Analysis of prospective new Business Locations in the city of Vancouver, Canada.

Introduction -

One of the key challenges faced by businesses today is to identify prospective locations to set up offices. Various factors come into play in taking this decision. The issue is especially demanding for the MNCs that frequently need to set up offices at new locations across the globe. Although the decision making process is highly subjective, the process largely depends on a common factor - “Businesses need the data regarding the amenities at different areas in a target city !!”

Getting such data is of use particularly for International Businesses that have never been players in that city/area. Getting this data will help in having an idea about the different factors that affect business. For an example, a Travel / Tourism start-up might prefer a location that is close to a major bus terminus, railway station, airport or a place where often tourism fairs are organised. A restaurant chain might look to expand its business in another part of the city where other providers are available and the footfall is high!! They might try to lure in customers of a competitive business by opening a branch nearby. But for all these to happen, they need authentic up-to-date information on the particulars of an area.

In this document, we will discuss one such case study where we develop a model to offer strategic business locations for businesses - both new and existing by choosing a particular area (Postal Code) in the city of Vancouver, Canada, as a prospective site for them to open office and expand their business.

The data used –

As discussed in the last paragraph, we concluded that we will develop a model that is able to showcase the characteristics of a particular area in the city of Vancouver. For that to happen, we need a list of all the areas in that city. In our case, our first aim is to obtain the postal code addresses of the

city of Vancouver from Zip-Codes.com. We then convert the relevant data into a Data Frame using pandas. Now, we need to obtain the latitude and longitude values of these postal areas. For that we use Google's Geocoding API. We add the newly fetched latitude and longitude values into the Data Frame against relevant Postal Codes and pass each such row to the Foursquare API to obtain the Venues around those locations. In this way, we are able to gain insights into the Postal Areas by looking at the venue categories.

We can then cluster the Data Frame into groups based on the data pertaining to the venues and when the user inserts his/her priority (such as location near banks, rental spaces, event spaces, or restaurants), the best

locations against the respective clusters are displayed, thus helping him/her decide the best venue for his/her business.

