Coursera Data Science Capstone Project

Location recommendation to start Indian Restaurant in Toronto city

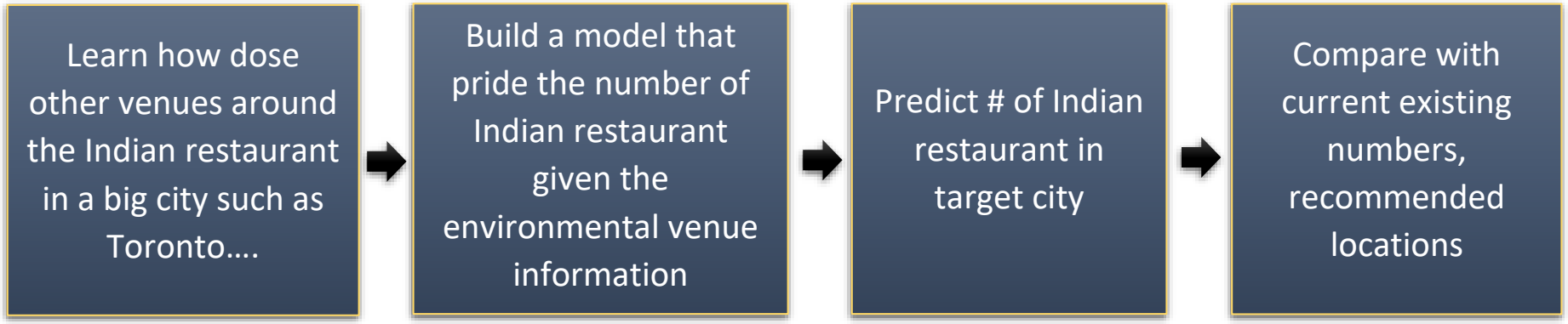
Table of content

I. Introduction/Business Problem

- Let say if you are a business manager who want to invest in Indian 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, you have to build a model get some recommendations where to start your business.

- Therefore, we will learn a model from a mature city/metropolitan city since we believe that it is more developed and your city will become a metropolis someday.
- Another thing you believe is that any one of business venue does not exist alone, and "Indian restaurant" always tends to be found with some other type of shops, because neighborhood's have "cultures" to like them both.

ii. Solution/Methodology



Data Import (Toronto)

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

Neighborhood information get from Wiki

WIKIPEDIA
The Free Encyclopedia

Main page
Contents
Current events
Random article
About Wikipedia
Contact us
Donate

Contribute
Help
Learn to edit
Community portal
Recent changes
Upload file

Tools
What links here
Related changes
Special pages
Permanent link
Page information

Not logged in

Article | Talk | Read | Edit | View history | Search Wikipe

List of postal codes of Canada: M

From Wikipedia, the free encyclopedia

This is a list of postal codes in Canada where the first letter is M. Postal codes beginning with M are located within the city of Toronto in the province of Ontario. Only the first three character Forward Sortation Area.

Canada Post provides a free postal code look-up tool on its website,[1] via its applications for such smartphones as the iPhone and BlackBerry,[2] and sells hard-copy directories and CD-RO validation tools, which allow customers to properly match addresses and postal codes. Hard-copy directories can also be consulted in all post offices, and some libraries.

Toronto - 103 FSAs [edit]

Note: There are no rural FSAs in Toronto, hence no postal codes should start with M0. However, the postal code M0R 8T0 is assigned to an Amazon warehouse in Mississauga, suggesting reserved the M0 FSA for high volume addresses.

Postal Code	Borough	Neighbourhood
M1A	Not assigned	Not assigned
M2A	Not assigned	Not assigned
M3A	North York	Parkwoods
M4A	North York	Victoria Village

