

# Prague vs Paris- Data Science Project

Clustering neighborhoods of Prague and Paris

# General information

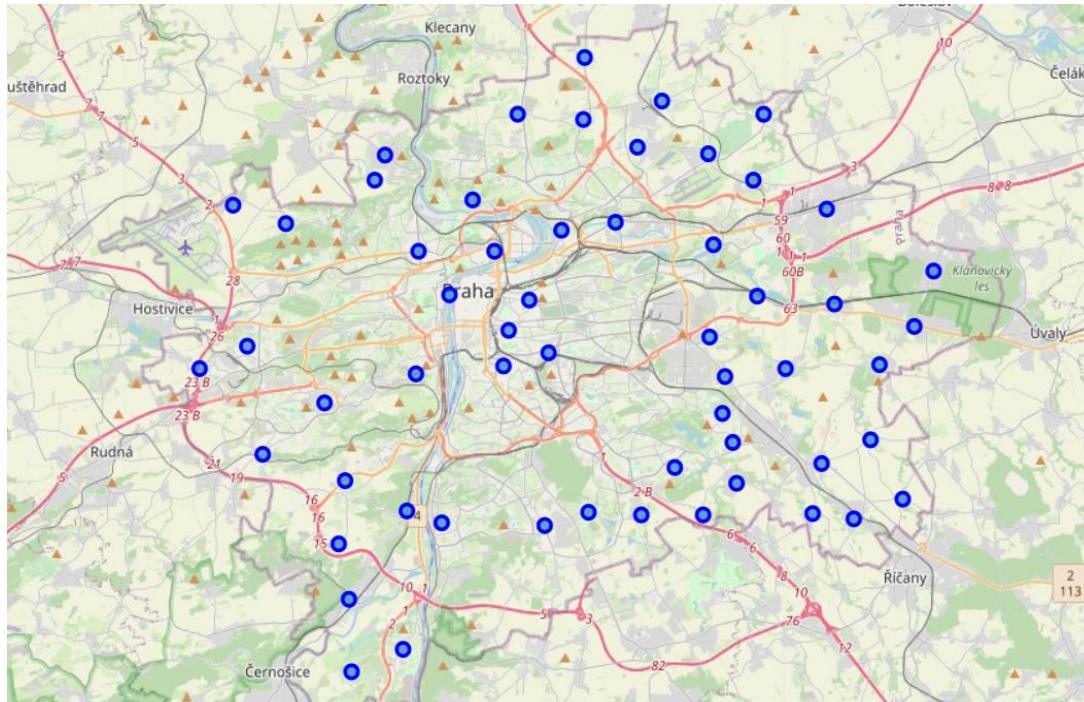
- Both cities are the capitals of their countries
- For this project I consider Paris as Paris Center and Paris Saint Deniz
- This project should help tourist, which of these two cities to visit, how long to stay there and where are the most interesting venues
- Helpful for people thinking about relocating
- Investors can find the best places to invest their money in these cities

# Datasets

- ▶ The Paris dataset was obtained here: <https://www.data.gouv.fr/fr/datasets/r/e88c6fda-1d09-42a0-a069-606d3259114e>
- ▶ The Prague Dataset was obtained here: <https://www.kaggle.com/konecfil/prague-neighborhoods-dataset>
- ▶ Then Foursquare API was used to get venues in the neighborhoods

