, between 2013 and 2014, there were comparatively few fatal injuries—nearly insignificant—which suggested that there weren't many fatal incidents. There was a discernible rise in fatal injuries beginning in 2015, and this trend continued to get worse through 2016 and 2017. This increasing pattern persisted, reaching a notable apex in 2018—the highest point of fatal injuries during the whole time. The number of fatal injuries decreased somewhat in 2019, but it was still comparatively high, indicating that fatal crashes remained a significant concern.

The trend increased again between 2020 and 2021, peaking at a secondary location, suggesting a resurgence of fatal crashes. Nonetheless, the number of fatal injuries gradually decreased beginning in 2022. This decrease demonstrated a notable drop in fatal injuries by 2023 and continued sharply into 2024. Hence, the graph shows two major peaks in 2018 and approximately in 2021, followed by a significant decline towards the end of the period, indicating recent advancements in traffic safety and a noteworthy drop in fatal crashes in the latter years.

X. Tableau Dashboard

Using a variety of visualizations, the dashboard "Charting the Route to Safety: Visualizing Chicago's Road Crash Data" offers a thorough summary of crash-related data. The incident types on the Tree Map are "No Injury / Drive Away" and "Injury and/or Tow Due to Crash," with the total of the posted speed limits indicated by the size and color of each category. The Packed Bubbles chart illustrates crash incidents according to lighting, emphasizing that the majority of crashes take place during bright daylight hours. The monthly distribution of fatal injuries across various location groups is depicted in the Area Map, with peaks occurring in the third and seventh months. The trend of fatal injuries over time is finally shown by the Line Graph, which shows notable peaks in 2018 and 2021 and a notable decline in 2024. The spatial data is effectively visualized by this dashboard.

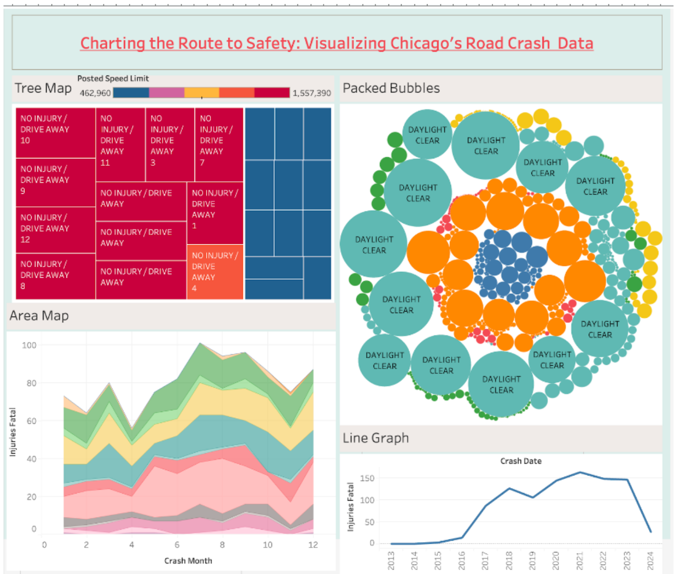

Fig. 9. Tableau Dashboard

A. *Line Graph for Fatal Injuries on Crash Dates*

The 12-year trend in fatal injuries from traffic crashes between 2013 and 2024 is depicted in the line graph. At first, between 2013 and 2014, there were comparatively few fatal injuries—nearly insignificant—which suggested that there weren't many fatal incidents. There was a discernible rise in fatal injuries beginning in 2015, and this trend continued to get worse through 2016 and 2017. This increasing pattern persisted, reaching a notable apex in 2018—the highest point of fatal injuries during the whole time. The number of fatal injuries decreased somewhat in 2019, but it was still comparatively high, indicating that fatal crashes remained a significant concern. The trend increased again between 2020 and 2021, peaking at a secondary location, suggesting a resurgence of fatal crashes. Nonetheless, the number of fatal injuries gradually decreased beginning in 2022. This decrease demonstrated a notable drop in fatal injuries by 2023 and continued sharply into 2024. Hence, the graph shows two major peaks in 2018 and approximately in 2021, followed by a significant decline towards the end of the period, indicating recent advancements in traffic safety and a noteworthy drop in fatal crashes in the latter years.

