

Applied Data Science Capstone Project

Finding the best location to start up a modern cuisine restaurant in Athens



Introduction

For Applied Data Science Capstone Project a hypothetical scenario of a modern cuisine restaurant in Athens, Capital of Greece is created. In this scenario a professional of food and drink gastronomy wants to start a modern cuisine area in Athens and looks for advice on selecting the best location for opening his restaurant in Athens. Project idea is selecting the best location to avoid unnecessary competition because of similar restaurants in proximity and exploiting the opportunity assuming that there are not many modern restaurants. To maximize profit attracting many customers he must open his restaurants in area where modern cuisine is popular, meaning that there are similar restaurants in the area, so he must make a very detailed decision and select the most suitable location.

Problem statement

The objective of this capstone project is to find the most suitable location for a gastronomy professional to start up a modern cuisine restaurant in Athens, capital of Greece. Using data science methodology and machine learning such as clustering, project aims in providing the best proposal to the question: where should a gastronomy professional, in vicinity of Athens, open a modern cuisine restaurant, considering that it should attract many customers who enjoy this type of cuisine and avoiding unnecessary competitive restaurants that it is possible.

Target audience

Gastronomy professionals who wish to open a modern cuisine restaurant in Athens city.

Data needed

This problem to be solved needs following data:

- List of neighborhoods Athens, Greece.
- Latitude and Longitude of these neighborhoods.
- Venue data related to modern cuisine restaurants to find the most suitable neighbourhood to open such a restaurant.

Methodology

At first step the list of Athens municipalities is get. It is extracted by webscraping using beautiful soup module from the wikipedia page https://en.wikipedia.org/wiki/Category:Municipalities_of_Attica

Data are collected in a list and their coordinates (latitude and longitude) are found using foursquare api. Municipalities names and their coordinates are stored inside a dataframe. A function utilizing try, except is used because some of these municipalities may not be detectable by foursquare. Geocoder data does not work well so foursquare solution is preferred. for practicality reasons search results are stored in a csv file inorder not to webscrape and collect data each time notebook is run.

Using foursquare api and modern greek cuisine, modern european cuisine categories all relevant restaurants near collected municipalities coordinates are found and stored in second dataframe and stored in another csv file for practicality reasons too. The data are further cleaned removing potential venues lieing outside of Attica (Athens) state and outside of Greece. Duplicate values are found and removed too. Unnecessary columns of the dataframe are removed too.

Using Folium restaurants are visualized on map and an initial estimation, based on geographical dispersion/ concentration is made about the optimum location that a restaurant should open.

Lastly, k-means clustering is performed. Trials with empirically chosen two and three clusters number (k) are made, while elbow method implementation indicates k=2 as the optimum number of clusters, that agrees with emperical results.

Collecting necessary data

Steps that will be followd to collect necessary data are:

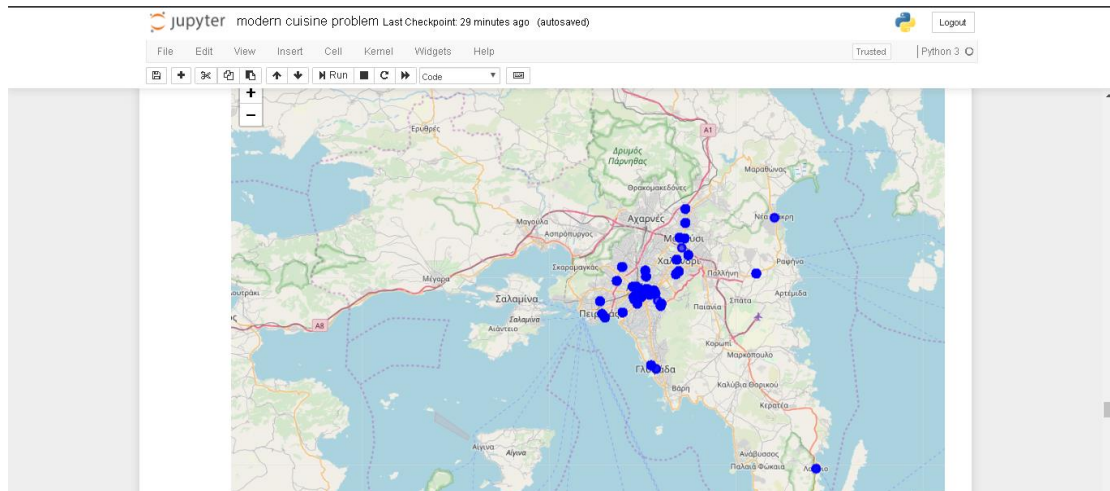
- Scrapping Athens neighborhoods from a Wikipedia page (https://en.wikipedia.org/wiki/Category:Municipalities_of_Attica) using Beuatiful Soup
- Getting the coordinates (Latitude and Longitude)of these neighborhoods via Geocoder package
- Using Foursquare API to get venue data related to these neighborhoods

