
Toronto Traffic Collision Prediction Model

By: Rakeen Ahmed

Agenda

-
- Project Overview
 - Data Overview
 - Data Exploration/EDA and Data Visualization
 - Modeling
 - Collision Risk Map
 - Feature Importance
 - Feature Interpretation
 - Recommendations
 - Questions
-

Project Overview

Background
Problem Statement
Objective



Background

Traffic Collisions: Major public health concern in cities worldwide today

- No. 1 cause of mortality between ages 15-29

Toronto Stats:

- 3rd leading cause of mortality
- 1,581 deaths in the last 10 years

Background

Vision Zero Program

Goal: Eliminate deaths and serious injuries resulting from collisions

Approach: “Data Driven insights based on past collision data to inform road design and infrastructure improvements”

Toronto lags behind behind major European and American cities in terms of road safety

Problem Statement

- **The project arose from a need to provide data-driven insights to inform road design and infrastructure improvements**
- **The goal is to identify the key drivers of collision risk in terms of street design and infrastructure**
 - **Data sources: road infrastructure**

Objective

- **Objective:** Develop a collision prediction model for the City of Toronto
- **Goal:**
 - Estimate and map the risk of traffic collisions on a given road and identify high-risk road segments
 - Identify the key design and infrastructure features driving collision risk to inform design changes.
- **Method:** Utilize road infrastructure (bike lanes, traffic cameras, speed limits), design, and historical collision data
- **Result:** Understanding key factors driving collision risk will inform effective design changes and lead to reduced collision

Data Overview

Data Sources

Merging Datasets

Exclusions

Data Sources

- **Toronto Centrelane Road Network**
- **All Traffic Collisions**
- **Traffic Volumes**
- Centrelane Lanes
- Automated Speed Enforcement
- Cycling Network
- Pedestrian Crossings
- Traffic Calming
- Traffic Cameras
- Posted Speed Limits
- Sensitive Zones
- KSI

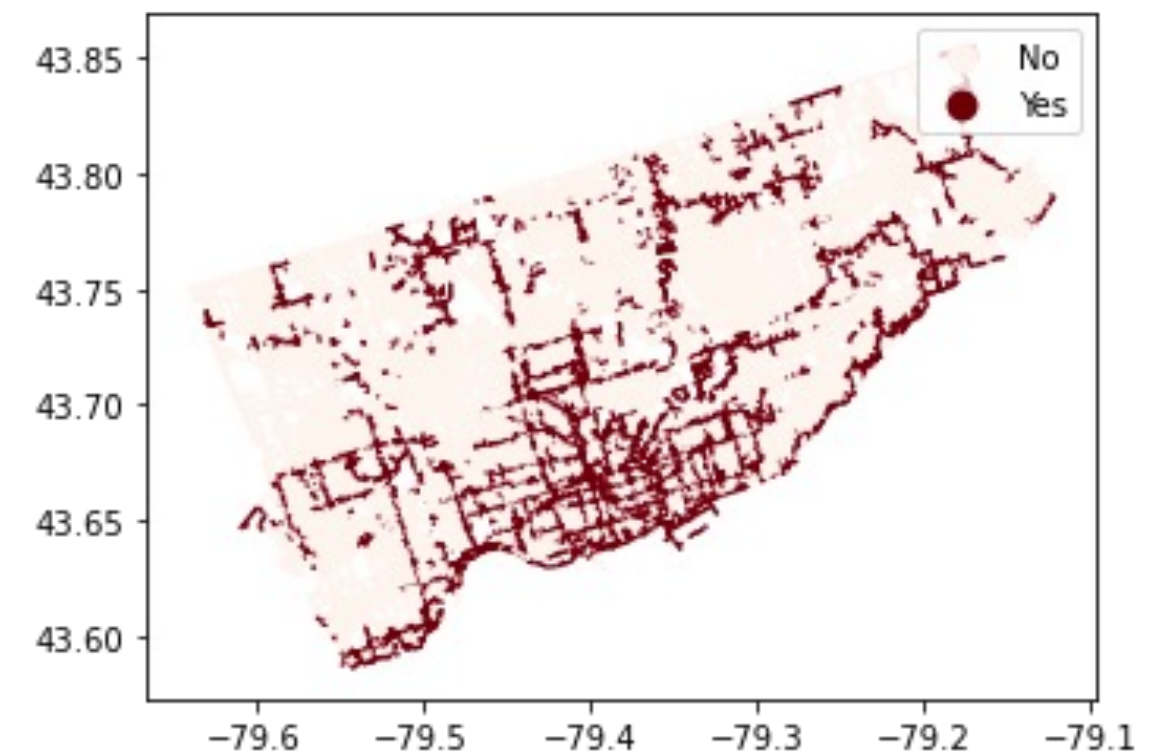
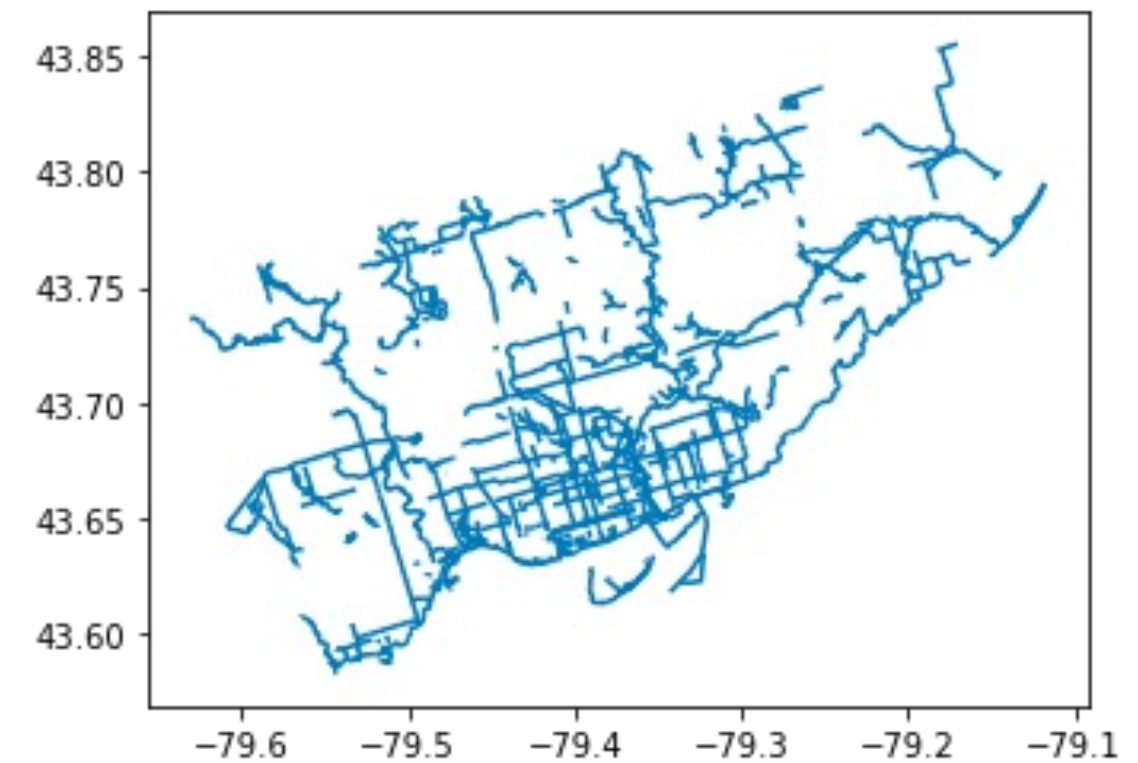


