Battle of Neighborhood

Suitable neighborhoods to open coffee shops

Outline

- Business Drive
- Data Source
- Methodology
- Results
- Discussion
- Conclusion

Business Drive

Introduction/Business Problem

- Coffee shop is one of the most common and popular shops in any city.
- The demand of coffee shops is large and neighborhood dependent
- Select a suitable neighborhood to open a coffee shop is the first and most fundamental step for either a chain enterprise or a personal start-up
- The objective of this project is to screen out the suitable neighborhoods to open coffee shops based on the neighborhoods' characteristics

Data Source

Foursquare location data

- The target city is Toronto
- The data of shops in the neighborhoods of Toronto is fetched from Foursquare
- The commercial ecosystem is largely reflected by the shops in the neighborhoods, especially mature neighborhoods/cities
- The coffee shop requirements will be predicted based on the existing shops in the neighborhoods

Methodology

Data Import and Processing

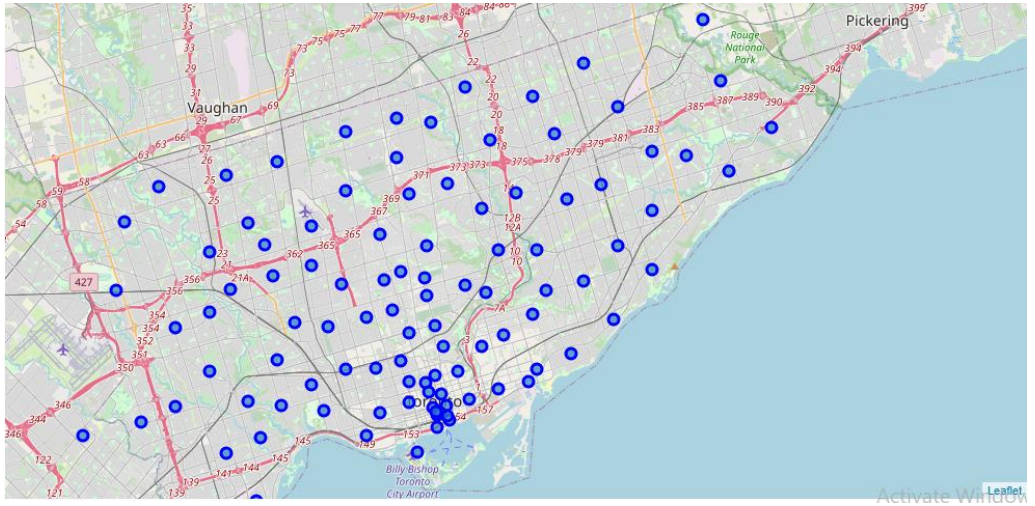
Neighborhood information get from Wiki

Postcode	Borough	Neighbourhood
M1A	Not assigned	Not assigned
M2A	Not assigned	Not assigned
M3A	North York	Parkwoods
M4A	North York	Victoria Village
M5A	Downtown Toronto	Harbourfront
M5A	Downtown Toronto	Regent Park
M6A	North York	Lawrence Heights
M6A	North York	Lawrence Manor
M7A	Queen's Park	Not assigned
M8A	Not assigned	Not assigned
M9A	Etobicoke	Islington Avenue

https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

Neighborhood table/location after processing

	PostalCode	Borough	Neighbourhood
0	M3A	North York	Parkwoods
1	M4A	North York	Victoria Village
2	M5A	Downtown Toronto	Regent Park, Harbourfront
3	M6A	North York	Lawrence Manor, Lawrence Heights
4	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government



Methodology_Cont'd

Get Venues Information from Foursquare

Foursquare API utilized to obtain the venues for each neighborhood around the specific location

	PostalCode	PostalZone Latitude	PostalZone Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	M3A	43.753259	-79.329656	Brookbanks Park	43.751976	-79.332140	Park
1	M3A	43.753259	-79.329656	Variety Store	43.751974	-79.333114	Food & Drink Shop
2	M4A	43.725882	-79.315572	Victoria Village Arena	43.723481	-79.315635	Hockey Arena
3	M4A	43.725882	-79.315572	Portugril	43.725819	-79.312785	Portuguese Restaurant
4	M4A	43.725882	-79.315572	Tim Hortons	43.725517	-79.313103	Coffee Shop

```
toronto_venues.shape
```

