San Diego Traffic Risk Estimator

Group 11

Andrew Obermiller, Haoran Li, Joey Liu, Tongqing Shi, Zachary Zheng

Problem Definition

- Existing route recommendation methods often look at traffic data only with an eye to avoid congestion and reduce travel time

- These systems do not inform on the risk of different routes

Objectives

- Examine San Diego traffic data in order to locate underlying trends correlated to increased collision rates
- Explore the relationship between street, vehicle type, time, and collision rate
- Explore the relationship between street, vehicle type, time, and risk of injury or death due to collision
- Create a method to estimate a driver's risk level in terms of collision and casualty likelihood when driving a vehicle along a particular street or route at a time

Datasets

Traffic counts

- CSV file containing traffic counts for selected streets in San Diego
- Multiple samples (vehicle count over 24 hours) taken for different roads between 2005-2023
- Includes:
  - Street name
  - Directional and total traffic counts
  - Date of sample

Collision Reports

- CSV file containing San Diego collision report information from 2015-2023
- Includes:
  - Collision location by street
  - Violation section by CA vehicle code
  - Injury and fatality counts
  - Vehicle type
  - Vehicle make and model
  - Date of report

Police Beats

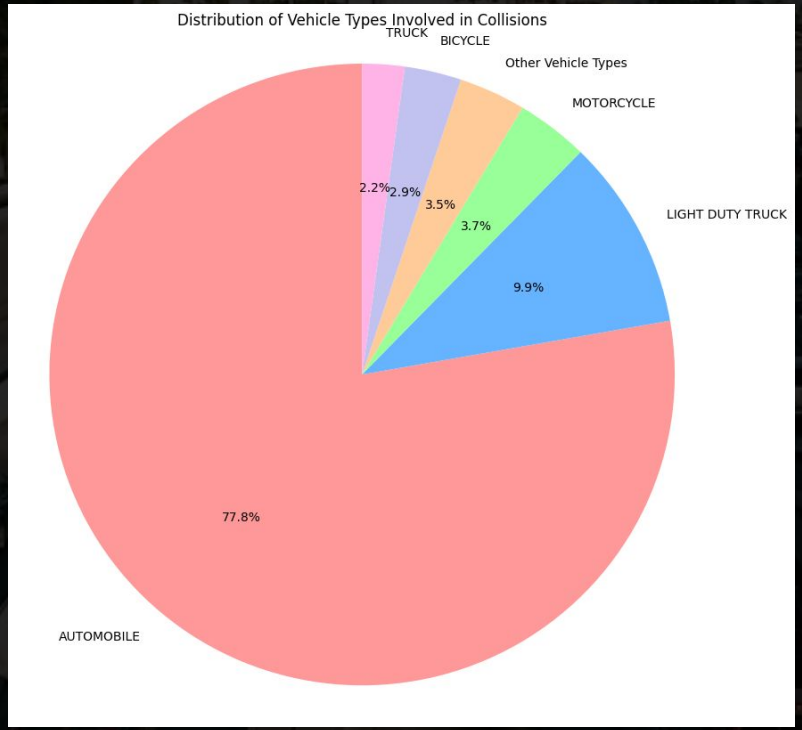
- CSV file relating police beat codes to neighborhood name

