Battle of Neighborhoods- Paris

Locations to setup New Restaurant in Paris

Introduction:

Paris is the capital and the most populous city of France and has a population of more than 2.1 million people. It is the second most expensive city in the world so the residents living closer to city center are likely to have higher income. And Paris is the most visited city in the world due to its famous heritage sites and museums. With it's diverse culture, comes diverse food items. There are many restaurants in Paris, each belonging to different categories like Chinese, Indian , French etc.

For this project I have selected the city of Paris to analyze its geospatial data to select a location to setup a new restaurant. The decision will have to be made using data driven approach where we will analyze the geospatial data and use the location of the existing restaurants, cafes, and bistros.

What type of clients or a group of people would be interested in this project?

1. Business personnel who wants to invest or open a restaurant. This analysis will be a comprehensive guide to start or expand restaurants targeting the large pool of office workers in Tokyo during lunch hours.

2. Freelancer who loves to have their own restaurant as a side business. This analysis will give an idea, how beneficial it is to open a restaurant and what are the pros and cons of this business.

3. New graduates, to find reasonable lunch/breakfast place.

4. Budding Data Scientists, who want to implement some of the most used Exploratory Data Analysis techniques to obtain necessary data, analyze it, and finally be able to tell a story out of it.

Data Acuisition and cleaning:

The data has been downloaded from https://opendata.paris.fr/explore/dataset/arrondissements/table/?dataChart and uploaded on the github account. And then again extracted it from my own github account.

I will use the following API for geolocation:

- Foursquare API: to find restaurant/venues

Cleaning: Cleaning this dataset involved changing the names of the columns for easy understanding and removing redundant columns to reduce the size of the dataset. French name of each neighborhood has also been included.

Exploratory data Analysis:

I have extracted the latitude and longitude of Paris. It will give us the exact location of Paris City Center. Then using this location, a map of Paris was created using Foursqaure API. Then different shops belonging to different categories were extracted and saved to use them further and visualize them:
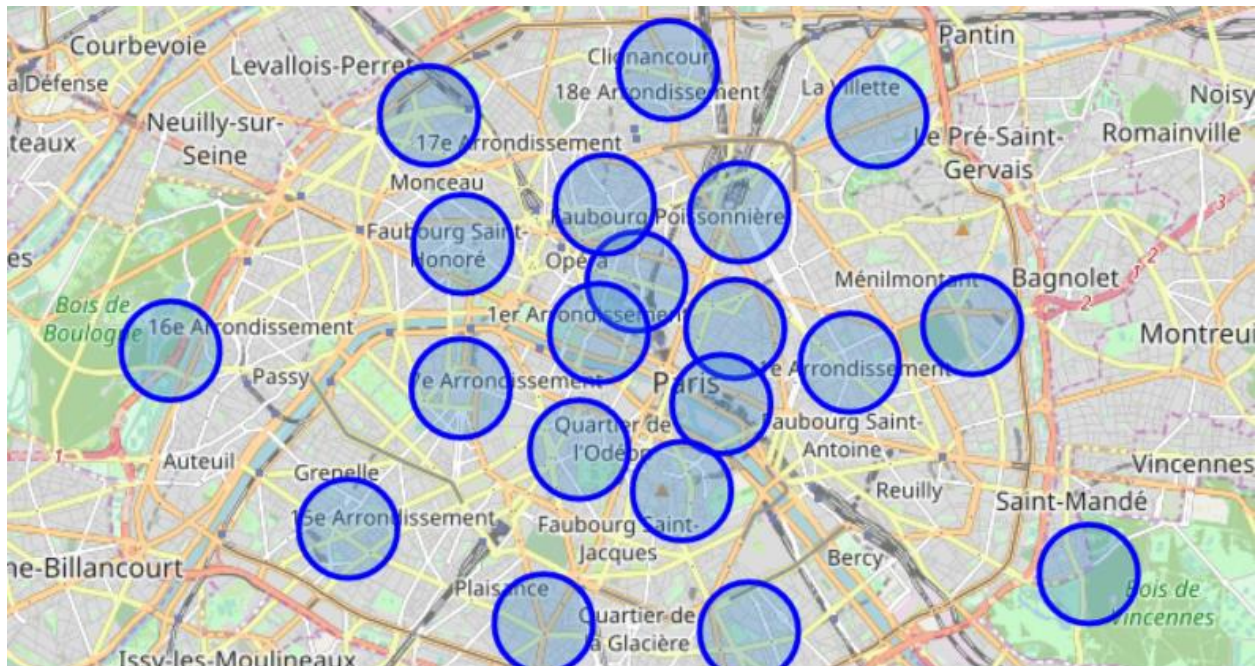
The data was extracted from Foursuare API which is geospatial data in which we will get information about different shops near Paris City Center. The frequency distribution of the shops in the city center looks as given below:
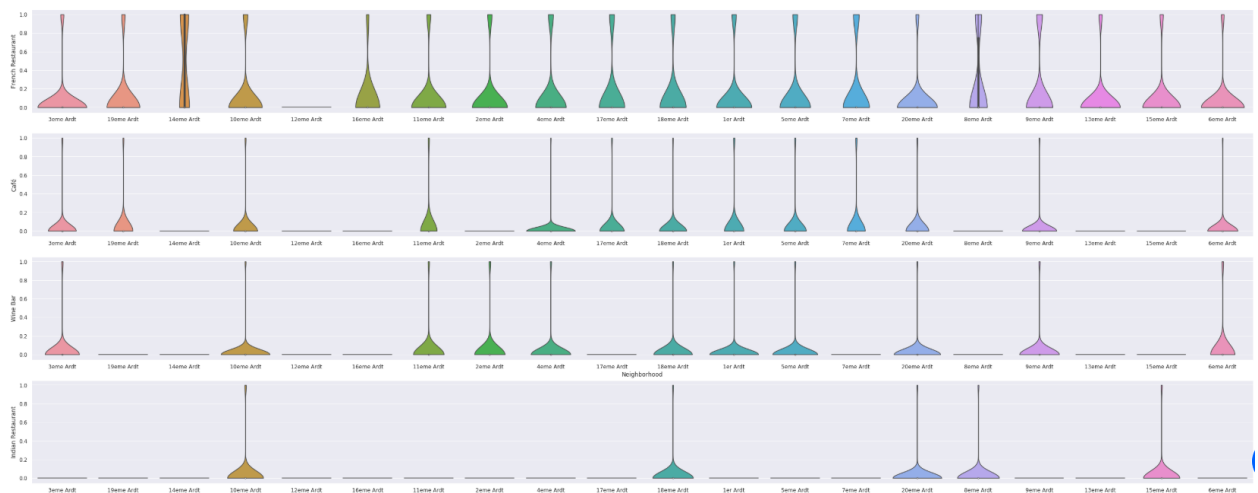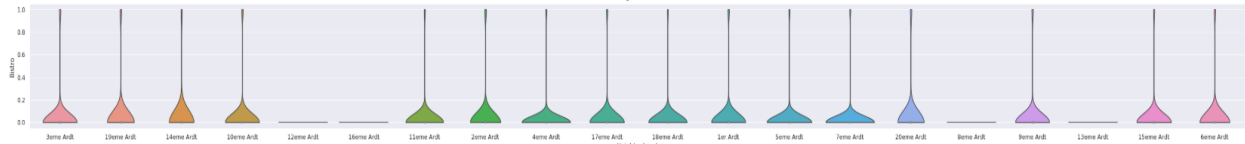
These frequency distributions show us that there 3 neighborhoods which have most number of eatries, which are Entrepot, Buttes-Montmartre and Menilmontant.

