

# Coursera Capstone Project

The Battle of Neighborhoods

Identifying a convenient place to Open a Shared Working place



By

Abdurahman Ali Mohammed

June, 2020

## Table of Contents

1. Introduction.....	1
2. Business Problem.....	1
Target Audience.....	1
3. Data.....	1
4. Methodology.....	2
5. Result.....	3
6. Discussion.....	5
7. Conclusion.....	5

## 1. Introduction

In our rapidly moving world, we all are busy with our jobs and striving to our goals. One thing that is widely believed to aid our productivity is a good workspace. Workspace refers to small premises provided, often by local authorities or economic development agencies, to help new businesses to establish themselves. These typically provide not only physical space and utilities but also administrative services and links to support and finance organizations, as well as peer support among the tenants. A continuum of sophistication ranges through categories such as 'managed work-spaces', 'business incubators' and 'business and employment co-operatives'. In cities, they are often set up in buildings that are disused but which the local authority wishes to retain as a landmark. At the larger end of the spectrum are business parks, virtual offices, technology parks and science parks.

## 2. Business Problem

In this document we will go through the process in which Machine learning is used to identify a convenient place to set up a shared workspace rental service in the city of Izmir, Turkey. There will some requirements put up by our client which we have to take into consideration while identifying the required area. According to the client the new workspace should be located in an area which is surrounded by other complimentary services such cafes.

Based on the above requirements, 3 neighborhoods will be recommended for the client to set up a Shared Workspace rental service in the city of Izmir, Turkey.

### Target Audience

The target audience for this project is a client who wants to set up a shared workspace rental business in the city of Izmir, Turkey. Based on the analysis that will be done in this project the clients will be handed locations at which they can setup their business.

## 3. Data

Our first data source will be Atlasbig.com (<https://www.atlasbig.com/tr/izmirin-ilceleri>). From there we scrap data of Izmir's neighborhoods with **BeautifulSoup**. Now that we have acquire the neighborhoods, we will use the technique of Geocoding to get the coordinates of the neighborhoods. For this purpose, we make use of **Geopy's Nominatim**. After preparing the neighborhoods and their coordinates, next thing to do is get venues within each neighborhood with the use of the **Foursquare API**. We do this in order to choose a prefect and convenient surrounding for the business to be set up.

## 4. Methodology

To identify a convenient place to set up the required business, we will go through the steps of acquiring data, cleaning and preparing the data, building a model and clustering the neighborhoods. To acquire the required data about the neighborhoods in Izmir, Turkey, we use the technique of web scraping using the BeautifulSoup tool. After that we create a data set and clean up the acquired data to prepare to build our model. Once our model is ready, we cluster the neighborhoods according to their convenience to the stated goal.

Since the clients want the workspace business to be opened in an area where cafes and restaurants are adequately abundant, we pick out neighborhoods which have renowned cafes and restaurants. With that, we construct a data set on which we do our clustering. Our clustering then results in different clusters according to prevalence of the desired venues in the neighborhoods. K-Means Clustering technique was used to do the clustering of the neighborhoods.

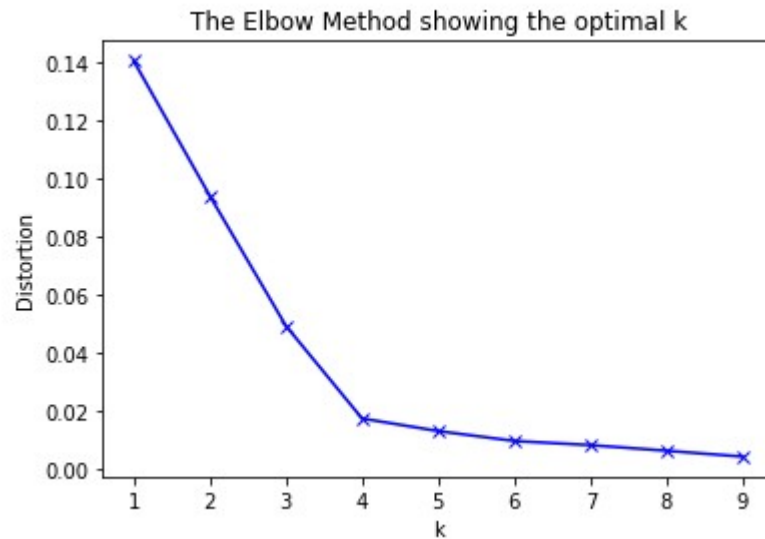
K-means clustering is a good place to start exploring an unlabeled data set. The K in K-Means denotes the number of clusters. This algorithm is bound to converge to a solution after some iterations. It has 4 basic steps:

1. Initialize Cluster Centroids (Choose those 3 books to start with)
2. Assign datapoints to Clusters (Place remaining the books one by one)
3. Update Cluster centroids (Start over with 3 different books)
4. Repeat step 2–3 until the stopping condition is met.

