Capstone Project - The Paris Job

Finding a new job in Paris

Business Problem

n this project we are trying to start a new job in Paris. Lets say that we want to open a shop of some jop. To help us deside what kind of shop to open, we will examine all Paris venues to see what is the most common type of shops, so we decide to start this kind of shops. Second, will try to find an optimal location this shop. Specifically, this report will be targeted to stakeholders interested in opening a **new shop** in **Paris**, France.

Data

Based on definition of our problem, factors that will influence our decission are:
   number of and distance to shops 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.
   ✍   number of restaurants and their type and location in every neighborhood will be obtained using **Foursquare API**
   ✍   coordinate of Paris center will be obtained manually using **Geo PyNominatim** that will point to the center of Paris

By analyzing the data we found that the most common category is **French Restaurant** so we will foucus in a place to start this job.
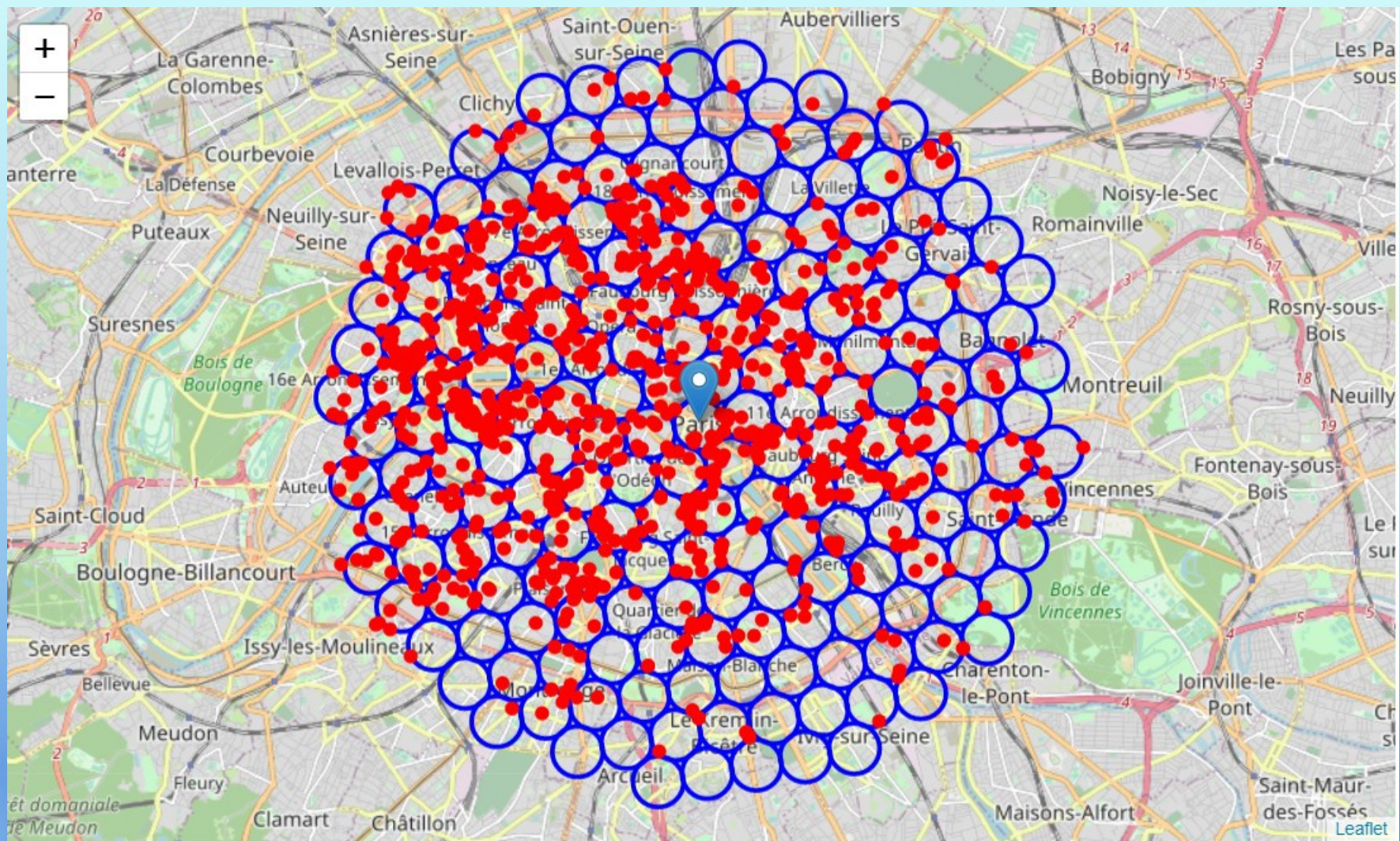
Methodology

n this project we will direct our efforts on detecting areas of Paris that have low French restaurant density. 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**. We have also **identified French restaurants** 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 French 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**: we will take into consideration locations **without French 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.

Analysis

Let's crete a map showing **heatmap / density of restaurants** and try to extract some meaningfull info from that. Also, let's show **borders of French boroughs** on our map and a few circles indicating distance of 1km, 2km and 3km from Paris center.

Looks like a few pockets of low restaurant density closest to city center can be found south, south-east and east from Paris center.

Let's define new, more narrow region of interest, which will include low-restaurant-count parts of south-east closest to Paris center.

