Business Location Explore in Toronto

Bin Yu

2020.04.21

Peer-graded Assignment for Course Applied Data Science Capstone, Week 4/5

Business Location Explore in Toronto

Contents

1. Introduction

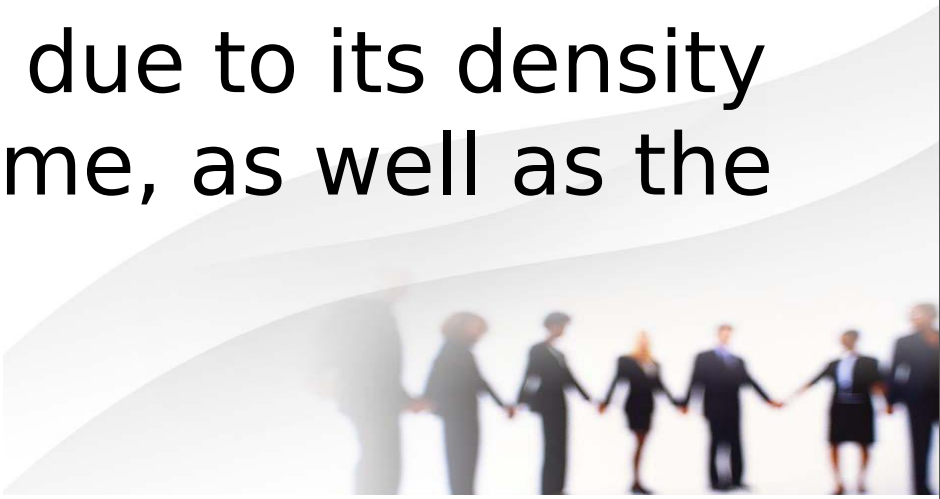
- 1.1 Business Requests
- 1.2 Analytic Approach

1.1 Business Requests

- Generate The location problem
- Location to start a new business

We assume:

An investor wants to start a new business to serve Chinese food in downtown Toronto due to its density of population, higher average income, as well as the diversity of culture.



1.2 Analytic Approach

A good location should satisfy the two criteria at the same time:

- (1) Sufficient demand
- (2) Insufficient support

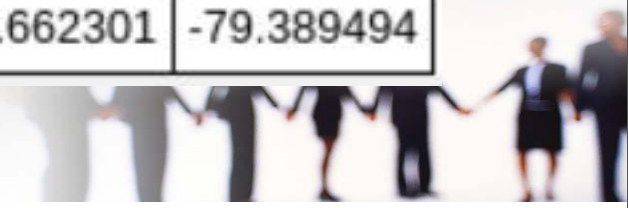
2. Data Collection

- 2.1 Scrape location info in Toronto
- 2.2 Fetch all 'FOOD' venues in Toronto
- 2.3 Show venues on map