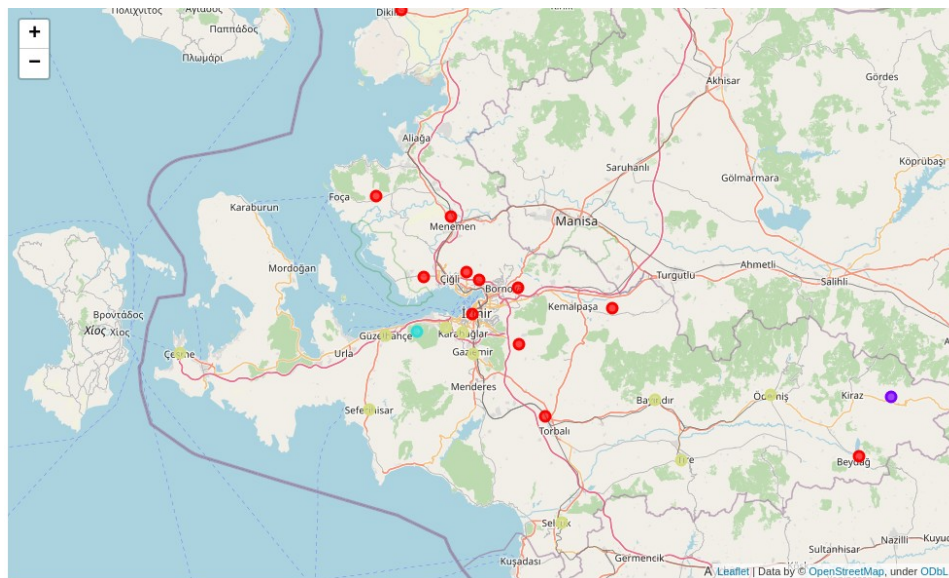
After clustering the neighborhoods, we will have insights about which clusters contain various options of cafés and restaurants to setup the business in. To provide more specific insights, we use the technique of data visualization to display the magnitude of the occurrence of Cafes and Restaurants. To perform this task, **Folium** library was used.

## 5. Result

So, the first observation is the number of clusters that will result in. Using the Elbow method, we understand that 4 Clusters would be the optimal number of clusters.



Therefore, K-Means Clustering was run on the data set and the following result was acquired.



As we can observe from the map, there are 4 clusters that are segmented according to the extent to how much crucial venues such as cafes and restaurants they have. For better understanding we display the resulting clusters' data and go through it.

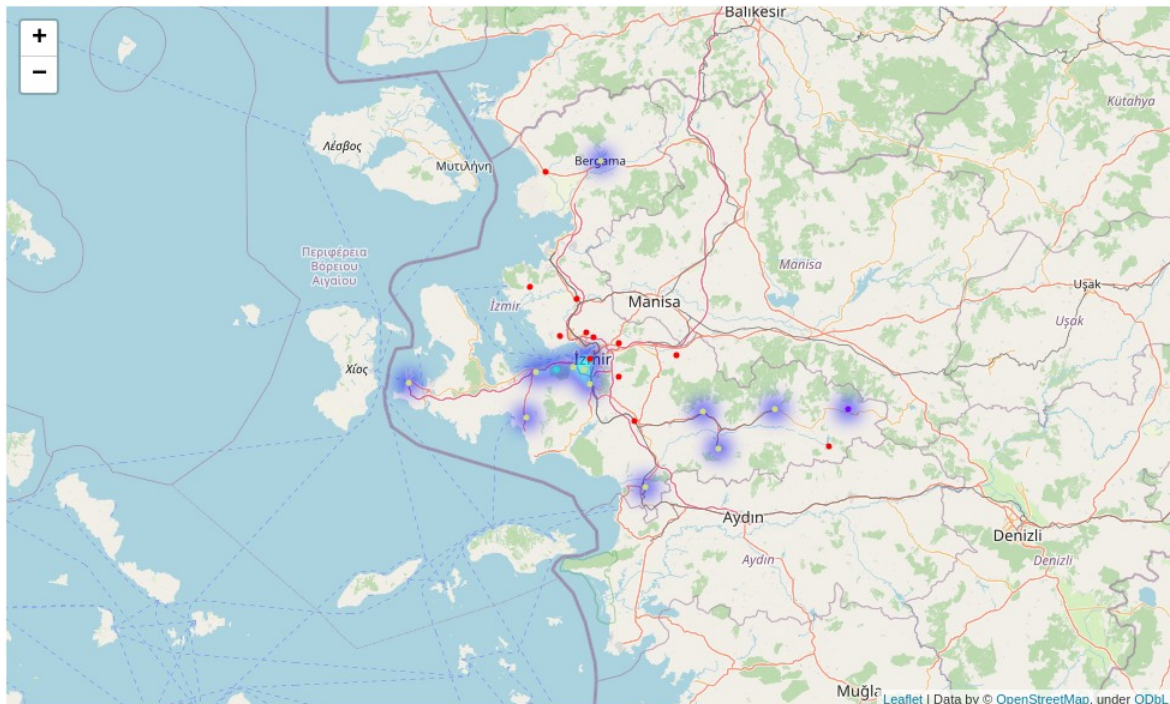
Now that we have clustered our neighborhoods in Izmir, Let's check how our clusters look like.

```
In [119]: 1 oc[izmir_merged['Cluster Labels'] == 0, izmir_merged.columns[[1] + list(range(5, izmir_merged.shape[1]))]].head()
```

