Clustering Paris and Marseille districts: restaurant location

Wojciech Stach

January 5, 2020

Outline

- Background
- Problem Statement
- Data
- Methods
- Results
- Conclusions

Background

- Business location can have huge impact on profitability
  - Especially important for restaurants as customers often use proximity as a very important factor in making decisions where to eat
  - Knowledge of neighborhoods is essential to shape menu and service to meet clients' needs
- This project focuses on two largest cities in France: Paris and Marseille
  - 20 districts (arrondissements) in Paris
  - 16 districts (arrondissements) in Marseille

Problem Statement and Audience

- Problem
  - Perform districts analysis in Paris and Marseille to identify similar and dissimilar districts in terms of types of venues
- Audience
  - New/future business owners
    - to help finding a good district for their business
    - to choose the restaurant type and menu that best address the clients' needs
  - Existing business owners
    - To help choosing location when expanding or moving their business to a different location

Data

- The following datasets were used
  - https://public.opendatasoft.com/explore/dataset/arrondissements-millesimes0/table/
    - Data for 20 districts in Paris and 16 districts in Marseille
      - names
      - longitude and latitude
  - https://developer.foursquare.com/
    - community-based venues data

Methods

- K-means clustering
  - Paris districts
  - Marseille districts
  - Paris & Marseille districts together
- Elbow method to find optimal number of clusters for each case
- Mark clusters on maps
- Find common characteristics of districts from the same clusters

Results - Paris
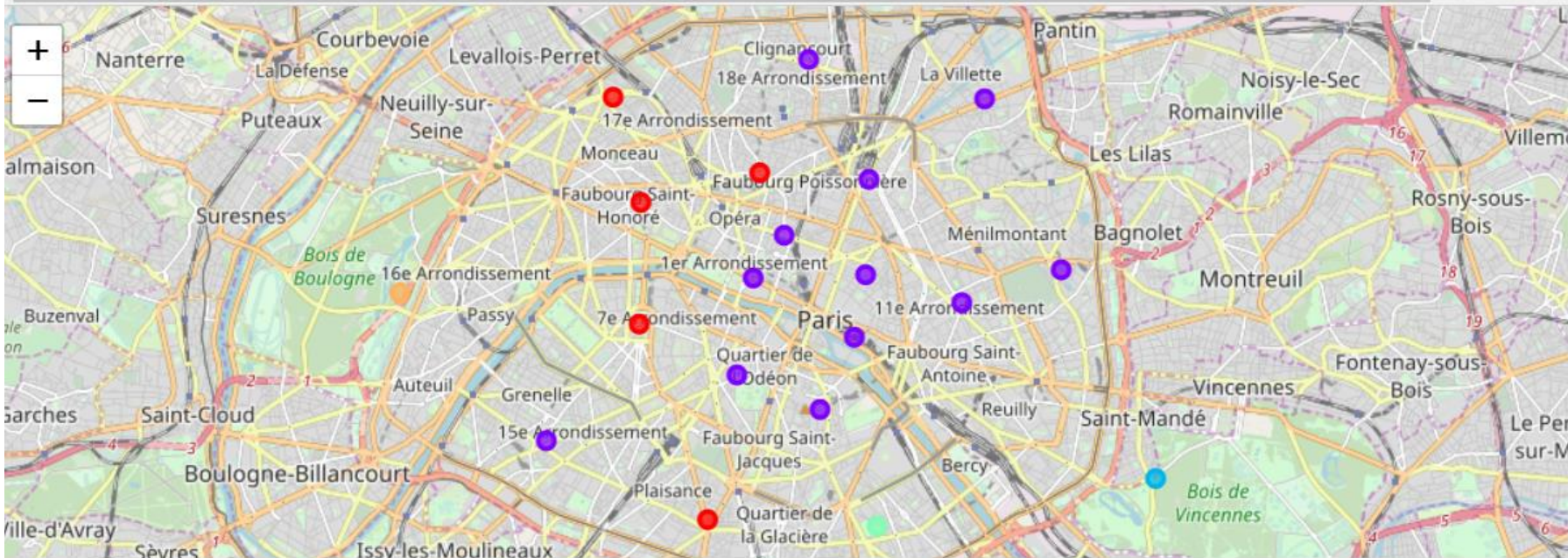
- Cluster 0 - red: Central-West districts of Paris (major venues: French restaurants and hotels)
- Cluster 1 - purple: Central-East districts of Paris and 15th district (that is in the West of Paris). Similar profile to Central-East district except for non-French restaurants showing quite high in terms of most popular venues
- Cluster 2 - blue: residential districts. Main venues include markets, stores, zoo
- Cluster 3 - green: multicultural districts, specifically including lots of Asian restaurants
- Cluster 4 - orange: upscale districts. Main venues include plaza, pool, lake.

