# Introduction: Business Problem

In this project we will try to find an optimal location for a restaurant. Specifically, this report will be targeted to stakeholders interested in opening an Italian restaurant in Paris, France.

Since there are lots of restaurants in Pariswe will try to detect locations that are not already crowded with restaurants. We are also particularly interested in areas with no Italian restaurants in vicinity. We would also prefer locations as close to city center as possible, assuming that first two conditions are met.

We will use our data science powers to generate a few most promissing neighborhoods based on this criteria. Advantages of each area will then be clearly expressed so that best possible final location can be chosen by stakeholders.

# Data

Based on definition of our problem, factors that will influence our decission are:

* number of existing restaurants in the neighborhood (any type of restaurant)

* number of and distance to Italian restaurants in the neighborhood, if any

* distance of neighborhood from city center

We decided to use regularly spaced grid of locations, centered around city center, to define our neighborhoods.

Following data sources will be needed to extract/generate the required information:

* centers of candidate areas will be generated algorithmically and approximate addresses of centers of those areas will be obtained using **Google Maps API reverse geocoding**

* number of restaurants and their type and location in every neighborhood will be obtained using **Foursquare API**

* coordinate of Paris center will be obtained using **Google Maps API geocoding** of well known Paris location (Place de la République)

# Methodology

In this project we will direct our efforts on detecting areas of Paris that have low restaurant density, particularly those with low number of Italian restaurants. We will limit our analysis to area ~6km around city center.

In first step we have collected the required **data: location and type (category) of every restaurant within 6km from Paris center** (Place de la République). We have also **identified Italian restaurants** (according to Foursquare categorization).

Second step in our analysis will be calculation and exploration of '**restaurant density**' across different areas of Paris - we will use **heatmaps** to identify a few promising areas close to center with low number of restaurants in general (*and* no Italian restaurants in vicinity) and focus our attention on those areas.

In third and final step we will focus on most promising areas and within those create **clusters of locations that meet some basic requirements** established in discussion with stakeholders: we will take into consideration locations with **no more than two restaurants in radius of 250 meters**, and we want locations **without Italian restaurants in radius of 400 meters**. We will present map of all such locations but also create clusters (using **k-means clustering**) of those locations to identify general zones / neighborhoods / addresses which should be a starting point for final 'street level' exploration and search for optimal venue location by stakeholders.

## Analysis

We perform some basic explanatory data analysis and derive some additional info from our raw data. First we count the **number of restaurants in every area candidate**:

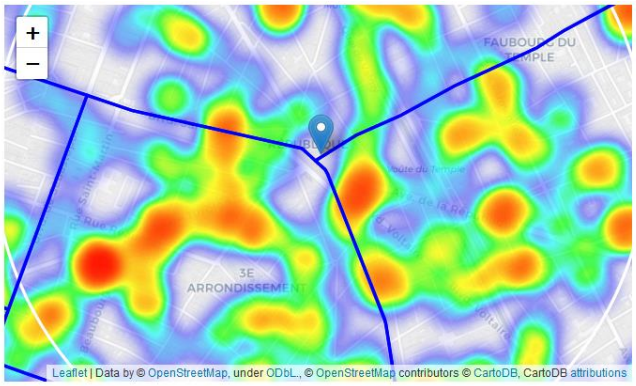
- ⇨ Average number of restaurants in every area with radius=300m: 9.755494505494505.

Then we calculated the **distance to nearest Italian restaurant from every area candidate center** (not only those within 300m - we want distance to closest one, regardless of how distant it is) :

- ⇨ Average distance to closest Italian restaurant from each area center: 420.93064385158203

We founded out that we can find an italien restaurant within 420 m on average.

We created a map showing **heatmap / density of restaurants** and extracted some meaningfull info from that. Also, we added **borders of Paris boroughs** on our map and a few circles indicating distance of 1km, 2km and 3km from Place de la République.




And the focus on italian restaurant :


This map indicates higher density of existing Italian restaurants directly north and west from Place de la République, with closest pockets of **low Italian restaurant density positioned east, south-east and south from city center**.