Merging Datasets

Datasets were merged using a spatial join using GeoPandas using exact or radius search.

Toronto Road Network was used as the base dataset

Get design and infrastructure features for all road segments for our prediction model.



Final Dataset & Exclusions

Final Dataset contained 13,979 rows and 26 columns. (All features not shown)

Excluded all collisions before 2016 from the collisions and KSI datasets.

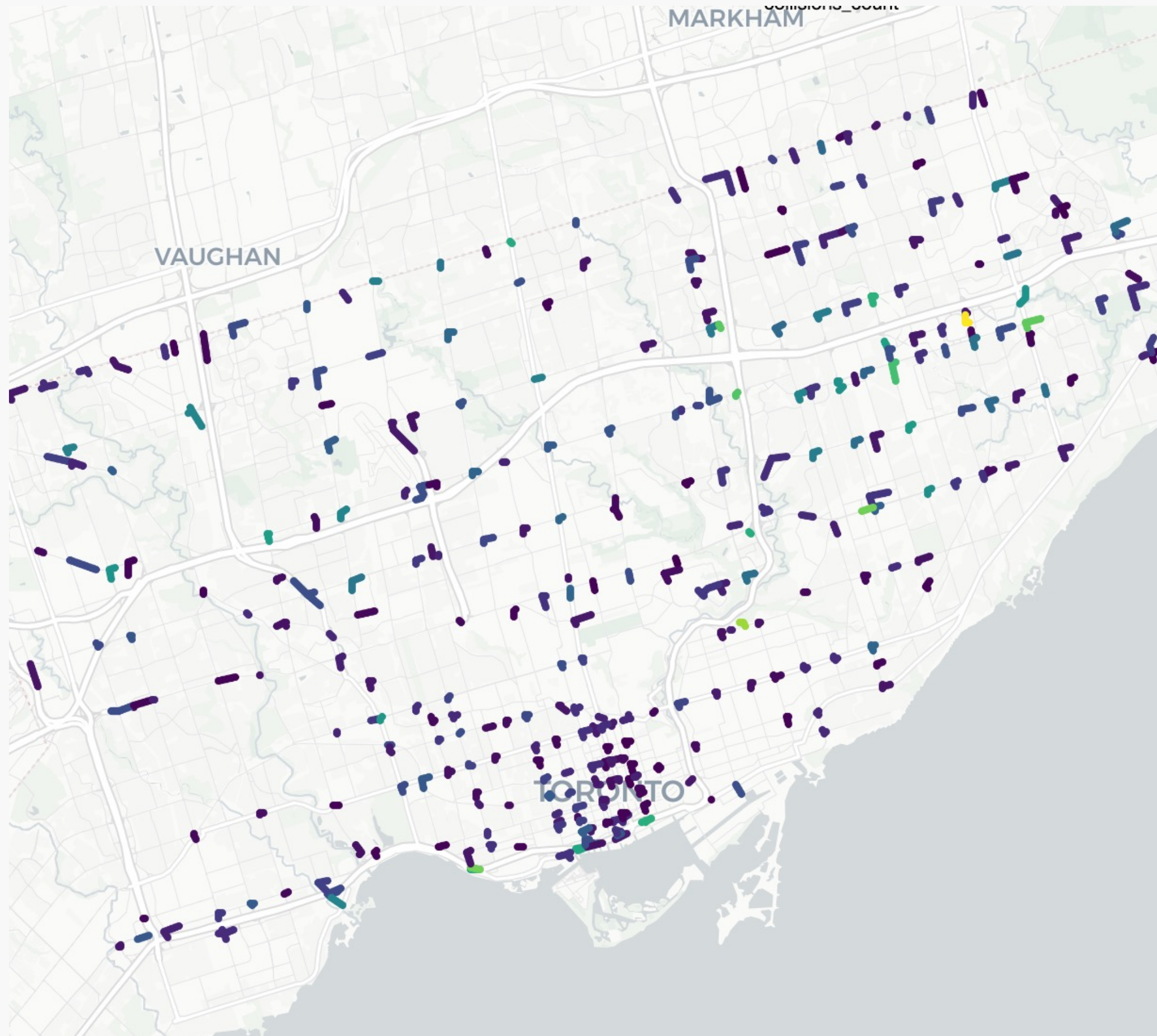
	road_classification	bikelane_present	number_of_lanes	pedestrian_crossover_present	speed_enforcement_p
treline_id					
914587	Major Arterial	No	2	No	
6735911	Major Arterial	No	2	No	
914588	Local	Yes	2	Yes	
9950476	Major Arterial	Yes	2	No	
3135620	Major Arterial	No	2	No	

EDA & Data Visualization

Feature Engineering
Data Visualization

Feature Engineering

- **Collision Rate:** The rate of collision of a road segment,
 - Derived from the total count of collisions divided by 1000 road users (pedestrians, vehicles, trucks etc.)
 - Measure of frequency of collision on each road segment by traffic volume
- **Collision Class:** Bins of collision counts for use in classification models.
- **Speed Hump Present:** Converted speed hump count to a flag

