

The Battle of Hotels in Athens

Evangelos Bezanis

2/21

Applied Data Science Capstone Week 5

1. Introduction

Athens, the capital of Greece, is a major touristic attraction and millions of tourists visit the city every year. For instance, in 2019 the number of tourists reached 5.5 Million, as many as the Athenian residents. Most of the visitors stay in the historic centre because there are many monuments (like Acropolis-Parthenon), museums, restaurants, bars and generally the centre offers a vibrant urban experience. Moreover, Athens is the European metropolis with the highest population density after Paris. Thus, the number of individuals visiting the centre leads to congestion making it hard even to stroll.

A guide for tourists or touristic agents will be presented. This guide depicts 100 major attractions and 50 hotels in the city centre. The hotels have been categorised in 3 clusters as hotels on the beaten path, off the beaten path and in between. The whole code was written in Jupiter Notebook with Python 3.

2. Data used

Initially, HTML table from Wikipedia was read [1] as dataframe. This table had information about the total and per km² population of every municipality in Greece. It was uploaded in Wikipedia by the local official authorities.

Secondly, GeoJSON file with boundaries of the Municipalities was read from the local official database [2].

Finally, from the Foursquare API data in JSON format was requested and converted to dataframe. These data include hotels-resort and venues to explore in Athens.

3. Methodology

The work done is separated in two main parts. The first concludes plotting a choropleth map of the 5 (in chapter is explained why not all) municipalities in Athens with the highest population density. The second part includes the illustration of hotels-venues and their clustering.

In the first part, the dataframe downloaded from Wikipedia, was cleaned. The numbers had additional characters for better visualization. For instance, the number 1000 was written as 1.000 and python stores

it as 1. Also, the JSON file, with the boundaries, was processed to include only the 5 municipalities with the highest population density.

In the second part, 3 API calls were made. The first two for hotels and resort around Athens. This 2 dataframes were connected and cleaned as usual. Thereafter, the last API call to explore venues around Athens was made and the data were cleaned as usual.

Finally, a dataframe with the distance that each hotel has from each venue was made. This dataframe have the real distance in meters (processed as lists), was trained to get the 3 clusters with the K-Means algorithm and is depicted in **Figure 1**.

	hotel name	National Theatre of Greece (Εθνικό Θέατρο)	Taf Coffee	Mr Bean Coffee Brewers	Πρωτοπορία	Της Θεάτρου Το Στέκι	Καραμανλίδικα του Φάνη	Μιράν	mama tierra	Anāna	I Cake You	2 γουλιές & 2 μπουκιές	The Handlebar	Polite Bookstore (Βιβλιοπωλείο Πολίτη)
0	Hapimag Resort Athens	0.238965	0.197914	0.200139	0.220952	0.177631	0.167101	0.165903	0.213374	0.130848	0.223643	0.129971	0.150540	0.1717
1	Polis Grand Hotel	0.059097	0.038234	0.038379	0.039610	0.102286	0.107356	0.106961	0.038927	0.109278	0.036954	0.105974	0.120065	0.0949
2	Titania Hotel	0.078316	0.009089	0.013603	0.041119	0.083383	0.083999	0.083010	0.032771	0.068568	0.042207	0.062491	0.091283	0.0600
3	Ambrosia Hotel & Suites	0.020583	0.060606	0.064317	0.076348	0.079613	0.087532	0.087681	0.073410	0.106915	0.074183	0.108373	0.103438	0.1228

Figure 1: Dataframe for clustering

4. Results

For the first the part, the choropleth is illustrated in **Figure 2**. For the second part the venues (in green) and the hotels (in blue) are placed on the map showed in **Figure 3**. In **Figure 4** the three cluster are depicted, where the “In between” hotels are in red, the “Off the beaten path” hotels in purple and the “On the beaten path” in light green. In **Figure 5** the list of hotels in each cluster is given.