B. *Tree Map*

By classifying incidents into "No Injury / Drive Away" and "Injury and/or Tow Due to Crash," with each category further subdivided by crash month, the tree map visualizes data on road crashes. The blocks are colored red to blue to represent the severity or impact of crashes, and their sizes are determined by adding up the posted speed limits. Much of the map is taken up by the "No Injury / Drive Away" category, which is primarily colored red and orange and indicates that there was either a higher frequency of these incidents or a higher total speed limit. The numbers in these blocks represent distinct months; they exhibit a regular distribution over a range of months with some variance in block sizes, signifying variations in the speed limits as posted.

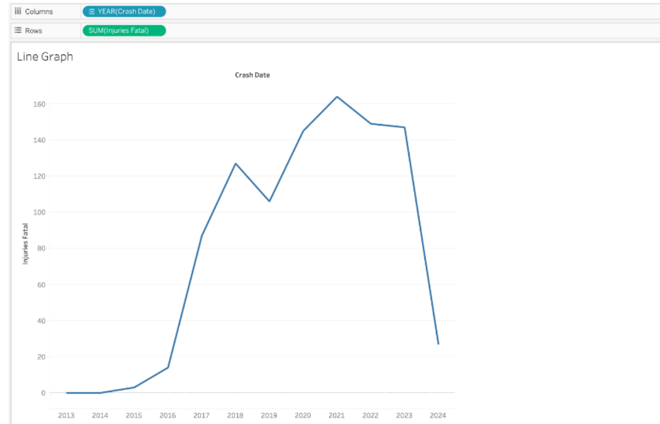

Fig. 10. Line Graph for Fatal Injuries on Crash Dates

On the other hand, the blue-hued "Injury and/or Tow Due to Crash" category shows fewer but noteworthy incidents involving injury or towing. These blocks are typically smaller than those in the "No Injury / Drive Away" category, indicating lower frequencies or overall speed limits. A clear visual comparison between the less severe and more severe crash outcomes is provided by the color gradient that runs from red to blue. The distribution and severity of crash incidents across various types and months are clearly shown by this tree map, with a noticeable preference for non-injury crashes over those that require towing or result in injuries.

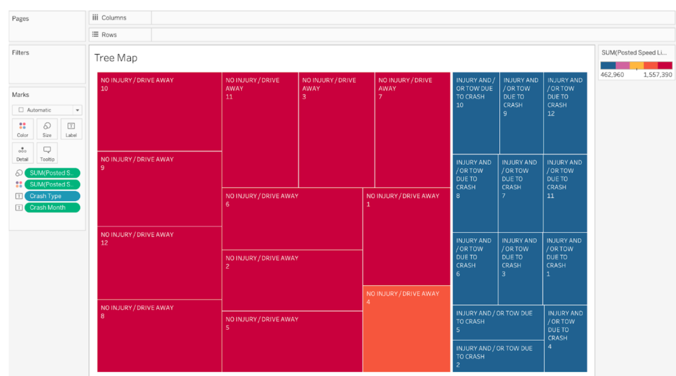

Fig. 11. Tree Map

The analysis of different classes of power rectifiers will next be analyzed based on the harmonic terminations presented to the rectifying element, and independent of the physical nonlinear device which performs the rectification.

XI. Power BI Dashboard

The Power BI dashboard offers a comprehensive examination of traffic crash data in Chicago, delivering invaluable insights into the multifaceted aspects of road safety. Through succinct card visuals, users gain immediate access to crucial metrics, including a comprehensive tally of 155,000 injuries, among which 969 resulted in fatalities. Furthermore, the dashboard meticulously breaks down injury severity, with 16,000