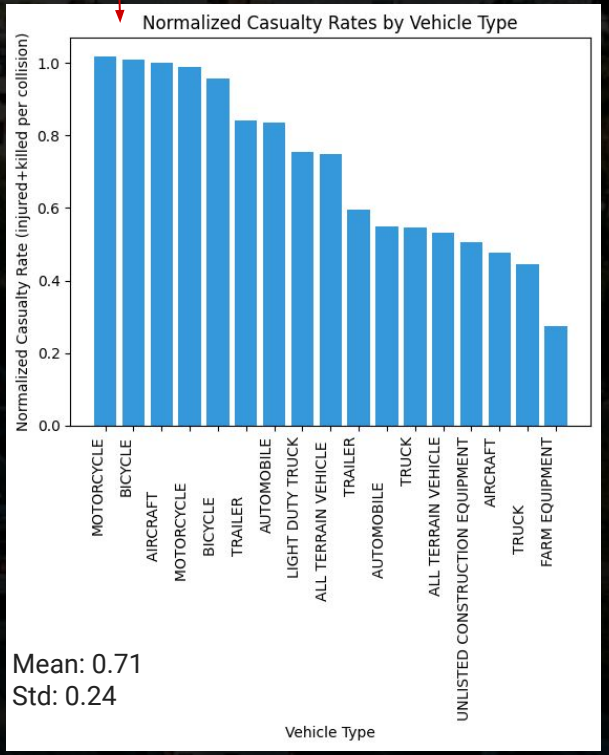
Data Preprocessing

- Convert all datasets to pandas dataframe
- Expand all street names using standard abbreviations
  - Ex: st, str, strt -> street
- Remove all streets not contained in traffic rates and collision reports datasets
- Average traffic counts across all samples for each street
- Eliminate unneeded columns (report ID, reporter role in collision report, etc.)

Impact of Vehicle Type

Collisions involving motorcycles or bicycles are the most dangerous from a casualty perspective, despite making being involved in a small fraction of total collisions

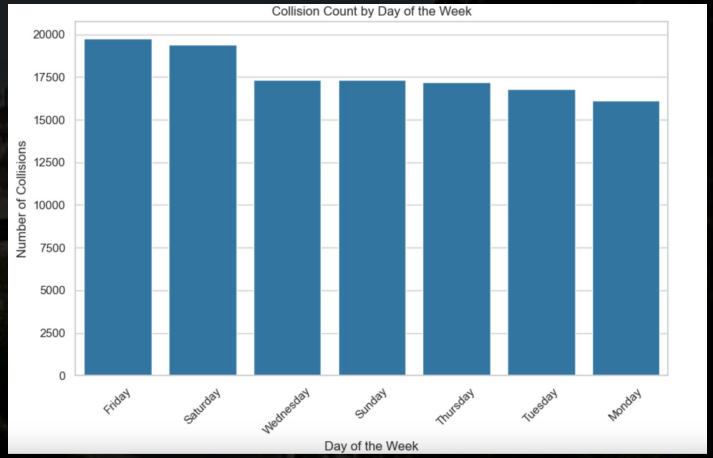


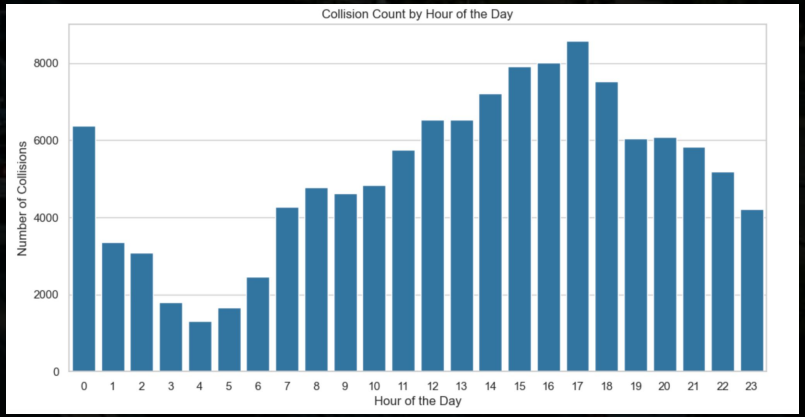


Mean: 0.71
Std: 0.24

Impact of Date/Time

- No significant correlation between collision rate and time of the year
- Most number of collisions around 5pm and another peak around 12am
- Most number of collisions on Fridays and Saturdays.

