

# **Analysis of new Restaurant Location in Attica, Greece**

Nikolaos Syrios

April 2020

## **1. Introduction**

### **1.1. Background**

Opening a new restaurant is an investment that has a lot of variables to be considered. The type of the restaurant, the cuisine, the employees are some of the most important aspects of it. However before making the very first recruitment, the location must be taken into consideration. The choice of the location can be proven very advantageous for a business, especially for a restaurant that needs to be known by customers and be relatively accessible. Therefore, the ability to predict to some extent, the possibility of a successful business with regards to the location chosen can be crucial in the beginning of operations as well as in the prospects.

### **1.2. Problem**

As it was briefly outlined in the introduction, the success of a restaurant relies heavily on its location. Opening a restaurant in Attica, Greece is certainly a move in the right direction, considering the cultural and historical value it has to offer. Tourism alone can offer up to 5 million travelers each year consisting of about 16% of Greece's total tourism. [1] The question posed is which municipality of Attica would be beneficial, yet financially reasonable to set up the new business.

### **1.3. Interest**

Obviously, this comprises a proof of concept of an analysis made before opening a new business. The outcomes do not provide any immediate call to action, but they give a general impression of the nature of the process and the potential of such analysis.

## **2. Datasets**

### **2.1. Data sources**

The dataset consisting of the Municipalities of Athens and their respective location coordinates has been acquired from Wikipedia using web-scraping techniques in each Municipality webpage. The overview page, containing the individual links can be found here: [https://en.wikipedia.org/wiki/Category:Municipalities\\_of\\_Attica](https://en.wikipedia.org/wiki/Category:Municipalities_of_Attica).

The data of property pricing in these areas as well as the property index fluctuation of the past few years has been provided by <https://www.spitogatos.gr>, the leading real estate website in Greece. These data reflect the actual value of properties sold or rented based on the real-world transactions made using the website from 2011 till this day.

### 3.1. Data acquisition and cleaning

manually and some others had to be removed. Namely the city of Athens was not initially in the Wikipedia Municipalities page and had to be inserted by hand, while some municipalities showed

limited amount of venues when tested with the Foursquare API which skewed disproportionately the results and had to be removed.

Next, after cleaning the dataset, the Foursquare API had to be called in order to get the nearby venues of each municipality. The radius determining the “nearby” venues was set to 1200m and

Table 2. Top ten venues for each municipality. Segment of the table shown here.

	Municipality	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
0	Acharnes	Café	Plaza	Supermarket	Mobile Phone Shop	Bakery	Meze Restaurant	Gym	Coffee Shop	Souvlaki Shop	Stadium
1	Agia Paraskevi	Café	Greek Restaurant	Coffee Shop	Bakery	Supermarket	Bar	Electronics Store	Plaza	Dessert Shop	Yoga Studio
2	Agia Varvara	Greek Restaurant	Café	Dessert Shop	Plaza	Supermarket	Taverna	Park	Bakery	Snack Place	Grilled Meat Restaurant
3	Agioi Anargyroi-Kamatero	Café	Greek Restaurant	Coffee Shop	Plaza	Meze Restaurant	Burger Joint	Bus Stop	Bar	Mobile Phone Shop	Supermarket
4	Agios Dimitrios	Bakery	Café	Coffee Shop	Supermarket	Snack Place	Greek Restaurant	Grilled Meat Restaurant	Lingerie Store	Furniture / Home Store	Ice Cream Shop

the API limit per municipality to 100. The venue type from the data acquired was then processed using one-hot encoding. This step assured that we could later perform K-means clustering to a categorical value dataset as the one we are currently handling. The dataset was then normalized using the StandardScaler algorithm from sklearn library because the data were sparse and not evenly distributed which could potentially tamper with the results. A table with the top ten venues per municipality was created for illustrative purposes.

