# Opening a co-working center in paris.

## Table des matières

# I. Introduction

Telecommuting appeared in the 1970s and has been strengthen by the information and communication technologies development. Today, an employee can work from home as if he were at work. The advantages of teleworking are multiples for both employees and companies, in example, you could significantly reduce your offices rental costs if all of your employees worked from home half of the time. We can also imagine a company without any office and only telecommuter. On the other side, for the employees, working from home is a life-changer, because you don't have to spend 1hrs per days (or more) in the public transports, it also provides them more flexibility and autonomy so it improves their productivity.

With the recent lockdown imposed by the COVID crisis, compagnies developed the teleworking to continue their activities, and not only the bigger ones but also the smallest !

Thus, more and more people are teleworking now, but few meetings cannot handle remotely or in the staff's lounge ! That is why the co-working places are the alternative to oversized premises.

But where should we implant our new co-working office ? That's the question we will try to answer.

# II. Business Case

We are a real estate company which want to surf on the teleworking wave and find a new location to open co-working office. We have to knowledge to build the best workplaces to allow our clients employees to work efficiently.

We want to open our first co-working office in Paris because with the recent Brexit a lot of financial companies are planning to move there, but there isn't enough business premises available.

For us, the office must be in Paris intra-muros. For information, Paris is divided in 20 arrondissement which are divided in 80 boroughs. So, we want a model that can cluster each borough and is also able to predict at which cluster a GPS Coordinates belongs to.

In addition, to define where is the best boroughs to settle our first office we have done a survey of our customers' employees preferred features :

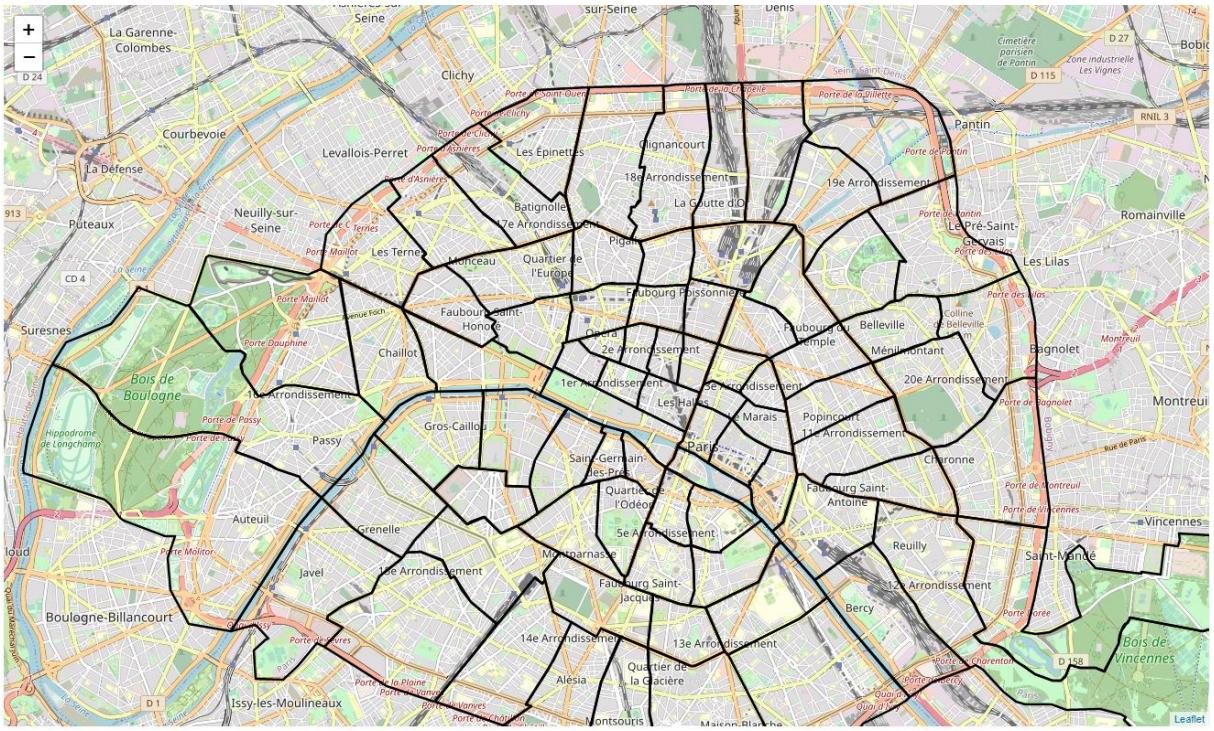
1) It must be easy to reach by public transport ( RER/METRO/TRAM/ Vélib)
2) They like to spend their break doing sport or doing shopping
3) They want to have a lot of restaurant choice
4) It must be easy to book a hostel near the office

# III. Data

## A. Paris' Neighborhoods

Paris is divided in 20 Arrondissements and each arrondissement is divided in 4 boroughs. To get a more precise model we decided to work on the 80 boroughs instead of the 20 arrondissements.

We can find the data on the town of paris website : https://parisdata.opendatasoft.com/explore/dataset/quartier_paris/information/




## B. Public Transport in Paris

We choose to focus on the fastest and easiest way of mobility in Paris : Metro, RER, tramway, navette(mostly airport). This dataset provided by the RATP give us all Paris Metro/RER/Train/Tramway stations. Please note that a station could be used by multiple Metro Lines and RER stations at the same time, a same place will be mentioned as many times as a Metro Line is operating there.

We will not delete the duplicated data because it adds mobility.

https://data.iledefrance-mobilites.fr/explore/dataset/emplacement-des-gares-idf/information/?location=14,48.85351,2.3991&basemap=jawg.streets

## C. Velib Station in Paris

'Vélib' is a public sharing bicycle service. Added to a great public transport service (RER/Metro), Vélib is a great way to do the last kilometers.

The data is provided by the Town of Paris website : https://parisdata.opendatasoft.com/explore/dataset/velib-emplacement-des-stations/information/

## D. Sports Centers

Because the opportunity to do sport during our break is a real advantage to boost the teams, we added their location to our model.