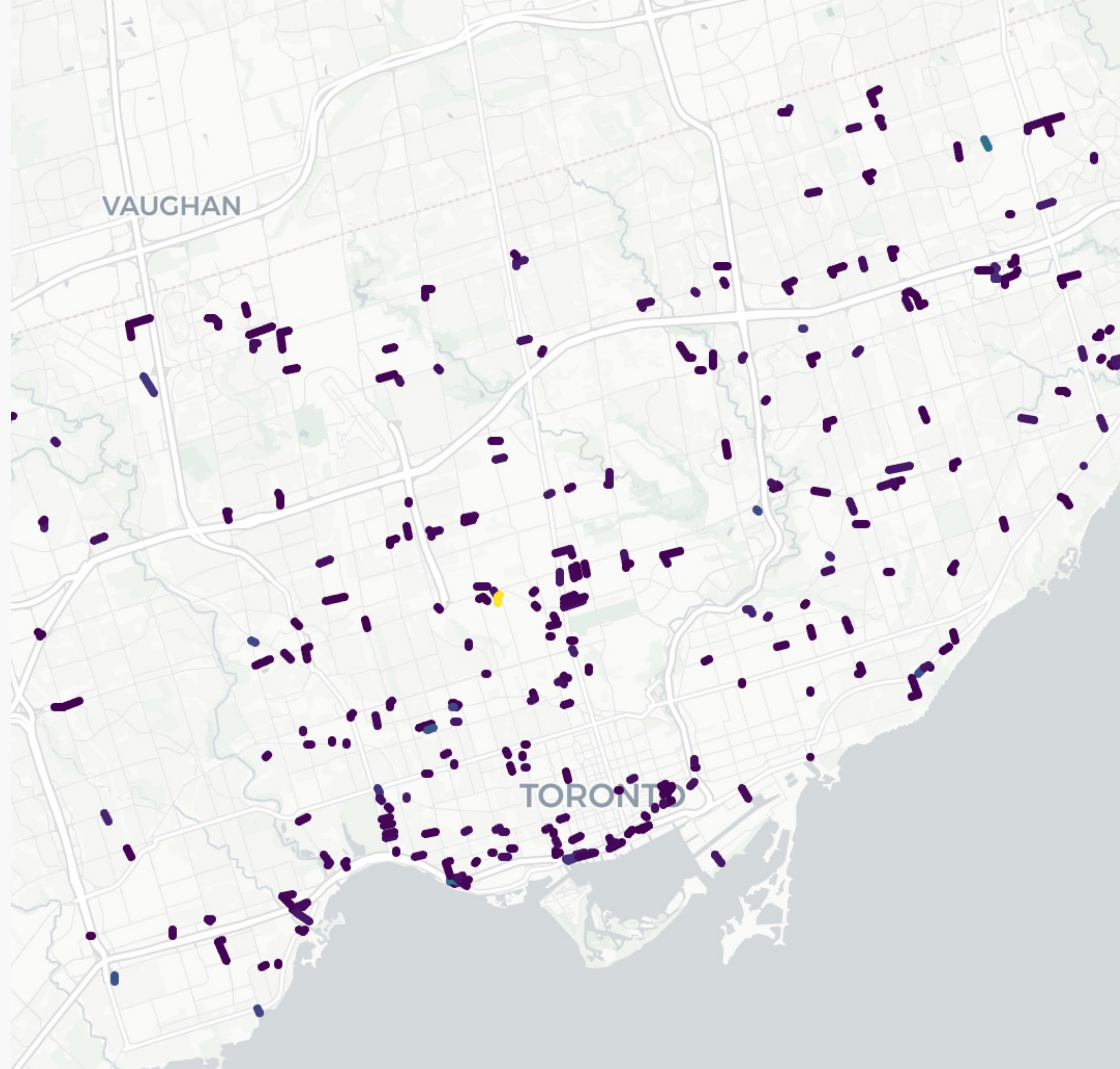


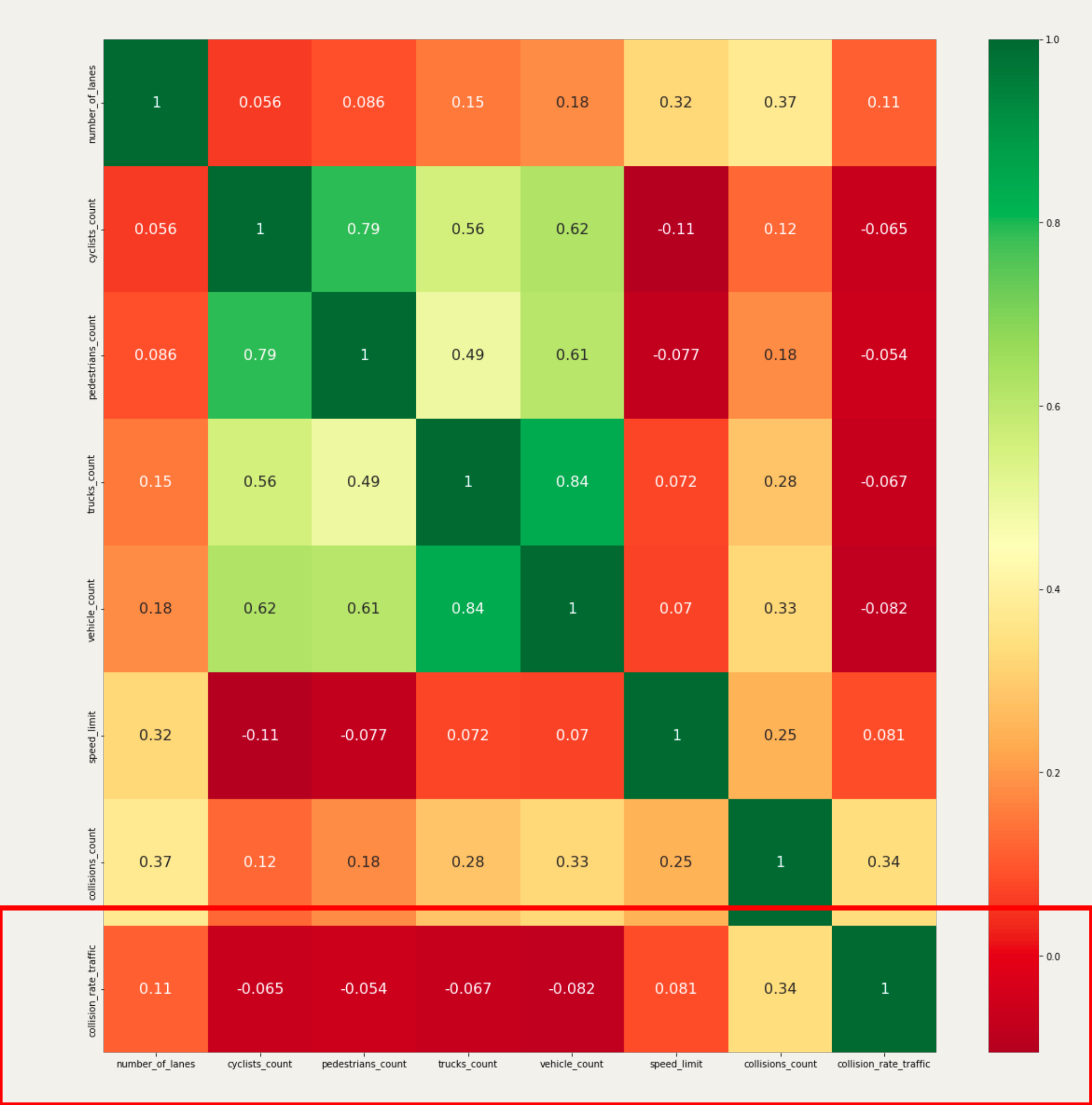
Data Visualization

Road Segments with
the highest number
of collisions

Data Visualization

Road Segments with the
highest collision rates.





Data Visualization

Correlation Matrix

