GROUP 19

Vehicle Collision Analysis in Toronto

Qi Zhao
Weijing Zhang
Jiabo Wang
Nancy Pham

CONTENTS

Toronto Vehicle Collisions -- Killed and Seriously Injured

- Description: 12557 entries, 58 features
- Data Highlights:
  - **Where:** Geographic Location (District, Ward, and Neighbourhood)

    Relative Location (Intersection, Mid-Block)

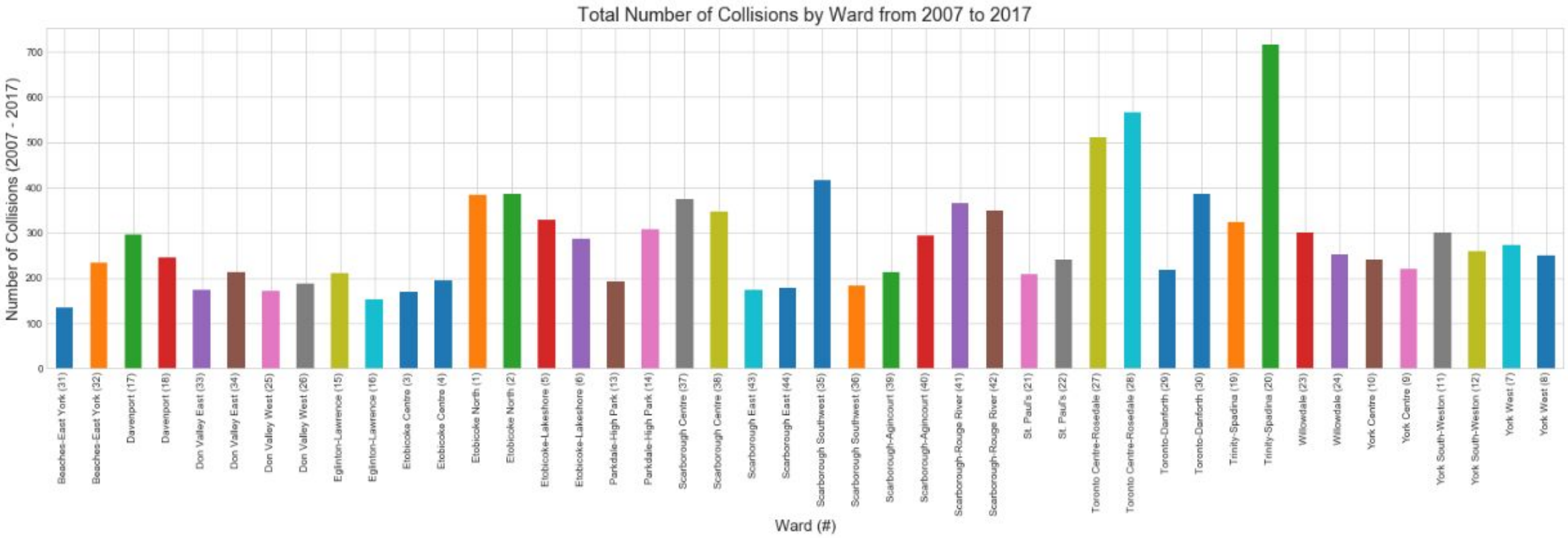
    Road Type (Highway, Major Street, Residential)
  - **When:** Date, Time
  - **Factors:** Lighting, Visibility, Road Conditions, Driver Conditions

	alcohol	speeding	redlight	visibility	light	rdsfcond	drivact	drivcond	initdir	invtype	manoeuver	ward_id	hood
0	No	No	No	Clear	Dark	Dry	Driving Properly	Normal	West	Driver	Going Ahead	27	
1	No	1	No	Clear	Dark, artificial	Dry	Other	Unknown	Unknown	Passenger	Other	40	
2	No	No	No	Clear	Dark, artificial	Dry	Other	Unknown	South	Pedestrian	Other	30	
3	No	1	No	Clear	Dark, artificial	Dry	Other	Unknown	Unknown	Vehicle Owner	Other	40	