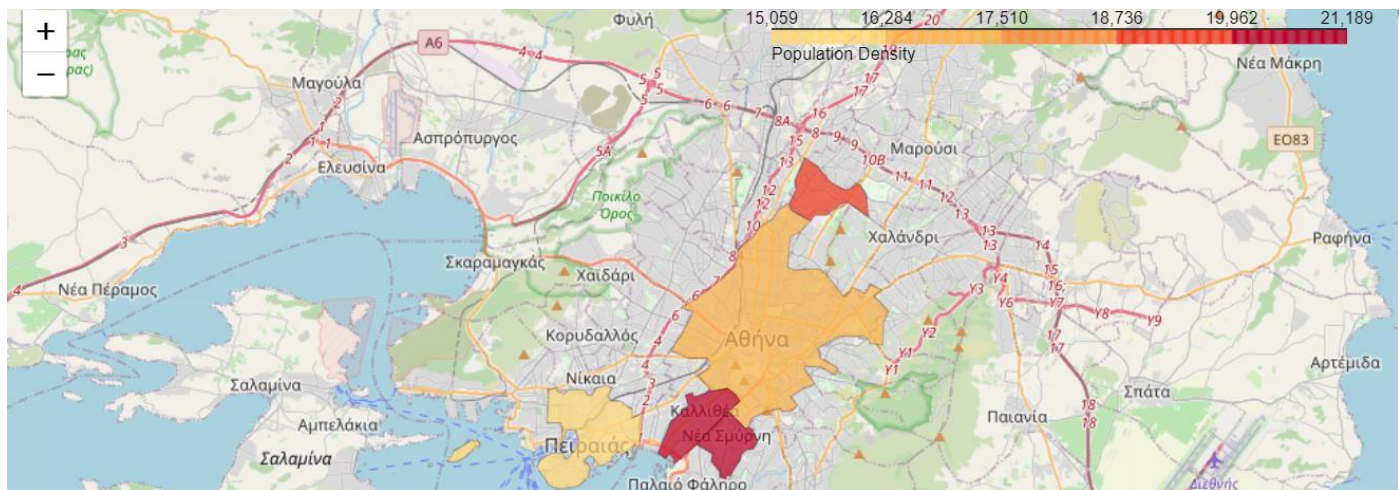


Figure 2: The five most dense municipalities (population/km²)

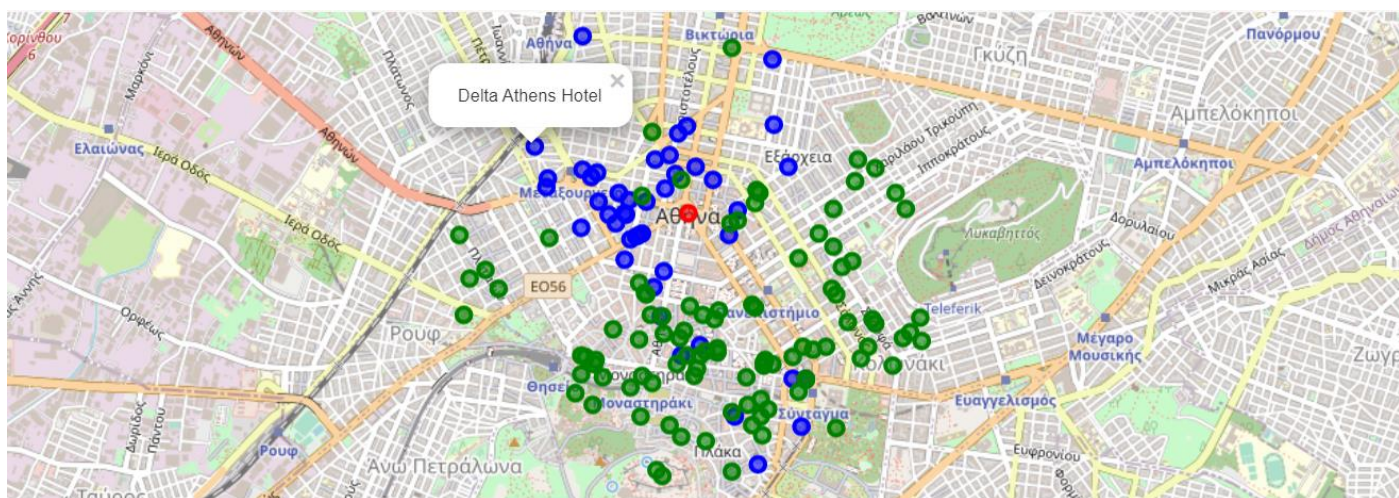


Figure 3: Venues (in green) and Hotels (in blue)

