The Battle of Neighborhoods | Business Proposal | Introduction

1-Introduction:

I North Africa There is a big confusion in the migration/tourist communities community for more than 10 years (2000-2010) big focus on the tourist to France but from 2016 there are new calls to migrate to Canada but the community they take serious and even today in covid-19 time they go in illegal migration to France. So we ant to help them see to the evolution of Canada as we are taking from the example of Tronto database and we will compare to Paris.

Toronto and Paris are quite popular tourist and vacation destinations for people all around the world. They are diverse and multicultural and offer a wide variety of experiences that is widely sought after. We try to group the neighborhood s of Toronto and Paris respectively and draw insights to what they look like now.

2-Business Problem:

The aim is to help North african and other tourists to choose their destinations depending on the experiences that the neighborhood s have to offer and what they would want to have. This also helps people make decisions if they are thinking about migrating to Toronto or Paris or even if they want to relocate neighborhood s within the city. Our findings will help stakeholders make informed decisions and address any concerns they have including the different kinds of cuisines, provision stores and what the city has to offer.

3-Data Description:

We require geographical location data for both Toronto and Paris. Postal codes in each city serve as a starting point. Using Postal codes we can find out the neighborhoods, boroughs, venues and their most popular venue categories.

- Toronto:

For the Toronto neighborhood data, a Wikipedia page exists that has all the information we need to explore and cluster the neighborhoods in Toronto.

Data Link: <https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M>

from week 3. Dataset consisting of latitude and longitude, zip codes.

- Paris:

To derive our solution, We leverage JSON data available at <https://www.data.gouv.fr/fr/datasets/r/e88c6fda-1d09-42a0-a069-606d3259114e>

The JSON file has data about all the neighborhood s in France, we limit it to Paris.

1. postal\_code : Postal codes for France
2. nom\_comm : Name of Neighborhood s in France
3. nom\_dept : Name of the boroughs, equivalent to towns in France
4. geo\_point\_2d : Tuple containing the latitude and longitude of the Neighborhood s.

-Foursquare API Data:

We will need data about different venues in different neighborhood s of that specific borough. In order to gain that information we will use "Foursquare" locational information. Foursquare is a location data provider with information about all manner of venues and events within an area of interest. Such information includes venue names, locations, menus and even photos. As such, the foursquare location platform will be used as the sole data source since all the stated required information can be obtained through the API.

After finding the list of neighborhood s, we then connect to the Foursquare API to gather information about venues inside each and every neighborhood . For each neighborhood , we have chosen the radius to be 500 meters.

The data retrieved from Foursquare contained information of venues within a specified distance of the longitude and latitude of the postcodes. The information obtained per venue as follows:

1. Neighborhood : Name of the Neighborhood
2. Neighborhood Latitude : Latitude of the Neighborhood
3. Neighborhood Longitude : Longitude of the Neighborhood
4. Venue : Name of the Venue
5. Venue Latitude : Latitude of Venue
6. Venue Longitude : Longitude of Venue
7. Venue Category : Category of Venue

Based on all the information collected for both Tornto and Paris, we have sufficient data to build our model. We cluster the neighborhood s together based on similar venue categories. We then present our observations and findings. Using this data, our stakeholders can take the necessary decision.

4-Methodology:

## For the methodology since the Torono already exist in the week 3. We will focus on the Paris dataset.

**4.1 Data Collection**

To collect data for Paris, we download the JSON file containg all the postal codes of France from <https://www.data.gouv.fr/fr/datasets/r/e88c6fda-1d09-42a0-a069-606d3259114e>

Using Pandas we load the table after reading the JSON file:

In [37]:

!wget -q -O 'france-data.json' https://www.data.gouv.fr/fr/datasets/r/e88c6fda-1d09-42a0-a069-606d3259114e

print('Data Downloaded!!')

paris\_raw = pd.read\_json('france-data.json')

paris\_raw.head()

It look like the figure above

## 

**4.2 Data Preprocessing**

For Paris, we break down each of the nested fields and create the dataframe that we need:

paris\_field\_data = pd.DataFrame()

