

# Where is the best place to open a cafe in Taksim / Istanbul?



# Business Problem

The aim of this project, the largest metropolitan area in Istanbul, is to find a suitable place to open a cafe in Turkey. I chose Taksim Square, the busiest and most crowded square in Istanbul, to open the cafe. However, Taksim Square is closed to car traffic. For this reason, customers have to use the subway to the cafe. The first requirement is that the new cafe is easily accessible to potential customers and more specifically is located near a metro station. If possible, the number of cafes currently available in an area should also be taken into account to avoid fierce competition. It can serve as the first starting point for locations to consider starting business.

In order to achieve the project goals, python geolocation libraries will be used together with the Foursquare API. Also, the K-Means machine learning clustering algorithm will be used to create clusters from similar candidate locations.

# Data

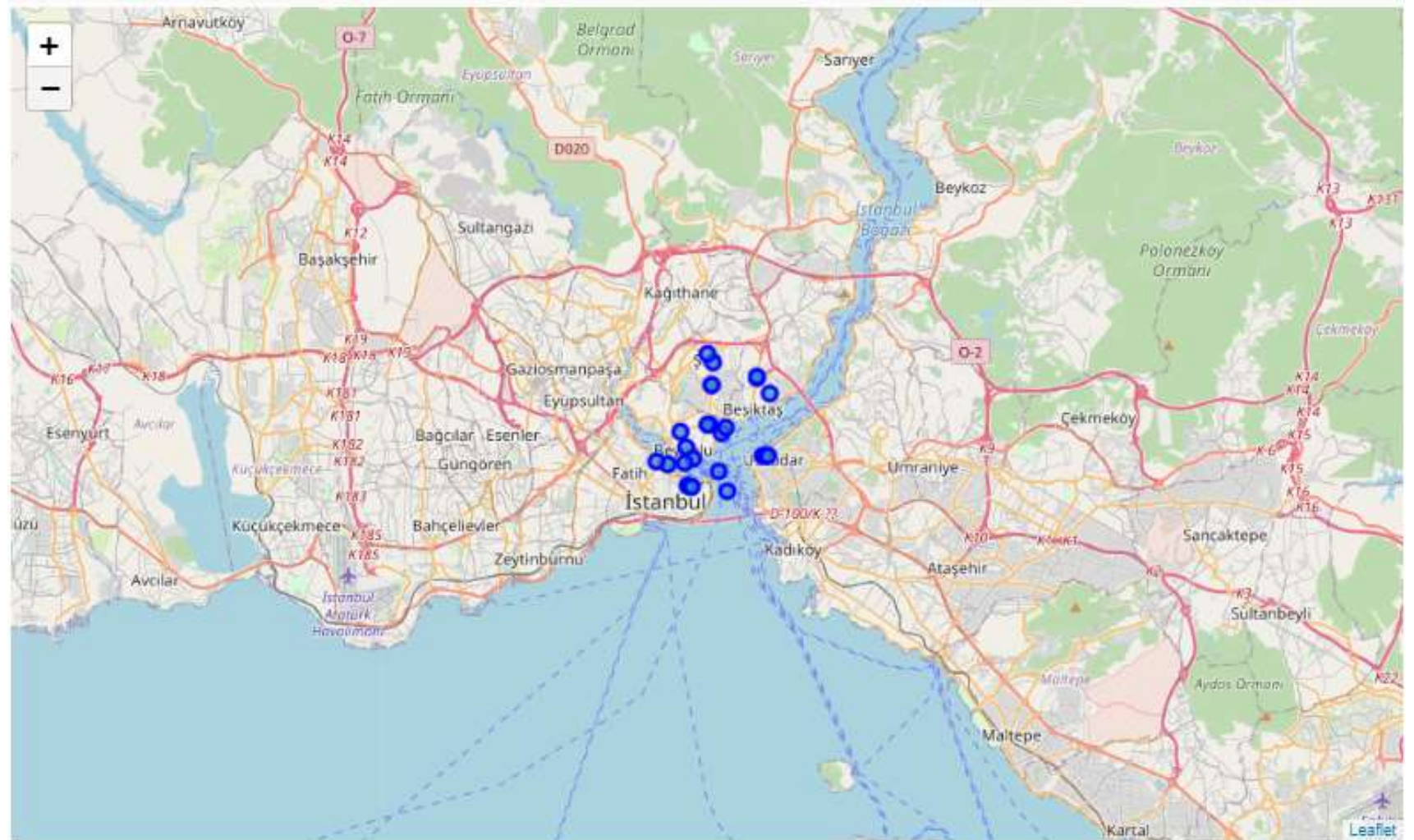
- Metro stations around Istanbul and Taksim Square Number of available cafes near each station.
- The distance to the nearest cafe will be used for each metro station.