Scrap from website and organize into Data frame

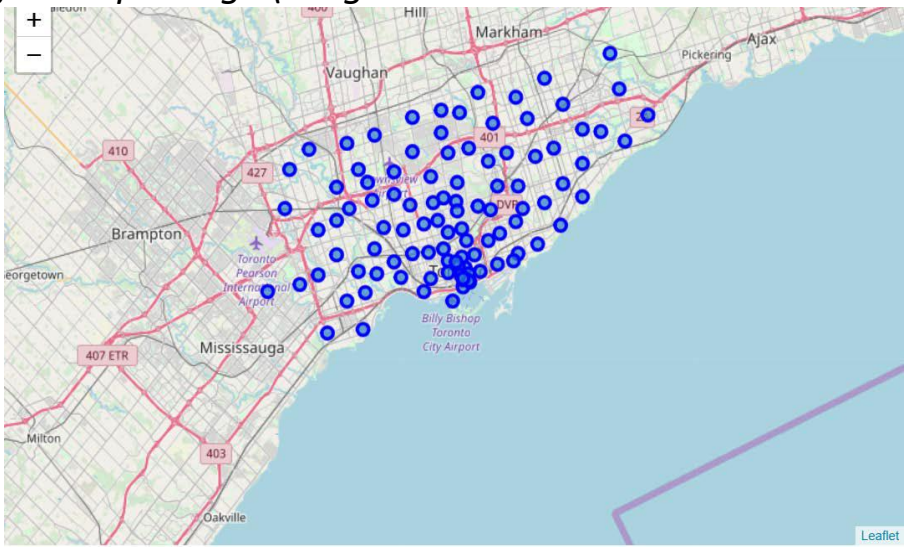
	PostalCode	Borough	Neighborhood
0	M1B	Scarborough	Rouge,Malvern
1	M1C	Scarborough	Highland Creek,Rouge Hill,Port Union
2	M1E	Scarborough	Guildwood,Morningside,West Hill
3	M1G	Scarborough	Woburn
4	M1H	Scarborough	Cedarbrae

Get location information

Use geocoder package to get location information:

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

Plot the location using folium package (Neighborhood location in Toronto on map)



Get venues information

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

	Neighbourhood	Neighbourhood Latitude	Neighbourhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Parkwoods	43.753259	-79.329656	Allwyn's Bakery	43.759840	-79.324719	Caribbean Restaurant
1	Parkwoods	43.753259	-79.329656	Graydon Hall Manor	43.763923	-79.342961	Event Space
2	Parkwoods	43.753259	-79.329656	Galleria Supermarket	43.753520	-79.349518	Supermarket
3	Parkwoods	43.753259	-79.329656	Naan & Kabob Halal	43.742903	-79.305148	Middle Eastern Restaurant
4	Parkwoods	43.753259	-79.329656	Starbucks Reserve Bar	43.735764	-79.344156	Coffee Shop

Create one-hot encoding for each catogory

Out[27]:

	Neighbourhood	Accessories Store	Afghan Restaurant	Airport	Airport Lounge	American Restaurant	Aquarium	Art Gallery	Arts & Crafts Store	Asian Restaurant	Athletics & Sports	Auto Dealership	Automotive Shop	BBQ Joint	Badminton Court	Bagel Shop	Bakery	Bank	Bar	Baseball Field	Baseball Stadium	Basketb Co
0	Agincourt	0	0	0	0	1	0	0	1	1	0	0	0	0	0	0	3	1	0	0	0	
1	Alderwood, Long Branch	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	4	0	0	0	0	
2	Bathurst Manor, Wilson Heights, Downsview North	0	0	1	0	0	0	0	1	0	2	0	0	0	0	1	3	0	0	0	0	
3	Bayview Village	0	0	0	0	1	0	0	0	0	0	1	0	1	0	3	5	2	0	1	0	
4	Bedford Park, Lawrence Manor East	0	0	1	0	0	0	0	1	0	0	0	0	1	0	0	3	0	0	0	0	

Build Model for prediction

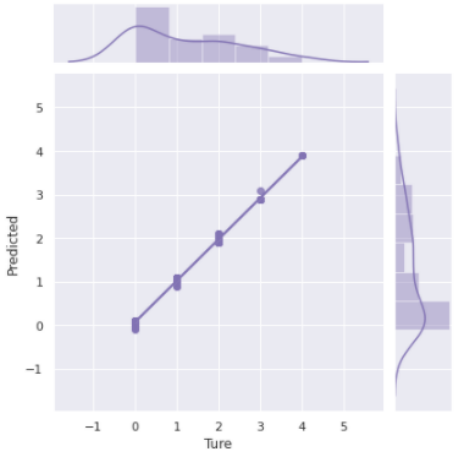
We will use number of venues in each neighborhood except Indian restaurant as inputs and number of Indian restaurants

Use SVR (rbf kernel) as learning algorithm output.