Performing K-means analysis to the cleaned data, required number of clusters we wanted to create. In order to determine this value, the elbow method was executed. The elbow method consists of iteratively running K-means analysis with different values of K and calculating the distortion. The distortion was calculated by averaging the squared distances from the cluster centers

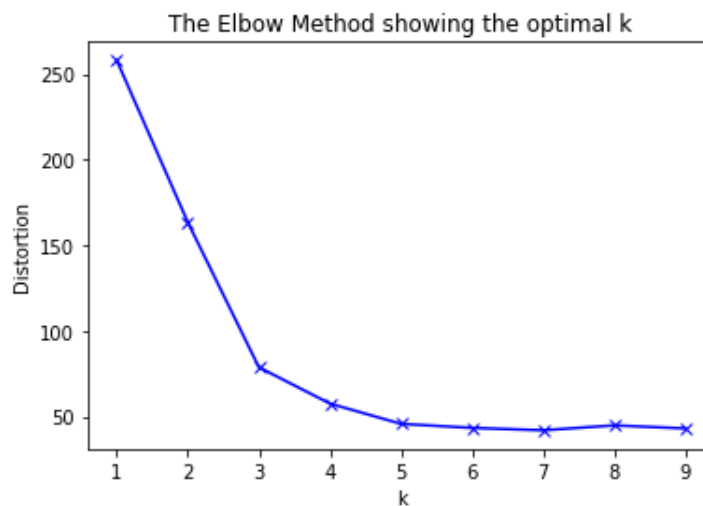


Figure 2. Determining the k-value with the Elbow method. Distortion is calculated with the Canberra distance. Optimal k-value was chosen as k=4 from the graph

of the respective clusters. The Euclidean distance seemed very unreliable as the graph representation

showed a linear fashion for all K-values in the range [1,10]. The Canberra distance was instead used to get the elbow chart, shown below in **Figure 2**. To determine the optimal number of clusters, we have to select the value of k at the “elbow” i.e. the point after which the distortion start decreasing in a linear fashion. So, the value of 4 was used for later analysis.

Performing the k-mean analysis with k value equal to 4, showed 2 of the 4 clusters accumulating more than 95% of the municipalities. This demonstrates the shortcomings associated with using the k-means algorithm for this dataset. Also running the algorithm for a couple iterations show inconsistent results owing to the heuristic nature of this approach. However, the results painted a very clear picture regarding the places with most restaurants.

Finally, the property pricing had to be determined in order to produce a meaningful summary of the results. The average pricing for renting and buying properties in each municipality was obtained from spitogatos.gr, the leading real estate listing website in Greece. The data unfortunately, were not readily available as an API endpoint but as a webpage. Again, using web-scraping techniques the average pricings (€/square meter) for the first quarter of 2020 were harvested for each municipality of Greece. The values regarding the municipalities of Attica were kept as well as the countrywide average for comparative purposes.

All the results were visualized using the folium library in a Map, illustrating the different clusters as well as the different pricing levels for renting and buying a property.

## 4. Results

### 4.1. Clusters

Using K-means algorithm with k value equal to 4, according to the elbow method, the dataset was redistributed in 4 clusters. As discussed in the methodology section, more than 95% of our data points resulted in 2 of the 4 clusters. This showcases the different disadvantages of using a clustering algorithm such as K-means in this dataset. However, this clustering provided us with some usable information regarding the restaurant distribution in the different municipalities of

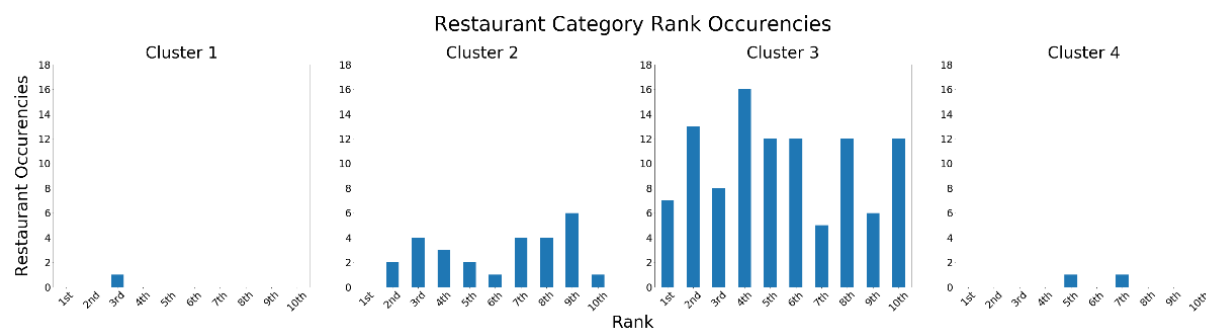


Figure 3. Restaurants occurrences in the different clusters' top ten venues. Cluster 3 shows a greater tendency in restaurant-like venues.

