

The Battle of Neighborhoods

Introduction

The problem we want to address is to identify the location that is suitable to open a new restaurant in the City of Toronto.

Data

The data we shall use is the neighborhoods of Toronto (from https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M (https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M)); the location information for each neighborhoods (from `geocoder`) and the venues information for each neighborhoods (from *FourSquare*).

We shall fit the Multi-dimensional Linear Regression where the features are the number of venues which is not a restaurant; and the dependent variable is the number of venues which is a restaurant. Then we use this model to predict the expected number of restaurants for each neighbourhood. If the actual value is lower than the expected value, then we shall consider it is a good location to open a new restaurant as there are still markets in that neighbourhood. In contrast, if their actual value is larger than expected value, then we shall consider it is a bad location as the competition is too high in that neighbourhood. Then we can sort the result to see which location is good to open a new restaurant.

Let's first fetch the neighborhood data from Wikipedia and clean it.

(103, 3)

	Postal Code	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

Then we shall fetch the location data. Since `geocoder` is unstable, we shall use the pre-fetch csv file.

(103, 3)

	Postal Code	Latitude	Longitude
0	M1B	43.806686	-79.194353
1	M1C	43.784535	-79.160497
2	M1E	43.763573	-79.188711
3	M1G	43.770992	-79.216917
4	M1H	43.773136	-79.239476

In the next steps, we can fetch venues data using FourSquare API.

(2153, 7)

	Postal Code	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	M1B	43.806686	-79.194353	Wendy's	43.807448	-79.199056	Fast Food Restaurant
1	M1B	43.806686	-79.194353	Interprovincial Group	43.805630	-79.200378	Print Shop
2	M1C	43.784535	-79.160497	Royal Canadian Legion	43.782533	-79.163085	Bar
3	M1E	43.763573	-79.188711	RBC Royal Bank	43.766790	-79.191151	Bank
4	M1E	43.763573	-79.188711	G & G Electronics	43.765309	-79.191537	Electronics Store

Let's convert it to one-hot encoding and sum them up.

(100, 270)

	Postal Code	Accessories Store	Afghan Restaurant	Airport	Airport Food Court	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	Ant S
0	M1B	0	0	0	0	0	0	0	0	
1	M1C	0	0	0	0	0	0	0	0	
2	M1E	0	0	0	0	0	0	0	0	
3	M1G	0	0	0	0	0	0	0	0	
4	M1H	0	0	0	0	0	0	0	0	

5 rows × 270 columns

Let's sum any columns contains Food or Restaurant and make it into a new column Restaurant

(100, 219)

	Postal Code	Accessories Store	Airport	Airport Lounge	Airport Service	Airport Terminal	Antique Shop	Aquarium	Art Gallery	Art Museum
0	M1B	0	0	0	0	0	0	0	0	0
1	M1C	0	0	0	0	0	0	0	0	0
2	M1E	0	0	0	0	0	0	0	0	0
3	M1G	0	0	0	0	0	0	0	0	0
4	M1H	0	0	0	0	0	0	0	0	0

5 rows × 219 columns

Let's merge above data to finish data preparation.

(103, 223)

	Postal Code	Borough	Neighbourhood	Latitude	Longitude	Accessories Store	Airport	Airport Lounge	Airport Service
0	M3A	North York	Parkwoods	43.753259	-79.329656	0.0	0.0	0.0	0.0
1	M4A	North York	Victoria Village	43.725882	-79.315572	0.0	0.0	0.0	0.0
2	M5A	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636	0.0	0.0	0.0	0.0
3	M6A	North York	Lawrence Manor, Lawrence Heights	43.718518	-79.464763	1.0	0.0	0.0	0.0
4	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government	43.662301	-79.389494	0.0	0.0	0.0	0.0

5 rows × 223 columns

Methodology

We shall use multi-linear regression to fit the data.

