# Finding the best borough for expatriation

Table des matières

[Finding the best borough for expatriation 1](#_Toc41558693)

[Introduction 1](#_Toc41558694)

[Data 1](#_Toc41558695)

1. Introduction / Business case

A friend of mine works for an international company that offers him to join one of their north American office. He has to choose between **Toronto**, **Montreal** and **New-York** office and ask me help for the decision. My friend lives in the 11th arrondissement (borough) of **Paris** and really likes his living environment. I will try to give some clues based on location data analysis to find the city that most similar to Paris. We will try to answer to the below question that will help my friend.

1. Which city between Toronto, Monreal or New-York is the most similar to Paris?
2. Which borough (arrondissement) is closet to Paris 11th arrondissement?

## Data

In this work, we will use the following datasets:

* **Paris\_Data** : this dataset provides the list of Paris’s boroughs and neighborhoods

|  |  |  |
| --- | --- | --- |
| **PostalCode** | **Borough** | **Neighborhood** |
| 75001 | Le Louvre | Saint-Germain-l'Auxerrois |
| 75001 | Le Louvre | Les Halles |
| 75001 | Le Louvre | Le Palais-Royal |
| 75001 | Le Louvre | La Place-Vendôme |

* **Montreal\_Data** : This is data exposed in Montreal official website : <http://donnees.ville.montreal.qc.ca/dataset/00bd85eb-23aa-4669-8f1b-ba9a000e3dd8/resource/e9b0f927-8f75-458c-8fda-b5da65cc8b73/download/limadmin.json>

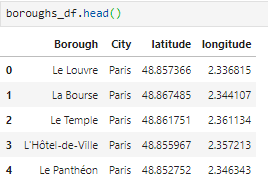
This json format file will be used to create Montreal dataframe in the model of Paris\_Data.

* **Toronto\_Data** : dataset containing Toronto boroughs and neighborhoods that we download from Wikipedia: <https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M>'
* **New\_york\_Data**: this contains New York list of boroughs and neighborhoods: <https://cocl.us/new_york_dataset>
* We will also use Foursquare API data to explore venues of each city and its boroughs

The dataset will be used with kmeans clustering algorithm to get how Paris is similar or dissimilar to each city. Then to identify which borough is closet to Paris 11th arrondissement.

## Methodology

After collecting data from above dataset, I have enhanced them with geographical data of each borough. Then I have created a dataframe containing all 4 cities boroughs:



Thanks to Foursquare API, I retrieve the main venues of each borough.

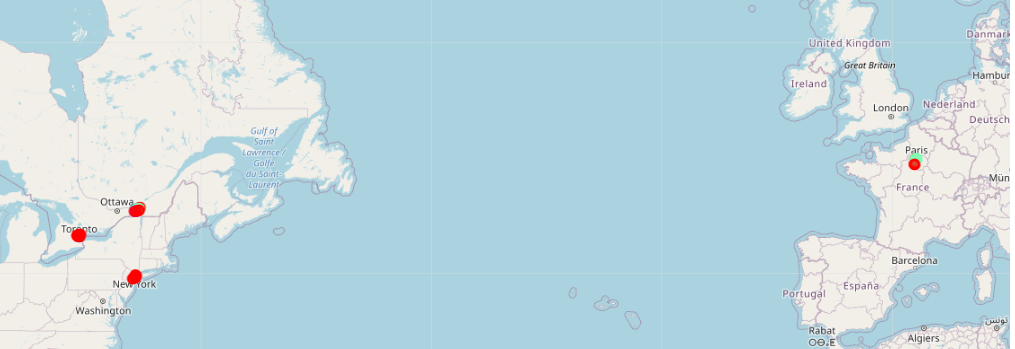


Since our aim is to compare cities and borough through the category of venue they have, I encode the venue categories to be able to apply machine learning technique.

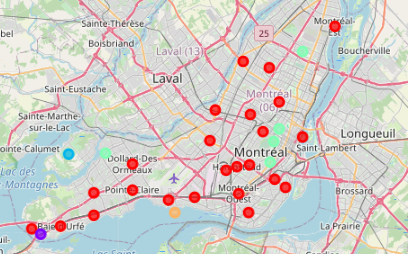
Kmeans clustering algorithm is applied to check how similar or dissimilar borough of Paris, Toronto, New York and Montreal are. I choose to dispatch boroughs into 5 clusters and use sklern.Kmeans method.

## Results

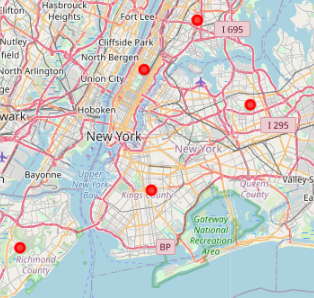
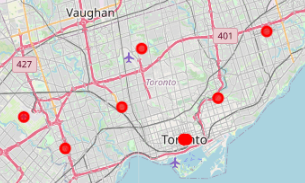
With folium, let’s display the 4 cities without zooming into the clusters.



Paris Monreal

New York Toronto

**Resultat1 :**

As we can see in the above capture, Montreal is the closest to Paris in terms of venues. So my friend should choose Montreal to have better chance to retrieve his living environment he has in Paris.

**Resultat2 :**

In addition to choosing Montreal, my friend should choose among Montreal 5 borough that are closest to Paris’s 11e arrondissement: Le Plateau-Mont-Royal, Mont-Royal, Westmount, Pierrefonds-Roxboro or Anjou.

## Discussions

As we see, in the Paris capture, Paris boroughs are similar except one. The same is observed in New York and Toronto. Result would be different if we chose to compare Neighborhoods instead of Borough.

## Conclusion

Even if the sharpness of the result is not here, using Kmeans clustering on venues categories of borough is enough to help my friend to chose his new destination.