Ingrid Nilsson

Coursera

Capstone Project – The Battle of Neighborhoods

Where is the best place for opening a French Restaurant in Paris?

Introduction

A French cook is looking to invest and start up a French Restaurant in Paris. The restaurant must be located centrally, and therefore the central areas of Paris and its most popular venues will be analyzed. The purpose of this project is to analyze venues in the central areas of Paris to come up with the best location of where to open a French Restaurant. The preferable neighborhood is an area where there is a shortage of French Restaurants. The most popular venues of each neighborhood will be analyzed. Each neighborhood will then be clustered to see where there is a shortage of French Restaurants. When a neighborhood has been found, the ratings of the other French Restaurant in the nearing will be examined in order to support the decision whether to open a restaurant there or not.

Data

The problem will be approached by using the following data:

- Json file of Paris Neighborhood data: This data contains borough, neighborhood/district, latitude and longitude. This data will be used to cluster neighborhoods that have similar venues.
- Python geocoder library: This data will be used to get geographical coordinates of neighborhoods
- Foursquare API: Will be used to get venues of every neighborhood in Paris

This data will used to analyze the neighborhoods and cluster them depending on their most popular venues. We will also be able to get ratings of restaurants and analyze the number of French Restaurants in each neighborhood.

Methodology

At first, the data was downloaded, and the right data was extracted and formed a data frame with the columns Borough, Neighborhood, Latitude and Longitude. The focus is on central areas of Paris and because of that all other boroughs was dropped, and only central Paris with its Neighborhoods was kept in a data frame.

In order to get the most popular venues for each neighborhood, the Foursquare API was used. For each neighborhood, the (max) 100 most popular venues were identified and added to a new data frame.

In order to get the popular frequency of each category for each neighborhood, dummy variables were used. The data frame with dummy variables was then grouped by neighborhood and the mean of each category was shown.

A data frame with the ten most common venues for each Neighborhood was created. Here we could already see that French Restaurants was the top 5 most common venues for many Neighborhoods.

The machine-learning method k-means clustering was used to cluster the neighborhood into five clusters. A cluster is a collection of data points aggregated together because of certain similarities. K-means clustering identifies k number of centroids and allocates every data point to the nearest cluster while keeping the cluster centroids as small as possible. This was done to see what neighborhoods that is similar to each other and what distinguishes different neighborhoods.

By using Foursquare API, the French Restaurants in a certain area could be analyzed and their rating extracted. A French Restaurant with a high rating would create hard competition.

To support the decision of locating a Restaurant in a certain neighborhood, the number of restaurants per inhabitant in the neighborhood was extracted. A low number of inhabitants per French Restaurant in our chosen neighborhood is a good result and indicates that there are room for more French Restaurants.

Results

The k-means clustering divided the neighborhoods into five clusters, and we could clearly see that neighborhood 13E was separated from the other clusters. 13E does not have French Restaurant as its most popular venue and there are possibilities of open a new French Restaurant here.

By investigating number of inhabitants per French Restaurant, the result showed that neighborhood 13E had a high number. This means that the competition of potential customers is lower in this area.

Discussion

Paris has a lot of French Restaurants for each Neighborhood. Three out of 20 Neighborhoods had French Restaurants as their 1st most common venue. The k-means clustering shows that Neighborhood 13E has French Restaurant as their second most popular venue, meaning there is a possibility to invest and start up a new French Restaurant. This area is not in the most central areas of Paris, meaning that the rent would be lower than investing in the central areas of Paris.

Looking closer at the restaurants in this area, we can see that the competing restaurants has yet not been rated or has a rather low rating. This indicates that the competition would be lower than if they had higher ratings.

Taking the population of each neighborhood and dividing this with the number of French Restaurants gives us the number thousands of people per French Restaurant for each Neighborhood. This shows us that Neighborhood 13E, 19E, 20E and 12E has a high number which could possibly indicate that the market of French Restaurants is not full in these areas.

Further analysis that could be made is to look at average income per person for each neighborhood which would give us an indication of how much each person would spend on venues and restaurant visits. Higher income per person in a neighborhood would possibly mean higher revenue for the restaurant, why this would be relevant to look more into. Further analysis could also include looking more into tourist patterns and general population habits to see where it would be most suitable to open a French Restaurant.

Based on the results and discussion, the recommendation is to invest in a French Restaurant in Neighborhood 13E.

Conclusion

The overall recommendation is to invest in a restaurant in area 13E, located in the southern parts of central Paris. The k-means clustering showed significant differences with other parts of Paris and number of people per French Restaurant is rather high in this area. However, further analysis is recommended. This could involve looking into average income per person for each neighborhood and investigate tourist data and patterns as Paris has a very high numbers of tourists visiting each year.