INTORDUCTION / BUSINESS PROBLEM STATEMENT

Find the best locations for starting a new Restaurant in Toronto City

A global restaurant chain company wants to start a restaurant shop in one of the best neighbourhoods in Toronto city. They want to find the best neighbourhood locations where there are no, or minimal restaurants currently exists in the area. They are looking for less competitors in the area but a more popular place to start with.

Since they are a multicuisine restaurant chains, they want to take into account all types of restaurants that provide different country dishes.

The problem is to find and filter such locations and apply the best statistical approaches to find the best location(s) and visualize with statistical proofs to confirm the analysis.

DATA DESCRIPTION

The primary data that will be needed is the popular locations within Toronto city and a reliable source of data.

We have the Canadian postal data from the wiki which we can leverage to find the popular boroughs and the neighbourhoods that exists in each borough.

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

The data primarily consists of the below details in tabular format

1. Postal Code
2. Borough Name
3. Neighbourhood(s)

M4A North York (Victoria Village)	**M5A** Downtown Toronto (Regent Park / Harbourfront)	**M6A** North York (Lawrence Manor / Lawrence Heights)
M4B East York (Parkview Hill / Woodbine Gardens)	**M5B** Downtown Toronto (Garden District, Ryerson)	**M6B** North York (Glencairn)
M4C East York (Woodbine Heights)	**M5C** Downtown Toronto (St. James Town)	**M6C** York (Humewood-Cedarvale)

SOLUTION APPROACH

The existing data is tabular format which we cannot input directly to any data analysis tools. Below is a step by step solution approach to leverage the table data to find suitable locations.

1. Scrap the Canada Postal data from the web page into processable input format
2. Get the location data for each neighbourhood/borough and update the input data
3. Visualize the neighbourhoods in Canada
4. Get the popular neighbourhoods within a defined geographic circle of Toronto
5. Explore the neighbourhoods in Toronto
6. Analyse each neighbourhood based on restaurant data
7. Cluster each neighbourhood based on restaurant data
8. Examine and selected the best clusters with that satisfies the restaurant location requirements

Detailed Solution Approach

1. *Scrap the Canada Postal data from the web page into processable input format*

   The webpage data in the tabular form needs to be read and converted into a suitable data structure that can be used as an input to any analysing tools. The web scraping libraries such as Scrapy, beautiful soup or python requests to convert the data into usable format. In this assignment we'll use the python requests. The expectation is to create a new tabular data structure in below format.

Postal Code	Borough	Neighbourhood(s)

2. *Get the location data for each borough and update the input data*

   The primary data must be updated with additional useful data, such as geographic location data with latitudes and longitudes to get more information about the places. We can leverage geopy library or third-party data sets to do this. In this assignment we'll use one of the third-party data sets to combine the location data to the input data set of the Canadian neighbourhoods.

   The above-mentioned in 1. table will be updated to below format.

Postal Code	Borough	Neighbourhood(s)	Latitude	Longitude

3. *Visualize the Canadian neighbourhoods*

   This is an optional step. However, it is useful to get a good high-level visual view of the initial input data set that is going to be analysed and refined in the further stages. We can leverage the existing visualization libraries such as Folium to get this view.

4. *Get the popular neighbourhoods within a defined geographic circle*

   In this step we'll get the popular neighbourhoods within a more defined geographical circle in and around Toronto. This requires two steps.

   a. First, the above table must be filtered for records of Toronto.
   b. Pass the required input data to a location data API such as the Foursquare API to get more details on the Toronto neighbourhoods which can be input to the next stages in the solution approach.

   The results from the location data API such as FourSquare can be used to get a lot of information about the explored locations such as its popularity and the reasons for its popularity. This can be

filtered to identify the suitable locations that can be taken as candidates for the problem statement. A sample response data is shown below for another location.

Below is a Sample Four Square Data of a Diner in Tibett Ave, US


```
{'reasons': {'count': 0,
   'items': [{'summary': 'This spot is popular',
     'type': 'general',
     'reasonName': 'globalInteractionReason'}]},
  'venue': {'id': '4b79cc46f964a520c5122fe3',
   'name': 'Tibbett Diner',
   'location': {'address': '3033 Tibbett Ave',
    'crossStreet': 'btwn 230th & 231st',
    'lat': 40.8804044222466,
    'lng': -73.90893738006402,
    'labeledLatLngs': [{'label': 'display',
      'lat': 40.8804044222466,
      'lng': -73.90893738006402}],
    'distance': 452,
    'postalCode': '10463',
    'cc': 'US',
    'city': 'Bronx',
    'state': 'NY',
    'country': 'United States',
    'formattedAddress': ['3033 Tibbett Ave (btwn 230th & 231st)',
     'Bronx, NY 10463',
     'United States']},
   'categories': [{'id': '4bf58dd8d48988d147941735',
     'name': 'Diner',
     'pluralName': 'Diners',
     'shortName': 'Diner',
     'icon': {'prefix': 'https://ss3.4sqi.net/img/categories_v2/food/diner_',
      'suffix': '.png'},
     'primary': True}],
   'photos': {'count': 0, 'groups': []}},
  'referralId': 'e-0-4b79cc46f964a520c5122fe3-2'},
```


5. *Explore the neighbourhoods in Toronto*

   The above step 4 will give you the neighbourhood data in Toronto with additional features and details. This data can be explored to find the suitable locations to identify where all restaurants exists.

6. *Analyse each neighbourhood based on restaurant data*

   We can analyse each neighbourhood based on additional data such as if the same venue has a more popularly visited additional place such as a hotels or parks. This increases the business chances for a restaurant.

7. *Cluster each neighbourhood based on restaurant data*

   In this step we'll cluster the results based on the above business criteria and select the best locations for listing in a tabular and visual format.

8. *Examine and selected the best clusters with that satisfies restaurant location requirements*

   Examine the best clusters and list them as suitable candidate places for starting new restaurant and locations.