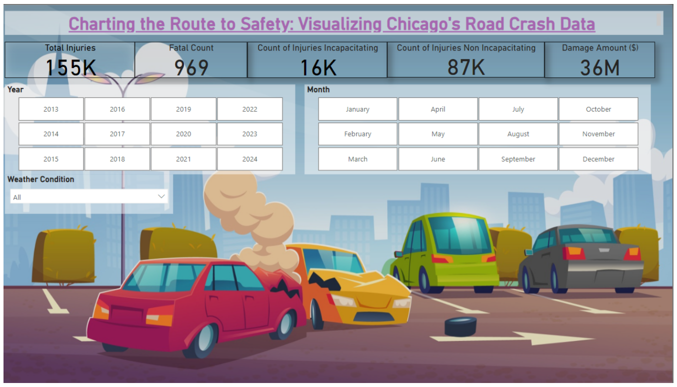

Fig. 12. PBI Dashboard 1

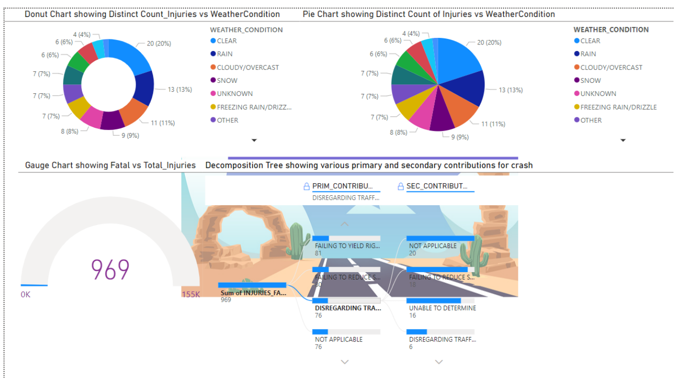

Fig. 13. PBI Dashboard 2

recorded as incapacitating and 87,000 as non-incapacitating, illuminating the diverse impact of crashes on victims. The financial ramifications are equally pronounced, with the total damages amounting to a staggering $36 million, underscoring the significant economic toll of road traffic incidents. Facil-

Fig. 14. PBI Dashboard 3

itating an interactive exploration of the data landscape, the dashboard incorporates slicers that empower users to delve into various dimensions of the data. With options to filter by year, spanning from 2013 to 2024, users can discern temporal trends and identify shifts in crash patterns over time. Similarly, slicers for month, street name in Chicago, weather conditions (including clear, cloudy, overcast, fog, rain, and drizzle), and location within the city afford stakeholders the flexibility to conduct granular analyses tailored to specific scenarios. This versatility facilitates a nuanced understanding of the underlying factors contributing to crash occurrences and enables stakeholders to pinpoint areas requiring targeted interventions.

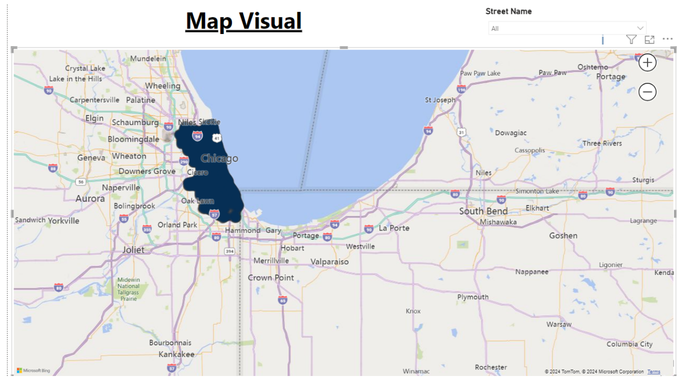

Fig. 15. PBI Dashboard 4

In addition to the structured presentation of data through card visuals and slicers, the dashboard employs intuitive visualizations such as donut charts, pie charts, and ribbon charts to illustrate complex relationships and trends. The distribution of injuries across different weather conditions is vividly portrayed, offering insights into weather-related risk factors and their impact on crash outcomes. Similarly, the ribbon chart provides a nuanced depiction of the correlation between lighting conditions and injuries, shedding light on visibility-related factors influencing crash severity.

