Real State Agent Helper

A machine learning tool that helps find the right home for you!!

Introduction. Business Problem

Real State Agents have to sometimes face specific client demands about the characteristics of the different locations. This information is not so easy to access and summarize in order provide useful information for the clients.

- Examples of different real state clients demands concerning locations are:
- How many primary, high schols are there for my kids? Are there parks around?
- Are there many bars and restaurants around in order to enjoy the night life?
- How about cultural locations. I love to go to the teather and cinemas. Are There many options around?

Introduction. Business Problem

The general question to ask is: How can we provide useful information about locations regarding a specific real state client demand. In other words, Is the future accommodation suitable for a client specific needs?

Scenarios

We defined three (3) kind of needs or scenarios:

- Family environments a cultural environment: is the location suitable for kids and families?. Are there any parks, schools, is it safe? Are there museums, cinemas and theaters?
- Nightlife environments: is the location suibale for a person that enjoys go out frequently. How many bars, restaurants, discos are there
- Service facilities: how many service facilities are there? such as hotels, banks, train stations, spa, gyms?

This solution also compares how any particular location is similar to another concerning these scenarios

Where and How?

- Where?

This solution is suitable for any kind of city that possess many kind of different locations in a defined area. For this project the city of Paris will be explored. Why Paris? It is one of the cities in which each location has its particularity and provides different kind of needs

- How?

For this project, a clustering machine learning solution will be needed in order to group the different neighborhoods for the 3 defined scenarios

Data sources and collection

In order to solve the problem, we have to collect, combine and analyze two kind of informations/datasets:

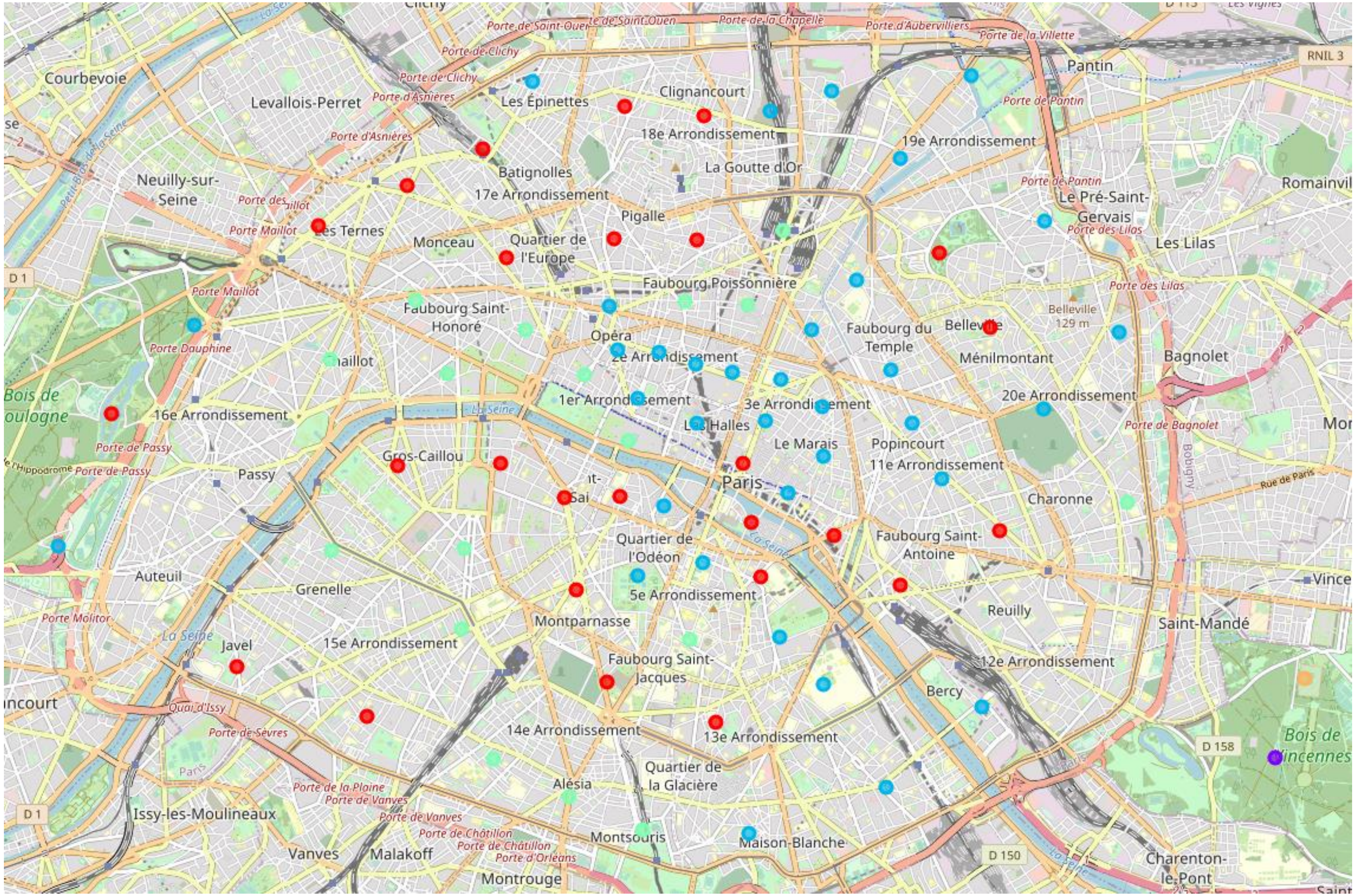
- *Demographic information: the city* of Paris is divided by (arrondissements) which are the districts and "quartiers" which we can translate by neighborhoods. A public dataset of the Paris cityhall website offers this information https://opendata.paris.fr/explore/dataset/quartier_paris/table/.
- *Location venues:* we need to get a dataset that provides all the different kinds of venues for a particular location. For this project we will use the foursquare API

