Capstone Project

Exploring Paris neighborhoods

Choosing location for a hotel

Stefanskaya Anna

June 2020

Introduction/Background

Paris is the biggest city in France with population over 2 million people. There were more than 2,000 hotels in Paris in 2018.

Introduction/Problem

Online hotel search services usually don’t have search criteria where we can choose nearby places’ categories that we prefer

Introduction/Interest

The algorithm of neighborhoods clusterisation based on nearby places may be used personally while choosing the best place for staying during vacation or business trip. In business it may be used by online hotel search services to improve their search criteria for their users.

Data/Sources

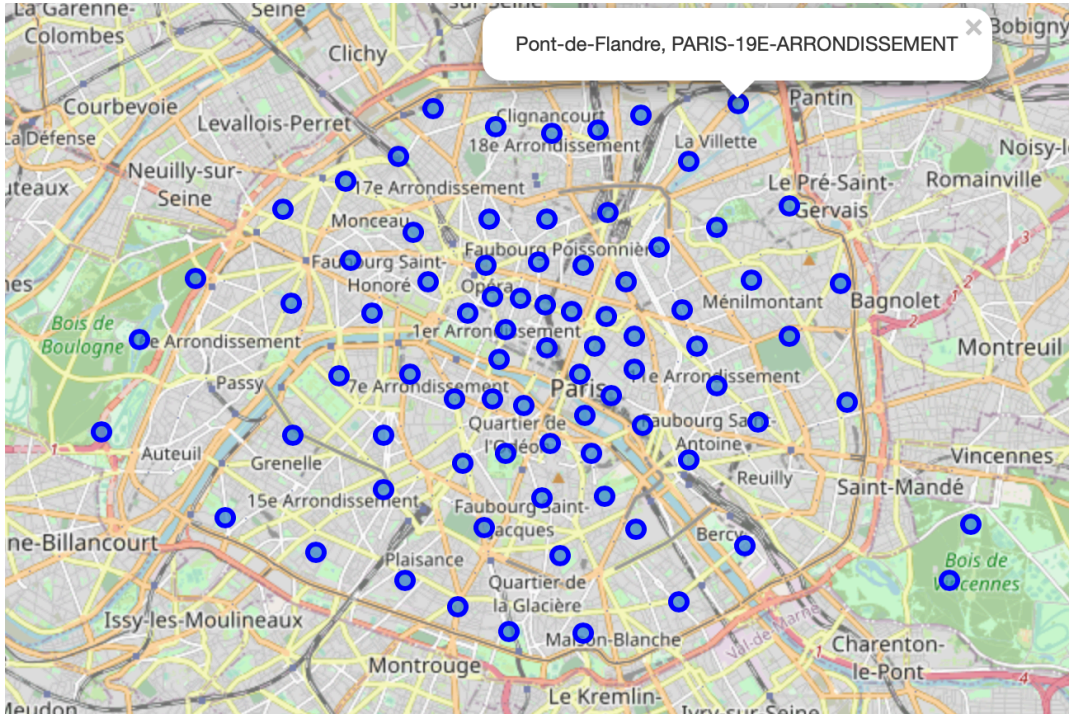
1. France boroughs' location data [1] – 20 boroughs
2. Paris neighborhoods' location data [2] – 80 neighborhoods
3. Data retrieved from Foursquare Places API: venue category, venue location, venue neighborhood – 6529 venues obtained for 80 neighborhoods