Let's also create new, more dense grid of location candidates restricted to our new region of interest (let's make our location candidates 100m appart).
We calculate the most important thing for each location candidate: **distance to closest French restaurant**.
We will **filter** those locations: we're interested only in **locations with no French restaurants in radius of 400 meters**.
We will 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.
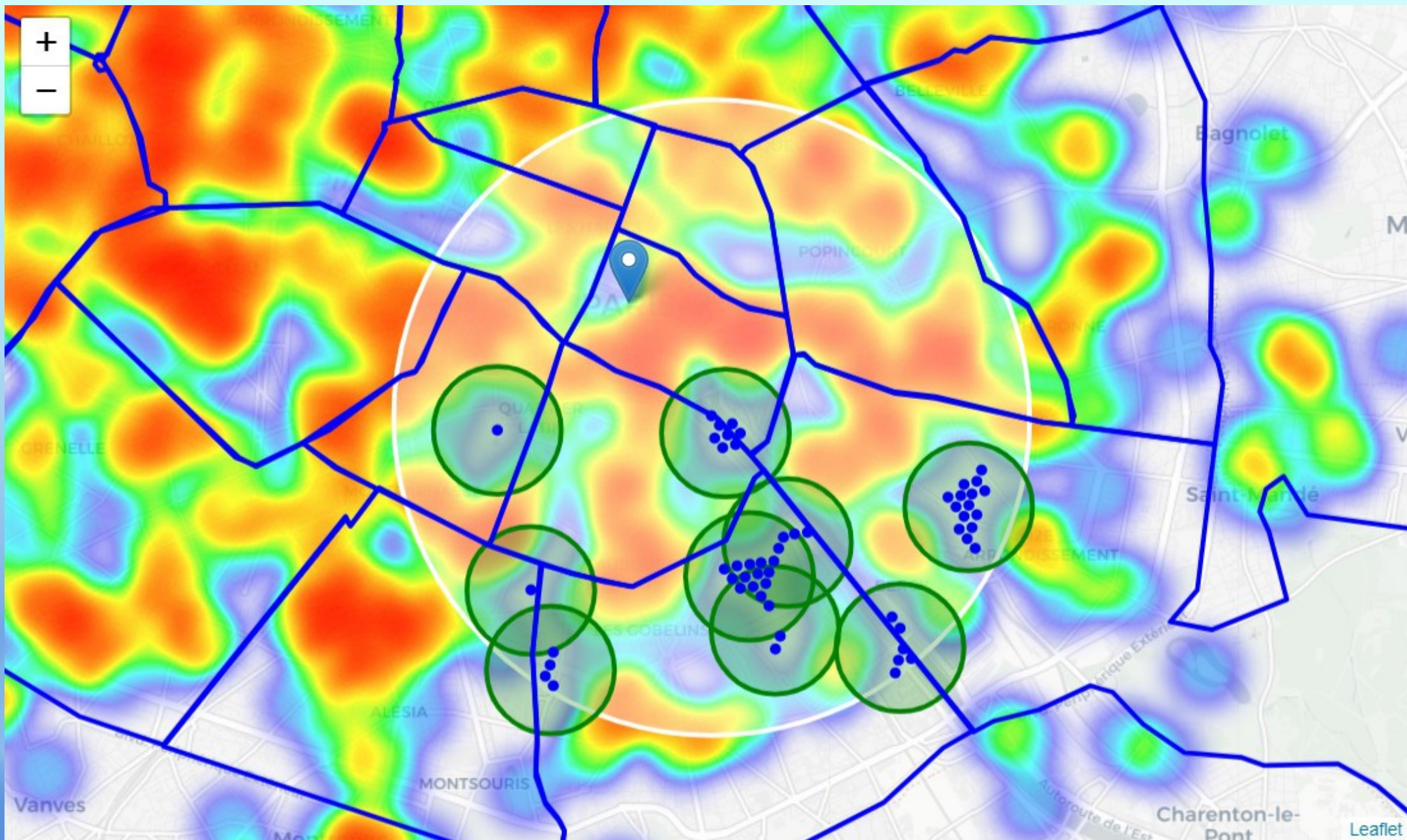
Results and Discussion

Our analysis shows that although there is a great number of restaurants in Paris, we found that the most common restaurant type if "French resaurant", so we decided to use this type of restaurants to start the new job in Paris.
Then we sarted to analyze the distribution of restaurants to find the best place to start the job. There are pockets of low restaurant density fairly close to city center. Highest concentration of restaurants was detected north and west from Paris center, so we focused our attention to areas south-east.

After directing our attention to this more narrow area of interest (covering approx. 5x5km south-east from Paris center) we first created a dense grid of location candidates (spaced 100m appart); those locations were then filtered so that those with French restaurant closer than 400m were removed.

Those location candidates were then clustered to create zones of interest which contain greatest number of location candidates.

Result of all this is 9 zones containing largest number of potential new restaurant locations based on number of and distance to existing venues - French restaurants particularly.

Conclusion

urpose of this project was to identify best type of jobs to start in Paris, thes to select areas close to center with low number of this type in order to aid stakeholders in narrowing down the search for optimal location for a new job.By analysing venues it was found that "French restaurant" is the most common venues in Paris. By calculating restaurant density distribution from Foursquare data we have first identified general boroughs that justify further analysis (south-east Paris center), 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 locations 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.