Classifying neighborhoods in Athens city center

by k-means clustering on the categories of parking stations

by

Panayiotis Tzitziris

May 2020

The present was created for "Applied Data Science Capstone" course which is part of "Applied Data Science Specialization" e-learning program provided by Coursera in cooperation with IBM.

1. Introduction

1.1 Description of the problem

In this project, we classify the neighborhoods in the city center of Athens by the method of k-means clustering depending on the number of parking stations as a criterion. The motivation to select this subject comes from the increasing demand for limited parking space in an urban area in which about 700k people reside and about 1.5m people work (City Population). Thus, we realize the high daily transport activity in the city center of Athens and its effects on the greater economy and the quality of daily life.

The aim of the project is to provide information through its results which can be useful in strategy planning and decision making for the Greek Ministry of Infrastructure and Transport, the authorities of the Municipality of Athens, and other entities and organizations which participate in the design of traffic networks and the management of relevant services.

1.2 Description of data

The data that we used for our analysis consist of the following:

- number of neighborhoods in the city center of Athens, which are obtained from Foursquare API (Foursquare). The sample includes 22 neighborhoods.
- categories of parking stations in the city center of Athens, which are obtained from Foursquare API (Foursquare).

In the following section, we make a k-means cluster analysis in order to classify neighborhoods in the city center of Athens based on the categories of parking stations in the same geographical area. All data process and statistical analysis for the purposes of the project were performed on a 64-bit Jupyter notebook by using Python 3.6 programming language and the code which was developed is available on Github repository (GitHub). Every stage of our analysis is described in detail in the next section.

2. Methodology

In this project, we classified neighborhoods in the city center of Athens by k-means clustering statistical method on the number and type of parking stations in these areas. Our methodology is divided in 6 steps, which are presented in detail below.

I. The first step includes the import of all the necessary Python libraries and packages for our analysis. The libraries and packages which were imported are *numpy* library for handling data in a vectorized manner, *pandas* library for the analysis of data contained in dataframes, *json* library for handling JSON files, *Nominatim* module from *geopy.geocoders* library for converting addresses into latitude and longitude coordinations, *requests* library to handle requests for urls, *json_normalize* module from

pandas.io.json library for transforming JSON files into pandas dataframe, *matplotlib.cm* and *matplotlib.colors* modules for creating plots, *KMeans* module from *sklearn.cluster* library for clustering, *make_blobs* module from *sklearn.datasets* library for visualization of neighborhoods, and *folium* library for rendering maps.

II. The second step includes the download of data with regard to the location of neighborhoods in the city center of Athens and the creation of this area's map. These data were available from Foursquare API by using a GET request url with the Category ID query set to 4f2a25ac4b909258e854f55f, the radius set to 1000m and the limit of returned results set to 30. 22 neighborhoods in the city Center of Athens were returned by Foursquare API and they are presented in Table 1 of Appendix I. In the fist column of Table 1 names of each neighborhood are displayed, while the second and the third column show the latitude and the longitude of each neighborhood, respectively. Based on these results, we created the map of Athens in which the blue markers show the location of each neighborhood, as shown in Figure 1 of Appendix II.

III. In the third step, we collect data from Foursquare API in order to find parking station in the city center of Athens. We set a radius of 1km from each neighborhood and the limit of returned results to 50. The Category ID which was inserted in GET request url was set to 4c38df4de52ce0d596b336e1. For the process, it was necessary to remove duplicate values of parking names when constructing the final dataframe which includes the parking stations, because some parking stations appeared to be in more than two neighborhoods. This problem was due to the radius size and the removal of duplicate values was necessary in order to minimize the probability of statistical errors that could affect the results of k-means clustering. We should mention that removal of duplicates lead also to removal of four neighborhoods (Metaxourgeio, Pedion Areos, Psyrri, Stathmos Larissis) as the parking stations which appeared in these neighborhoods were also present in other neighborhoods due to the extent of the radius from each neighborhood that was set. The numbers of parking stations in each neighborhood are shown in Table 2 of Appendix I. The number displayed in each row from the third to the last column of Table 2 represents the number of parking stations in the respective neighborhood.

IV. In the fourth step, we analyzed each neighborhood with respect to the number of parking stations in each neighborhood and their categories. The results of this step are shown in Table 3 of Appendix I. The numbers in Table 3 represent the mean frequency of occurrence of each parking category per neighborhood in the city center of Athens. Let us give some examples. According to the data in Table 3, in Akadimia 50% or ½ of the parking stations in this neighborhood is a hotel parking and the other is a parking in a building. In the neighborhood of Exarcheia, the only parking station is a common parking. In Kolonaki, the only parking station in this neighborhood is a parking of a college. In Neapoli, 2/6 parking stations or 33.33% of them are hotel parkings while the rest 4/6 or 66.67% of the parking stations in this neighborhood are common parkings.