Detailed Methodology –

To start the process, the first step is to obtain the values of the Postal Codes in the city of Vancouver. In this specific case, we use zip-codes.com to obtain this data.

```
['V5K 0A1', 'V5K 0A2', 'V5K 0A3', 'V5K 0A4', 'V5K 0A5', 'V5K 0A6', 'V5K 0A7', 'V5K 0A8', 'V5K 0A9', 'V5K 0B1',
'V5K 0B2', 'V5K 0B3', 'V5K 0B4', 'V5K 0B5', 'V5K 0B6', 'V5K 0B7', 'V5K 0B8', 'V5K 0B9', 'V5K 0C1', 'V5K 0C2',
'V5K 0C3', 'V5K 0C4', 'V5K 0C5', 'V5K 0C6', 'V5K 0C7', 'V5K 0C9', 'V5K 0E1', 'V5K 0E2', 'V5K 0E3', 'V5K 0E4',
'V5K 0E5', 'V5K 0E6', 'V5K 0E7', 'V5K 1A1', 'V5K 1A4', 'V5K 1A5', 'V5K 1A6', 'V5K 1A7', 'V5K 1A8', 'V5K 1A9',
'V5K 1B1', 'V5K 1B2', 'V5K 1B3', 'V5K 1B4', 'V5K 1B5', 'V5K 1B6', 'V5K 1B7', 'V5K 1B8', 'V5K 1B9', 'V5K 1C1',
'V5K 1C3', 'V5K 1C4', 'V5K 1C5', 'V5K 1C6', 'V5K 1C7', 'V5K 1C8', 'V5K 1C9', 'V5K 1E1', 'V5K 1E2', 'V5K 1E3',
'V5K 1E4', 'V5K 1E5', 'V5K 1E6', 'V5K 1E7', 'V5K 1E8', 'V5K 1E9', 'V5K 1G1', 'V5K 1G2', 'V5K 1G3', 'V5K 1G4',
'V5K 1G5', 'V5K 1G6', 'V5K 1G7', 'V5K 1G8', 'V5K 1G9', 'V5K 1H1', 'V5K 1H2', 'V5K 1H3', 'V5K 1H4', 'V5K 1H5',
'V5K 1H6', 'V5K 1H7', 'V5K 1H8', 'V5K 1H9', 'V5K 1J1', 'V5K 1J2', 'V5K 1J3', 'V5K 1J4', 'V5K 1J5', 'V5K 1J6',
'V5K 1J7', 'V5K 1J8', 'V5K 1J9', 'V5K 1K1', 'V5K 1K2', 'V5K 1K3', 'V5K 1K4', 'V5K 1K5', 'V5K 1K6', 'V5K 1K7',
'V5K 1K8', 'V5K 1K9', 'V5K 1L1', 'V5K 1L2', 'V5K 1L3', 'V5K 1L4', 'V5K 1L5', 'V5K 1L6', 'V5K 1L7', 'V5K 1L8',
'V5K 1L9', 'V5K 1M2', 'V5K 1M3', 'V5K 1M4', 'V5K 1M5', 'V5K 1M6', 'V5K 1M7', 'V5K 1M8', 'V5K 1M9', 'V5K 1N1',
'V5K 1N2', 'V5K 1N3', 'V5K 1N4', 'V5K 1N5', 'V5K 1N6', 'V5K 1N8', 'V5K 1N9', 'V5K 1P1', 'V5K 1P2', 'V5K 1P3',
'V5K 1P4', 'V5K 1P5', 'V5K 1P6', 'V5K 1P7', 'V5K 1P8', 'V5K 1P9', 'V5K 1R1', 'V5K 1R2', 'V5K 1R3', 'V5K 1R4',
'V5K 1R5', 'V5K 1R8', 'V5K 1R9', 'V5K 1S1', 'V5K 1S2', 'V5K 1S3', 'V5K 1S4', 'V5K 1S5', 'V5K 1S6', 'V5K 1S7',
'V5K 1S8', 'V5K 1S9', 'V5K 1T1', 'V5K 1T2', 'V5K 1T3', 'V5K 1T4', 'V5K 1T5', 'V5K 1T8', 'V5K 1T9', 'V5K 1V1',
```

After obtaining the Postal Codes for the city of Vancouver, Canada, we need to obtain the latitude and longitude values against each such Postal Area in order for us to get details / specifics about a particular Postal Area from the Foursquare API. For this task, we take the help of the Google Geocoding API. We pass each and every Post Code from the list we created

to the Google API and obtain the respective co-ordinates. Thereafter, we create one final Data Frame as follows -

	postalcode	neighborhood	latitude	longitude
0	V5K 0A1	Vancouver	49.291018	-123.040352
1	V5K 0A2	East Vancouver	49.285023	-123.028471
2	V5K 0A3	East Vancouver	49.280953	-123.028838
3	V5K 0A4	East Vancouver	49.280918	-123.051041
4	V5K 0A5	East Vancouver	49.272429	-123.052170

The next step is to obtain the details/specifics pertaining to each such Postal Area – such as the most popular venues around each location. For this, we pass the latitude and longitude co-ordinates to the Foursquare API and

obtain the resulting JSON file. We then parse that file and make another data frame comprising the postal areas along with the venues in those areas.

