Report

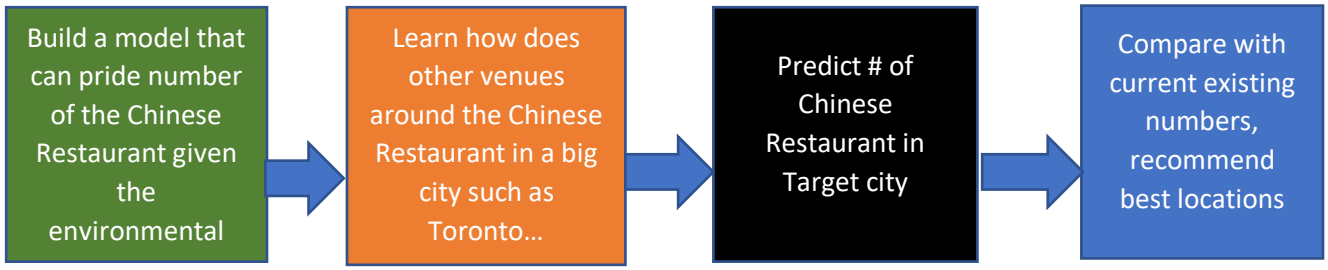
Location recommendation

Table of content

1.Introduction/Business Problem

Let say a business manager who want invest a Chinese restaurant in your resident city , You are live in the mid-size city which has fast growth, you have to decide where or which neighborhoods to open the restaurant. In order to answer this question, we can build a model get some recommendations where to start the business.

2. Solution / Methodology



3.Data Import

- Neighborhood information get from Wiki

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

- Scrap from website and organize into DataFrame

	PostalCode	Borough	Neighborhood
0	M1B	Scarborough	Rouge,Malvern
1	M1C	Scarborough	Highland Creek,Rouge Hill,Port Union
2	M1E	Scarborough	Guildwood,Morningside,West Hill
3	M1G	Scarborough	Woburn
4	M1H	Scarborough	Cedarbrae

4 .Get location information

- by use geocoder package

PostalCode	Borough	Neighborhood	Latitude	Longitude
M1B	Scarborough	Rouge,Malvern	43.806686	-79.194353
M1C	Scarborough	Highland Creek,Rouge Hill,Port Union	43.784535	-79.160497
M1E	Scarborough	Guildwood,Morningside,West Hill	43.763573	-79.188711
M1G	Scarborough	Woburn	43.770992	-79.216917
M1H	Scarborough	Cedarbrae	43.773136	-79.239476

- Plot the location using folium package

5.Get venues information

• Use FoursquareAPI, we can explore the venues around on specific location, so we could achieve venues' name and category

https://foursquare.com

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Rouge,Malvern	43.806686	-79.194353	African Rainforest Pavilion	43.817725	-79.183433	Zoo Exhibit
1	Rouge,Malvern	43.806686	-79.194353	Toronto Pan Am Sports Centre	43.790623	-79.193869	Athletics & Sports
2	Rouge,Malvern	43.806686	-79.194353	Toronto Zoo	43.820582	-79.181551	Zoo
3	Rouge,Malvern	43.806686	-79.194353	Polar Bear Exhibit	43.823372	-79.185145	Zoo
4	Rouge,Malvern	43.806686	-79.194353	penguin exhibit	43.819435	-79.185959	Zoo Exhibit

• Create one-hot encoding for each category

	Neighborhood	ATM	Afghan Restaurant	African Restaurant	Airport	Airport Lounge	American Restaurant	Aquarium	Art Gallery	Arts & Crafts Store	Asian Restaurant	Athletics & Sports	Auto Dealership
0	Adelaide,King,Richmond	0	0	0	0	0	1	1	2	1	0	0	0
1	Agincourt	0	0	0	0	0	1	0	0	1	1	0	0
2	Agincourt North,L'Amoreaux East,Milliken,Steel...	0	0	0	0	0	1	0	0	1	2	0	0
3	Albion Gardens,Beaumond Heights,Humbergate,Jam...	1	1	0	0	0	1	0	0	0	3	0	0
4	Alderwood,Long Branch	0	0	0	0	0	1	0	0	1	1	0	0

6 .Build Model for prediction

use number of venues in each neighborhood except Chinese restaurant as inputs and number of Chinese restaurants as output.