The data is provided by the French Government on : https://www.data.gouv.fr/fr/datasets/recensement-des-equipements-sportifs-espaces-et-sites-de-pratiques-2/

In this DataSet we have got all of the sports equipment in France ordered by department / usage / user.

Only few data interest us :

- DepCode : Department code (Paris intra-muros is 75)
- Utilisateurs : Filtered on : 'Individuel(s) / Famille(s)' ( Only the free-access equipment interests us)
- InsNom : Equipment name

Note : We will clean the duplicated data because we just want to know how many Sportive Centers there is in the borough and not each sportive associations and activities. Moreover, some sports equipment is not inside the Paris intra-muros area and will need to be remove.

### E. Population of Paris per borough

This data represents the population of Paris per borough (in 1999). We didn't find a more recent dataset, however since 1999 the Paris population increased by 3.74% (from 2 125 246 to 2 206 488.)

https://fr.wikipedia.org/wiki/Liste_des_quartiers_administratifs_de_Paris

### F. Foursquare information

We will use the foursquare database to get some information on each borough :

- NB of Hostel per boroughs : 4bf58dd8d48988d1fa931735
- NB of Restaurant per boroughs : 4bf58dd8d48988d1c4941735
- NB of shop/services/malls : 4d4b7105d754a06378d81259

## IV. Methodologie

### 1. Data Cleaning

#### *a) Paris borough*

In this DataSet we have all of the information concerning the 80 borough of Paris, and the most important are :

- NBH_Num
- NBH_AREA_m2NBH_Name
- Arrondissement
- Geometry
- NBH_Latitude
- NBH_Longitude

We will use it to draw a map of paris borough and dispatch every venues/Public Transport Station per borough.

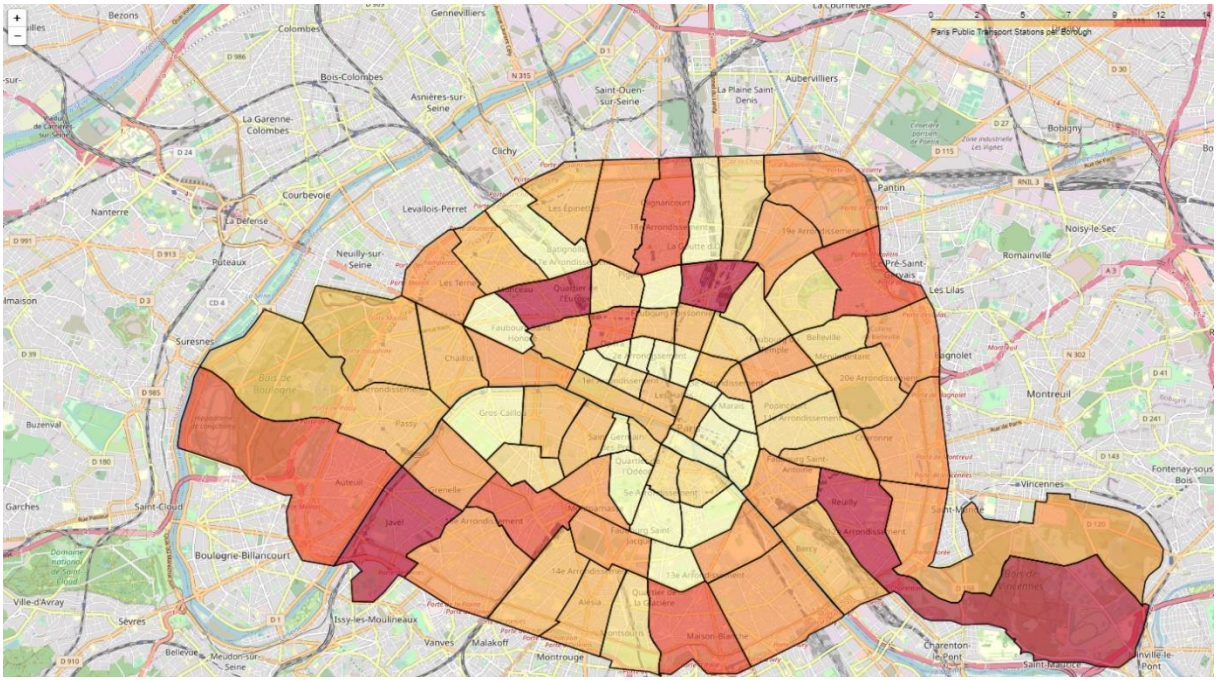
### *b) Public Transport Station*

In this DataSet we have all of the paris public transport station, we will extract :

- Nomlong = Name
- Mode_ = Metro/train/tram
- Latitude
- Longitude

And then, we will determine in which borough are the stations.