	Postal Areas	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	V5K 0A1	49.291018	-123.040352	New Brighton Park	49.289614	-123.038484	Park
1	V5K 0A1	49.291018	-123.040352	New Brighton Park Beach	49.290694	-123.039160	Beach
2	V5K 0A1	49.291018	-123.040352	New Brighton Park Pool	49.289343	-123.036673	Pool
3	V5K 0A1	49.291018	-123.040352	Cascadia Terminal	49.290940	-123.033511	Pier
4	V5K 0A2	49.285023	-123.028471	7-Eleven	49.280783	-123.027842	Convenience Store
5	V5K 0A2	49.285023	-123.028471	Tim Hortons	49.280771	-123.027835	Coffee Shop
6	V5K 0A2	49.285023	-123.028471	Cassiar Tunnel	49.286071	-123.031796	Tunnel
7	V5K 0A2	49.285023	-123.028471	Bus Stop 50956 (14,28,130,135)	49.281944	-123.029489	Bus Station
8	V5K 0A2	49.285023	-123.028471	McDonald's	49.280857	-123.030541	Fast Food Restaurant
9	V5K 0A2	49.285023	-123.028471	Tiffany N.Y. Bridal	49.280988	-123.026729	Bridal Shop
10	V5K 0A2	49.285023	-123.028471	Kootenay Bus Loop	49.281145	-123.025666	Bus Station
11	V5K 0A2	49.285023	-123.028471	Empire Field	49.282886	-123.034338	Athletics & Sports
12	V5K 0A3	49.280953	-123.028838	Taz Comics	49.281108	-123.025065	Hobby Shop
13	V5K 0A3	49.280953	-123.028838	Wooden Roller Coaster	49.281744	-123.035128	Theme Park Ride / Attraction
14	V5K 0A3	49.280953	-123.028838	La Fontana Caffe	49.281170	-123.023447	Café

But we need to get the frequency of each such venue in a particular area and not just how many or where they are!! We therefore add the code to transform the data frame to include the venues as columns and then take their mean of the venue-categories to determine the frequency.

	Postal Areas	American Restaurant	Amphitheater	Asian Restaurant	Athletics & Sports	BBQ Joint	Bakery	Bank	Bar	Beach	...	Tea Room	Tennis Court	Thai Restaurant	Theater	Theme Park	Theme Park Ride / Attraction	Thrift / Vintage Store	Trail	Tunnel	Vietnamese Restaurant
0	V5K 0A1	0.000000	0.000000	0.000000	0.000	0.000000	0.000000	0.000000	0.0	0.25	...	0.000000	0.0	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	0.000	0.000000
1	V5K 0A2	0.000000	0.000000	0.000000	0.125	0.000000	0.000000	0.000000	0.0	0.00	...	0.000000	0.0	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	0.125	0.000000
2	V5K 0A3	0.000000	0.000000	0.058824	0.000	0.000000	0.000000	0.000000	0.0	0.00	...	0.000000	0.0	0.000000	0.000000	0.000000	0.058824	0.000000	0.0	0.000	0.000000
3	V5K 0A4	0.000000	0.000000	0.000000	0.000	0.000000	0.033333	0.033333	0.0	0.00	...	0.000000	0.0	0.033333	0.000000	0.000000	0.000000	0.000000	0.0	0.000	0.133333
4	V5K 0A5	0.000000	0.000000	0.000000	0.000	0.000000	0.000000	0.000000	0.0	0.00	...	0.000000	0.0	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	0.000	0.125000
5	V5K 0A6	0.000000	0.000000	0.000000	0.000	0.000000	0.000000	0.000000	0.0	0.25	...	0.000000	0.0	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	0.000	0.000000