02 Objectives

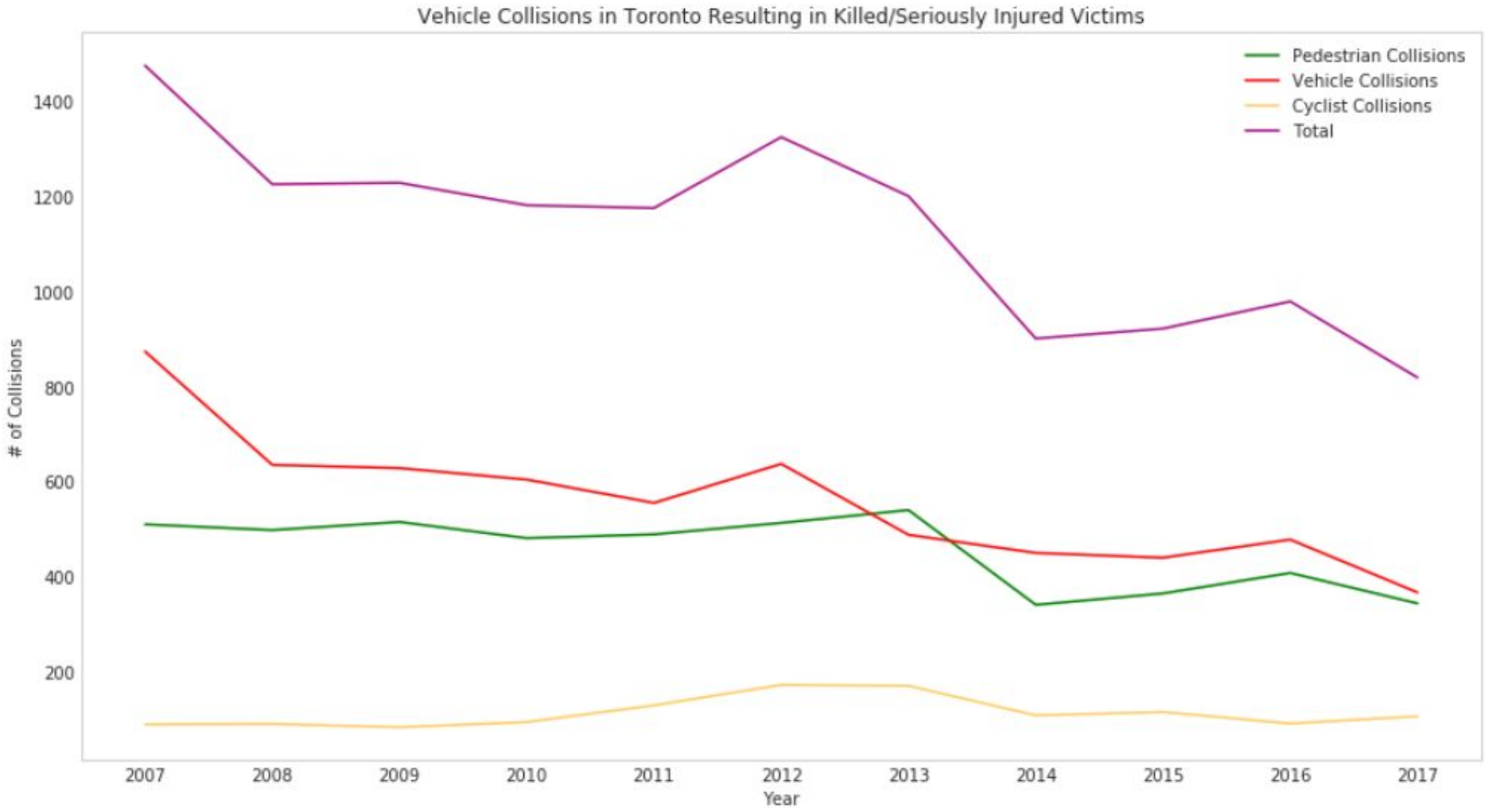
- Find the correlations in the data.
- Build model to predict the types of collisions that might occur given certain conditions.
- The purpose is to help the city to be more prepared as well as to help take preventative measures.