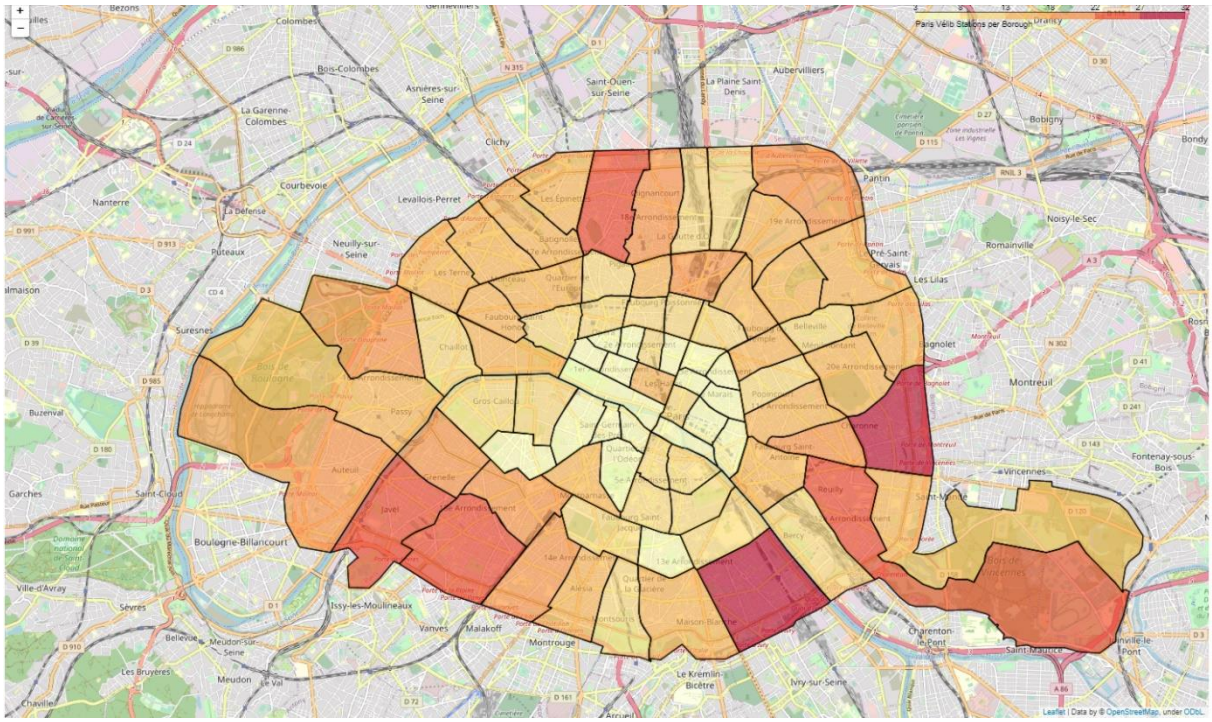
Finally, we will have a Dataframe with the number of Public transport stations per borough.




### *c) Paris vélib stations*

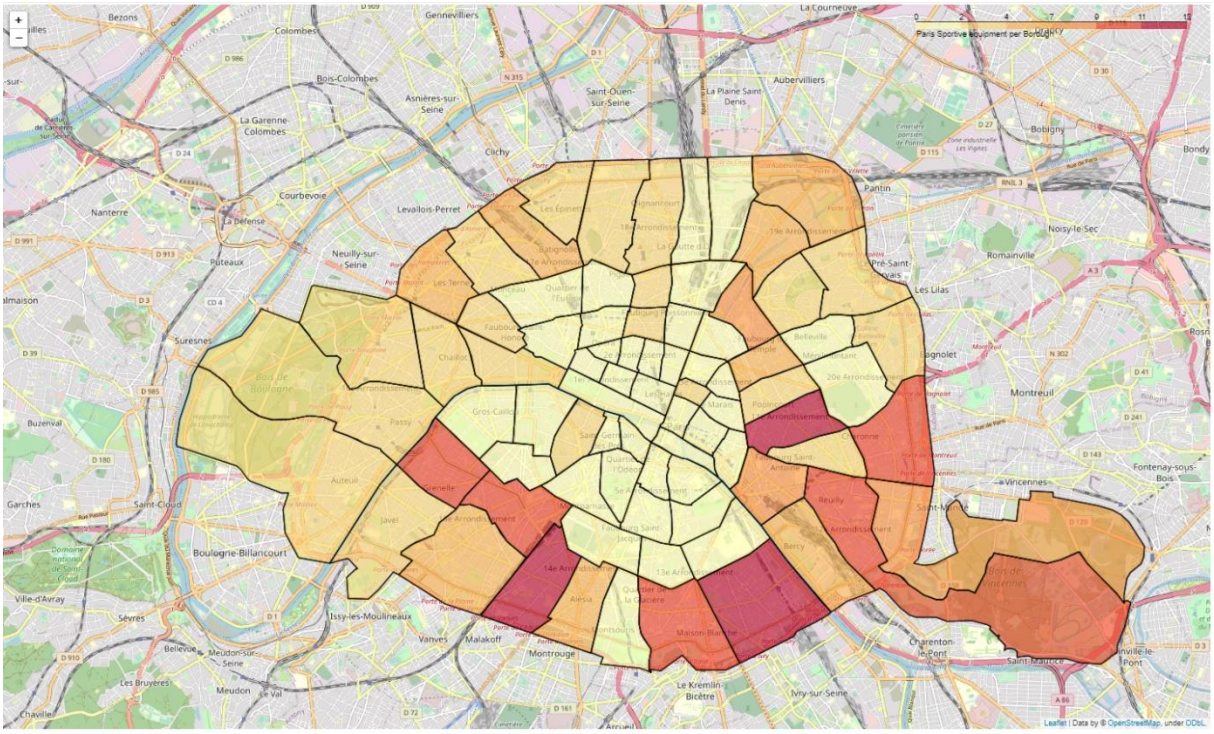
In this Data Set, we will extract all Vélib stations, and determine in which borough they are. However, some stations are not in Paris Intra Muros, so we will delete those rows.

Finally, we will have a Dataframe with the number of Vélib stations per borough.