Methodology

The goal is to propose to the clients the neighborhoods that will satisfy certain need. For that we need the grouped the similar neighborhoods that share the same frequency of venues for each of the category.

- **Since we want to talk about grouping or segment different elements based on similarities or dissimilarities, we will use an unsupervised machine learning technique called Clustering**
- **The algorithm that we are going to use is the K-means algorithm**: K-Means is a type of partitioning clustering, that divides the data into K non-overlapping subsets or clusters without any cluster internal structure or labels. We do not use the DBSCAN algorithm because we are not searching for anomalies. We do not want to leave any neighborhood behind

Clustering the city considering all the venues



Clustering the city considering all the venues

Cluster 1: we can say that in this cluster the most important venues are french restaurants, bar and hotels. This neighbourhoods seem to be very visited by locals and tourist.

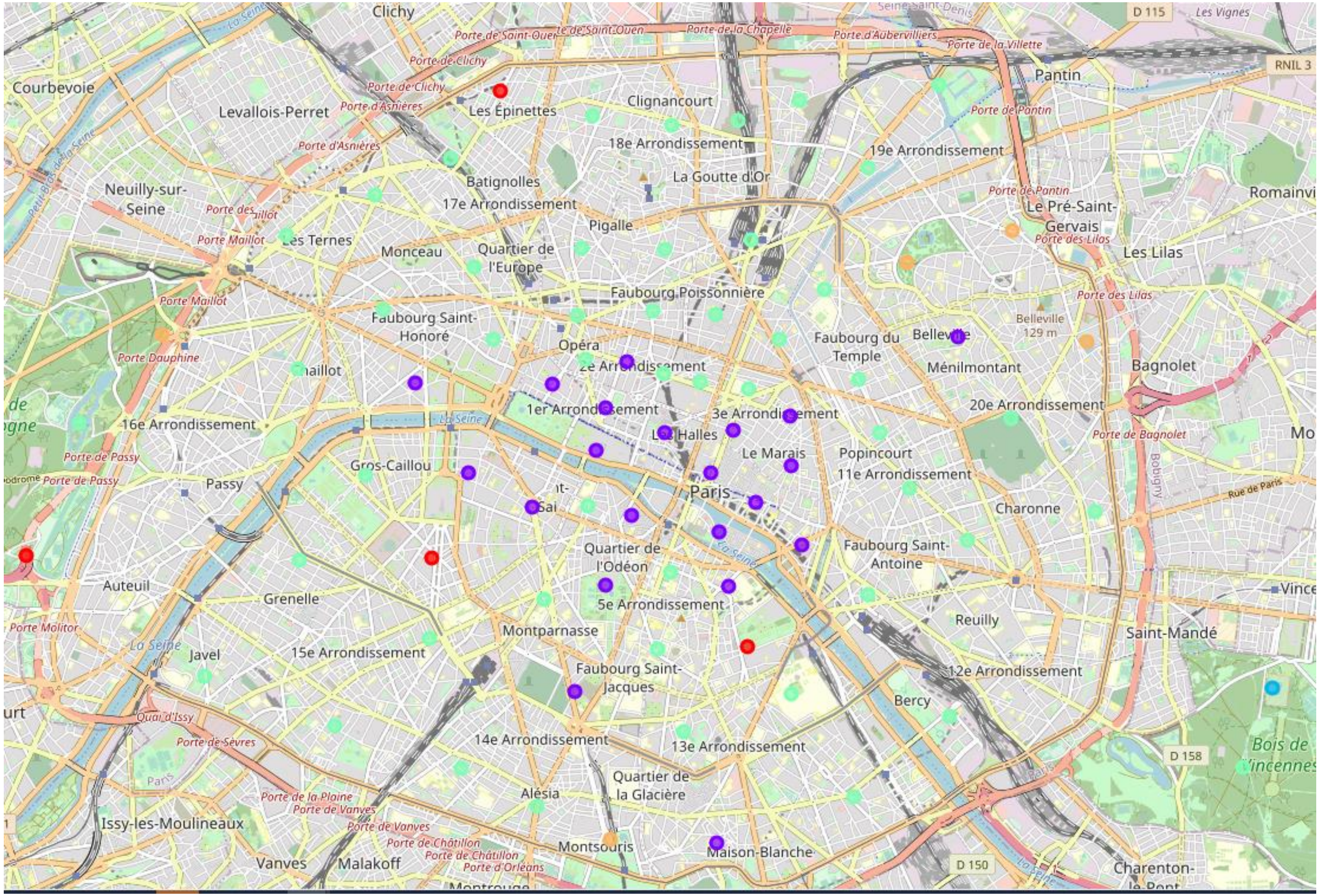
Cluster 2: we only one neighbourhood in this cluster. The most important venues are shops and services. We can see there are open spaces like pedestrian plazas and parks. It is not a crowded area.

Cluster 3: this cluster is similar to the first cluster. However the difference with the first cluster is that we see more varieties of services, shops and open spaces like parks and gardens.

Cluster 4: this cluster is and hybrid of the first and third cluster. We can find a lot of restaurants and hotels like the first cluster. This cluster has also many shops and services facilities like the third cluster but less open spaces.

Cluster 5: As Cluster 2. This cluster has one neighbourhood. It is like cluster 2 because there are many open spaces like playgrounds and plazas. However, there are more restaurants and bars than the cluster 2 and the shops are different

Scenario 1: Clustering the city considering the family, cultural venues



Scenario 1: Clustering the city considering the family, cultural venues

Cluster 1: we have found 4 neighbourhoods. If you love gardens and general museums these neighbourhoods are four you