Results - Marseille
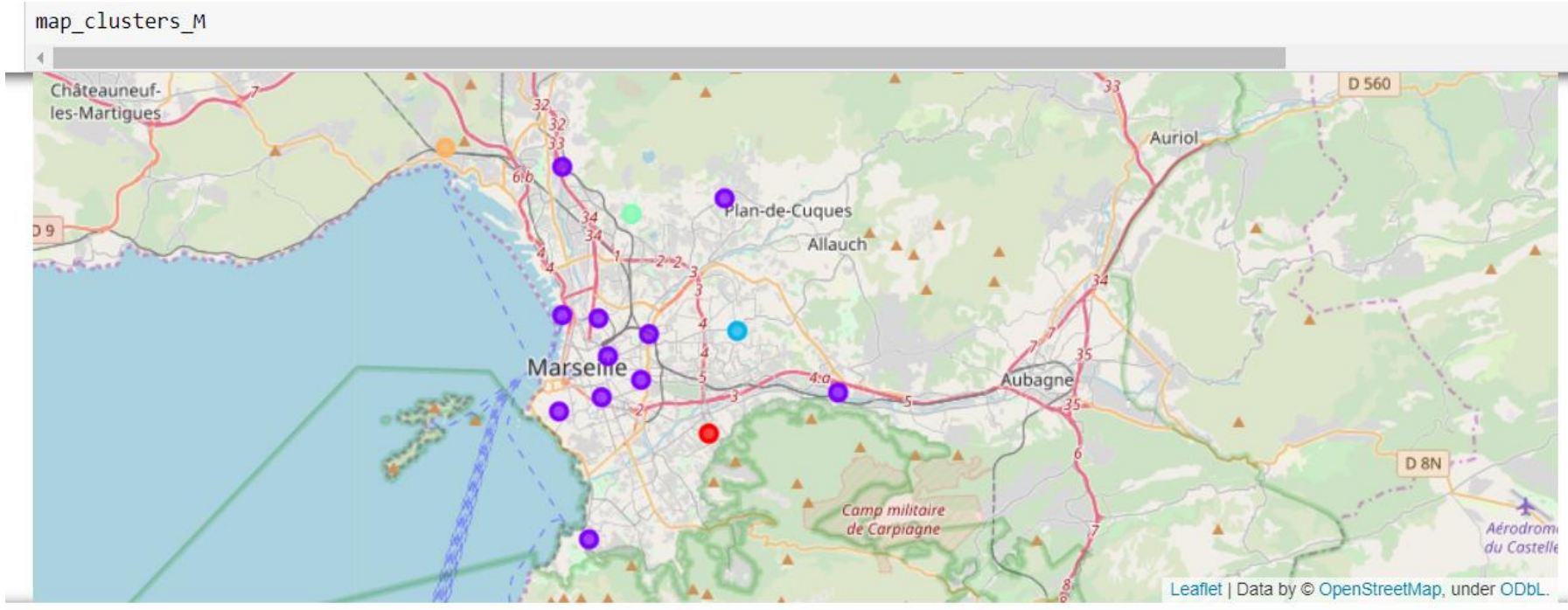
- Cluster 1 - purple: Center area of Marseille. Most popular venues include hotels and restaurants. Interestingly, districts 8th, 11th, 13th, and 15th that are further from the center of Marseille are in this cluster, too
- Clusters 0, 2, 3, 4 – other colors: Other venues are more popular, including shops tram stations, etc. More knowledge about the city is necessary to provide better distinctions between these clusters