	Accessories Store	Airport	Airport Lounge	Airport Service	Airport Terminal	Antique Shop	Aquarium	Art Gallery	Art Museum	Arts & Crafts Store
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2	0.0	0.0	0.0	0.0	0.0	1.0	0.0	1.0	0.0	0.0
3	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0

5 rows × 217 columns

```
0    1.0
1    1.0
2    5.0
3    1.0
4    2.0
```

Name: Restaurant, dtype: float64

```
0    0.214769
1    0.985277
2    4.999925
3    0.840990
4    1.983324
dtype: float64
```

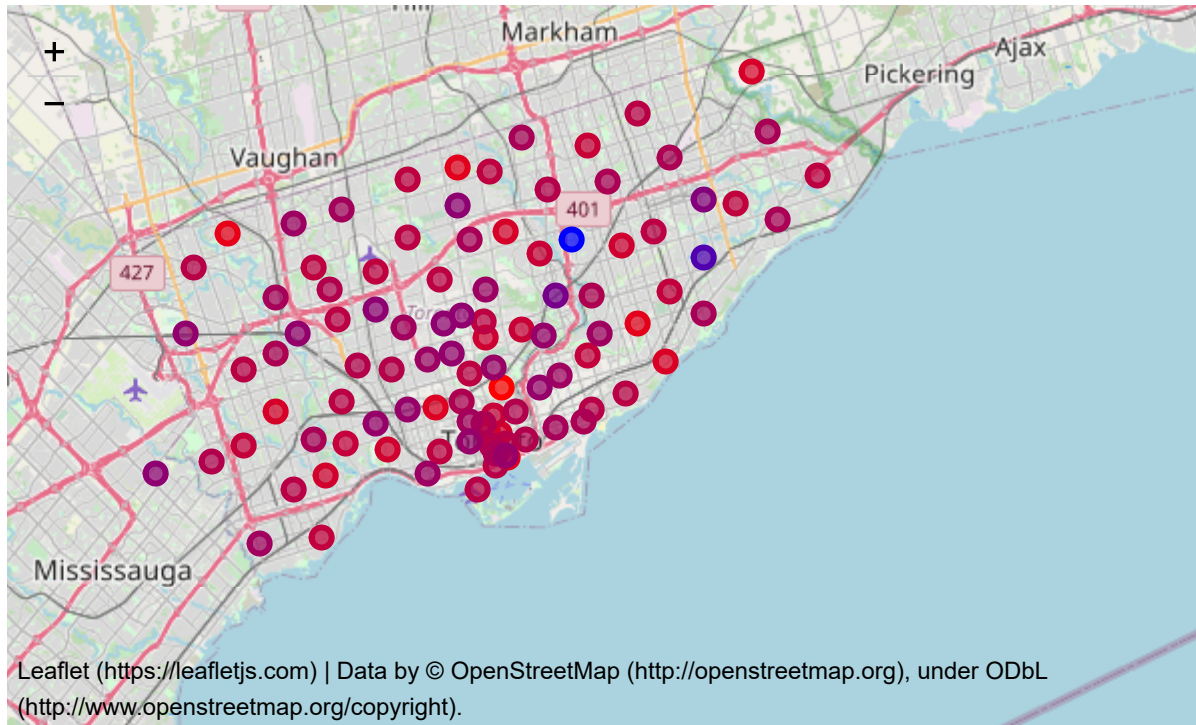
Result

Let's see the capacity of opening a new restaurant for each neighbor.

	Postal Code	Borough	Neighbourhood	Latitude	Longitude	Accessories Store	Airport	Airport Lounge	A S
91	M4W	Downtown Toronto	Rosedale	43.679563	-79.377529	0.0	0.0	0.0	
50	M9L	North York	Humber Summit	43.756303	-79.565963	0.0	0.0	0.0	
97	M5X	Downtown Toronto	First Canadian Place, Underground city	43.648429	-79.382280	0.0	0.0	0.0	
44	M1L	Scarborough	Golden Mile, Clairlea, Oakridge	43.711112	-79.284577	0.0	0.0	0.0	
20	M5E	Downtown Toronto	Berczy Park	43.644771	-79.373306	0.0	0.0	0.0	
...	
3	M6A	North York	Lawrence Manor, Lawrence Heights	43.718518	-79.464763	1.0	0.0	0.0	
26	M1H	Scarborough	Cedarbrae	43.773136	-79.239476	0.0	0.0	0.0	
13	M3C	North York	Don Mills	43.725900	-79.340923	0.0	0.0	0.0	
32	M1J	Scarborough	Scarborough Village	43.744734	-79.239476	0.0	0.0	0.0	
0	M3A	North York	Parkwoods	43.753259	-79.329656	0.0	0.0	0.0	

103 rows × 225 columns

We can see the map below, where red means good location and blue means bad location.



Discussion

As we can see above, M4W is a great place to open a new restaurant, and M9L, M5X, M1L, M5E, M2M can also be good alternatives. However, M1J and M3A are not suitable to open a new restaurant due to its competition.

Conclusion

Thus, the recommendation is to open a new restaurant at neighborhood M4W.