Capstone Project - The Battle of Neighborhoods

─────────

JUNE 9

─────────

Coursera Project
Authored by: Prashanth H Siddaraju

Problem Description

Lot of stores are selling coffee in New York and Toronto, but there are very fewer coffee shops which only sell coffee (example- Starbucks).

1) If any coffee company or a coffeehouse chain is planning to open a new store in New York or Toronto, where can they open the same? Which are the probable Neighborhoods in which coffee is already being sold by a non-Coffee store, but have a possibility to have a coffee store?

2) Which coffee drinking neighborhoods in New York and Toronto are similar, so the same business model can be implemented in the other city as well?

Background of the issue

According to thecoffeebump statistics

- 58% of Americans over the age of 18 drink coffee every single day
- The US alone spends $4 billion importing coffee each year
- The average coffee drinker spends $164.71 on coffee each year
- 100 million people in the US drink coffee each day
- 25% of Americans skip breakfast, yet 50% will still consume a cup of coffee in the morning

Specialty coffee sales are increasing by 20% per year and account for nearly 8% of the 18 billion-dollar U.S. coffee market. Major coffee companies like Starbucks, Espresso house are planning to open more and more stores. But coffee is already being sold in Neighborhood by general shops, like Donut store, Mc Donald's, KFC etc.… which proves a fact that there is an increasing demand for coffee in these areas. If major coffee companies can plan to open their outlets in these places, it could lead to better sales and profit.

Second issue being which business model to be selected for a specific neighborhood while opening the store. Business model / Sales strategy place a vital role in starting and improving the business in any place. There is a need to know which business model would help to improve the business here depending on coffee drinking habits, by comparing the same which is already implemented in a different city.

Description of Data and Solution

We would be needing the Neighborhood information of 2 Cities here

1) New York
2) Toronto

New York neighborhood data with latitude and longitudes are already available in the link - https://cocl.us/new_york_dataset

This data is json format, this must be converted into Tabular format.

Toronto dataset is not readily available, but we can find the neighborhood information with zip codes Wikipedia page

https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

This has to be web scraped and cleansed before the same can be used.

- Drop the postcodes with Borough's 'Not Assigned'
- Group the Neighborhood with more than one post code
- Replace Neighborhood names which are 'Not Assigned' with respective Borough names.

Latitude and longitude data for Toronto neighborhoods can be found in this location

http://cocl.us/Geospatial_data

Once the above 2 data sets are gathered, we would require Venue details which are selling coffee around 500 meters of the above-mentioned neighborhoods. This data can be acquired from foursquare using explore api with relevant filters.

Once the data is cleansed and Venue details are captured, the same can be split into 2 data frames

1) coffee companies selling coffee
2) Other Outlets selling coffee

The same needs to be visualized using Folium on a New York map to see the distribution of Non-Coffee shops, and the same can be used to estimate the possibility of opening a coffee shop in the vicinity.

We can also do the same analysis for Toronto, which later can be used to cluster with New York to see which Coffee Drinking Neighborhood in Toronto is similar to a neighborhood in New York.