

# Selecting the Best Borough to Open a Restaurant

Engin ANKARALI

3 March, 2020

## 1. Introduction

### 1.1. Background

Izmir is the third largest city in Turkey with a population of around 4 million, the second biggest port after Istanbul, and a very good transport hub. It has total of 30 districts and 11 of them are counts as a central. The fact that almost half of its population of 4 million are under the age of 30, makes Izmir a city full of life. The city hosts tens of thousands of university students, educates scientists, artists, business leaders and academics. It is a rapidly growing city on the Central Aegean coast of Turkey. [1][2]

### 1.2. Problem description

In a highly populated city like Izmir, opening a restaurant is always more risky, costful and time consuming. Main goal of this project is, using KMeans clustering model and Foursquare locate which boroughs in Izmir suits well to opening a new restaurant in Izmir. This project will provide information about most common venues of boroughs and their characteristics. With that information stakeholders can choose which borough(s) suits well to them.

## 2. Data Description

### 2.1. Izmir

I found Turkey - Subnational Administrative Boundaries from HDX [3] the .xlsx file has all neighborhoods, refugee camps, cities of Turkey and their latitude and longitude values in English and Turkish. I limit it to city of Izmir and cleaned data from unnecessary columns and created new csv file named "izmir\_data" which includes borough name, latitude of borough, longitude of borough. Original file and cleaned file can be reachable from my github repo.

### 2.2. Nominatim

I used Nominatim from Geopy to get latitude and longitude value of Izmir.

Geopy is a Python client for several popular geocoding web services. Geopy makes it easy for Python developers to locate the coordinates of addresses, cities, countries, and landmarks across the globe using third-party geocoders and other data sources. [4]

### 2.3. Foursquare API

I used Foursquare API to get the most common venues in boroughs of Izmir.

The Places API offers real-time access to Foursquare's global database of rich venue data and user content to power your location-based experiences in your app or website.[5]

