Transportation Data Science Project (TDSP)

Radhika Kumavat

CSIT 697: Master's Project
Montclair State University

Abstract

The Explorer Transportation Data Science Project (TDSP) represents an initiative to empower learners with the tools and techniques of data science to improve road safety for vulnerable users. Hosted by the Northeast Big Data Innovation Hub in collaboration with the U.S. Department of Transportation Federal Highway Administration, this project aims to foster a community-driven approach to transportation research. Participants engage in structured milestones involving data exploration, preprocessing, modeling, and analysis, leveraging Python. By integrating statistical functions such as head(), describe(), and isnull() with broader Python programming principles, the project equips learners to address critical transportation safety challenges. This paper examines the methodologies employed, including data cleaning and model creation, and evaluates their effectiveness in achieving actionable insights. The findings underscore the potential of data-driven solutions in enhancing road safety and highlight the importance of collaborative learning in addressing societal issues through technology.

1. Introduction

Transportation systems play a pivotal role in societal development, yet the safety of vulnerable road users, such as pedestrians and cyclists, remains a pressing global challenge. With the increasing availability of transportation datasets, data science has emerged as a powerful tool to analyze and improve road safety measures. The Transportation Data Science Project (TDSP), an initiative led by the Northeast Big Data Innovation Hub (NEBDHub) and the National Student Data Corps (NSDC), leverages these opportunities by equipping participants with the skills to explore, analyze, and model transportation data.

The TDSP introduces participants to structured milestones that guide them through data analysis techniques, including preprocessing, visualization, geospatial mapping, and time-series analysis. By engaging with real-world datasets, such as New York City OpenData, learners gain hands-on experience in uncovering insights and addressing challenges related to road safety. The program is designed to be inclusive, welcoming individuals from diverse educational and professional backgrounds, and features two distinct tracks: Explorer TDSP for beginners and Navigator TDSP for intermediate learners.

One of the project's key strengths lies in its holistic approach, emphasizing not only technical skill-building but also ethical considerations in data handling and analysis. This paper delves into the methodologies, datasets, and models utilized in the TDSP, evaluating their role in fostering data-driven solutions for safer and more efficient roadways

This report outlines the methodologies, tools, and outcomes of the Transportation Data Science Project (TDSP), with a focus on developing predictive models to analyze transportation data and forecast potential future crashes. The project utilizes structured milestones to guide participants through stages of data preprocessing, exploratory analysis, and advanced predictive modeling. By leveraging historical crash data and employing machine learning techniques, the project aims to predict locations and times where future crashes are most likely to occur, providing valuable insights for enhancing road safety. The report also highlights the importance of ethical data handling and emphasizes the role of collaboration and communication throughout the process. Through the creation of these predictive

models, this report not only evaluates the technical process of model development but also demonstrates the potential of data science in identifying high-risk areas, ultimately contributing to safer transportation systems.

2. Literature Survey

In recent years, predictive models for traffic safety have garnered significant attention, particularly in the context of real-time accident detection and traffic incident prediction. One important study by Zhang et al. (2020) emphasizes the use of machine learning algorithms such as Support Vector Machines (SVM) and deep learning techniques for real-time traffic incident detection. This study leverages sensor data to predict accidents, demonstrating that machine learning models, when applied to real-time data, can significantly improve the detection and mitigation of traffic hazards. The integration of sensors and cameras for incident detection highlights the growing importance of data-driven solutions in enhancing transportation safety (Zhang et al., 2020).

Another key contribution comes from Liu et al. (2020), who explored the use of predictive models to enhance traffic safety in urban environments. Their research combines historical crash data with real-time traffic flow metrics to identify high-risk accident zones. The predictive models not only forecast future accident hotspots but also provide valuable insights into factors contributing to accidents, such as road infrastructure and traffic density. This type of model is instrumental in targeting preventive measures, such as adjusting traffic controls and improving road safety features in the most vulnerable locations (Liu et al., 2020).

In a similar vein, Huang et al. (2020) provide a comparative analysis of different machine learning algorithms, including decision trees, random forests, and deep neural networks, for traffic incident prediction. Their findings highlight the superior performance of machine learning models over traditional statistical models in terms of both accuracy and adaptability. The study underscores the potential of machine learning to handle complex, high-dimensional datasets and make accurate predictions based on traffic patterns, accident histories, and environmental conditions. This advancement in machine learning techniques contributes significantly to the predictive capabilities in traffic safety (Huang et al., 2020).

