Data Science Capstone Coursera

Final project : My Italian restaurant grand opening in Paris

Context

ITALIAN RESTAURANTS ARE TURNING TO BE TRENDING LOCATIONS IN PARIS

THEY COMPETE TO PROPOSE THE MOST AUTHENTIC EXPERIENCE AT AFFORDABLE PRICES IN NICE DECORATED ROOMS

HOW COULD WE USE DATA TO FIND OUT WHICH LOCATION IN PARIS IS THE BEST TO START SUCH A BUSINESS ?

Requirements

The restaurant should be located in one of the **trendiest zones in Paris** to be sure to reach the targeted audience

It will propose a **"slow food"** experience

→ The restaurant will therefore open **on the evening only** and should be located in an area where there is much activity in the evening

However the restaurant should not be located in an area which is already saturated with other venues of the same type

Data

To map the surface of Paris, the official neighborhood classification of the city is adopted (80 neighborhoods)

To identify trendy areas in Paris, a clustering analysis is performed through the kmeans method

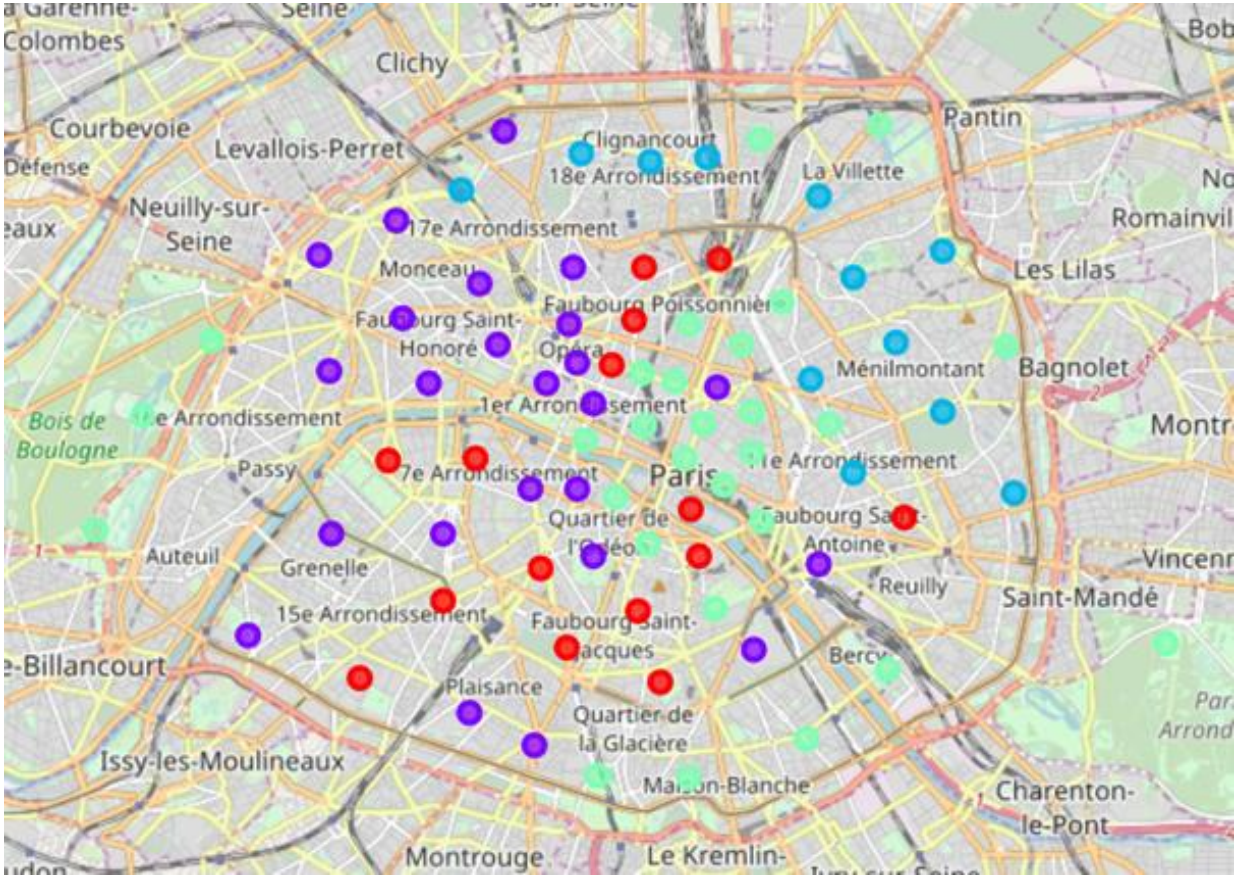
Venues are collected with automated requests to Foursquare for each neighborhood	Duplicated venues were identified, only the closest to its neighborhood center was kept	The contents of all cluster is analyzed and the one that matches the most the concept of the restaurant is kept