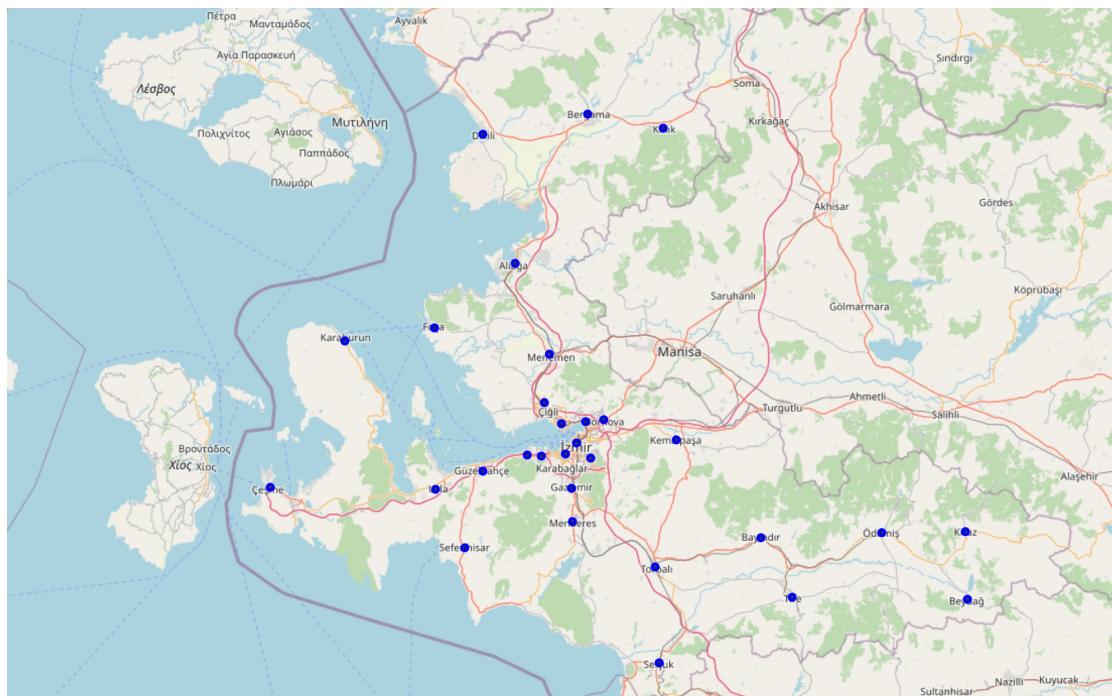
### 3. Methodology

The data from I found from HDX has information about cities of Turkey, their neighborhoods, latitude and longitude values in English and Turkish format. I limit it to boroughs of Izmir and dropped some of columns which I don't need.

Our data after cleaning process.

	Borough	Latitude	Longitude
0	ALIAGA	38.802491	26.975580
1	BALCOVA	38.392353	27.048095
2	BAYINDIR	38.218886	27.648303
3	BAYRAKLI	38.467495	27.167934
4	BERGAMA	39.118371	27.173508

I used folium to create map of Izmir using latitude and longitude values. In next step I will be getting venues using Foursquare API.

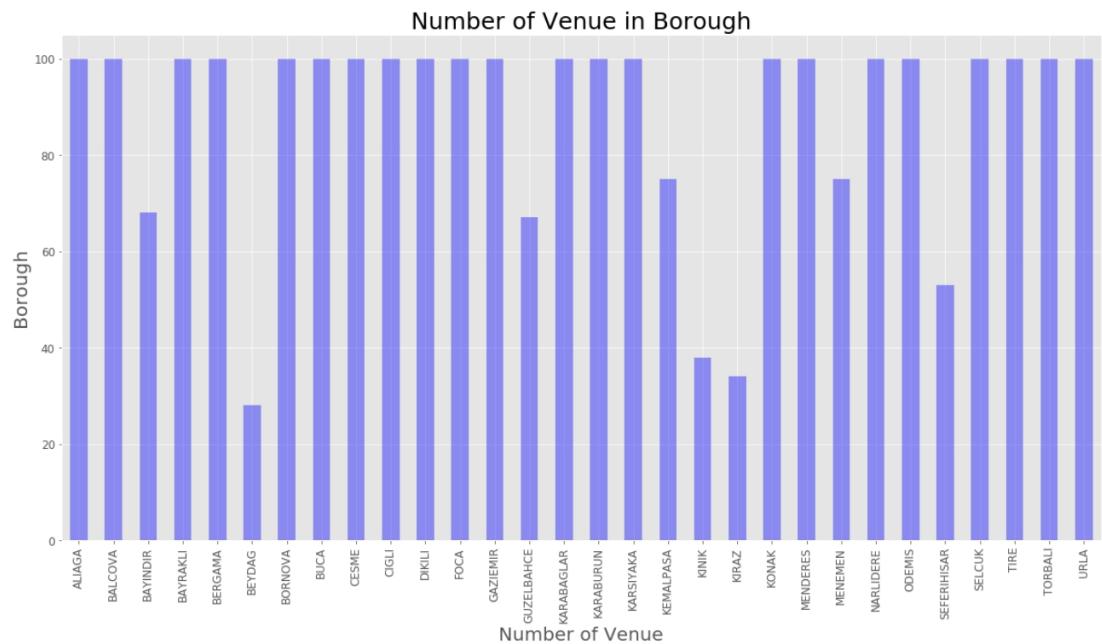


To get venues I used Foursquare API. I designed the limit as 100 venue and the radius 1500 for each borough from their given latitude and longitude informations. 272 unique categories returned. Here is our merged table of boroughs and venues.