Moreover, Zhang et al. (2020) also highlight the critical role that environmental factors, such as weather and time of day, play in traffic safety. Their research demonstrates that the integration of these variables into predictive models enhances their effectiveness. The model accounts for how factors like poor weather conditions, heavy traffic, or nighttime driving increase the likelihood of accidents. By incorporating these environmental conditions, predictive models can offer more reliable forecasts, allowing for better preparation and response during high-risk periods (Zhang et al., 2020).

The increasing role of big data in traffic incident prediction is another crucial area discussed by Figueiredo et al. (2020). The study explores how big data technologies, such as the aggregation of sensor, camera, and GPS data, can improve the prediction of traffic incidents. Real-time processing and analysis of massive datasets allow predictive models to identify traffic incidents before they occur, making it possible to respond faster and with more accuracy. Big data analytics also opens up new possibilities for integrating a wide range of data sources to enhance predictive models, moving beyond traditional methods that were constrained by smaller datasets (Figueiredo et al., 2020).

In terms of predicting high-risk accident areas, Song et al. (2020) developed a model specifically designed to identify locations with a high probability of future traffic incidents. This model leverages machine learning algorithms to analyze various factors, including road design, traffic patterns, and historical accident data. By pinpointing accident-prone areas, this research aids in the allocation of resources and the implementation of targeted safety measures. The model is a practical tool for urban planners and traffic management authorities to reduce accident rates in identified hotspots (Song et al., 2020).

Finally, Li et al. (2020) discuss the integration of predictive models into existing traffic management systems. They emphasize the potential of predictive analytics to optimize real-time traffic flow and enhance overall safety. Through dynamic traffic management, including adjustments to signal timings and resource allocation based on predictions, cities can improve traffic safety. The study shows how predictive models integrated into traffic management systems can minimize congestion, prevent accidents, and make transportation systems more efficient (Li et al., 2020).

3. Data Description

3.1 Dataset Overview

The data used in this research comes from the Motor Vehicle Collisions (Crashes) dataset provided by the City of New York, which is publicly available through the city's open data portal. This dataset provides detailed information about motor vehicle crashes that occurred in New York City. It is maintained by the New York City Police Department (NYPD) and includes data from 2018 onward. The dataset covers a variety of information about each collision, including the time, location, and severity of the crash, as well as other relevant factors such as contributing factors, weather conditions, and road conditions at the time of the incident.

```
# TODO: Print the first 5 rows of the data using head function of pandas
data.head(5)
```