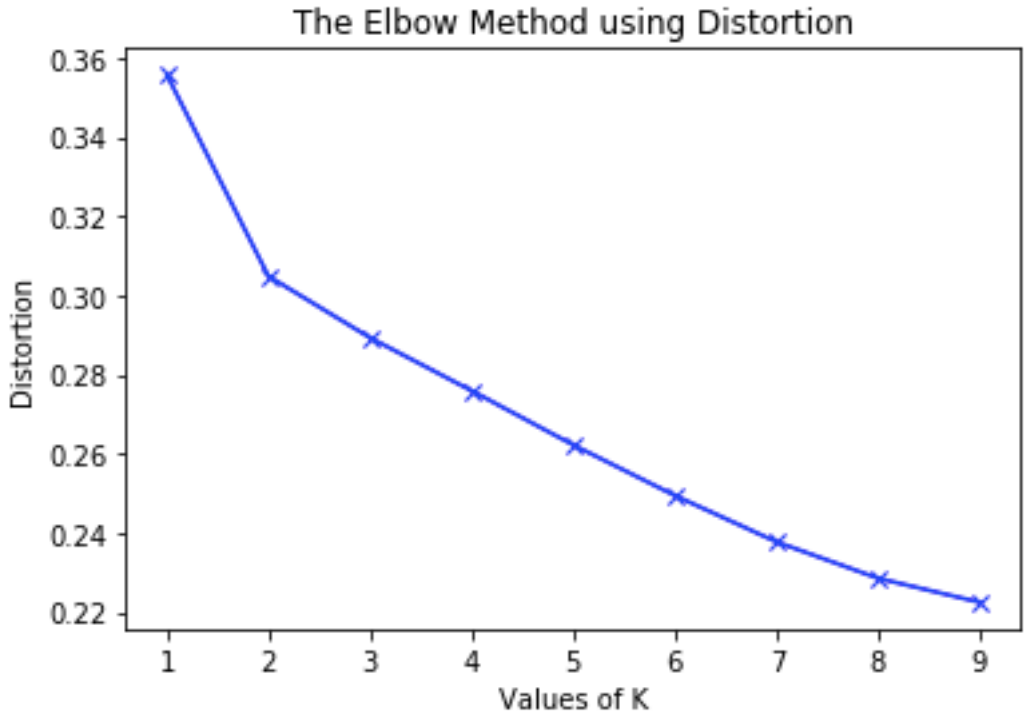
Grouping the data in this manner will let us determine the top venues in a Postal Area.

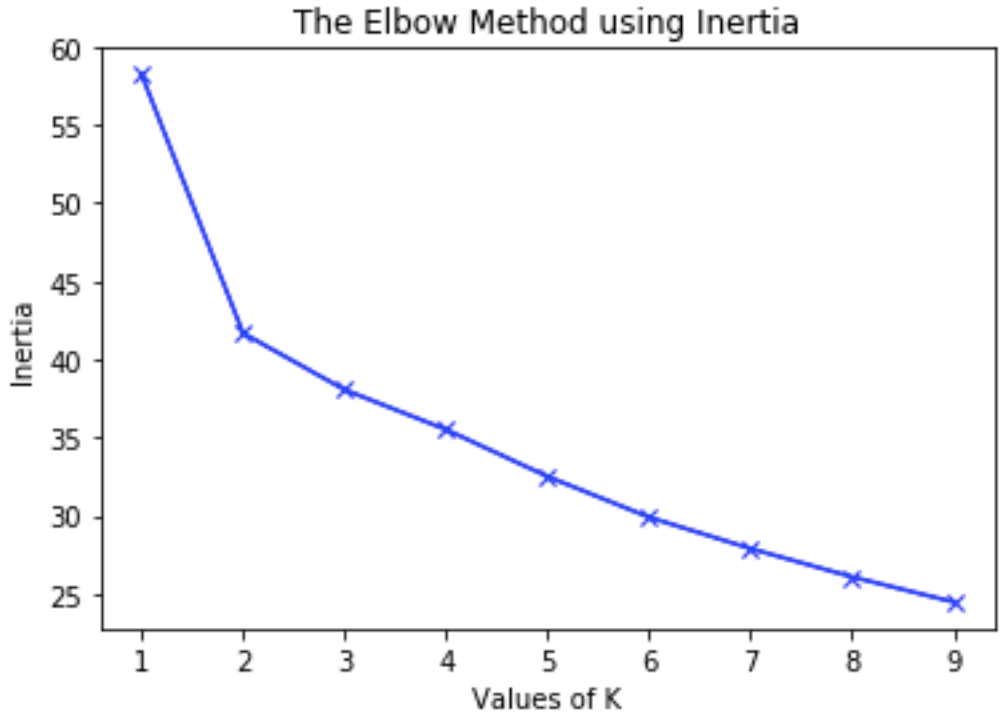
```
----V5K 0A1----
                  venue  freq
0                  Pool  0.25
1                  Pier  0.25
2                 Beach  0.25
3                  Park  0.25
4  American Restaurant  0.00


----V5K 0A2----
                   venue  freq
0            Bus Station  0.25
1                 Tunnel  0.12
2    Athletics & Sports  0.12
3  Fast Food Restaurant  0.12
4     Convenience Store  0.12


----V5K 0A3----
```