2.1 Scrape location info in Toronto

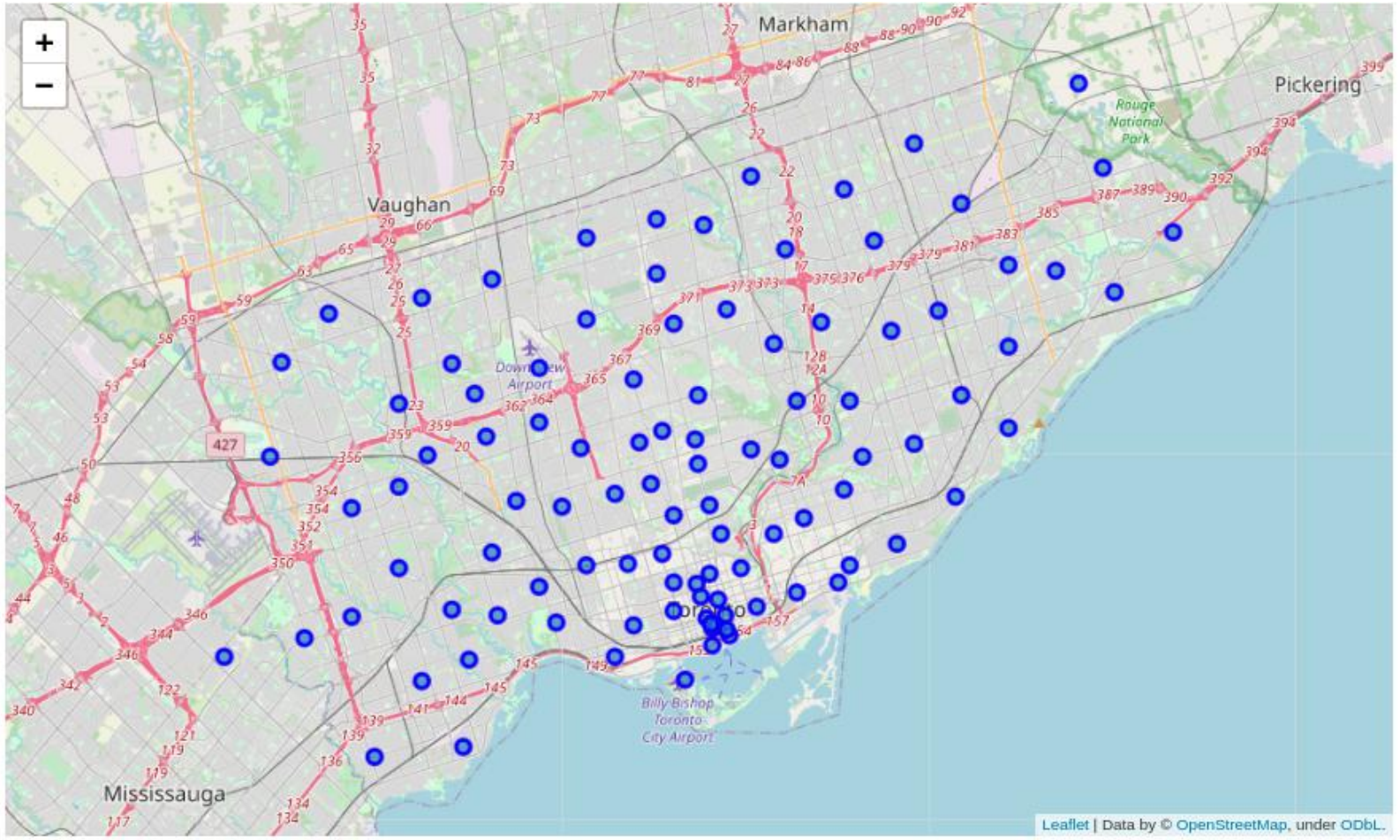
- Step(1) Fetch postal code in Toronto
- Step(2) Attaching geo info for each postal code

	Borough	Neighborhood	Latitude	Longitude
Postal Code				
M3A	North York	Parkwoods	43.753259	-79.329656
M4A	North York	Victoria Village	43.725882	-79.315572
M5A	Downtown Toronto	Regent Park / Harbourfront	43.654260	-79.360636
M6A	North York	Lawrence Manor / Lawrence Heights	43.718518	-79.464763
M7A	Downtown Toronto	Queen's Park / Ontario Provincial Government	43.662301	-79.389494