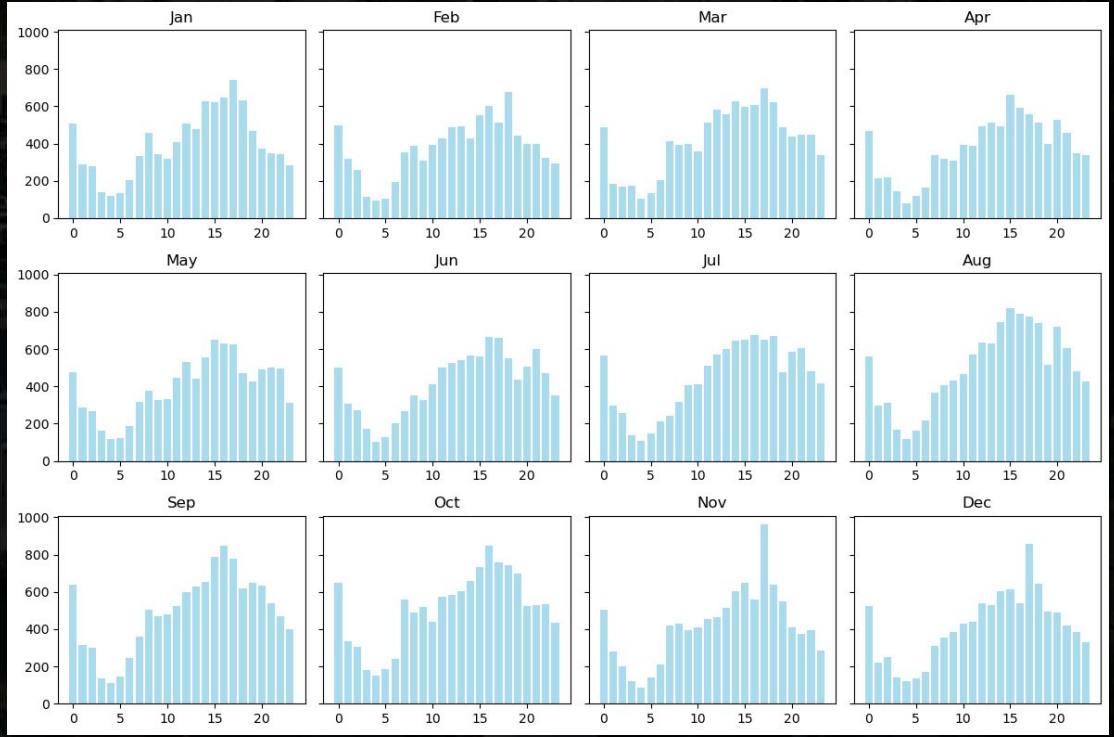




Impact of Date/Time

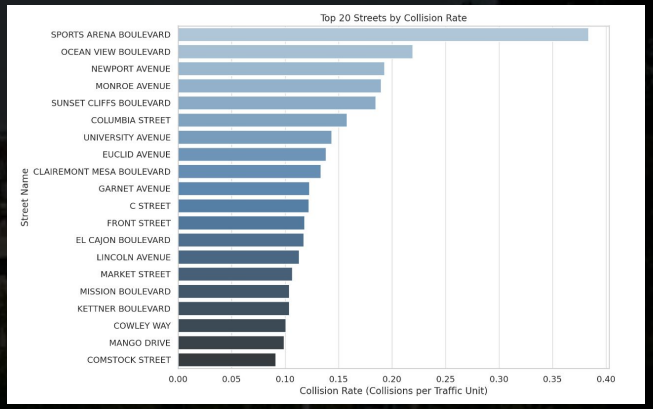
- Collision within a day based on different month
- Roughly submit to Gaussian Distribution
- Number of collisions decrease during spring and the start of the summer

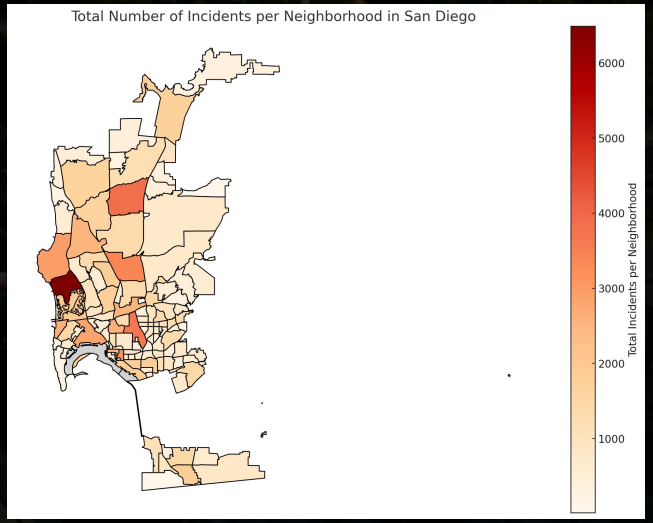
Monthly Distribution of Collision Counts over the 24 Hours of Each Day