Methodology and Approach:

The project has been performed using IBM Watson using python. And used different libraries like pandas, numpy, K-means for modelling and for visualizing seaborn and matplotlib have been used.

- Collect the Paris data from the github link.
- Using FourSquare API we will find all venues for each neighborhood.
- Filter out all venues that are eatries or restaurants in paris.
- Find location, tips and like count for each Resturants using FourSquare API.
- Using location of neighborhood for each restaurant, we will sort that data.
- Visualize the Ranking of neighborhoods using folium library(python).
- The make blobs of different categories
- Verify it with k-means clustering

K-means clustering is the most used unsupervised machine learning algorithm for partitioning a given data set into a set of k groups (i.e. k clusters), where k represents the number of groups pre-specified by the analyst. It classifies objects in multiple groups (i.e., clusters), such that objects within the same cluster are as similar as possible (i.e., high intra-class similarity), whereas objects from different clusters are as dissimilar as possible (i.e., low inter-class similarity). In k-means clustering, each cluster is represented by its center (i.e, centroid) which corresponds to the mean of points assigned to the cluster.

K-means clustering is performed for 3 neighborhoods which have most number of restaurants, bars, etc..
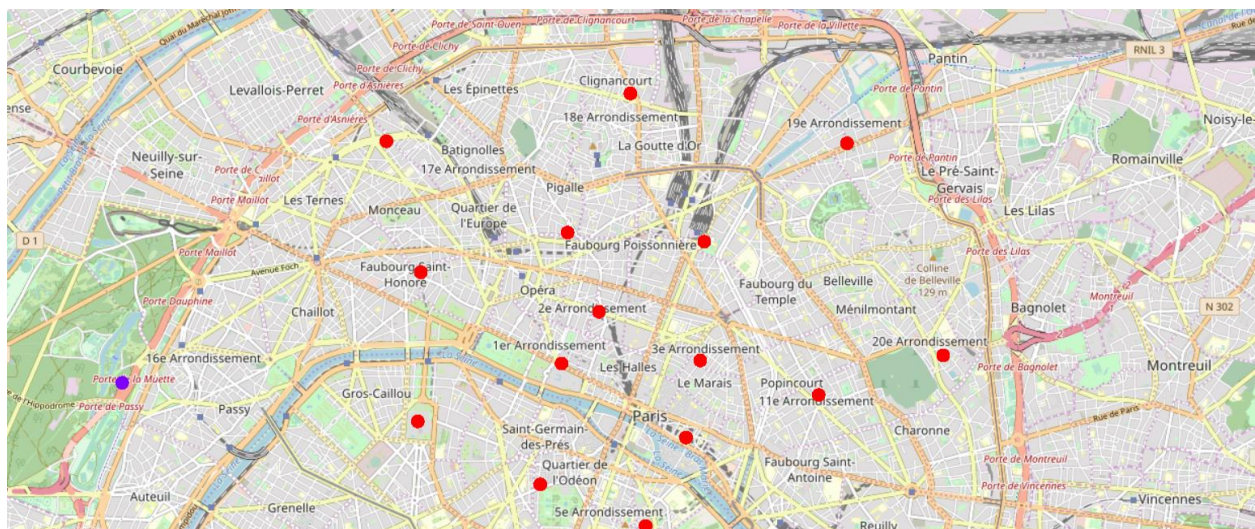
Conclusion:

As a result of this analysis, we were able to figure out 3 main regions or neighborhoods which are suitable for setting up a new restaurant as these places visited by tourists and locals. Further based of the analysis we can also decide which specific cuisine should be served by the restaurant so that to maximize profits.

Limitations:

In this analysis only the concentration of shops have been used as deciding factor considering that these places will be most visited by the customers. To go deeper we can further analyse and include more factors like reviews and likes given by the customers.

Future Use:

Like I mentioned in the Limitations more factors can be included in the analysis. And due to thses factors the result might come as different clusters.