Borough	Borough Latitude	Borough Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0 ALIAGA	38.802491	26.97558	Brew Mood Coffee & Tea	38.803533	26.973963	Coffee Shop
1 ALIAGA	38.802491	26.97558	HURACAN	38.802982	26.974550	Gym / Fitness Center
2 ALIAGA	38.802491	26.97558	Kitabe	38.801566	26.974845	Bookstore
3 ALIAGA	38.802491	26.97558	Sofrada Balik	38.801844	26.972237	Seafood Restaurant
4 ALIAGA	38.802491	26.97558	Mozaik Sanat	38.804507	26.975254	Dance Studio

From this table we can see how many venues returned for each borough. As you can see, Bayindir, Beydag, Guzelbahce, Kemalpasa, Kinik, Kiraz, Menemen and Seferihisar didn't reach their limit unlike others.

Also I should say that this graph is not %100 accurate. Longitude and latitude values used as center of boroughs and from that center I created circle which has 1500 radius so, our circle may not be cover all of borough or may exceed bordor of borough.

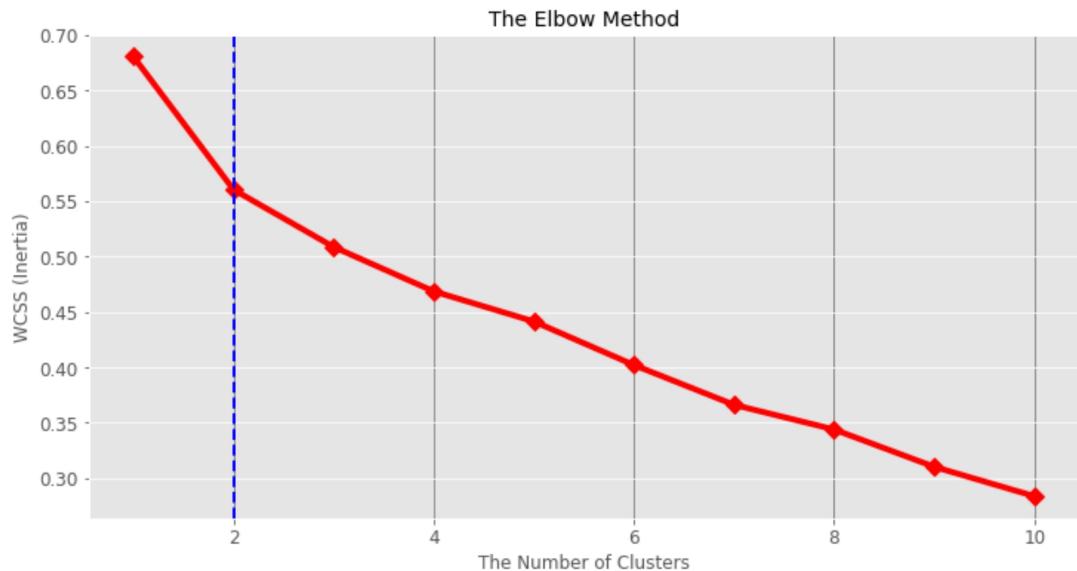


In summary of this graph 280 unique categories were returned by Foursquare API, then I created a table which shows a list of top 10 venue category for each borough in below table.