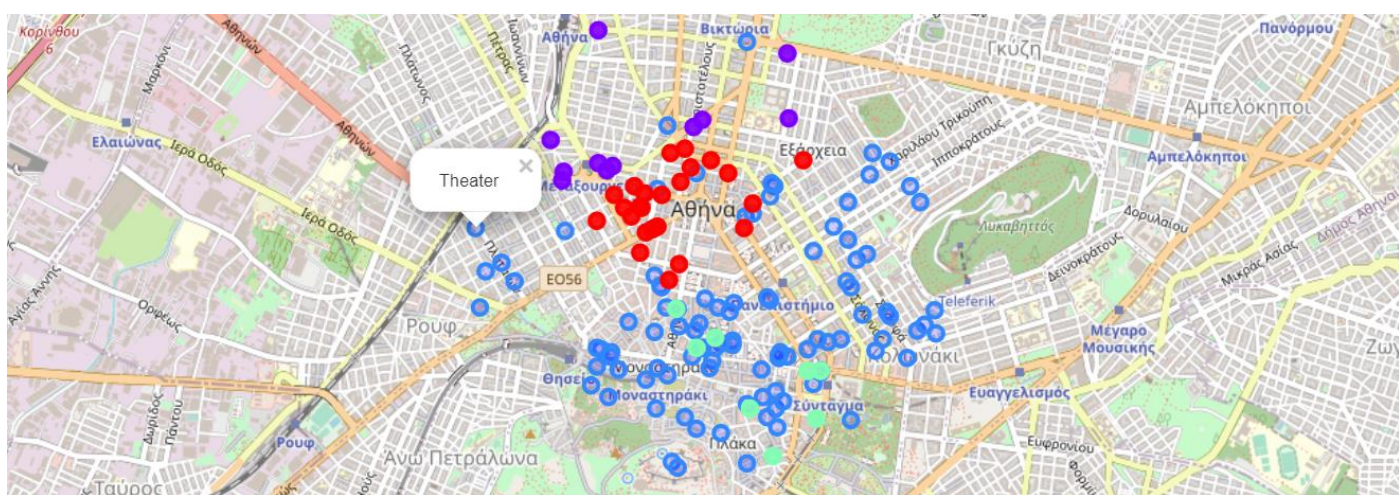


Figure 4: Clusters as “In between” in red, “Off the beaten path” in purple and “On the beaten path” in light green

cluster0 # <i>In between hotels</i>	cluster1 # <i>Off the beaten path</i>	cluster2 # <i>On the beaten path</i>
['Polis Grand Hotel', 'Titania Hotel', 'Ambrosia Hotel & Suites ****', 'Dorian Inn Hotel', 'Hotel Ritsi', 'Athens Center Square Hotel', 'Iniohos Hotel', 'Diros Hotel', 'Parnon Hotel', 'Vienna Hotel', 'Fresh Hotel', 'Elite Hotel', 'Ionis Hotel', 'Epidavros Hotel', 'Delphi Art Hotel', 'Elysium Hotel', 'CHIC Hotel', 'Joker Hotel', 'Marina Hotel', 'Achillion Hotel', 'Best Western My Athens Hotel', 'Sparta Team Hotel', 'Stalis Hotel', 'Exarchion Hotel', 'King Jason Hotel']	['Katerina Hotel', 'Apollo Hotel', 'Best Western Museum Hotel', 'Art Hotel', 'Moka Hotel', 'Nafsika Hotel', 'Novus City Hotel', 'Delta Athens Hotel', 'Radisson Blu Park Hotel', 'Rio Hotel', 'Oscar Hotel']	['Hapimag Resort Athens', 'Hotel Grande Bretagne', 'NJV Athens Plaza Hotel', 'Cecil Hotel', 'Amalia Hotel', 'Central Athens Hotel', 'Carolina Hotel', 'Tempi Hotel']

Figure 5: List with the Hotels

5. Discussion

After processing all the data and by observing the clusters an interesting deduction can be made by a person that knows the area. The hotels that are on the beaten path are far more expensive than these off the beaten path. Because the supply is almost the same the demand to lodge on the beaten path is significantly higher.

6. Conclusion and difficulties

To address the problem of congestion in Athens the choropleth map of the population density was given only for 5 municipalities. The data for the boundaries [2] were not in accordance with the data for the population density [1]. Even if the boundaries [2] were uploaded by the legal local authorities they are bugged. They are according to an elder segmentation even if in the webpage is stated that they are aligned with the latest amendment. However, the Top-5 municipalities that have been depicted are correct. In addition to this, the data that Foursquare can provide for free (or generally) are not enough. Hence, the initial idea to make a guide for “Where to build a hotel in Greece”, was abandoned.

7. References

[1] [Population Density](#)

[2] [Boundaries](#)