Segmenting the Neighborhoods of Calgary and Vancouver

A

Capstone Project

for

IBM Data Science Professional Certification

By

Nicholas Olusegun IBITOKUN

13 April 2021

Table of Contents

1. INTRODUCTION

Canada is the second largest country on earth with an area of 10 million square kilometers. However, possibly due to its geographical location and cold climate, its population is about a tenth of its neighbor, the USA. To stimulate economic growth and development, the government of Canada over the years rolled-out immigration policies with the intent of attracting the best talents from all over the world. The country has been very successful in its point-based immigration system, so much so, population has been on the steady increase. This is evident in populations growths across several boroughs and cities in Canada.

In addition to being popular tourist and vacation destinations, Calgary and Vancouver, the biggest western cities in the provinces of Alberta and British Columbia respectively, are also leading immigration locations. The cities are diverse and multicultural and offer numerous attractions for potential immigrants and tourists.

Using available online public data, this project clusters and segments the neighborhoods of Calgary and Vancouver with the objective of drawing insights useful for immigrants and tourists interested in landmarks including restaurants, schools, residential areas, parks, museums, churches, etc.

2. BUSINESS PROBLEM

Immigrants are usually concerned about which neighborhood to live, available restaurants and cuisine, schools to attend, proximity to hospitals, places of worship and public transportation, amongst other things.

The objective of this project is to unearth useful insights for immigrants to select their new place of abode. In addition, it is expected it will allow temporary visitors (tourists) to choose their destinations depending on their interests and amenities on offer in different locations.

3. DATA DESCRIPTION

The main data source for this project was the Wikipedia website. The site contains useful information about the postal codes of boroughs and neighborhoods in Canada. Its also contains some geographical location data for the cities of interest (Calgary & Vancouver) in this project. Specifically, existing alphanumeric postal codes served as proxies to extract information about neighborhoods and boroughs.

In cases where location coordinates were missing from the Wikipedia website, ArcGIS package was deployed to obtain the required longitudes and latitudes and the resulting data wrangled into a cleaned data-frame using standard Python merging codes.

Location data of venues in proximity to the neighborhoods of interest were obtained using the Foursquare API service.

4. METHODOLOGY

The techniques described in this section were deployed to extract required information from the identified sources. Also captured are manipulation techniques used to pre-process, clean and wrangle information into a useful data-frame.

4.1 WebScraping

WebScraping technique was employed to extract information from Wikipedia website. The following links were used for the respective cities:

i.   Calgary: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_T
ii. Vancouver: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_V

The Calgary link contains the postal codes, boroughs, neighborhoods and their respective coordinates. The Vancouver link also contains the postal code, borough and neighborhood information, however the coordinate data (longitude and latitude) are missing. The ArcGIS library was installed in the Python notebook and used to extract the longitude, latitude information for Vancouver. The coordinate data was merged with the Vancouver table obtained from the Wikipedia link, using standard Python data-frame merging techniques with Postal Code as the linking parameter (key).

4.2 Foursquare API

Several venues are available in different Canadian boroughs and neighborhoods. These pieces of information were obtained using the Foursquare location data. Foursquare is a location technology platform offering business solutions and consumer products through a deep understanding of location including all manner of venues and events within an area of interest. The location data from foursquare was used since it contains robust and relevant information critical to build a clustering model. The data was readily extracted via the Foursquare API.

After the listing of the boroughs, neighborhoods and their corresponding coordinates, the Foursquare API was employed to pull information about venues in proximity to each neighborhood. The API calls were set such that venues location data pulled were within 700m radius of the subject neighborhoods.

4.3 Data Cleaning & Preprocessing

Ordering, sorting, grouping, merging, renaming column labels, normalisation were some of the techniques used to pre-process, clean and wrangle required information into a final data-frame.

4.4   Imported Libraries

All required libraries for the project were imported at the beginning. Also required packages such as Folium and ArcGIS were also installed. Below is a listing of imported libraries and installed packages.