Attica, as shown in the charts of **Figure 3**. Cluster 3 showed a greater tendency towards restaurant-like venues. The distribution of the individual clusters can be viewed in **Figure 4**.

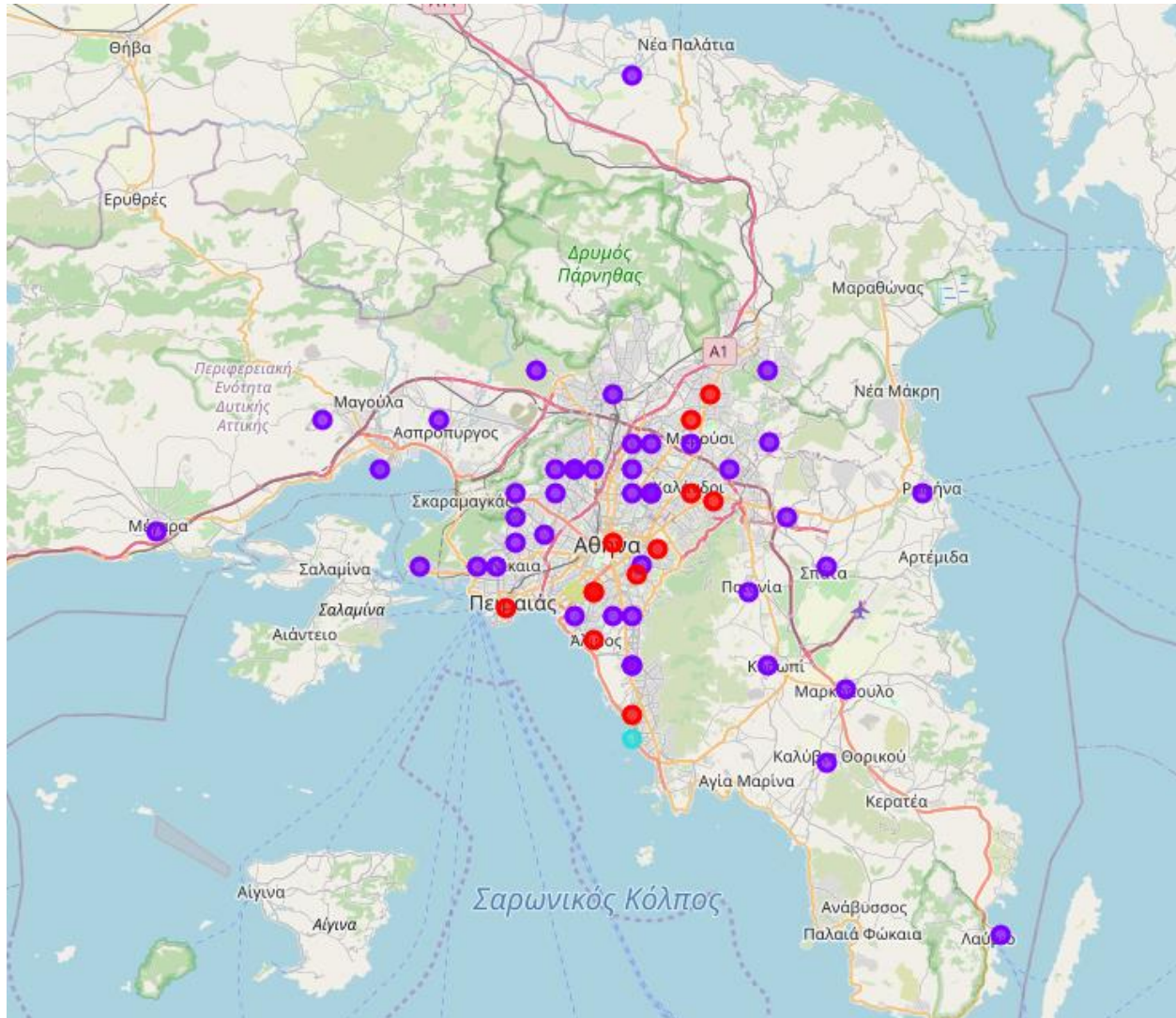


Figure 4. Distribution of clusters in the map of Attica.