Results

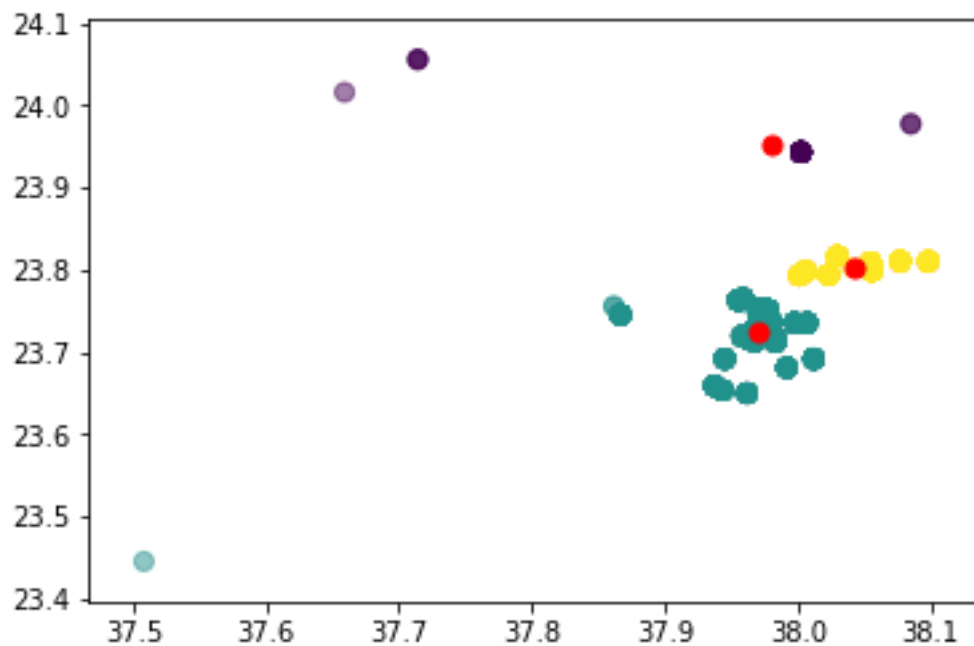
Visualize restaurants

The following map presents restaurants in Athens offering modern Greek and European cuisine geographical dispersion in Athens area. Next using k-means clustering the best location to open a restaurant of modern cuisine will be selected. The problem is solved for

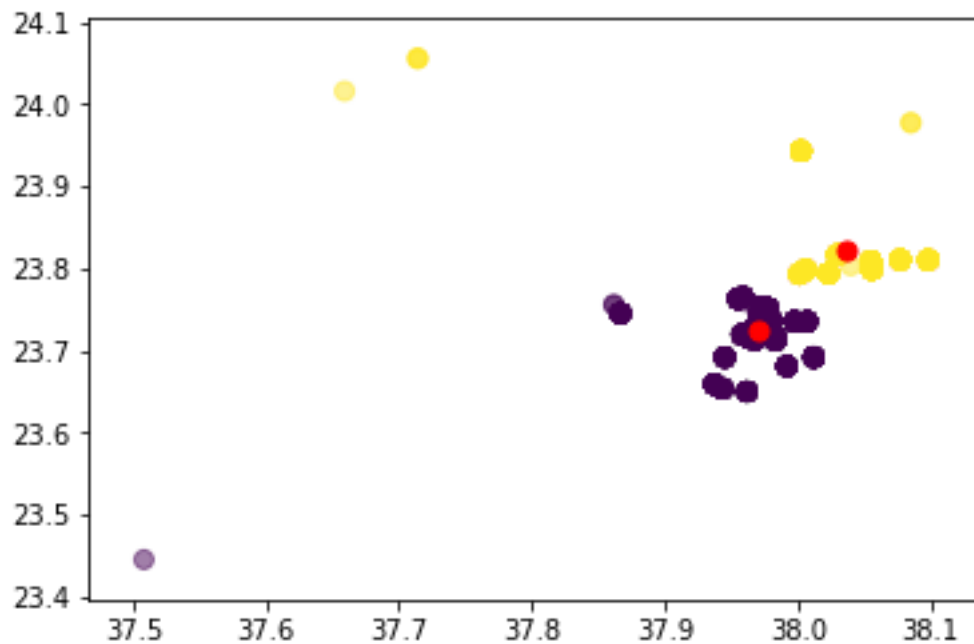
two and three clusters. Using the elbow method, it is proved that optimal number of clusters is $k=2$, something that empirically studying the map above seems correct