V. The fifth step regards the classification of the neighborhoods in the city center of Athens by using the method of k-means clustering. We chose to set k = 4 clusters based on the fact that there are 4 main

categories of parking stations (Building, College Academic Building, Hotel, Parking) as it is shown in Table 3. The cluster in which each neighborhood is contained is shown in the fifth column named 'Cluster' in Table 4 of Appendix I.

VI. In the final step, we created a map in order to visualize the different clusters of neighborhoods in Athens city center. This visualization is represented by Figure 2 in Appendix II.

3. Results

The results which were obtained from the above method are summarized below:

- Cluster 0 represented by red color in Figure 2 of Appendix II contains neighborhoods in which there are only common parking stations.
- Cluster 1 represented by purple color in Figure 2 of Appendix II contains neighborhoods in which there are hotel and common parking stations.
- Cluster 2 represented by blue color in Figure 2 of Appendix II contains only the neighborhood of Kolonaki in which there is a college parking station.
- Cluster 3 represented by beige color in Figure 2 of Appendix II contains only the neighborhood of Akadimia in which there are one hotel parking station and one building parking station.

From the obtained clusters, we observe that Cluster 0 contains neighborhoods which are considered as residential and located at the outer center of the city while Cluster 1 contains neighborhoods which are considered as more tourist-attractive and located at the oldest central part of Athens and this may be the reason for the presence of hotel parkings in the neighborhoods of Cluster 1. Kolonaki and Akadimia which are contained in Clusters 2 and 3, respectively, can be considered as special cases, because these neighborhoods are described by both residential and touristic characteristics while there are also many places of academic and educational interest in these two neighborhoods.

4. Discussion

Two points of discussion are worthy to mention after our analysis. Both points are relevant to the perspective of reliability of the results. First, we ascertain that the obtained results align with the characteristics that describe each neighborhood in the city center of Athens. In other words, the clusters produced by k-means method based on the categories of parking stations are in agreement with the characteristics of the neighborhoods which are contained in each cluster. However, with regard to the second point of discussion, reliability of the results may be influenced by the fact that we have to increase the scale

of the radius which we set in the GET request url when we download parking stations data from Foursquare API in order to encompass as many parking stations near the neighborhoods as we can. The problem with the increase in the scale of the radius is that we obtained duplicate values of parking stations. As a consequence, the same parking station was possible to appear in different neighborhoods. Therefore, in order to minimize statistical errors produced by the above duplication, we had to remove duplicate values of parking stations as it was mentioned in the previous section with regard to description of methodology.

5. Conclusion

In this work, we analyzed the neighborhoods in the city center of Athens by using k-means clustering method on the categories of parking stations in this urban area. The motivation was the fact that the city of Athens is a region with high population residing or working in it, while the aggregate parking space in the city is limited. In addition, the city attracts many tourists every year due to the great number of sights to visit in Athens. Therefore, efficient strategies in the planning and management of transport and parking services in the city of Athens are needed.

All the data that we used in our analysis came from Foursquare API. As it was described above, 4 clusters of neighborhoods were obtained. Cluster 0 contains neighborhoods with residential characteristics, Cluster 1 contains neighborhoods characterized by high tourist interest, and Cluster 2 and 3 are two special cases. Based on the obtained results, we can claim that they are quite reliable with respect to the characteristics of each neighborhood. The only point of objection is that this reliability may be affected by setting high scale of the radius in the GET request url when downloading parking stations data from Foursquare API which leads to get duplicate values or more precisely the same parking station in more than one neighborhoods. Thus, we needed to pay attention to the downloaded values of parking stations and remove the respective duplicate values.

The results of the present project could be deployed in combination with other indicators or venues in the city of Athens, such as population, metro and bus stations, taxi stands etc., for further future research.