**for** f **in** paris\_raw.fields:

dict\_new = f

paris\_field\_data = paris\_field\_data.append(dict\_new, ignore\_index=**True**)

paris\_field\_data.head()

**4.3 Feature Selection**

For both of our datasets, we need only the borough, neighborhood , postal codes and geolocations (latitude and longitude). So we end up selecting the columns that we need by:

df\_paris = paris\_data[['postal\_code', 'nom\_comm', 'nom\_dept', 'geo\_point\_2d']]

**4.4 Feature Engineering**

Both of our Datasets actually contain information related to all the cities in the country. We can narrow down and further process the data by selecting only the neighborhood s pertaining to 'Paris'

df\_paris = df\_paris[df\_paris['nom\_dept'].str.contains('PARIS')].reset\_index(drop=True)

df\_paris.head()

for our Paris dataset, we already have it stored in the geo\_point\_2d column as a tuple in the df\_paris dataframe.

We just need to extract the latitude and longitude for the column:

paris\_lat = paris\_latlng.apply(**lambda** x: x.split(',')[0])

paris\_lat = paris\_lat.apply(**lambda** x: x.lstrip('['))

paris\_lng = paris\_latlng.apply(**lambda** x: x.split(',')[1])

paris\_lng = paris\_lng.apply(**lambda** x: x.rstrip(']'))

paris\_geo\_lat = pd.DataFrame(paris\_lat.astype(float))

paris\_geo\_lat.columns=['Latitude']

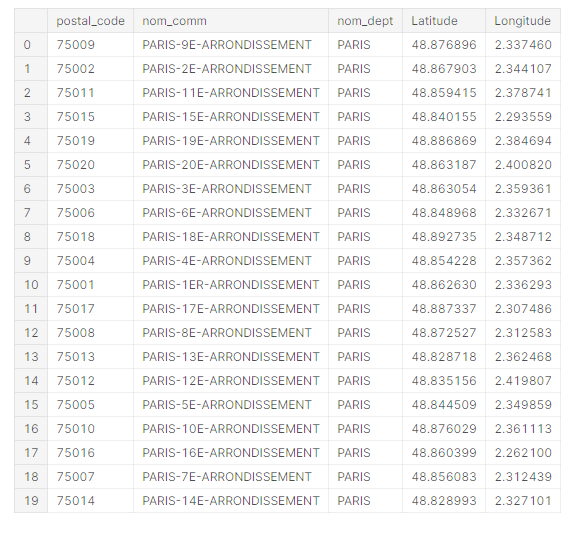
paris\_geo\_lng = pd.DataFrame(paris\_lng.astype(float))

paris\_geo\_lng.columns=['Longitude']

We then create our Paris dataset with the required information:

paris\_combined\_data = pd.concat([df\_paris.drop('geo\_point\_2d', axis=1), paris\_geo\_lat, paris\_geo\_lng], axis=1)