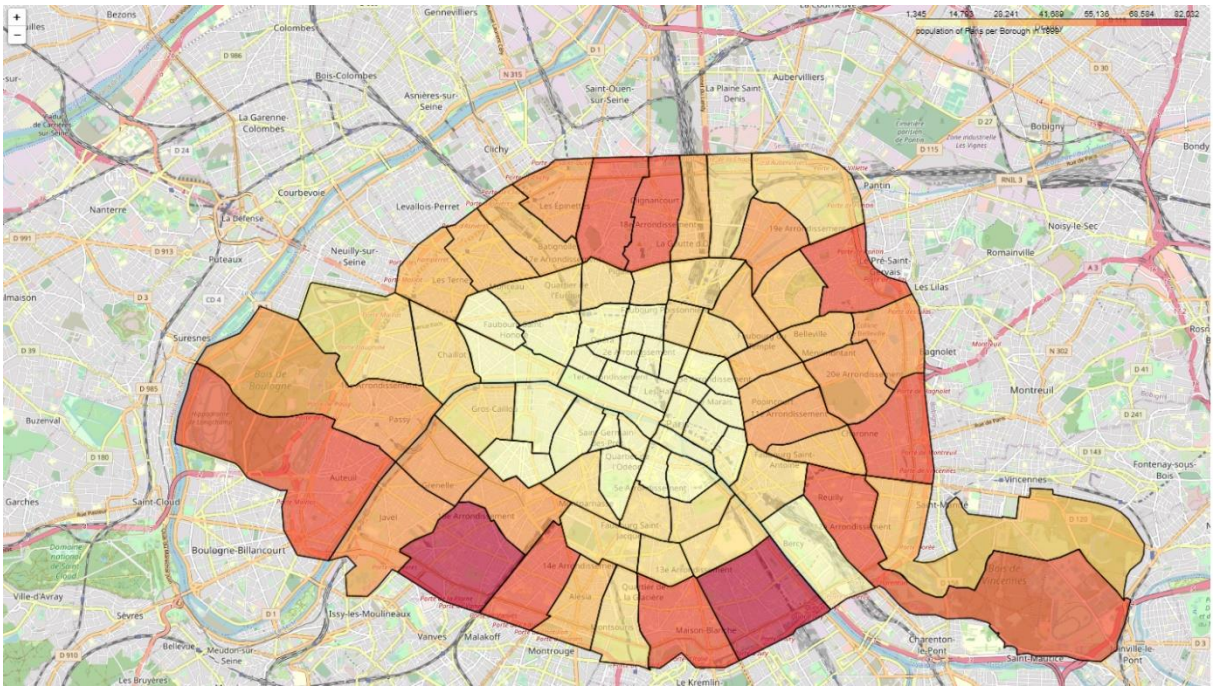

### *d) Sports center*

This Data Set contains all activities in all France's sports centers. So we only want the Paris intra-muros ones, and determine in which borough they are so we could have a dataframe counting the number of sports centers by Borough.




### *e) Population of paris per Borough*

This dataset has been extracted from the Paris wikipedia's page. We will extract the borough and population.