1. *import pandas as pd*
2. *import numpy as np*
3. *import requests*
4. *from pandas.io.json import json_normalize*
5. *!pip install folium*
6. *import folium # for plotting*
7. *from bs4 import BeautifulSoup*
8. *from sklearn.cluster import KMeans*
9. *import matplotlib.cm as cm*
10. *import matplotlib.colors as colors*
11. *!pip install arcgis*

4.5   Clustering Model

A K-means clustering model was built and used to segment the Calgary and Vancouver neighborhoods into clusters. The cities were grouped into 5 clusters each with delineations and description of the popular venues in proximity to the different neighborhoods of each cluster.

4.6   Visualization

The Folium library was used to generate insightful maps of Vancouver and Calgary in 2 phases. In the first phase, maps generated highlighted the different neighborhoods as wrangled into data-frames. For the second phase, Folium maps were also used to visualize the resulting clusters after the K-Means clustering was completed.

4.7   Evaluation of Clusters

Finally, five different clusters obtained each for Calgary and Vancouver were reviewed to draw insights with the intent of assisting potential customers (immigrants and tourists) in making informed choices.

5. RESULTS

5.1 Calgary

The results for the city of Calgary are presented in this section. Below is a map of Calgary generated using Folium. It captures the segmentation of the neighborhood into 5 clusters.
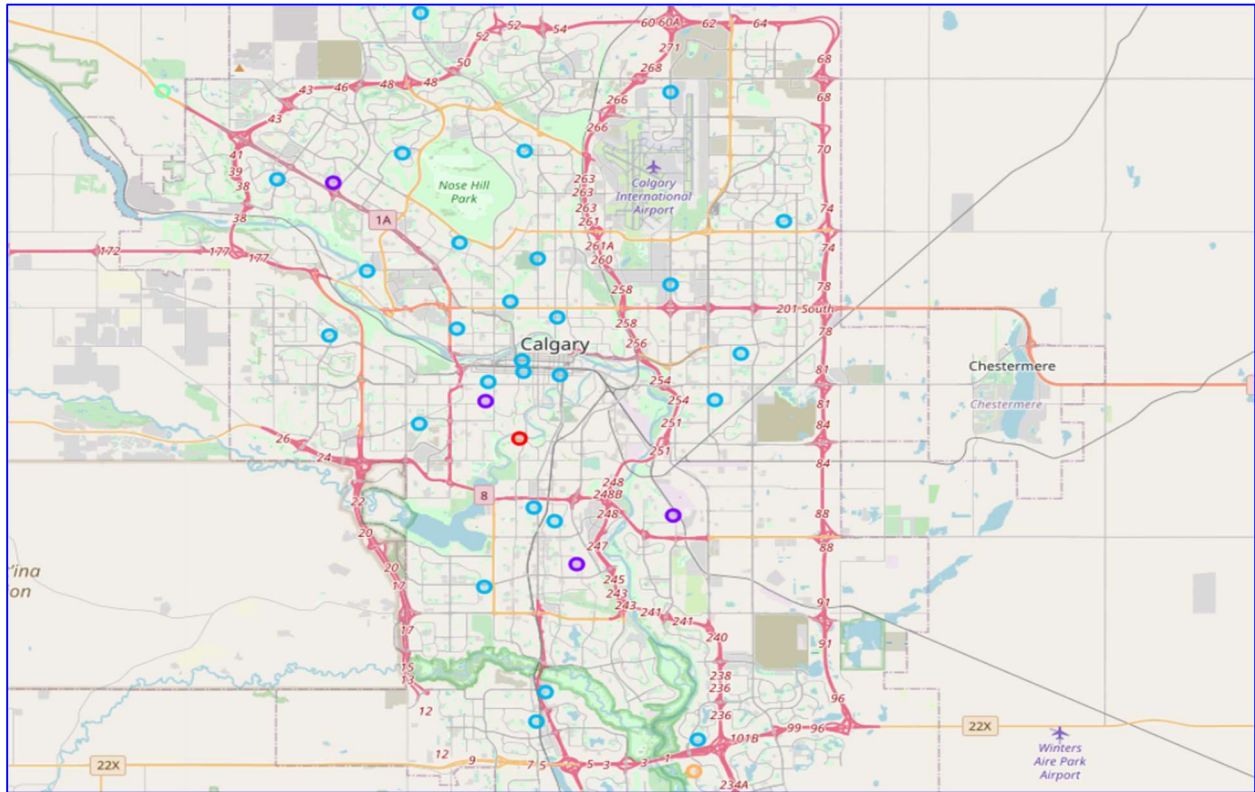

