

Office Supplier Expansion

New Location Data Analysis

Coursera Capstone Project

Background and description of the problem

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 clients 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.

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

The data science problem was 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 question one, which city borough would offer the greatest catchment area for the business?

Data Description

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.

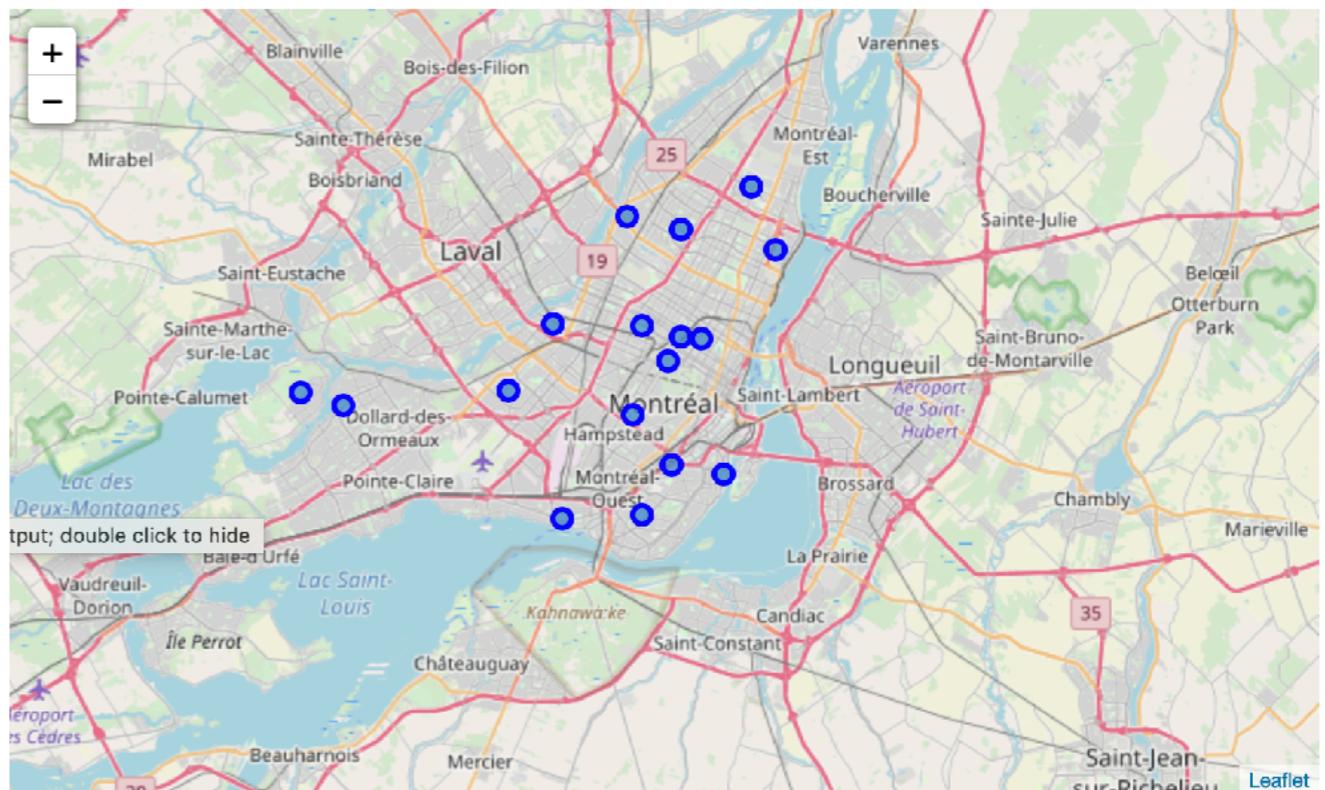
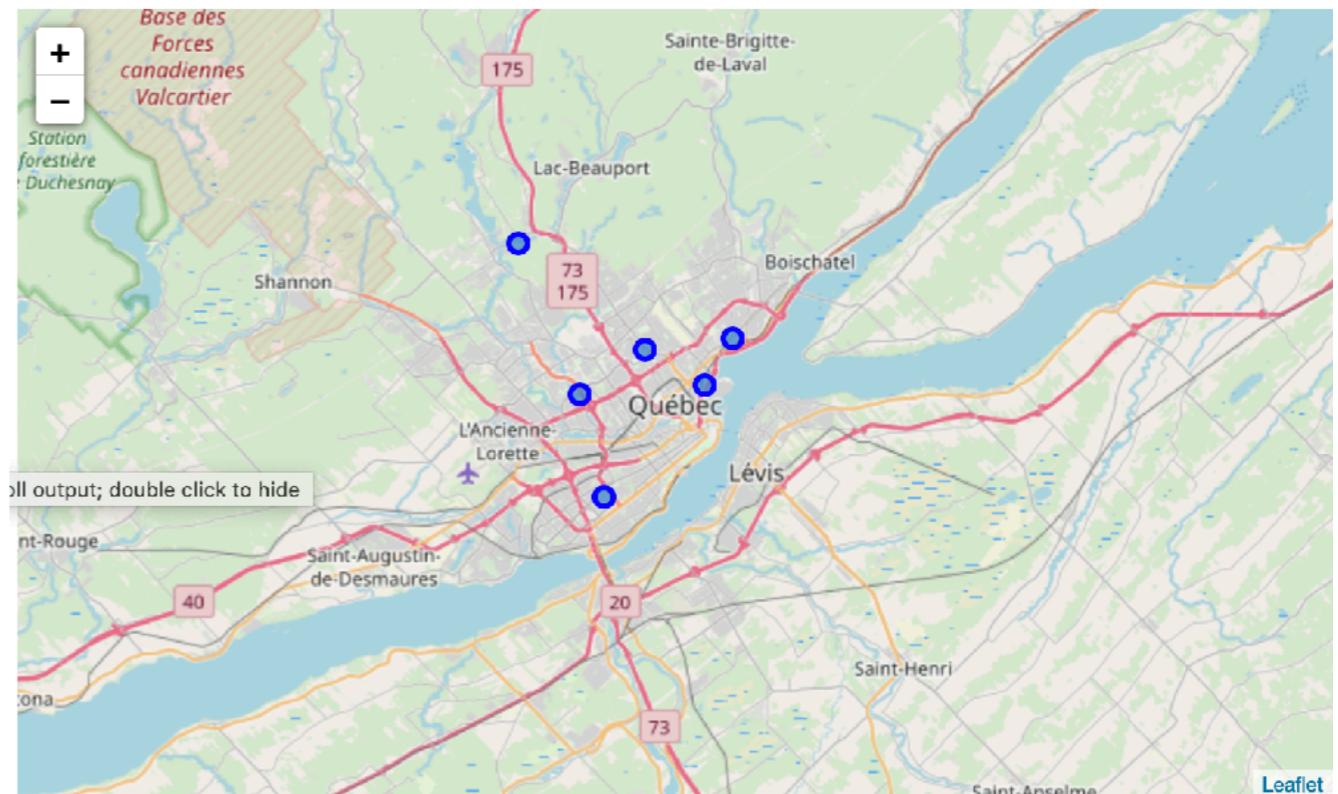
Data extracted included features such as city/ borough location, business type, name along with their longitude and latitude co-ordinates. Borough postal codes were also scraped from wikipedia. Collected data was then prepared and cleaned before analysis using data frames and mapping techniques using 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 clients business expansion.

