Coursera Capstone

IBM Applied Data Science Capstone

Opening a New Sushi Restaurant in Athens, Greece

By: Spyridon Milas

Introduction

Over the last few years the capital city of Greece, Athens has attracted many international kitchens with great response from the local customers as well as tourists. Dining areas developers are taking advantage of this trend, establishing brand new restaurants and hiring experienced chefs in order to correspond to this increasing demand. As a result, the location of a significant investment in a dining establishment is a serious consideration taking into account the competition that this industry is facing.

Business Problem

The objective of the capstone project is to analyze and afterwards select the best neighborhood locations in Athens, Greece to open a new sushi restaurant. Using Machine learning techniques like clustering, and Foursquare APIs with a combination of raw text data sets from open sources, this project aims to provide solution to the following question: If a food service developer would want to open a new Sushi Restaurant in Athens, which neighborhood would most likely provide the higher revenue income?

Data

To solve the business problem described above, we will make use of the following data:

- List of neighborhoods of Athens, capital of Greece. This will define the locational scope of the project.
- Latitude and longitude coordinates of these neighborhoods. These will be necessary to plot the city map and map the relevant venues.
- Venue data about sushi restaurants initially. Then venue data about every dining area within the most favorable cluster in order to end up to the resulting neighborhood with the fewest competition threat.

Sources of data

The Wikipedia page (https://en.wikipedia.org/wiki/Category:Neighbourhoods_in_Athens) contains a list of neighborhoods in Athens, with a total of 63 neighborhoods. Web scrapping techniques will be used in order to extract that data from the Wikipedia page, and with the assistance of Python requests and beautifulsoup packages will be converted to an editable data frame. Consequently, using Python Geocoder package we will get the latitude and longitude geographical coordinates of the neighborhoods in order to be used by the Foursquare API to get the nearby venues.

Using the Foursquare API, we will get the necessary data for the venues around our neighborhoods. Initially, we will isolate our data results for Sushi Restaurants in order to form our necessary clusters. Afterwards, venue data concerning dining areas in general will be used in order to locate the neighborhood with the fewest restaurants in the cluster, we determined that it is the most appropriate. In the next section, we will further analyze the steps that we followed in order to get to the final result and the machine learning techniques that will be used.

Methodology

Firstly, we need to acquire the data, we described above. Therefore, we initialize the process by gathering the list of neighborhoods of Athens from the Wikipedia page (https://en.wikipedia.org/wiki/Category:Neighbourhoods_in_Athens). Using web scraping technique and with the assistance of Python requests and beautifulsoup packages, we extract the list of neighborhoods data. Eventually, through the Geocoder packages we are mapping these neighborhoods with geographical coordinates.

After these initial steps, we will populate the data into a Pandas data frame and then visualize the neighborhoods in a map using Folium package. This visualization technique will also verify that the coordinates were map accordingly along the neighborhoods of Athens. Using the Foursquare APIs, we will get in JSON format venue names, categories and latitude, longitude values. With this data, we can see how many venues returned per neighborhood and examine the number of the unique venue categories. Afterwards, we will analyze each neighborhood by grouping the data frame by neighborhood and take the mean frequency of each venue category. The resulting data is prepared for our clustering Machine Learning technique of KMeans.

Since we are interested for sushi bars we will isolate the “Sushi Restaurant” category and create a new data frame consisting of only this relevant information. Eventually, we will perform clustering on the data using the above technique and result in the neighborhoods that present the highest and lowest occurrence of Sushi Restaurants. After isolating the neighborhood with absence of sushi bars, we repeat the initial steps in the neighborhood