Figure 1: Map of Calgary showing different Neighborhood Clusters

Details of the resulting Calgary clusters are captured in the tables below. Sample tables shown reflect the different neighborhoods and the most common venues within 700m radius of it. Full tables for all the 5 Calgary clusters are available in Git-Hub link (https://github.com/Segun-Ibitokun/Coursera_Capstone/blob/4b9b852a82085708265333814c6aa93b58fcce79/Clustering%20and%20Segmenting%20%20Calgary%20and%20Vancouver.ipynb).

Table 1: Calgary Cluster 1 Showing most popular Venues

	Neighborhood	Cluster Label	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
4	Lynnwood Ridge, Ogden, Foothills Industrial, G...	1	Pizza Place	Theater	Clothing Store	Convenience Store	Diner	Eastern European Restaurant	Dog Run	Furniture / Home Store	Frozen Yogurt Shop	Fried Chicken Joint
9	Hawkwood, Arbour Lake, Citadel, Ranchlands, Ro...	1	Pizza Place	Convenience Store	Miscellaneous Shop	Donut Shop	Flea Market	Furniture / Home Store	Frozen Yogurt Shop	Fried Chicken Joint	French Restaurant	Food Truck
12	Queensland, Lake Bonavista, Willow Park, Acadia	1	Pizza Place	Convenience Store	Chinese Restaurant	Construction & Landscaping	Insurance Office	Dry Cleaner	Flea Market	Frozen Yogurt Shop	Donut Shop	Fried Chicken Joint
28	South Calgary (Altadore / Bankview / Richmond)	1	Pizza Place	Health & Beauty Service	Convenience Store	Coffee Shop	Candy Store	Café	Spa	Pet Store	Athletics & Sports	Fried Chicken Joint

Table 2: Calgary Cluster 2 Showing most popular Venues (Full table in Git-Hub link)

	Neighborhood	Cluster Label	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Penbrooke Meadows, Marlborough	2	Convenience Store	Vietnamese Restaurant	Noodle House	Bank	Grocery Store	Yoga Studio	Filipino Restaurant	Frozen Yogurt Shop	Fried Chicken Joint	French Restaurant
1	Dalhousie, Edgemont, Hamptons, Hidden Valley	2	Gas Station	Asian Restaurant	Convenience Store	Café	Yoga Studio	Flea Market	Furniture / Home Store	Frozen Yogurt Shop	Fried Chicken Joint	French Restaurant
2	Forest Lawn, Dover, Erin Woods	2	Convenience Store	Sandwich Place	Smoke Shop	Coffee Shop	Indonesian Restaurant	Fried Chicken Joint	Falafel Restaurant	Bar	Frozen Yogurt Shop	French Restaurant
3	Montgomery, Bowness, Silver Springs, Greenwood	2	Clothing Store	Coffee Shop	Restaurant	Park	American Restaurant	Men's Store	Shopping Mall	Furniture / Home Store	Luggage Store	Snack Place
5	Rosscarrock, Westgate, Wildwood, Shaganappi, S...	2	Mexican Restaurant	Pub	Vietnamese Restaurant	Pizza Place	Sandwich Place	Indian Restaurant	Fast Food Restaurant	Grocery Store	Cosmetics Shop	Bookstore
6	Bridgeland, Greenview, Zoo, YYC	2	Chinese Restaurant	Japanese Restaurant	Middle Eastern Restaurant	Pub	Park	Sandwich Place	Scenic Lookout	Noodle House	Fast Food Restaurant	Coffee Shop
7	Lakeview, Glendale, Killarney, Glamorgan	2	Bowling Alley	Japanese Restaurant	Sports Bar	Sandwich Place	Supermarket	Sushi Restaurant	Pizza Place	Bakery	Chinese Restaurant	School