Use SVR (rbfkernel) as learning algorithm

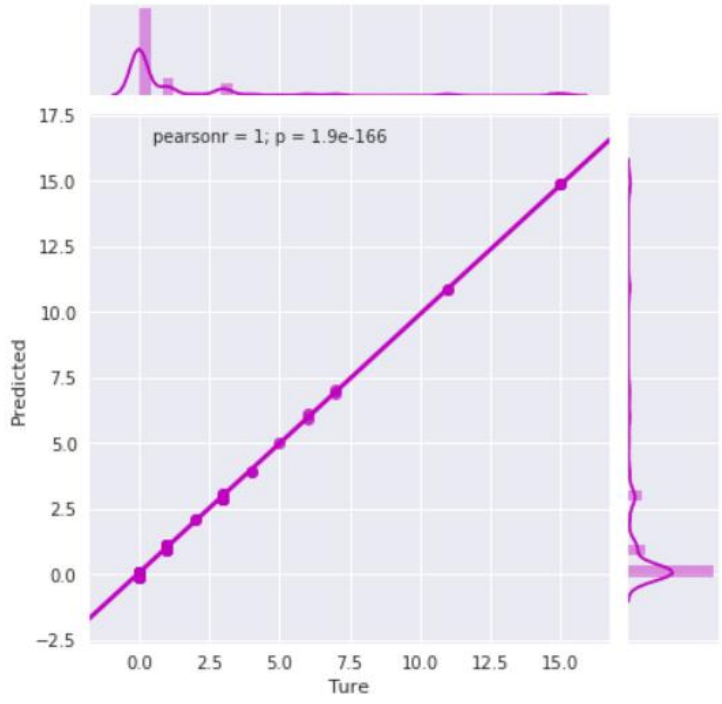
Step 1. optimize the hyperparameter using GridSearchCVon parameter ‘gamma’ and ’C’. 5 fold cross validation is used.

```
svr_rbf = GridSearchCV(SVR(kernel='rbf', gamma=0.1), cv=5,
                       param_grid={"C": [1e0, 1e1, 1e2, 1e3],
                                   "gamma": np.logspace(-2, 2, 5)})
```

Best parameter is:

```
SVR(C=100.0, cache_size=200, coef0=0.0, degree=3, epsilon=0.1, gamma=0.01,
  kernel='rbf', max_iter=-1, shrinking=True, tol=0.001, verbose=False)
```