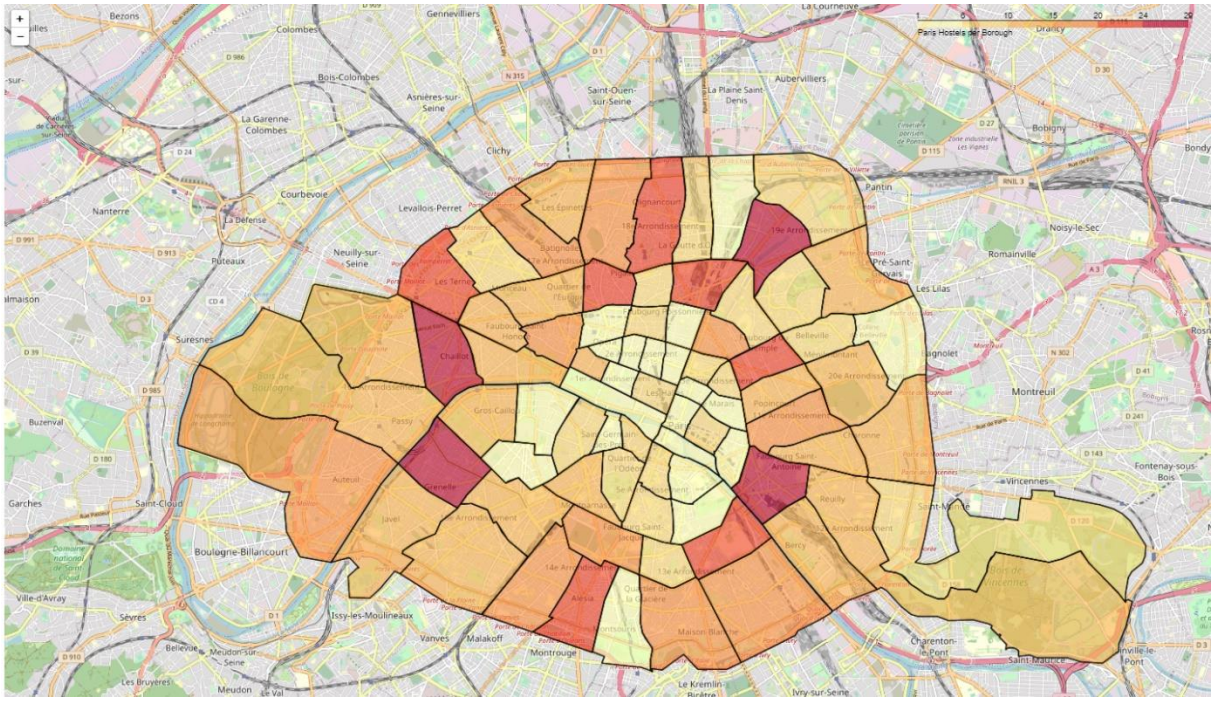



### *f) Foursquare informations*

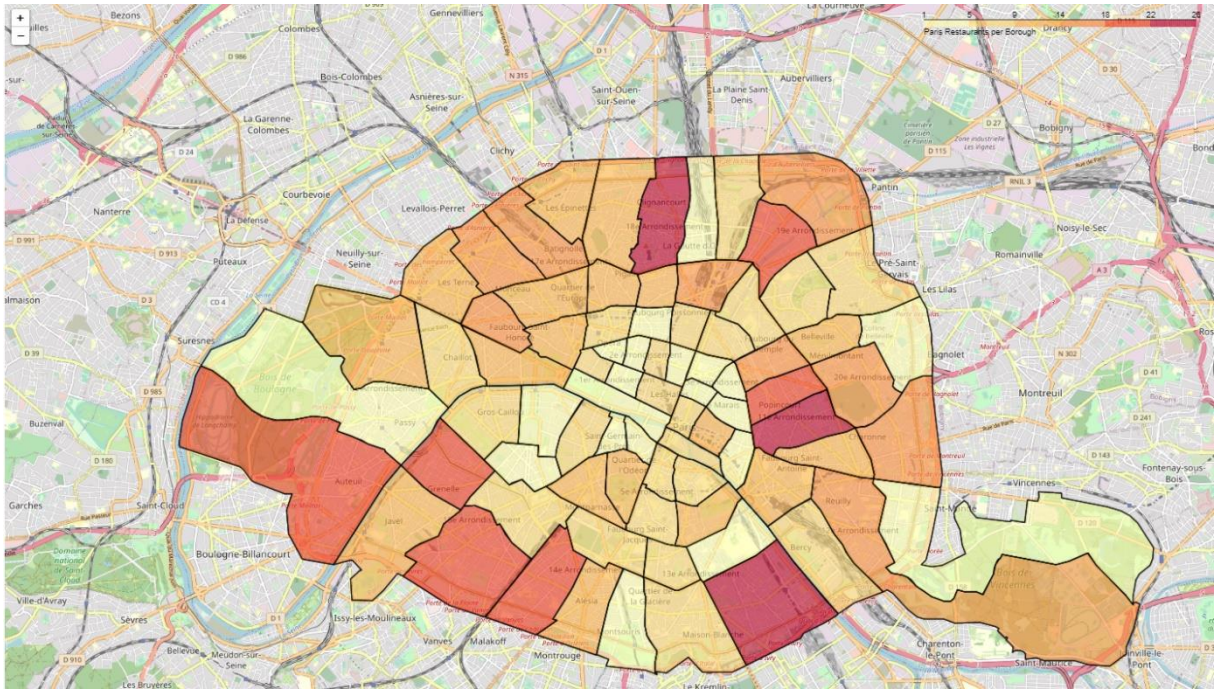
Using the Foursquare database and the paris center of each borrows combinated to a huge radius (1200), we have been able to store every hostels, restaurants, shops/services in a .CSV per feature.

Then we deleted the duplicated rows and affected each venues to a borough.