References

- City Population. GREECE: Greater Athens. https://www.citypopulation.de/en/greece/athens/
- Foursquare. Foursquare Developers. Explore Our Endpoints. https://developer.foursquare.com/docs/places-api/endpoints/
- GitHub. https://github.com/Panos1987/Athens_kmeans_clustering

APPENDIX I. TABLES

Table 1: Neighborhoods in the city center of Athens

	Neighborhood	NeighborhoodLat	NeighborhoodLong
0	Lofos Strefi (Λόφος Στρέφη)	37.987975	23.738004
1	Omonoia (Ομόνοια)	37.984194	23.728503
2	Kapnikareas Square (Πλατεία Καπνικαρέας)	37.976312	23.728692
3	Akadimia (Ακαδημία)	37.981125	23.733117
4	Metaxourgeio (Μεταξουργείο)	37.985889	23.722645
5	Psyrri (Ψυρρή)	37.978439	23.724986
6	Exarcheia (Εξάρχεια)	37.986144	23.735487
7	Monastiraki Square (Πλατεία Μοναστηρακίου)	37.976108	23.725814
8	Monastiraki (Μοναστηράκι)	37.976715	23.726306
9	Pefkakia Lykavittou (Πευκάκια Λυκαβηττού)	37.984003	23.739073
10	Syntagma Square (Πλατεία Συντάγματος)	37.975336	23.734982
11	Pedion Areos (Πεδίον Άρεως)	37.992227	23.732855
12	Anafiotika (Αναφιώτικα)	37.973430	23.727864
13	Syntagma (Σύνταγμα)	37.975358	23.734139
14	Stathmos Larisis (Σταθμός Λαρίσης)	37.992080	23.721047
15	Plaka (Πλάκα)	37.972474	23.730403
16	Thissio (Θησείο)	37.975496	23.720094
17	Neapoli (Νεάπολη)	37.986997	23.742700
18	Kolonaki (Κολωνάκι)	37.977631	23.741781
19	Victoria (Βικτώρια)	37.995774	23.730269
20	Kerameikos (Κεραμεικός)	37.979605	23.712732
21	Agios Panteleimonas (Άγιος Παντελεήμονας)	37.997704	23.729671

Table 2: Number of parking stations in each neighborhood of the city center of Athens

	Neighborhood	NeighborhoodLat	NeighborhoodLong	Parking	Parking Lat	Parking Long	Parking Category
0	Agios Panteleimonas (Άγιος Παντελεήμονας)	2	2	2	2	2	2
1	Akadimia (Ακαδημία)	2	2	2	2	2	2
2	Anafiotika (Αναφιώτικα)	4	4	4	4	4	4
3	Exarcheia (Εξάρχεια)	1	1	1	1	1	1
4	Kapnikareas Square (Πλατεία Καπνικαρέας)	1	1	1	1	1	1
5	Kerameikos (Κεραμεικός)	3	3	3	3	3	3
6	Kolonaki (Κολωνάκι)	1	1	1	1	1	1
7	Lofos Strefi (Λόφος Στρέφη)	1	1	1	1	1	1
8	Monastiraki (Μοναστηράκι)	2	2	2	2	2	2
9	Monastiraki Square (Πλατεία Μοναστηρακίου)	3	3	3	3	3	3
10	Neapoli (Νεάπολη)	6	6	6	6	6	6
11	Omonoia (Ομόνοια)	2	2	2	2	2	2
12	Pefkakia Lykavittou (Πευκάκια Λυκαβητού)	2	2	2	2	2	2
13	Plaka (Πλάκα)	1	1	1	1	1	1
14	Syntagma (Σύνταγμα)	2	2	2	2	2	2
15	Syntagma Square (Πλατεία Συντάγματος)	2	2	2	2	2	2
16	Victoria (Βικτώρια)	1	1	1	1	1	1

Table 3: Mean frequency of occurrence of each parking category per neighborhood in the city center of Athens

	Neighborhood	Building	College Academic Building	Hotel	Parking
0	Agios Panteleimonas (Άγιος Παντελεήμονας)	0.0	0.0	0.000000	1.000000
1	Akadimia (Ακαδημία)	0.5	0.0	0.500000	0.000000
2	Anafiotika (Αναφιώτικα)	0.0	0.0	0.000000	1.000000
3	Exarcheia (Εξάρχεια)	0.0	0.0	0.000000	1.000000
4	Kapnikareas Square (Πλατεία Καπνικαρέας)	0.0	0.0	0.000000	1.000000
5	Kerameikos (Κεραμεικός)	0.0	0.0	0.000000	1.000000
6	Kolonaki (Κολωνάκι)	0.0	1.0	0.000000	0.000000
7	Lofos Strefi (Λόφος Στρέφη)	0.0	0.0	0.000000	1.000000
8	Monastiraki (Μοναστηράκι)	0.0	0.0	0.500000	0.500000
9	Monastiraki Square (Πλατεία Μοναστηρακίου)	0.0	0.0	0.333333	0.666667
10	Neapoli (Νεάπολη)	0.0	0.0	0.333333	0.666667
11	Omonoia (Ομόνοια)	0.0	0.0	0.500000	0.500000
12	Pefkakia Lykavittou (Πευκάκια Λυκαβηττού)	0.0	0.0	0.000000	1.000000
13	Plaka (Πλάκα)	0.0	0.0	0.000000	1.000000
14	Syntagma (Σύνταγμα)	0.0	0.0	0.500000	0.500000
15	Syntagma Square (Πλατεία Συντάγματος)	0.0	0.0	0.500000	0.500000
16	Victoria (Βικτώρια)	0.0	0.0	0.000000	1.000000