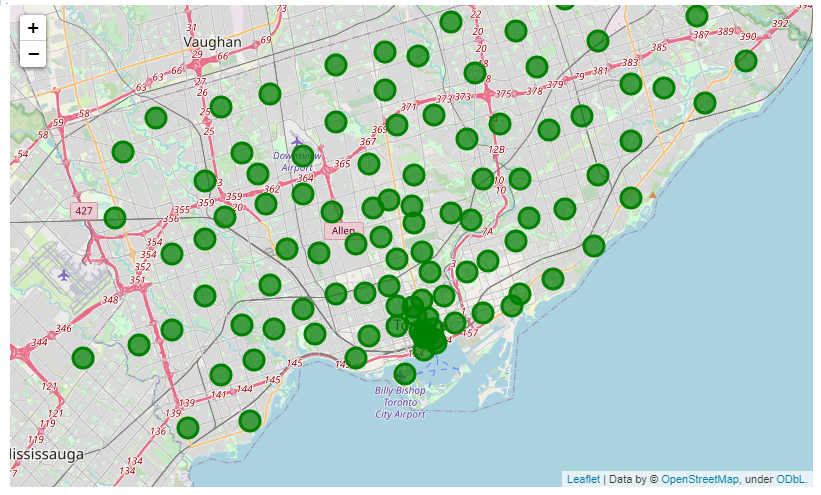
paris\_combined\_data



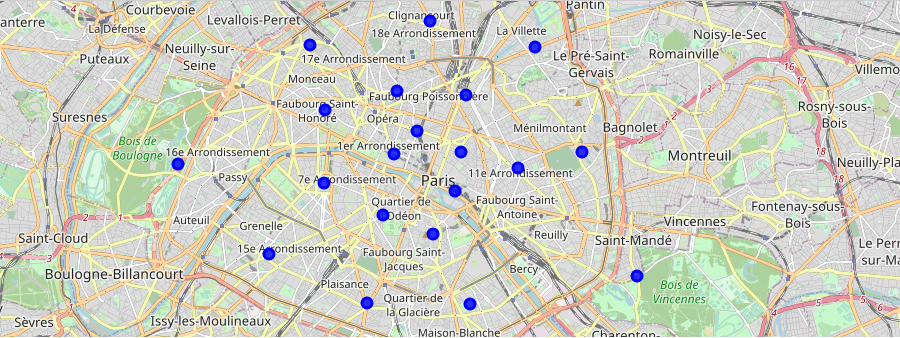
**4.5 Visualizing the Neighborhood s of Toronto and Paris**

Now that our datasets are ready, using the Folium package, we can visualize the maps of Toronto and Paris with the neighborhood s that we collected.

Neighborhood map of Toronto:



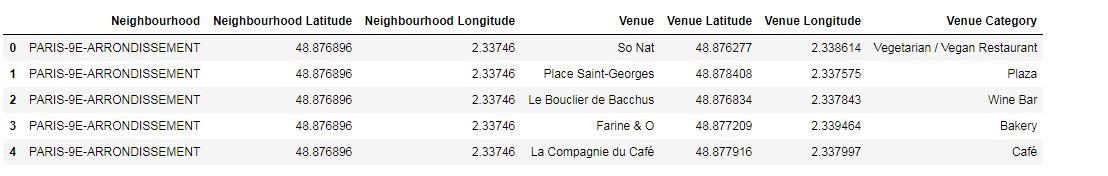
Neighborhood map of Paris:



Now that we have visualized the neighborhood s, we need to find out what each neighborhood is like and what are the common venue and venue categories within a 500m radius.

This is where Foursquare comes into play. With the help of Foursquare we define a function which collects information pertaining to each neighborhood including that of the name of the neighborhood , geo-coordinates, venue and venue categories.

Resulting data looks like:

