

# Coursera Capstone

## IBM Applied Data Science Course

### Opening a Vegetarian/ Vegan Restaurant in Istanbul, Turkey

Sinan Sonmez

April, 2021



## Introduction

Vegetarian cuisine is based on food that meets vegetarian standards by not including meat and animal tissue products (such as gelatin or animal-derived rennet). The strictest form of vegetarianism is veganism, which excludes all animal products, including dairy, honey, and some refined sugars if filtered and whitened with bone char.

There are also partial vegetarians (flexitarians), such as pescetarians who may eat fish but avoid other types of meat.

Turkish cuisine is largely the heritage of Ottoman cuisine, which can be described as a fusion and refinement of Mediterranean, Balkan, Middle Eastern, Central Asian, Eastern European and Armenian cuisines. Meat-based foods such as kebabs are common in Turkish cuisine.

Over the last few years, more and more people prefer to have vegetarian/ vegan dishes. Therefore, it is a challenging task to decide which neighbourhood is the best place to open a new vegetarian/ vegan restaurant.

## Business Problem & Target Audience

Using data science, this project aims to answer entrepreneurs which is the best location to open a vegetarian/ vegan restaurant in Istanbul, Turkey.

Target audience of this project is local and international entrepreneurs who would like to open a vegetarian/ vegan restaurant in Istanbul, Turkey. Every day more and more people become aware of the environmental effects of consuming meat and meat related products. Many people object to eating meat out of respect for sentient life. Such ethical motivations have been codified under various religious beliefs, as well as animal rights advocacy. Therefore there is a good business opportunity to open such restaurants.

## Data

To solve the problem, we need the following data

1. list of districts in Istanbul
2. average annual household income data for each district
3. coordination (latitude and longitude) information for each district

## Data Sources and Extraction Methods

1. List of districts in Istanbul: The data is available in Wikipedia (please see the link below). The data is extracted from Wikipedia using web scraping technique and BeautifulSoup library. Istanbul has 39 districts.  
Data Link: [https://en.wikipedia.org/wiki/List\\_of\\_districts\\_of\\_Istanbul](https://en.wikipedia.org/wiki/List_of_districts_of_Istanbul)
2. Average Annual House Income (\$): This data is also available in the link above. The same web scraping technique is used to extract the data.

	District	Population	Area(km2)	Density(per km2)	Annual Household Income(\$)
0	Adalar	16033	11.05	1,451	10,978
1	Arnavutköy	296709	450.35	659	3,350
2	Ataşehir	422594	25.23	16,750	10,854
3	Avcılar	436897	42.01	10,400	6,064
4	Bağcılar	737206	22.36	32,970	5,295

3. Coordination: The coordination info (latitude and longitude) is obtained for each district using the Python Geocode package.

	District	Population	Area(km2)	Density(per km2)	Annual Household Income(\$)	Latitude	Longitude
0	Adalar	16033	11.05	1,451	10,978	40.86306	29.12423
1	Arnavutköy	296709	450.35	659	3,350	41.18558	28.74147
2	Ataşehir	422594	25.23	16,750	10,854	40.99248	29.12777
3	Avcılar	436897	42.01	10,400	6,064	40.97813	28.72101
4	Bağcılar	737206	22.36	32,970	5,295	41.03323	28.86351

After 3 data are obtained, we used the Foursquare API to get the venue data for each district. Foursquare API provides us the number of revenues in specific categories (e.g. turkish restaurant, italian restaurant, cafe, bar etc) for each district. We used this information which region has already vegetarian/ vegan restaurants.

## Methodology

In order to start the analysis, we need the list of districts of Istanbul, Turkey. This data is already available in Wikipedia. We leverage web scraping technique using Python and BeautifulSoup to extract the information. Fortunately, Wikipedia has additional information which we can use later on such as Population and Annual Household Income (\$). Finally, we need the coordinates of each district as latitude and longitude. Coordinate information is important because we are using coordinate information to get the venue information from Foursquare API. We used Python's

Gecoder package to convert the address into geographical coordinates. After all the data is extracted, we created a Pandas data frame to be able to work on it. We also manually checked some of the coordinates (randomly selected) to make sure that we have the correct information.

