Accident Metadata for Enhanced Analysis

Team Accident Analysis Avengers (AAA)

Our Team

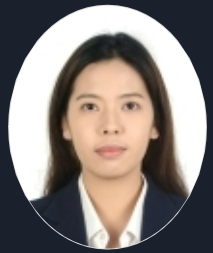
Ansh D.

Balaji U.

Ei A.

Harshit K.

Joshua M.



Lirong H.

Sanskriti Y.

Introduction

- According to the Centers for Disease Control and Prevention (CDC), road traffic crashes are a leading cause of death in the United States for people ages 1-54.
- In San Francisco, there are over 30,000 collisions annually or over 80 per day.
- These collisions have claimed nearly 4,000 lives in the last 18 years.

Costs to Society:

- Medical costs of $4.6 Billion
- Property damage of $2.8 Billion
- Lost productivity impacting economy

Project Overview

- We will analyze 2 datasets that contain data from all crashes resulting in injury in San Francisco, California from 2005 till 2023.

- Our Mission:
  - We aim to analyze the data in order to:
    - Provide valuable information related to public safety
    - Create predictive models to reduce the risk of collisions
    - Ultimately, help the U.S. government make more informed decisions.

- Desired Outcomes:
  - Reduce loss and risk of life due to avoidable collisions
  - Minimize health and economic impact on society
  - Guide infrastructure improvements and public policy decisions

Dataset Overview

Traffic Crashes Dataset (2005-2023)

- 56,010 incidents records
- 3.3 million data points
- 60+ variables including:
  - Unique crash IDs
  - Location (latitude, longitude, intersection etc.)
  - Date and time
  - Collision severity
  - Weather, lighting and road conditions

Parties Involved Dataset

- 118,087 detailed victim records
- 9.4 million data points
- 80+ variables covering:
  - Vehicle makes, models and manufacturing year
  - Safety equipment in use during collision
  - Injuries sustained
  - Actions prior to collision (text descriptions)

Connection between Datasets

- Correspond fatality and injury outcomes to vehicle types
- Relate detailed party behaviors to incident locations and conditions

Questions of Interest

In this project, we aim to answer the following 3 main questions:

1. Are there any geographic hotspots or any other patterns in collisions?
2. How do vehicle brands, their safety equipment, and their year of manufacturing affect collisions?
3. What is the best time of the year for the San Francisco Government to do construction so that the collision occurring on a particular road is of least severity?

Data Cleansing

Key Tasks:

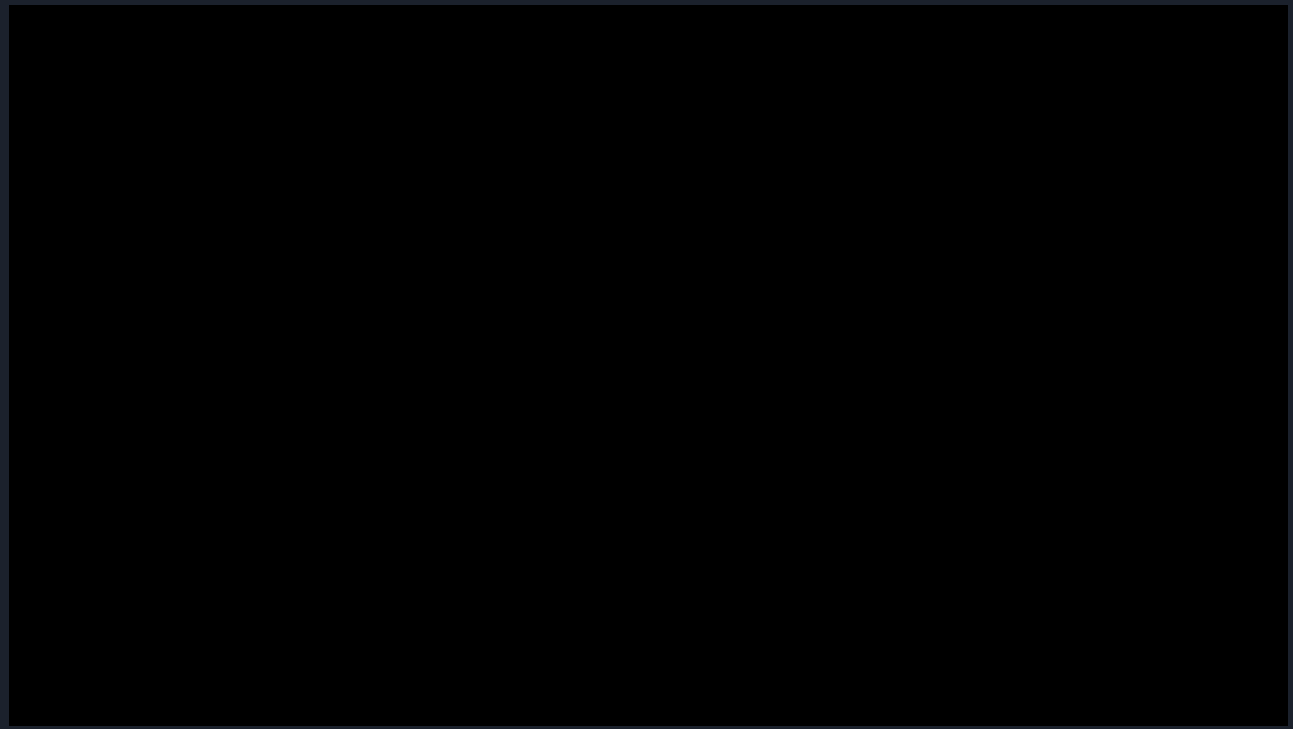
- Removing duplicates
- Dropping unnecessary columns
- Renaming columns for clarity
- Handling missing values
- Transforming DateTime column
- Ensuring consistency in categories