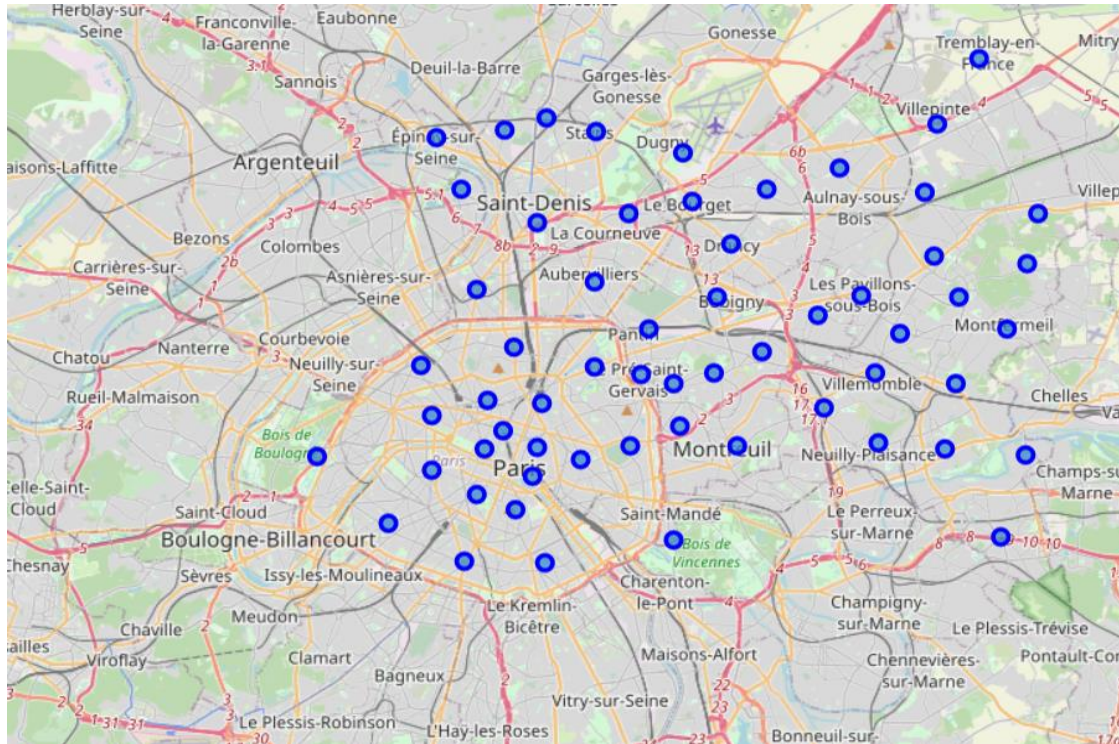
# Data Visualization

- Neighborhoods of Prague: There are 57 neighborhoods in Prague



# Data Visualization

- Neighborhoods of Paris: There are 60 neighborhoods in Paris



# Data Analysis

- ▶ There are 1142 venues in Prague neighborhoods and 230 unique venues
- ▶ There are 1508 events in Paris neighborhoods and 237 unique venues
- ▶ We want to know the frequency of occurrence for each venue in the neighborhoods-one-hot encoding

