IBM Data Science Capstone Project

Best Culinary Experience In Montreal

Marc Enriel Quilatan

May 30 2020

1. **Introduction**

Food is an important aspect in our human lives. Every country has different kinds of cuisine. The culinary world offers a variety of cuisines and can leave you speechless after one bite. In our modernise world where restaurants are dispersed all around a city. Where should we stay to have the best culinary experience.

Montreal is one of the most ethnically and culturally diverse cities in the world. With this diversity, also brings an abundance of different cuisine around the city. This project will encompass the different neighborhoods of Montreal to find out where is the best neighborhood to stay for the best foodie experience.

By geolocating where would be the best place to stay, this would help **tourist**, **Montreal natives** like myself, **companies** ( like Airbnb, hotels…), **social media influencers**, etc. to find out where they should strategically place themselves to introduce themselves to the different food amenities that Montreal has to offer.

1. **Data**

To complete the task ahead, multiple data sources will been needed:

1. The use of the postal codes to find the different neighborhoods of Montreal. Found here: <https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_H>
2. The used of the combination of the **geopy.geocders** and **Nominatim** API to find the longitude and latitude of the different neighborhoods
3. The **Foursquare API** to find the different venues of each Neighborhood
4. **Methodology**
   1. Gathering the Data

Finding the data for the postal codes was as simple as googling for it. It lead me to the Wikipedia site.



Fig. 1 Postal Codes of Montreal

* 1. Putting the Data into a DataFrame

This proves to be one of the simplest and more difficult part of the project. Since the table Wikipedia has is not of a regular format. Beautiful Soup Package was used to wrangle the data. Unfortunately, I wasn’t able to wrangle the data cleanly and I had to clean the dataframe multiple times manually. This is due to the fact a lot of the postal codes wrangle were not used as shown in Figure 1 and also the fact that the data I wrangled contained sometimes multiple links ( which beautiful soup takes in )

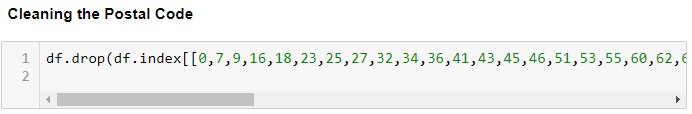


Fig 2. Cleaning Postal Code Data

In the end, I ended up with a useable dataframe as shown below:

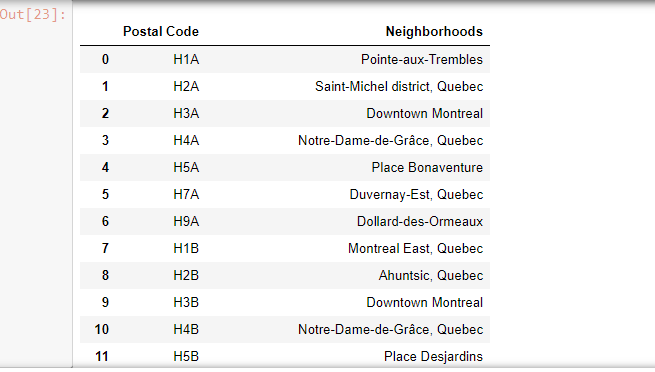


Fig. 3 Neighborhood of Montreal Dataframe

* 1. Finding the Longitude and Latitude of Every Neighborhood

In this step we imported Nominatim with geopy to get the latitude and longitude of each postal code. Nominatim is an API the is a search engine for OpenStreetMap. This site was used to find out why some of the postal codes were not outputting any coordinates.

<https://nominatim.openstreetmap.org/>

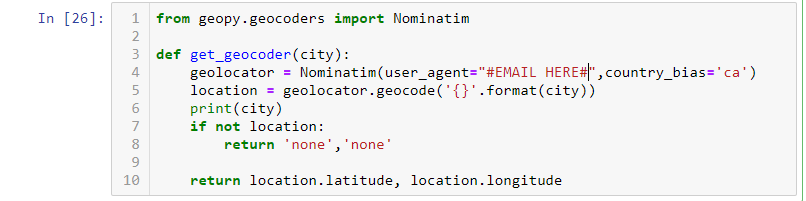


Fig. 4 Function to get Longitude and Latitude of the Postal Codes

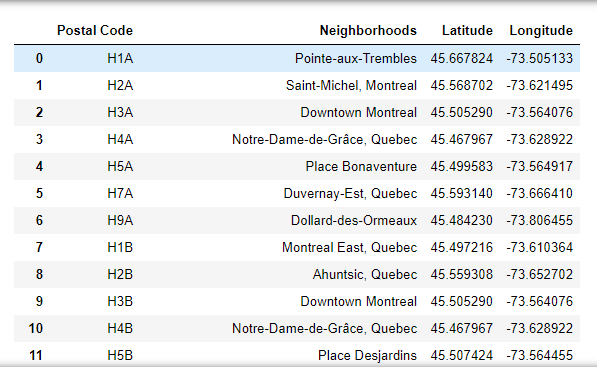


Fig. 5 Data Frame with Longitude and Latitude

Folium was used to generate a map with all the neighborhoods on top. Take note that I considered Laval as part of Montreal.