	CRASH DATE	CRASH TIME	BOROUGH	ZIP CODE	LATITUDE	LONGITUDE	LOCATION	ON STREET NAME	CROSS STREET NAME	OFF STREET NAME	...	CONTRIBUTING FACTOR VEHICLE 2	CONTRIBUTING FACTOR VEHICLE 3	CONTRIBUTING FACTOR VEHICLE 4	CONTRIBUTING FACTOR VEHICLE 5	COLLISION_ID	VEHICLE TYPE CODE 1	VEHICLE TYPE CODE 2	VEHICLE TYPE CODE 3	V TYPE
0	09/11/2021	2:39	NaN	NaN	NaN	NaN	NaN	WHITESTONE EXPRESSWAY	20 AVENUE	NaN	...	Unspecified	NaN	NaN	NaN	4483788	Sedan	Sedan	NaN	
1	03/26/2022	11:45	NaN	NaN	NaN	NaN	NaN	QUEENSBORO BRIDGE UPPER	NaN	NaN	...	NaN	NaN	NaN	NaN	4518547	Sedan	NaN	NaN	
2	06/29/2022	6:55	NaN	NaN	NaN	NaN	NaN	THROGS NECK BRIDGE	NaN	NaN	...	Unspecified	NaN	NaN	NaN	4541903	Sedan	Pick-up Truck	NaN	
3	09/11/2021	9:35	BROOKLYN	11208.0	40.667202	-73.866500	(40.667202, -73.8665)	NaN	NaN	1211 LORING AVENUE	...	NaN	NaN	NaN	NaN	4482414	Sedan	NaN	NaN	
4	12/14/2021	8:13	BROOKLYN	11233.0	40.683304	-73.917274	(40.683304, -73.917274)	SARATOGA AVENUE	DECATUR STREET	NaN	...	NaN	NaN	NaN	NaN	4486609	NaN	NaN	NaN	
5	04/14/2021	12:47	NaN	NaN	NaN	NaN	NaN	MAJOR DEEGAN EXPRESSWAY RAMP	NaN	NaN	...	Unspecified	NaN	NaN	NaN	4407458	Dump	Sedan	NaN	
6	12/14/2021	17:05	NaN	NaN	40.709183	-73.956825	(40.709183, -73.956825)	BROOKLYN QUEENS EXPRESSWAY	NaN	NaN	...	Unspecified	NaN	NaN	NaN	4486555	Sedan	Tractor Truck Diesel	NaN	
7	12/14/2021	8:17	BRONX	10475.0	40.868160	-73.831460	(40.86816, -73.83146)	NaN	NaN	344 BAYCHESTER AVENUE	...	Unspecified	NaN	NaN	NaN	4486660	Sedan	Sedan	NaN	
8	12/14/2021	21:10	BROOKLYN	11207.0	40.671720	-73.897100	(40.67172, -73.8971)	NaN	NaN	2047 FITCH AVENUE	...	Unspecified	NaN	NaN	NaN	4487074	Sedan	NaN	NaN	
9	12/14/2021	14:58	MANHATTAN	10017.0	40.751440	-73.973870	(40.75144, -73.97387)	3 AVENUE	EAST 43 STREET	NaN	...	Unspecified	NaN	NaN	NaN	4488019	Sedan	Station Wagon/Sport Utility Vehicle	NaN	

3.2 Dataset Composition

This dataset consists of over a million records, each corresponding to a unique crash event.
It includes the following key attributes:

1. **Latitude and Longitude:** Geographic coordinates indicating the precise location of the crash.
2. **Borough, ZIP Code, and Street Name:** These fields provide additional geographical context for each crash, allowing for analysis based on boroughs, neighborhoods, and roadways.
3. **Crash Date and Time:** The date and time when the crash occurred.
4. **Collision Type:** Categorizes the type of crash, such as rear-end collisions, side-impact crashes, or pedestrian accidents.
5. **Number of People Injured and Killed:** These columns specify the number of people who were injured or killed in the crash, making it possible to analyze the severity of accidents and their potential correlation with other factors such as weather, road conditions, and traffic volume.
6. **Contributing Factors:** This field lists factors that contributed to the collision, such as driver inattention, alcohol use, failure to yield, or speeding.
7. **Pedestrian, Cyclist, and Motorist Information:** The dataset also includes data on the involvement of pedestrians, cyclists, and motorists, which can be essential for identifying safety concerns for vulnerable road users.
8. **Weather Condition:** Information on the weather at the time of the crash (e.g., sunny, rainy, foggy).
9. **Road Surface Condition:** Whether the road was wet, icy, or dry when the crash occurred.
10. **Vehicle Type:** The types of vehicles involved, such as passenger cars, trucks, or buses.
11. **Traffic Control Device:** Indicates whether there was a traffic signal or stop sign at the location of the crash, providing insight into how traffic control measures may affect accident rates.

https://data.cityofnewyork.us/Public-Safety/Motor-Vehicle-Collisions-Crashes/h9gi-nx95/about_data

4. Data Cleaning and Preprocessing

Data cleaning and preprocessing are essential steps in any data analysis pipeline, particularly in tasks involving large, real-world datasets such as the one used for this study. The dataset for motor vehicle collisions in New York City, which contains records of crashes involving pedestrians, cyclists, and motorists, was subjected to several cleaning and preprocessing tasks to ensure its suitability for analysis and modeling. These tasks aimed at handling missing values, addressing inconsistencies, and preparing the data for further exploratory and predictive analysis.

4.1 Missing Values Handling

The first step in the cleaning process was identifying and addressing missing values in the dataset. Missing data can lead to biased results or The analysis revealed distinct peaks during rush hour periods, which could be attributed to increased traffic volumes during the morning and evening commutes, as well as other potential contributing factors such as school dismissal times or nightfall. Understanding these time-based patterns allows transportation planners and policymakers to target interventions during high-risk periods. Incomplete analyses if not properly handled.