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)})
```

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



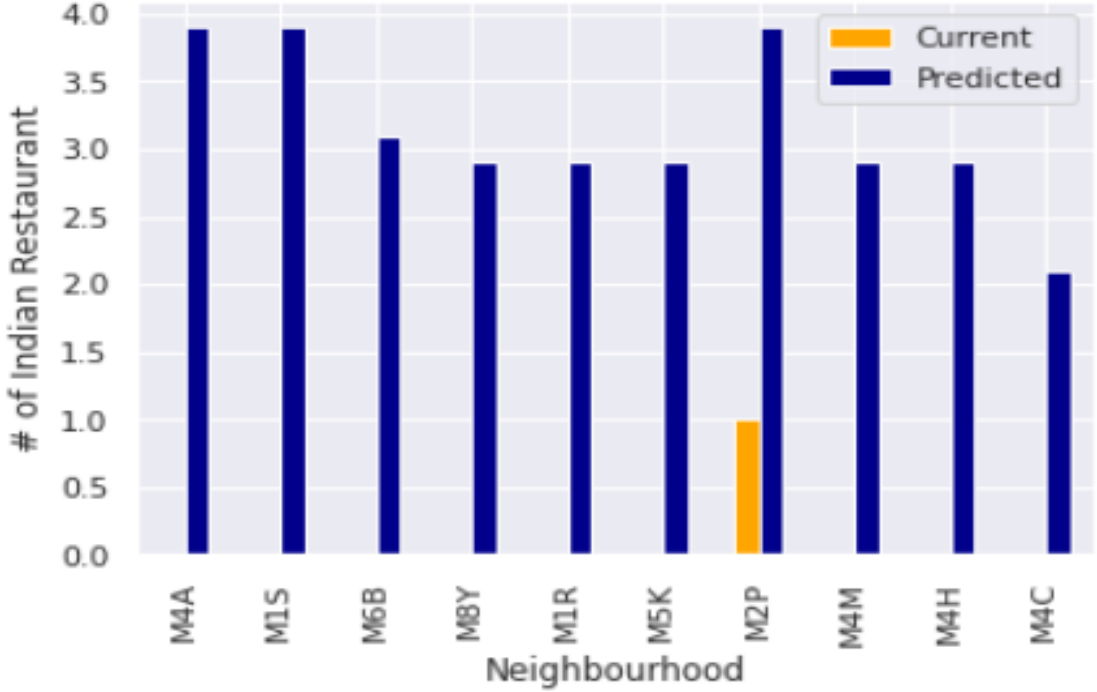
Get information of target city

Get venues information from Toronto.

	Neighbourhood	Neighbourhood Latitude	Neighbourhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	M1B	43.753259	-79.329656	Allwyn's Bakery	43.759840	-79.324719	Caribbean Restaurant
1	M1B	43.753259	-79.329656	Graydon Hall Manor	43.763923	-79.342961	Event Space
2	M1B	43.753259	-79.329656	Galleria Supermarket	43.753520	-79.349518	Supermarket
3	M1B	43.753259	-79.329656	Naan & Kabob Halal	43.742903	-79.305148	Middle Eastern Restaurant
4	M1B	43.753259	-79.329656	Starbucks Reserve Bar	43.735764	-79.344156	Coffee Shop

Predict using trained model

Top 10 recommendations locations to start Indian Restaurant in Toronto



Top 10 recommendations locations to start Indian Restaurant in Toronto (Map)

iii. Result and Discussion

- Used the Foursquare API get the venues information on given locations
- Build predictive models with SVR algorithm
- Top 10 recommendations of location to invest "Indian Restaurant" in Toronto

iv. Conclusion

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 limited venues exploration on free version, it is better to Conclude all of the venues to avoid bias coming from the sampling

v. Acknowledgement

In this project, we have to acknowledge the data science course provided by IBM powered by Coursera