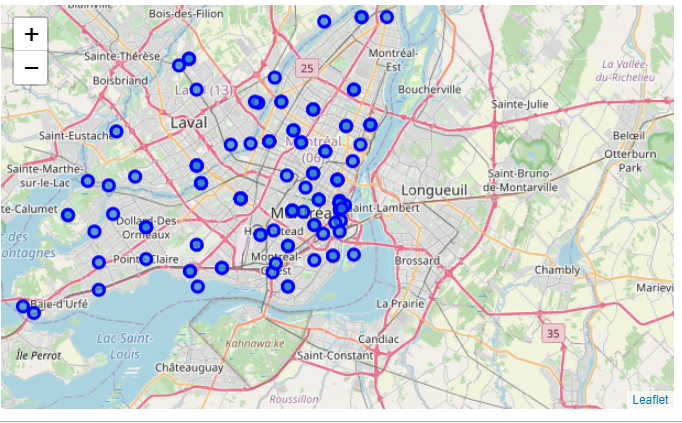


Fig. 6 Map of Montreal with Every Neighborhood

* 1. Obtain the Different Venues in Each Neighborhood with Foursquare

To obtain the different Venues in each neighborhood, the Foursquare API was used with the coordinates we have gotten with Geolocator



Fig. 7 Function used to Find Different Venues around the Neighborhood

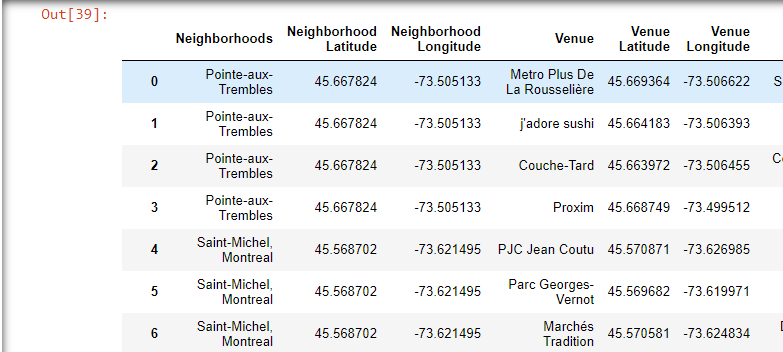


Fig. 8 Dataframe Containing the Different Venues

* 1. **Clustering the Neighborhoods**

In this step, we took every single neighborhood and find out which ones were most prominent to neighborhoods. One-Hot Encoding was used to find out which neighborhoods were prominent for what type of venue.

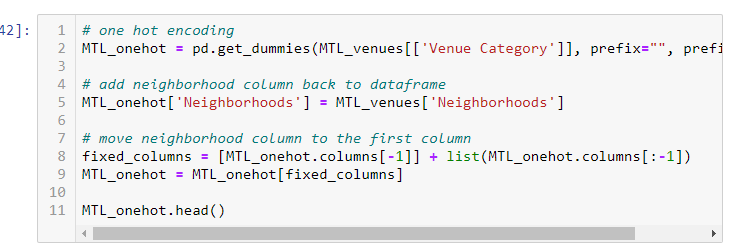


Fig 9. One hot Encoding Function

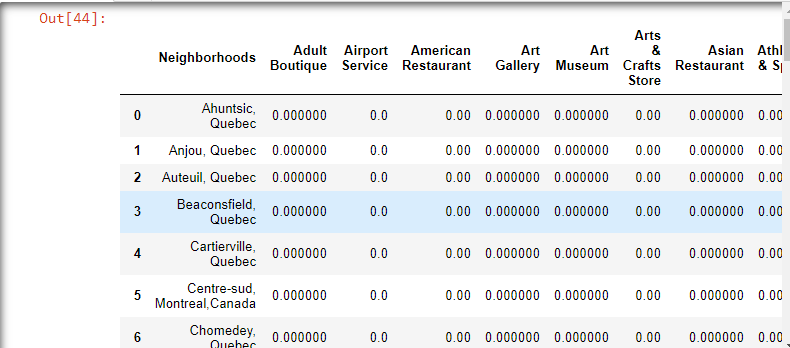


Fig. 10 Neighborhoods and Venues One Hot Encoded and Indexed

With this, we are able to find out the top 5 Venues every neighborhood had. K-Means clustering is then use to cluster different Neighborhoods together. K-Means Clustering is the best method to use in this case because, we want the data to be clustered in terms of what type of venues they offer.