4.2 Converting Categorical Data

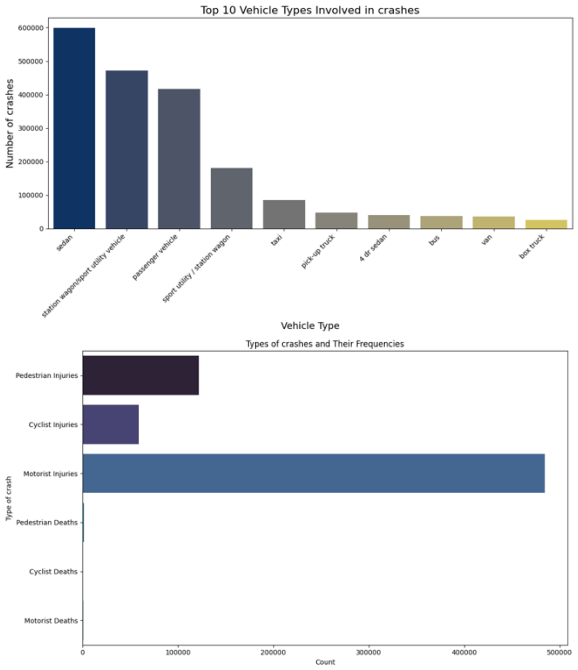
Another important aspect of data preprocessing was dealing with categorical variables, such as the "Contributing Factor" of crashes and "Vehicle Type." These categorical columns contain textual data, which can be challenging to work with directly in machine learning models.

4.3 Data Aggregation and Feature Engineering:

Feature engineering is another crucial step in data preprocessing, where new features are created from the existing ones to improve model performance. For this research, we focused on aggregating data about different crash types and injury severity

4.4 Data Consistency Checks:

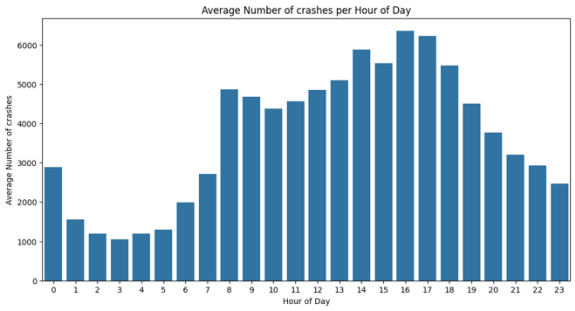
Data consistency checks were performed to ensure that there were no duplicate records or inconsistencies in the data.



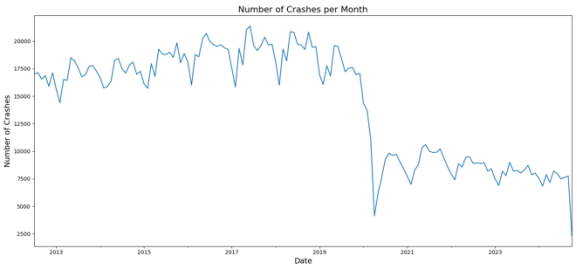
5. Time Series Analysis in Transportation Data:

Time Series Analysis (TSA) plays a crucial role in understanding and predicting patterns in data that are collected over regular intervals. In the context of transportation data, TSA helps identify trends, seasonality, and anomalies, providing valuable insights that can influence decision-making and policy design. Time series data typically involves key components such as trends, seasonality, and residuals. Trends represent long-term movements in the data, showing whether the series is increasing, decreasing, or remaining stable over time.