Table 4: Mean frequency of occurrence of each parking category per neighborhood and clusters of neighborhoods in the city center of Athens

	Neighborhood	NeighborhoodLat	NeighborhoodLong	Cluster	Building	College Academic Building	Hotel	Parking
0	Lofos Strefi (Λόφος Στρέφη)	37.987975	23.738004	0	0.0	0.0	0.000000	1.000000
1	Omonoia (Ομόνοια)	37.984194	23.728503	1	0.0	0.0	0.500000	0.500000
2	Kapnikareas Square (Πλατεία Καπνικαρέας)	37.976312	23.728692	0	0.0	0.0	0.000000	1.000000
3	Akadimia (Ακαδημία)	37.981125	23.733117	3	0.5	0.0	0.500000	0.000000
6	Exarcheia (Εξάρχεια)	37.986144	23.735487	0	0.0	0.0	0.000000	1.000000
7	Monastiraki Square (Πλατεία Μοναστηρακίου)	37.976108	23.725814	1	0.0	0.0	0.333333	0.666667
8	Monastiraki (Μοναστηράκι)	37.976715	23.726306	1	0.0	0.0	0.500000	0.500000
9	Pefkakia Lykavittou (Πευκάκια Λυκαβητού)	37.984003	23.739073	0	0.0	0.0	0.000000	1.000000
10	Syntagma Square (Πλατεία Συντάγματος)	37.975336	23.734982	1	0.0	0.0	0.500000	0.500000
12	Anafiotika (Αναφιώτικα)	37.973430	23.727864	0	0.0	0.0	0.000000	1.000000
13	Syntagma (Σύνταγμα)	37.975358	23.734139	1	0.0	0.0	0.500000	0.500000
15	Plaka (Πλάκα)	37.972474	23.730403	0	0.0	0.0	0.000000	1.000000
17	Neapoli (Νεάπολη)	37.986997	23.742700	1	0.0	0.0	0.333333	0.666667
18	Kolonaki (Κολωνάκι)	37.977631	23.741781	2	0.0	1.0	0.000000	0.000000
19	Victoria (Βικτώρια)	37.995774	23.730269	0	0.0	0.0	0.000000	1.000000
20	Kerameikos (Κεραμεικός)	37.979605	23.712732	0	0.0	0.0	0.000000	1.000000
21	Agios Panteleimonas (Άγιος Παντελεήμονας)	37.997704	23.729671	0	0.0	0.0	0.000000	1.000000