Results – combined Paris (left) & Marseille (right)
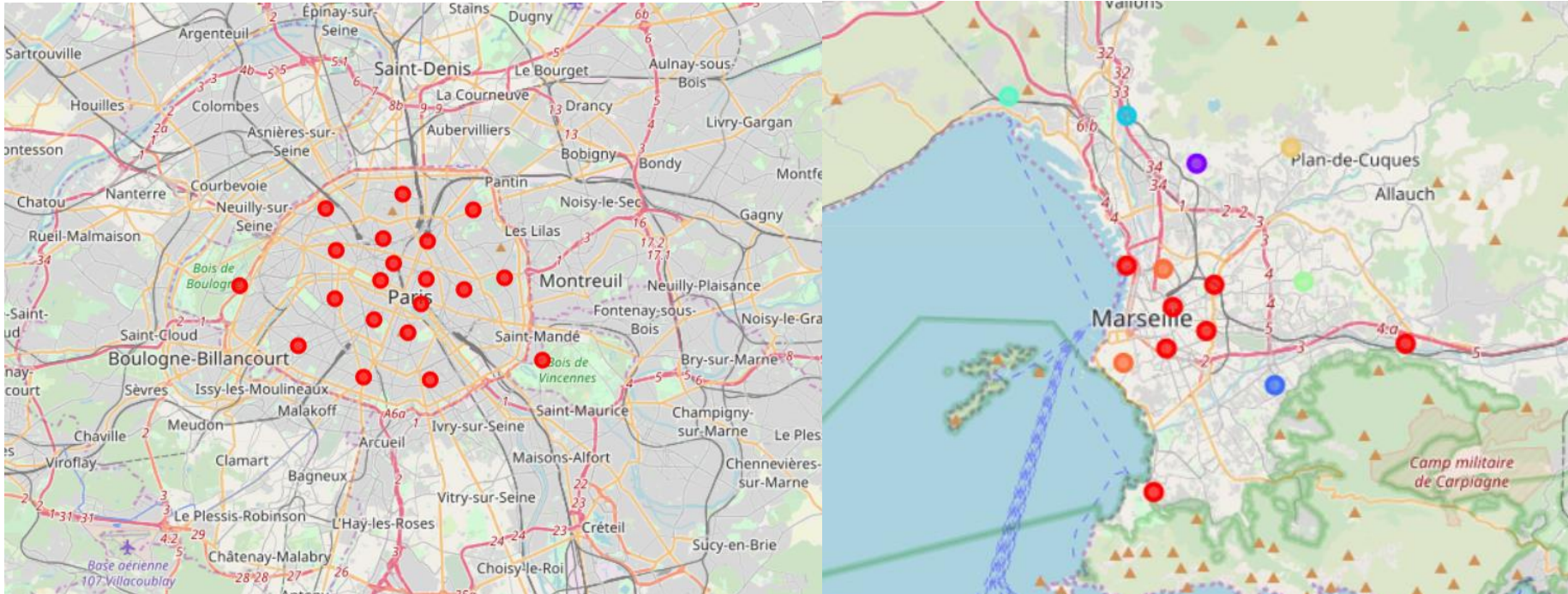
- Cluster 0 - red: Includes all districts from Paris and seven districts from Marseille (all of them from cluster 1 in Marseille clustering). This suggests that all Paris districts are quite like each other when compared to Marseille districts
- Clusters 1-7: Outskirts of Marseille. More knowledge about the city is necessary to provide better distinctions between these clusters

Conclusions and Future Work

- Clustering is capable of grouping districts based on multiple features
  - Paris clusters match intuitive classification of central districts vs. outskirts
    - Interestingly, the 15th district show more similarities to the central districts
  - Marseille clusters show similar clustering in terms of central districts
    - 8th, 11th, 13th, and 15th districts are more similar to the central districts despite being geographically closer to the outskirts
  - Paris and Marseille clusters show that Paris clusters are very similar to each other when compared with Marseille
  - It would be very beneficial to have a local expert to provide more insights and explanations of the clustering results
- The clustering analyses can be helpful in making decisions about choosing the location for a restaurant
  - Looking at the most popular venues and similar districts can provide interesting information
- Future work may include adding population, socio-economic as well as demographic data to provide more comprehensive analysis