Analyze and Compare Tech Hubs in Canada

Presentation by: Shivani Sheth

Abstract

In this project, we take leverage of the Foursquare API to find the most popular places in each of the neighborhoods consisting of British Columbia. A place is marked as “happening” by the Foursquare API according to the number of people present at a given place and hence the place is updated in real-time; it might change every few minutes. We then cluster the neighborhoods based upon their preferred places in the surrounding area. This will give a clear picture of the aura or the vibe of the place, which can help an individual know what to expect in the neighborhood, and hence decide upon a suitable location according to their preference. Finally, we also compare it with Toronto, another widely preferred city among prospective students, and brief upon the similarities and dissimilarities based upon its neighborhood.

Data Acquisition and Cleaning

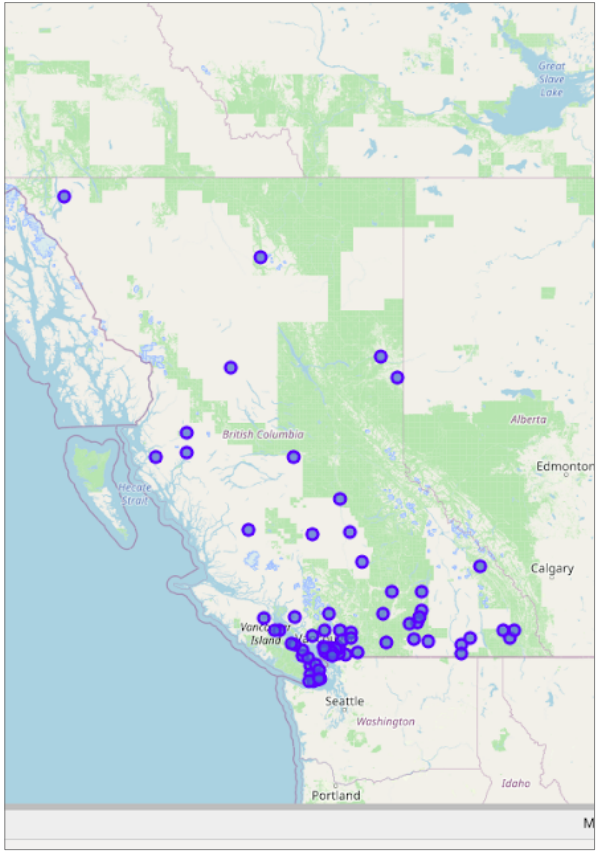
- Table scrapped from libraryhttps://www.geonames.org/postal-codes/CA/BC/british-columbia.html using the read_html function of the Pandas.
- Extract alternate address rows of this table in a separate data frame, say ‘temp_df’. For this, we loop over the number of records in a multiple of two and fetch each row. This will give us a total of 192 rows containing the neighborhood address of British Columbia.
- Delete all except one column for the coordinates and merge the temp_df with the original df.
- Separate latitude and longitude values from the coordinates string and convert into float.

Cleaned and Pre-processed Data

	Coordinates	Place	Code	Country	Province	Latitude	Longitude
0	49.323/-122.863	Port Moody	V3H	Canada	British Columbia	49.323	-122.863
1	49.221/-122.69	Pitt Meadows	V3Y	Canada	British Columbia	49.221	-122.69
2	49.026/-122.806	White Rock	V4B	Canada	British Columbia	49.026	-122.806
3	49.481/-119.586	Penticton	V2A	Canada	British Columbia	49.481	-119.586
4	49.866/-119.739	Westbank	V4T	Canada	British Columbia	49.866	-119.739

Methodology

- Outline the British Columbia map and superimposing all of the neighborhoods, from our data frame, onto the map.
- Utilize the Folium library to visualize the map.
- Extract the coordinates of the British Columbia province through the geocoder package of the geopy library.



Methodology - Explore

- Here, we leverage the Foursquare API in this section to explore the nearby venues of each neighborhood.
- For each set of neighborhood names and coordinates, we prepare a URL request string and send a GET request. The output is the JSON file, from which we extract the nearby locations and store them in a data frame.
- Check the number of venues returned for each neighborhood and observe that we have 76 unique venues.

:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Abbotsford	49.0625	-122.3125	Discovery Trail	49.060245	-122.315565	Trail
1	Abbotsford	49.0625	-122.3125	Grandmas Market Gladwin Rd	49.066149	-122.313659	Grocery Store
2	Burnaby	49.2500	-123.0000	BCITSA's Stand Central SE2	49.251424	-123.001384	Snack Place
3	Burnaby	49.2500	-123.0000	BCIT Bookstore	49.251548	-123.001364	Bookstore
4	Burnaby	49.2500	-123.0000	The Rix @ BCIT	49.251153	-123.000636	Coffee Shop

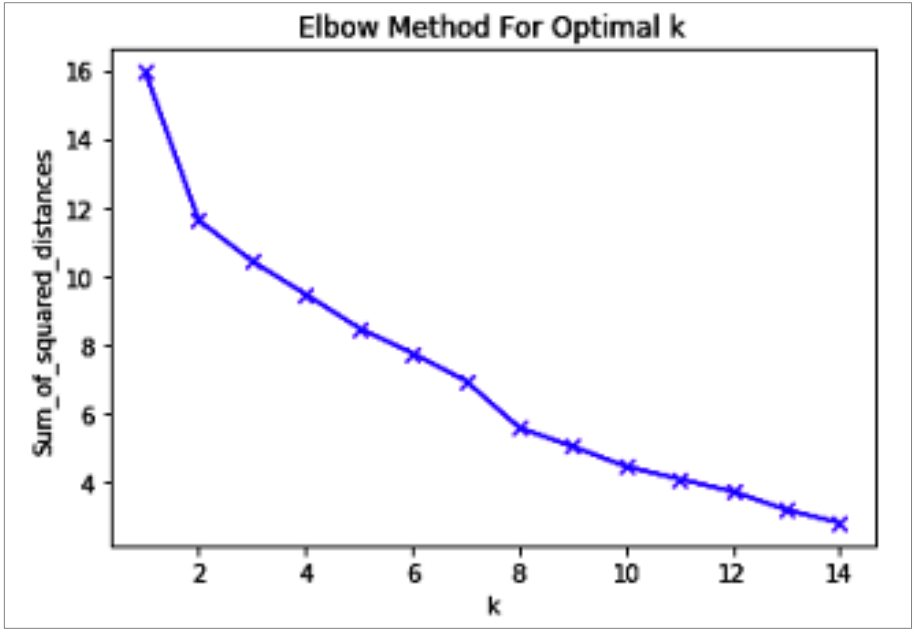
Methodology - Analyze

- To analyze each neighborhood, we use the one-hot encoding technique and map the nearby venue categories against each neighborhood into a series of 0's and 1's. The categories that are present in the neighborhood are marked by 1, and the ones that aren't present in that neighborhood are marked by 0.
- Add the neighborhood names corresponding to their values in the data frame and calculate the mean of venue categories present.
- Arrange the top ten venues of each neighborhood and display it in a separate data frame

	Neighborhood	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	Abbotsford	Grocery Store	Trail	Falafel Restaurant	Coffee Shop	Construction & Landscaping	Convenience Store	Dessert Shop	Dim Sum Restaurant	Dog Run	Elementary School
1	Burnaby	Bus Stop	Bookstore	Snack Place	Park	Bus Station	Burger Joint	Sandwich Place	Coffee Shop	Fast Food Restaurant	Falafel Restaurant
2	Comox	Fast Food Restaurant	Coffee Shop	Pharmacy	Sandwich Place	Juice Bar	Elementary School	Construction & Landscaping	Convenience Store	Dessert Shop	Dim Sum Restaurant
3	Coquitlam	Asian Restaurant	Convenience Store	Golf Course	Gas Station	Coffee Shop	Zoo	Falafel Restaurant	Dessert Shop	Dim Sum Restaurant	Dog Run
4	Cranbrook	Construction & Landscaping	Zoo	Fast Food Restaurant	Convenience Store	Dessert Shop	Dim Sum Restaurant	Dog Run	Elementary School	Falafel Restaurant	Fish & Chips Shop