5.2   Vancouver

The results for the city of Vancouver are presented in this section. Below is a map of Vancouver generated using Folium. It captures the segmentation of the neighborhood into 5 clusters.
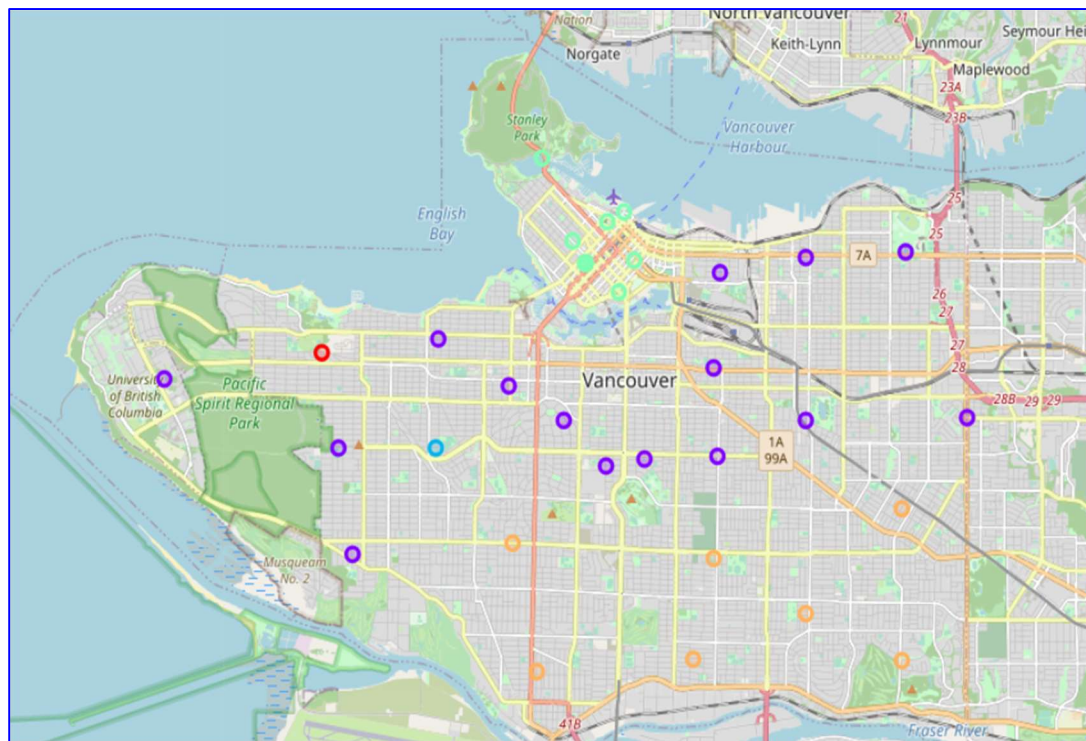

Figure 2: Map of Vancouver showing different Neighborhood Clusters

Details of the resulting Vancouver clusters are captured in the tables below. Sample tables shown reflect the different neighborhoods and the most common venues within 700m radius of it. Full tables for all the 5 Vancouver clusters are available in Git-Hub link (https://github.com/Segun-

).

Table 3: Vancouver Cluster 1 Showing most popular Venues (Full table in Git-Hub link)

