

Finding the best place to open a restaurant in Attica, Greece

Konstantinos Anathreptakis

June, 2021

1. Introduction

1.1 Background

Attica is the largest region in Greece and the population is approximately above the 50% of the country. The region of Attica includes Megaris as part of the regional unit West Attica, and the Saronic Islands and Cythera, as well as the municipality of Troizinia on the Peloponnesian mainland, as the regional unit Islands. It is very difficult to define which location will have the best prospect to open a new restaurant using a traditional approach, as the area of Attica is extremely wide.

We will explore Latitude and Longitude data in order to find the best place to open a restaurant. We have to utilize foursquare location data based on the most frequent venues.

1.2 Problem

In this project we try to utilize Foursquare data in order to find locations that restaurants are the most frequently venues in that location.

Moreover, it is also important to answer the following question "Which neighborhood exhibits the same characteristics based on venues?" In order to answer the above mentioned question a clustering method is applied. The neighborhoods will be categorized into 5 clusters that have similar characteristics. This project aims to choose the location that is best for restaurants per cluster.

1.3 Interest

Many businesses in the food industry as well advertising companies would be very interested in the above stated problems. For instance, a franchise fast-food restaurant wants to know where

to open restaurants. The best strategy would be to open one restaurant in each unique cluster. Moreover, the knowledge of the location with the most frequently restaurants venues are very important for the business plan of a company in in the food industry.

2. Data acquisition and cleaning

2.1 Data sources

First of all, we need to find data for all the areas of Greece. Noted that Greece has 408 sub regions based on data from <https://simplemaps.com>. The data for this project was obtained from public web site such as Foursquare and simplemaps. The initial dataset was downloaded from (<https://simplemaps.com/data/gr-cities>). The initial dataset was in csv and then was imported to Jupiter through pandas.

2.1 Data cleaning

First all, the data combined into dataframe. The initial dataset had 4 columns (city, lat, lng and Region). We chose the region to be Attika and we renamed the columns as following (Neighborhood, Latitude, Longitude, Region,).

Afterwards we can easily run folium.Map in order to visualize a map with all sample locations in Attica. Finally, we request data with Foursquare API for all locations to get the top 50 venues per location within a radius of 500 meters. Noted that our dataset consist of 82 Neighborhoods.

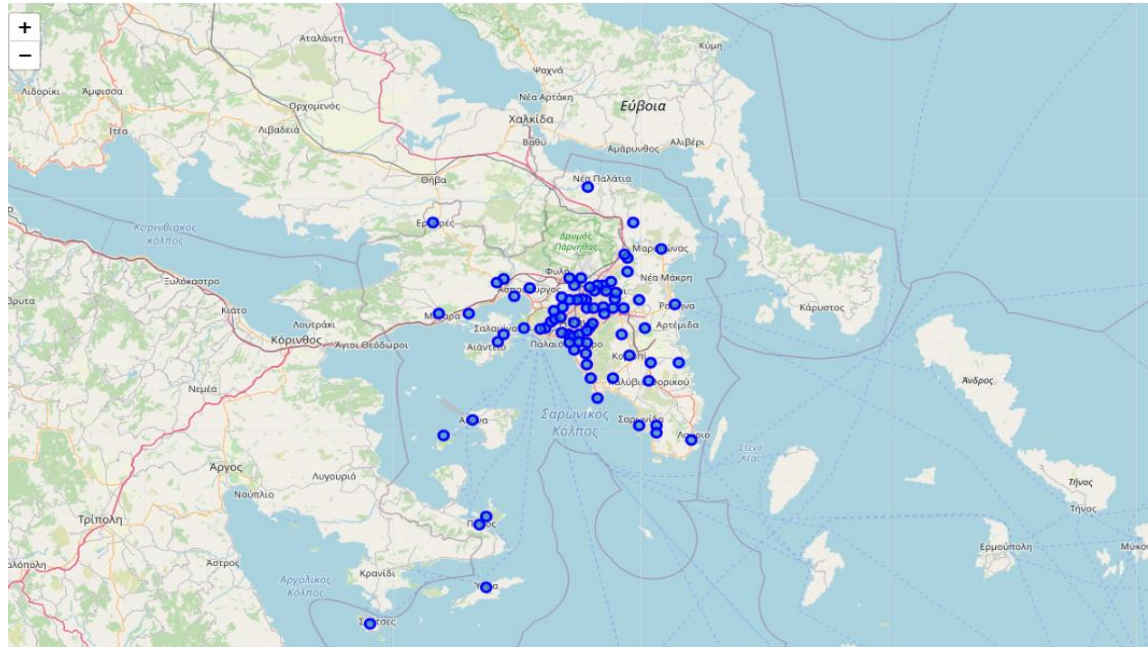
2.3 Feature selection

In this project we will apply an unsupervised machine learning algorithm in order to cluster the locations, so we only have features (not label data needed). Our features will be the venues, such as Restaurant, Coffee Shop, Wine Shop, Gym, etc. In this project the number of features was 197.