Impact:

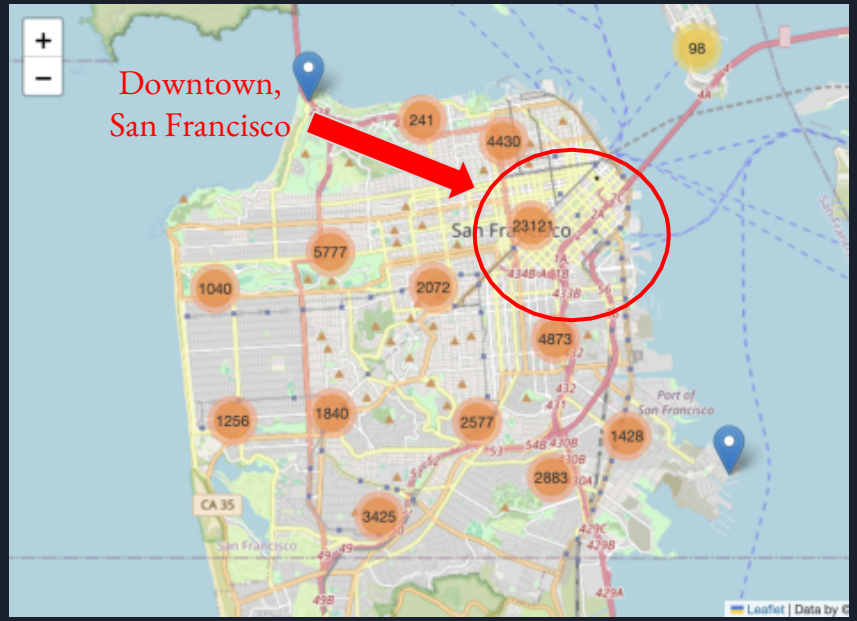
- Increased accuracy and reliability of analysis
- Supported better visualizations and modeling
- Revealed early insights into data issues
- Identified data collection gaps

Question 1:

Are there any geographic hotspots or any other patterns in collisions?



Where Do Most of the Collisions Occur?