To obtain the data, a combination of the geopy Python library and Foursquare API will be used:

1. "Taksim square" will be accepted as the center. I will get the geospatial coordinates using the Geopy library
2. Foursquare API, which has the coordinates of Taksim Square, will be used to obtain data for all metro stations within a 3 km radius.
3. To find existing cafes near subway stations, the Foursquare API will be used again for each station. I will get data for all cafes within a radius of 500 meters of each metro station Using the data collected, I will calculate the number of available cafes near each station. I will also be able to determine the minimum distance to a cafe for each metro station. This minimum distance from a cafe to each subway station, along with the number of cafes currently available near the station, will be used as input to the K-Means clustering algorithm to obtain clusters of areas (metro stations).

# Data transformation and cleaning

- 27 metro stations were found close to the Taksim location. (in a radius of 750 m)





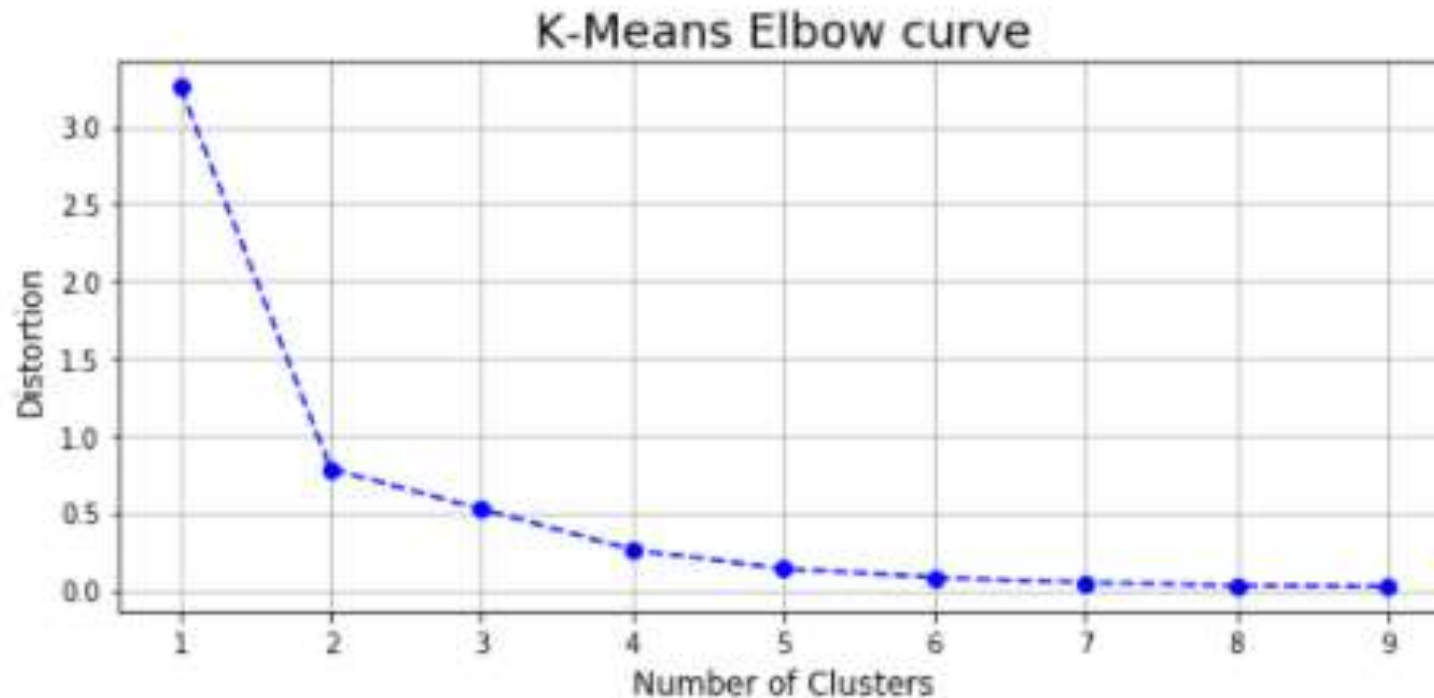
- There were 23 subcategories under the cafe category. The sub-categories of Café, Coffee Shop, Dessert Shop and Chocolate Shop were filtered because the cafe we were considering to open was about coffee and dessert.