Modeling

Algorithms

Modeling Breakdown

Metrics

Model Comparison

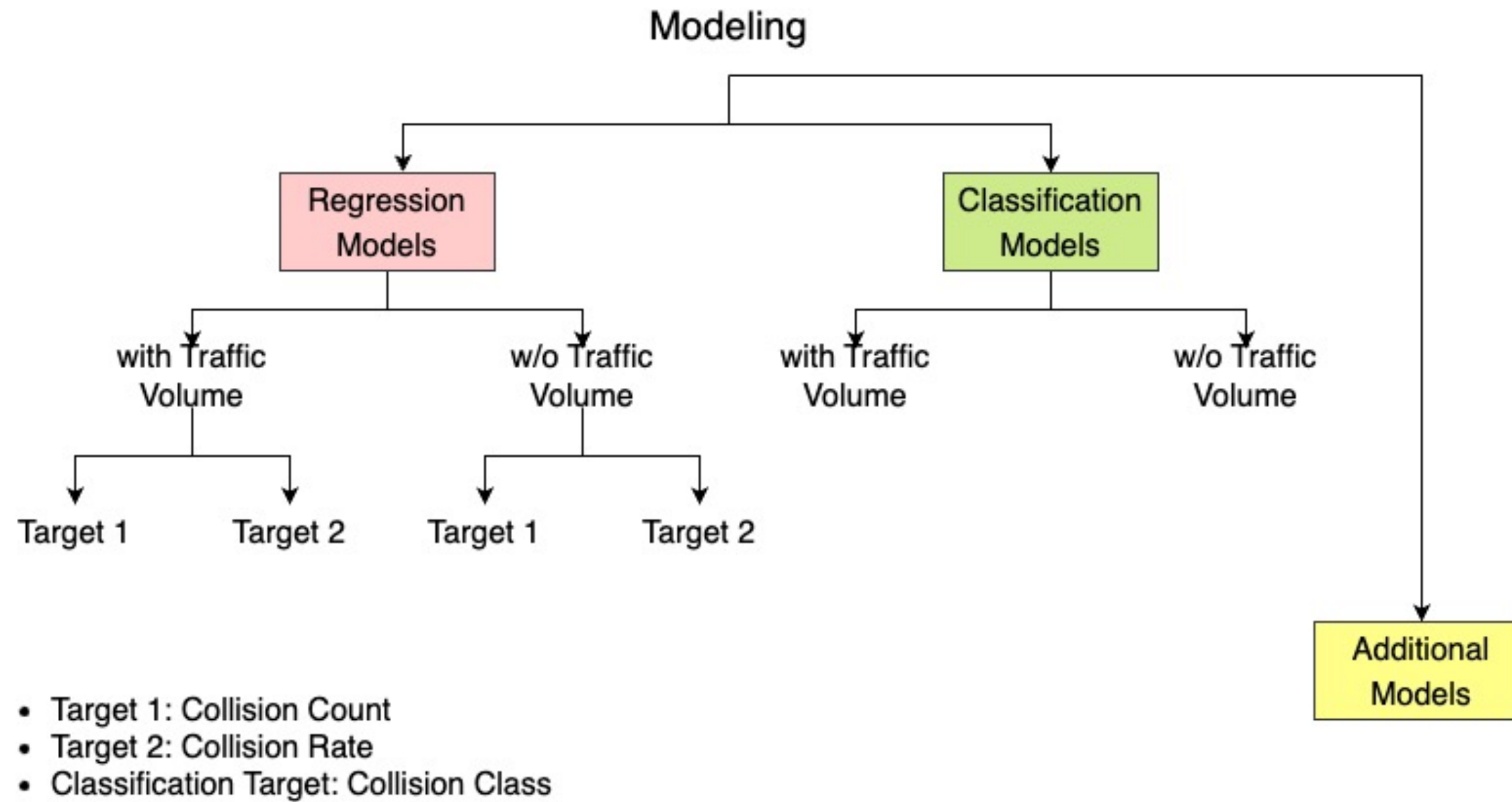
Model Selection

Algorithms

-
- Decision Tree
 - Random Forest
 - Neural Network
 - AdaBoost
 - Linear Regression
 - Logistic Regression
 - KNN
 - XGBoost



