

# Coursera Capstone Project: Office Supplier Expansion - New Location Data Analysis

## 1. Introduction

The client is an Office Supplier in Toronto which services small to medium offices. They have saturated the market in their own area and are now looking to expand into a nearby city. The two options under consideration by the client are Quebec and Montreal. As this will be the client's first business expansion, it is important that out of the two possible locations, they choose the city offering the best possible outcome.

During this process another aspect of the client's business has also been taken into consideration. This is the fact that potential customers expect delivery within the hour, so any business location would have to be close to a number of customer sites.

The data science problem can be described using the following questions:

PART ONE: Which out of two possible locations has the greater number of potential customers?

PART TWO: In the city chosen in answer to part one, which city borough would offer the greatest catchment area for the business?

***This analysis is also targeted at any business looking to expand into a new location as the same process of data analysis would apply. It is vital for any business planning an expansion to look for locations with established potential customers in order to make it a success***

## 2. Data

Data collection utilised Foursquare data. The API was used to gather potential customer data in both cities.

For example: Foursquare data can be filtered to find potential customers using category ID's such as Offices, Business Centres etc. The extracted data can then be imported as .json files into data frames ready for preparation.

Data extracted included features such as city/borough location, business type, name along with their longitude and latitude co-ordinates. Borough postal codes were also scraped from Wikipedia. Collected data was then prepared and cleaned before analysis using data frames and mapping techniques using Folium. These processes established which city has a greater number of potential customers.

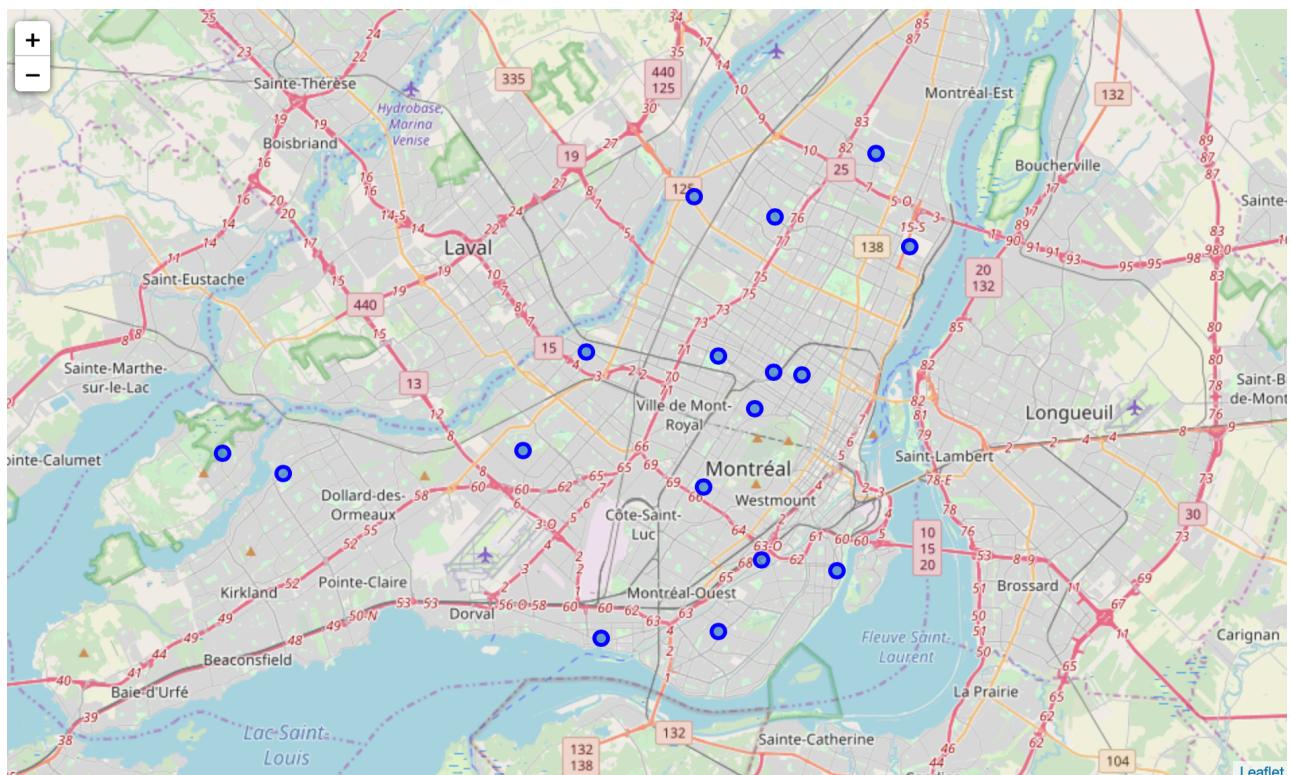
For the second part of the Data Science problem, potential customers within the selected city were then segmented and clustered. The results overlaid on a map of the city. The data maps and clustering results established the ideal borough within the city for the client's business expansion.

