Assignment 3: Solution Design

Safe LA: Enhancing Traffic Safety Through Data-Driven Insights

Team 8

Alexandra Gladkova

Christina Saju

Hao Lun Rong

Kushwanth Sai Kolli

Ladan Asempour

Capstone Project

BIA-5450-0LA

Samer Al-Obaidi

Group Member Contribution

Topic	Contributor
Process Data diagram	Christina Saju
Existing IT Architecture	Ladan Asempour
Solution’s design	Hao Lun Rong – Kushwanth Sai Kolli
The fit of the new solution into the existing IT architecture	Hao Lun Rong
Challenges and future enhancement opportunities	Alexandra Gladkova

Process Data diagram

Project Proposal: Enhance public safety in LA by analyzing traffic collision data. Generate insights supplemented with visualizations for use by the primary stakeholders.
Deliverables: Comprehensive Reports, Data Visualizations, Actionable Insights

↓

Data Acquisition: LA Traffic Collision Dataset collected from data.gov features information regarding date, time, location, crime codes, and victims’ characteristics

↓

Data Cleaning and Preprocessing:
- Handling missing values
- Identifying key features for analysis
- Transformation/ Creation of features for geographical analysis
- Transformation of MO Codes into readable values for analysis

↓

Data Exploration:
- Checking for correlations between the features
- Checking frequency of occurrences geographically and demographically
- Solution design for all proposed questions

↓

Reporting:
- Create visualizations (plots, tables, heatmaps, etc.) that show the trends/ patterns of the collisions.
- Create a dashboard with key visualizations to present to stakeholders

↓

Solutions:
- Present findings of the analysis and recommendations based on the same to stakeholders, enabling them to make data driven decisions to enhance traffic safety management.

Existing IT Architecture

The IT architecture for analyzing traffic collisions in Los Angeles includes key components such as data sources, software tools, and integration mechanisms to compile the components into a structured dataset.

Major Components:

- **Data Sources:** The Traffic Collision Data from 2010 to Present from Data.gov (https://catalog.data.gov/dataset/traffic-collision-data-from-2010-to-present) , contains details regarding location, time, involved parties, and contributing factors. The official data dictionary (https://data.lacity.org/Public-Safety/Traffic-Collision-Data-from-2010-to-Present/d5tf-ez2w/about_data) and MO Codes document (https://data.lacity.org/) are available to interpret and classify data.

- **Software Tools:** Python has been utilized for data exploration, processing, cleaning, and statistical and visual analysis to identify trends in collision data.

Integration and Ecosystem:

The components all work in together to analyze collision data. The data is downloaded manually from Data.gov, cleaned, and processed using Python. Insights are derived from the data to inform traffic safety decisions. APIs are not utilized in the data ingestion and integration process but integrates data through manual downloads and processing.

Connections:

- **Data Flow:** Data is downloaded from Data.gov and processed using Python.
- **Security and Privacy:** Data cleaning techniques and privacy guidelines ensure the protection of sensitive information.

This IT setup enables data-driven decision-making for City of Los Angeles Department of Transportation (LADOT) and Law Enforcement (LAPD) and Traffic Monitoring Agencies, enhancing road safety and urban mobility in Los Angeles.

Solution's design

The solution design for this project consists of developing visualizations and dashboards to answer key components of our analysis. This includes the key analytical questions that were proposed earlier which include:

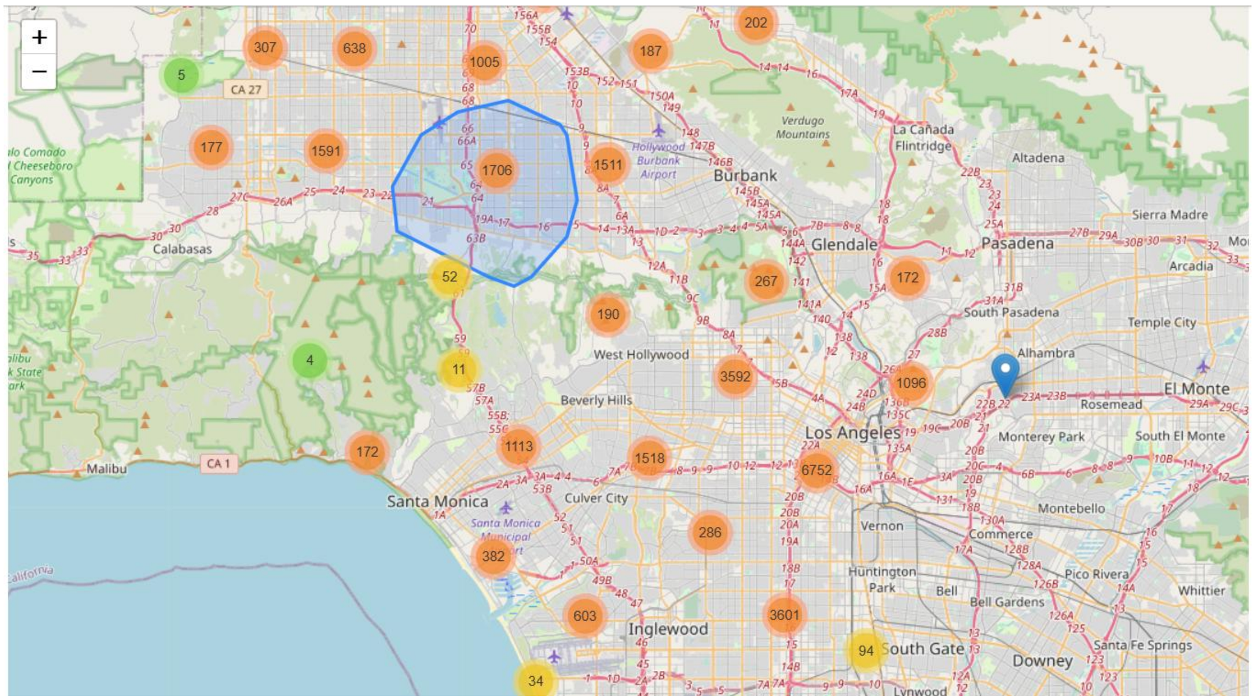
- *Which locations have the highest density of collisions?*
- *What periods of the day and days of the week have the most collisions?*
- *Are certain demographics more frequently involved in traffic collisions?*
- *What types of traffic collisions (e.g., vehicle vs. vehicle, vehicle vs. pedestrian, vehicle vs. property) occur most frequently?*

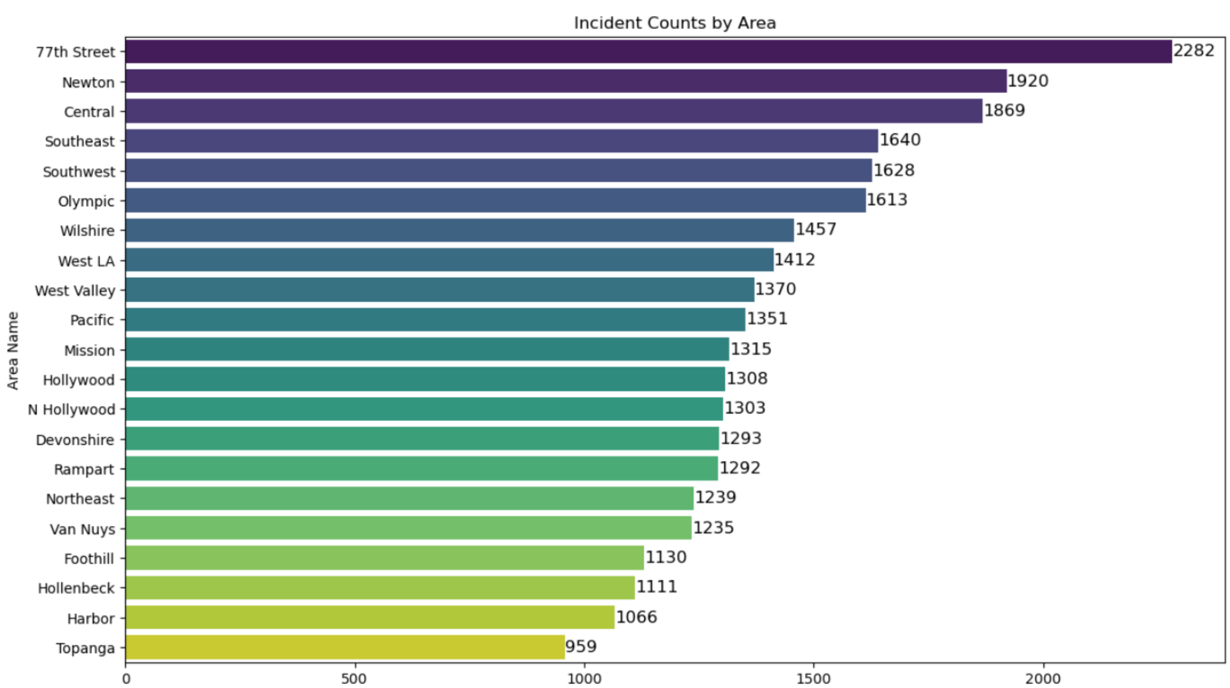
By answering these questions, patterns, trends, and other variables that contribute to the reported accidents are hoped to be uncovered, which will allow for stakeholders to have better access and understanding to make more informed decisions.