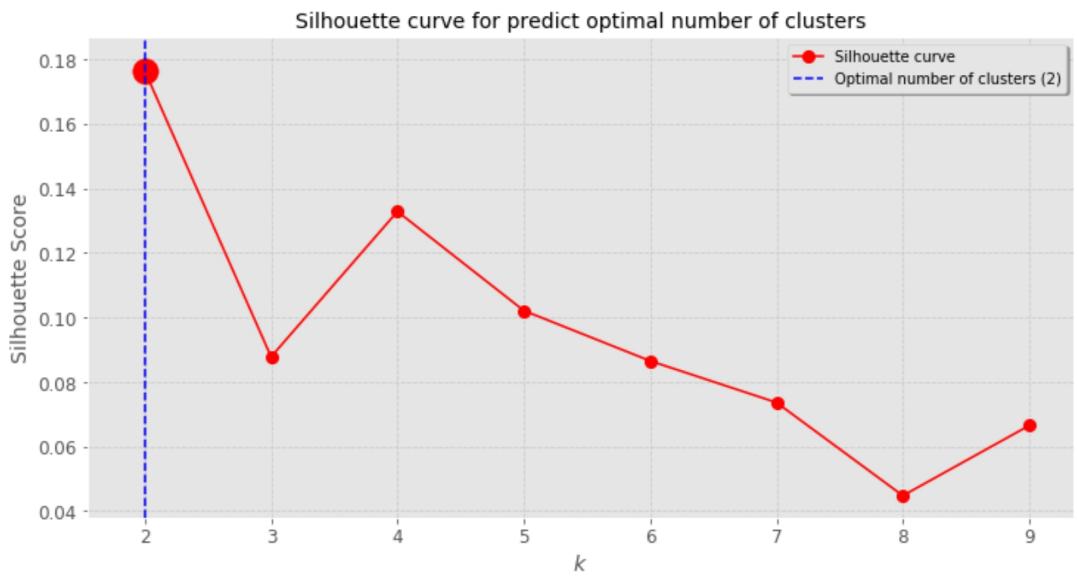
Borough		1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	ALIAGA	Café	Restaurant	Bar	Seafood Restaurant	Gym / Fitness Center	Turkish Restaurant	Park	Burger Joint	Bagel Shop	Beach
1	BALCOVA	Clothing Store	Coffee Shop	Gym / Fitness Center	Café	Sporting Goods Shop	Dessert Shop	Restaurant	Burger Joint	Bakery	Furniture / Home Store
2	BAYINDIR	Café	Restaurant	Pizza Place	Turkish Restaurant	Hotel	Park	Bar	Diner	Mountain	Smoke Shop
3	BAYRAKLI	Café	Bakery	Pide Place	Grocery Store	Restaurant	Turkish Restaurant	Nightclub	Park	Plaza	Gym
4	BERGAMA	Café	Historic Site	Hotel	Turkish Restaurant	Bar	Breakfast Spot	Dessert Shop	Lounge	Steakhouse	Food Court
5	BEYDAG	Park	Café	Plaza	Botanical Garden	Bakery	Convenience Store	Sandwich Place	Scenic Lookout	Business Service	Bus Station
6	BORNOVA	Café	Coffee Shop	Gym / Fitness Center	Dessert Shop	Meyhane	Restaurant	Hookah Bar	Park	Turkish Restaurant	Pharmacy
7	BUCA	Café	Gym / Fitness Center	Coffee Shop	Gym	Bakery	Fast Food Restaurant	Turkish Restaurant	Bar	Sandwich Place	Restaurant
8	CESME	Hotel	Café	Bed & Breakfast	Seafood Restaurant	Ice Cream Shop	Pub	Turkish Restaurant	Kumru Restaurant	Gym / Fitness Center	Lounge
9	CIGLI	Park	Café	Bakery	Convenience Store	Turkish Restaurant	Restaurant	Grocery Store	Buffet	Pide Place	Gym / Fitness Center

To find similar boroughs I choosed KMeans algorithm to clustering. K-Means algorithm is unsupervised and one of the most common cluster method.

First, I used elbow method to find optimal cluster value for my data.



Then I used silhouette score to validate my k value from elbow method. As we can see from graph silhouette score and elbow method suggest us to choose cluster value as 2. Therefor I decided to create 2 cluster.



Here is merged data with cluster labels as 1 and 0. When we examine our data we can see cluster 0 represents mostly “Seasonal & Touristic” boroughs and cluster 1 represents mostly “Cafes & Restaurants”.