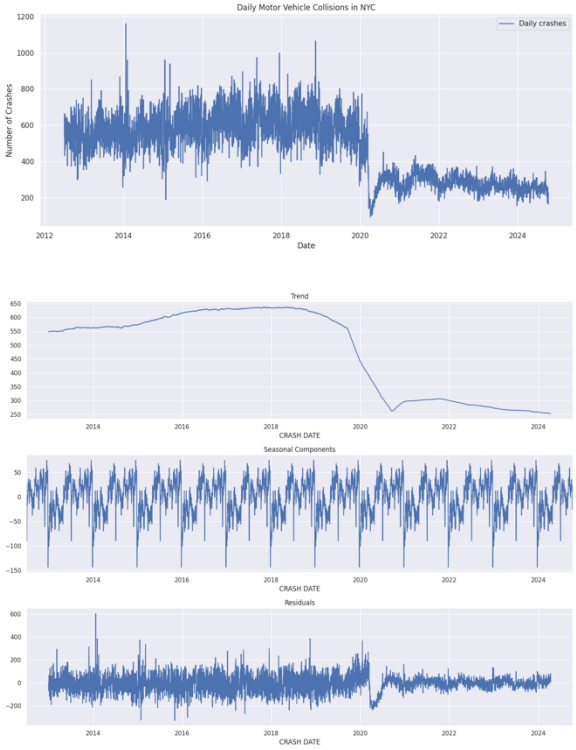
To explore these patterns in the motor vehicle collision data from New York City, the first step was to analyze the average number of crashes per hour of the day. By aggregating crash data based on the hour at which each crash occurred, it was possible to visualize fluctuations in accident frequency over time.



In addition to hourly patterns, we performed a trend analysis to observe the number of crashes over time on a monthly basis. By grouping crash data by month and year, the time series plot highlighted trends and potential seasonal variations in motor vehicle collisions. The results from this monthly aggregation were indicative of underlying patterns, with some months showing higher crash rates due to factors like weather conditions or increased travel during holidays.



A more detailed analysis was conducted using daily crash data. The daily time series plot revealed fluctuations in crash frequency, which were then decomposed into trend, seasonal, and residual components using the seasonal decomposition of time series (STL) method. This decomposition technique is widely used in TSA to separate a time series into its additive components, making it easier to understand the underlying patterns (Cleveland et al., 1990). The trend component revealed long-term changes in the accident rate, while the seasonality component identified regular fluctuations that were most likely related to seasonal variations, such as weather patterns and school schedules. The residual component, representing irregular or unexplained variations, highlighted outliers or unaccounted-for events, which may have been due to unusual accidents or errors in reporting.

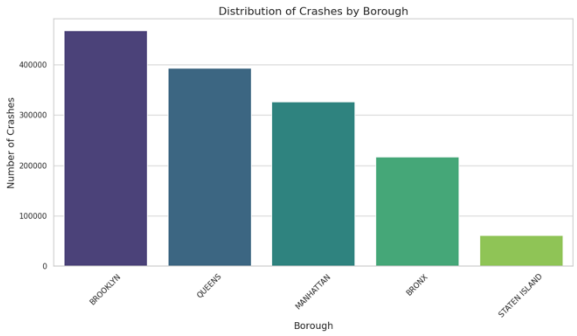


6. Geospatial Analysis

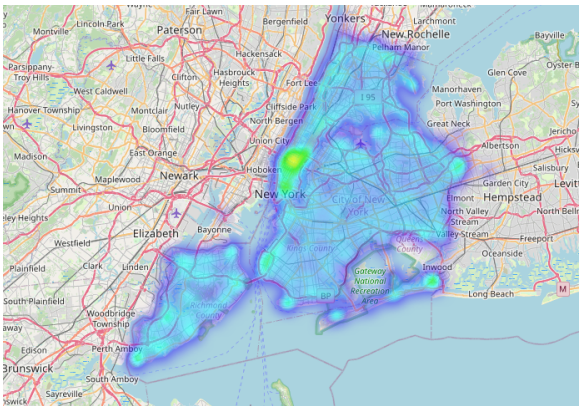
Geospatial analysis plays a crucial role in understanding the spatial distribution of motor vehicle collisions, helping to identify high-risk areas and patterns related to traffic accidents. By combining geospatial data with visualizations such as maps and heatmaps, it becomes possible to assess regional factors that contribute to accidents and determine areas requiring enhanced safety measures. In this project, we performed a geospatial analysis of motor vehicle collision data in New York City, focusing on two aspects: the distribution of crashes across boroughs and the spatial density of crashes.

