Battle of neighborhood : Toronto city Starting a new coffee shop

Business Problem Toronto is Canada's largest city with a population of more than 2,7 million and a density of 4,334.4 people per square kilometer. The city is renowned as one of the most multicultural cities globally due to its large population of immigrants from all over the globe.This also means that the market is highly competitive.Thus, any new business venture or expansion needs to be analysed carefully. The insights derived from analysis will give a good understanding of the business environment which help in strategically targeting the market. This will help in reduction of risk. And the Return on Investment will be reasonable.

Problem Overview Now, imagine that you have a coffee franchise called coffee Costa that has been doing business successfully in New York. you plan to expand the business and decide to look for a city that shares the same trait as New York, and one of the cities is Toronto.To ensure this project's success, the team requires insights into the demographics, neighboring businesses, and crime rates. For each neighborhood, we can ask: How many other cafes

exist?What are the most popular hangouts? Can we get information about the other recreational spots? What is the neighborhoods' crime rate? Thus, the project goal is to figure out the best locations for opening up a new coffee shop in Toronto City. Other factors needed to be taken in consideration are -Toronto Population, Toronto City Demographics

Are there any other venues like, Entertainment zones, Parks etc nearby where floating population is high etc

Who are the competitors in that location?

Cuisine served / Menu of the competitors

Segmentation of the Borough

Untapped markets

Saturated markets etc

3. Target Audience
Entrepreneurs who are passionate about opening a coffee shop in a metropolitan city would be very interested in this project. The project is also for business owners and stakeholders who want to invest in the franchise chain.

Data Description 1. Data Requirements and Collection
We need historical data about crime incidents, busiest roads, and popular venues. Luckily, Toronto has an open data portal that makes it public. We can also leverage Foursquare Location data to compare neighborhoods in terms of service. Hence, the followings are data sources that we can use for this project-: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M The list of Toronto neighborhoods represented by postal codes and their boroughs

https://ckano.cf.opendata.inter.prod-toronto.ca/en/dataset/traffic-signal-vehicle-and-pedestrian-volumes   The most updated record of traffic signal vehicle and pedestrian volumes in Toronto City

https://tinyurl.com/toronto-mci The most updated record of crime incidents reported in Toronto City provided by Toronto Police Services. :

The popular or most common venues of a given neighborhood in Toronto.

Data Cleaning and Feature Extraction The first is a Wikipedia page about Toronto postal code. We will scrape the page and create a data frame consisting of three columns; PostalCode, Borough, and Neighborhood. We remove any rows that do not have a borough assigned. Then, we will be using the Geocoder python package to retrieve the postal code's coordinates. It will return 103 rows and 5 columns.

The second data is in a CSV file. It contains 2280 rows and 11 columns. The data is typically collected between 7:30 a.m. and 6:00 p.m at intersections where there are traffic signals. Each intersection holds vehicle and pedestrian volumes data, along with its coordinates. We will focus on 5 columns; those are Main, 8 Peak Hr Pedestrian Volume, 8 Peak Hr Vehicle Volume, Latitude, and Longitude. We will use these features to diagnose each main road's characteristics and locate the busiest main roads in the city.

The third data is also in a CSV file. It contains 206,435 rows and 9 columns. The rows represent crime incidents that were reported from 2014 to 2019. It has 5 Major Crime Indicators (MCIs) scattered to 17 divisions and 140 listed neighborhoods. We will group the data based on division and get statistics about crime rates

The fourth data is stored inside Foursquare Location Data, and we will use Foursquare API to access it. We utilize the postal coordinates to retrieve popular venues around a specific radius. As a result, the same venue categories will be returned to different neighborhoods. We can use this idea to cluster the neighborhoods based on their venues representing services and amenities.

We will run the k-Means algorithm to perform this clustering with a different number of clusters (k). The features will be the mean of the frequency of occurrence of each venue category. Finally, we can visualize the cluster model using the Folium module.

To sum up, we will use the 2nd and 3rd data to analyze the pedestrian/vehicle volume and crime rates. Then, we load the 1st data to obtain the exact coordinates for each neighborhood based on the postal code, allowing us to explore and map the city. Finally, we will use the coordinates and Foursquare credentials to access the 4th data source through its API and retrieve the popular venues along with their details, especially for coffee shops. The venue frequency in each neighborhood will be the features of the clustering model.