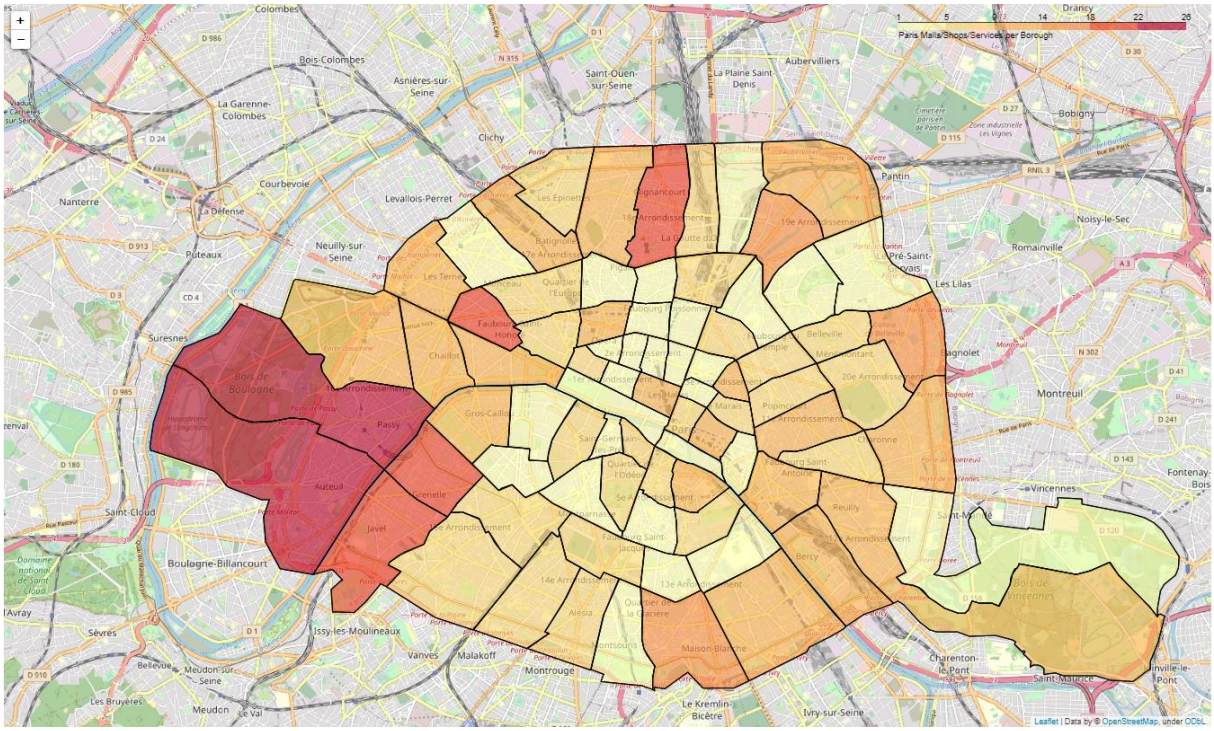
#### (1) Hostels



### (2) Restaurants



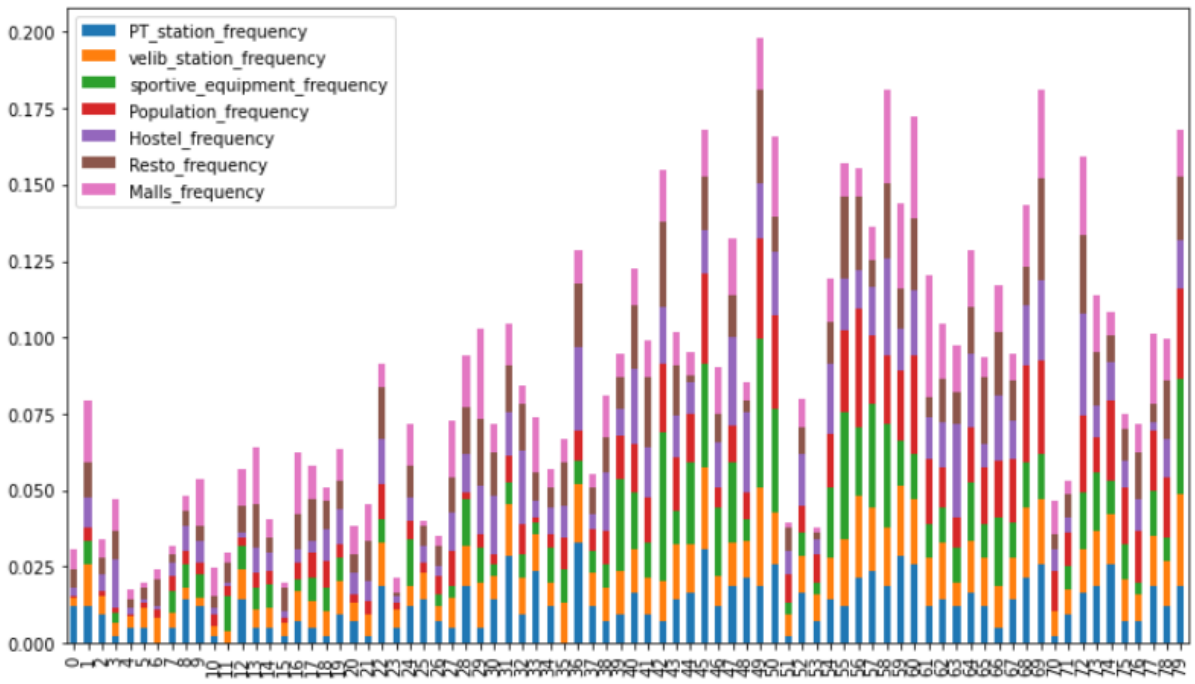

### (3) Shop /Services



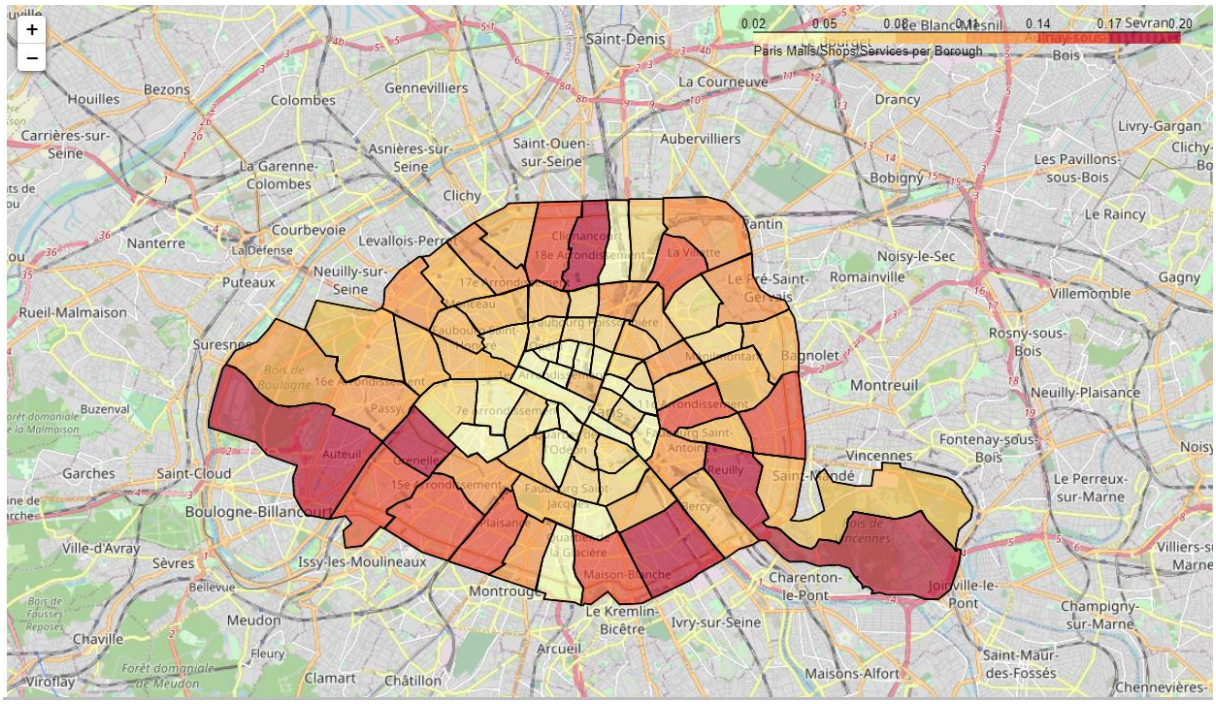

## 2. Data Analysis

### *a) Features cumulated frequencies*

Here are the cumulated frequencies of each features per borough. We can see that the borough 50 is the best one. Also, we can see on this graph that few of our features are located in the 32$^{th}$ first borough. ( 1$^{st}$ – 8$^{th}$ Arrondissement, which correspond to the historical center)