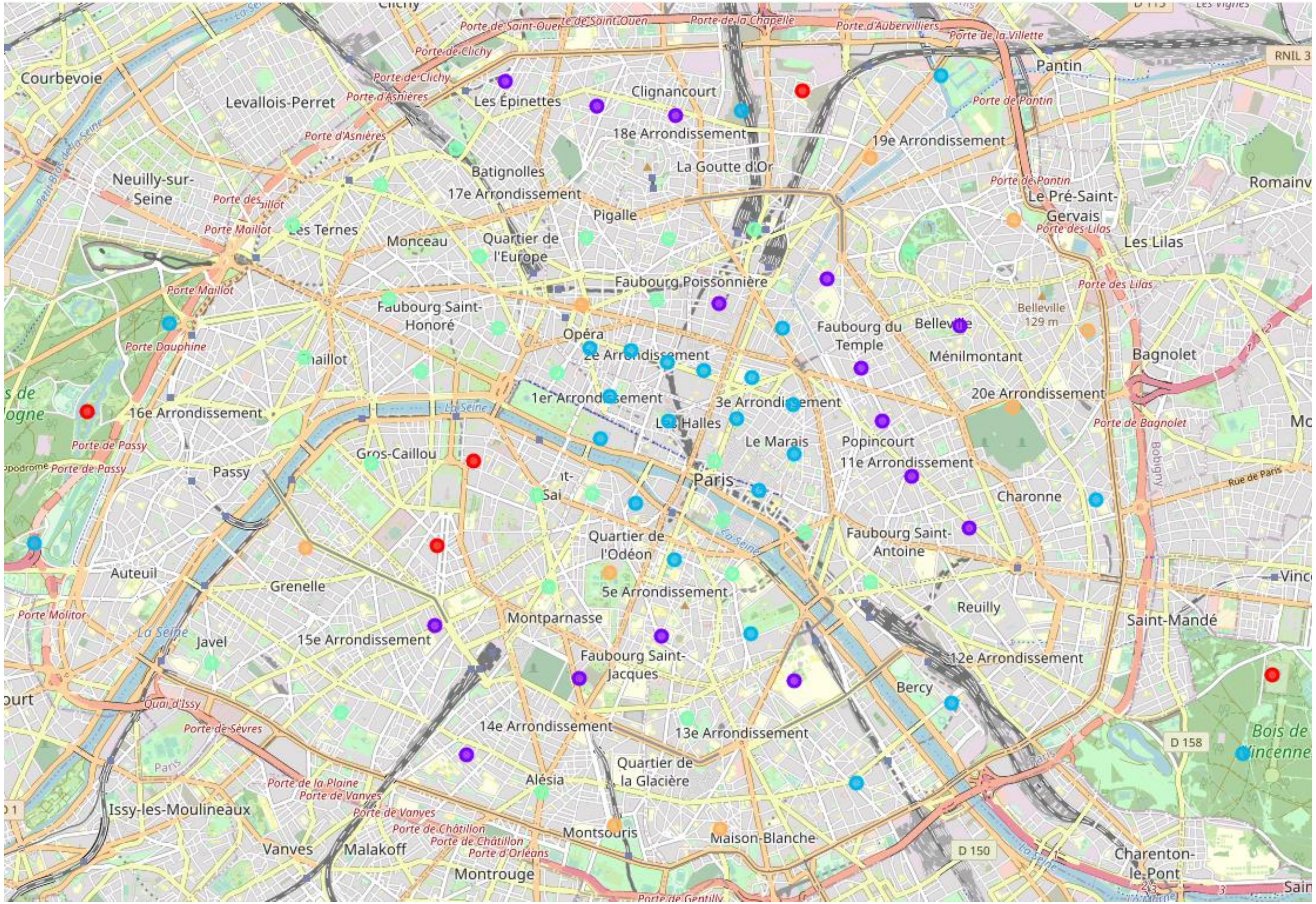
Cluster 2: this is the zone for the art galleries and art museums

Cluster 3: this cluster has only one neighbourhood. We do not find a lot museums nearby. But this neighbourhood is good if you like open spaces, concerts and general cultural activities.