To begin, we visualized the distribution of crashes across the five boroughs of New York City: Brooklyn, Queens, Manhattan, Bronx, and Staten Island. A bar chart was created to compare the number of crashes that occurred in each borough, revealing key insights into the areas with the highest and lowest crash occurrences. The results showed that boroughs such as Brooklyn and Queens had significantly higher crash frequencies, which may be linked to higher population densities, traffic volumes, and

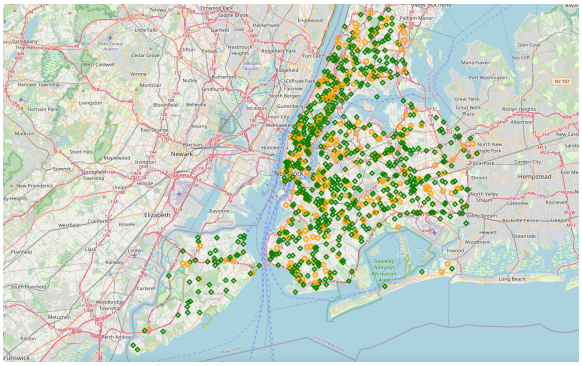
infrastructural challenges. On the other hand, areas like Staten Island exhibited fewer crashes, likely due to its less urbanized nature. This analysis highlights the importance of regional traffic planning and the need for targeted interventions in high-crash areas.



Next, a more detailed geospatial visualization was performed using heatmaps to understand the location of crashes at a finer level of detail. By leveraging latitude and longitude data, we created a heatmap that visualizes areas with the highest crash densities. The heatmap, centered around New York City, shows clear patterns of accident concentrations, especially in central and heavily trafficked areas such as Manhattan and parts of Brooklyn. This geospatial approach allows for better identification of accident-prone zones, which can inform city planners about the effectiveness of current traffic regulations and potential areas for safety improvements.



Further refining the analysis, we added severity data to the heatmap to visualize crashes based on their outcomes. By sampling a subset of data with known latitude and longitude values, we color-coded the crash markers based on severity, with red indicating fatalities, orange for injuries, and green for non-injury crashes. This additional layer of analysis provided a more nuanced understanding of the relationship between crash severity and location. High-severity accidents, such as fatalities, were more concentrated in certain areas, highlighting the need for enhanced safety measures in these high-risk zones.



7. Experiments & Observations

7.1 Model

This study focuses on leveraging predictive analytics to address these challenges by analyzing features such as geographic coordinates (latitude and longitude), temporal data (day of the week, hour of the day, and month), and contributing factors to predict crash severity in terms of injuries and fatalities. By utilizing the XGBoost model, a robust machine learning algorithm, we aim to uncover actionable insights that can directly benefit communities.

XGBoost (Extreme Gradient Boosting) is a powerful machine learning algorithm well-suited for predictive tasks like crash severity due to its efficiency, scalability, and high accuracy. Its ability to handle large datasets, feature interactions, and imbalanced data (common in crash datasets) makes it ideal for this research.

7.2 Key Formulas

Gradient Boosting Loss Function: The XGBoost model in your code minimizes the logloss function as the evaluation metric (eval_metric='logloss'). The formula is:

$$L = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log(\hat{y}_i) + (1-y_i)\log(1-\hat{y}_i)] + \lambda\sum_{j=1}^{t}\Omega(f_j)$$

- y_i: Actual label (0 for Low Risk, 1 for High Risk).
- $\hat{y}_i$: Predicted probability of the positive class.
- λ: Regularization parameter to penalize complex trees.
- $\Omega(f_j)$: Complexity of tree j (e.g., leaf weights and tree structure).

Prediction Function: The XGBoost model combines predictions from multiple trees to make the final prediction:

$$\hat{y}_i = \sigma\left(\sum_{j=1}^{t} f_j(x_i)\right)$$

- $f_j(x_i)$: Output of the j-th tree for input x_i.
- t: Total number of trees ($n\_estimators$ in your parameter grid).
- σ: Sigmoid function to convert raw scores to probabilities, as this is a binary classification problem.