### 3. Methodology

For **PART ONE** the process for each city was to data scrape from wikipedia the boroughs and postcodes of each city location importing them as a .csv file before creating a data frame. Longitude and Latitude co-ordinates for each borough were also imported into a data frame. Both data frames were then merged together (see montreal\_merged data frame below).

	Postal code	Borough	Latitude	Longitude
0	H3M	Ahuntsic-Cartierville	45.53856	-73.69266
1	H1K	Anjou	45.60933	-73.54508
2	H3W	Côte-des-Neiges–Notre-Dame-de-Grâce	45.49069	-73.63321
3	H8S	Lachine	45.43665	-73.68510
4	H8N	Lasalle	45.43890	-73.62583

Geopy library was utilised to get the latitude and longitude values of each city which was then used along with the merged Borough data frame to create a map with the boroughs superimposed on top.



The Foursquare API was then used to create a .json file of potential customers within a radius of 5 miles of the centre of each city. I chose this radius as it was large enough to give an indication of potential customer numbers within each city. The resulting output was achieved by using a specific category ID within the foursquare URL, in this case 'Offices' the foursquare category included subcategories of 'Co-working Space, Tech Startups, Advertising Agency, Campaign Office, Corporate Amenity'. The results were then imported into its own data frame. Making particular note of the data frames size each time.

		name	categories	lat	lng
0		Absolu	Office	46.814336	-71.223545
1		Le Soleil	Office	46.813975	-71.224300
	I output; double click to hide	CGI Québec	Office	46.813938	-71.224311
3	Festival d'été de Québec - siège social		Office	46.815324	-71.221039
4	CSN Conseil Central Québec		Office	46.812067	-71.227938
<b>nearby_offices_quebec.shape</b>					
: (43, 4)					

Following the same process for each city location, resulted in a clear choice the city of Montreal to explore further using data analysis in PART TWO.

**PART TWO** was focused on the analysis of the chosen city's individual boroughs in order to find the greatest catchment area of potential customers. Once again Foursquare API was utilised, using the specific category ID of 'Offices' but this time for each borough within Montreal.

Borough	1st Most Common Office	2nd Most Common Office	3rd Most Common Office
8	Rosemont-La Petite-Patrie	Office	Coworking Space
9	Saint-Laurent	Office	Advertising Agency
10	Verdun	Office	Tech Startup
11	Ville-Marie	Office	Tech Startup
12	Villeray-Saint-Michel-Parc-Extension	Office	Advertising Agency

Each borough was segmented and then clustered using K-means clustering. To find the borough with the highest potential customer numbers and their business type.

Borough	Latitude	Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category	Cluster Labels	1st Most Common Office	2nd Most Common Office	3rd Most Common Office
0	Ahuntsic-Cartierville	45.53856	-73.69266	OSHARA INC	45.536596	-73.693092	Office	3	Office	Coworking Space
1	Ahuntsic-Cartierville	45.53856	-73.69266	Peter K Photography - (Montreal Wedding Photog...)	45.541345	-73.695517	Office	3	Office	Coworking Space
2	Ahuntsic-Cartierville	45.53856	-73.69266	santé de pieds podologie	45.542016	-73.694763	Coworking Space	3	Office	Coworking Space
3	Côte-des-Neiges–Notre-Dame-de-Grâce	45.49069	-73.63321	TD Canada Trust	45.491823	-73.631786	Office	0	Office	Tech Startup
4	Lachine	45.43665	-73.68510	Vitrerie Chatelle/Simard	45.438562	-73.681563	Office	3	Office	Coworking Space

A map was created of Montreal with the clusters overlaid. This and the resulting data clearly revealed the borough's with the highest business potential for the client and through analysis of the results narrowed the choice to one borough in particular (see Results section).