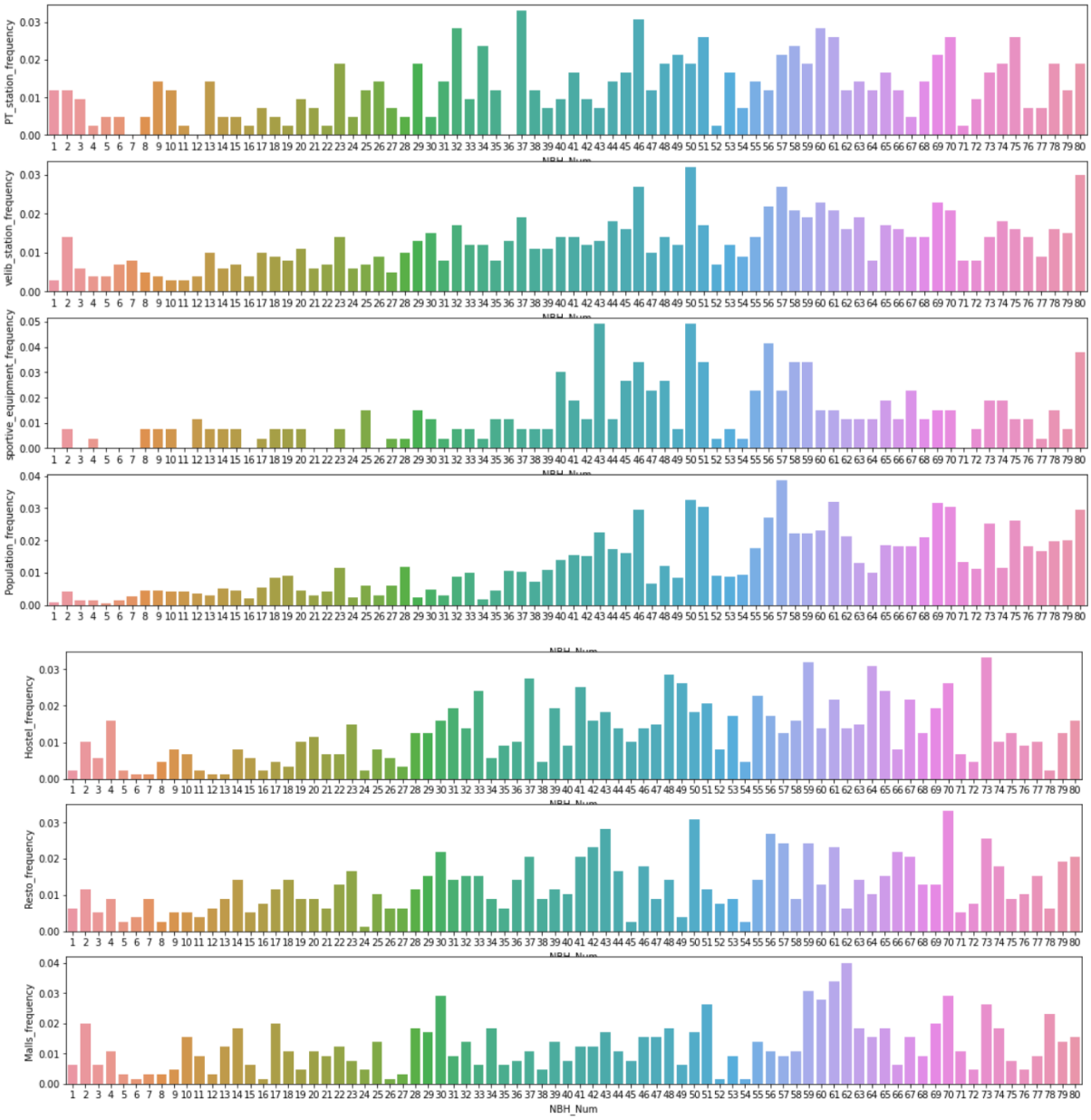

*b) Map of cumulated frequencies per borough*




This map represents the cumulated frequencies of each Borough.

Because we want to maximize each features, this map give us an idea of which borough got the best potential score. However, it don't tell us the composition of each features. In example, 'Auteuil' seems to be a good candidate but is it because 30% of the pop lives here or is there any sports centers near ?

### c) *Detail of each borough features frequencies*




Some Boroughs seems to be good candidates for implenting a Co-Working office like the 50, 80, 70,46 but we would like to know their similarities and determine clusters which we are not able to do by now.

# V. Model selection and details

## A. Model : kMEANS

KMean is an unsupervised iterative algorithm which randomly defines a centroid and measure the distance between every points and it.

The distance between all point and their centroïd represent the model's accuracy. (Because we want to minimize it by calculating the Sum Squared Error).

Also, we want to clearly differentiate each clusters, so the distance between should be maximized. (Silhouette score)

Thus, K means try to minimize the intra cluster distances and maximize the inter-cluster distance.

Because kMEANS is an iterative algorithm, we need to define the best k iteratively first.

Also, our features are very different so we will standardize its data in order to avoid one feature overtake others.

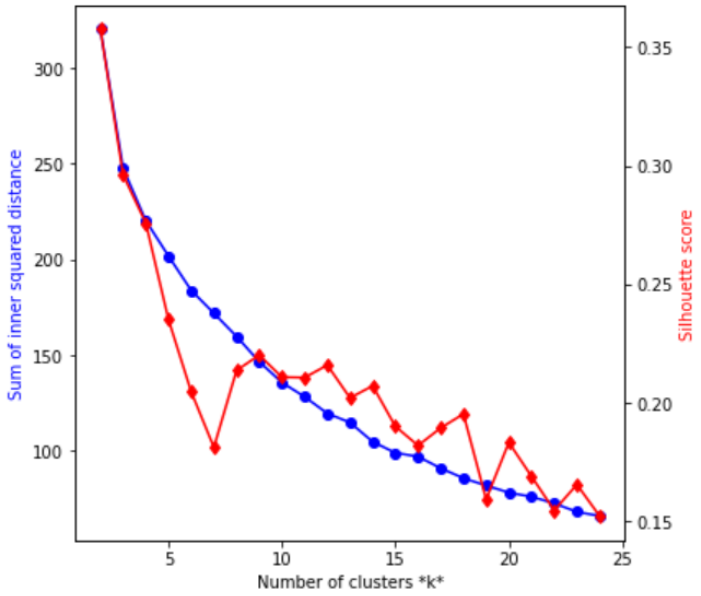
## B. kMEANS : Find the best k (Sum of squared error + silhouette test)

So, how to find out the best k ?

There is two main scorers to evaluate a kMEAN model accuracy :

- Sum of Squared Error ( The sum of the distance between each point and their centroid), it must be minimized.