	station	lat	lng	Venue	Venue Latitude	Venue Longitude	Distance from Station
Venue Category							
Bar	1	1	1	1	1	1	1
Bistro	2	2	2	2	2	2	2
Bookstore	11	11	11	11	11	11	11
Breakfast Spot	4	4	4	4	4	4	4
Café	2193	2193	2193	2193	2193	2193	2193
Chocolate Shop	2	2	2	2	2	2	2
Coffee Shop	80	80	80	80	80	80	80
Dessert Shop	3	3	3	3	3	3	3
Diner	1	1	1	1	1	1	1
Fast Food Restaurant	1	1	1	1	1	1	1
Flower Shop	2	2	2	2	2	2	2
Food & Drink Shop	3	3	3	3	3	3	3
Gourmet Shop	3	3	3	3	3	3	3
Hookah Bar	4	4	4	4	4	4	4
Hotel	13	13	13	13	13	13	13
Ice Cream Shop	3	3	3	3	3	3	3
Italian Restaurant	2	2	2	2	2	2	2
Kebab Restaurant	2	2	2	2	2	2	2
Lounge	2	2	2	2	2	2	2
Pub	2	2	2	2	2	2	2
Restaurant	25	25	25	25	25	25	25
Scenic Lookout	3	3	3	3	3	3	3
Tea Room	9	9	9	9	9	9	9

