New York City - Traffic collision Spatial Analysis

Nikita Aware
UC Riverside
naawar001@ucr.edu

Nimisha Gulati
UC Riverside
ngula003@ucr.edu

Gandharva Deshpande
UC Riverside
gdesh002@ucr.edu

Vishv Patel
UC Riverside
vpate062@ucr.edu

Mihir Patel
UC Riverside
mpate125@ucr.edu

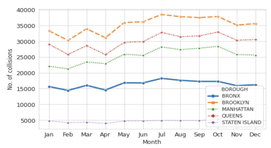
Siddhant Poojary
UC Riverside
spooj003@ucr.edu

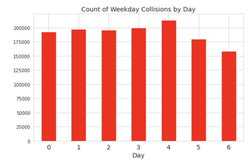
Abstract

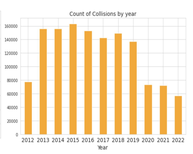
- Performed Spatio-Temporal Analysis of NYC collision Data to garner insights and interesting trends.
- Determined the impact of collisions by analyzing fatality, injuries in the boroughs of New York city using Spatial Analysis
- Developed Machine Learning models viz Logistic Regression and KNN Classification to predict safety of a location and analyze collision hotspots in NYC based on spatio-temporal factors.

Preprocessing and EDA

- Classified Collisions into 5 sub-types
- Removed more than ~20% missing values using ARCGis reverse geocode
- Engineered spatio-temporal Features







Temporal Analysis

- Serious collisions trended between 2013 to 2019
- Brooklyn has the highest number of collisions followed by Queens and Manhattan
- Collisions dropped sharply in 2020, we assume the reason is the response to COVID-19 and reduced travel
- Most collisions occurred in the evening
- Collisions are least during late night hours on weekdays but are significant on weekends.
- Fatal collisions are a smaller percentage of the dataset compared to Collisions with injuries.

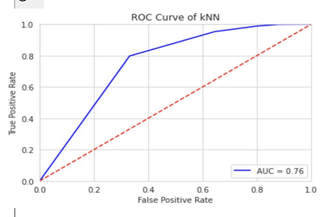
Spatial Analysis

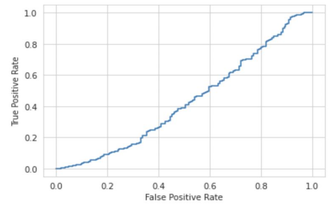
Machine Learning Results

Hotspot detection:

- Logistic regression: 0.924
- KNN classification: 0.937

- Location Safety Prediction:
  - KNN Classification - 0.909
  - Logistic regression: 0.919.
  -
  -





Conclusions

- In this project we performed Spatio-Temporal Analysis of NYC Collision Data and discovered insights and intriguing patterns.
- We further analyzed fatalities and injuries in the boroughs of New York City using spatial analysis to determine the impact of collisions.
- We also developed machine learning methods, such as logistic regression and KNN classification, to assess collision hotspots in New York City and forecast a location's safety based on spatiotemporal characteristics.