APPENDIX II. FIGURES

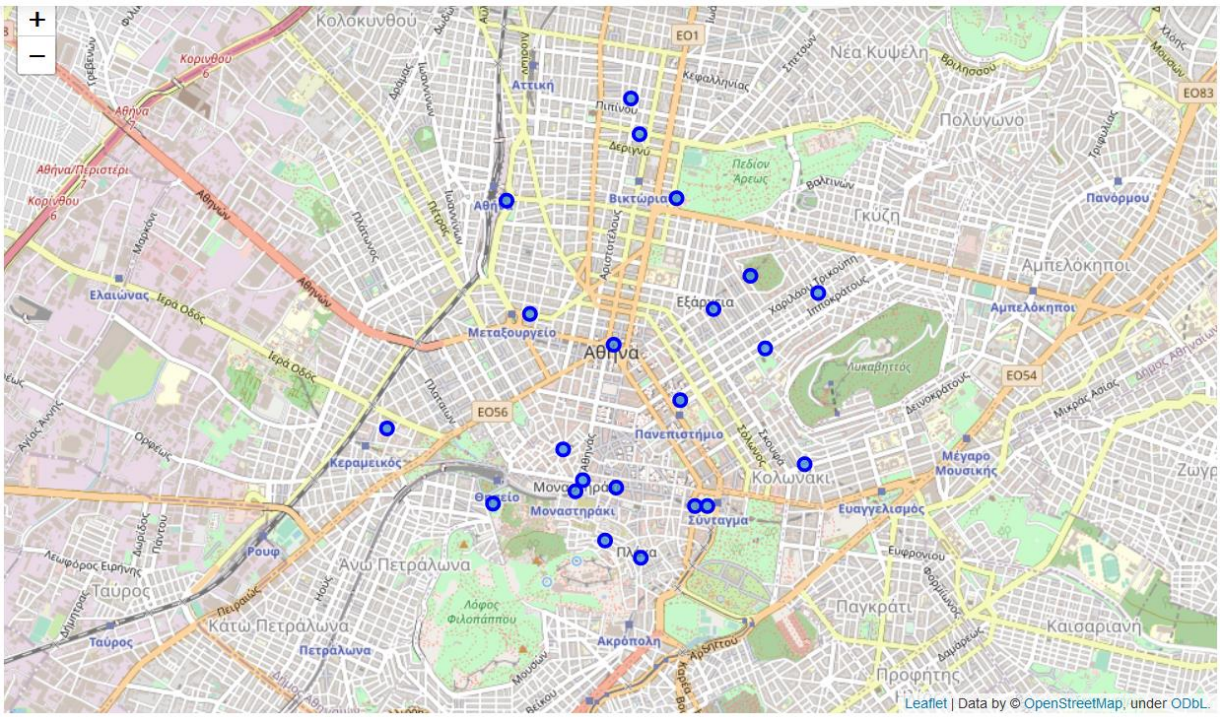

Figure 1: Map of the 22 neighborhoods in the city center of Athens
which were returned by Foursquare API

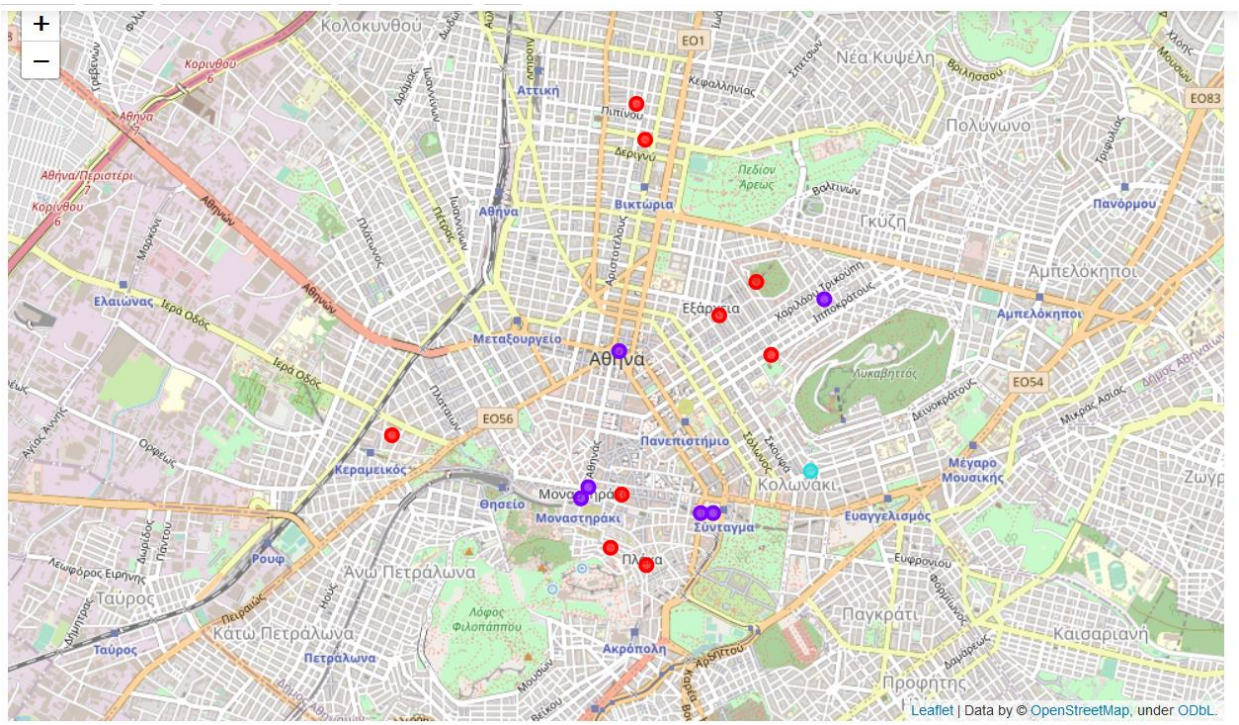

Figure 2: Map of the classified neighborhoods in the city center of Athens by the method of k-means clustering on the categories of parking stations in each neighborhood