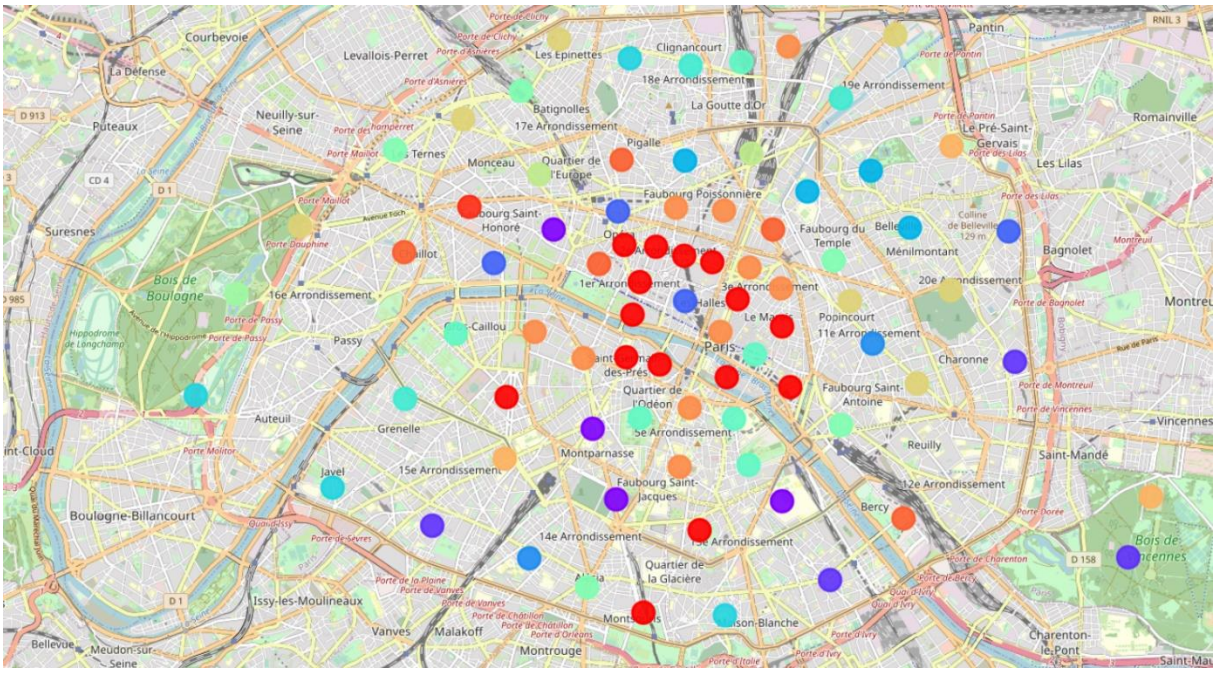
- Silhouette : It represent the distance between each cluster, it must be maximized.




**So, in our case, the best k is : 17.**

# VI. Results

## A. Map of each cluster

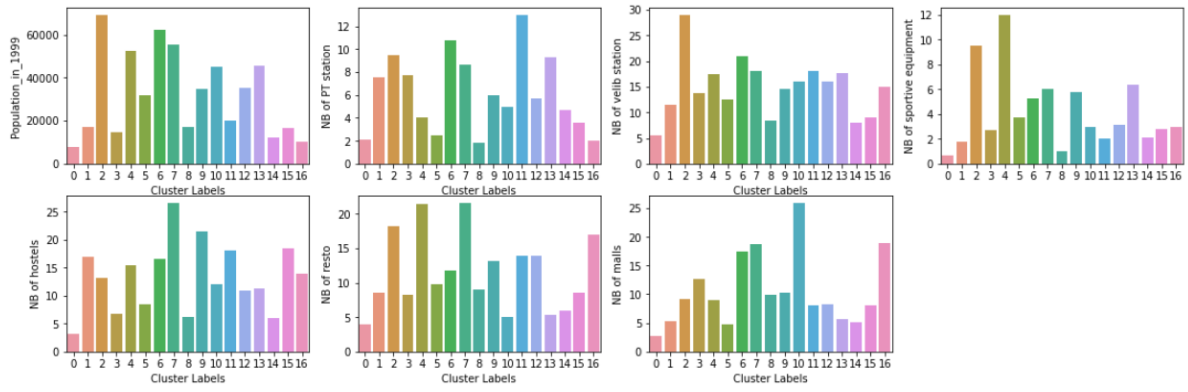



Now we have an idea of the clusters' position in Paris, but which is the best one to implement our Co-working office ?

## B. Clusters analysis

We want to choose the one which maximize every features.

So we will calculate the average of each features by borough.




Following those charts, the **cluster 2 is the best one** to implement our co-working office because it maximized every features. Followed by the 7 and 6 one.

The cluster 2 is composed by :

- Picpus (46)
- Gare(50)
- Saint-Lambert(57)
- Charonne(80)


## VII. Discussion and conclusion

In this study, we have merged a lot of information to describe a Borough : Public Transport Station, Velib Station, Population, Nb of Hostels, Restaurant and Malls.

Because of the huge Number of Borough in Paris, it wasn't easy at the first look to define where is the best place. Even if few Borough seemed to be good candidates.

Next, we have set up a kMEANS model to segement each Borough and we have qualified them. To do it, we have calculated the average of each features by cluster and determined which one had the best score.

That how we clearly defined Cluster 2 as the best one to set our co-working office. It is composed by the following borough : Picpus (46), Gare(50), Saint-Lambery(57), Charonne(80)

Also, it wasn't easy to clean each Data Set because they came from very different sources (wikipedia, French Open Data Project, Foursquare), but it was challenging and if we could have more data on the Foursquare result we could have got a more precise model. Indeed, we have the number of hostel, restaurant, services but we don't know their reviews/rating. If we did, we could have determined the most 'qualitative' borough.

To conclude, I have checked the last news on the cluster 2's borough and for example, Charonne has recently been selected to develop the new Green Borough in Paris. Saint-Lambert is a very cultural borough... So, our model precision is very satisfying.