Modeling Breakdown



Metrics

Regression Models

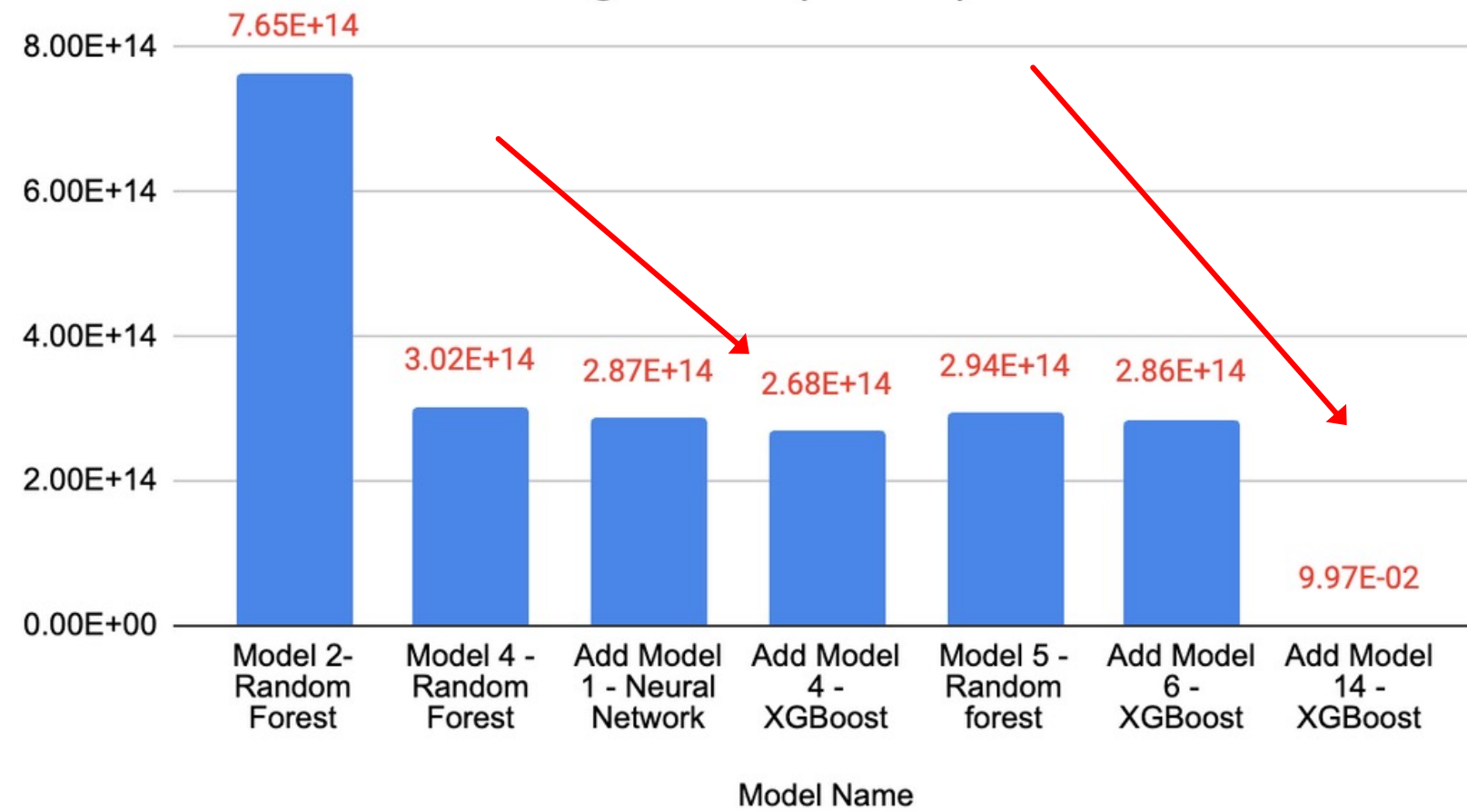
- **Mean Absolute Percentage Error**
 - Average percentage difference between the actual and predicted values
- **Normalized Mean Absolute Error**
 - MAE measures the average absolute difference between the actual and predicted values
 - MAE divided by the range of y-values to normalize the error
 - Normalization enables comparison between diff. target variables