	Borough	Latitude	Longitude	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	ALIAGA	38.802491	26.975580	1	Café	Restaurant	Bar	Seafood Restaurant	Gym / Fitness Center	Turkish Restaurant	Park	Burger Joint	Bagel Shop	Beach
1	BALCOVA	38.392353	27.048095	1	Clothing Store	Coffee Shop	Gym / Fitness Center	Café	Sporting Goods Shop	Dessert Shop	Restaurant	Burger Joint	Bakery	Furniture / Home Store
2	BAYINDIR	38.218886	27.648303	1	Café	Restaurant	Pizza Place	Turkish Restaurant	Hotel	Park	Bar	Diner	Mountain	Smoke Shop
3	BAYRAKLI	38.467495	27.167934	1	Café	Bakery	Pide Place	Grocery Store	Restaurant	Turkish Restaurant	Nightclub	Park	Plaza	Gym
4	BERGAMA	39.118371	27.173508	0	Café	Historic Site	Hotel	Turkish Restaurant	Bar	Breakfast Spot	Dessert Shop	Lounge	Steakhouse	Food Court

### Cluster 0: Seasonal & Touristic Boroughs

	Borough	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
4	BERGAMA	Café	Historic Site	Hotel	Turkish Restaurant	Bar	Breakfast Spot	Dessert Shop	Lounge	Steakhouse	Food Court
8	CESME	Hotel	Café	Bed & Breakfast	Seafood Restaurant	Ice Cream Shop	Pub	Turkish Restaurant	Kumru Restaurant	Gym / Fitness Center	Lounge
10	DIKILI	Beach	Café	Bar	Turkish Restaurant	Beach Bar	Park	Coffee Shop	Ice Cream Shop	Hotel	Seafood Restaurant
11	FOCA	Hotel	Beach	Café	Bed & Breakfast	Seafood Restaurant	Bar	Harbor / Marina	Convenience Store	Ice Cream Shop	Coffee Shop
15	KARABURUN	Beach	Hotel	Café	Seafood Restaurant	Bed & Breakfast	Restaurant	Dive Spot	Ice Cream Shop	Motel	Beach Bar
20	KONAK	Hotel	Dessert Shop	Söğüts Place	Historic Site	Coffee Shop	Turkish Restaurant	Music Store	Motorcycle Shop	Kofte Place	Café
26	SELCUK	Hotel	Café	Turkish Restaurant	Restaurant	Coffee Shop	Bed & Breakfast	Fast Food Restaurant	Park	Convenience Store	Auto Garage