03 Data Analysis



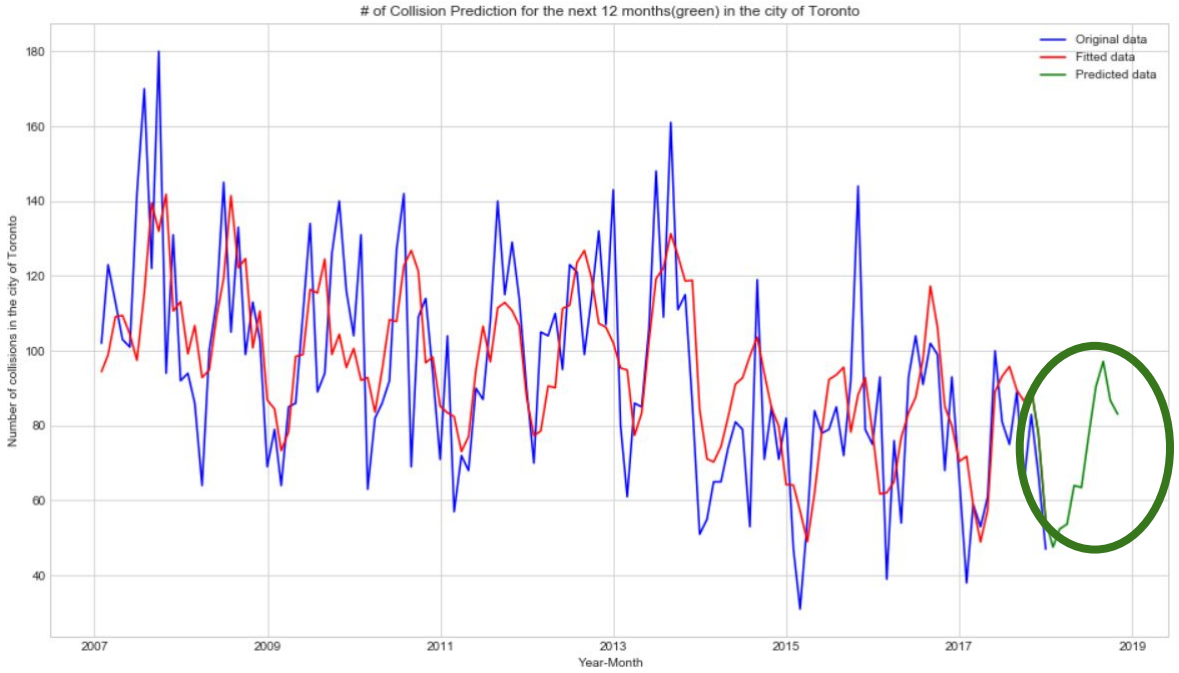
Total number of collisions by ward from 2007 to 2017

03 Data Analysis

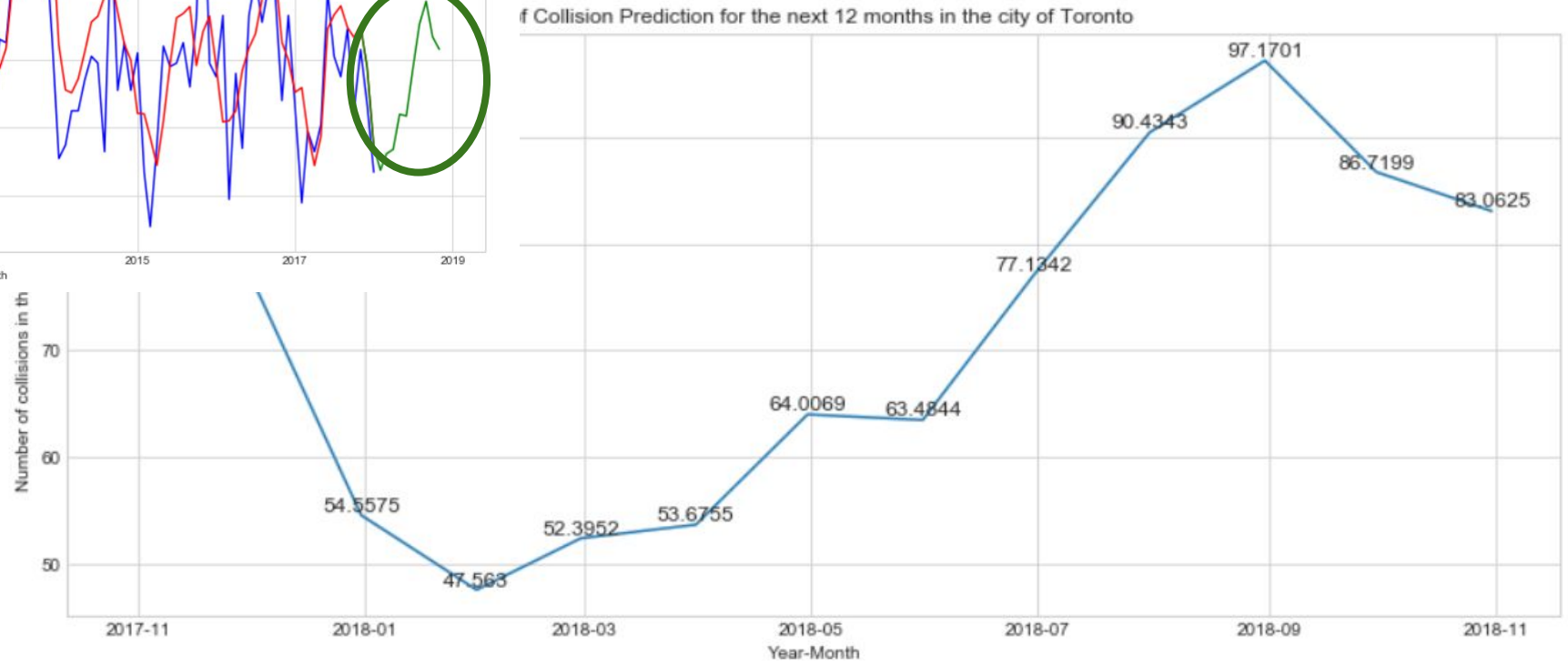


Vehicle Collisions in Toronto Resulting in killed/seriously injured Victims

03 Time-series Prediction Result




```
from statsmodels.tsa.arima_model import ARIMA
model = ARIMA(Toronto_df_count, order=(9, 0, 9))
results_AR_Toronto = model.fit(disp=-1)
plt.plot(Toronto_df_count)
plt.plot(results_AR_Toronto.fittedvalues, color='red')
plt.title('Prediction')
```