Classification Models

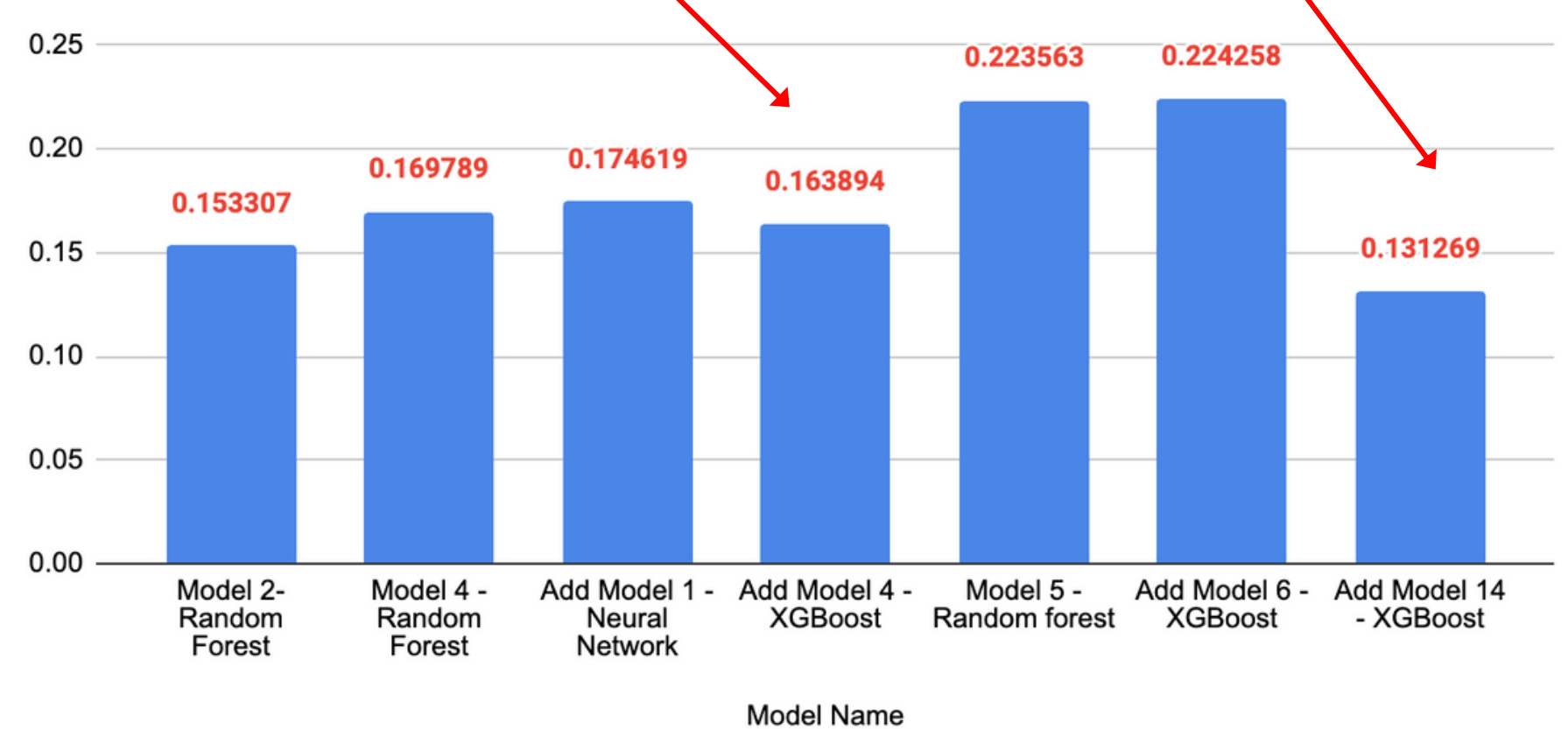
- Accuracy

Model Comparison - Regression

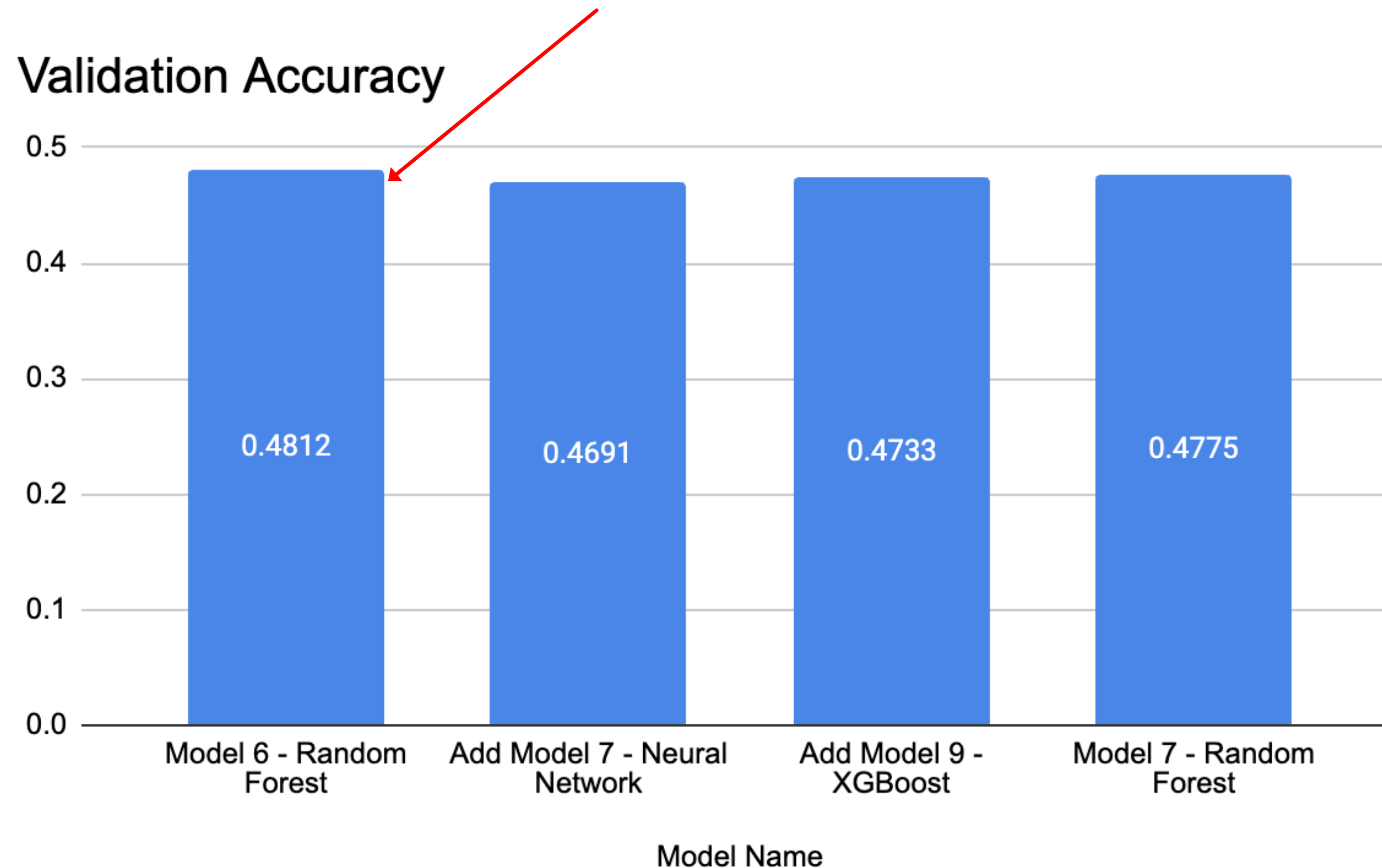
Mean Absolute Percentage Error (MAPE)