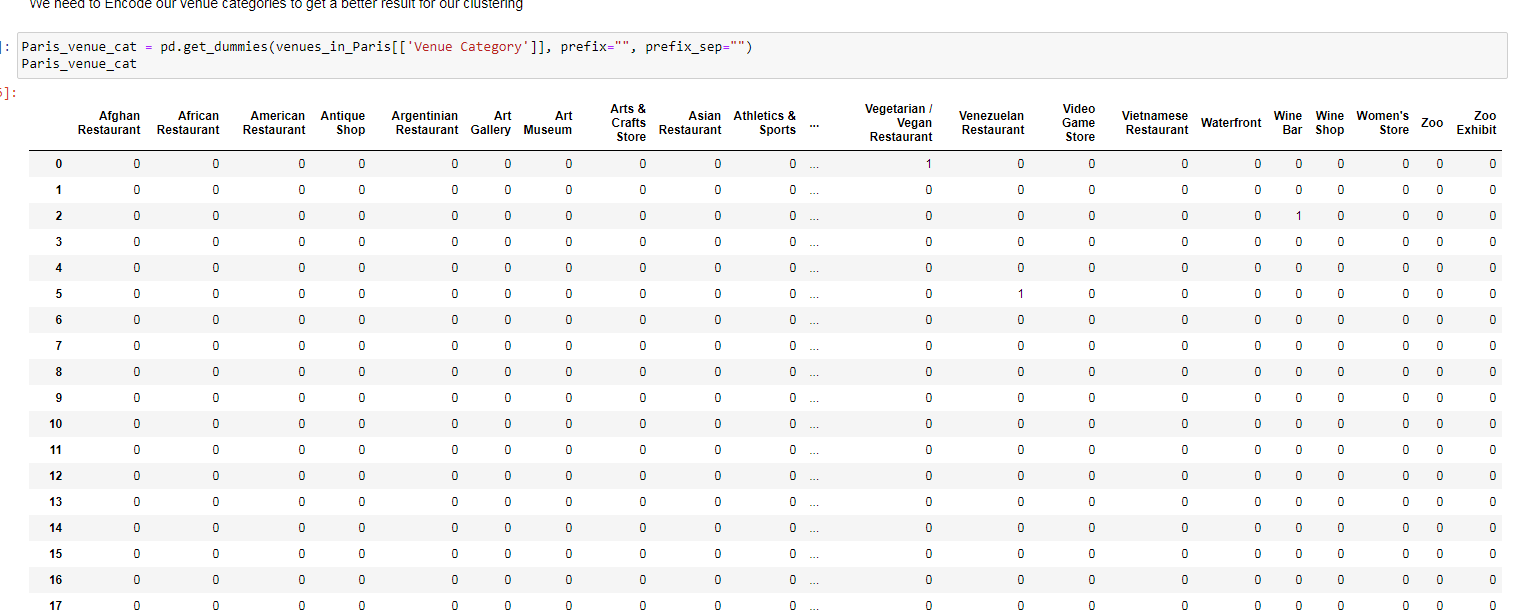


**4.6 One Hot Encoding**

Since we are trying to find out what are the different kinds of venue categories present in each neighborhood and then calculate the top 10 common venues to base our similarity on, we use the One Hot Encoding to work with our categorical datatype of the venue categories. This helps to convert the categorical data into numeric data.

We won't be using label encoding in this situation since label encoding might cause our machine learning model to have a bias or a sort of ranking which we are trying to avoid by using One Hot Encoding.

We perform one hot encoding and then calculate the mean of the grouped venue categories for each of the neighborhood s.

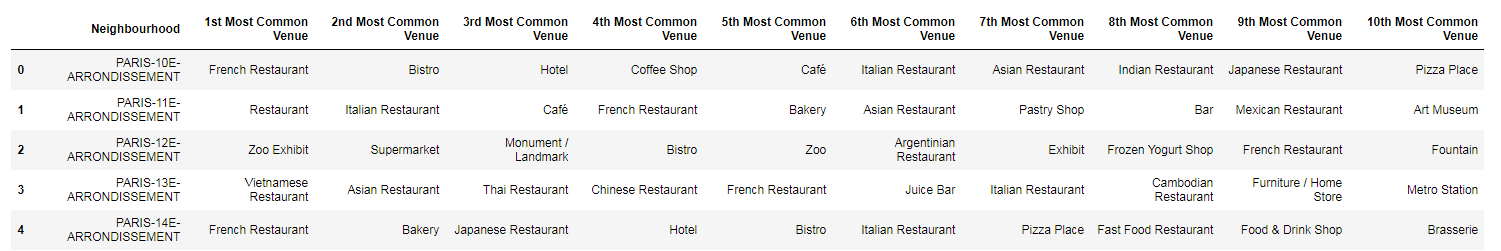
Resulting data looks like:

**4.7 Top Venues in the Neighborhood s**

In our next step, We need to rank and label the top venue categories in our neighborhood.

Let's define a function to get the top venue categories in the neighborhood

Result



**4.8 Model Building - KMeans**

Moving on to the most exicitng part - **Model Building!** We will be using KMeans Clustering Machine learning algorithm to cluster similar neighborhood s together. We will be going with the number of clusters as 5.