## 4. Results

### PART ONE: Choice of City Location

Results of Part One	
:	<pre>{} potential customers were returned by Foursquare in Montreal.'.format(nearby_offices_montreal.shape[0]))\nprint('{} potential customers were returned by Foursquare in Quebec.'.format(nearby_offices_quebec.shape[0]))</pre>
	100 potential customers were returned by Foursquare in Montreal. 43 potential customers were returned by Foursquare in Quebec.

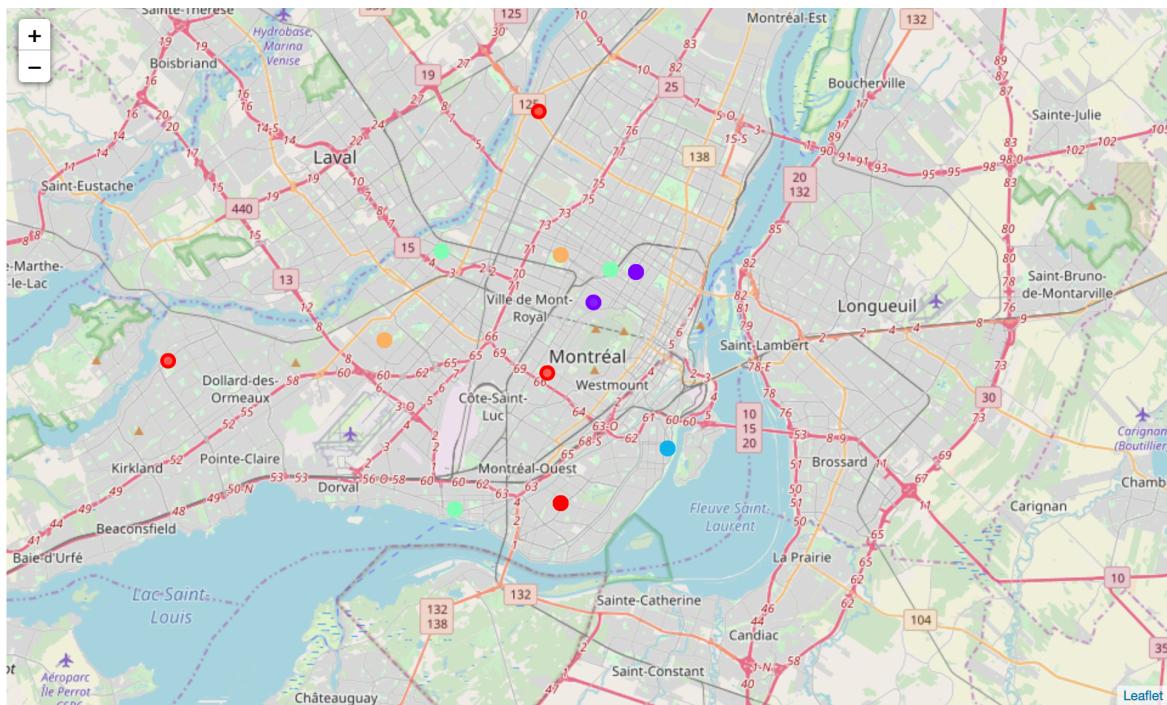
Following data analysis of the two location choices of Montreal and Quebec, there was a clear winner. **Montreal** having a larger percentage of potential customer numbers available to the client.

### PART TWO: Choice of City Borough Location

Further analysis of the boroughs of Montreal revealed a cluster containing more potential customers than the others. This was 'Cluster 4' (see table below) . This particular cluster contained the greatest concentration of potential customers for the client. Potential customers in this cluster are in the form of Tech Startups and Co-working Offices. Both customer types fall within the small/ medium office preferred by the client.