Cluster 4: this is the cluster with the most number of neighbourhoods. Here you can find many cultural shops like bookstores, as well as general museums and theatres. It has many cultural centers too and gardens. In general a very versatile cluster a typical downtown cluster

Cluster 5: If you are looking for parks and great open areas these are the neighbourhoods for you.

Scenario 2: Clustering the city considering the nightlife venues



Scenario 2: Clustering the city considering the nightlife venues

Cluster 1: we have found 5 neighbourhoods. This first cluster is for restaurant lovers only. We can find mostly French restaurants but many other options to eat too. However, no bars nearby

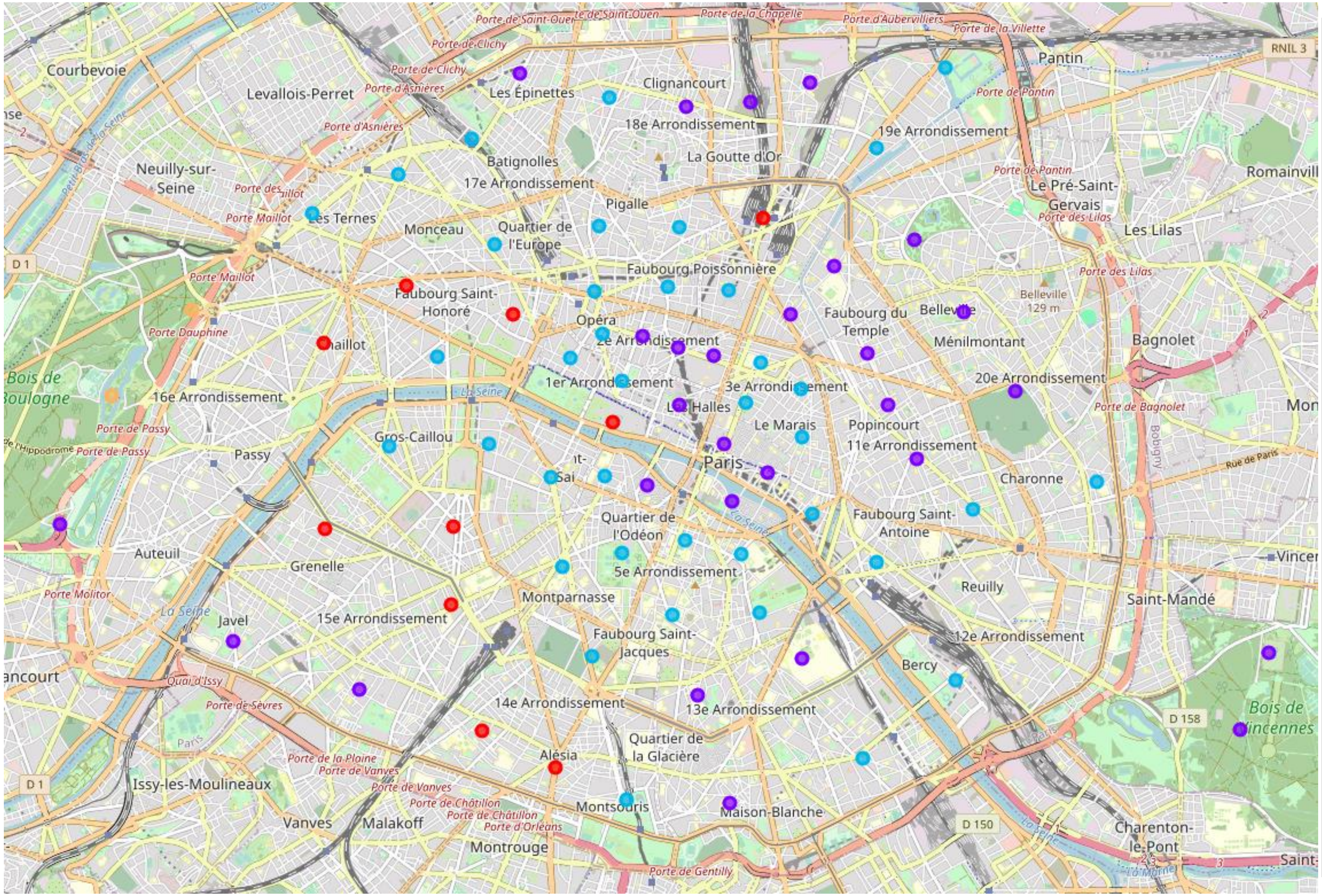
Cluster 2: in this second cluster we find also many French restaurants and a well variety of international restaurants. The difference with the first cluster is that we found a lot of bars too. It seems the neighbourhood to have fun