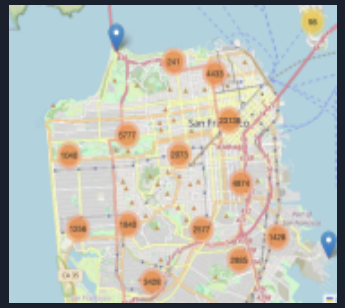


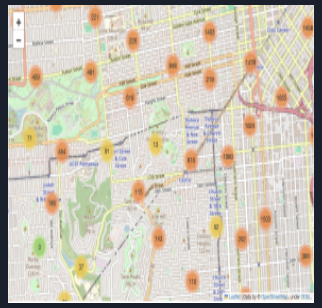
Q1 Analysis

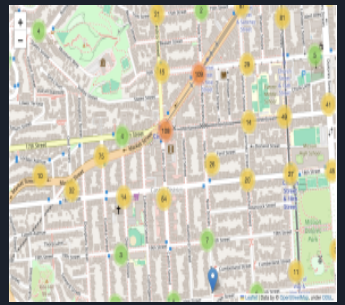
Geographic Patterns in San Francisco Traffic Collisions

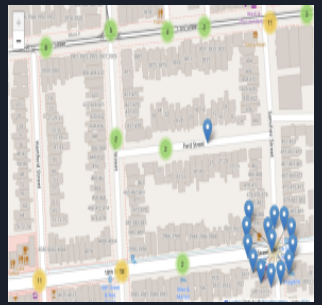
Key Observations:

- Clustering of collisions in downtown/central areas
- Potentially due to:
  - Urban density and traffic volume
  - Traffic congestion
  - Tourism and Visitor activity
  - Pedestrian involvement
  - Infrastructure challenges







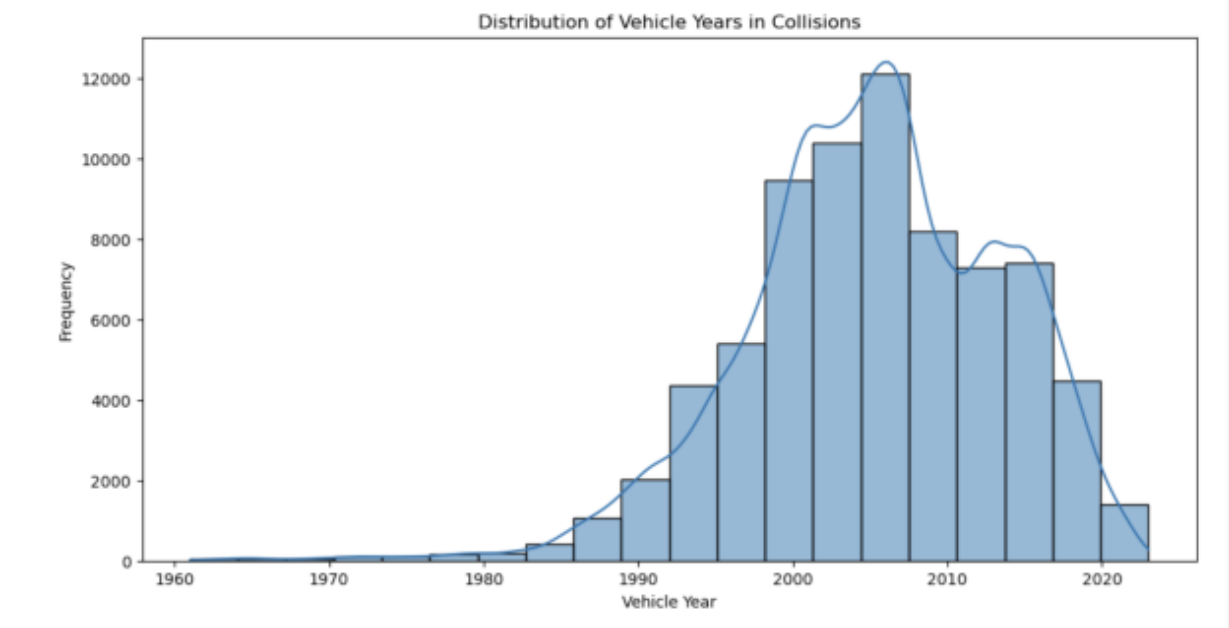


Implications:

- Underscores need for safety improvements in high risk areas
- Informs future urban planning and traffic management
- Urges consideration of congestion charges or vehicle restrictions during peak periods.

Question 2:

How do vehicle brands, their safety equipment, and their year of manufacturing affect collisions?



Q2 Analysis

How do vehicle brands, their safety equipment, and their year of manufacturing affect collisions?