Number of Collision Prediction for the next 12 months in the city of Toronto

03 Data Analysis

Predictive Modelling

- Target Variable: Type of Collision -- Pedestrian, Cyclist, or Vehicle-Related
- Input Features:
  - **Where:** Geographic Location (Ward)

    Relative Location (Intersection, Mid-Block)

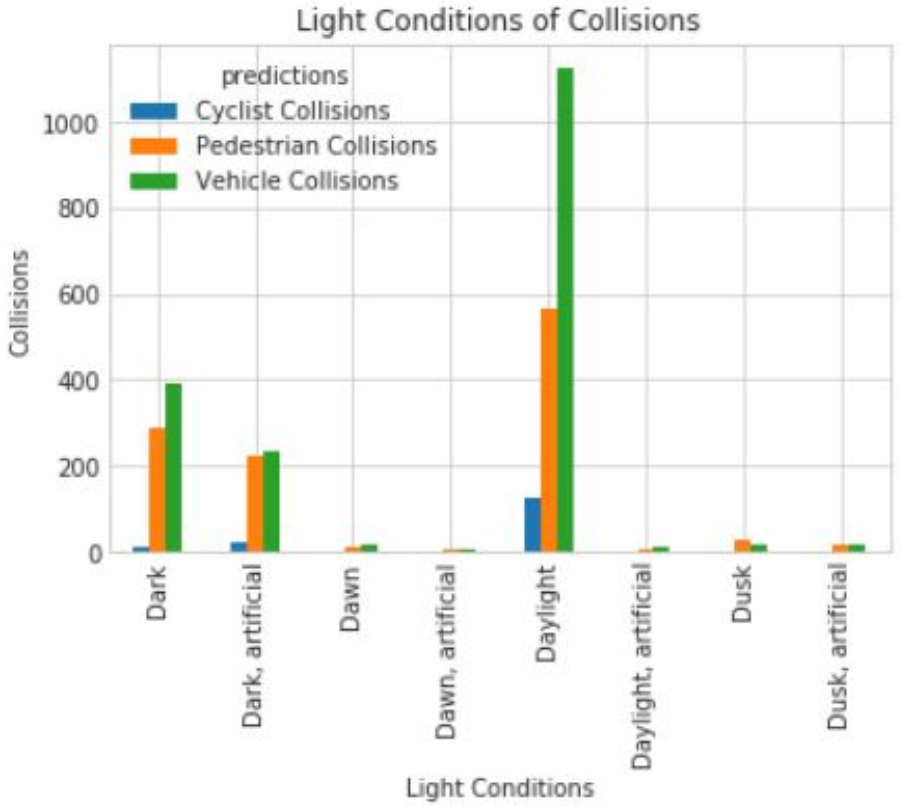
    Road Type (Highway, Major Street, Residential)
  - **When:** Time of Day, Day of Week
  - **Factors:** Lighting, Visibility, Road Conditions
- Algorithm: Random Forest