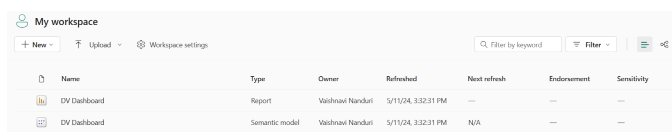

Fig. 16. Workspace

URL - https://app.powerbi.com/groups/me/reports/436c7e03-b198-4e7f-ae31-de68543c2720/ReportSection?experience=power-bi

Others might not be able to access the URL because we need Premium Subscription.

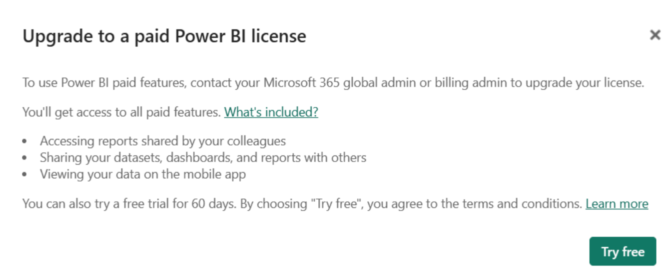

Fig. 17. Premium Image

In summary, the Power BI dashboard serves as a sophisticated analytical tool, empowering stakeholders to navigate through vast troves of data and derive actionable insights. By leveraging its interactive features and intuitive visualizations, stakeholders can gain a comprehensive understanding of road safety dynamics in Chicago, identify areas of concern, and formulate evidence-based strategies to enhance road safety outcomes and protect lives.

A. *Donut vs Pie Chart showing relation between Injuries count and Lighting condition*

In the Power BI dashboard, the donut and pie charts are employed to visually represent the distribution of injuries resulting from road crashes across various weather conditions. While both charts convey the same underlying data, they present it in slightly different formats, each offering distinct advantages in data visualization. The donut chart displays the data in the form of a ring, with each segment representing a specific weather condition such as clear, rainy, cloudy, overcast, or snowy. The size of each segment corresponds to the proportion of injuries attributed to that particular weather condition, allowing for quick and easy comparison of the relative contribution of each condition to the total number of injuries. On the other hand, the pie chart presents the data in a circular format, with each slice representing a weather condition and the size of each slice indicating the proportion of injuries. While both charts effectively convey the distribution of injuries across weather conditions, the pie chart may be more suitable for visualizing fewer categories due to its simplicity and ease of interpretation.

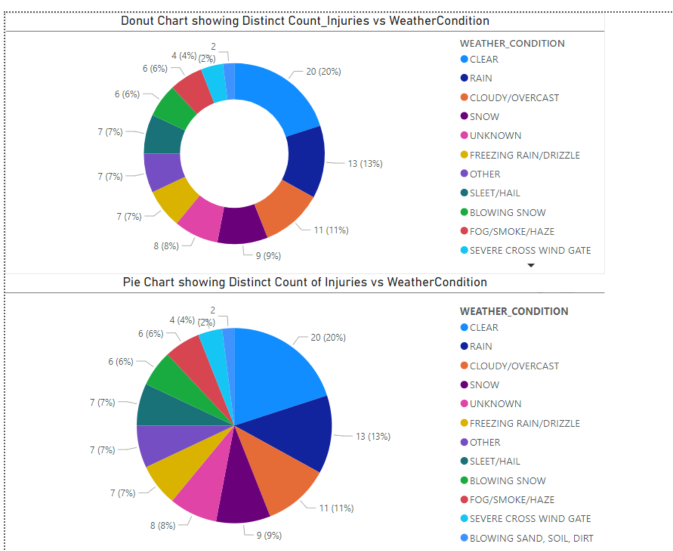

Fig. 18. Donut vs Pie Chart

In the specific graph showcasing the count of injuries in road crashes versus weather condition, stakeholders are presented with a detailed breakdown of injuries attributed to different weather conditions prevalent during road crash incidents. Clear weather conditions emerge as the dominant factor, contributing to the largest proportion of injuries at 20% of the total. Following closely behind are rainy conditions, which account for 13% of injuries, while both cloudy and overcast conditions collectively contribute 11% of injuries. Snow conditions, though less frequent, still have a significant impact, accounting for 9% of the total injuries recorded. This visualization help with valuable insights into the correlation between weather patterns and road crash injuries, facilitating a deeper understanding of the factors influencing road safety outcomes. Such insights enable the development of targeted interventions and safety measures aimed at mitigating the risks associated with specific weather conditions, ultimately contributing to more effective road safety strategies and initiatives.