Key Observations:

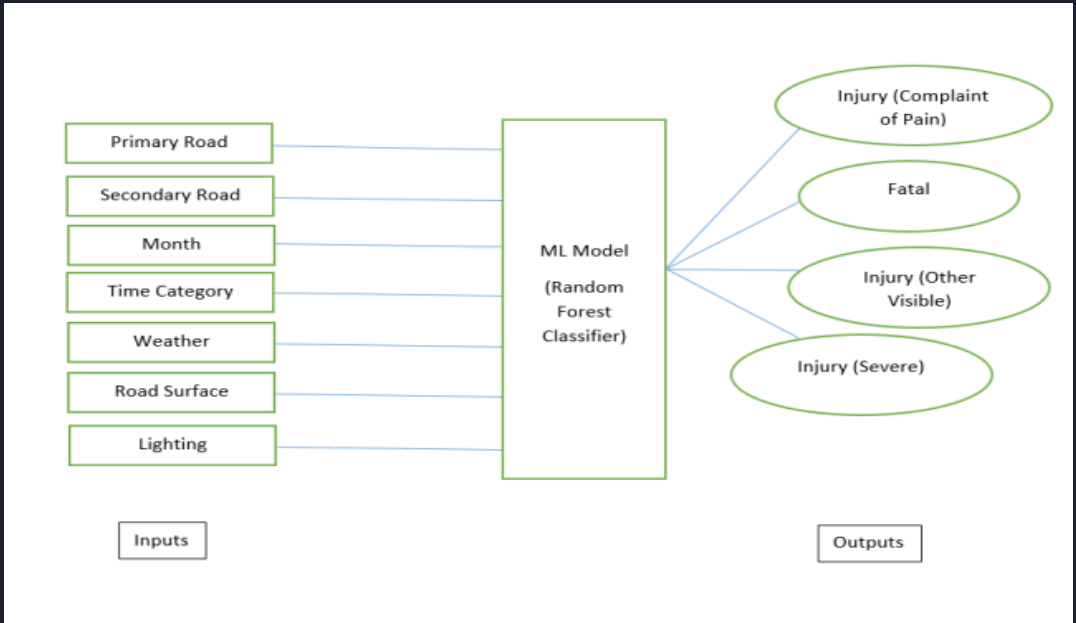
- Peak in collisions involving vehicles manufactured in the early 2000s.
- This could be due to :
  - High volume of 2000s vehicles on the road
  - Aging Vehicle Fleet
  - Lack of technological advancements
  - Poor maintenance and upkeep
  - Risky driver behaviour

Implications:

- These findings may prompt policymakers and regulatory bodies to reevaluate safety standards for vehicles
- Public awareness campaigns about the regular maintenance of 2000s vehicles.
- Insurance rates for the 2000s model could be raised due to higher risks

Question 3:

What is the best time of the year for the San Francisco Government to do construction so that the collision occurring on a particular road is of least severity?



Q3: The Model Analysis

What is the best time of the year for the San Francisco Government to do construction so that the collision occurring on a particular road is of least severity?

For the sample data = { 'Primary_Road': '***NAPLES ST***', 'Secondary_Road': '***EXCELSIOR AVE***', 'Month': '***March***', 'Time_Category': '***2:01 pm to 6:00 pm'***, 'Weather': ***'Clear'***, 'Road_Surface': '***Slippery***', 'Lighting': '***Daylight***'},

We got collision severity as Severe.

Changed the month to ***January*** and ***December*** to get collision severity as complaint of pain and other visible injury.

Conclusion

Key Findings

- Heatmap shows collisions concentrated downtown
- Vehicle brand & lack of safety equipment linked to more severe crashes
- Modeling reveals slippery roads increase collision risk
- Collisions have multifactorial causes

Recommendations

- Gov't can improve infrastructure & traffic systems to prevent collisions
- Manufacturers can enhance safety standards & tech
- Buyers can reference safety record when purchasing
- Reliance on tech (electric vehicles) limited - human behavior is critical
- Gov't can address this infrastructure issue to reduce crashes
- Collaboration needed across orgs & individuals for a safer future

Thank You!

Any Questions?