# Capstone Project - Coursera

Location Recommendation for *Indian Restaurant* in New York

# Content

1. Introduction / Business Problem
2. Solution / Methodology
3. Result / Discussion
4. Conclusion
5. Acknowledgement

# 1. Introduction / Business Problem

Lets say If you are a Business Manager who want to invest for '*Indian Restaurant*' in your resident city. You are live in a city which has fast growth.

Now the question is " Where or Which *Neighborhood* to open restaurant ? "

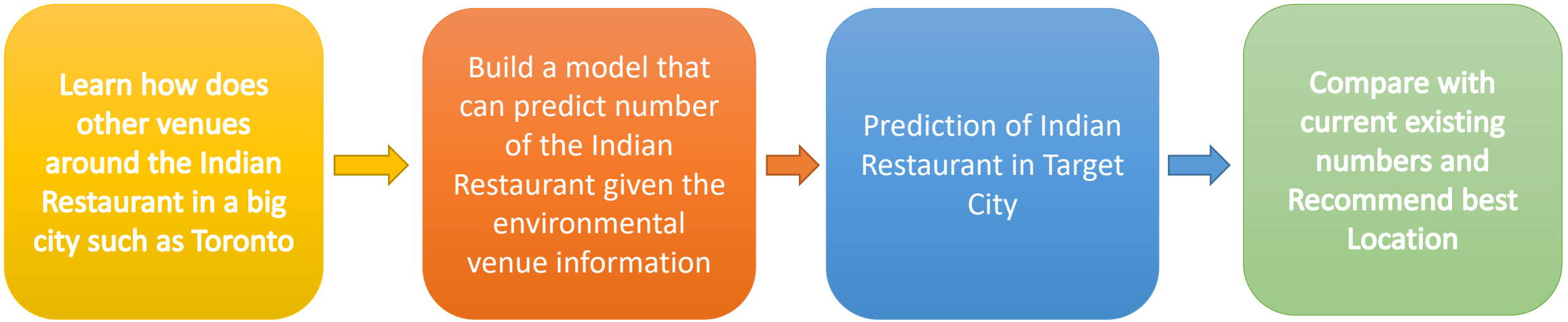
In order to get answer of this question, you have to build **MODEL** to get some recommendations where to open the Restaurant.

- Therefore, firstly we will learn model of a Mature / Metropolitan city since we believe that it is more developed.

  Your city become more bigger some day.
- Another thing you believe is that any business venue does not exist alone and Indian restaurants always tends to be find some other type of shops because neighborhood have 'Cultures' to like them both.

# 2. Solution / Methodology

## FLOWCHART



# Data Import (Toronto)

## 1. Neighborhood Information from Wikipedia

### 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 characters are listed, corresponding to the 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-ROMs. Many vendors also sell 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 that Canada Post may have reserved the M0 FSA for high volume addresses.

| Postal Code | Borough | Neighborhood |
|---|---|---|
| M1A | Not assigned | Not assigned |
| M2A | Not assigned | Not assigned |
| M3A | North York | Parkwoods |
| M4A | North York | Victoria Village |
| M5A | Downtown Toronto | Regent Park, Harbourfront |
| M6A | North York | Lawrence Manor, Lawrence Heights |
| M7A | Downtown Toronto | Queen's Park, Ontario Provincial Government |

2. Scrap from website (https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M) and organize into dataframe

| | Postal Code | Borough | Neighborhood |
|---|---|---|---|
| 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 |

# 3. Location Information

- Use geocoder package to get location information:

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

- Plot the location using folium package

Neighborhood location in Toronto on Map

# 4. Get Venues Information

- Use FOURSQUARE API , we can explore the venues around on specific location, so we could achieve venues, name and category. (https://www.foursquare.com)

| | Neighborhood | Neighborhood Latitude | Neighborhood Longitude | Venue | Venue Latitude | Venue Longitude | Venue Category |
|---|---|---|---|---|---|---|---|
| 0 | Regent Park, Harbourfront | 43.65426 | -79.360636 | Roselle Desserts | 43.653447 | -79.362017 | Bakery |
| 1 | Regent Park, Harbourfront | 43.65426 | -79.360636 | Tandem Coffee | 43.653559 | -79.361809 | Coffee Shop |
| 2 | Regent Park, Harbourfront | 43.65426 | -79.360636 | Cooper Koo Family YMCA | 43.653249 | -79.358008 | Distribution Center |
| 3 | Regent Park, Harbourfront | 43.65426 | -79.360636 | Body Blitz Spa East | 43.654735 | -79.359874 | Spa |
| 4 | Regent Park, Harbourfront | 43.65426 | -79.360636 | Dominion Pub and Kitchen | 43.656919 | -79.358967 | Pub |

- Create **one-hot coding** for each category

| | Neighborhood | Afghan Restaurant | Airport | Airport Food Court | Airport Gate | Airport Lounge | Airport Service | Airport Terminal | American Restaurant | Antique Shop | ... | Toy / Game Store | Trail | Tra Sta |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | Berczy Park | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | ... | 0 | 0 | 0 |
| **1** | Brockton, Parkdale Village, Exhibition Place | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | ... | 0 | 0 | 0 |
| **2** | Business reply mail Processing Centre, South C... | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | ... | 0 | 0 | 0 |
| **3** | CN Tower, King and Spadina, Railway Lands, Har... | 0 | 1 | 1 | 1 | 1 | 3 | 2 | 0 | 0 | ... | 0 | 0 | 0 |
| **4** | Central Bay Street | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | ... | 0 | 0 | 0 |

# 5. Build Model for Prediction

We will use number of venues in each neighborhoods except Indian restaurants as Input and number of Indian restaurants as Output.

Use SVR(rbf kernel) as learning algorithm.