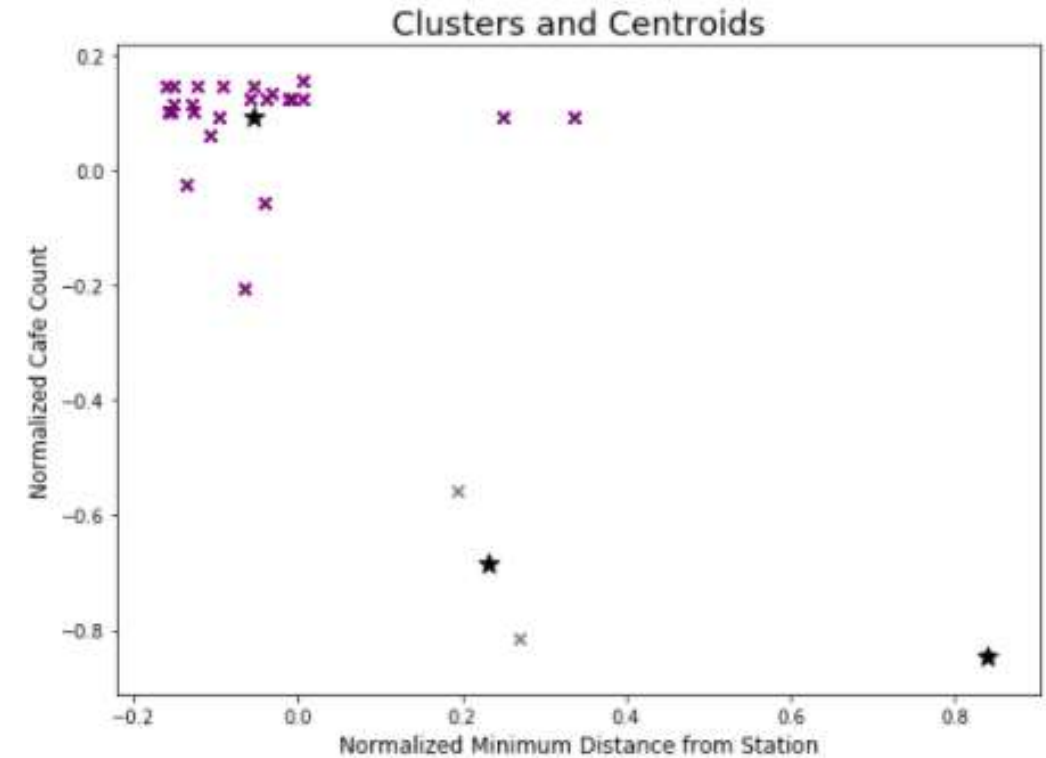
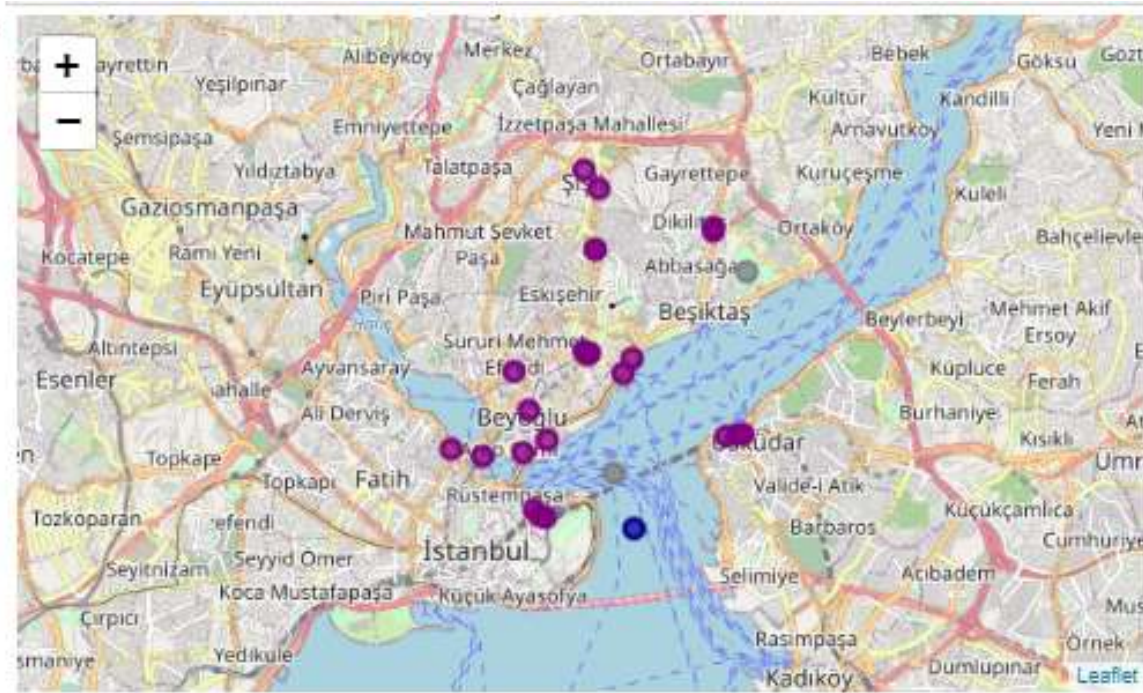
	station	lat	lng	Venue	Venue Latitude	Venue Longitude	Distance from Station
Venue Category							
Café	2193	2193	2193	2193	2193	2193	2193
Chocolate Shop	2	2	2	2	2	2	2
Coffee Shop	80	80	80	80	80	80	80
Dessert Shop	3	3	3	3	3	3	3

# K-Means clustering algoritim

- K-Means clustering algorithm was used to create clusters of stations.



# Results



As can be seen on the map, Cluster 2 has the highest potential. (purple color)