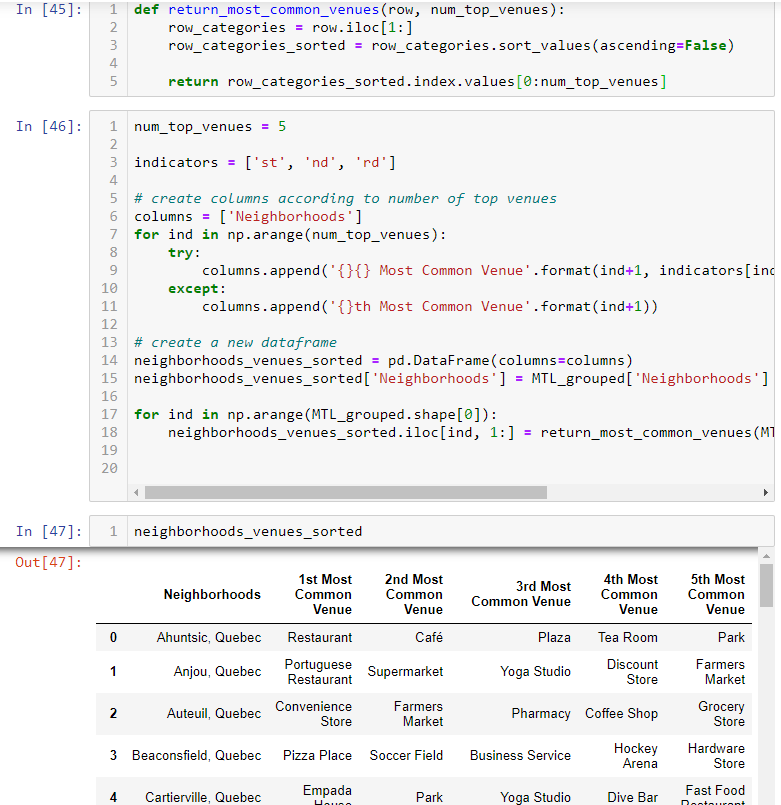
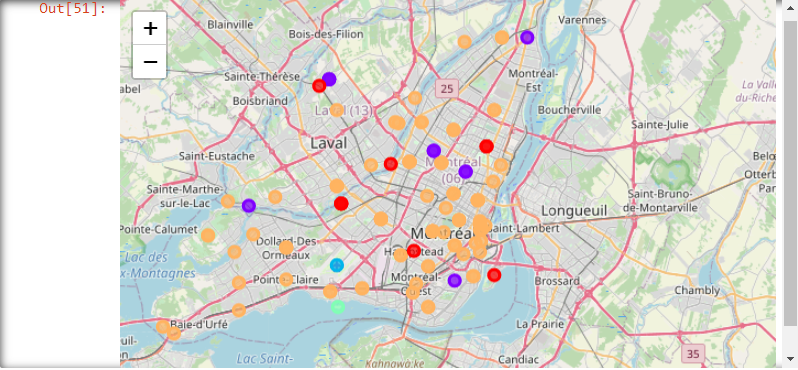


Fig 11. Top 5 Venues of each Neighborhood

Once K-means is performed, each neighborhood will be clustered into 5 separate groups



*Fig 12. Cluster Labelled Neighborhoods*



*Fig 13. Map of the 5 Clustered Neighborhoods*

Looking at the result of the 5 clusters, one of the clusters were most prominent with restaurants and that was the cluster that was more looked into. The clusters were separated in terms of Parks, Convenience/Grocery Stores, Airport, Boat and Ferry and finally the one we looked at Café’s and Restaurants. The Neighborhoods in that cluster were then investigated.

