

BakerZ, Inc. Expansion Strategy for Toronto and New York

Shail Rajput

November 22, 2020

1. Introduction/Business Problem

BakerZ, Inc. is a multi-national corporation that specializes in wholesale manufacturing and distribution of fine baked goods to coffee shops including sandwiches, cakes, pastries, and cookies. Quality, freshness, and punctual delivery of the product have been their key success factors in sustaining their customers. These are the factors that BakerZ does not compromise on.

BakerZ is now looking at expanding their business to New York, NY in the USA and Toronto, ON in Canada. They have tasked us to determine the best neighborhood locations in Toronto and New York where they can set up distribution centers.

Their requirements for each distribution center are as follows:

- It serves multiple neighborhoods that are closest to each other
- It should attempt to target the neighborhoods with the most coffee shops but have at least 10 coffee shops as the minimum
- It should itself be located in the neighborhood with the maximum number of coffee shops

Based on these requirements, BakerZ, Inc. has tasked us in determining how many distribution centers they should establish in Toronto and New York, and their neighborhood locations. For each city, they also want us to recommend a priority order in establishing the distribution centers so they have a better return on investment.

2. Data

We primarily need the geolocations (latitude and longitude) for all neighborhoods in Toronto, ON and New York, NY to get started.

Once we have the geolocations of each neighborhood, we shall use the Foursquare API to retrieve the data on coffee shops in each neighborhood of each city.

We shall then use this data to solve our business problem as stated above in Section 1.

Data sources for Toronto, ON

We have two sources of data for Toronto.

Neighborhood data shall be scraped from Wikipedia located at the following URL:
https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

This data does not have geolocation information for each neighborhood so we shall use a second source that maps postal codes to geolocations to get the complete data we need to

solve our problem. The second source is located in a CSV file called `toronto_postal_code_geoloc.csv` at the following URL:
https://cocl.us/Geospatial_data

We shall then combine these two data sources using the postal code as key to create a single data frame that holds all the data we need for Toronto, mapping neighborhoods to geolocations.

Data sources for New York, NY

We shall use a single source for New York for the data mapping neighborhoods to geolocations. This data is available in a file called located at the following URL:

https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDeveloperSkillsNetwork-DS0701EN-SkillsNetwork/labs/newyork_data.json

Similar to Toronto, this data shall be compiled into a data frame that maps New York neighborhoods to their geolocations.

Sourcing Coffee Shop data using Foursquare API

Getting the total number of coffee shops for each neighborhood in Toronto and New York is critical to solving our business problem.

We shall use the Foursquare Venues Search feature (<https://api.foursquare.com/v2/venues/search/<parameters>>) and use the geolocation of each neighborhood to determine all the coffee shops in that neighborhood, count them and add that number to our city data frame in a new column to capture the number of coffee shops.

Using these data source coupled with data extracted using the Foursquare API satisfies our basic data needs to solve our business problem.