Cluster 3: the neighbourhoods of this cluster has a very good balance between restaurants and bars

Cluster 4: as cluster number 1, neighbourhoods on cluster 4 offers mostly restaurants. No so many bars. However, we found more italian restaurants here

Cluster 5: this cluster has a versatile food option: french, italian and japanese restaurants, bistros and more calm venues such as cafés.

Scenario 3: Clustering the city considering the service venues



Scenario 3: Clustering the city considering the service venues

Cluster 1: the neighbourhoods of this cluster are suitable for tourism due to the variety of hotels

Cluster 2: The neighbourhoods of this cluster provide an interesting variety of services. We can find hotels, many supermarkets, and bakeries. It is a typical downtown neighbourhood

Cluster 3: This cluster is a typical downtown cluster too. It is very suitable for tourists as cluster 1 due to the variety of hotels. The difference from cluster 1 is that cluster 3 seem to offer more bakeries and supermarkets

Cluster 4: the two neighbourhoods of this cluster offers many services, but they are not suitable for visitors due to the lack of hotels offers

Cluster 5: like Cluster 4, the neighbourhoods of this cluster are not suitable for tourist. It is not a typical downtown cluster because we do not find supermarkets and bakeries

Results and discussion

- This tool has stablished 4 types of pool of clusters. Bellow we can describe the results:

a) General clustering:

- In this first analysis we have seen that there are neighbourhoods more crowded than others.
- We identify neighbourhoods that offers more open spaces than others (playgrounds, gardens and plazas)
- We identify crowded areas that offers several kind of restaurant venues and services like shops

b) Family and cultural clustering:

- We identify cluster of neighbourhoods that offers open spaces like gardens and museums
- We identify a cluster in which art is essential in the form of art museums and galleries
- There are other neighbourhoods in which we do not find art and museums but it offers other cultural activities like concerts
- There is a cluster in which we found many bookstores and theatres

Results and discussion

c) Nightlife clustering:

- Neighbourhoods that offers more options to eat than to drink (not so many bars)
- We have found also the opposite. Clusters that offers a very good variety of bars and nightclubs
- Neighbourhoods that offers the two options: a more calm venues like restaurants and cafés and more "loud" venues like bars and nightclubs

d) Services clustering:

- There are neighbourhoods more suitable for the tourism which offers a good variety of hotels
- There are neighbourhoods that offers more "downtown" kind of life with a good variety of supermarket, shops and bakeries
- There are neighbourhoods less crowded that do not offer venues like supermarket, groceries stores and bakeries

Conclusions

This tool helps real state agents to aim and search the most suitable neighbourhoods considering many criterias. It also collects, summarizes and segment the group of neighbourhoods that offer similar venues.

With this tool we can get:

- The neighbourhoods more crowded and less crowded
- The neighbourhoods that offers more open spaces like gardens and parks
- The neighbourhoods that offers the amount and kind of cultural offer like art galleries, museums and theatres
- The neighbourhoods that offers more restaurants and cafés.
- The neighbourhoods that are more suitable for tourism
- The neighbourhoods that are more suitable for nightlife activities (bars and nightclubs)
- Finally, as we use a cluster algorithm. The neighbourhoods are grouped. This could help a client if they want to go to a neighbourhood that is similar to the current one or a total different one.