[1] https://www.data.gouv.fr/fr/datasets/r/e88c6fda-1d09-42a0-a069-606d3259114e

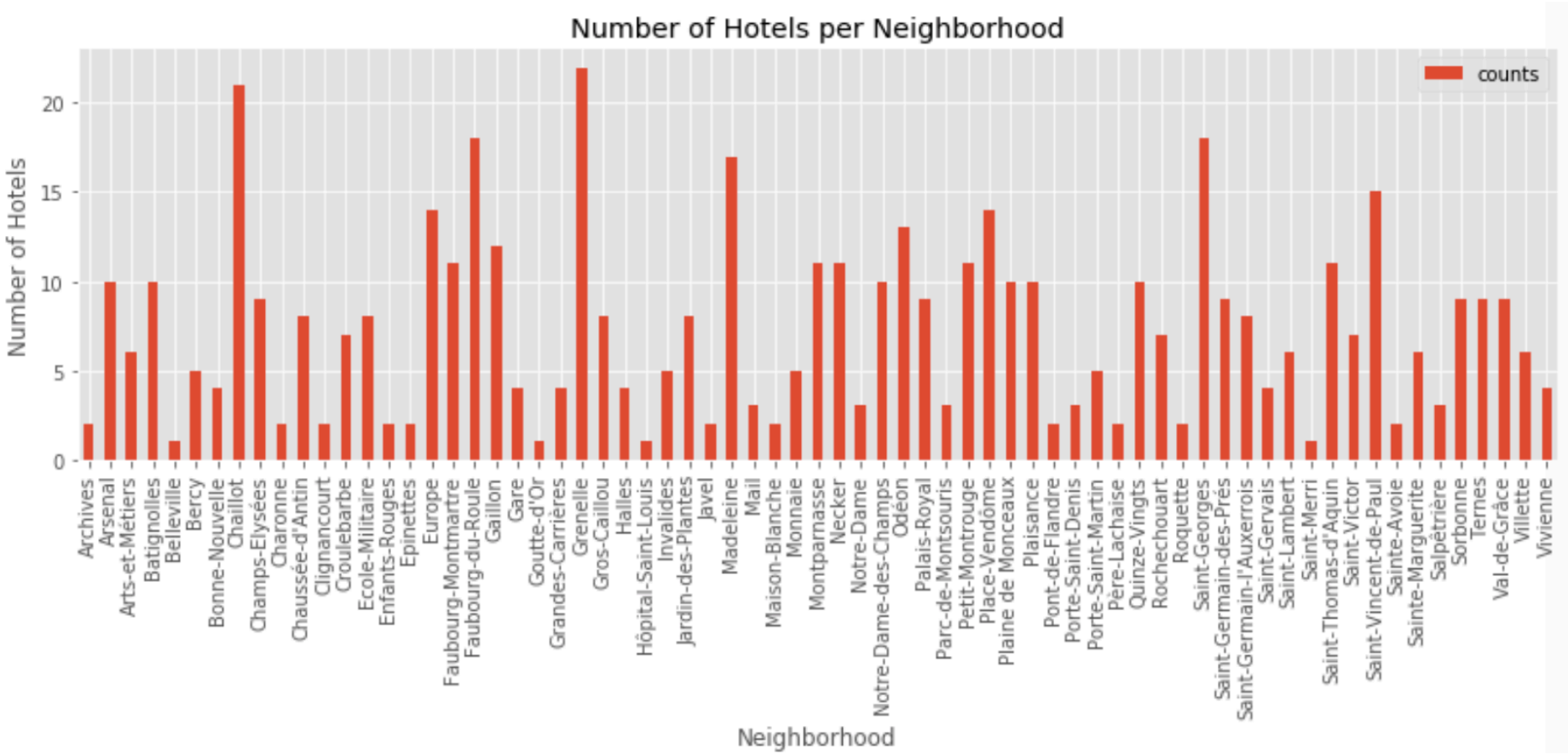
[2] https://opendata.paris.fr/explore/dataset/quartier_paris/download/?format=json&timezone=Europe/Berlin

Methodology/Exploratory Data Analysis

Map of Paris with marked neighborhoods – build with the help of Folium libraries.