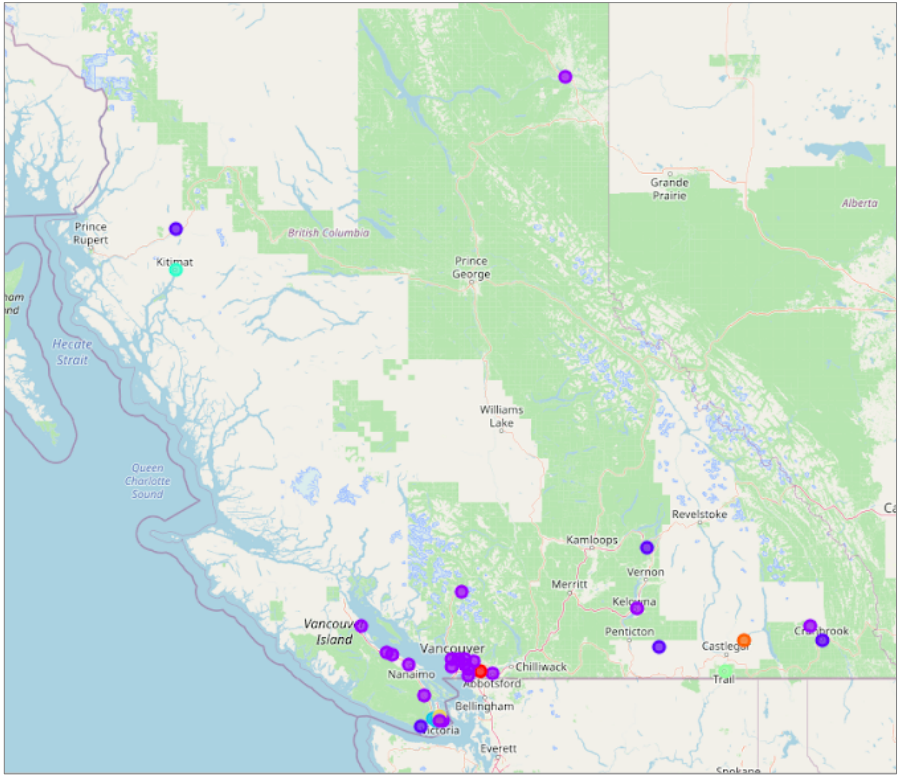
Methodology - Cluster

- Cluster neighborhoods according to their top most common venues using the k-means clustering algorithm.
- Use the elbow method for choosing the optimal number of clusters, or ‘k’.
- The elbow point is k=8. We run the k-means to cluster the neighborhood into eight different clusters.



Methodology - Cluster

- Extract the cluster labels for each neighborhood and append it to the ‘top venues’ data frame.
- Append the postal codes and coordinates from our original data frame into this new ‘top venues’ data frame.
- Visualize the neighborhood clusters superimposed onto the British Columbia map marked by different cluster colors.



Results

British Columbia Cluster 1:

(1, 11)

	Place	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Langley City	Baseball Field	Zoo	Fish & Chips Shop	Convenience Store	Dessert Shop

British Columbia Cluster 2:

(20, 11)

	Place	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Abbotsford	Grocery Store	Trail	Falafel Restaurant	Coffee Shop	Construction & Landscaping
1	Burnaby	Bus Stop	Bookstore	Snack Place	Park	Bus Station
2	Comox	Fast Food Restaurant	Coffee Shop	Pharmacy	Sandwich Place	Juice Bar
3	Coquitlam	Asian Restaurant	Convenience Store	Golf Course	Gas Station	Coffee Shop
4	Duncan	Convenience Store	Gas Station	Dog Run	Zoo	Fast Food Restaurant

Results

British Columbia Cluster 3:

(5, 11)

	Place	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Cranbrook	Construction & Landscaping	Zoo	Fast Food Restaurant	Convenience Store	Dessert Shop
1	Salmon Arm	Construction & Landscaping	Zoo	Fast Food Restaurant	Convenience Store	Dessert Shop
2	Sooke	Construction & Landscaping	Zoo	Fast Food Restaurant	Convenience Store	Dessert Shop
3	South Okanagan	Construction & Landscaping	Zoo	Fast Food Restaurant	Convenience Store	Dessert Shop
4	Terrace	Construction & Landscaping	Zoo	Fast Food Restaurant	Convenience Store	Dessert Shop

British Columbia Cluster 4:

(2, 11)

	Place	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Esquimalt	Boat or Ferry	Fish & Chips Shop	Convenience Store	Dessert Shop	Dim Sum Restaurant
1	Highlands	Zoo	Theme Park	Boat or Ferry	Wine Shop	Auto Workshop

Results

British Columbia Cluster 5:

(1, 11)

	Place	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Kitimat	Business Service	Zoo	Fast Food Restaurant	Convenience Store	Dessert Shop

British Columbia Cluster 6:

(1, 11)

	Place	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Trail	Pub	Falafel Restaurant	Coffee Shop	Construction & Landscaping	Convenience Store

British Columbia Cluster 7:

(1, 11)

	Place	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Saanich Central	Bank	Zoo	Fish & Chips Shop	Convenience Store	Dessert Shop

Conclusion

- The neighborhood clusters define the type of locality and can be used by different people to choose their preferred area of interest accordingly. For example, clusters involving a locality that has baseball fields, zoos in their most popular nearby venues and hence would be more suitable for a younger generation, preferably kids in their middle schools. On the contrary, the clusters involving a locality that has grocery shops, fast food shops, and convenience stores as their popular nearby locations would be more suitable for college students, who would need all of these on an almost daily basis.

- Hence, these clusters provide a basic guide or a map on the different types of neighborhoods in British Columbia and give an idea of what to expect in that neighborhood.