2.1 Scrape location info in Toronto

- Step (3) Visualization areas on map



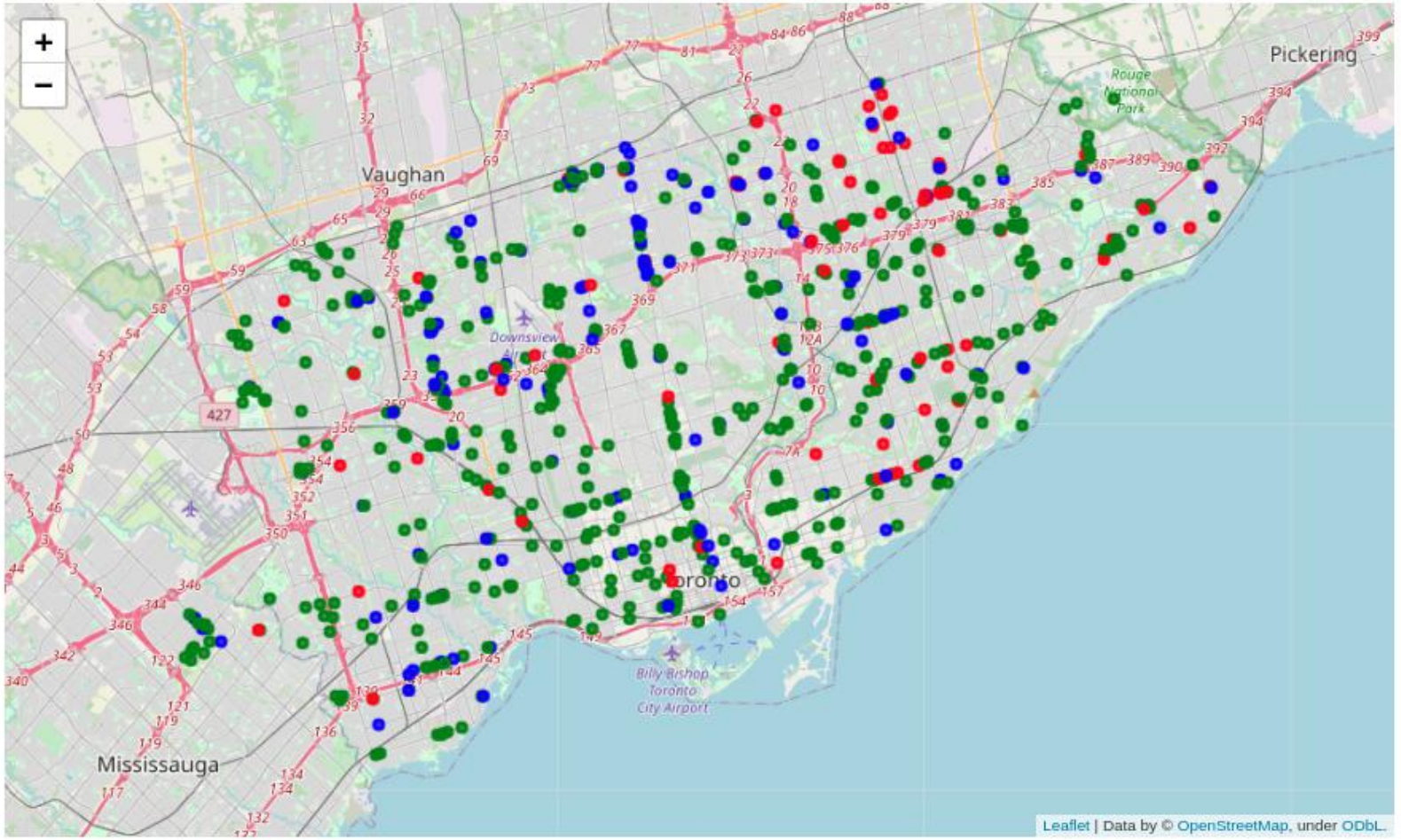
2.2 Fetch all 'FOOD' venues in Toronto

- Step(1) First we set API credentials
- Step(2) Fetch data via FourSquare API
- Step(3) Review venues contains Restaurant in their Name

2.3 Show venues on map

- Step (1) Find the center of the map
- Step (2) Mark the venues on maps

Mark the Restaurants on maps



3. Methodology

- 3.1 Definition of Sufficiency and InSufficiency
- 3.2 Isolation Forest Anomaly Dection

3.1 Definition of Sufficiency and InSufficiency

Sufficient Demand :

a bigger average restaurant provider over the area

Insufficient Support:

a smaller average restaurant business over the area

3.2 Isolation Forest Anomaly Dection

- Step (1) Define and Fit the model

```
IF_cols = ['Avg_ChineseR','Avg_AsianR','Avg_OtherR','Avg_AnyR']

model=IsolationForest( n_estimators=50, max_samples='auto', contamination=float(0.1),
max_features=4 )

model.fit( df_avg[])
```