	Neighborhood	Cluster Label	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	North Hastings-Sunrise	1	Theme Park Ride / Attraction	Park	Vietnamese Restaurant	Coffee Shop	Beer Garden	Event Space	Theme Park	Pizza Place	Sandwich Place	Stadium
1	North Grandview-Woodland	1	Coffee Shop	Brewery	Café	Chinese Restaurant	Vietnamese Restaurant	Bakery	Grocery Store	Breakfast Spot	Theater	Pizza Place
2	South Hastings-Sunrise , North Renfrew-Colling...	1	Coffee Shop	Sandwich Place	Indian Restaurant	Chinese Restaurant	Latin American Restaurant	Fast Food Restaurant	Burger Joint	Candy Store	Bank	Bus Stop
3	South Grandview-Woodland , NE Kensington-Cedar...	1	Café	Park	Sushi Restaurant	Lake	Bookstore	Diner	Liquor Store	Skating Rink	Farmers Market	Beach
7	East Mount Pleasant	1	Coffee Shop	Pizza Place	Sushi Restaurant	Vietnamese Restaurant	Mexican Restaurant	Bakery	Gym	Breakfast Spot	Arts & Crafts Store	Ethiopian Restaurant
8	West Kensington-Cedar Cottage , NE Riley Park-...	1	Coffee Shop	Park	Vietnamese Restaurant	Bakery	Pie Shop	Pizza Place	Chinese Restaurant	Sandwich Place	Shoe Store	Café
11	West Mount Pleasant , West Riley Park-Little M...	1	Coffee Shop	Japanese Restaurant	Skating Rink	Café	Arts & Crafts Store	Vegetarian / Vegan Restaurant	Restaurant	Farmers Market	Chinese Restaurant	Dessert Shop
12	East Fairview , South Cambie	1	Coffee Shop	Bus Stop	Chinese Restaurant	Bubble Tea Shop	Park	Bank	Sporting Goods Shop	Juice Bar	Outdoor Sculpture	Sandwich Place

Table 4: Vancouver Cluster 3 Showing most popular Venues (Full table in Git-Hub link)

	Neighborhood	Cluster Label	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
14	NE Downtown , Gastown , Harbour Centre , Inter...	3	Hotel	Restaurant	Coffee Shop	Taco Place	Café	Clothing Store	Breakfast Spot	Italian Restaurant	Burger Joint	Plaza
15	Waterfront , Coal Harbour , Canada Place	3	Hotel	Restaurant	Coffee Shop	Café	Steakhouse	Boat or Ferry	Cosmetics Shop	Bookstore	Plaza	Food Truck
16	SE West End , Davie Village	3	Hotel	Japanese Restaurant	Bakery	Dessert Shop	Food Truck	Cosmetics Shop	Restaurant	Seafood Restaurant	Gay Bar	Coffee Shop
17	NW West End , Stanley Park	3	Japanese Restaurant	Hotel	Italian Restaurant	Garden	Noodle House	Korean Restaurant	Coffee Shop	Café	Aquarium	Pub
28	SW Downtown	3	Hotel	Japanese Restaurant	Italian Restaurant	Bakery	Cosmetics Shop	Seafood Restaurant	French Restaurant	Restaurant	Dessert Shop	Sandwich Place
29	Outer East	3	Hotel	Japanese Restaurant	Italian Restaurant	Bakery	Cosmetics Shop	Seafood Restaurant	French Restaurant	Restaurant	Dessert Shop	Sandwich Place
30	Inner East	3	Hotel	Japanese Restaurant	Italian Restaurant	Bakery	Cosmetics Shop	Seafood Restaurant	French Restaurant	Restaurant	Dessert Shop	Sandwich Place
31	East Central	3	Hotel	Japanese Restaurant	Italian Restaurant	Bakery	Cosmetics Shop	Seafood Restaurant	French Restaurant	Restaurant	Dessert Shop	Sandwich Place
32	North Central	3	Hotel	Japanese Restaurant	Italian Restaurant	Bakery	Cosmetics Shop	Seafood Restaurant	French Restaurant	Restaurant	Dessert Shop	Sandwich Place

6. DISCUSSION

Calgary is a bubbling cosmopolitan city located in the western Canadian province of Alberta. It is located in the southern part of the province about 300km south of the provincial capital, Edmonton. Its rapid growth has been partly due to it being the center of the Canadian oil industry. The current population is approximately 1.6 million. It is also a major attraction for teeming immigrant populations looking to settle in vibrant urban communities, surrounded by nature and scenic beauty. There are several immigrant and tourist attractions in the city. These include diverse cuisines, restaurants, coffee shops, playgrounds, bars, pizza place, parks, hotels, liquor stores, pubs, gas stations, etc. The cuisine on offer reflects the multicultural leanings of the city. Variants available include American, Mexican, Chinese, Japanese, Asian, Vietnamese and Italian.