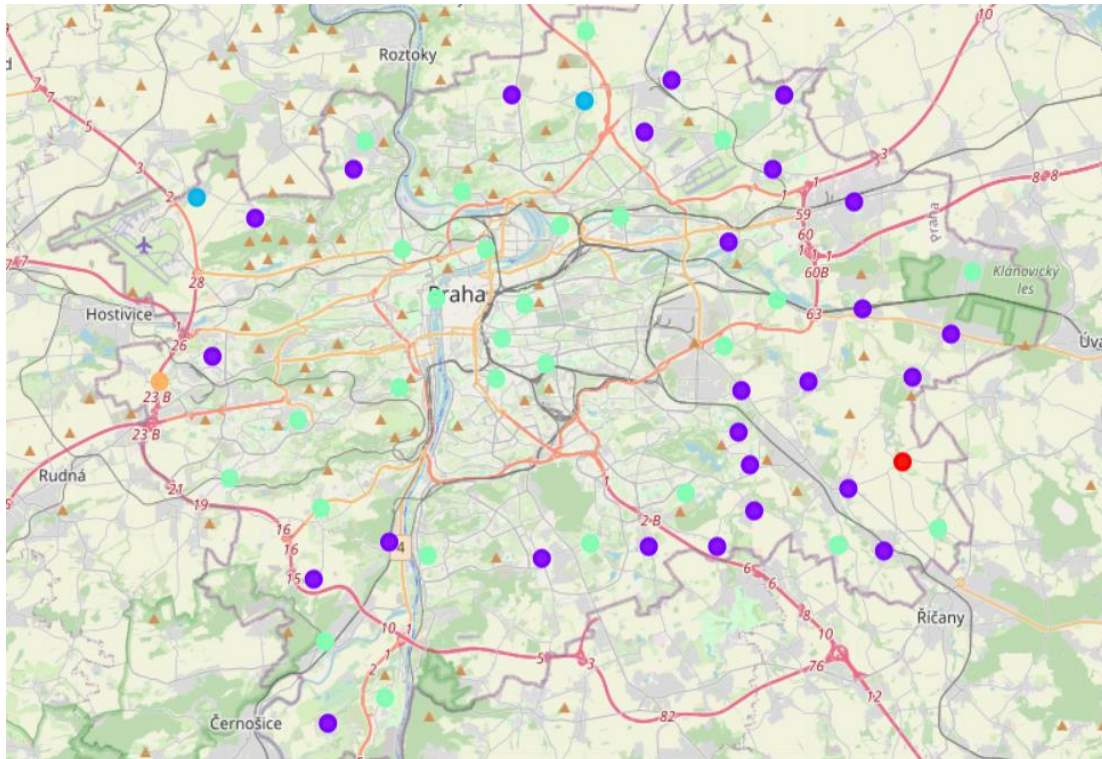
# Clustering Neighborhoods

- ▶ We will use k-means algorithm
- ▶ The number of clusters will be 5
- ▶ Clustering based on the frequency of occurrence for each venue in the neighborhoods



# Visualization Clusters- Prague

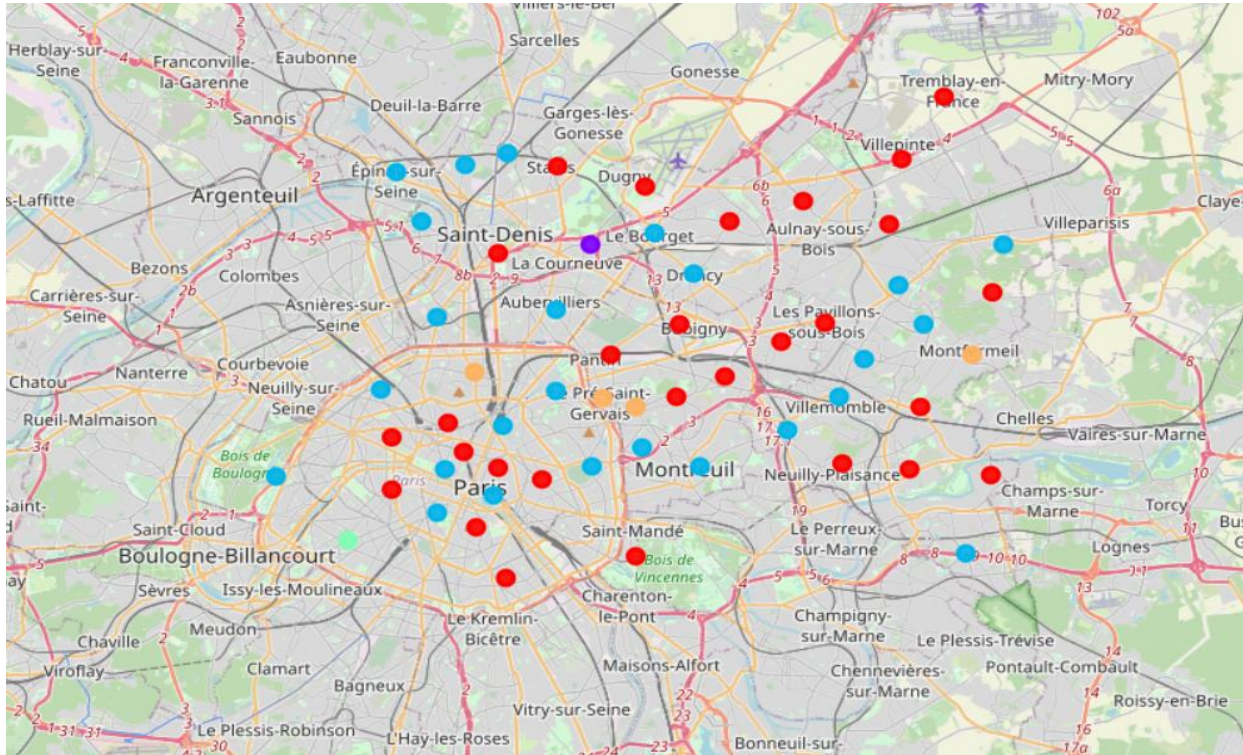
## ► Prague Clusters:





# Visualization Clusters- Paris

## ► Paris Clusters



# Discussion/ Conclusion

- ▶ Two main clusters are mixed. Then we have 3 very specific clusters.
- ▶ Clusters in Prague are very logical, the center of the city is one cluster, the suburbs is a second cluster. Then we have some very specific clusters.
- ▶ In both cities there are 2 main clusters and 3 very specific clusters.