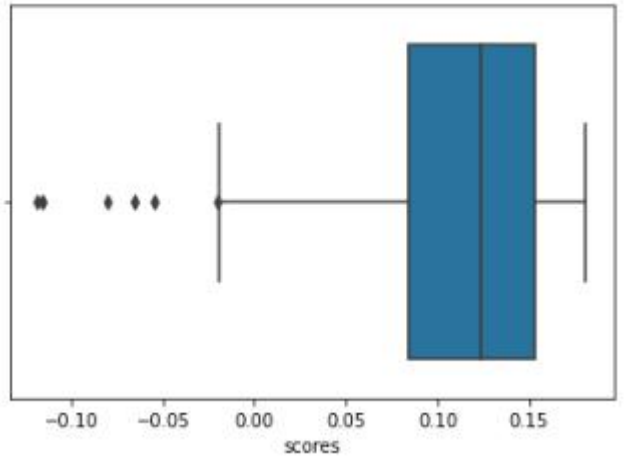
3.2 Isolation Forest Anomaly Dection

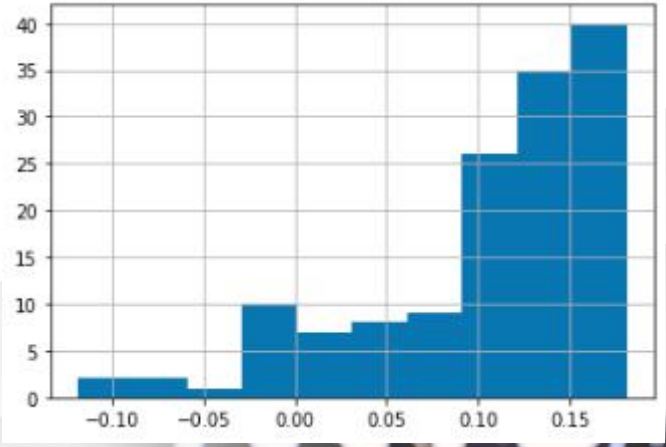
- Step (2) Attached Scores and Anomaly Column

```
result_cols = ['AREA_ID', 'scores','anomaly', 'NAME', 'LATITUDE','LONGITUDE' ]
df_avg['scores'] = model.decision_function(df_avg[IF_cols])
df_avg['anomaly']=model.predict(df_avg[IF_cols])
df_result = df_avg.merge( df )[result_cols]
df_result.sort_values('scores', inplace=True)
```

3.2 Isolation Forest Anomaly Dection

	AREA_ID	scores	anomaly	NAME	LATITUDE	LONGITUDE
27	4649	-0.118717	-1	North St.James Town (74)	43.669623	-79.375247
37	4660	-0.115751	-1	Regent Park (72)	43.659992	-79.360509
2	4623	-0.079746	-1	Yonge-St.Clair (97)	43.687859	-79.397871
42	4665	-0.065879	-1	Rouge (131)	43.821201	-79.186343
40	4663	-0.055065	-1	Roncesvalles (86)	43.646123	-79.442992
...	...	...	...	...	...	...
24	4646	0.173633	1	Newtonbrook West (36)	43.785830	-79.431422
113	4739	0.174701	1	Forest Hill South (101)	43.694526	-79.414318
39	4662	0.177534	1	Rockcliffe-Smythe (111)	43.674790	-79.494420
41	4664	0.180706	1	Rosedale-Moore Park (98)	43.682820	-79.379669
9	4630	0.181015	1	Leaside-Bennington (56)	43.703797	-79.366072





