Coursera Capstone -

Opening a New Restaurant in Montreal, Canada

IBM Applied Data Science Capstone

Introduction – Business Problem

1. Introduction

Montreal is the most populous, diverse city in the Canadian province of Quebec and the second-most populous city in Canada. It is an old city and has a lot of attraction place for tourist. Each year millions of tourists visits Montreal so, it has multibillion-dollar restaurant business.

There are many restaurants in the city of Montreal and many more are being built. Of course, the market is highly competitive and as a well-developed city, the cost of doing business is also one of the highest in the country.

The choosing location of the restaurant is one of the most important decisions that will determine whether the mall will be a success or a failure. In this project we will find the best location to start a restaurant using data science techniques.

1.1 Business Problem

The objective of this capstone project is to analyze and select the best locations in the city of Montreal, Canada to open a new restaurant. As a startup, they need to choose their first starting location carefully for the points highlighted above and more importantly, if this is successful, they will be able to replicate the same success fairly quickly; so, first mover advantage is critical for this business and thereby the choice of location (i.e. neighborhood) is important to them.

1.2 Target Audience

This project is particularly useful to property developers and investors looking to open or invest in new restaurant in the city of Montreal. This project is timely as the city is currently suffering from oversupply of shopping malls. However, the number of visitors is increasing each year and city has initiated to many measures to attract more visitor. Therefore, city needs new restaurant and hence opening a new restaurant would be a very profitable business.

2. Data Description

To solve the problem, we will need the following data:

• As we need to explore, segment, and cluster the boroughs in the city of Montreal, the Montreal suburbs or borough data is key for this project.

• Latitude and longitude coordinates of those suburbs. This is required in order to plot the map and to get the venue data.

• Venue data, particularly data related to shopping malls. We will use this data to perform clustering on the neighborhoods.

	Borough	Latitude	Longitude
0	Ahuntsic-Cartierville	45.540030	-73.681850
1	Anjou	45.602180	-73.559600
2	Côte-des-Neiges–Notre-Dame-de-Grâce	45.490730	-73.629480
3	Lachine	45.432970	-73.681120
4	LaSalle	45.429650	-73.662830
5	Le Plateau-Mont-Royal	45.529190	-73.578620
6	Le Sud-Ouest	45.474260	-73.581580
7	L'Île-Bizard–Sainte-Geneviève	45.497340	-73.891770
8	Mercier–Hochelaga-Maisonneuve	45.582110	-73.530660
9	Montréal-Nord	45.594900	-73.639480
10	Outremont	45.514130	-73.610970
11	Pierrefonds-Roxboro	45.496420	-73.847030
12	Rivière-des-Prairies–Pointe-aux-Trembles	45.651756	-73.562551
13	Rosemont–La Petite-Patrie	45.552820	-73.579270
14	Saint-Laurent	45.507350	-73.683560
15	Saint-Léonard	45.586240	-73.596010
16	Verdun	45.459270	-73.571740
17	Ville-Marie	45.529740	-73.553610
18	Villeray–Saint-Michel–Parc-Extension	45.560280	-73.604660

There are total 19 boroughs (suburbs) in Montreal has been found form this Wikipedia page.

2.1 Data Features

We will be leveraging on features in a reliable location information provider such as the Foursquare.com to explore the various types of venues and its categories available in each neighborhood. We will also need to understand the trending of these venues in the respective neighborhood. The information obtained per neighborhood will be as such like below and must be in a structured format.

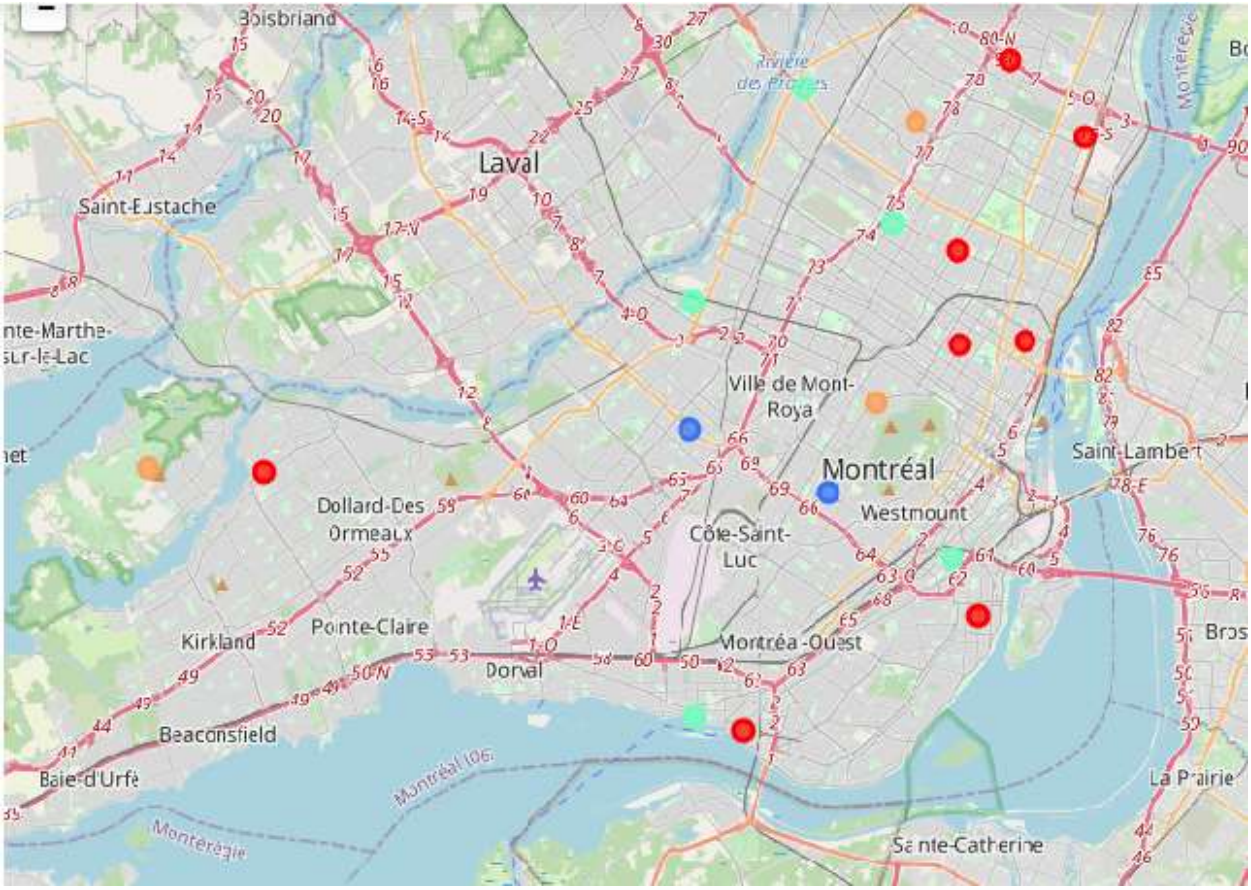
	Borough	Latitude	Longitude	VenueName	VenueLatitude	VenueLongitude	VenueCategory
0	Ahuntsic-Cartierville	45.54003	-73.68185	Sapori Di Napoli	45.540799	-73.685707	Italian Restaurant
1	Ahuntsic-Cartierville	45.54003	-73.68185	Parc Marcelin-Wilson	45.540585	-73.685730	Park
2	Ahuntsic-Cartierville	45.54003	-73.68185	Subway	45.537975	-73.679188	Sandwich Place
3	Ahuntsic-Cartierville	45.54003	-73.68185	Aréna Marcellin-Wilson	45.540662	-73.685701	Hockey Arena
4	Ahuntsic-Cartierville	45.54003	-73.68185	Restaurant Dima	45.540614	-73.685741	Middle Eastern Restaurant