PART ONE

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).

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 maps with the boroughs superimposed on top.



PART ONE

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
2	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

```
nearby_offices_quebec.shape  
(43, 4)
```

	name	categories	lat	lng
95	MPC	Office	45.498711	-73.555387
96	Morgan Stanley	Office	45.498053	-73.555189
97	Camden	Office	45.502593	-73.555518
98	Saputo Head Office Canada	Office	45.522396	-73.654251
99	WB Games Montréal	Office	45.516739	-73.559559

```
: nearby_offices_montreal.shape  
(100, 4)
```

PART ONE

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.

Results of Part One

```
: print('{} potential customers were returned by Foursquare in Montreal.'.format(nearby_offices_montreal.shape[0]))
print('{} potential customers were returned by Foursquare in Quebec.'.format(nearby_offices_quebec.shape[0]))
```

100 potential customers were returned by Foursquare in Montreal.
43 potential customers were returned by Foursquare in Quebec.

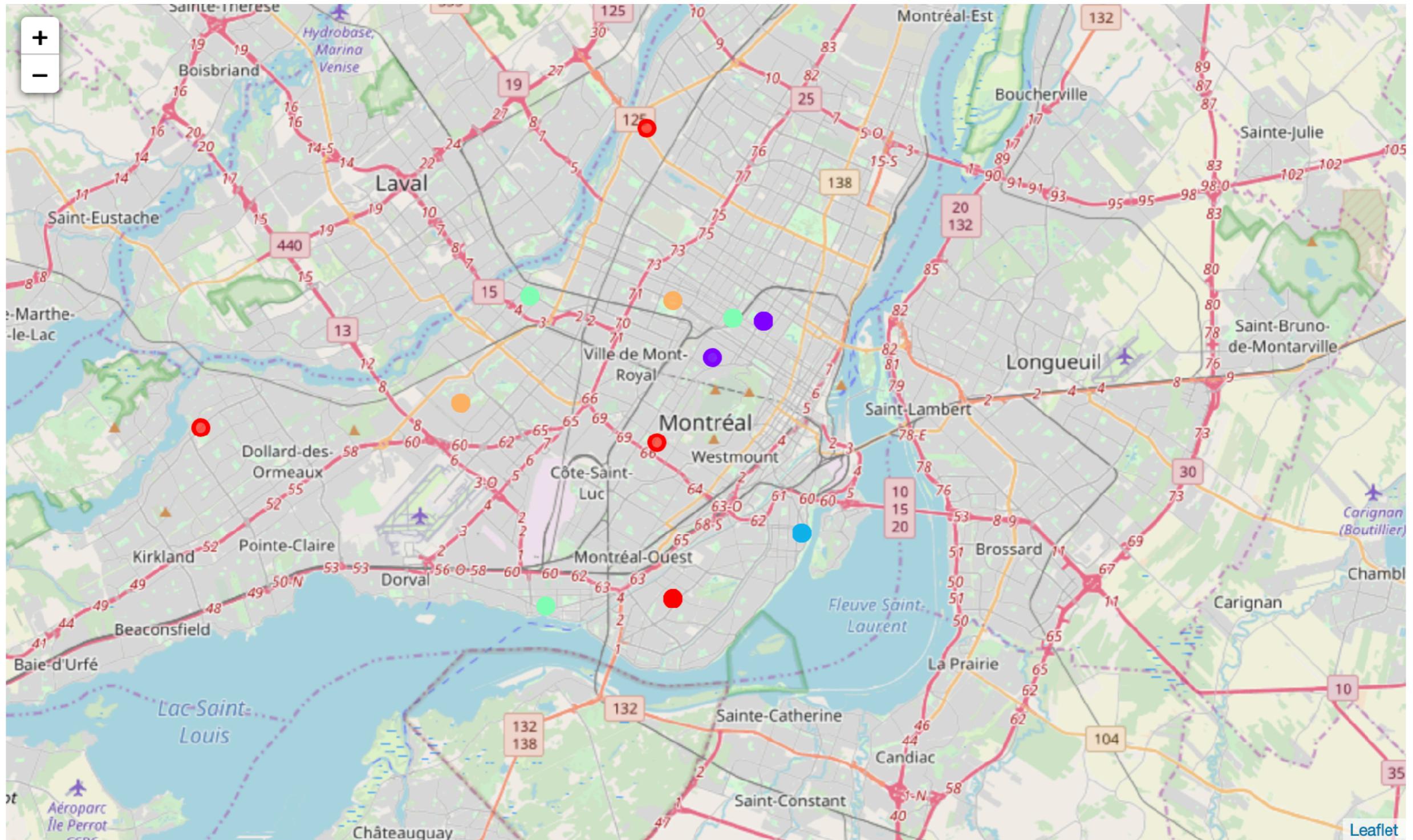
PART TWO

Further analysis of the boroughs of Montreal revealed a cluster containing more potential customers than the others. ‘Cluster 4’ contained the greatest concentration of potential customers for the client. 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’ 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)