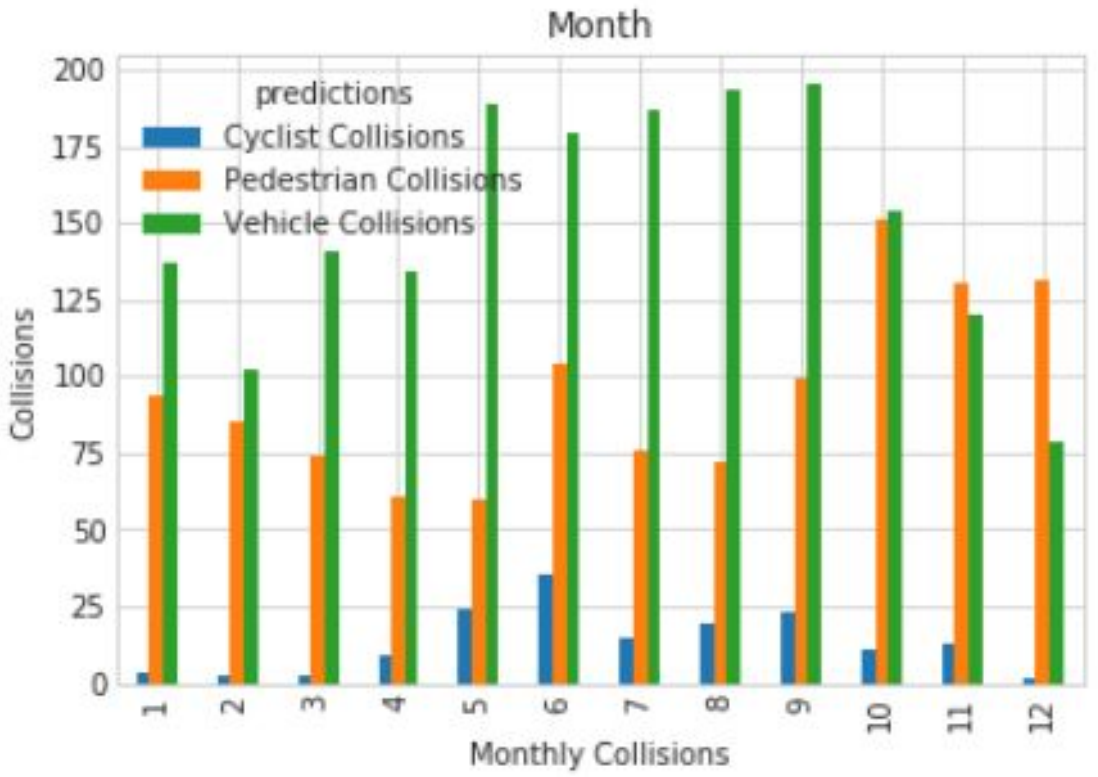
Accuracy for Random Forest for Vehicle Collision Classification: 71.76%

	precision	recall	f1-score	support
Cyclist Collisions	0.49	0.25	0.33	312
Pedestrian Collisions	0.72	0.74	0.73	1237
Vehicle Collisions	0.74	0.79	0.77	1556
avg / total	0.71	0.72	0.71	3105

04 Model Output

- We can split the the types of collisions to understand better the contribution from the factors.