3. Technology and Methodology

We will use web scraping techniques and BeautifulSoup to extract the data from the Wikipedia page. Geocoder API will be used to get the latitude and longitude geographical coordinates of the suburbs. This is important so that we can input this information into the location information provider such as Foursquare.com to obtain venue information in these neighborhoods, and this is precisely what we have done for it in this project.

We will use Foursquare API to get the venue data for those suburbs. Panda library for data cleaning, data wrangling, whereas Scikit-learn library for machine learning (K-means clustering) and Folium for map visualization will be used.

Finally, with all these technology and methodologies, we will then be able to come up with a best recommendation to the owner of a new restaurant to their problem which is where is the best regions of suburb for them to first start off to offer their services based on suburb similarities, high intensity of restaurant and low competition (i.e. fewer restaurant). In other words, we will not want to recommend to the new owner to enter a suburb whereby there is already a high concentration of restaurant available and lower demand in the suburb.



4. Results

The results from the k-means clustering show that we can categorize the Suburbs into 4 clusters based on the frequency of occurrence for “Restaurant”:

- Cluster 0: Suburbs with moderate number of restaurants
- Cluster 1: Suburbs with high concentration of restaurants
- Cluster 2: Suburbs with between high and moderate concentration of restaurants
- Cluster 3: Suburbs with low number to no existence of restaurants

The results of the clustering are visualized in the map below with cluster 0, cluster 1, cluster 2, cluster 3 in red, blue, aqua and orange color, respectively.

5. Discussion

As observations noted from the map in the Results section, most of the restaurants are concentrated in the central area of Montreal, with the highest number in cluster 1 and moderate number in cluster 0 and between high and moderate number in the cluster 2. On the other hand, cluster 3 has no restaurants in the suburbs. This represents that cluster 3 has a great opportunity and high potential to open new restaurant as there are no competition from existing restaurant. Meanwhile, restaurants in cluster 1 are likely suffering from intense competition due to oversupply and high concentration of restaurants. From another perspective, the results also show that the oversupply of restaurants mostly happened in the central area of the city, with the suburb area still have very few restaurants. Therefore, this project recommends new restaurants owner to capitalize on these findings to open a new restaurant in suburbs in cluster 3 with no competition. New owner with unique selling propositions to stand out from the competition can also open new restaurant in suburbs in cluster 0 with moderate competition. Lastly, new owners are advised to avoid suburbs in cluster 1 and 2 which already have high concentration of restaurant and suffering from intense competition.

6. Limitations

In this project, we only consider one factor i.e. frequency of occurrence of restaurant, there are other factors such as population, income of residents, tourist sight, hotel that could influence the location decision of a new restaurant. However, due lack of data availability, we were unable to use those parameters. As we are using free version of Foursquare API that came with limitations as to the number of API calls and results returned. That is why cluster 3 has no restaurant which are not realistic. Future research could make use of paid account to bypass these limitations and obtain more results.

7. Conclusion

A finding process to identify the best location in the suburbs of the Montreal City to open a new restaurant by using data science technology and methodology has been presented. In this

project, we have gone through the process of identifying the business problem, specifying the data required, extracting and preparing the data, performing machine learning by clustering the data into 4 clusters based on their similarities, and lastly providing recommendations to the new restaurant owner to open a new restaurant. Less intense areas are given preference to select area to choose. According to that we have found that cluster 3 are the most preferred locations to open a new restaurant. The findings of this project will help the relevant stakeholders to capitalize on the opportunities on high potential locations while avoiding overcrowded areas in their decisions to open a new restaurant.