Finding suitable locations for a Restaurant in Downtown Toronto, Canada

Ashwini Anand

August 22, 2021

1. Introduction:

1.1. Business Problem:

This project aims at identifying a suitable location to open a new restaurant in the city of Toronto, Canada, specifically Downtown Toronto. Through this project, I aim to ease the process of shortlisting the most suitable locations for such a venture using map data from Foursquare.

Although a lot of factors must be considered in such cases, our primary goal is to find areas with a low-to-moderate density of restaurants, which are also close to the city centre. This will be helpful to ensure good footfall in our restaurant, while also maintaining a healthy level of competition with other restaurants for maximum growth of our business.

1.2. Interest:

This project will be of great use to stakeholders interested in opening a new restaurant in the Downtown Toronto area, and especially to those looking to bounce back after the coronavirus pandemic has pushed many out of business, and to those looking to cash in on the surge of customers getting out of their homes after the pandemic scare and repeated lockdowns.

2. Data and Sources:

2.1. Data Acquisition:

First, the list of boroughs and neighbourhoods in Toronto city can be found on this Wikipedia page. I scraped the page and created a data frame of all neighbourhoods by postal codes and boroughs. However, this data frame lacks coordinates. The location data of these postal codes can be found here. The location data, in csv format, is already sorted by postal codes with coordinates to the centre of each area. Then, I merged the two data frames to create a data frame having the postal codes, borough, neighbourhoods, and the latitude and longitude data of each postal code in Toronto.

The remaining data about restaurants in each neighbourhood was collected using Foursquare API, which will be covered in detail in the methodology section.

2.2. Data Cleaning:

The final data frame was already well-organised and not much cleaning was required. Some borough names were erroneous, which were replaced with the corrected names. In the end, a complete data frame was created with 103 rows.