Step 2. Train the dataset the plot prediction form the model and True value



7. Get information of target city

- Neighborhoods information is get from

http://www.city-data.com/nbmaps/neigh-Columbus-Ohio.html

And GPS location is get from

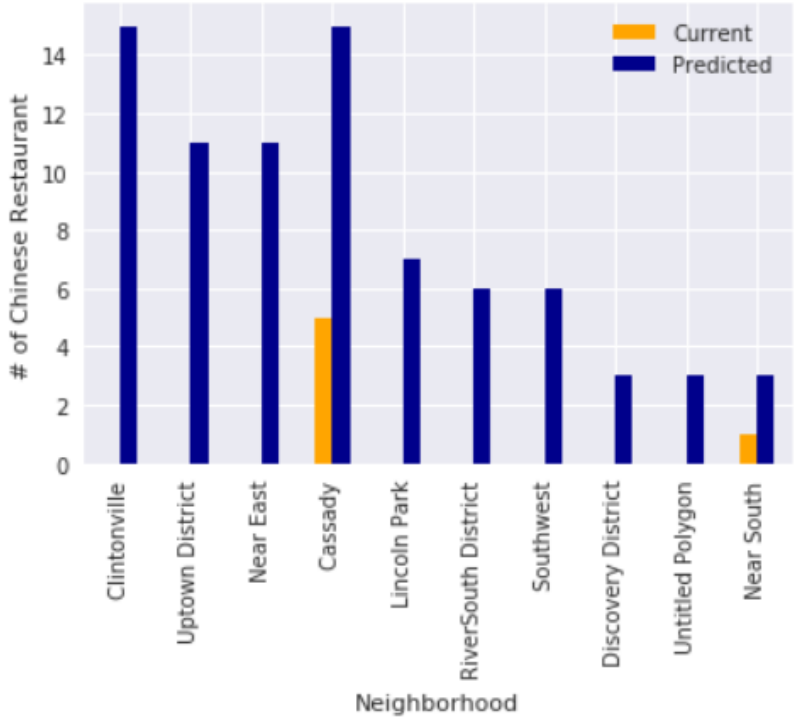
https://www.gps-coordinates.net/

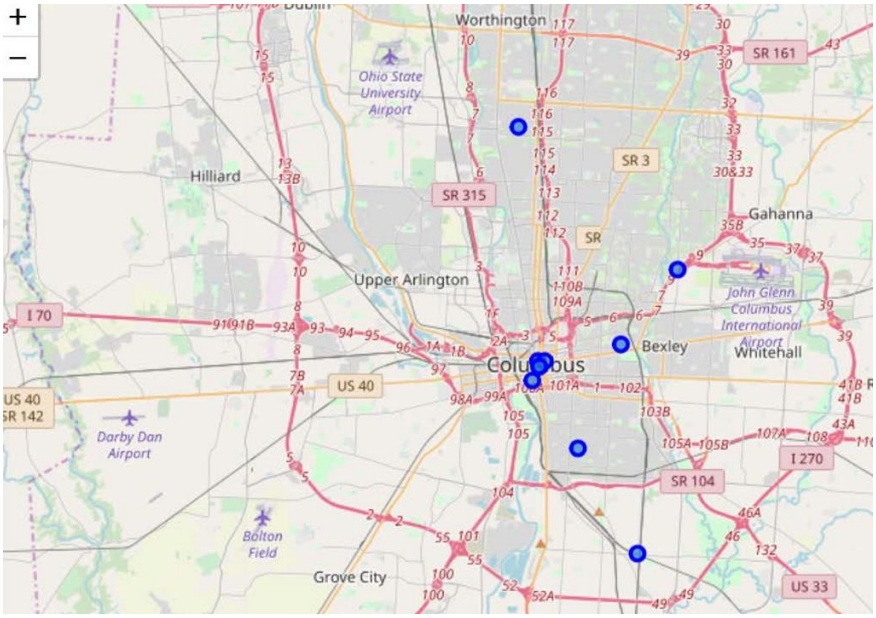
	Neighborhood	Latitude	Longitude
0	Arena District	39.968959	-83.005251
1	Brewery District	39.951159	-83.001111
2	Cassady	39.998355	-82.930182
3	Clintonville	40.052178	-83.009280
4	Discovery District	39.963340	-82.996524

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Arena District	39.968959	-83.005251	Nationwide Arena	39.969293	-83.005936	Hockey Arena
1	Arena District	39.968959	-83.005251	North Market	39.971952	-83.004338	Market
2	Arena District	39.968959	-83.005251	Hilton Columbus Downtown	39.970600	-83.002435	Hotel
3	Arena District	39.968959	-83.005251	Hot Chicken Takeover	39.971527	-83.004470	Fried Chicken Joint
4	Arena District	39.968959	-83.005251	Jeni's Splendid Ice Creams	39.971903	-83.004322	Ice Cream Shop

8. Predict using trained model

Top 10 recommendations for start your business in Columbus





9. Conclusion

- We used the Foursquare API get the venues information on given locations
- We build predictive models with SVR algorithm
- We get the top 10 recommendations of location to invest “Chinese Restaurant” in Columbus

These things can make it better:

1. This model is built on the assumption that the target city will have a trend to grow to “big-city” like we used into model training.

2. The training dataset still very small, if we can get more data from more big cities, we can make the model better

3. Foursquare app can only give 100 venues exploration on free version, it is better to conclude all of the venues to avoid bias coming from the sampling