

Restaurant in Toronto – An Opportunity¶

1. Introduction¶

1.1 Background

Toronto is the provincial capital of Ontario and the most populous city in Canada, with a population of more than 2,954 thousand. The diverse population of Toronto reflects its current and historical role as an important destination for immigrants to Canada. Due to which different types of restaurants are there be it Indian, Chinese, American etc. This fact shows us that Toronto is a promising market for investors/business owners to start their business. It is advantageous for business owners to accurately predict where they should open the restaurant and what are the driving factor for opening of restaurant.

1.2 Problem

Major challenge faced in starting new business is to choose the location of the business. This project aims to predict and decide which neighbourhoods are most suitable to opening of the restaurant.

1.3 Interest

Obviously, anyone who is planning to set the restaurant in Toronto will be interested in this data for competitive advantage over others who are planning for the same.

2. Data Acquisition and Cleaning

2.1 Data Sources

- City of Toronto Neighbourhood Profiles for providing an overview of the neighbourhoods in Toronto:
 - Wikilink for fetching the postal codes
 - Geospatial Data for coordinates
- City of Toronto Neighbourhood Shapes for mapping: GeoJSON File
- Foursquare API to collect information:
 - Venues in the Neighbourhood
 - Venue Details like Rating, Tips etc.

2.2 Data Cleaning

Data downloaded or scraped from multiple sources were combined into one table.

2.3 Feature Selection

3. Exploratory Data Analysis

Map View

Let's view every neighbourhood in Toronto with a map. We determine the coordinate of Toronto as the centre of map. Then, we mark each neighbourhood as a point in the map using markers.

Nearby Neighbourhoods

To determine nearby venues for each neighbourhood, we use the Foursquare Location Data to list all venues that are nearby using Foursquare API.

Neighbourhood, Neighbourhood Latitude, Neighbourhood Longitude, Venue ID, Venue, Venue Latitude, Venue Longitude, Venue Category are the fields that are fetched using API. Then Venue ID of each venue of the neighbourhood is used to fetch the details of that venue, fields like Venue ID, Venue, Venue Latitude, Venue Longitude, Venue Tips Count, Venue Rating, Likes Count, Venue Price Tier.

Hot coding needs to be done for calculating ratio based on venue category.

Because our analysis focusses on each neighbourhood as each record entry / row, we need to aggregate the venue data. There are 5 features that we want to analyse:

- Restaurants: # of restaurants in each neighbourhood
- Venues: # of all venues in each neighbourhood
- Restaurant_Ratio: Ratio of restaurants corresponding to # of venues
- Average_Rating: Average restaurant rating
- Average_Tips: Average restaurant tips

4. Modelling

K- MEANS CLUSTERING - elbow method to determine the number of clusters.

Since the data is standardized, we can't really see the exact value of each column so need to join multiple data frames to get the needed data frame. We need to fetch the below details from the clustered data frame:

- Top Venue Restaurants
- No of Restaurant
- No of Venues
- Restaurant_Ratio
- Average_Rating
- Average_Tips

5. Conclusion:

We need to analyse descriptive statistics for each column, starting with Top Venue Restaurants.

Cluster 2 have the highest top venue restaurant, with mean of 3.8. This means that in top 10 venues for each neighbourhood in Cluster 2, there are roughly 4 restaurants that get into that top 10 venues list. Cluster 4 is slightly behind. In cluster 0,1 and 3, on average 2.5 restaurant that made it to top 10 venue list.

From this table we can conclude that Cluster 3 and 4 have less Restaurant in its neighbourhoods compared to other Cluster. Cluster 0 has max restaurants and Cluster 1 and 2 have similar number of restaurants in its neighbourhood, and almost close to Cluster 0.

Cluster 2 has highest number of venues and Restaurants.

Cluster 3 has highest number of tips followed by Cluster 1 and then 2. This shows tips are there in other clusters too, but quality/services is not that good in terms that people pay tips.

As a business owner that just recently start the business, we want to build our restaurant in places with restaurants where people are not completely satisfied with the restaurants in that area. This strategy is chosen because we want to grab markets where people are looking for new options and they are willing to pay to as tips are also coming for many restaurants.

Based on the characteristics mentioned above, it's obvious that business owner should choose Cluster 2 to build their first restaurant.