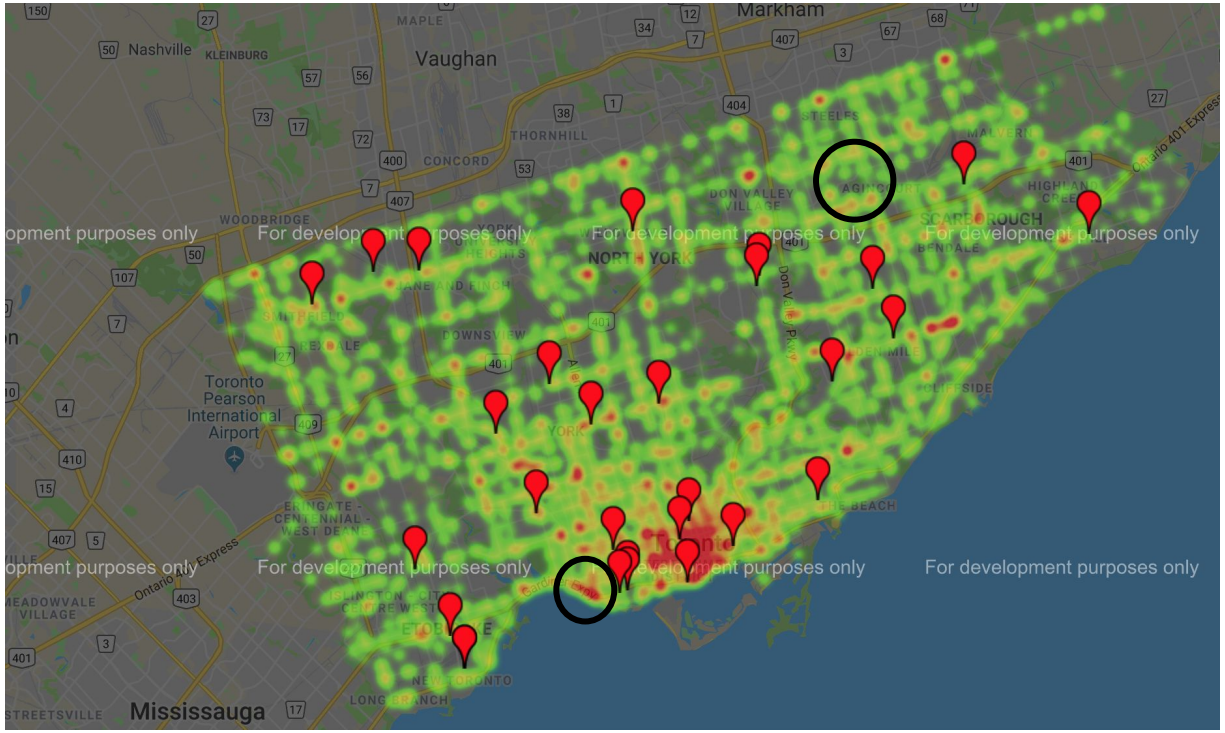




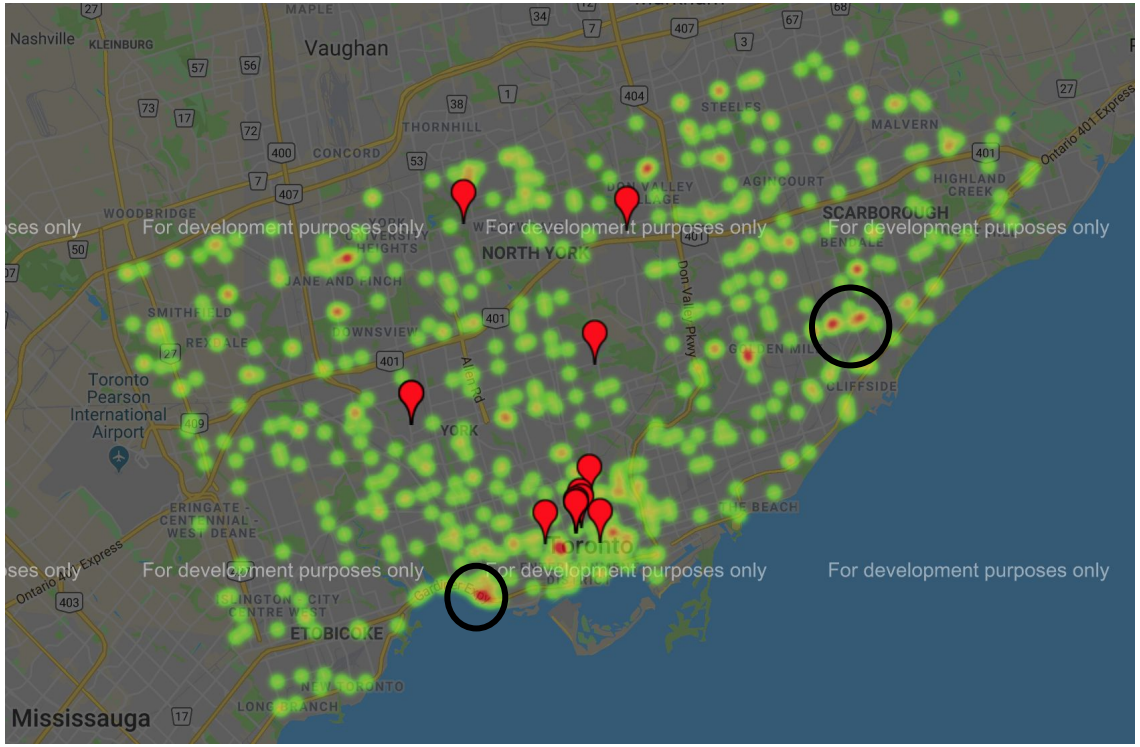
04 Problem

Mapping on Google Maps

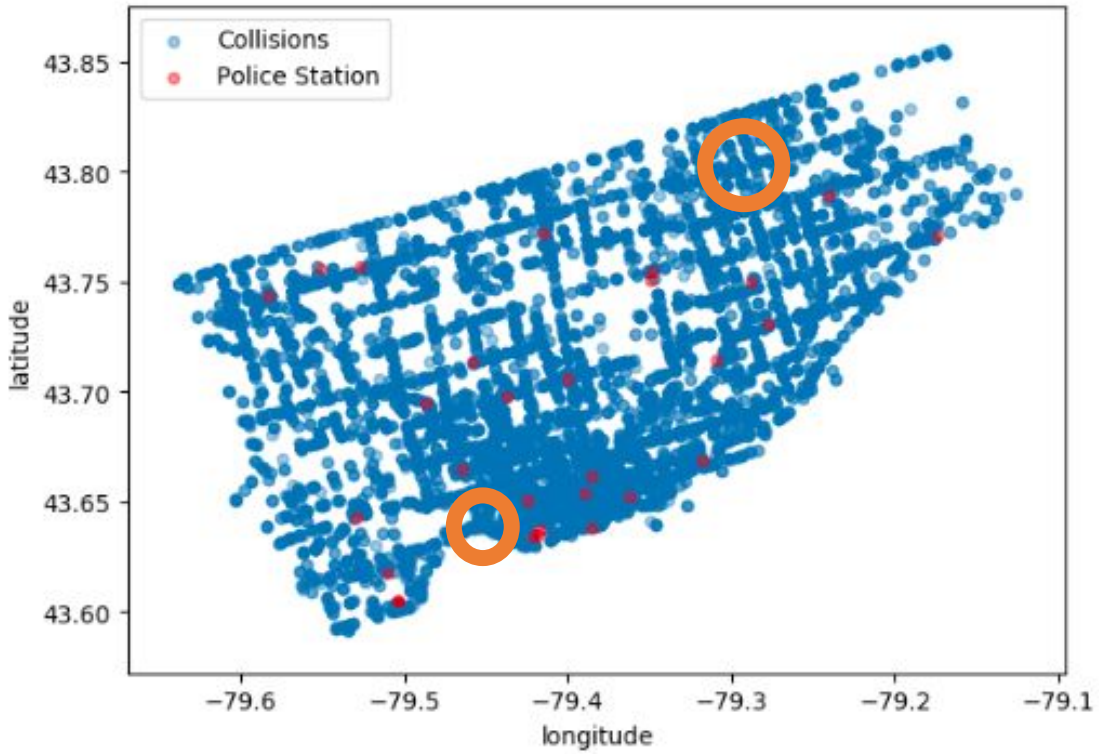
Police Division

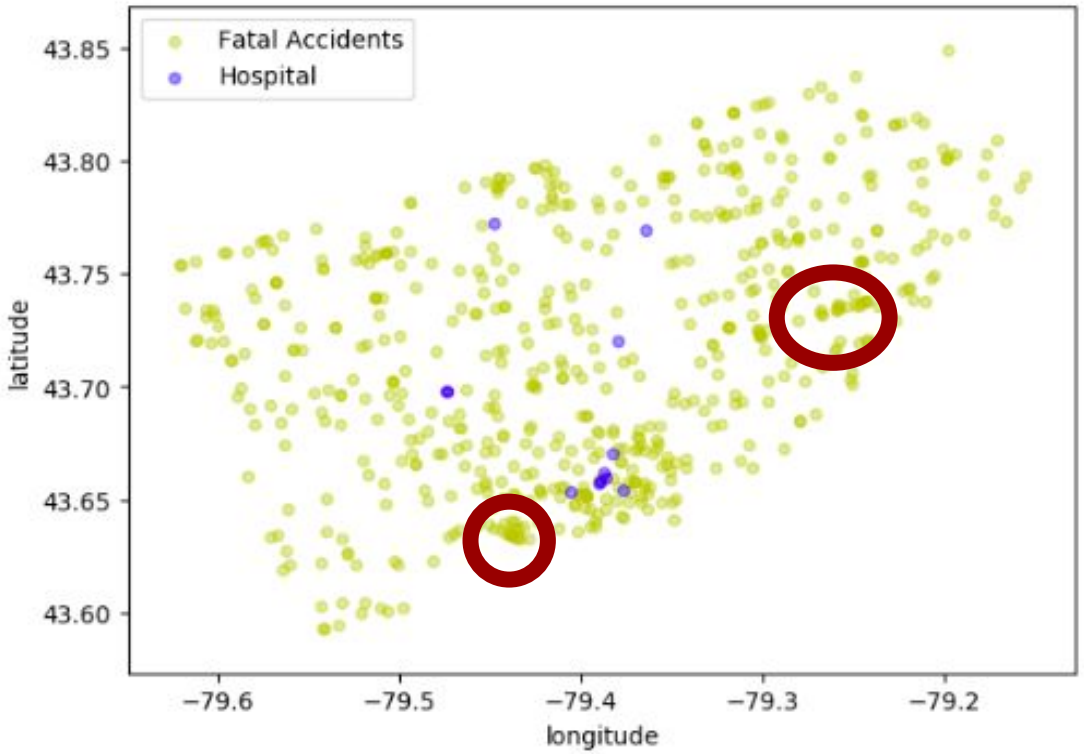


Hospital



04 Proposed Solution





Police Division