Solution 1. Identifying areas with high collision density

To analyze collision density, both a marker cluster map and a bar chart will be created to highlight streets with high collision frequency. The marker cluster map utilizes latitude and longitude data to display individual collision points, grouping them into clusters based on proximity. Larger clusters with higher numbers indicate areas with more frequent collisions. To supplement this, a bar chart will also be created, showing a breakdown of streets and intersections with the highest record of collisions. This will provide further insight into which intersections are at higher risk, aligning with the higher density indicators of the marker cluster map. Together, these visuals will allow high-risk areas to be pinpointed, supported by specific areas of risk in each of them.

Visualizations:





Insights:

From the image, it is observed that Downtown Los Angeles, North Hollywood, and Burbank exhibit high collision densities, with large orange and red clusters representing thousands of incidents. Meanwhile, areas such as Calabasas and Malibu show significantly fewer collisions, as indicated by smaller green clusters. This spatial visualization helps pinpoint high-risk locations for targeted safety improvements.

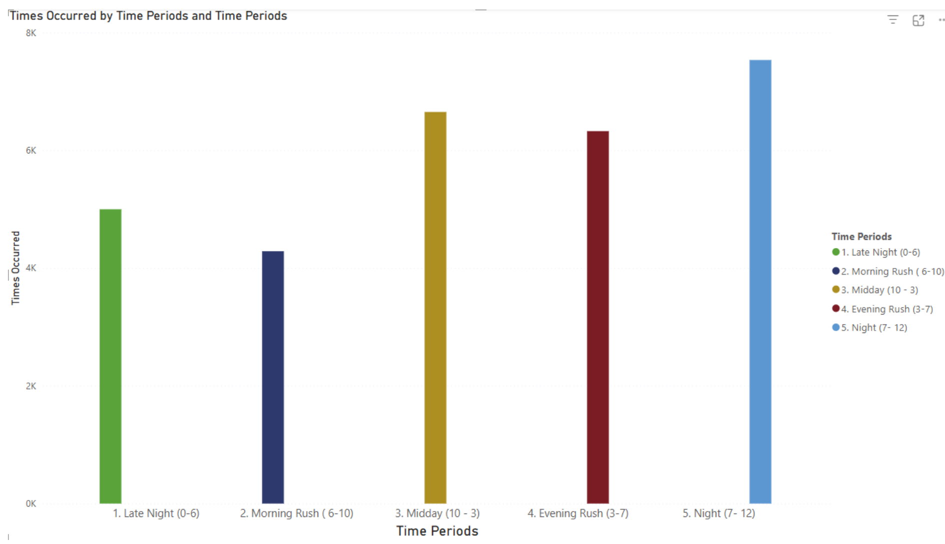
The bar chart illustrates incident counts across various areas, with 77th Street reporting the highest number of incidents at 2,282, followed by Newton (1,920) and Central (1,869). These areas stand out as potential hotspots requiring further investigation into crime patterns or public safety measures. In the mid-range, areas like Wilshire (1,457), West LA (1,412), and West Valley (1,370) show moderately high incident counts, suggesting that while they are not the most critical zones, they still warrant attention. Towards the lower end, Topanga (959), Harbor (1,066), and Hollenbeck (1,111) record the fewest incidents, indicating relatively safer regions or possibly areas with less activity. The overall trend shows a gradual decrease in incidents from the highest to the lowest areas. These insights could guide law enforcement and policymakers in prioritizing resource allocation, implementing targeted safety measures, or analyzing underlying factors contributing to the varying incident rates.