Out[119]:

	location	altitude	Cluster Labels	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	(Buca, İzmir, Ege Bölgesi, Türkiye, (38.34446095, 27.262771990750778))	0.0	0.0	Garden	Farm	Concert Hall	Wings Joint	Electronics Store	Fishing Spot	Fish Market	Fish & Chips Shop	Fast Food Restaurant	Farmers Market
2	(Bornova, İzmir, Ege Bölgesi, Türkiye, (38.4678978, 27.260804924467173))	0.0	0.0	Stadium	Mountain	Wings Joint	Electronics Store	Fish Market	Fish & Chips Shop	Fast Food Restaurant	Farmers Market	Farm	Factory
3	(Konak, İzmir, Ege Bölgesi, Türkiye, (38.410958, 27.1294534))	0.0	0.0	Bakery	Borek Place	Kokoreç Restaurant	Flower Shop	Turkish Restaurant	Park	History Museum	Plaza	Soccer Field	Kofte Place
4	(Karşıyaka, İzmir, Ege Bölgesi, Türkiye, (38.5034711, 27.113482575513878))	0.0	0.0	Mountain	Market	Garden	Farm	Wings Joint	Factory	Fish Market	Fish & Chips Shop	Fast Food Restaurant	Farmers Market
5	(Bayraklı, İzmir, Ege Bölgesi, Türkiye, (38.4867391, 27.149264834063008))	0.0	0.0	Cosmetics Shop	Pool Hall	Bike Trail	Fast Food Restaurant	Gaming Cafe	Factory	Fishing Spot	Fish Market	Fish & Chips Shop	Farmers Market

Based on the clustering that we conducted we have got the information that clusters 0 and Cluster 3 contain the type of neighborhoods that we are looking for. But since we are looking for better information to make our decision, we have constructed a Heat-Map.



One important addition for the potential neighborhood to have is being located at the heart of the city. That can be clearly seen from the heat map that we have constructed. On the Heat-Map we can see the Neighborhood 'Konak' which is under Cluster 0 has



good amount of the desired venues as well as proximity to the center of the city. Furthermore, it is located closer to the Sea which will give customers fresh breath as well as amazing views.

## 6. Discussion

In general, the use of Data science techniques along with Geo-spatial data acquired from Foursquare API has played a huge role in aiding the decision to set up a shared workspace rental business in Izmir, Turkey. From acquiring the neighborhoods in the city, through fetching venues in each neighborhood, up to using a model to cluster and identify potential locations can clearly show how data science is powerful as well as applicable in our daily lives.

One challenge that was encountered when carrying out this project was lack registered venues in cities. Sometimes the scarcity of data might lead to wrong decisions. Therefore, it is advised to carry out such projects on top of adequate data. In addition to this, Other factors should also be considered when making business decisions.

## 7. Conclusion

To conclude, with the help of data science techniques and tools with the collaboration of Foursquare API, an optimal location for a shared workspace rental business in Izmir, Turkey was identified. Even though further research is needed before starting the business, a potential neighborhood (Konak) has been selected for the business.