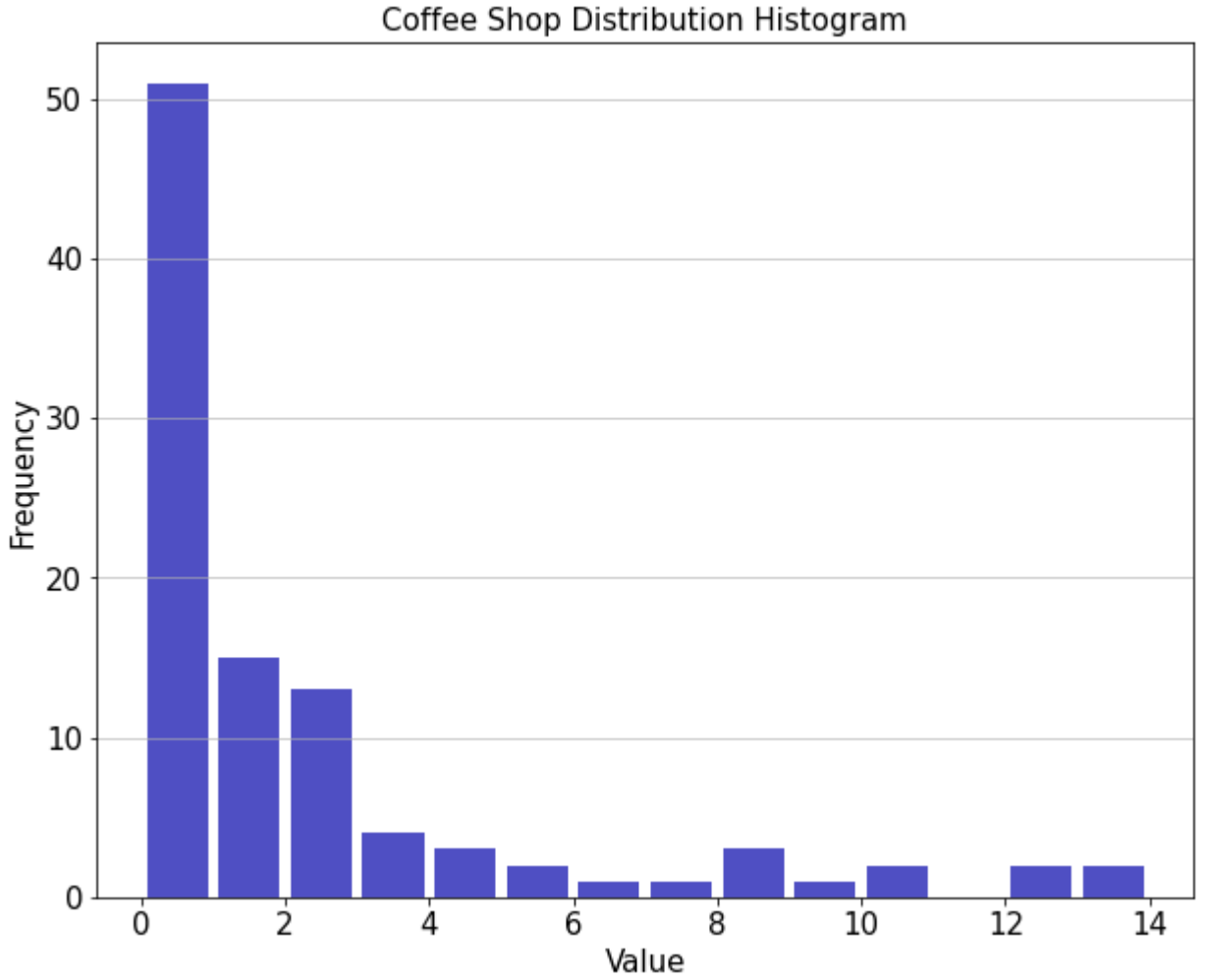
```
(2166, 7)
```

- One hot encoding to create the data frame for analysis
- A total of 277 unique venues/shops is listed