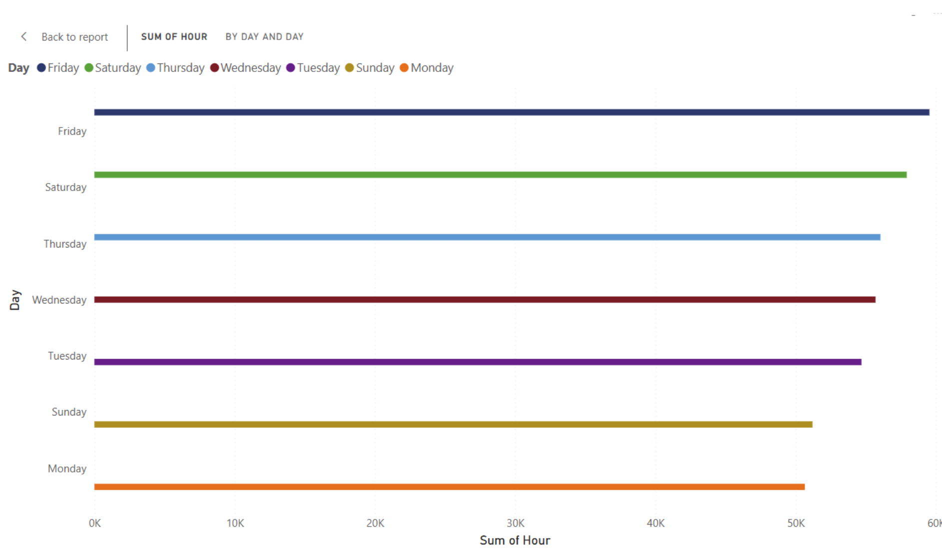
Solution 2. What time of day and days of the week have the most collisions?

To analyze when the collisions occur throughout the week, bar charts are created to show the frequency distribution of time-based visual trends. These charts will include the average total collisions based on day of the week. Information on whether the collisions are more frequent during the workweek or during weekends will also be extracted.

To supplement this, times of the day will also be analysed with specific categories included (Rush hours, Afternoon, Evening, Night). This will help determine the peak periods where most incidents occur.

Visualisations:





Insights:

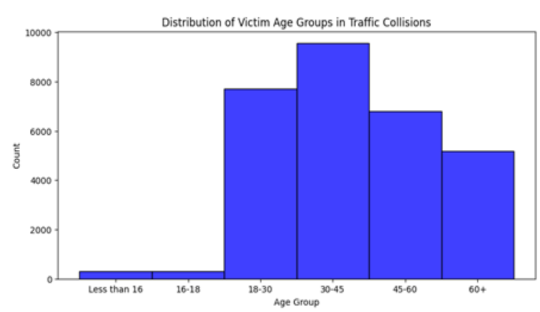
These visualizations show patterns in the behaviour of drivers in LA. Contrary to popular belief that morning and evening rush hours were the times where collisions would most likely occur, it appears that night-time commutes are the ones where collisions are most frequently reported.