Based on this we will now focus our analysis on areas *south-west, south, south-east and east from Paris center* - we will move the center of our area of interest and reduce it's size to have a radius of **2.5km**. This places our location candidates mostly in boroughs **Le Marais** or **Quartier Picpus**

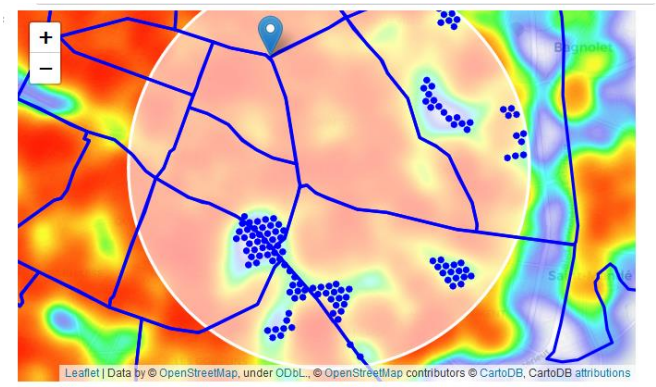
## Focus on Le Marais & Picpus

We calculated two most important things for each location candidate: **number of restaurants in vicinity** (we'll use radius of **250 meters**) and **distance to closest Italian restaurant**.

| | Latitude | Longitude | X | Y | Restaurants nearby | Distance to Italian restaurant |
|---|---|---|---|---|---|---|
| 0 | 48.833218 | 2.378992 | -425162.407865 | 5.486209e+06 | 2 | 317.037739 |
| 1 | 48.833366 | 2.380322 | -425062.407865 | 5.486209e+06 | 2 | 397.993054 |
| 2 | 48.833164 | 2.371485 | -425712.407865 | 5.486296e+06 | 6 | 131.962373 |
| 3 | 48.833312 | 2.372815 | -425612.407865 | 5.486296e+06 | 6 | 116.236013 |
| 4 | 48.833460 | 2.374144 | -425512.407865 | 5.486296e+06 | 3 | 170.345308 |
| 5 | 48.833608 | 2.375474 | -425412.407865 | 5.486296e+06 | 3 | 254.016214 |
| 6 | 48.833756 | 2.376804 | -425312.407865 | 5.486296e+06 | 4 | 307.700049 |
| 7 | 48.833904 | 2.378133 | -425212.407865 | 5.486296e+06 | 2 | 350.092650 |
| 8 | 48.834052 | 2.379463 | -425112.407865 | 5.486296e+06 | 2 | 412.856399 |
| 9 | 48.834200 | 2.380793 | -425012.407865 | 5.486296e+06 | 0 | 375.563755 |

Then we **filtered** those locations: we're interested only in **locations with no more than two restaurants in radius of 250 meters**, and **no Italian restaurants in radius of 400 meters**.

On a heatmap it gives us




We have a bunch of locations fairly close to Place de la République, and we know that each of those locations has no more than two restaurants in radius of 250m, and no Italian restaurant closer than 400m. Any of those locations is a potential candidate for a new Italian restaurant, at least based on nearby competition.

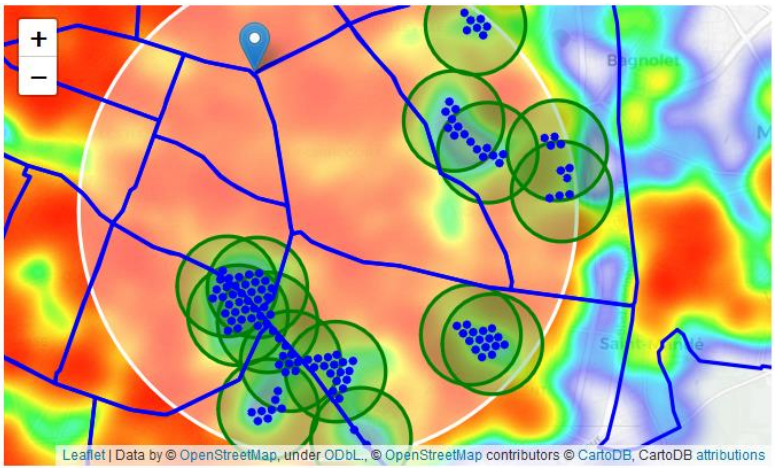
The potential places are then


What we have is a clear indication of zones with low number of restaurants in vicinity, and *no* Italian restaurants at all nearby.

It's interesting to see that a zone of "Cimetière du Père Lachaize" is identified as good place... For sur it is a famous graveyard in Paris !

Let us now **cluster** those locations to create **centers of zones containing good locations**. Those zones, their centers and addresses will be the final result of our analysis.

We used a Kmeans analysis to identify Cluster of location candidates.




Picpus and Le Marais appeared as good place for an italian restaurant.