Methodology_Cont'd

Existing Coffee Shop Histogram

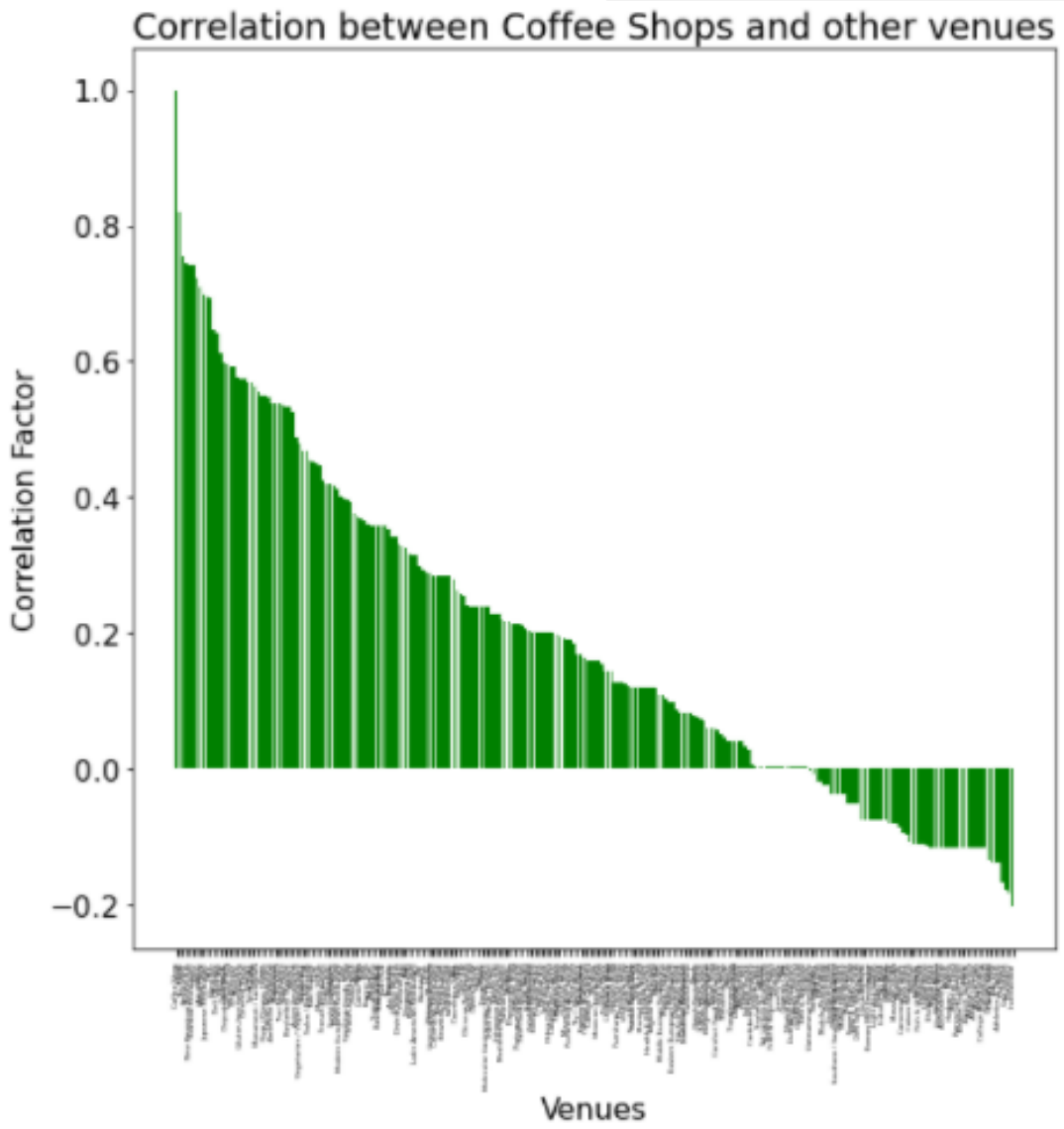
- The chart shows the histogram of the existing coffee shops in Toronto
- As it indicates, there are large number of neighborhoods which don't have coffee shop
- The dataset is splited into two sets:
  1. The neighborhoods **with** coffee shops (used for model building)
  2. The neighborhoods **without** coffee shops (used for candidate selection)



Methodology_Cont'd

Correlation Matrix

- The chart shows the correlation between coffee shop numbers with the other venues
- There are a total of 244 categories having correlation with coffee shops
- For the purpose of simplification and negative correlation with some venues, all these venues are taken into account