B. *Gauge Chart showing Fatal Count vs Injuries*

In Power BI, the gauge chart is recognized as a versatile and visually engaging tool that surpasses its traditional role of tracking progress towards targets or presenting key performance indicators (KPIs). Its primary function is to provide stakeholders with an intuitive and immediate understanding of a single value within a defined range, enabling them to grasp the relative significance or severity of a metric or outcome. This dynamic visualization tool offers a clear depiction of data, simplifying the identification of trends, anomalies, and areas requiring attention.

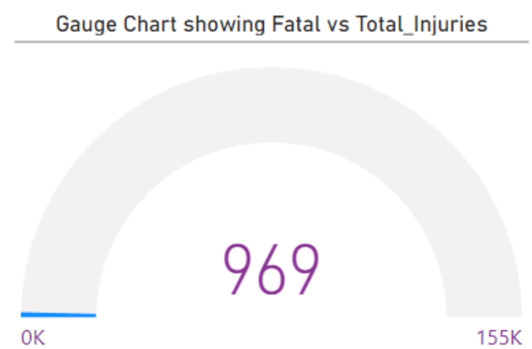

Fig. 19. Gauge Chart showing Fatal Count vs Injuries

In our specific application, the gauge chart serves a distinct purpose by illustrating the count of fatal injuries against the entire range of total injuries, which extends from 0 to 155,000. Unlike scenarios where a specific target value is established, here, the gauge chart acts as a comparative tool rather than an indicator of goals. By visually juxtaposing the fatal count of 969 within the broader spectrum of total injuries, stakeholders gain valuable insights into the gravity of road safety outcomes and the proportion of fatalities relative to the overall injury count. This approach provides a nuanced understanding, emphasizing the significant gap between fatal and total injuries and highlighting the urgency of interventions aimed at reducing fatalities and improving road safety measures.

In summary, the gauge chart plays a crucial role in conveying complex data in a clear and succinct manner, enabling stakeholders to make informed decisions and take meaningful action towards enhancing road safety outcomes. Its ability to present data within a visual framework facilitates deeper insights and more effective strategies for addressing road safety challenges and ultimately saving lives on our roadways.

C. Decomposition Tree showing the Primary and Secondary Contributions

The Decomposition tree feature within Power BI presents a sophisticated analytical approach, offering an in-depth and hierarchical perspective on the factors influencing specific outcomes or metrics. Unlike conventional charts, the Decomposition tree allows for a nuanced examination of both primary and secondary drivers contributing to a given result, providing stakeholders with valuable insights into the multifaceted nature of the analyzed data.

In our utilization of the Decomposition tree, the focus is directed towards investigating the primary and secondary factors contributing to fatal injuries resulting from road crashes. By categorizing these factors into primary and secondary contributions, the tree facilitates a granular understanding of the main causes of fatal injuries, as well as the additional factors exacerbating their severity.

For instance, among the identified primary contributions such as failure to reduce speed to avoid a crash, improper lane usage, failure to yield right-of-way, and weather conditions, stakeholders can delve deeper to explore specific instances and contexts associated with each factor. Similarly, the secondary contributions including behaviors like operating the vehicle erratically or recklessly, and issues related to road construction or maintenance, provide further context and insights into the complexities surrounding fatal injuries.