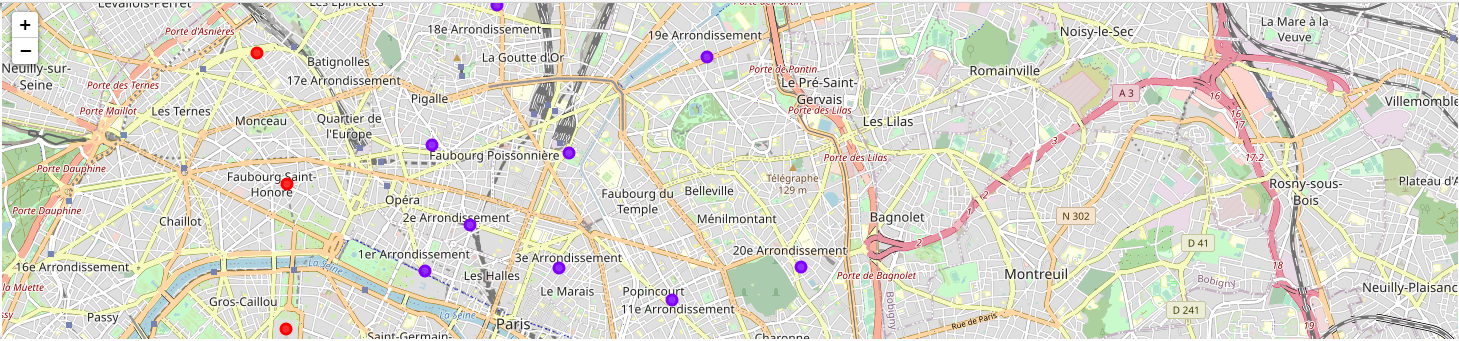
Result



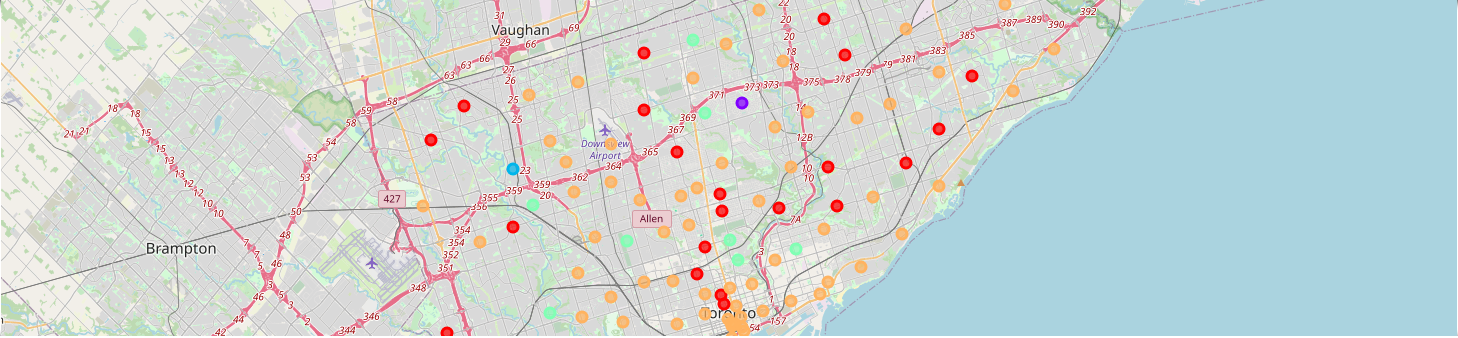
## 4.9 Visualizing the clustered Neighborhood s

Our data is processed, missing data is collected and compiled. The Model is built. All that's remaining is to see the clustered neighborhood s on the map. Again, we use Folium package to do so.

