

Capstone Project - Coursera

The Battle of neighborhoods

Ines Ben Ameer

May 15, 2019

Summary

1 Introduction

2 Explored Data

- Data specification
- Data samples

3 Methodology

- Prepare data
- Clustering

4 Results

- View Results
- Discuss Results

Introduction

Business-Problem

- The objective of this capstone project is to help an Italian entrepreneur to look for the most fitting location to open a new Italian restaurant in Athens, Greece.
- By using machine learning methods such as clustering and a location data platform like Foursquare, we will extract features and characteristics of neighborhoods in Greece, cluster them and finally determine the ideal place for opening the restaurant.

Business-Problem

- The objective of this capstone project is to help an Italian entrepreneur to look for the most fitting location to open a new Italian restaurant in Athens, Greece.
- By using machine learning methods such as clustering and a location data platform like Foursquare, we will extract features and characteristics of neighborhoods in Greece, cluster them and finally determine the ideal place for opening the restaurant.

Business-Problem

- In order to find this ideal location, we will try to detect locations that are not already crowded with Italian restaurants. We are also particularly interested in areas with attractions in vicinity (movie theatres, Amphitheater, etc).

Explored Data

Data specification

Data specification

- List of neighborhoods in Athens: It is necessary to collect the neighborhoods of the city of Athens to apply the clustering.
- Latitude and Longitude of the neighborhoods: We need the geolocation coordinates of each neighborhood in order to extract the venues located in each one.
- Venues: Finally we need to extract the list of Italian restaurants along with different types of attraction venues in each neighborhood to complete our data collection. For the first step of data collection, we will load all sort of venues provided by Foursquare, then we will filter the places with the types we desire.

Data specification

- List of neighborhoods in Athens: It is necessary to collect the neighborhoods of the city of Athens to apply the clustering.
- Latitude and Longitude of the neighborhoods: We need the geolocation coordinates of each neighborhood in order to extract the venues located in each one.
- Venues: Finally we need to extract the list of Italian restaurants along with different types of attraction venues in each neighborhood to complete our data collection. For the first step of data collection, we will load all sort of venues provided by Foursquare, then we will filter the places with the types we desire.

Data specification

- List of neighborhoods in Athens: It is necessary to collect the neighborhoods of the city of Athens to apply the clustering.
- Latitude and Longitude of the neighborhoods: We need the geolocation coordinates of each neighborhood in order to extract the venues located in each one.
- Venues: Finally we need to extract the list of Italian restaurants along with different types of attraction venues in each neighborhood to complete our data collection. For the first step of data collection, we will load all sort of venues provided by Foursquare, then we will filter the places with the types we desire.

Explored Data

Data samples

We filter the locations and keep only Italian restaurants and attractions:

Neigh	Neigh.Lat	Neigh.Long	Venue	V.Lat	V.Long	V.Cat
Aerides	36.147765	22.989737	Belvedere	36.148118	22.989718	Italian Restaurant
Akadimia	37.980285	23.734528	Piadina L' Umbro	37.979424	23.735630	Italian Restaurant
Akadimia	37.980285	23.734528	La Pasteria	37.978384	23.736132	Italian Restaurant
Akadimia	37.980285	23.734528	Frankie	37.980583	23.737988	Italian Restaurant
Akadimia	37.980285	23.734528	Il Postino	37.982145	23.736906	Italian Restaurant

Methodology

Prepare data

Steps of obtaining the data:

- Loading Neighborhoods from source page.
- Neighborhoods' coordinates: Get the geolocation coordinates of each neighborhood.
- Collect Venues: use Foursquare API to collect all the venues in each neighborhood.

Steps of obtaining the data:

- Loading Neighborhoods from source page.
- Neighborhoods' coordinates: Get the geolocation coordinates of each neighborhood.
- Collect Venues: use Foursquare API to collect all the venues in each neighborhood.

Steps of obtaining the data:

- Loading Neighborhoods from source page.
- Neighborhoods' coordinates: Get the geolocation coordinates of each neighborhood.
- Collect Venues: use Foursquare API to collect all the venues in each neighborhood.

Steps of obtaining the data:

- Filter Venues: As we mentioned in the previous sections, we're only going to need to count the number of italian restaurants per location. Therefore we should get rid of the unneeded venues and keep a list containing the following attractions:
 - Art Gallery
 - Concert Hall
 - Movie Theater
 - Music Venue
 - Museum
 - Pub
 - Performing Arts Venue

Steps of obtaining the data:

- Form a new list out of this one, containing the number of attractions per location. We get a table like this one:

	Neighborhood	Number of attractions
0	Akadimia	1.0
1	Akadimia Platonos	1.0
2	Anafiotika	2.0
3	Asteroskopeio	3.0
4	Ellinoroson	1.0

Steps of obtaining the data:

- Using the same procedure, we create a new column containing the number of Italian restaurants:

	Neighborhood	Number of Italian Restaurants
0	Aerides	1.0
1	Akadimia	1.0
2	Ellinoroson	2.0
3	Kallimarmaro	6.0
4	Kallimarmaro	6.0

Methodology

Clustering

Unsupervised learning: KMeans

The final part would be to apply Kmeans in order to cluster the neighborhoods to 3 groups based on the features we calculated:

- Number of attractions per neighborhood
- Number of italian restaurants per neighborhood

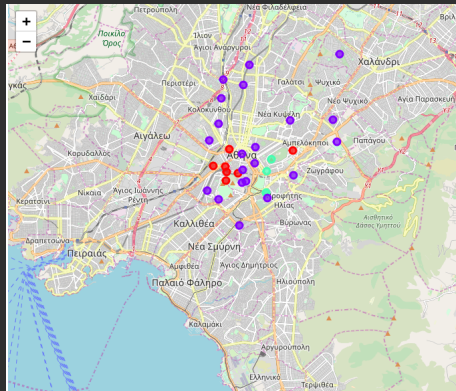
Results

View Results

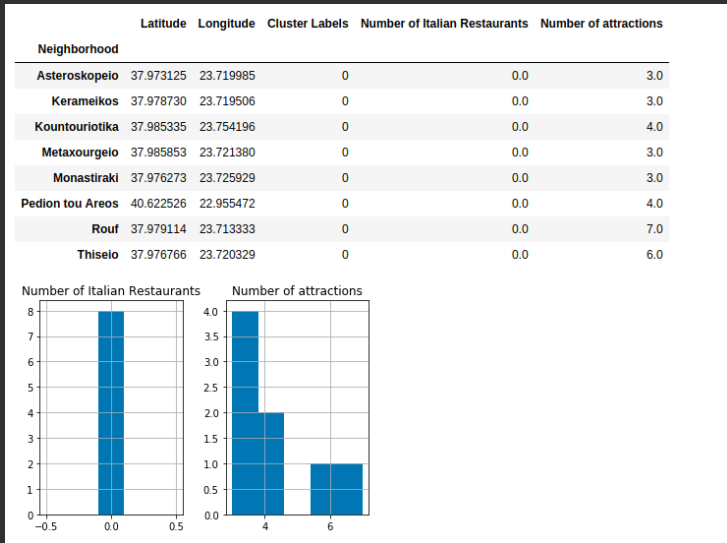
Partitioned data:

	Neighborhood	Latitude	Longitude	Cluster Labels	Number of Italian Restaurants	Number of attractions
0	Aerides	36.147765	22.989737	1	1.0	0.0
1	Agios Eleftherios	38.020044	23.731724	1	0.0	0.0
2	Agios Panteleimonas	37.607478	26.096458	1	0.0	0.0
3	Akadimia Platonos	37.989357	23.711217	1	0.0	1.0
4	Akadimia	37.980285	23.734528	1	1.0	1.0

Clusters on the map



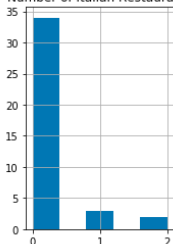
Cluster 1:



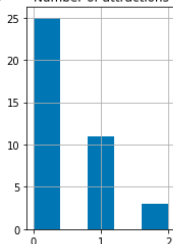
Cluster 2:

Neighborhood	Latitude	Longitude	Cluster Labels	Number of Italian Restaurants	Number of attractions
Aerides	36.147765	22.989737	1	1.0	0.0
Agios Eleftherios	38.020044	23.731724	1	0.0	0.0
Agios Panteleimonas	37.607478	26.096458	1	0.0	0.0
Akadimia Platonos	37.989357	23.711217	1	0.0	1.0
Akadimia	37.980285	23.734528	1	1.0	1.0

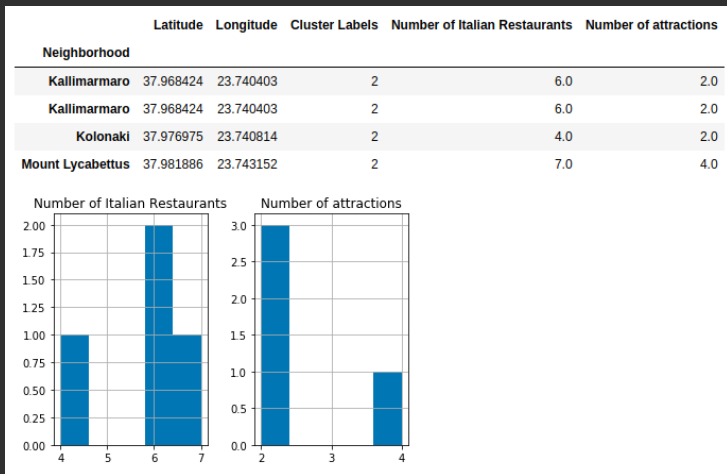
Number of Italian Restaurants



Number of attractions



Cluster 2:



Results

Discuss Results

Discuss Observations

Based on the graphs shown above, we can clearly notice that we obtained 3 types of neighborhoods:

- Cluster 1: This cluster is characterized by a low number of italian restaurants (0) and a high number of attractions
- Cluster 2: This group is not characterized by the presence of italian restaurants nor attractions.
- Cluster 3: Finally, we have a group of neighborhoods that has a high number of italian restaurants but not a very high number of attractions.

Obviously, our Italian entrepreneur should consider the neighborhoods within **cluster 1** to setup his business.