To identify trending areas at night, a Foursquare "trending" request was performed for each neighborhood

Only the 50% with the most activity were kept	**Among the final list, the neighborhood with the least Italian restaurants is chosen**

Results - clustering

The k-means was performed with the maximum value of k for which a cluster has at least 2 neighborhoods



4 clusters

Content :

- Cluster 1 : Touristic places
- Cluster 2 : Cultural places & business
- Cluster 3 : Trendy places – bars, international restaurants ...
- Cluster 4 : balanced categories

→ Cluster 3 fits the best for our application

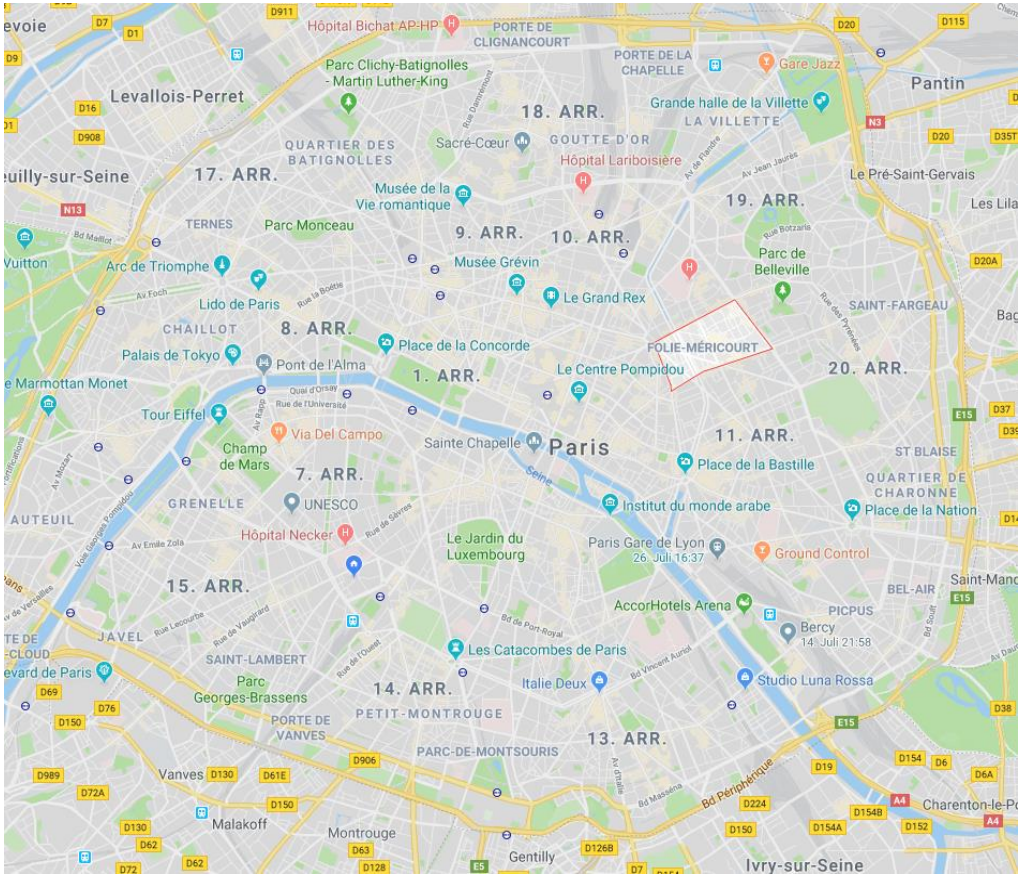
Results – Neighborhood selection

Trending analysis - Foursquare

- The median number of trending venues per neighborhood is 0
- All neighborhoods with at least one trending venue were kept

Neighborhood selection

- For all remaining neighborhoods, the one with the least occurrence of venues labelled as « Italian Restaurant » and « Pizza Place » was kept
- The winner is the neighborhood of **« Folie Méricourt »**



Limitations of the study

Mapping of Paris

- The neighborhood mapping is set in question as the the **neighborhoods centers are not equally spaced**
  - In the center of Paris the distance is **smaller**
  - In the periphery of Paris the distance is **larger**
- The radius was set to **1 km** for the Foursquare request → some venues were probably missed on peripheric neighborhoods → **source of error for the clustering**

FOURSQUARE

Poor number of trending venues

- Foursquare is not really popular in France
- It is mainly used by tourists
- Yet tourists is not the main target of the study

→ This could explain why Foursquare generally outputs very few trending venues

→ **This is also a source of error as we partly base our choice on this result**