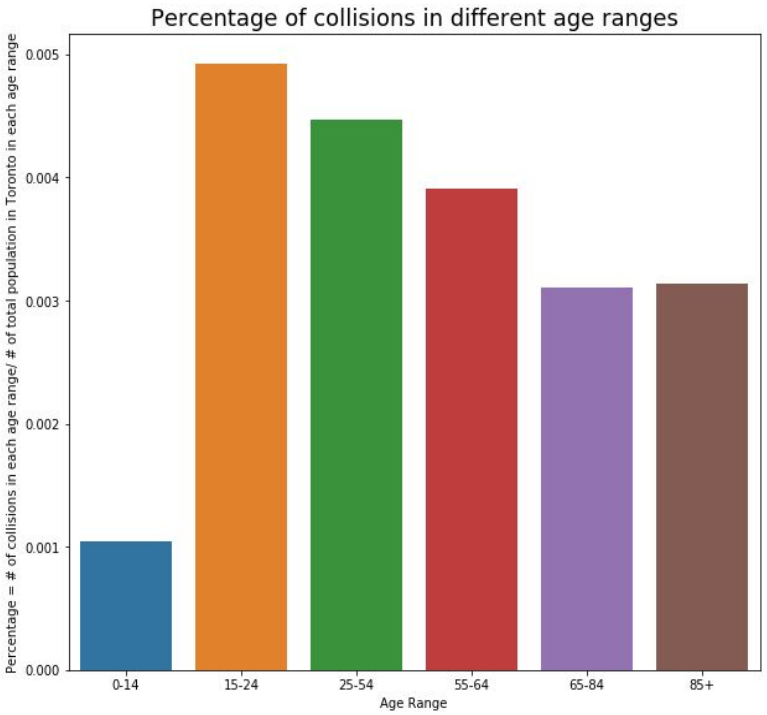
Solution: Build new police stations in the orange circle area.

Hospital

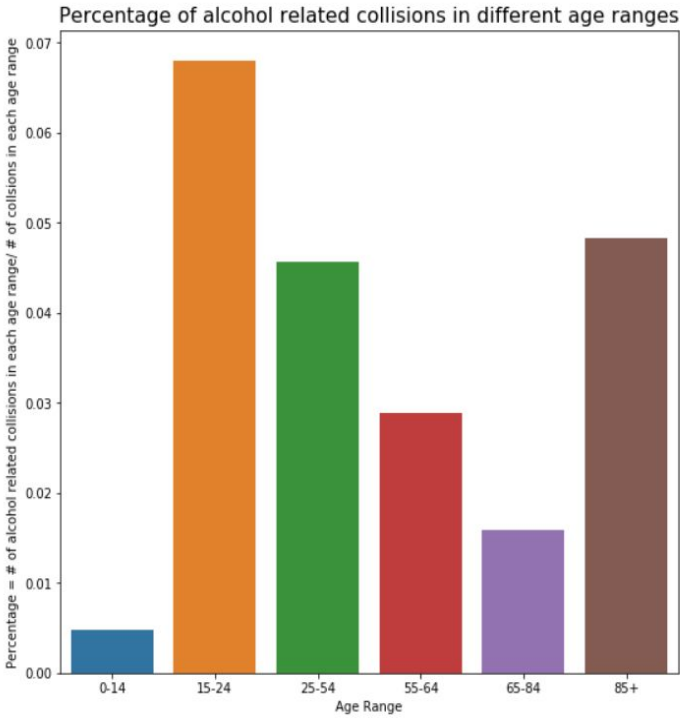
Solution: Build new hospitals in the red circle area.

04 Problem Analysis

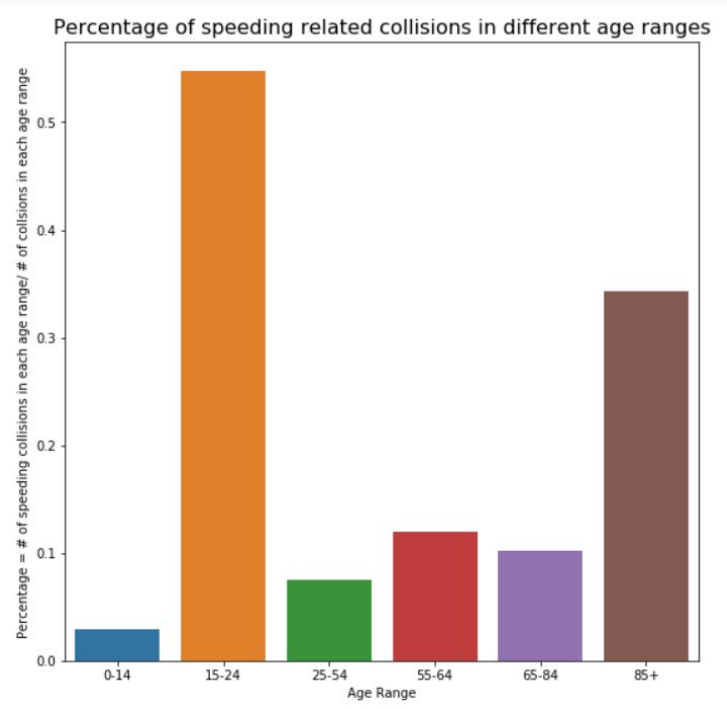
Different Age Ranges



- Percentage of collisions



- Percentage of alcohol related collisions



- Percentage of speeding related

04 Proposed Solution

BAC Level < 0.05

Increase Driving license min-age from 16 to 18

18-month restricted permit for 15- 24 years drivers that allowed no driving from 22:00 to 05:00

Thank You !