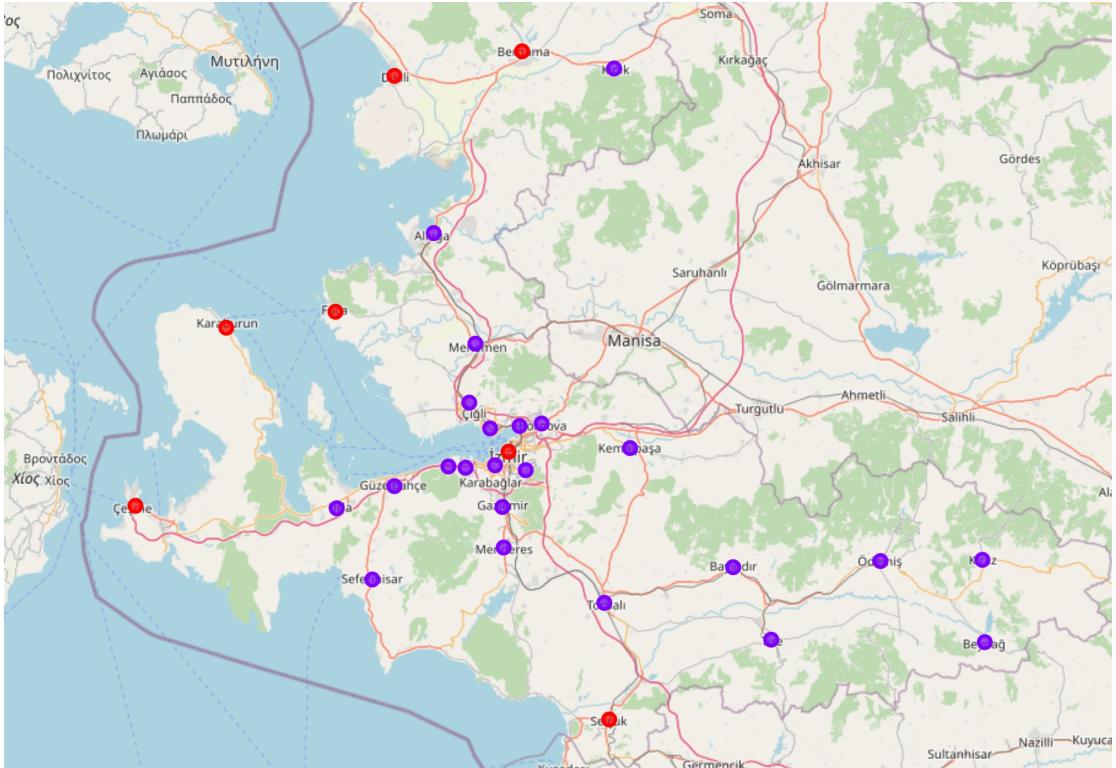
## Cluster 1: Mostly Cafes & Restaurant

	Borough	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	ALIAGA	Café	Restaurant	Bar	Seafood Restaurant	Gym / Fitness Center	Turkish Restaurant	Park	Burger Joint	Bagel Shop	Beach
1	BALCOVA	Clothing Store	Coffee Shop	Gym / Fitness Center	Café	Sporting Goods Shop	Dessert Shop	Restaurant	Burger Joint	Bakery	Furniture / Home Store
2	BAYINDIR	Café	Restaurant	Pizza Place	Turkish Restaurant	Hotel	Park	Bar	Diner	Mountain	Smoke Shop
3	BAYRAKLI	Café	Bakery	Pide Place	Grocery Store	Restaurant	Turkish Restaurant	Nightclub	Park	Plaza	Gym
5	BEYDAG	Park	Café	Plaza	Botanical Garden	Bakery	Convenience Store	Sandwich Place	Scenic Lookout	Business Service	Bus Station
6	BORNOVA	Café	Coffee Shop	Gym / Fitness Center	Dessert Shop	Meyhane	Restaurant	Hookah Bar	Park	Turkish Restaurant	Pharmacy
7	BUCA	Café	Gym / Fitness Center	Coffee Shop	Gym	Bakery	Fast Food Restaurant	Turkish Restaurant	Bar	Sandwich Place	Restaurant

Also we can see distribution of clustered boroughs from our map. Almost all of “Seasonal & Touristic Boroughs” close to sea as you can notice.

Red circles: Cluster 0, Seasonal & Touristic Boroughs

Purple circles: Cluster 1, Mostly Cafes & Restaurant



## 4. Results and Discussion

In this project, I focused on the most common venues in boroughs. In order to make an insightful data-driven decision. Of course there are other factors can be used to get more accurate results like, neighborhood clustering, rent rate, annual household income, population, etc. for running a restaurant in Izmir. Unfortunately I couldn't find these datas for Izmir. First I decided to use neighborhoods in central Izmir but APIs like Nominatim or ArcGis didn't return latitude and longitude values correctly. I spent couple of days to figure it out and it always returned incorrect values about

couple of neighborhoods. For future research paid APIs like Google may perform better to get these values. Also I need to overcome API calls limitation from Foursquare and this means I need paid account for it.

## 5. Conclusion

By looking at the clustered data, we can see that cluster 1 which is "Mostly Cafes & Restaurant" is the one that we are most interested in If stakeholder doesn't want to run a restaurant in seasonal and touristic places.

## 6. References

- [1] (<https://wikitravel.org/en/Izmir>)
- [2] (<https://en.wikipedia.org/wiki/Izmir>)
- [3] (<https://data.humdata.org/dataset/ac768a5c-db29-4872-9792-bd3179e29c45/resource/f721e3c5-26e4-4750-b237-68ac1ac2c08b/download/common-operational-dataset.xlsx>)
- [4] (<https://geopy.readthedocs.io/en/stable/>)
- [5] (<https://developer.foursquare.com/docs/places-api/>)