Legend: Green -> Cluster 1, Red -> Cluster 2, Blue -> Cluster 3, Cyan -> Cluster 4

#### 4.2. Property average price distribution

Another factor that plays an important role in determining the place of a new business is the property pricing in this area. The average prices, rent and sale, were fetched by the leading real estate listing website in Greece (spitogatos) for each municipality as well as the average of the countrywide property price. The property pricing tendency is shown in **Figure 5**. The tendency of





Figure 5. Property average prices in different regions of Attica. Source <https://en.spitogatos.gr/>

price levels seems very similar when it comes to renting or buying a property, with some minor exceptions. The distribution in the map regarding the clusters is shown in **Figure 6** for renting a property and in **Figure 7** for buying one.

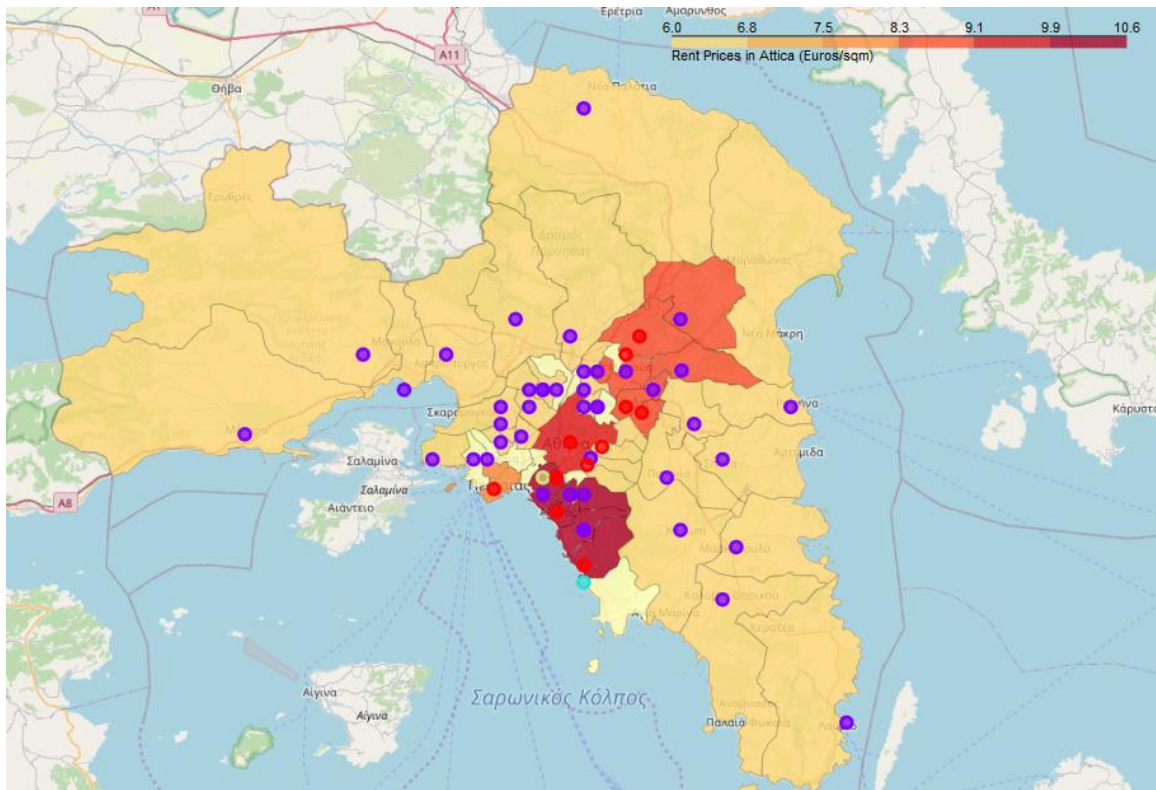


Figure 6. Average renting property prices distribution in Attica

Markers Legend: Green -> Cluster 1, Red -> Cluster 2, Blue -> Cluster 3, Cyan -> Cluster 4