3. Exploratory Data Analysis

3.1 Visualization Map

Firstly, we get the geographical coordinates of Attica locations. The folium library is used to obtain the below picture. We can see that location are around 80 and are in a very wide area.



3.2 Venues per borough

Now we want to find the top venues per Neighborhood that are within a 500 meters radius. We create the appropriate url request and we get the result for each Neighborhood. In the below table we show the first five sun regions with the 10 most common venues based on frequency. We see that words restaurant, Taverna and Souvlaki are shown in top venues many times, so restaurant is a very frequent venue.

Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
Acharnes	Café	Mobile Phone Shop	Supermarket	Creperie	Souvlaki Shop	Taverna	Seafood Restaurant	Plaza	Cosmetics Shop	Dance Studio
Agia Paraskevi	Bakery	Pharmacy	Pizza Place	Clothing Store	Café	Plaza	Baby Store	Cosmetics Shop	Coffee Shop	Furniture / Home Store
Agia Varvara	Grocery Store	Fish Taverna	Café	Optical Shop	Greek Restaurant	Restaurant	Bakery	Bus Stop	Fast Food Restaurant	Betting Shop
Agkistri	Hotel	Nightclub	Hotel Bar	Cocktail Bar	Harbor / Marina	Café	Greek Restaurant	Women's Store	Farmers Market	Food & Drink Shop
Aigaleo	Café	Bar	Meze Restaurant	Coffee Shop	Burger Joint	Souvlaki Shop	Mobile Phone Shop	Donut Shop	Bakery	Snack Place

3.4 Number of restaurants in every Neighborhood (top 10)

Furthermore, we calculate the number of venues containing the words restaurant, Taverna and Souvlaki in each sub region. In order to calculate that number, we use pandas funtions such as str.contains and join search for the words aurant, Taverna and Souvlaki and then we sum the number of venues per Neighborhood. It is important to kwon in which locations have opened many restaurants, as it is an indication that the competition in food away from home industry is intense and that people frequent have lunch or dinner outdoors.

In the below table we see that the island of ydra, the sub regions Nea Filadelfeia and Argyroupoli are the areas that operating more restaurants compared to other locations. The above information is very useful, but it may lead us to biased conclusions as we have to take into account more variables in order to find the best candidate to open a new restaurant.

Neighborhood	Number_of_Restraaurants
ydra	21
Nea Filadelfeia	19
Argyroupoli	17
Kaisariani	15
Chalandri	15

Palaia Fokaia	14
Ilioupoli	14
Melissia	12
Aigina	12
alimos	12

3.3 Neighborhood contain restaurant in top 3 Most Common Venue venues

In the final dataset of all locations we try to find in which locations the words restaurant, Taverna and Souvlaki are containing in the top three most common venues. This will help us to find which Neighborhoods are the best places to open a restaurant as the high frequency of restaurant in that area, indicating that a new restaurant will have positive outlook and that people in that location frequently go to food industries.

Among all Neighborhood in our dataset, restaurant venues as the first most common venue are found for 12 sub regions, indicating that these locations are a very good candidate to open a new restaurant. We also observe that the most frequent venue among restaurant is the Greek Restaurant, so it would be a good choice to open a Greek Restaurant as it seems to be very attractive destination in the region of Attica.

Neighborhood	1st Most Common Venue
Dafni	Greek Restaurant
Elefsina	Cretan Restaurant
Kalyvia Thorikou	Taverna
Koropi	Meze Restaurant
Mandra	Taverna
Melissia	Greek Restaurant
Oropos	Greek Restaurant
Palaia Fokaia	Seafood Restaurant
Porto Rafti	Fish Taverna
Zografos	Greek Restaurant
agios Dimitrios	Souvlaki Shop

ydra	Greek Restaurant
------	------------------

In the search for the second most common venue, we found that 23 sub regions as shown in the bellow table. We conclude that Greek Restaurant and Souvlaki Shop are the most popular restaurants categories; consequently a new restaurant in order to be profitable and famous should be in that category of industry.

Neighborhood	2nd Most Common Venue
Agia Varvara	Fish Taverna
Aigina	Greek Restaurant
Chaidari	Greek Restaurant
Dafni	Meze Restaurant
Galatas	Greek Restaurant
Galatsi	Greek Restaurant
Gerakas	Grilled Meat Restaurant
Ilioupoli	Meze Restaurant
Kaisariani	Meze Restaurant
Kallithea	Souvlaki Shop
Kalyvia Thorikou	Greek Restaurant
Kifisia	Japanese Restaurant
Megara	Grilled Meat Restaurant
Moschato	Greek Restaurant
Nea Filadelfeia	Souvlaki Shop
Nea Peramos	Greek Restaurant
Paiania	Greek Restaurant
Palaia Fokaia	Greek Restaurant
Peristeri	Souvlaki Shop
Petroupoli	Greek Restaurant
Piraeus	Souvlaki Shop
Porto Rafti	Greek Restaurant
alimos	Greek Restaurant

Moreover, we are doing the same exercise to find the Neighborhoods that restaurants in the third most common venue in our sample. We found that in 15 sub regions the restaurant venue is the third most common (see table below).