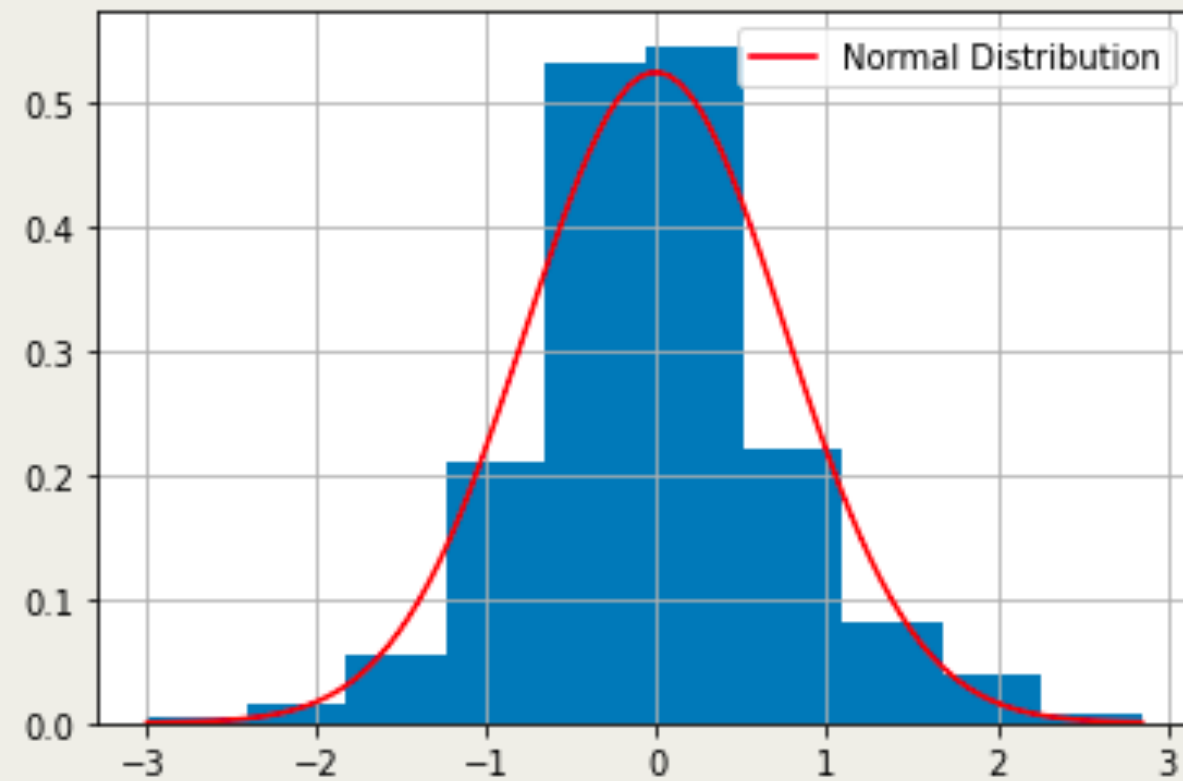


Normalized Mean Absolute Errors



Model Comparison - Classification





Model Selection

- **Add Model 4 – XGBoost** was selected as the best model as it had a combination of low MAPE score as well as low Normalized MAE.
- Regression model preferred.
- XGBoost Models are very fast to train.
- Residuals of the model are normally distributed.

Regression Statistics:

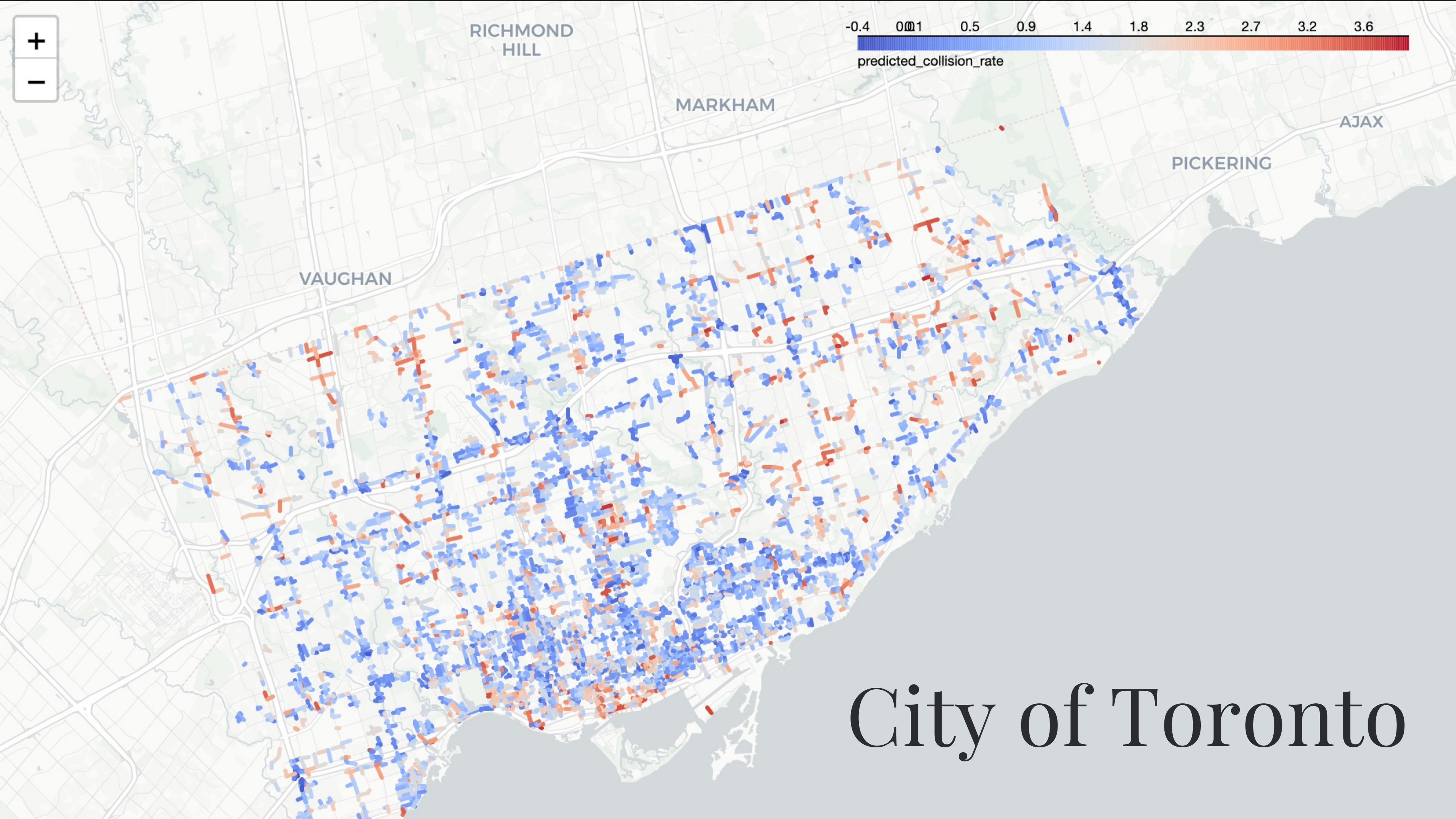
Mean Squared Error (MSE): 0.5766868238

Mean Absolute Error (MAE): 0.566978475

Mean Absolute Percentage Error (MAPE):