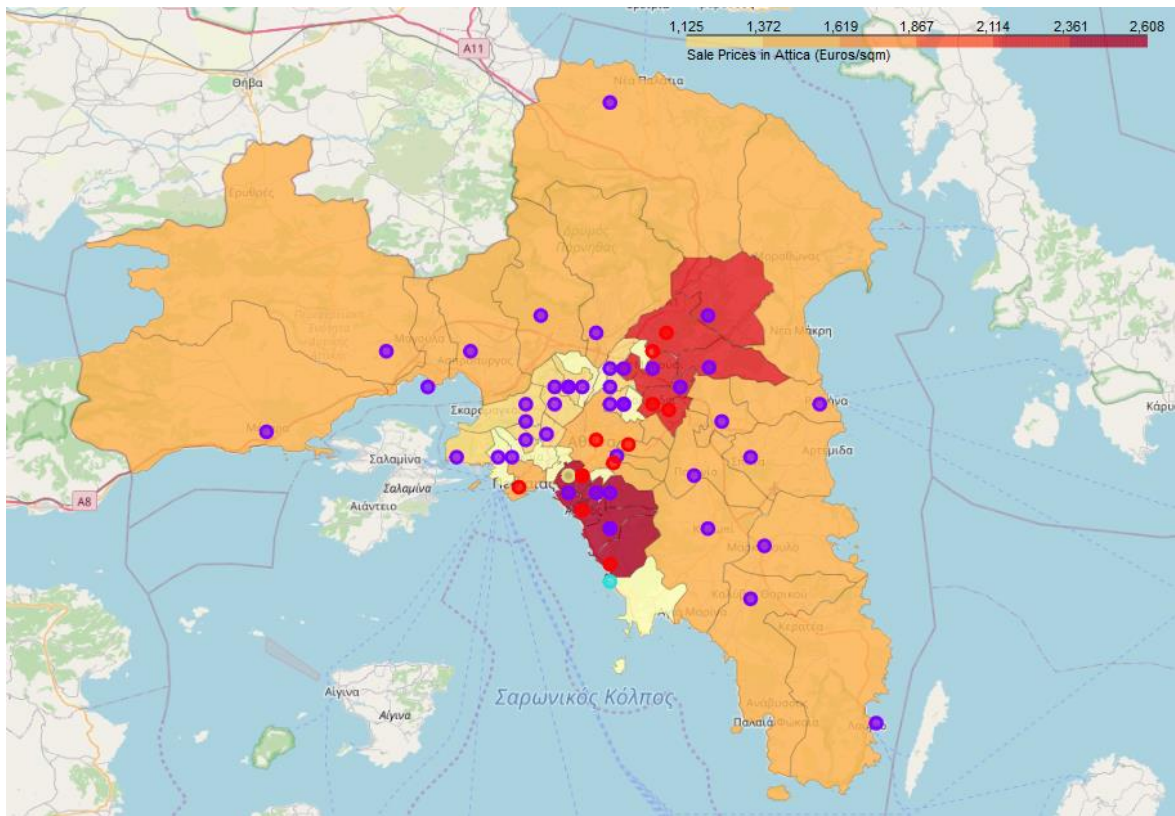


Figure 7. Average buying property prices distribution in Attica  
 Markers Legend: Green -> Cluster 1, Red -> Cluster 2, Blue -> Cluster 3, Cyan -> Cluster 4

## 5. Discussion

The segmentation of a big and crowded Greek region such as Attica can provide many challenges. The choice of the segmentation process and algorithm could potentially yield very different results. The information source can have a great impact on that too. The use of k-means algorithm provides some quick results that can give a good impression of the place. However, being a heuristic algorithm can converge to a local minimum instead of a global one, losing the "bigger picture" and skewing the results. The use of the Foursquare API may limit the results to certain venues that local people feel more comfortable reporting on.

There are certainly limitations to this analysis that may or may not be overcome with different data sources and/or processing algorithms. However, we can get a descriptive snapshot of the different Municipalities and their venues that certainly give us plentiful of information. We can draw several conclusions about the real estate status and the different venues in each municipality.

There are also prospects for this analysis, that could include broader data sources regarding the venues and a more detailed real estate pricing status for each municipality. After picking the

desired municipalities, we could proceed analyzing the respective neighborhoods in order to pinpoint the exact location for the new restaurant. Other factors must be taken into consideration, such as the type of restaurant, whether the property will be bought or rented as well as the target group and cuisine.

## **6. Conclusion**

In this study, I analyzed the fitness of all the municipalities of Attica in opening a new restaurant. I identified each municipality's venues and average pricing for buying and renting properties. I clustered the municipalities using a classification algorithm showcasing the relations between the venues in each cluster. Later these clusters were combined with the data about the average property pricing. This analysis could be beneficial for someone desiring to open a new restaurant in Attica. Changing a few parameters, the methodology used could be extrapolated to a wide variety of businesses. The same methodology could be tweaked to work also in different places in the world classifying the individual municipalities or neighborhoods of a region.