Furthermore, in contract with the previous results, we see that types of restaurant are more diversified. For example in Aigaleo sub region the most common restaurant is Meze Restaurant, in Ilioupoli is Kebab Restaurant and in Kapandriti in Seafood Restaurant. So is very difficult to identify which kind of restaurant will be the best choice to open, based on our sample and assumptions.

Neighborhood	3rd Most Common Venue
Aigaleo	Meze Restaurant
Anavyssos	Greek Restaurant
Argyroupoli	Souvlaki Shop
Ilioupoli	Kebab Restaurant
Kalyvia Thorikou	Souvlaki Shop
Kapandriti	Seafood Restaurant
Kitsi	Greek Restaurant
Mandra	Souvlaki Shop
Metamorfosi	Greek Restaurant
Nea Filadelfeia	Meze Restaurant
Perama	Meze Restaurant
Saronida	Seafood Restaurant
Vyronas	Souvlaki Shop
Zefyri	Souvlaki Shop
agios Stefanos	Souvlaki Shop

4. Methodology

The methodology we will apply is Clustering. Clustering is a type of unsupervised machine learning that is used to organize data points into similar groups called clusters. A cluster contains a subset of observations from the data set such that all the observations within the same cluster are considered “similar.” The aim is to find a good clustering of the data meaning

that the observations inside each cluster are similar or close to each other and the observations in two different clusters are as far away from one another or are as dissimilar as possible.

Clustering algorithms are particularly useful in the many investment problems and applications in which the concept of similarity is important.

4.1 Clustering model (K-means)

K-means is a relatively old algorithm that repeatedly partitions observations into a fixed number, k , of non-overlapping clusters.

The number of clusters, k , is a model hyperparameter whose value must be set initially. In this project we assume that k is 5. The number of features for the Clustering algorithm will be 197. Each cluster is characterized by its centroid, and each observation is assigned by the algorithm to the cluster with the centroid to which that observation is closest.

5. Results

Through our analysis we are in position to answer important questions about venues frequency in each location in Attica Greece. We found that the restaurant, Taverna and Souvlaki are very common venues in most of our sample sub regions. Some potential good candidates to open a restaurant should be the island of ydra and the sub regions Nea Filadelfeia and Argyproupoli, but probably will not be the best choice as the competition is very high in that locations. In order to find the best location to open a new restaurant we should take into account the Neighborhoods that restaurants are the most frequently venues such as Melissia, Elefsina, Zografos as restaurants is the first most common venue.

Result of all locations in our sample this is around 80 sub regions is very large, so we now can have 5 clusters with approximately some characteristics. We found that the first two clusters consist of many Neighborhoods. While the clusters 3, 4 and 5 include only one sub region as the location is island. In the below table we show each location in its cluster.

Clusters

Cluster 1				Cluster 2			
Agia Paraskevi	Kalyvia Thorikou	Pefki	agios Dimitrios	Mobile Phone Shop	Kapandriti	Paiania	Galatas
Agia Varvara	Keratsini	Perama	agios Stefanos	Aigaleo	Korydallos	Palaio Faliro	Galatsi
Agkistri	Kifisia	Peristeri	alimos	Aigina	Kythira	Petroupoli	Irakleio
Argyroupoli	Kitsi	Piraeus	ano Liosia	Anavyssos	Lykovrysi	Vyronas	Moschatos
Athens	Koropi	Poros	anoixi	Aspropyrgos	Magoula	ydra	Nea Peramos
Chalandri	Mandra	Porto Rafti	ilion	Chaidari	Markopoulo	Nikaia	
Cholargos	Marathonas	Psychiko	Kallithea	Mobile Phone Shop	Kapandriti	Paiania	
Dafni	Marousi	Rafina	Palaia Fokaia				
Dionysos	Megara	Saronida	agioi Anargyroi				
Elefsina	Melissia	Sparta	Kaisariani				
Erythres	Metamorfosi	Voula	Ntrafi				
Gerakas	Nea Filadelfeia	Vouliagmeni	Zografos				
Glyfada	Nea Ionia	Vrilissia					
Ilioupoli	Nea Smyrni	Zefyri					

Cluster 3	Cluster 4	Cluster 5
Lavrio	Spetses	Oropos

6. Discussion further development

There is place for future development of this project, as we can add more variables such as economic development, demographic data and financial data about the potential customer in each sub region. Moreover, we can run the same analysis for more locations versus for each Neighborhood.

Noted that we chose to analyze sub regions of Attica and not the whole county of Greece, in order to deal with the limit of foursquare API calls volume.

7. Conclusions

Identify locations in which restaurants is the most frequently venue is a very important information for many stakeholders. It is not an easy problem as many factors can impact our decision. In this project we conclude which neighborhoods appear to be more favorable to open a new restaurant. Finally throughout the clustering algorithm we applied we can categorize all sub regions in Attica into 5 categories based of the most common venues. The clustering analysis is very useful as a franchise restaurant company can utilize these categories in order to open restaurants in all clusters.