Methodology/Exploratory Data Analysis



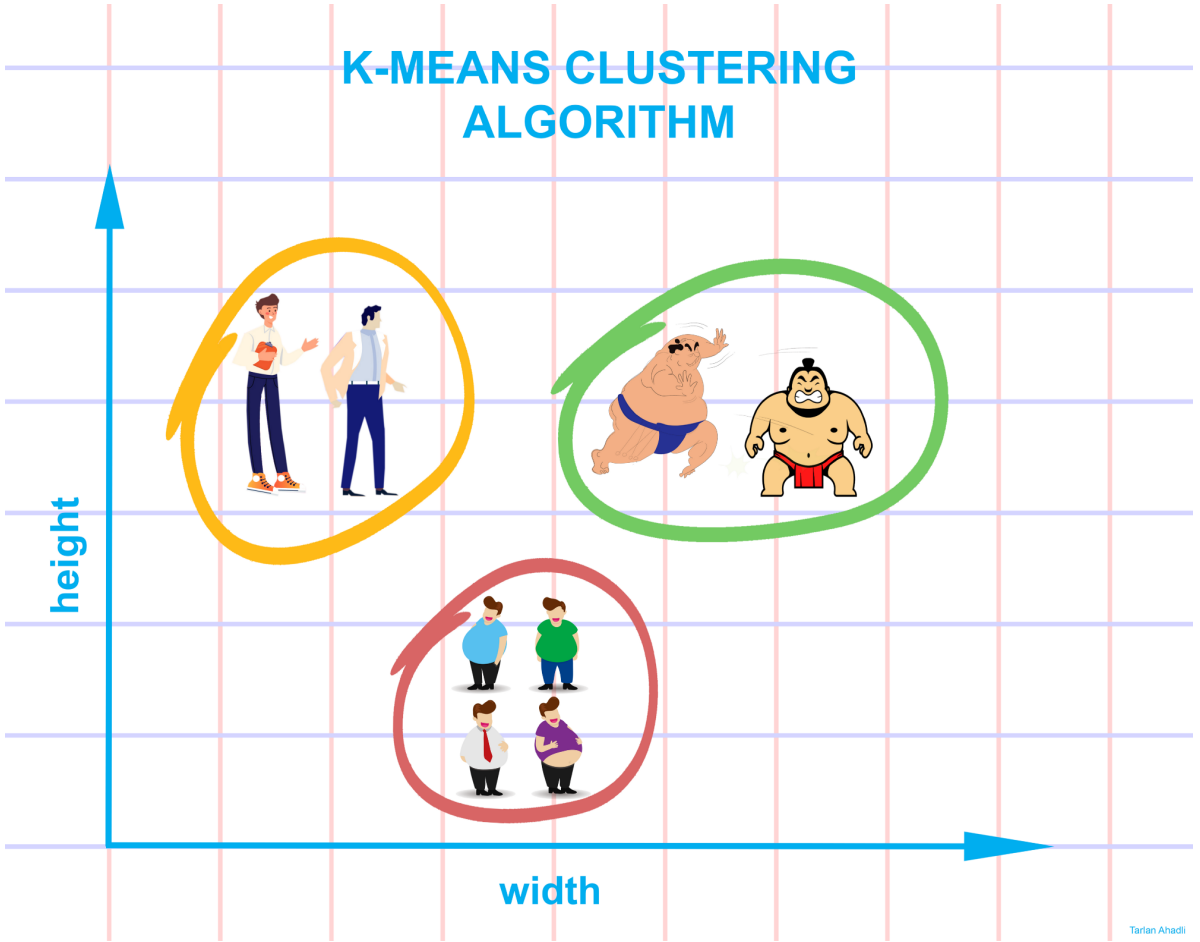
Total neighborhoods – 80; Neighborhoods with hotels – 69

(according to information retrieved using Foursquare Places API given API limitations)

Methodology/Machine Learning Algorithm

We will use:

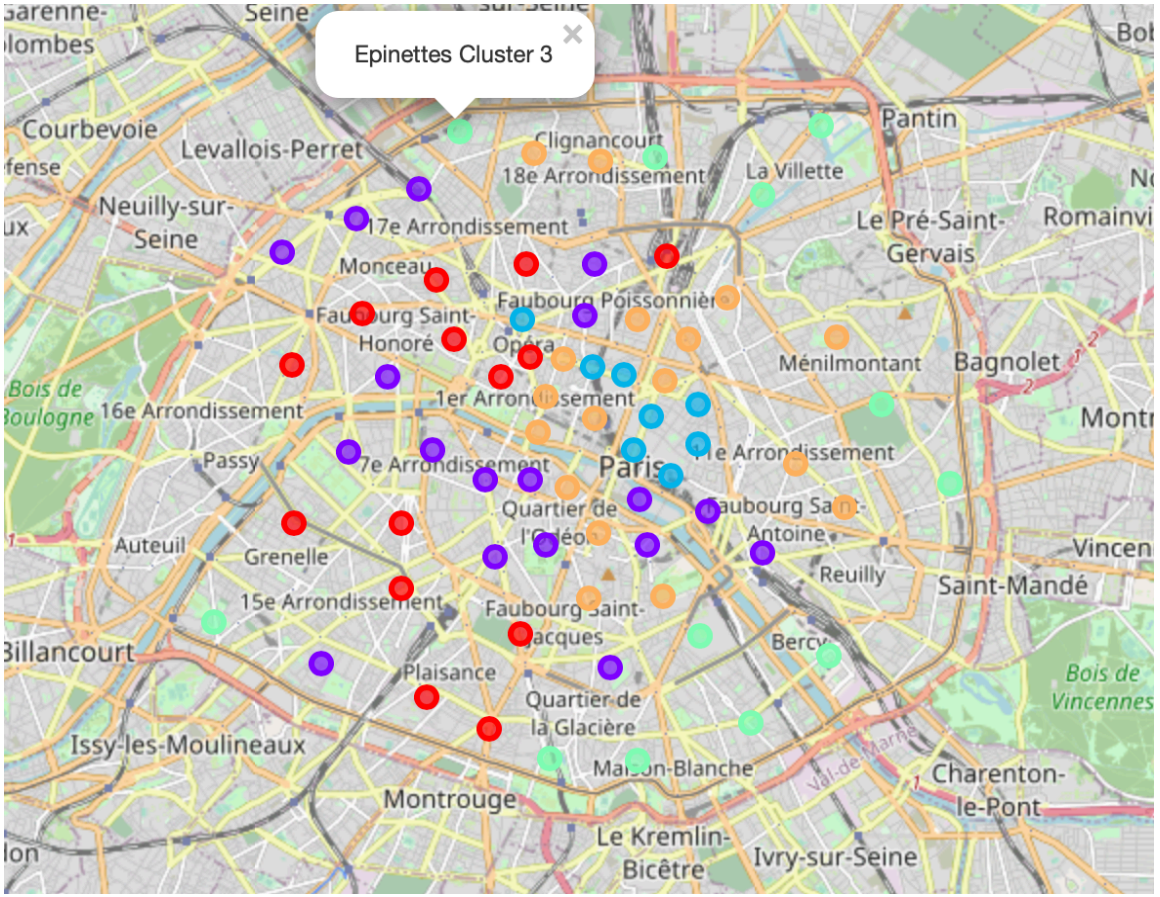
- the *k*-means clustering algorithm to segment the neighborhoods based on nearby venues categories
- k=5 to obtain 5 clusters
- from 69 neighborhoods



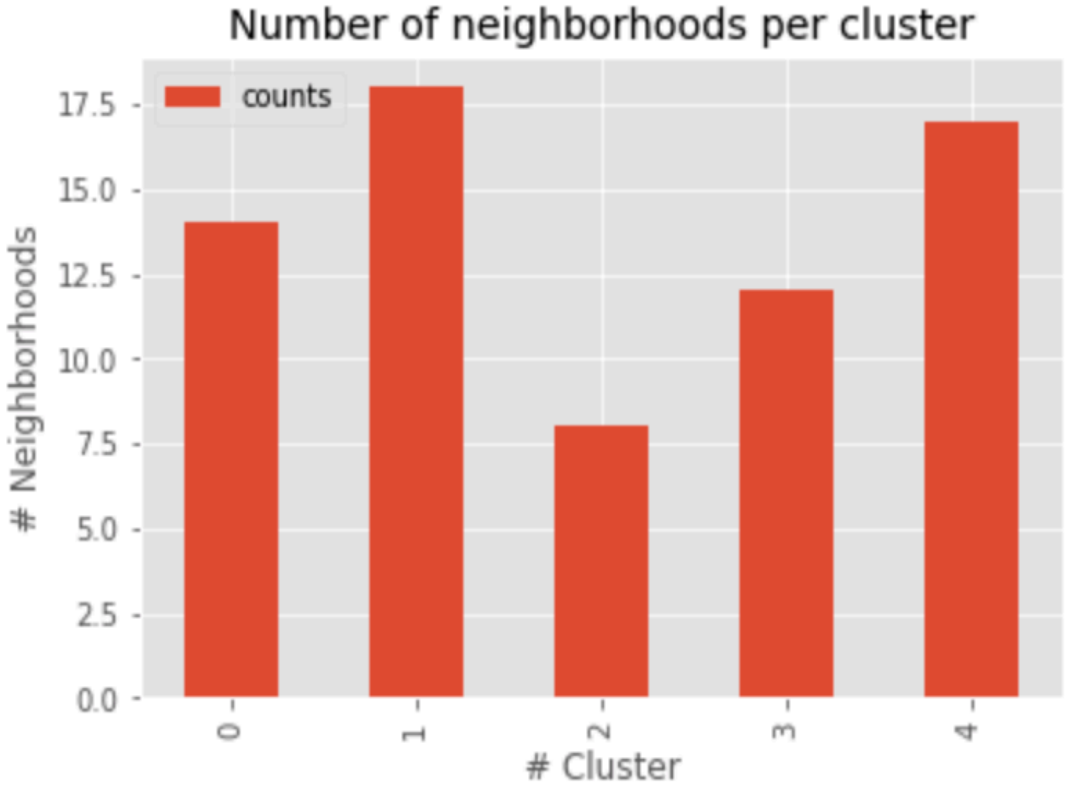
Source: https://medium.com/@tarlanahad/a-friendly-introduction-to-k-means-clustering-algorithm-b31ff7df7ef1