4. Result

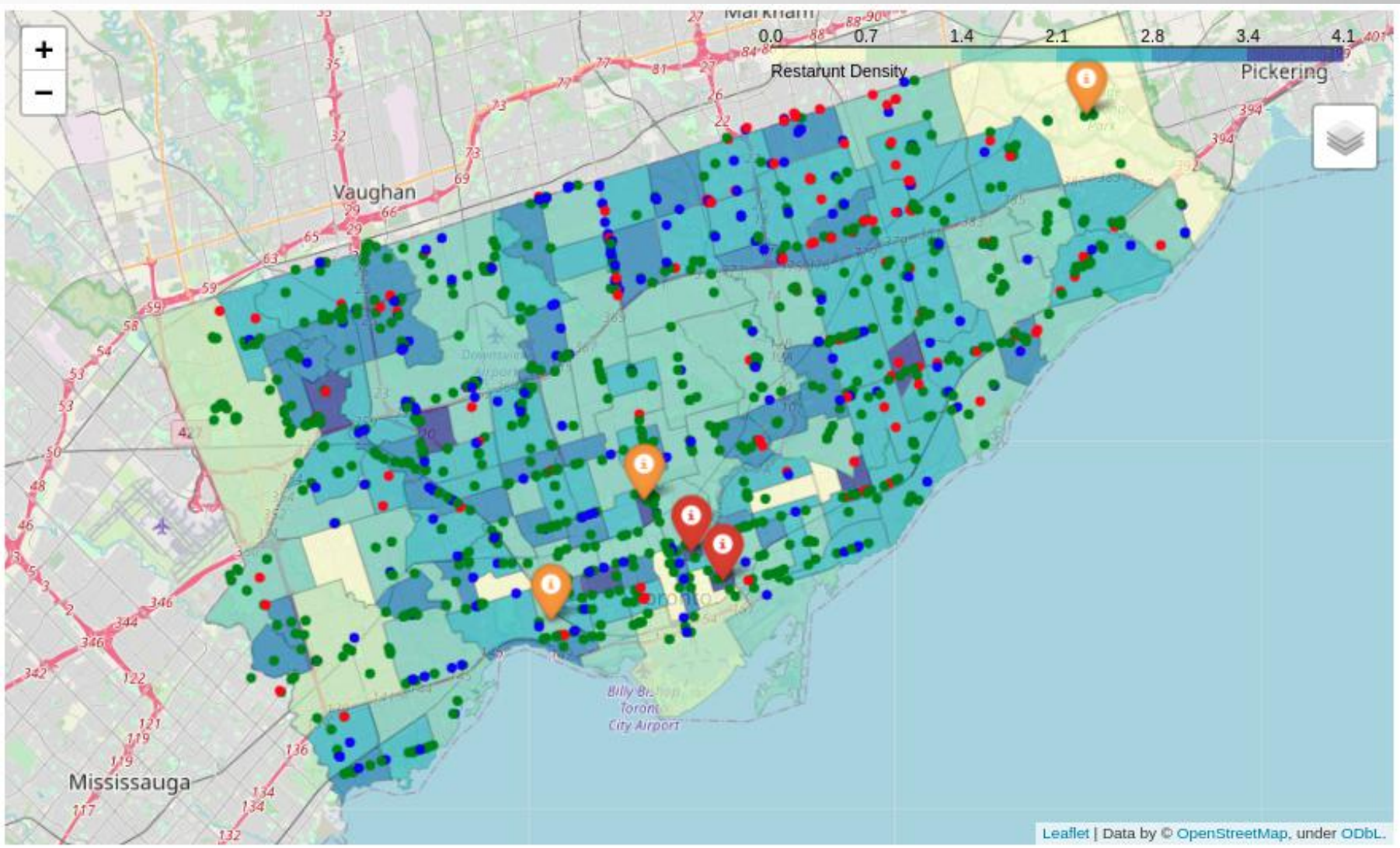
To open a new Chinese restaurant, we have two locations with potentially highest opportunites, marked as Red on above map:

- North St.James Town
- Regent Park

Plus the other three locations may also have moderate oppertunites, marked as Orange on above map:

- Yonge-St.Clair
- Rouge
- Roncesvalles

4. Result


0.0
0.7
1.4
2.1
2.8
3.4
4.1
Restarunt Density
Pickering
Vaughan
Mississauga
Leaflet | Data by © OpenStreetMap, under ODbL.

5. Discussion

- We could add more features into the model, such as the rating of the venues, size of the business, etc.
- Also, it would be better if we could fetch more data from different data sources, along with FourQuares, it may help us to build a more accurate model.
- Plus, We could introduce other dimensional data like population, Demographics, income, etc., for this information also has an impact on the consuming market.

6. Conclusion

Isolation Forest model help us to find outliers in the restaurant business, and find out location-based significantly different from those majority of the other locations

Business Location Explore in Toronto

Thank You!

This is Peer-graded Assignment for Course Applied Data Science Capstone, Week 4/5