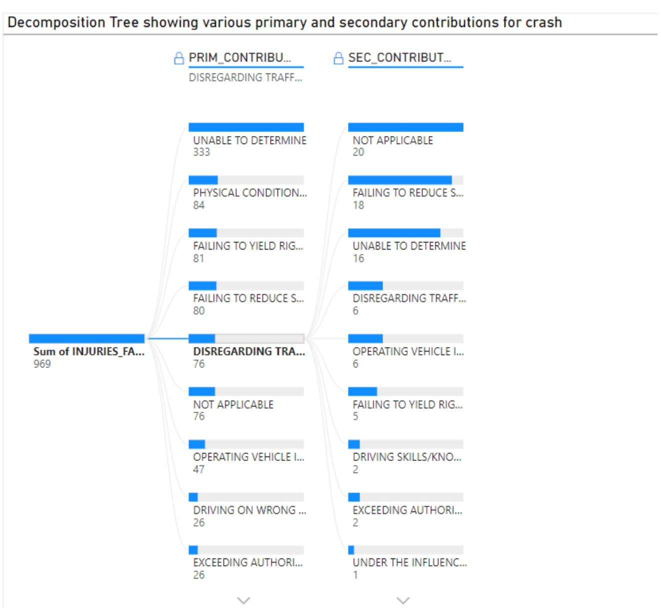

Fig. 20. Decomposition Tree

D. Ribbon, Line clustered column chart showing Fatal Injuries vs Lighting Condition

The Ribbon chart offers a dynamic visualization of trends over time for various categories. In this representation, each category is visualized as a ribbon, with the thickness of the ribbon reflecting the magnitude of the category's value at specific points in time. The overall shape of each ribbon provides an intuitive depiction of the trend for that category, allowing viewers to easily discern patterns and fluctuations over time.

On the other hand, the Line clustered column chart combines multiple line and column charts onto the same axes, enabling the comparison of different categories or data series over time. Each line or column within the chart represents a single category, with the y-axis value plotted against the x-axis representing time. This chart format facilitates a clear comparison of the values of each category at specific time points, aiding in the identification of trends and anomalies.

In our specific dataset, the highest number of fatal injuries occurred under conditions categorized as darkness, particularly on lighted roads, with a total count of 496. Following closely, daylight conditions recorded the second-highest number of fatal injuries, totaling 356. In contrast, darkness, dusk, and dawn contributed to comparatively lower counts of fatal injuries, with 64, 31, and 16 fatalities respectively. This insightful analysis highlights the varying impact of different lighting conditions on the severity of road accidents, emphasizing the importance of addressing factors such as visibility and safety measures to mitigate the risk of fatalities.

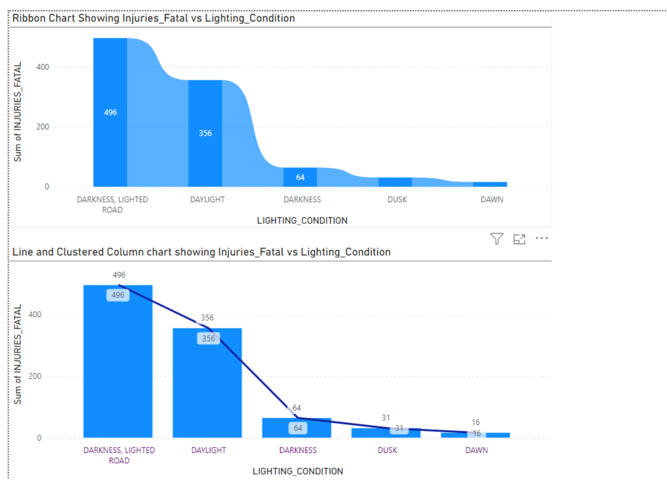

Fig. 21. Ribbon, Line clustered column chart

XII. Conclusion

Using modern data visualization techniques, the project "Charting the Route to Safety: Visualizing Chicago's Road Crash Data" has successfully produced a thorough analysis of road crashes in Chicago. The project has produced detailed visualizations, such as line graphs, tree maps, area maps, packed bubbles, and map views, that highlight important patterns and trends in crash frequency, severity, and contributing factors by integrating and cleaning a variety of datasets. The understanding of how different elements like time, location, weather, and lighting conditions affect road safety has been greatly improved by these visual aids. This project's interactive dashboard gives stakeholders instant access to vital crash data, enabling prompt decision-making and preventative safety measures.