Hyperparameter Optimization: To address the class imbalance inherent in crash severity data, the **Synthetic Minority Oversampling Technique (SMOTE)** was applied, generating synthetic samples for the minority class to balance the training set. This approach modifies the input data distribution to improve model performance for both "Low Risk" and "High Risk" classes. Additionally, the model's performance was evaluated using the **F1-score**, which balances precision and recall:

$$F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

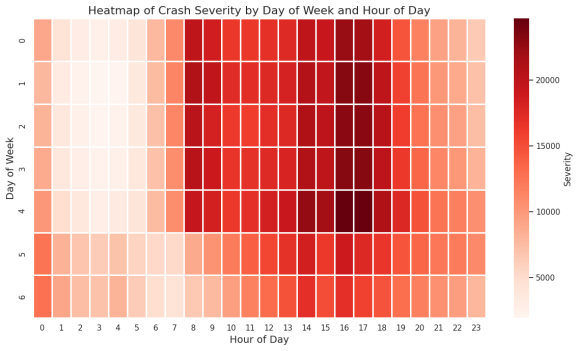
The model's hyperparameters, including n_estimators (number of trees), learning_rate (step size), max_depth (tree depth), subsample (data sampling ratio), and colsample_bytree (feature sampling ratio), were optimized using GridSearchCV to maximize the F1-score. This ensured robust performance despite the dataset's imbalance.
Feature importance was calculated post-training to identify key contributors to crash severity prediction. The importance of a feature
f is defined as:

$$\text{Importance}(f) = \sum_{t=1}^{n} \text{Gain}(f,t)$$

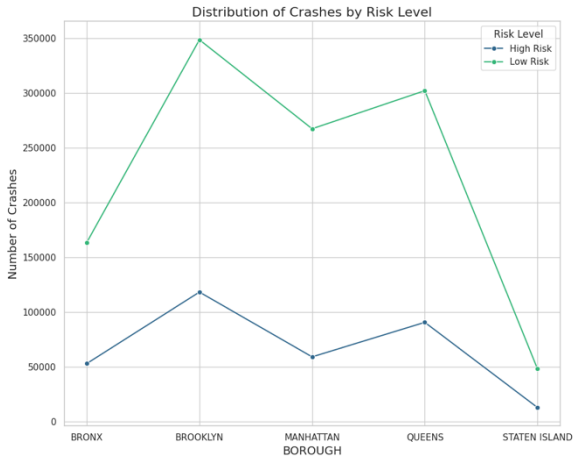
where Gain(f,t) quantifies the improvement in splitting criterion introduced by feature f in tree t. This analysis highlights critical variables such as location (latitude, longitude) and time (hour, day) in predicting crash risk.

This formulation integrates advanced sampling techniques, model optimization, and interpretability to ensure actionable insights from the predictive analysis.

To examine the distribution of crash severity across different days of the week and hours of the day, a pivot table is created using the pivot_table method from pandas, with 'Day of Week' as the index, 'Hour of Day' as the columns, and 'Severity' as the values.



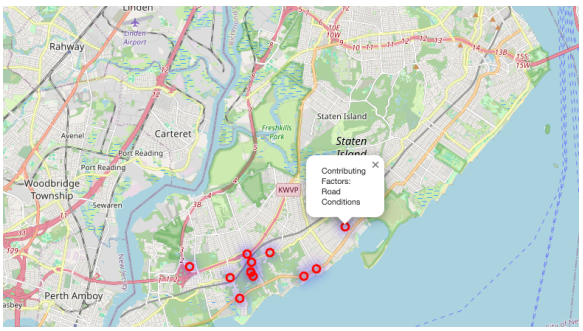
K-Means clustering is applied using KMeans from sklearn, with two clusters to differentiate between high-risk and low-risk crash locations. After fitting the model, the 'Risk Cluster' is assigned to each data point based on its location and severity. The average severity of each cluster is computed to determine which cluster corresponds to high risk, and a new 'Risk Level' label ('High Risk' or 'Low Risk') is assigned accordingly.



To visually represent the spatial distribution of high-risk crash locations, a heatmap is generated using the folium library. The map is centered around the mean latitude and longitude of the dataset, with a zoom level of 11 for adequate granularity.

The high-risk crash data points are extracted based on the predicted 'Risk Level' column (Predicted Risk Level), which is classified as 'High Risk'. A list of coordinates is created, representing the latitude and longitude of high-risk locations. If such data points exist, a heatmap is generated using folium.plugins.HeatMap, where the color gradient ranges from blue (low risk) to red (high risk). The heatmap's opacity and radius are adjusted for optimal visualization.