STEP 1 – Optimize the hyperparameter using GridSearchCV on 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 here

```
SVR(C=10.0, 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)
```

# 6. Get information of Target city

- Neighborhoods information is get from [https://cocl.us/new_york_dataset](https://cocl.us/new_york_dataset)

| | Borough | Neighborhood | Latitude | Longitude |
|---|---|---|---|---|
| 0 | Bronx | Wakefield | 40.894705 | -73.847201 |
| 1 | Bronx | Co-op City | 40.874294 | -73.829939 |
| 2 | Bronx | Eastchester | 40.887556 | -73.827806 |
| 3 | Bronx | Fieldston | 40.895437 | -73.905643 |
| 4 | Bronx | Riverdale | 40.890834 | -73.912585 |

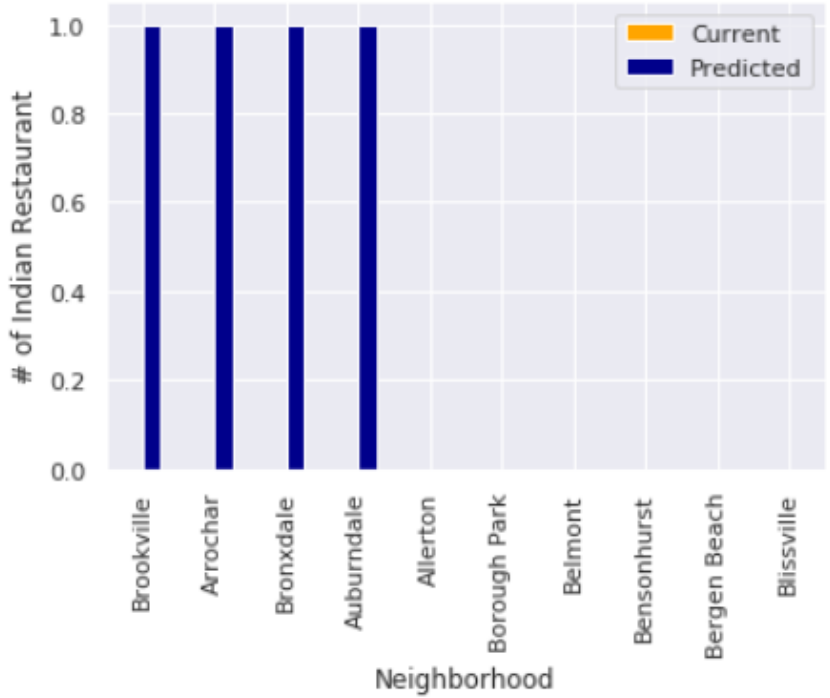
- Get venues information similar to Toronto

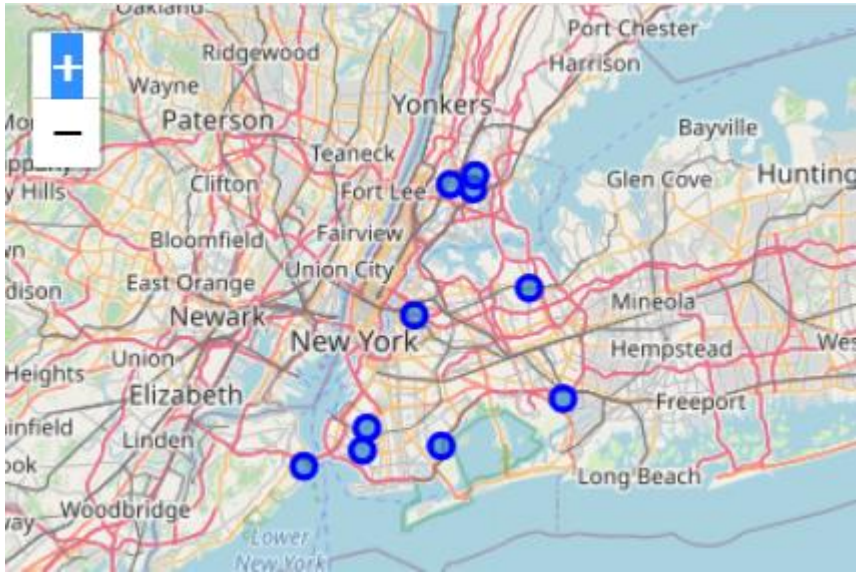
| | Neighborhood | Neighborhood Latitude | Neighborhood Longitude | Venue | Venue Latitude | Venue Longitude | Venue Category |
|---|---|---|---|---|---|---|---|
| 0 | Wakefield | 40.894705 | -73.847201 | Lollipops Gelato | 40.894123 | -73.845892 | Dessert Shop |
| 1 | Wakefield | 40.894705 | -73.847201 | Walgreens | 40.896528 | -73.844700 | Pharmacy |
| 2 | Wakefield | 40.894705 | -73.847201 | Carvel Ice Cream | 40.890487 | -73.848568 | Ice Cream Shop |
| 3 | Wakefield | 40.894705 | -73.847201 | Rite Aid | 40.896649 | -73.844846 | Pharmacy |
| 4 | Wakefield | 40.894705 | -73.847201 | Dunkin' | 40.890459 | -73.849089 | Donut Shop |

# 3. Result

## Predict using Trained model

- Top 4 recommendations for start your business




Top recommendations for start your business in New York

# 4. Conclusion

- We used FOURSQUARE API to get the venues information on given locations.
- We build predictive model with SVR algorithm.
- We get Top recommendations of location to invest “Indian restaurant” in New York.

➢ **These things can make model better:-**

1. This model is build on the assumption that the target city will have a trend to grow to much bigger city.
2. The training dataset very small, if we 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.

# 5. Acknowledgment

In this project, I acknowledge that the Data Science course provided by IBM powered by Coursera.