Methodology_Cont'd

ML Regression Models

6 Models are evaluated in order to have a higher confidence prediction

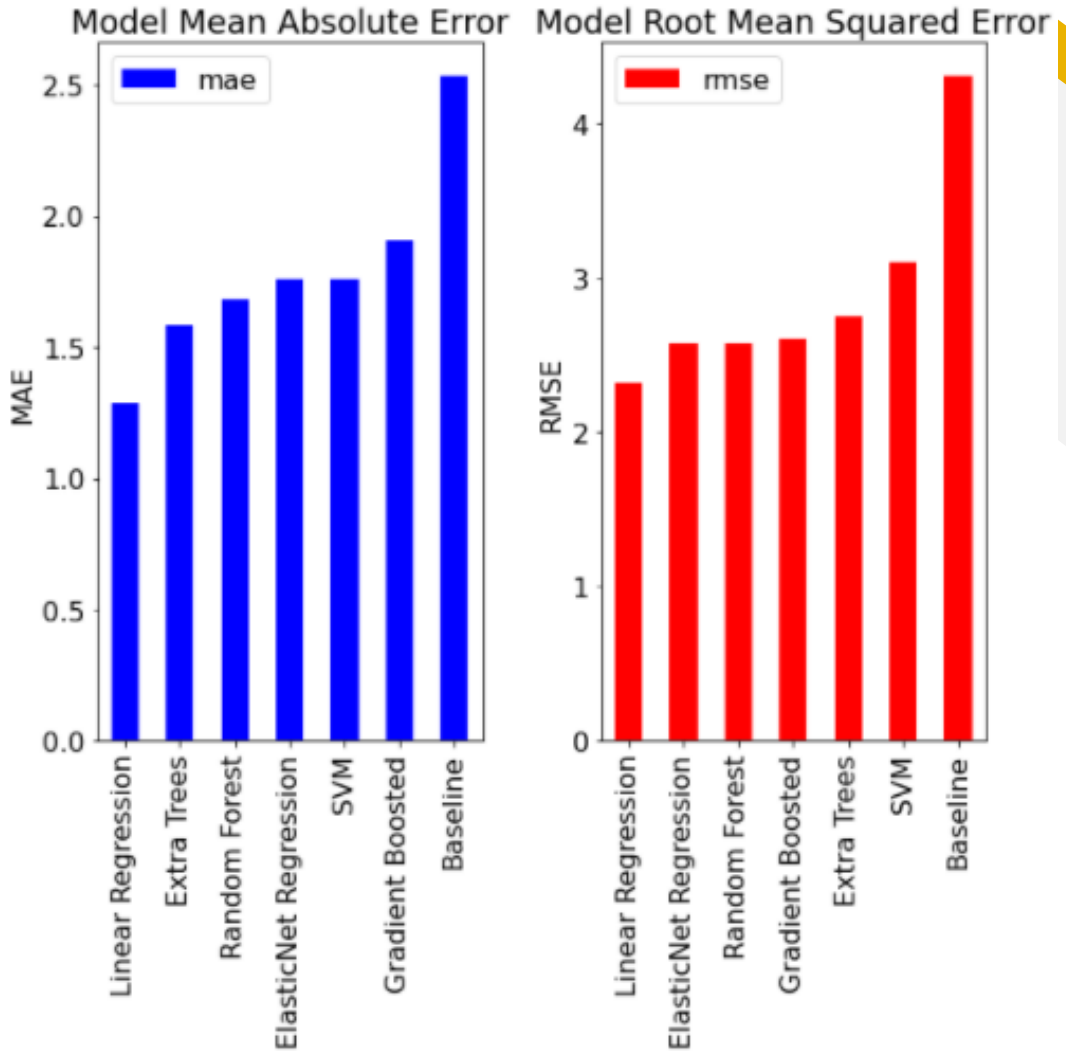
```
model1 = LinearRegression()
model2 = ElasticNet(alpha=1.0, l1_ratio=0.5)
model3 = RandomForestRegressor(n_estimators=50)
model4 = ExtraTreesRegressor(n_estimators=50)
model5 = SVR(kernel='rbf', degree=3, C=1.0, gamma='auto')
model6 = GradientBoostingRegressor(n_estimators=20)
```

- Training and Testing data split: 70/30

Results

- MAE and RMSE utilized to compare the model accuracy
- The evaluation shows that Linear Regression has the best accuracy