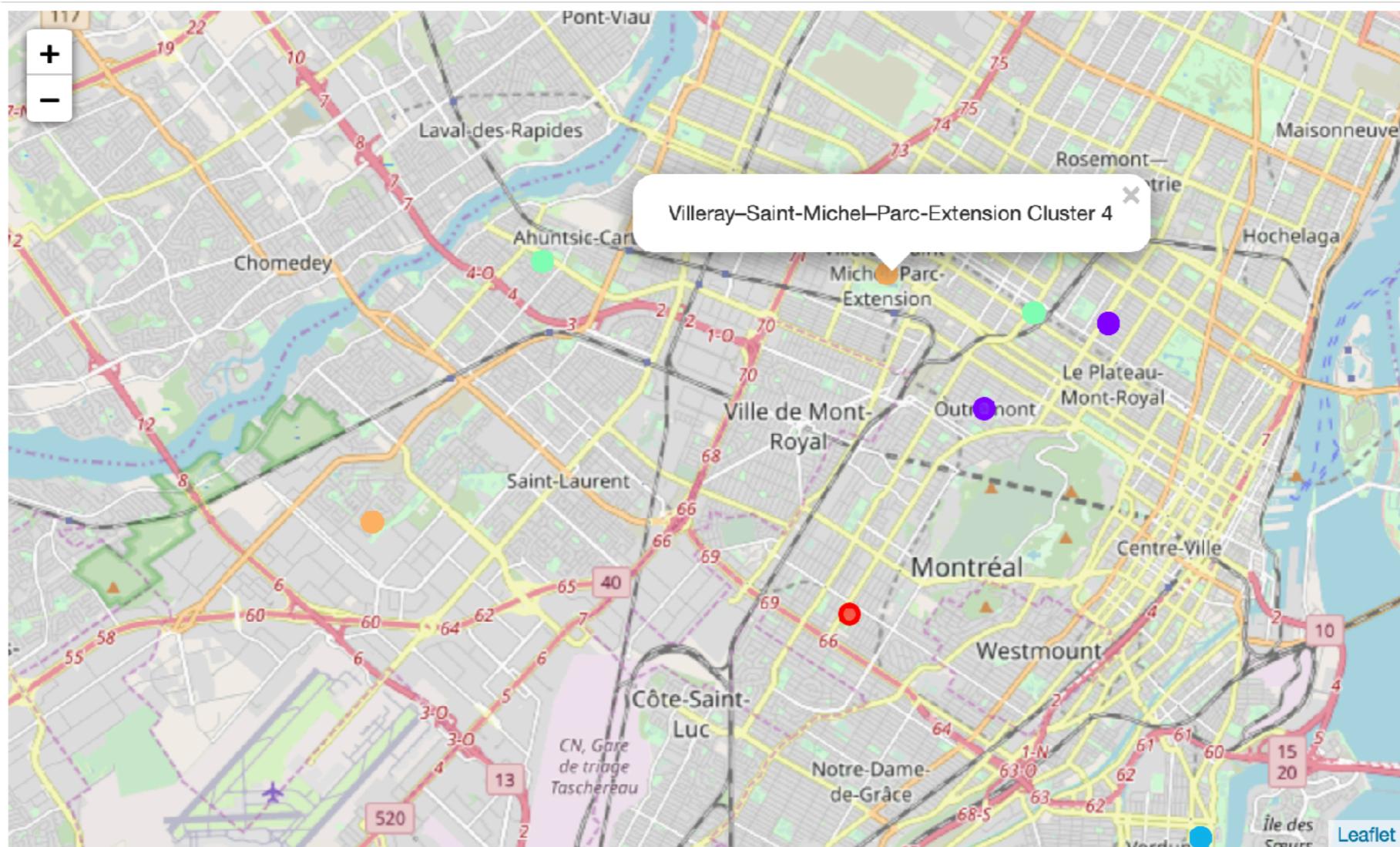
From this data, the city borough which would offer the greatest catchment area for the clients business is **Villeray–Saint-Michel–Parc-Extension**. This borough was chosen from the two possibilities because of its close proximity to three other cluster centres (see map above). The borough of Saint-Laurent,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

PART TWO



PART TWO

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.



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**.

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.