Thus, we are now ready to make a new Data Frame with the top 10 venues in a particular postal area as follows -

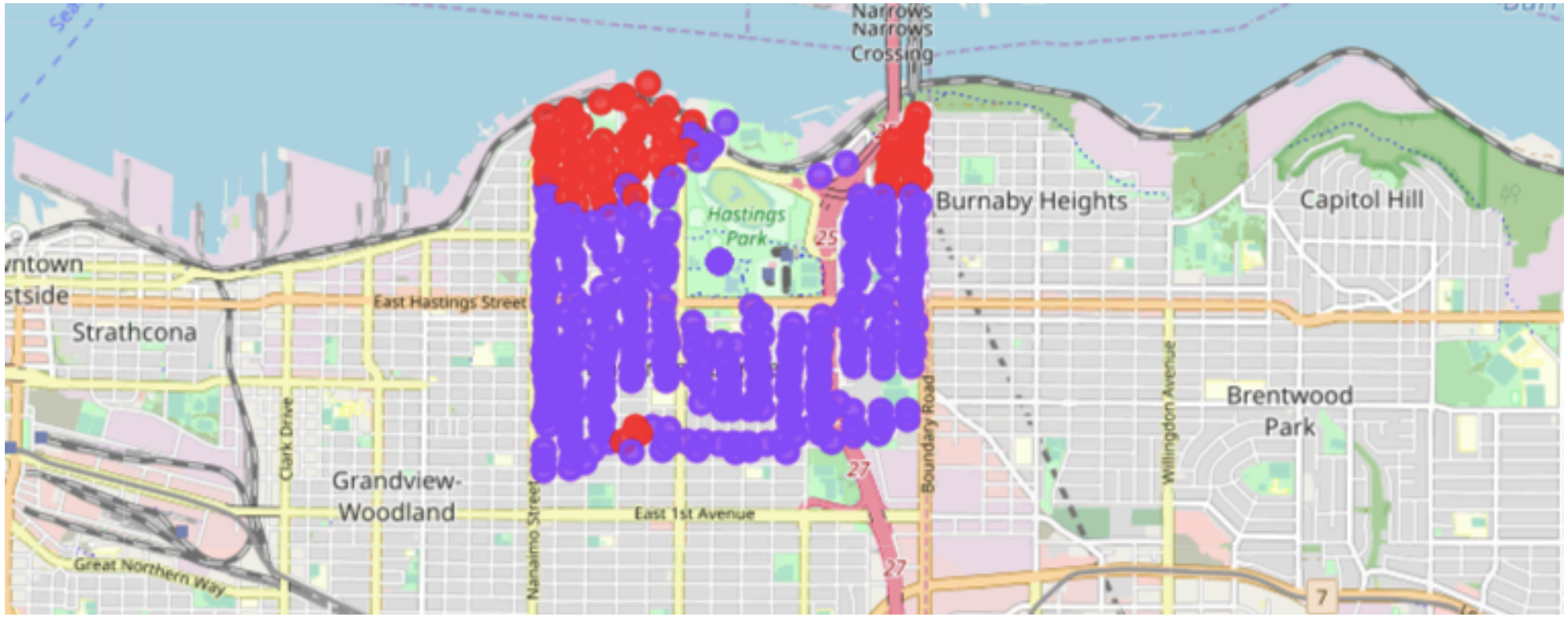
	Postal Areas	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	V5K 0A1	Pool	Pier	Park	Beach	Fast Food Restaurant	Dim Sum Restaurant	Diner	Dog Run	Donut Shop	Event Space
1	V5K 0A2	Bus Station	Convenience Store	Bridal Shop	Fast Food Restaurant	Coffee Shop	Tunnel	Athletics & Sports	Diner	Donut Shop	Event Space
2	V5K 0A3	Soccer Field	Bus Stop	Sporting Goods Shop	Fast Food Restaurant	Park	Coffee Shop	Chinese Restaurant	Café	Pub	Bus Station
3	V5K 0A4	Vietnamese Restaurant	Coffee Shop	Café	Sushi Restaurant	Pharmacy	Burger Joint	Chinese Restaurant	Bank	Diner	Sandwich Place
4	V5K 0A5	Vietnamese Restaurant	Sporting Goods Shop	Gas Station	Massage Studio	Park	Pizza Place	Deli / Bodega	Convenience Store	Bar	Fast Food Restaurant