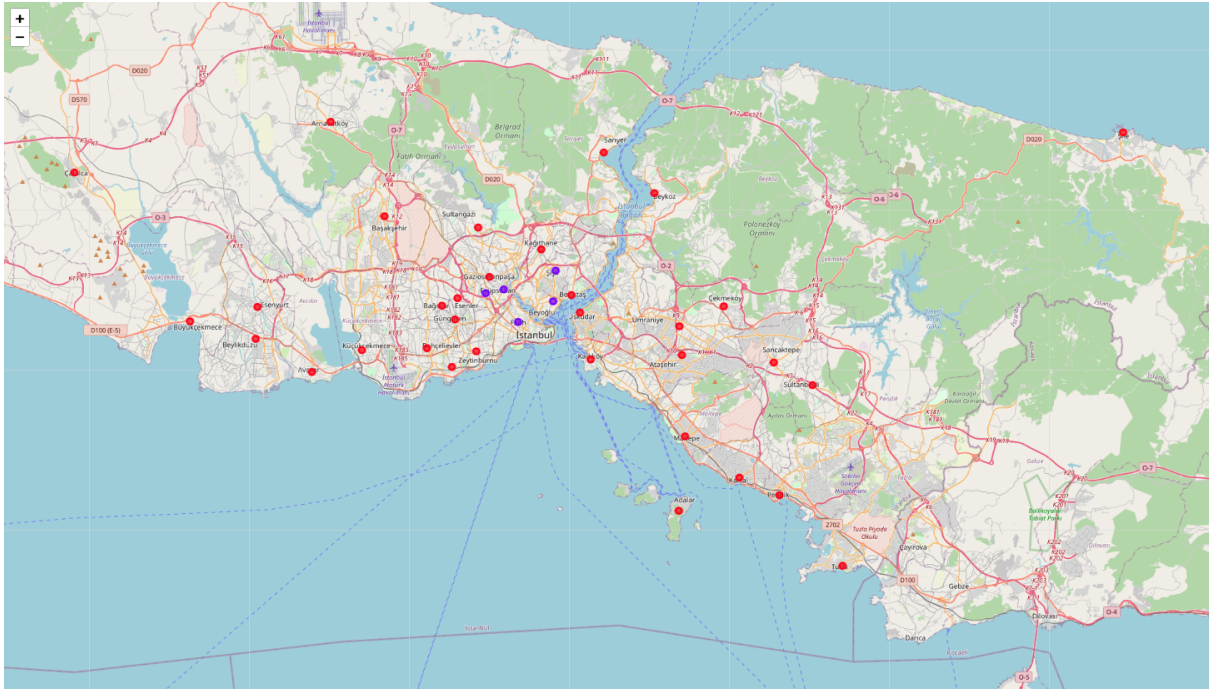
Next,

Finally we use k-means clustering technique to differentiate clusters based on vegetarian/ vegan restaurant availability. We cluster the districts into 2 clusters. One cluster includes districts which already have a vegetarian/ vegan restaurant and the other cluster includes the remaining districts. These clusters help us to focus only districts without a vegetarian/ vegan restaurant. Then we sort these districts based on annual household income to understand which district has no vegetarian/ vegan restaurant but has a high income.

## Results

Analysis show that we have 2 different clusters based on the vegetarian/ vegan restaurant availability

- Cluster 0 (red): Districts with zero vegetarian/ vegan restaurant
- Cluster 1 (purple): Districts with at least one vegetarian/ vegan restaurant



## Discussion

The purpose of this project is to suggest a district to an entrepreneur to open a vegetarian/ vegan restaurant in Istanbul, Turkey. As you can see in the map below, vegetarian/ vegan restaurants are concentrated in the center of Istanbul and European side of the city.

## Conclusion

In this project, we went through the process of identifying the business problem, specifying the data required, extracting and preparing the data, performing machine learning by clustering the data into 2 clusters based on their similarities, and lastly providing recommendations to the entrepreneurs the best locations to open a new vegetarian/ vegan restaurant.

The answer to the business based on the analysis: The districts in cluster 0 don't have any vegetarian/ vegan restaurant. When we deep dive into these districts, we see that the top 3 districts based on Annual Household Income(\$): Besiktas (Europe), Kadikoy (Asia) and Bakirkoy (Europe) don't have any vegetarian/ vegan restaurants. These districts are also big in terms of population. Therefore these three districts are the best choices for vegetarian/ vegan restaurants.