

Select Neighborhood in Toronto to Open New Restaurant

Coursera Capstone

A dark blue diagonal gradient bar that starts from the bottom left and extends towards the top right, covering the lower half of the slide.

Introduction

Business Problem:

For restaurants to be successful, location is one of the most determining factors. A good location can bring foot traffic to a new restaurant, which is crucial in the early stages. Location is also a factor that is difficult/costly to change in the future, so it is very important for the business owner to make an informed, calculated decision.

Target Audience:

This project aims to help business owners to select the most appropriate neighbourhood to open a new restaurant in the city. Toronto will be used for demonstration.

Dataset

To measure these two factors, we will use the following metrics:

1. List of Toronto neighbourhoods, which we will extract from Wikipedia
2. Postal codes, Latitudes and longitudes of each neighbourhood, which we can get from the csv file provided in previous week's assignment

We will then retrieve the following data using Foursquare API

3. number of restaurants within 1000m of radius from the center of the neighbourhood
4. % of venues that are restaurants within 1000m of radius from the center of the neighbourhood
5. number of tips given to the restaurants
6. average rating scores of the restaurants

Methodology – Step 1

Data Scraping

Use Pandas to scrape Toronto neighbourhood information from website and online file.

	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

Methodology – Step 2

Acquiring Data from Foursquare

Use Foursquare API to get all the nearby venue information within the radius of 1000m from the center of each neighbourhood.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category	ID	Rating	Number of Tips
0	Parkwoods	43.753259	-79.329656	Allwyn's Bakery	43.759840	-79.324719	Caribbean Restaurant	4b8991cbf964a520814232e3	8.8	16
1	Parkwoods	43.753259	-79.329656	Brookbanks Park	43.751976	-79.332140	Park	4e8d9dcdd5fbbb6b3003c7b	7.2	4
2	Parkwoods	43.753259	-79.329656	Tim Hortons	43.760668	-79.326368	Café	57e286f2498e43d84d92d34a	7.0	1
3	Parkwoods	43.753259	-79.329656	A&W	43.760643	-79.326865	Fast Food Restaurant	58a8dcaa6119f47b9a94dc05	6.8	1
4	Parkwoods	43.753259	-79.329656	Bruno's valu-mart	43.746143	-79.324630	Grocery Store	4bafa285f964a5203a123ce3	6.6	4

Methodology – Step 3

Data Cleaning and Wrangling

Clean, transform and summarize data in a way that can be processed in K Mean Clustering.

Neighborhood	Restaurant Count	Restaurant Blend	Restaurant Number of Tips	Restaurant Rating
Agincourt	15	0.500000	0.0	0.0
Alderwood, Long Branch	1	0.033333	0.0	0.0
Bathurst Manor, Wilson Heights, Downsview North	4	0.133333	0.0	0.0
Bayview Village	4	0.266667	0.0	0.0
Bedford Park, Lawrence Manor East	11	0.366667	0.0	0.0

Methodology – Step 4

K Means Clustering

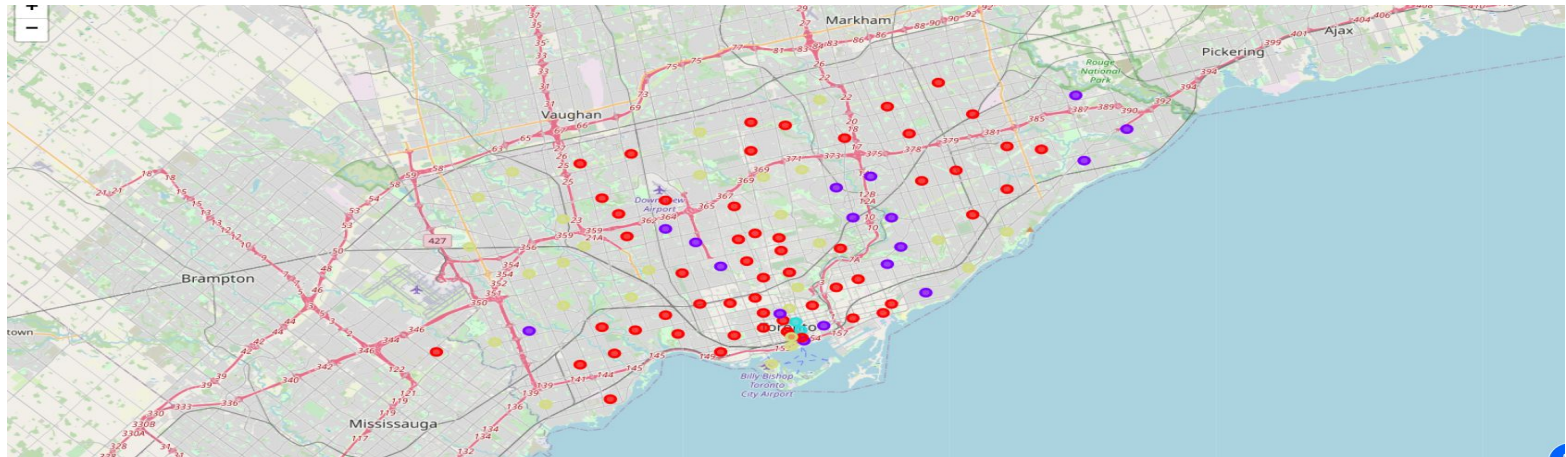
use K Means Clustering method to cluster the neighbourhoods into 4 clusters based on the metrics shown in previous step.

	Restaurant Count	Restaurant Blend	Restaurant Number of Tips	Restaurant Rating	Latitude	Longitude
Cluster Labels						
0	8.964286	0.301516	0.000000	0.000000	43.702568	-79.395410
1	6.470588	0.206481	12.223704	7.149935	43.712823	-79.342320
2	7.500000	0.250000	67.205357	8.329464	43.654328	-79.377177
3	1.777778	0.084035	0.000000	0.000000	43.702519	-79.443866

Methodology – Step 5

Data Visualization

Visualize the grouping using Folium



Results

- Cluster 0: Neighbourhoods with a decent amount of restaurants but no ratings and tips, indicating that customers have low level of motivation to endorse the restaurants.
- Cluster 1: Neighbourhoods with a decent amount of restaurants and average tips amount and rating.
- Cluster 2: Neighbourhoods decent amount of restaurants and a high volume of tips and high rating.
- Cluster 3: Neighbourhoods with a minimum representation of restaurants. No tips and ratings.

Discussions

Foot traffic and competition are the two most important factors to be considered when selecting restaurant locations. Ideally, we would want to open the restaurant at a place that has a lot of foot traffic and relatively low competition.

- Restaurant count: more restaurants mean a high level of competition but also indicate more foot traffic due to the economy of scale effect.
- Restaurant blend: higher blends indicate greater competition as most venues fall into the same category.
- Restaurant rating: high ratings indicate high qualities of existing restaurants and a higher level of competition
- Restaurant number of tips: more tips given by users indicate that customers are more likely to give feedback and recommendations, increasing the online presence of the restaurants.

Conclusions

Cluster 1 and 2 seem to be the most suitable areas to open a new restaurant, with healthy foot traffic suggested by the restaurant count, restaurant blend.

Out of these two clusters, cluster 1 is better than cluster 2 for three reasons:

- 1) Restaurants represent lower percentage of all venues in the neighbourhood, which means less competition is present.
- 2) Existing restaurants have lower ratings, also suggesting less competition.
- 3) Cluster 2 neighbourhoods are located in the center of the city. The prime real estate usually cost more to rent.

Based on the above reasoning, business owners should consider opening new restaurants in neighbourhoods in **cluster 1**