Now is the time to cluster this Data Frame into groups based on the venue categories using the K-Means machine learning algorithm. But for that, we need to determine the value of K. We determine this value through the Elbow method using both Inertia and Distortion methodologies.


The Elbow Method using Distortion
Distortion
0.36
0.34
0.32
0.30
0.28
0.26
0.24
0.22
1
2
3
4
5
6
7
8
9
Values of K



As is evident through both the above charts, the optimum value of k is 2. Thus we will cluster our dataset using 2 clusters and plot the clusters on the map of Vancouver city.



So now we are able to plot the most popular venues in the city of Vancouver on the map. However, merely having this data is insufficient to help out businesses whose aim to set up offices and expand their business

in this city! Therefore, to help them, we first need to understand the priority of the user.

If it's a restaurant business that is trying to expand in this part of the globe, then it needs to know the areas where existing restaurant chains are operating. It can then choose to establish a base of operations nearer to those existing places and in a way aim to draw a portion of their customer base that frequents that area or choose an alternate strategy based on other factors – such as amusement parks, malls, vegetable stores, etc. – to selectively establish their base of operations at places that does not have good restaurant options.

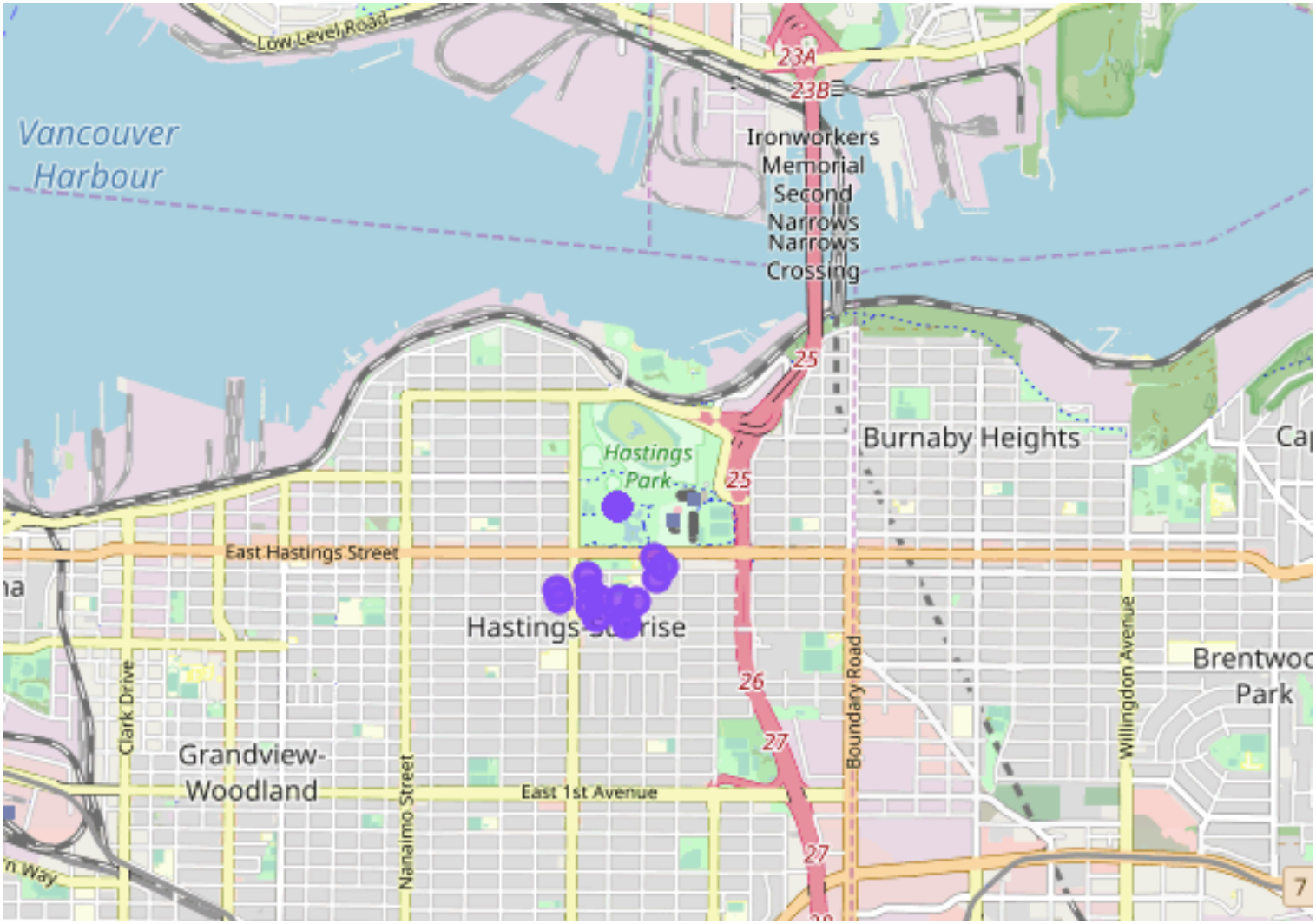
Similarly, if it's an advertising company looking to set foot in Vancouver, a locality that has good Rental Services, Office space availability, Auditoriums to organise seminars and hotels to host parties and events might be a good choice. Since the priorities change with each Business, it is better to know it first.

Hence, we change the model to incorporate the user's priority that will be searched in each cluster, and plotted on the map of the city.

Results -

```
What is your first priority ? NOTE: Enter the words as in the list - data is case sensitive !!Event Space