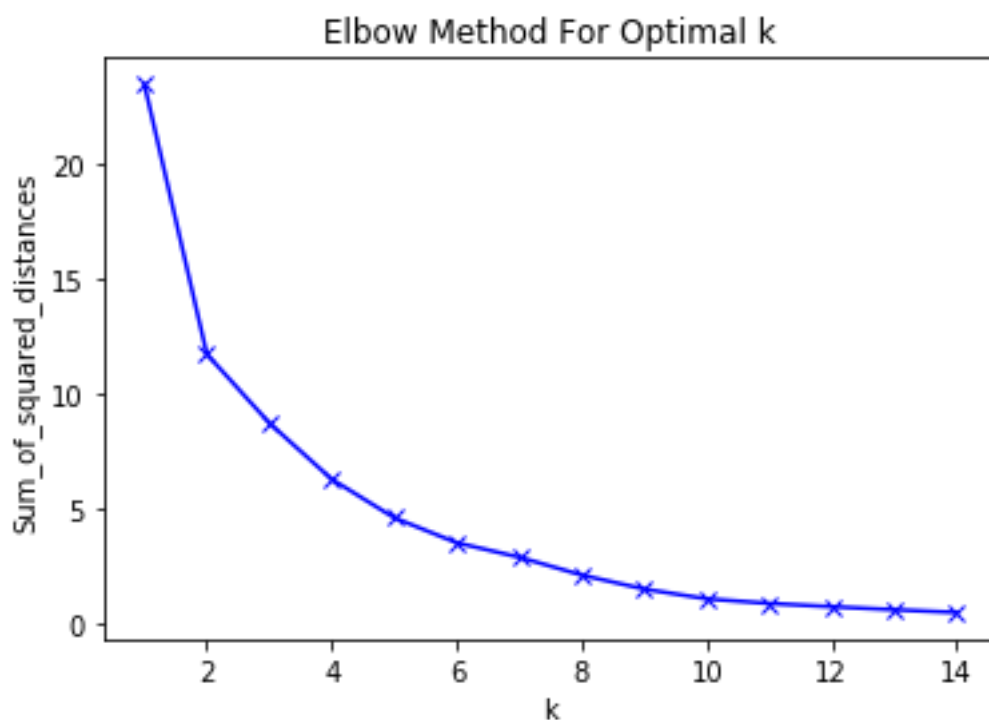
Following picture presents solving the business problem in scope with k-means method using three centroids (presented with red colour).



While, the following picture presents the k-means solution with two centroids:



Two clusters appearing indicate the optimum solution. The optimum selection of two clusters is mathematically proven using the elbow method:



Discusion - recommendations

K-means clustering indicates that exist two location-clusters (using $k=2$). The cluster of lower concentration should be chosen as the solution; a location belongint to that cluster (of yellow colour) should be selected. Strategically, restaurants owners should select a location where much customer traffic is expected. So, all restaurant ownwers would make similar selections about the location of their restaurant. The result of this behavior would be that many restaurants would open nearby in the same area to exploit that area

strategical characteristics. So, one should select one cluster with high concentration, ideally the smallest of the two to avoid some of the competition.

Conclusion

In this project, we have gone through the process of identifying the business problem, specifying the data required, extracting and preparing the data, performing the machine learning by utilizing k-means clustering and providing recommendation to the stakeholder.

References

List of municipalities of Athens:

[1]:https://en.wikipedia.org/wiki/Category:Municipalities_of_Attica

Foursquare documentation:

[2]:https://en.wikipedia.org/wiki/Category:Municipalities_of_Attica

Foursquare resources' categories:

[3]:<https://developer.foursquare.com/docs/resources/categories>