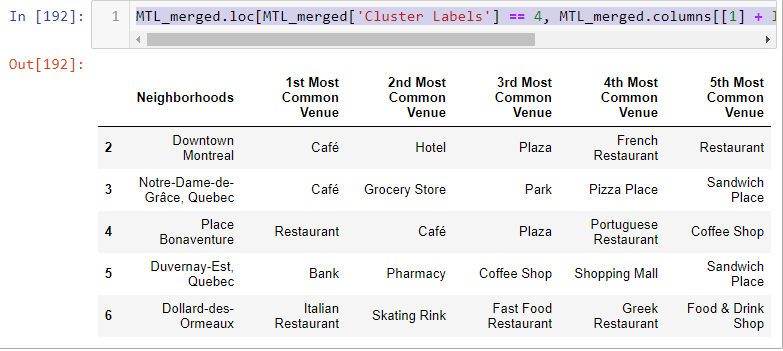


Fig 14. Restaurant Cluster

* 1. Analysis of the 4th Label

Looking into the label, we found which neighborhoods were prominent with Restaurants but how can we deduce which are the best neighborhoods? To do this, **we looked at the quantity of restaurants per Neighborhood in that specific cluster.** By doing this, the final table we get shows:

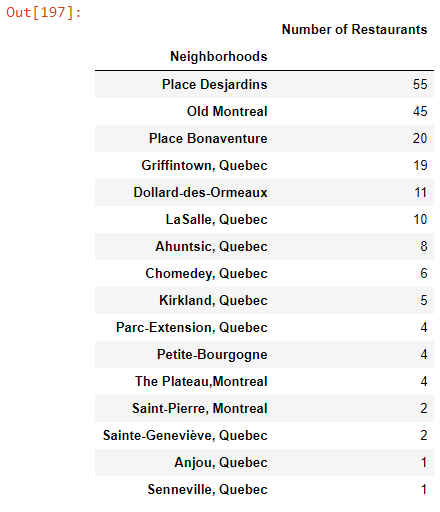


Fig. 15 Number of Restaurant per Neighborhood

1. **Results & Discussion**

Observing Fig. 15, the we can see that the best neighborhood overall that would be have the best culinary experience based on the quantity of restaurants in that neighborhood is the **Place Desjardins** neighborhood that contains 55 restaurants in total that is not including fast food restaurants. A close second with be Old Montreal that contains a high 45 Restaurants also. To explore more widely, we checked where the Neighborhoods are these neighborhoods situated geographically

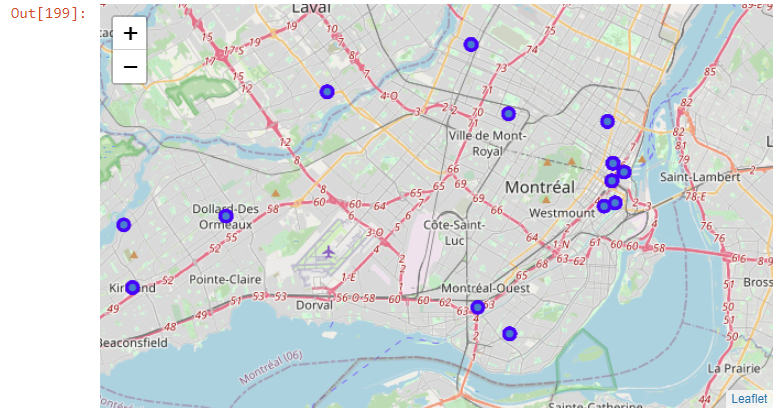


Fig. 16 Map of Neighborhoods that are Predominantly Restaurants

Looking at the map, we can see that the top 3 places, **Place Bonaventure, Place Desjardins and Old Montreal.** Are clustered together. This will be the holy trinity of places stakeholders should stay at to have the best culinary experience in Montreal

However, as much as this are the results I’ve gotten in this research. I feel like many venues are missing. I was personally surprised that plateau only has 4 restaurants. As a Montreal Native, I know that the Plateau has been getting a lot of attention in the food scene as it is a hipster neighborhood. I think what caused this can be that the Foursquare API doesn’t contain all the recent data of restaurants that has opened. As those 3 Neighborhoods are in fact great for food, I believe that there are many venues missing.

1. **Conclusion**

In conclusion, the best neighborhoods to stay in for the best culinary experience is Old Montreal, Place Desjardins and Place Bonaventure. However, take this with a grain of salt as per the foursquare API is missing many venues that should be considered while doing this report.

To improve upon this report as I grow in my data scientist skillset, I would’ve maybe used another API that would encompass more of the data of restaurants in Montreal. In addition, instead of using a quantity per neighborhood to judge where is the best culinary experience, I would try and include different weighting factors as in distance from the neighborhood and also the different ratings of the restaurants.