The first cluster in Calgary boasts several neighborhoods where residents can buy fresh Pizza. Also, on offer are several convenience stores with provisions and care products. Neighborhoods in this cluster include Lynnwood Ridge, Ogden, Hawkwood, Queensland, Lake Bonavista, Willow Park, Acadia, Altadore Bankview and Richmond. This segment seems not to have several public transportation features like bus stops. Hence arrangements to rent cars rental or use taxis may be unavoidable.

The second cluster of neighborhoods in Calgary is the biggest. The attractions and amenities on offer reflect a multicultural segment of the city with diverse cuisines, pubs, coffee shops, pizza place, convenience stores, etc. For example, bowling alleys are very common in Lakeview, Glendale, Killarney and Glamorgan neighborhoods. Hotels are the most common venues for Inglewood, Burnsland, Chinatown and East Victoria. Other locations with hotels include Northeast Calgary, Rundle, Whitehorn and Monterey Park. Coffee lovers will certainly appreciate the proximity of neighborhoods such as Brentwood, Collingwood, Nosehill, Symons Valley, Evergreen and Somerset to coffee shops.

A seasonal flea market is available in Northwest Calgary, a neighborhood in the third cluster, while other neighborhoods in clusters 3, 4 and 5 namely Elbow Park, Park Hill, Britannia and Mission will be of particular interest to Yoga enthusiasts, given the prevalence of Yoga studios in the area.

Vancouver is a vibrant west coast seaport in the Canadian province of British Columbia. It is one of the cities with the highest population density and most ethnically diverse. The current population is approximately 2.6 million. Due to it scenic horizons and beautiful beaches, Vancouver is a popular filming destination. It is surrounded by highlands and also has thriving art, theatre and music scenes. Attractions in Vancouver include art galleries with works by regional artists and museums boasting renowned First Nations collections. Similar to Calgary, it also includes diverse cuisines, restaurants, coffee shops, playgrounds, bars, pizza place, parks, hotels, liquor stores, pubs, gas stations, etc. The cuisines on offer reflect the multicultural nature of the city with exciting options including American, Mexican, Chinese, Japanese, Asian, Vietnamese and Italian meals.

Coffee shops the most common venue in the first Vancouver cluster. Neighborhoods such as North Grandview, East Mount Pleasant, UBC, West Kerrisdale, East Fairview , South Cambie have several coffee shops. North Hastings-Sunrise is a neighborhood with Theme park rides. Golf Courses are available in the vicinity of West Kerrisdale , South Dunbar-Southlands and Musqueam.

The third cluster in Vancouver has abundant hotels offerings. Also, within this cluster, there is a remarkable prevalence of Japanese and Italian cuisines. Other attractions are cosmetic shops, bakeries and parks. The Pacific Center neighborhoods offer delicious sea food options for visitors.

In the fourth Vancouver cluster, the availability of several bus stops in SE Oakridge , East Marpole , South Sunset and Killarney neighborhoods is indicative of the relative ease of commuting in these locations. The generally high population density also suggests car rentals or taxis may not be efficient forms of transportation and may attract premiums.

7. CONCLUSION

The objective of this project was to review the attractions in Calgary and Vancouver with the aim of gaining insights useful for assisting potential immigrants and tourists in making informed decisions on where to stay and what to do while in those cities. During this project, the postal codes of the city neighborhoods were used as proxies to ultimately obtain information on the most common venues in the neighborhoods. The K-means algorithm was used to segment the cities into clusters highlighting the main attractions in the different clusters.

It was evident from the result of the project that the neighborhoods of the two cities are similarly diverse and provide a wide range of attractions for immigrants and tourists. From the review of the different clusters, examples of neighborhoods and their attractions (within 700m radius), were clearly highlighted giving potential visitors examples of available amenities and activities.

Factors such as climate, average income, cost of living, availability of seaports and beaches are other reasons that can possibly influence customers. A comparison of the two cities in light of these factors is an interesting area for further data gathering and analyses.

REFERENCES

1. Postal Codes for Calgary: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_T
2. Postal Codes for Vancouver: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_V
3. Foursquare API: https://developer.foursquare.com/
4. ArcGIS API: https://developers.arcgis.com/python/