'Event Space'

The geograpical coordinate of Vancouver are 49.2808724, -123.0339529.
The Cluster 0 Data based on User's Priority -
Empty DataFrame
Columns: [postalcode, 1st Most Common Venue, 2nd Most Common Venue, 3rd Most Common Venue, 4th Most Common Venue, 5th Most Common Venue]
Index: []


The Cluster 1 Data based on User's Priority -
    postalcode         1st Most Common Venue  2nd Most Common Venue  \
265    V5K 2K7                   Event Space  Portuguese Restaurant
291    V5K 2N6                   Event Space  Portuguese Restaurant
199    V5K 2A3  Theme Park Ride / Attraction            Event Space
217    V5K 2C4  Theme Park Ride / Attraction            Event Space
238    V5K 2G7         Vietnamese Restaurant            Event Space
239    V5K 2G8         Vietnamese Restaurant            Event Space
240    V5K 2G9  Theme Park Ride / Attraction            Event Space
263    V5K 2K5         Vietnamese Restaurant            Event Space
264    V5K 2K6         Vietnamese Restaurant            Event Space
8      V5K 0A9  Theme Park Ride / Attraction             Theme Park
11     V5K 0B3  Theme Park Ride / Attraction             Theme Park
12     V5K 0B4  Theme Park Ride / Attraction             Theme Park
13     V5K 0B5  Theme Park Ride / Attraction             Theme Park
14     V5K 0B6  Theme Park Ride / Attraction             Theme Park
15     V5K 0B7  Theme Park Ride / Attraction             Theme Park
16     V5K 0B8  Theme Park Ride / Attraction             Theme Park
17     V5K 0B9  Theme Park Ride / Attraction             Theme Park
18     V5K 0C1  Theme Park Ride / Attraction             Theme Park
19     V5K 0C2  Theme Park Ride / Attraction             Theme Park
20     V5K 0C3  Theme Park Ride / Attraction             Theme Park
21     V5K 0C4  Theme Park Ride / Attraction             Theme Park
22     V5K 0C5  Theme Park Ride / Attraction             Theme Park
23     V5K 0C6  Theme Park Ride / Attraction             Theme Park
24     V5K 0C7  Theme Park Ride / Attraction             Theme Park
```