Results

After performing cluster segmentation based on top-10 most common venues in each neighborhood we got 5 clusters



Results



The greatest number of neighborhoods is in cluster 1 and cluster 4. Let's analyze each cluster in detail.

Results

Cluster	0	1	2	3	4
Location	Different locations, mainly in southeast and northwest	Different locations	Center	Far from center	Mainly on the center
1st Most Common Venue	French Restaurant Hotel	French Restaurant	Art Gallery Clothing Store Cocktail Bar French Restaurant	Bakery Bar Bike Rental French Restaurant Grocery Store Hotel Japanese Restaurant Supermarket	Bar Coffee Shop French Restaurant Japanese Restaurant
Neighborhoods	Chaillot Ecole-Militaire Europe Faubourg-du-Roule Gaillon Grenelle Madeleine Montparnasse Necker Petit-Montrouge Place-Vendôme Plaisance Saint-Georges Saint-Vincent-de-Paul	Arsenal Batignolles Champs-Elysées Croulebarbe Faubourg-Montmartre Gros-Caillou Invalides Notre-Dame Notre-Dame-des-Champs Odéon Plaine de Monceaux Quinze-Vingts Rochechouart Saint-Germain-des-Prés Saint-Lambert Saint-Thomas-d'Aquin Saint-Victor Ternes	Archives Bonne-Nouvelle Chaussée-d'Antin Enfants-Rouges Mail Saint-Gervais Saint-Merri Sainte-Avoie	Bercy Charonne Epinettes Gare Goutte-d'Or Javel Maison-Blanche Parc-de-Montsouris Père-Lachaise Pont-de-Flandre Salpêtrière Villette	Arts-et-Métiers Belleville Clignancourt Grandes-Carrières Halles Hôpital-Saint-Louis Jardin-des-Plantes Monnaie Palais-Royal Porte-Saint-Denis Porte-Saint-Martin Roquette Saint-Germain-l'Auxerrois Sainte-Marguerite Sorbonne Val-de-Grâce Vivienne

Results

Let’s consider that we want to book a hotel not far from city center – then Cluster 3 is not an option. But what about other clusters?

Cluster 0 (red markers)

First most common venues are Hotels and French restaurants. There are also Japanese and Italian restaurants, shops, cafes and some bars.

Cluster 1 (purple markers)

First most common venue is French restaurants, second most common venue is Hotel. This cluster is very similar to Cluster 0, but it also has many Bakeries, some galleries and museums, parks and wine places.

Results

Let’s consider that we want to book a hotel not far from city center – then Cluster 3 is not an option. But what about other clusters?

Cluster 2 (blue markers)

This cluster consists only of 8 neighborhoods and 4 boroughs. 1st most common venues are French restaurants, cocktail bars, clothing stores and art galleries.

Cluster 4 (orange markers)

The most common venues are French restaurants, coffee shops, bars and Japanese restaurants. However there are also many places with Italian food, wine places, some shops and museums.