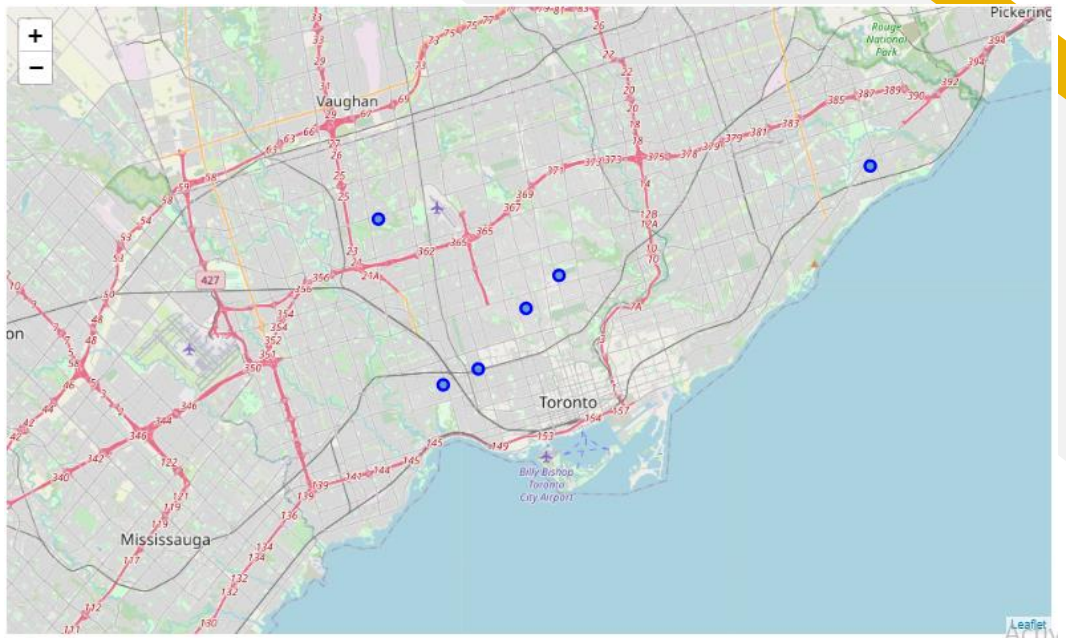
	mae	rmse
Linear Regression	1.28699	2.3239
ElasticNet Regression	1.75553	2.56829
Random Forest	1.68267	2.56917
Extra Trees	1.58133	2.74418
SVM	1.75892	3.09648
Gradient Boosted	1.90635	2.60797
Baseline	2.53333	4.30504