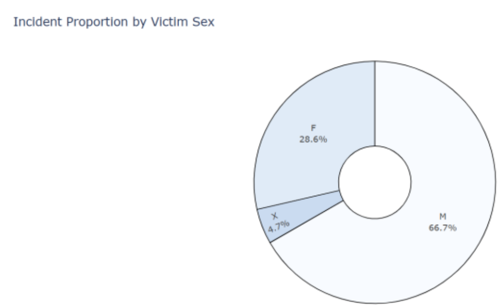
For the other table, daily collisions fluctuate, with Monday and Sunday showing significantly fewer collisions reported daily compared to the more eventful days such as Friday and Saturday.

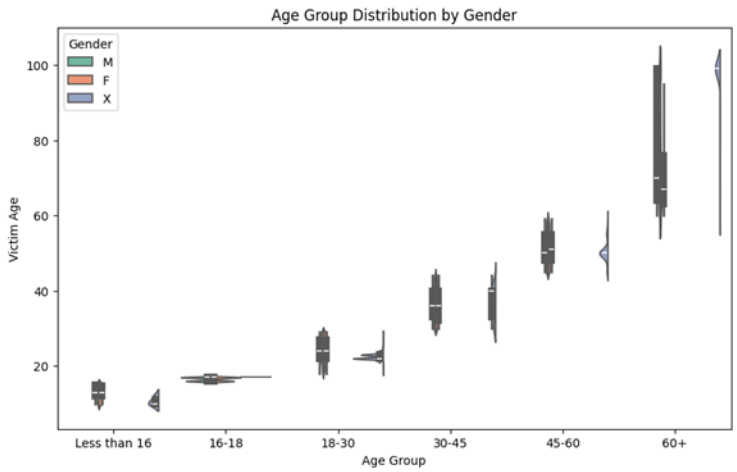
Solution 3. Are certain demographics more frequently involved in traffic collisions?

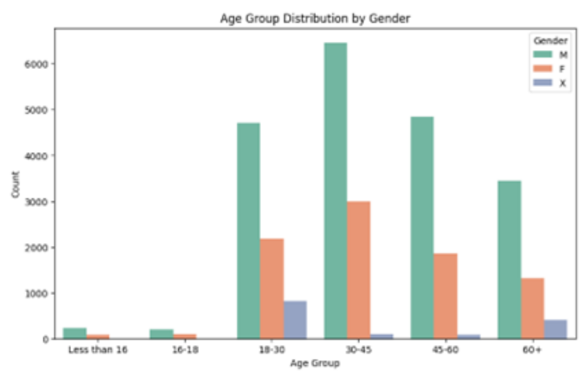
To analyze whether certain demographics are more frequently involved in traffic collisions, a series of visualizations will be created to break down categories of gender, ethnicity, and age groups to identify if there is a higher likelihood or risk of certain demographics being involved in accidents.

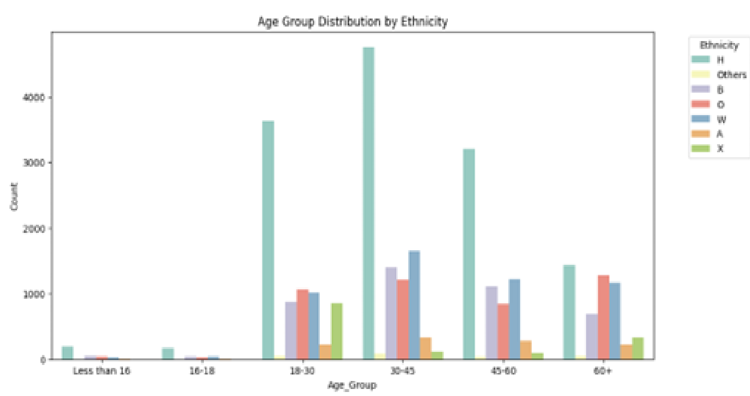
Visualizations:

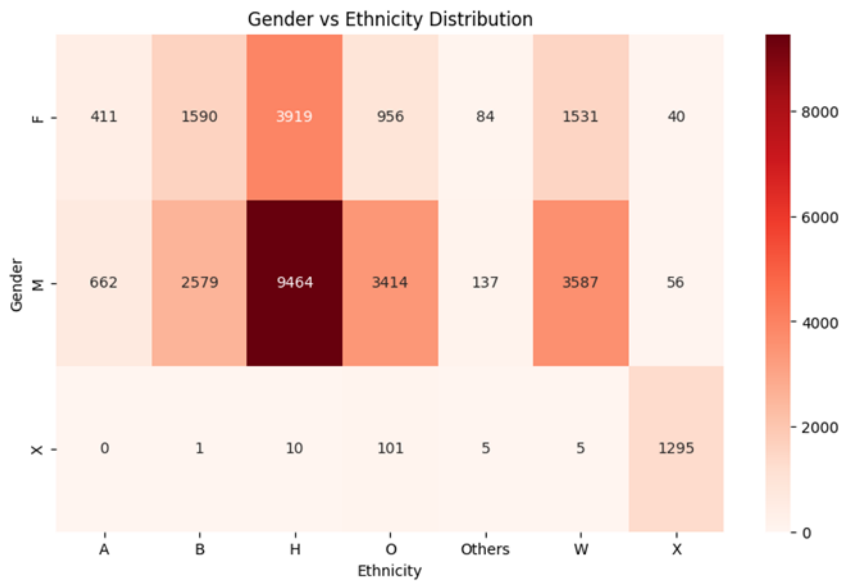





Age Group Distribution by Gender
Gender
M
F
X
Victim Age
100
80
60
40
20
Less than 16
16-18
18-30
30-45
45-60
60+
Age Group


Age Group Distribution by Gender
Gender
M
F
X
Count
6000
5000
4000
3000
2000
1000
0
Less than 16
16-18
18-30
30-45
45-60
60+
Age Group


Age Group Distribution by Ethnicity
Ethnicity
H
Others
B
O
W
A
X
Count
4000
3000
2000
1000
0
Less than 16
16-18
18-30
30-45
45-60
60+
Age_Group



Insights:

These visualizations show the key demographic patterns of traffic collision victims. Young people, particularly those aged 18–45, are the most affected, and this points to the need for targeted safety measures for pedestrians and drivers in these age groups. Although the 60+ population forms a smaller proportion of victims, they still account for a significant number of cases, and this necessitates safety initiatives for this age group. Gender disparities can also be observed, with males tending to get involved in accidents more frequently than women, especially among the age group 16–30. This calls for safety campaigns targeting young male drivers specifically.

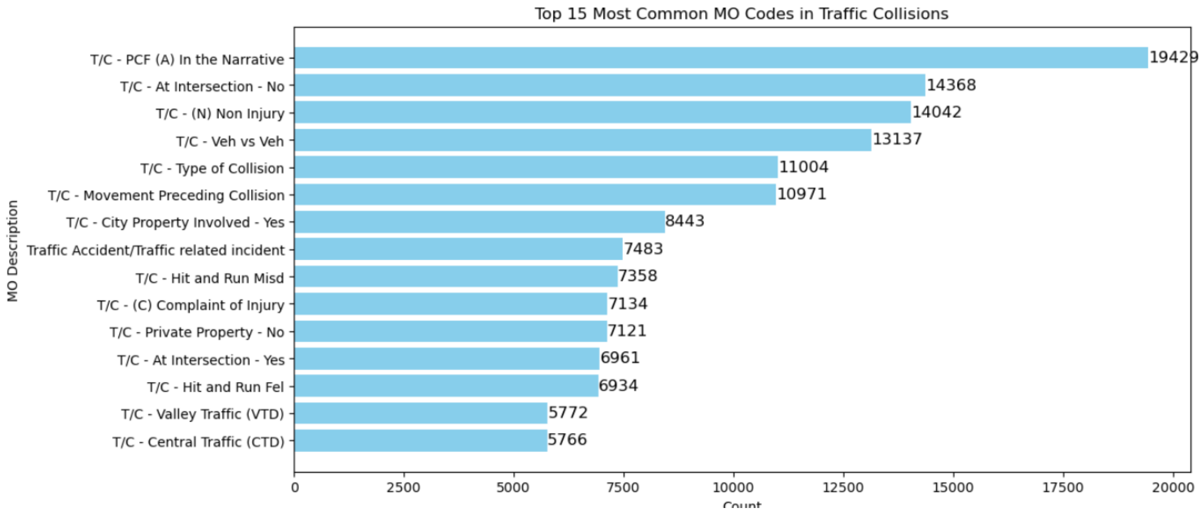
Ethnicity is also highly important, with the White and Hispanic population showing a high rate of traffic collision victims, i.e., 44.9% for the Hispanic population. In contrast, Asian and Black populations show lower accident involvement. The findings indicate the necessity of culturally appropriate traffic safety measures for the most affected ethnic groups. In addition, different age-specific risks exist within ethnic groups, such as higher numbers of 30–45-year-old victims in Hispanic and White groups, and large numbers of young age groups in other ethnic groups. These results suggest interventions based on age and ethnicity to address different risk factors.