Finally we found out some adresses recommanded for further analysis.

```
============================================================
Addresses of centers of areas recommended for further analysis
============================================================

Pont d'Austerlitz, Pont d'Austerlitz, 75012 Paris  => 2.5km from Place de
la République
7 Rue du Clos, 75020 Paris                         => 3.3km from Place de
la République
40 Rue de Picpus, 75012 Paris                      => 3.4km from Place de
la République
16 Rue des Rigoles, 75020 Paris                    => 2.3km from Place de
la République
Cimetière Père Lachaise, 16 rue du repos, 71ème division, Avenue Circulai
re, 75020 Paris => 2.1km from Place de la République
11 Port de la Gare, 75013 Paris                    => 3.9km from Place de
la République
Division Pinel, 75013 Paris                        => 3.4km from Place de
la République
11 Bis Quai Saint-Bernard, 75005 Paris             => 2.2km from Place de
la République
66 Quai d'Austerlitz, 75013 Paris                  => 2.9km from Place de
la République
Cimetière Père Lachaise, 16 rue du repos, 43ème division, Avenue Transver
sale n°1, 75020 Paris => 2.5km from Place de la République
7 Rue Victor Segalen, 75020 Paris                  => 3.2km from Place de
la République
9 Boulevard Morland, 75004 Paris                   => 2.2km from Place de
la République
58 Rue de Picpus, 75012 Paris                      => 3.7km from Place de
la République
57 Rue Cuvier, 75005 Paris                         => 2.5km from Place de
la République
28 Quai de la Rapée, 75012 Paris                   => 3.1km from Place de
la République
```

This concludes our analysis. We have created 15 addresses representing centers of zones containing locations with low number of restaurants and no Italian restaurants nearby, all zones being fairly close to city center (all less than 4km from Place de la République, and about half of those less than 2km from Place de la République). Although zones are shown on map with a radius of ~500 meters (green circles), their shape is actually very irregular and their centers/addresses should be considered only as a starting point for exploring area neighborhoods in search for potential restaurant locations. Most of the zones are located in

Picpus and Le Marais boroughs, which we have identified as interesting due to being popular with tourists, fairly close to city center and well connected by public transport.

## Results and Discussion

Our analysis shows that although there is a great number of restaurants in Paris (~2000 in our initial area of interest which was 12x12km around Place de la République), there are pockets of low restaurant density fairly close to city center. We focused our attention to areas south, south-east and east, corresponding to boroughs Picpus and Le Marais. Our attention was focused on Picpus and Le Marais which offer a combination of popularity among tourists, closeness to city center, strong socio-economic dynamics *and* a number of pockets of low restaurant density.

After directing our attention to this more narrow area of interest (covering approx. 5x5km south-east from Place de la République) we first created a dense grid of location candidates (spaced 100m appart); those locations were then filtered so that those with more than two restaurants in radius of 250m and those with an Italian restaurant closer than 400m were removed.

Those location candidates were then clustered to create zones of interest which contain greatest number of location candidates. Addresses of centers of those zones were also generated using reverse geocoding to be used as markers/starting points for more detailed local analysis based on other factors.

Result of all this is 15 zones containing largest number of potential new restaurant locations based on number of and distance to existing venues - both restaurants in general and Italian restaurants particularly. This, of course, does not imply that those zones are actually optimal locations for a new restaurant! Purpose of this analysis was to only provide info on areas close to Paris center but not crowded with existing restaurants (particularly Italian) - it is entirely possible that there is a very good reason for small number of restaurants in any of those areas, reasons which would make them unsuitable for a new restaurant regardless of lack of competition in the area. Recommended zones should therefore be considered only as a starting point for more detailed analysis which could eventually result in location which has not only no nearby competition but also other factors taken into account and all other relevant conditions met.

## Conclusion

Purpose of this project was to identify Paris areas close to center with low number of restaurants (particularly Italian restaurants) in order to aid stakeholders in narrowing down the search for optimal location for a new Italian restaurant. By calculating restaurant density distribution from Foursquare data we have first identified general boroughs that justify further analysis (Picpus and Le Marais), and then generated extensive collection of locations which satisfy some basic requirements regarding existing nearby restaurants. Clustering of those locations was then performed in order to create major zones of interest (containing greatest number of potential locations) and addresses of those zone centers were created to be used as starting points for final exploration by stakeholders.

Final decission on optimal restaurant location will be made by stakeholders based on specific characteristics of neighborhoods and locations in every recommended zone, taking into consideration additional factors like attractiveness of each location (proximity to park or water), levels of noise / proximity to major roads, real estate availability, prices, social and economic dynamics of every neighborhood etc.