	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
265	Theme Park Ride / Attraction	Park	BBQ Joint
291	Theme Park Ride / Attraction	Park	BBQ Joint
199	Theme Park	Portuguese Restaurant	Stadium
217	Theme Park	Pizza Place	Soccer Field
238	Bus Line	Inn	Gas Station
239	Burger Joint	Inn	Gas Station
240	Theme Park	Pizza Place	Bridal Shop
263	Bus Line	Inn	Gas Station
264	Sushi Restaurant	Inn	Gas Station
8	Event Space	Hockey Arena	Park
11	Event Space	Hockey Arena	Park
12	Event Space	Hockey Arena	Park
13	Event Space	Hockey Arena	Park
14	Event Space	Hockey Arena	Park
15	Event Space	Hockey Arena	Park
16	Event Space	Hockey Arena	Park
17	Event Space	Hockey Arena	Park
18	Event Space	Hockey Arena	Park
19	Event Space	Hockey Arena	Park
20	Event Space	Hockey Arena	Park
21	Event Space	Hockey Arena	Park


Low Level Road
23A
23B
Vancouver Harbour
Ironworkers Memorial Second Narrows
Narrows Crossing
25
Hastings Park
Burnaby Heights
East Hastings Street
Hastings Sunrise
26
27
Boundary Road
Willingdon Avenue
Brentwood Park
Clark Drive
Grandview-Woodland
Nanaimo Street
East 1st Avenue
7

In the above example, the user entered “Event Space” as his priority, and the system returned all those Postal Areas where Office is in the 1st, 2nd and 3rd most popular venues in all the clusters (0 and 1). A map is plotted with the resulting data and the client is now able to visualize the locations having the most popular "Event Space" destinations.

Discussions -

Note that along with this information, two (02) additional columns as the 4th and 5th most popular destinations have been shown to acquaint users with other popular facilities in that Postal Area and help them in taking a wise choice. For example, in the above result, Office spaces in areas having

a Café might be a wiser choice in order to let employees have some occasional relaxation amidst their workload. Similarly, an ice cream / dessert shop adjoining a restaurant without dessert options might make it a happening spot as well.

Conclusion -

Doing business is a complicated task and more complex is the process to identify prospective locations to expand the business. An entire venture may be ruined if a poor selection is made as a choice for the base of operations. Hence, pre-analysing prospective locations has been a hotbed of debate and research for decades. Any sound help in this area will be

welcomed by any prospective start-up or an existing business looking to expand. Although we did this exercise based on the data pertaining to the city of Vancouver, Canada, this methodology may be applied for almost any city in the world as long as we are able to fetch the Foursquare data for that location.

P.S. - I do not claim to be an expert in the field and this model may be improved in a myriad of ways. As such, if the reader identifies any loop holes and suggests an improvement, it will help me become a better analyser. Please feel free to mail your valuable feedback at – arunabeshc@rediffmail.com

Acknowledgements –

1) www.zip-codes.com

2) Google Geocoding API

3) Foursquare API

4) Coursera Teaching Staff