Neighborhood of Paris

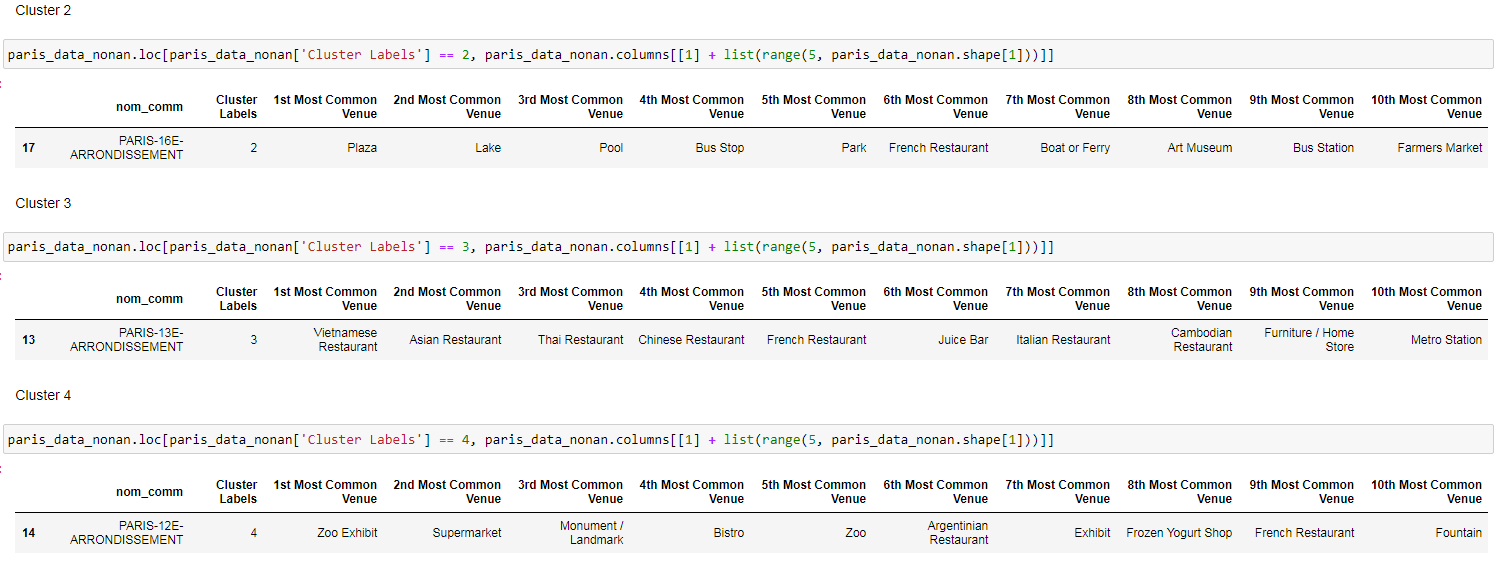


Neighborhood of Toronto



### **- Examining our Clusters**

We could examine our clusters by expanding on our code using the Cluster Labels column:



## Results and Discussion

Toronto's neighborhoods are very multicultural. There are many types of cousins such as Tunisian, Arabic, Italian, Turkish and Chinese. Toronto seems to take a step further in this direction by offering plenty of restaurants, juice bars, cafes, a Fish and Chips store and breakfast bars. It also has many shopping opportunities with signage. The main modes of transport seem to be Buses and trains. For leisure, there are many parks in the districts.

Overall, the city of Toronto offers a multicultural, diverse, and certainly fun experience.

Paris is relatively small geographically. However, there is a wide variety of cuisines and dishes, including French, Tunisian, Algerian, Moroccan, Cambodian, Asian, Chinese and so on. There are many restaurants, including many restaurants. There are many bistros in Paris. Various means of public transport in Paris, including buses, bicycles, boats or ferries. There are many squares, roads, parks, historical sites, clothing stores, art galleries and museums for leisure and sightseeing. Paris seems to be a relaxing holiday destination with a mix of lakes, historical sites and a wide variety of cousins.

## Conclusion

The aim of this project was to explore the cities of Toronto and Paris and see how attractive it is for tourists and potential travelers. We surveyed the two cities according to their postal codes and then extrapolated common venues locations that were present in each neighborhood to finally sort by similar neighborhoods.

We have seen that each district in both cities has a diverse experience to offer that is unique in its own way. Cultural diversity is quite noticeable which also gives a sense of a sense of belonging.

Paris and Toronto seem to offer leisure and beautiful scenery.

Overall, it is up to stakeholders to decide which experience they prefer and which would be best for them.