Recommendations for Stakeholders:

1. Launch targeted campaigns to educate young male drivers, particularly those aged 16–30 years, on road safety risks and prevention.
2. Implement culturally specific outreach programs for Hispanic and White populations that address their particular risk factors.
3. Create culturally suitable educational content to achieve optimal effectiveness for traffic safety initiatives.

Solution 4 - What types of traffic collisions (e.g., vehicle vs. vehicle, vehicle vs. pedestrian, vehicle vs. property) occur most frequently?

In this analysis, the top 15 most common MO codes associated with traffic collisions are being visualized. MO codes categorize different characteristics of incidents, such as whether they occurred at intersections, involved injuries, were hit-and-run cases, or impacted public/private property. The horizontal bar chart helps understand the frequency of various types of incidents by displaying the count of occurrences for each MO description.



Insights:

The most frequent MO code is “T/C - PCF (A) In the Narrative” (19,429), suggesting that many incidents have an identified Primary Collision Factor contributing to the crash. A large number of collisions occur away from intersections (14,368 cases) and are classified as non-injury accidents (14,042 cases), indicating that many incidents may be lower severity. Additionally, vehicle vs. vehicle collisions appear to be very common, as seen in "Veh vs Veh" (13,137 cases). Hit-and-run cases—both misdemeanor (7,358) and felony (6,934)—also rank high, highlighting the need for better enforcement or public awareness. These insights can help traffic authorities prioritize safety interventions in high-risk scenarios.

The fit of the new solution into the existing IT architecture

The proposed solutions seamlessly compliment the IT architecture which is currently being used. Taking advantage of the fully explored and engineered dataset and giving insights and meaning through not only being able to see the numbers through tables, but various types of visualizations to understand not only, location, time, identification, as well as risk factors police have documented associated with these collisions.

By transforming the processed data which has been a part of our framework into meaningful insights, the solution provides a comprehensive way for stakeholders to assess road safety concerns in LA.

Challenges and future enhancement opportunities

There are several challenges which were faced as well as some which may be relevant in the future. The first among them is the data quality of the reported collisions. Throughout the data exploration process, there was a lot of valuable information such as location details or demographics which were not reported. This could have been due to human error or lack of thorough reporting by the police records.

Another challenge faced is the sheer number of variables and volume of data, within just the last two years there are more than tens of thousands of reported incidents, with numerous variables which may have contributed to the collision, this includes time of day, driver demographics, and the MO codes which were given, however it could also be attributed to data which were not provided in the original data, such as road conditions or environmental factors affecting the driving scenarios that day. Because of this, identifying meaningful trends while avoiding misleading conclusions can prove to be a significant challenge.

Future Enhancement Opportunities

Moving forward with the project, future enhancements could very well improve the overall analysis. For the current solution, gathering insights about the location data could prove a beneficial part of the analysis, understanding each neighbourhood unique to LA could contribute to a greater understanding of certain collisions hotspots.

In terms of the technological aspects, the analysis could be further enhanced by further updates to the database, including real time integration of the information as well as proposing that the LAPD expands their police reports to include several other factors as previously mentioned such as weather or road conditions during the incident.