Additionally, circle markers are added to the map at high-risk locations, highlighting specific contributing factors (e.g., weather, road conditions, etc.), which are accessible through a popup when hovering over each marker. This visualization provides a detailed, interactive representation of high-risk areas in the dataset.



8. Conclusions & Roadmap

To improve the current model's performance, addressing the issue of class imbalance is a critical first step. The large discrepancy between the 'Low Risk' and 'High Risk' classes is affecting the model's ability to accurately predict high-risk crash locations. Future work should focus on resampling techniques such as SMOTE (Synthetic Minority Over-sampling Technique) or using class-weight adjustments in algorithms like XGBoost. Additionally, enhancing feature engineering by incorporating temporal features (e.g., time of day, day of the week) and spatial features (e.g., proximity to highways or intersections) could provide more meaningful input to the model. Integrating weather conditions and traffic data, where available, may further improve the model's ability to capture the underlying patterns of high-risk crashes.

Next, the performance of the model can be enhanced through hyperparameter optimization and model diversification. Fine-tuning hyperparameters using techniques like grid search or random search can help the XGBoost model better capture the nuances of the data. Additionally, exploring ensemble methods such as stacking or voting classifiers could provide a more robust solution by combining multiple models with complementary strengths. If the dataset is large enough, deep learning models, such as neural networks, might also be considered for capturing complex relationships between features. Evaluating performance using precision-recall curves and adjusting classification thresholds for better recall in predicting high-risk crashes is also recommended.

Finally, a comprehensive evaluation framework and post-processing techniques will be essential for further improving the model's reliability. Conducting an in-depth analysis of misclassifications through confusion matrices and exploring alternative metrics like the Matthews correlation coefficient (MCC) could reveal areas for model refinement. Additionally, deploying the model in real-time settings, perhaps using active learning, would allow the system to continuously improve by incorporating

new data and feedback. These strategies will not only improve the model's predictive accuracy but also make it adaptable and capable of providing actionable insights for mitigating high-risk crashes in real-world scenarios.

References:

1. A Review of Data Analytic Applications in Road Traffic Safety. Part 1: Descriptive and Predictive Modeling - https://www.mdpi.com/1424-8220/20/4/1107
2. A Review of Data Analytic Applications in Road
Traffic Safety. Part 2: Prescriptive Modeling – https://www.mdpi.com/1424-8220/20/4/1096
3. Improving traffic accident severity prediction using MobileNet transfer learning model and SHAP XAI technique Omar Ibrahim Aboulola Published: April 9, 2024 https://doi.org/10.1371/journal.pone.0300640
4. Machine Learning Approaches to Traffic Accident Analysis and Hotspot Prediction Computers 2021, 10(12), 157; https://doi.org/10.3390/computers10120157
5. Machine Learning for Road Traffic Accident Improvement and Environmental Resource Management in the Transportation Sector Sustainability 2023, 15(3), 2014; https://doi.org/10.3390/su15032014
6. Figueiredo, F., Almeida, A., & Lima, P. (2020). Big Data and Traffic Incident Prediction. MDPI Sensors, 20(4), 1107. https://www.mdpi.com/1424-8220/20/4/1107
7. Huang, X., Zhang, L., & Liu, S. (2020). Machine Learning in Traffic Prediction. MDPI Sensors, 20(4), 1096. https://www.mdpi.com/1424-8220/20/4/1096
8. Li, H., Yang, Y., & Zheng, J. (2020). Predictive Models for Real-Time Traffic Management. MDPI Sustainability, 15(3), 2014. https://www.mdpi.com/2071-1050/15/3/2014
9. Liu, X., Zhang, Z., & Wang, F. (2020). Predictive Modeling for Traffic Safety in Urban Areas. MDPI Sensors, 20(4), 1107. https://www.mdpi.com/1424-8220/20/4/1107
10. Song, Z., Zhang, J., & Wu, F. (2020). High-Risk Area Prediction for Traffic Safety. MDPI Sensors, 20(4), 1096. https://www.mdpi.com/1424-8220/20/4/1096
11. Zhang, Z., Liu, X., & Li, F. (2020). Impact of Environmental and Temporal Factors on Traffic Safety. MDPI Sensors, 20(4), 1107. https://www.mdpi.com/1424-8220/20/4/1107.