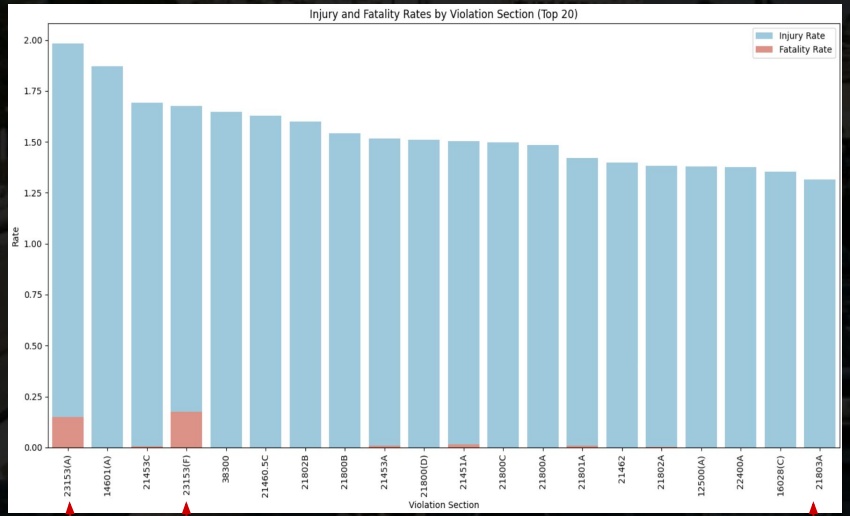
Impact of Location

- Location a significant risk factor
- Collison/injury/fatality rates varies throughout neighborhoods
  - Pacific Beach/North Park
- Collision rates among streets varies significantly
  - Some streets 10x or 20x more likely to see collisions





Visualization of other Risk Factors



DUI (Alcohol)

DUI (Drugs)

Failure to yield

- DUI violations are directly correlated to higher injury/fatality rates
- Fatality rate similar among other violations

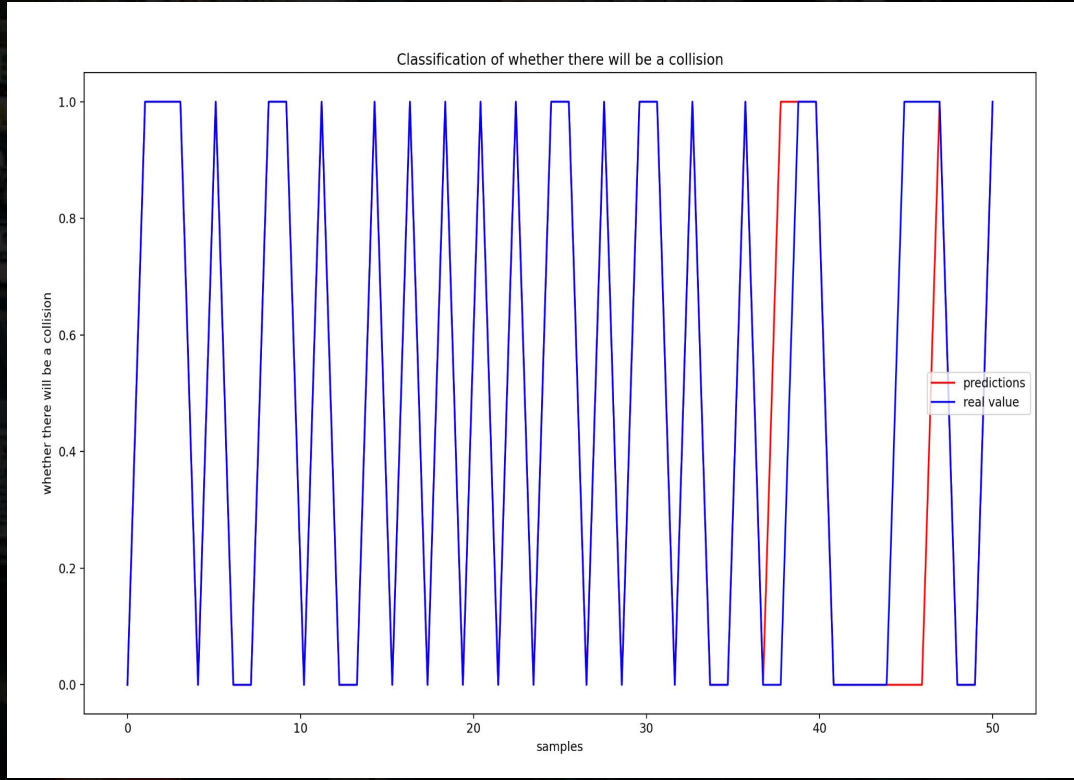
Street Risk Estimator Model 1: Classification

Input: vehicle model, road name and time.