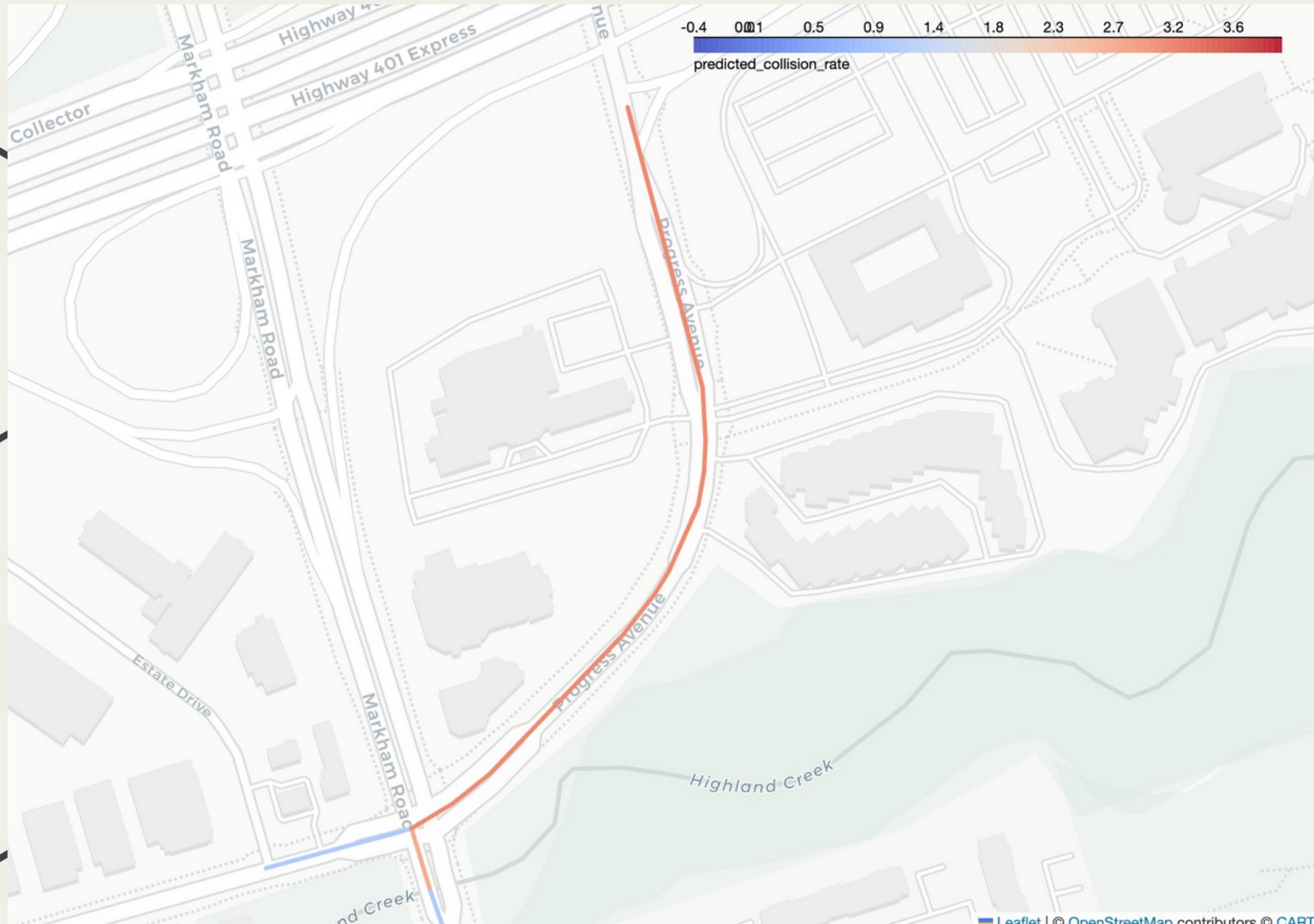
Normalized MAE:collision_rate_traffic
dtype: float64

Collision Risk Map



City of Toronto

Centennial College



COLLISION RATE

Predicted: 3.01 per 1000 road users

Actual: 3.44 collisions per 1000 road users

COLLISION COUNT

Predicted: 150 (Collision Cap)

Actual: 624

Feature Importance

Best Model

High Collision Rate Segments

Predictors	Coef	Odds Ratio
road_classification_Expressway Ramp	0.115145	1.122036121
road_classification_Local	0.102791	1.108259759
road_classification_Major Arterial	0.099045	1.104115984
traffic_cameras_present_Yes	0.098958	1.10401993
road_classification_Collector Ramp	0.092863	1.097311393
vehicle_count	0.081711	1.085142158
number_of_lanes	0.051327	1.052667059
road_classification_Minor Arterial	0.045857	1.04699469

Feature Importance – Best Model

Feature
Importance
of Add
Model 4 –
XGBoost

FEATURE IMPORTANCE – HIGH COLLISION RATE

Feature
Importance of
Add Model 14 -
XGBoost

Predictors	Coef	Odds Ratio
road_classification_Collector	0.127857	1.136390487
road_classification_Major Arterial Ramp	0.108899	1.115049725
road_classification_Expressway	0.091324	1.09562393
pedestrians_count	0.078018	1.081142119
vehicle_count	0.064473	1.06659678
trucks_count	0.059852	1.061679406
bikelane_present_Yes	0.059549	1.061357766
road_classification_Pending	0.059516	1.061322742
road_classification_Minor Arterial	0.050338	1.051626486
cyclists_count	0.04574	1.046802207

Feature Interpretation

Feature Interpretation

Road Classification: Major Arterial – 1.10

Collision rates are 10% higher on Major Arterial Roads as compared to all other road classes.

Road Classification: Minor Arterial – 1.04

Collision rates are 4% higher on roads classified as Minor Arterials.

Vehicle Count – 1.08

For each additional 1000 vehicles on the road segment, the collision rate increases 8%.

Number of Lanes – 1.05

For each additional lane on the road segment, the collision rate increases 5%

Recommendations

Recommendations
Future Work

Recommendations

Toronto Traffic Services should focus on these road segments:

- Road Classification: Major or Minor Arterial
- High Vehicle Count
- High Number of Lanes

Use the collision prediction model to identify road segments that require design and safety improvements.



Future Work

Include additional road features in the model – Avg. Traffic Speeds, Traffic Signals, 311 requests

Develop separate collision models – Roads with high collision counts or high pedestrian counts, downtown vs suburbs.

Map collision risk for all road segments in Toronto.

Thank you!