The project's insights are critical for shaping data-driven policy choices, traffic management plans, and urban planning that try to lower the number of traffic accidents. Through the identification of high-risk locations, times, and circumstances, the project offers practical suggestions for targeted actions aimed at increasing road safety. The project's future scope, which will incorporate real-time data, predictive analytics, and more data sources, should improve safety precautions and further our understanding of crash dynamics. This project is a major step forward in the use of data visualization for road safety, opening the door to more effective solutions and making the urban environment safer for all users of the roads. This project's continued development and growth will be essential in lowering the frequency and intensity of road crashes.

XIII. FUTURE SCOPE

A. *Integration of Real-Time Data*

Implement real-time crash data to give current insights and improve dashboard flexibility.

Provide alert systems so that stakeholders are informed right away of notable incidents or trends.

B. *Growth in Data Sources*

Add more data sources, such as reports on traffic conditions from social media, GPS data from cars, and feeds from traffic cameras.

Combine road maintenance schedules with weather forecasts to give a complete picture of all the factors affecting road safety.

C. *Better Analytical Skills*

Make use of statistical models to forecast crash times and high-risk locations.

To predict the effects of suggested safety measures or alterations to traffic laws, apply statistical analysis.

D. *Improved Map reading*

To find micro-level hotspots, use Geographic Information System (GIS) tools to perform more detailed spatial analysis.

Check the effects of particular road features on the number and severity of crashes, such as intersections and pedestrian crossings.

E. *Expansion to Other Regions*

Expand the project to include more cities or areas, customizing the dashboard to address unique traffic patterns and public safety issues.

Provide a tool for comparative analysis so that various areas' road safety metrics can be compared.

F. *Analysis of Environmental Impact and responsibility*

Analyze the effects of crashes on the environment, including any fuel spills and emissions brought on by traffic.

Create plans to reduce environmental risks connected to traffic accidents.

G. *Longitudinal Studies*

To evaluate the long-term efficacy of safety measures that have been put in place, conduct longitudinal research.

Examine patterns over long time spans to comprehend the changing dynamics of Chicago traffic safety.

H. *Community Engagement and Feedback*

Provide networks via which community members can use the dashboard to report risks or make enhancement suggestions.

Include feedback loops so that the dashboard can be improved over time in response to user experiences and recommendations.

I. *Creation of Mobile Apps*

Develop mobile applications to give drivers real-time safety updates and alerts.

Provide tools that allow users to report problems or dangers straight from their mobile devices.

J. *Incorporation of Driver Behavior Analysis*

Combine information from driver behavior monitoring systems and mobile technology to examine variables such as speeding, hard braking, and distraction.

Provide an understanding of how driving habits affect the risk of crashes and suggest focused education efforts.

REFERENCES

[1] Shahi, Sachita. (2022). SPATIAL ANALYSIS OF ROAD TRAFFIC CRASHES AND ASSESSMENT OF ROAD SAFETY ISSUES FROM ROAD USERS' PERSPECTIVE: A CASE STUDY OF ROTTERDAM AND THE SURROUNDING URBAN CONURBATION. 10.13140/RG.2.2.18314.54721.
[2] Bačkalić, Svetlana. (2013). Temporal analysis of the traffic accidents occurrence in province of Vojvodina. Transport Problems. 8. 87-93.
[3] Naik, Bhaven Tung, Li-Wei Zhao, Shanshan Khattak, Aemal. (2016). Weather impacts on single-vehicle truck crash injury severity. Journal of Safety Research. 58. 10.1016/j.jsr.2016.06.005.
[4] Hale, A.R. Guldenmund, Frank Loenhout, P.L.C.H. Oh, J.. (2010). Evaluating safety management and culture interventions to improve safety: Effective intervention strategies. Safety Science. 48. 1026-1035. 10.1016/j.ssci.2009.05.006.
[5] https://www.youtube.com/watch?v=oAIubTqg-Kw
[6] https://www.youtube.com/watch?v=-n5JKmnSpbA
[7] https://www.youtube.com/watch?v=4Sx3VQg7LgI
[8] https://www.youtube.com/watch?v=0iDiPxACkWQ