Model 1: A classification model using Logistic Regression to predict whether there will be a collision.

Output: 0 or 1, where 0 means safe and 1 means dangerous.

The accuracy of this model is about 0.95.



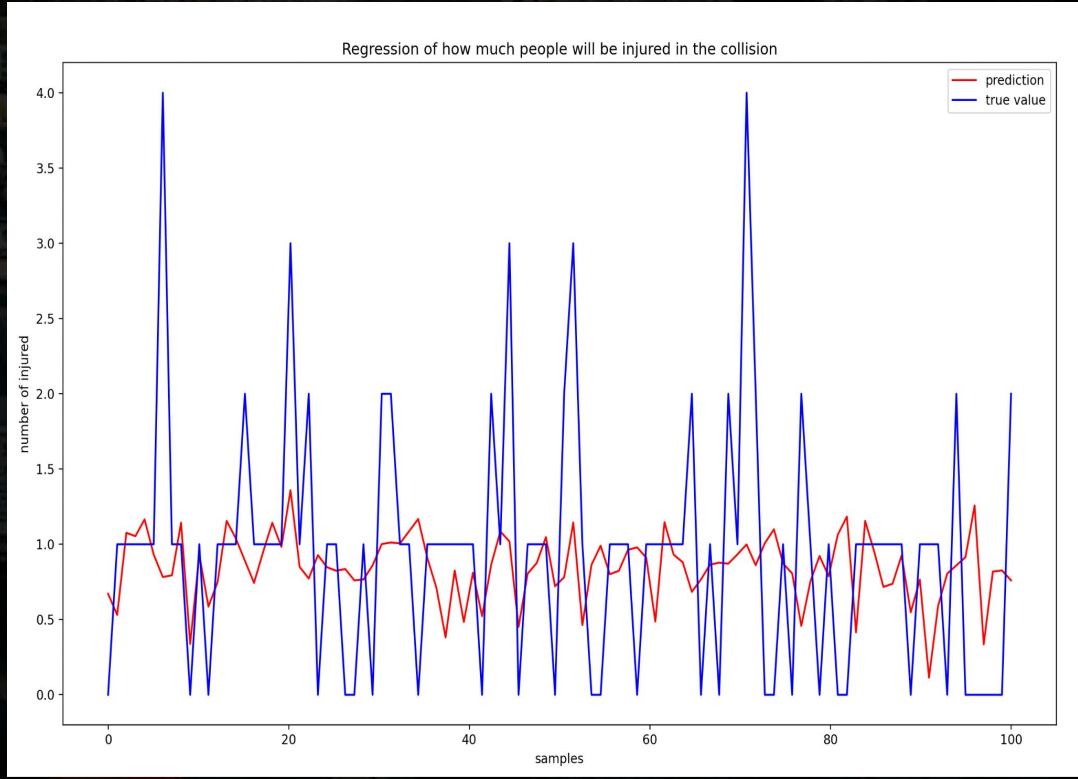
Street Risk Estimator Model 2: Regression

Input: vehicle model, road name and time.

Model 2: A regression model using Linear Regression to predict the number of people who will be injured by the collision.

Output: A real value, which means the number of people who will be injured.

The MSE of this model is about 1.



Conclusion

- Our analysis revealed relationships existing relationships between vehicle type, street location, and time to collision and casualty likelihood
  - From this we were able to determine the vehicle types, locations, and times of year corresponding to higher accident and casualty rates
- Our two models to predict accident rate and accident severity can be used to predict the likelihood of an accident and accident severity
  - Can be used as another criteria in route selection
- Future investigation should consider the frequency of unreported collisions and their potential impact on our models

References

Police beats. City of San Diego Open Data Portal. (2023a, December 5). https://data.sandiego.gov/datasets/police-beats/

Traffic collisions - people and vehicles involved. City of San Diego Open Data Portal. (2023b, December 4). https://data.sandiego.gov/datasets/police-collisions-details/

Traffic volumes. City of San Diego Open Data Portal. (2023c, April 4). https://data.sandiego.gov/datasets/traffic-volumes/