Testing the model

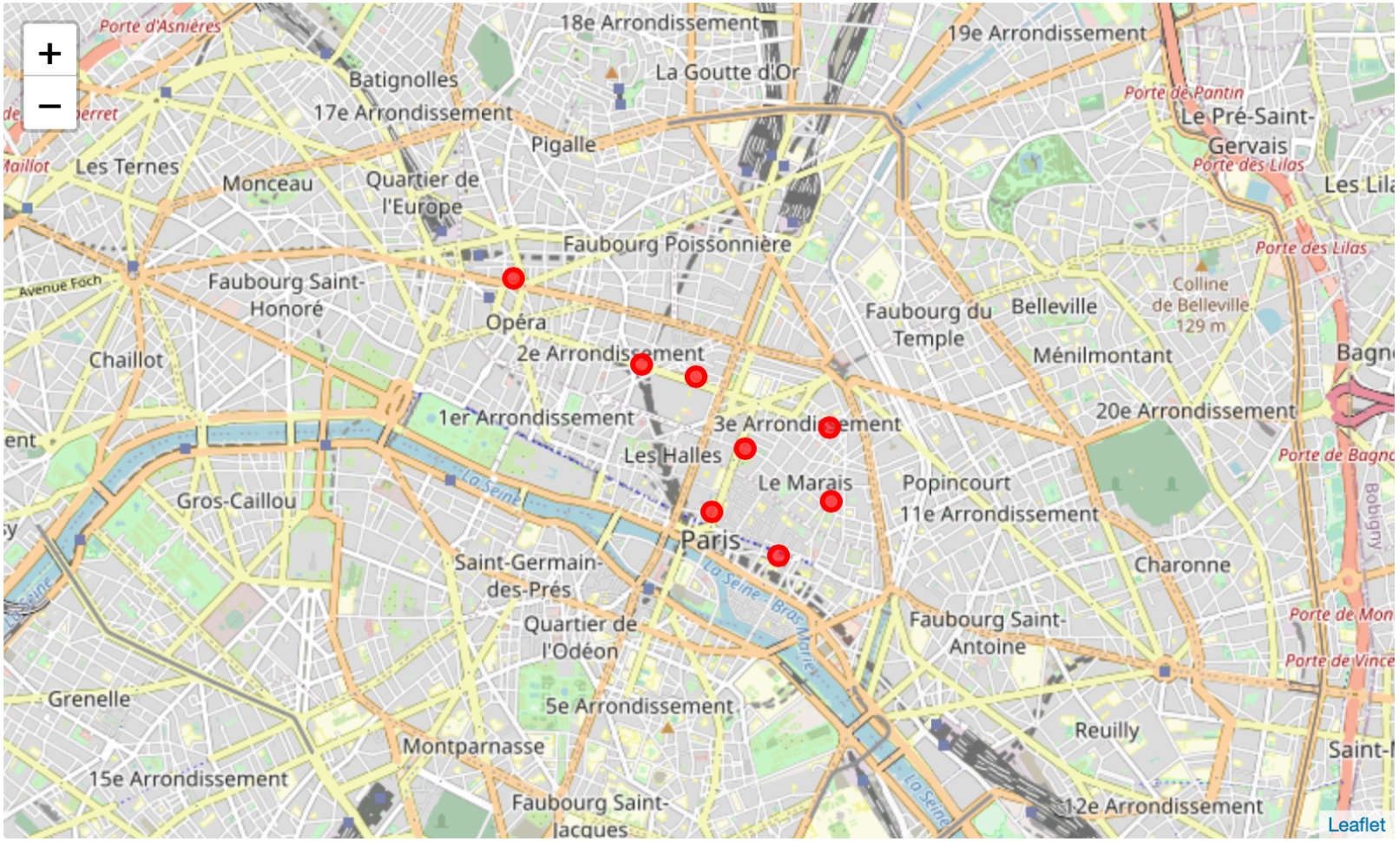
Let’s model a situation when you and you partner are choosing the best neighborhoods to book a hotel for your next vacation.

Your preferences are:

1) In the city center
2) French food
3) Italian food
4) Cocktail bars
5) Art Galleries
6) Shopping

Hotels located in Neighborhoods from Cluster 2 – is the best choice for you!

Testing the model



List of Neighborhoods from Cluster 2:

1) Archives
2) Bonne-Nouvelle
3) Chaussée-d'Antin
4) Enfants-Rouges
5) Mail
6) Saint-Gervais
7) Saint-Merri
8) Sainte-Avoie

Conclusion

Based on your preferences on what places you like to visit during your vacation and with the help of machine learning algorithms you may choose the best suitable hotel in any city.

Therefore, using clustering algorithm based on nearby places you can analyze neighborhoods for many different purposes, for example it may help you to choose the location for your new business or exploring the boroughs for buying or renting an apartment.