Discussion

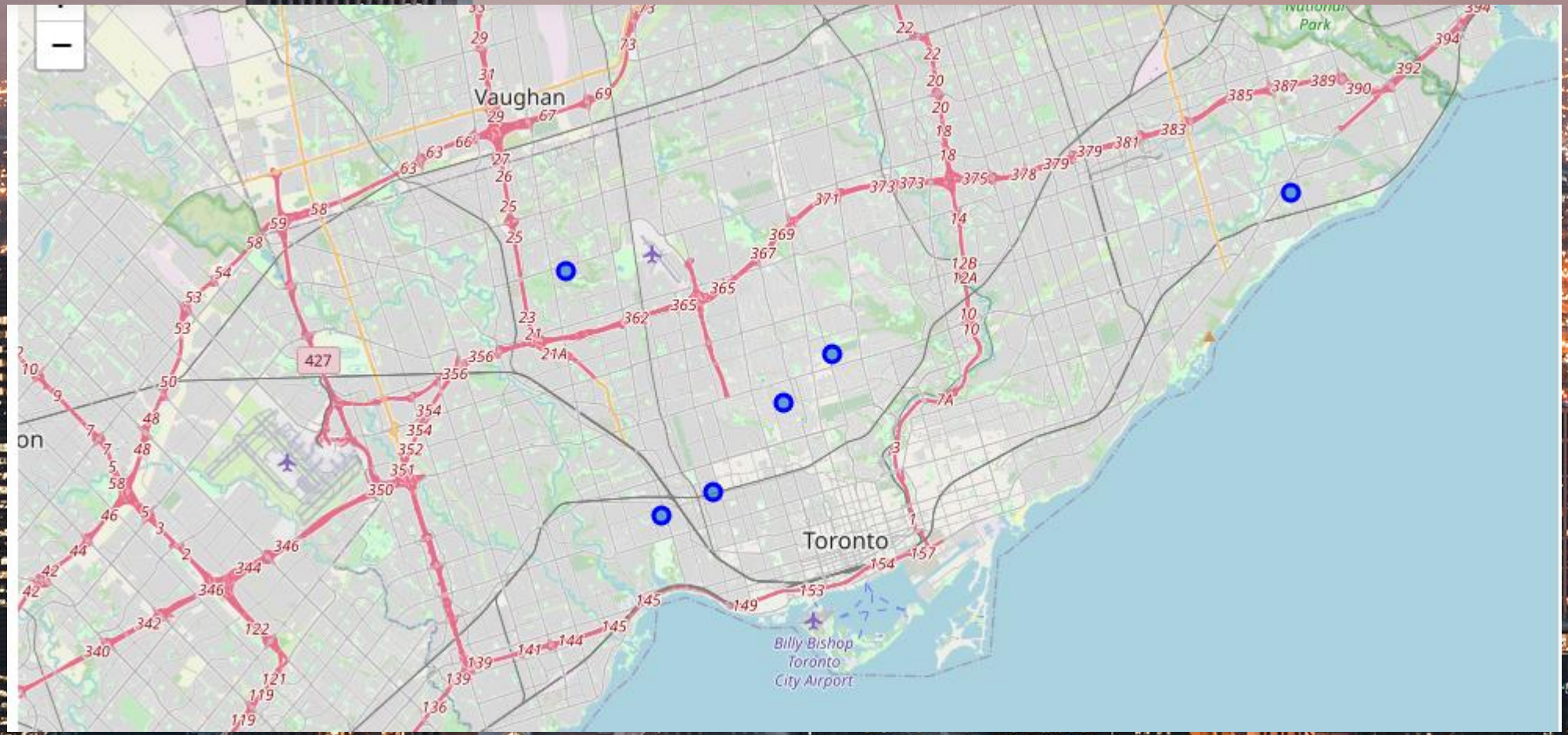
Listing the best neighborhoods

- The Best neighborhoods (top 6) are listed
- Based on the commercial ecosystem, each of these neighborhoods is predicted to be suitable to have at least two coffee shops
- The analysis provides a good insight for store location
- Indeed, the best place to open a coffee shop should also taking into account time-lapse data of shops, rental, etc.



	PostalCode	Linear Regression	ElasticNet Regression	Random Forest	Extra Trees	SVM	Gradient Boosted	Borough	Neighbourhood	Latitude	Longitude
0	M6P	3.897320	2.641193	2.10	2.36	2.390410	3.555138	West Toronto	High Park, The Junction South	43.661608	-79.464763
1	M1E	2.576944	2.274831	1.70	1.82	1.788216	1.859084	Scarborough	Guildwood, Morningside, West Hill	43.763573	-79.188711
2	M6H	2.160494	2.497412	1.86	2.02	1.964447	2.959219	West Toronto	Dufferin, Dovercourt Village	43.669005	-79.442259
3	M4P	2.138635	3.188268	4.38	1.66	1.852964	5.928621	Central Toronto	Davisville North	43.712751	-79.390197
4	M5P	2.118647	2.237996	1.62	1.84	1.698490	1.859084	Central Toronto	Forest Hill North & West, Forest Hill Road Park	43.696948	-79.411307
5	M3L	2.025981	3.188268	4.20	1.58	1.759907	6.124065	North York	Downsview	43.739015	-79.506944

6 Best Neighborhood to Open a Coffee Shop



Conclusion

- The project predicts the best neighborhoods to open a coffee shop in Toronto
- Foursquare API is utilized to obtain all the venues in the neighborhoods
- Multiple ML models (6 models) utilized to evaluate the predictions
- Top 6 locations indicates good potential to open the coffee shop

Way Forward for Real World Location Selection:

- More data should be incorporated, including time-lapse data of the shops, rental information, policy/government information, etc.
- More detailed location analysis (block or street level)

Thank You

Battle of Neighborhood

https://github.com/WendelZhao/Coursera_Capstone/blob/main/Battle_of_Neighborhoods.ipynb