Cluster 4: Coworking/ Tech Startup								
In [140]:	i.loc[montreal_merged['Cluster Labels'] == 3, montreal_merged.columns[[1] + list(range(5, montreal_merged.shape[1]))]]							
Out[140]:	Latitude	Venue	Longitude	Venue Category	Cluster Labels	1st Most Common Office	2nd Most Common Office	3rd Most Common Office
0	45.538560		-73.693092	Office	3	Office	Coworking Space	Tech Startup
1	45.538560		-73.695517	Office	3	Office	Coworking Space	Tech Startup
2	45.538560		-73.694763	Coworking Space	3	Office	Coworking Space	Tech Startup
4	45.436650		-73.681563	Office	3	Office	Coworking Space	Tech Startup
5	45.436650		-73.682269	Office	3	Office	Coworking Space	Tech Startup
6	45.436650		-73.690605	Office	3	Office	Coworking Space	Tech Startup
7	45.436650		-73.683516	Coworking Space	3	Office	Coworking Space	Tech Startup
8	45.436650		-73.681848	Office	3	Office	Coworking Space	Tech Startup
27	45.531567		-73.599026	Office	3	Office	Coworking Space	Tech Startup
28	45.531567		-73.598421	Office	3	Office	Coworking Space	Tech Startup
29	45.531567		-73.598641	Office	3	Office	Coworking Space	Tech Startup
30	45.531567		-73.597508	Office	3	Office	Coworking Space	Tech Startup
31	45.531567		-73.596386	Office	3	Office	Coworking Space	Tech Startup
32	45.531567		-73.603017	Office	3	Office	Coworking Space	Tech Startup
33	45.531567		-73.596758	Office	3	Office	Coworking Space	Tech Startup
34	45.531567		-73.596854	Office	3	Office	Coworking Space	Tech Startup
35	45.531567		-73.603105	Office	3	Office	Coworking Space	Tech Startup
36	45.531567		-73.597240	Office	3	Office	Coworking Space	Tech Startup
37	45.531567		-73.596552	Office	3	Office	Coworking Space	Tech Startup
38	45.531567		-73.598069	Office	3	Office	Coworking Space	Tech Startup
39	45.531567		-73.598134	Office	3	Office	Coworking Space	Tech Startup
40	45.531567		-73.596160	Office	3	Office	Coworking Space	Tech Startup
41	45.531567		-73.595443	Office	3	Office	Coworking Space	Tech Startup

'Cluster 4' consists of two boroughs which are Villeray–Saint-Michel–Parc-Extension and Saint-Laurent. The client only wanted one borough location, so I overlaid each cluster onto a map of



Montreal. ('Cluster 4' boroughs are in orange on the map)

From this data, the city borough which would offer the greatest catchment area for the clients business is **Villeray–Saint-Michel–Parc-Extension**. This borough was chosen from the two possibilities because of its close proximity to three other cluster centres (see map above). The borough of Saint-Laurent, while having a large concentration of potential customers is isolated and would reduce opportunities for future business growth.

## 5. Discussion

Through using the Foursquare API , I was able to analyse a large data set for each of my locations. This would ensure that any recommendations made as a result of the data analysis could be made with confidence. However for more detailed analysis the data set could be explored further, even down to neighbourhood level. I have found that maps are also useful for visualising data when presenting your findings to the client. Representing in a single image extensive data analysis.

### Recommendation based on the results:

After data collection, cleaning and analysis I can now answer the initial data science problem posed in the introduction of this project.

**PART ONE:** Which out of two possible locations (Quebec and Montreal) has the greater number of potential customers?

**Solution:** Montreal as it offers the greater number of potential customers for the client.

**PART TWO:** In the city chosen in answer to PART TWO, in this case Montreal, which city borough would offer the greatest catchment area for the business?

**Solution:** The city borough of Montreal which would offer the greatest catchment area for the clients business is Villeray–Saint-Michel–Parc-Extension.

## **6. Conclusion**

In conclusion I have demonstrated how, through the use of data science, businesses can leverage data when making important decisions such as where to relocate, or expand their business. It is clear that with analysis, this is no longer reliant on gut instinct but a reasoned decision process, producing the desired outcome in a timely manner.