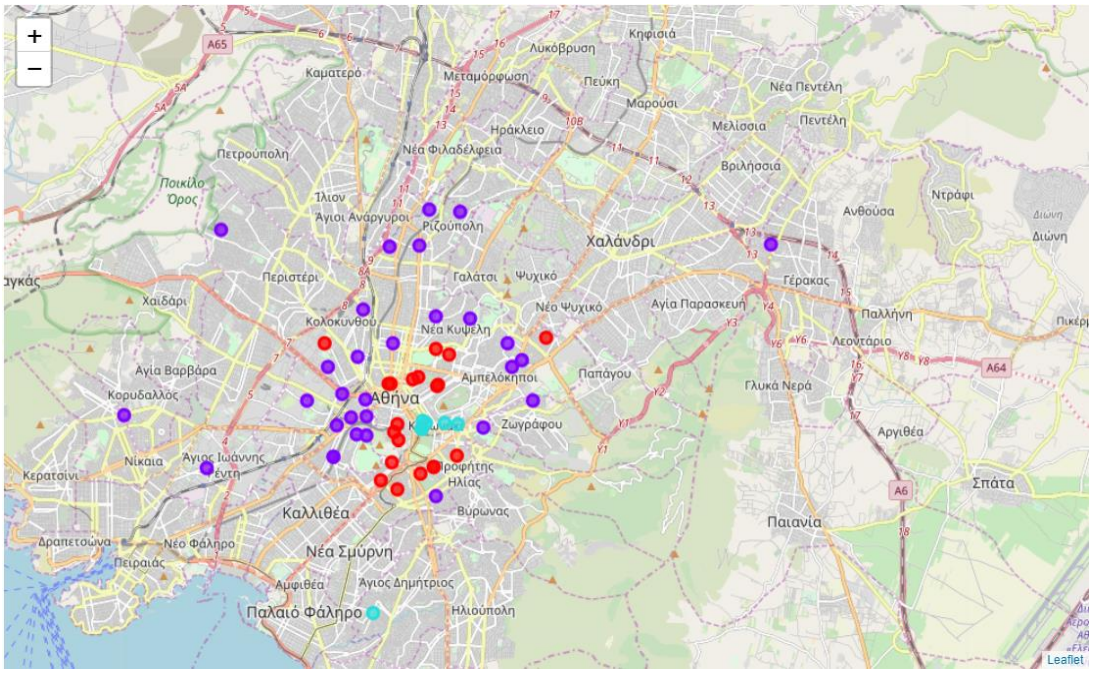
mentioned to determine which specific neighborhood has the fewest dining establishments. This will allow us to find a location that will strive in the absence of competition from supplemental food services.

Result

The results from the KMeans clustering show that we end up categorizing the neighborhoods into 3 clusters based on occurrence frequency of “Sushi Restaurants”:

- Cluster 0: Neighborhoods with moderate number of sushi bars visualized in red color.
- Cluster 1: Neighborhoods with absence of sushi bars visualized in purple color.
- Cluster 2: Neighborhoods with high concentration of sushi bars visualized in green color.

Below we can see the resulting folium map with the characterizations we described above:



Observations

As observed from the visualization of our clusters the clusters 0 and 2 that present medium to high occurrence, besides some exceptions reside in the central area of Athens. Some neighborhoods populated in cluster 1 with insignificant occurrence of Sushi Restaurants happen to be close to the city center as well but are not populated by sushi restaurants most likely because pure central areas were prioritized in the initial establishment of this foreign kitchen. Therefore, this project recommends that these areas will avoid the competition of nearby existing restaurants and thrive profitably. Property developers should avoid by all means the neighborhoods in cluster 2 with intense dining competition and consider an opening in cluster 0, only in occasions of unique selling positions. In the conclusion section of this report, we will end up with our final recommendation for the most favorable neighborhood for such an establishment.

Conclusion

In this project, we went through the process of identifying the business problem of evaluating the best location to open a new Sushi Restaurant. From our clustering technique, cluster 1 is populated by the preferred locations. Within that cluster, the neighborhood with the fewest dining establishments seems to be Gyzi. Eventually, our recommendation is, that an opening of a Sushi restaurant in this particular neighborhood would thrive as a result from the lack of competition as long as it meets the clients' expectations.

Limitations

This report would be furtherly enhanced if the factors of mean income and mean age of residents for each neighborhood could be taken into consideration. Sushi restaurants in neighborhoods with higher residents' salaries and younger aged adults would have higher traffic than their peers. Unfortunately, these data were not readily available.

References

Category: Neighborhoods in Athens. Wikipedia. Retrieved from

https://en.wikipedia.org/wiki/Category:Neighbourhoods_in_Athens

Foursquare Developers. Foursquare. Retrieved from

https://foursquare.com/developers/apps