WHERE TO OPEN UP A NEW GYM IN ISTANBUL AFTER THE OUTBREAK ?

Source: Getty Images

1) INTRODUCTION

The unfortunate event, the Covid—19 pandemic, that our world has been experiencing for the last two months had many people locked down at their homes. This disastrous happening essentially prevents people from going to their gyms to relieve the accumulated stress and burn excess calories to get into shape.

Considering the negative atmosphere that Covid—19 brought upon us and the fact that most people can only do some light activities at their homes, individuals dream about getting back to their gyms as soon as possible. Even the folks with no gym history yearn for registering to a gym to neutralize the effect of stress-eating which they have been doing since the very start of this pandemic.

So, as a young entrepreneur living in Istanbul, who has only enough money for the equipment and the rent of a couple of months, I would like to have an aimed-shot rather than a random one. That is, I have to find the best district to open up my gym by considering various factors which may affect the success of my place one way or another.

2) DATA

Data forms the essence of this project as Istanbul is one of the biggest cities in the world with history that dates back to the first ages. That is to say, a person living in this city can't grasp all of its details just by herself / himself. Therefore, in this project, I make use of 3 different data sources to carry out a detailed district analysis:

Wikipedia: To get the district names of Istanbul properly as there are nearly 40 districts.

	Districts
0	Adalar
1	Arnavutköy
2	Ataşehir
3	Avcılar
4	Bağcılar
5	Bahçelievler
6	Bakırköy
7	Başakşehir

Figure 1.1: District names obtained and put into Python's DataFrame format

Open Street Map: To get the exact coordinates of each district to pass them to Foursquare API.

	Districts	Latitude	Longitude
0	Adalar	40.87625945	29.091027262109563
1	Arnavutköy	41.184182	28.7407289
2	Ataşehir	40.9847487	29.1067199
3	Avcılar	40.9801353	28.7175465
4	Bağcılar	41.0338992	28.8578982
5	Bahçelievler	41.0002895	28.8637451
6	Bakırköy	40.9835414	28.8679735
7	Başakşehir	41.0976935	28.8061626
8	Bayrampaşa	41.0357375	28.9122605
9	Beşiktaş	41.0428465	29.0075283
10	Beykoz	41.1239355	29.1083151

Figure 1.2: Appended Latitudes & Longitudes to districts with the help of Open Street Map DB

Foursquare: Foursquare database composes the main source of information for this project. Within this database, we look for places such as universities, high schools, shopping malls etc. which may have sizable effects on the success ratio.

	District	Latitude	Longitude	Universities	High Schools	Supplement Shops	Shopping Malls	Sporting Goods Shops	History Museums	Fitness Centers
0	Adalar	40.87625945	29.091027262109563	6.0	13.0	1.0	7.0	1.0	7.0	13.0
1	Arnavutköy	41.184182	28.7407289	4.0	27.0	1.0	47.0	2.0	0.0	29.0
2	Ataşehir	40.9847487	29.1067199	49.0	50.0	20.0	50.0	48.0	21.0	50.0
3	Avcılar	40.9801353	28.7175465	50.0	50.0	10.0	50.0	50.0	11.0	50.0
4	Bağcılar	41.0338992	28.8578982	47.0	50.0	25.0	50.0	49.0	32.0	50.0
5	Bahçelievler	41.0002895	28.8637451	46.0	50.0	22.0	50.0	49.0	30.0	50.0
6	Bakırköy	40.9835414	28.8679735	47.0	50.0	17.0	50.0	49.0	17.0	50.0
7	Başakşehir	41.0976935	28.8061626	44.0	50.0	3.0	50.0	43.0	10.0	50.0
8	Bayrampaşa	41.0357375	28.9122605	50.0	50.0	21.0	50.0	49.0	49.0	50.0
9	Beşiktaş	41.0428465	29.0075283	50.0	50.0	15.0	50.0	49.0	50.0	50.0
10	Beykoz	41.1239355	29.1083151	46.0	47.0	2.0	47.0	14.0	7.0	48.0
11	Beylikdüzü	41.0010788	28.642054	49.0	50.0	24.0	50.0	49.0	8.0	50.0
12	Beyoğlu	41.0284233	28.9736808	50.0	50.0	15.0	50.0	49.0	50.0	50.0

Figure 1.3: Venue details obtained from Foursquare DB added to the Istanbul Districts DataFrame

3) METHODOLOGY

First of all, I started the project by installing & importing the necessary libraries such as Pandas, Numpy, Folium, GeoPy, and BeautifulSoup. While the typical libraries like Pandas and Numpy allowed me to work with the data with ease, I think the most exceptional ones that worth mentioning here are Folium and BeautifulSoup. Essentially,

Folium allowed me to visualize the data on interactive maps that depicted both distribution of the venues as well as the clustering

BeautifulSoup is another noteworthy library that gives the comfort of querying data from obscure XML and HTML datasets with ease.

```
import pandas as pd #To be able to handle data with ease (by employing dataframes)
import numpy as np #To be able to handle data in vectors

pd.set_option('display.max_columns', None)
pd.set_option('display.max_rows', None)

import requests
from bs4 import BeautifulSoup

import matplotlib.cm as cm
import matplotlib.colors as colors

from sklearn.cluster import KMeans

!pip install geopy
from geopy.geocoders import Nominatim # convert an address into latitude and longitu

!pip install folium
import folium # Map tool for Data Visualization

print("All the required libraries have been installed.")
```

Figure 1.4: The libraries that are used

After obtaining data from the Wikipedia page with the help of BeautifulSoup, I needed it to do some data cleaning to have the data in a clear and meaningful format. In Fig 1.4, I give the raw and final format of the data respectively.

```
["Princes' Islands", 'Arnavutköy (district)', 'Ataşehir', 'Avcılar, Istanbul',
'Bağcılar', 'Bahçelievler', 'Bakırköy', 'Başakşehir', 'Bayrampaşa', 'Beşiktaş',
'Beykoz', 'Beylikdüzü', 'Beyoğlu', 'Büyükçekmece', 'Çatalca', 'Çekmeköy', 'Esen
ler', 'Esenyurt', 'Eyüp', 'Fatih', 'Gaziosmanpaşa', 'Güngören', 'Kadıköy', 'Kağ
ıthane', 'Kartal', 'Küçükçekmece', 'Maltepe, Istanbul', 'Pendik', 'Sancaktepe',
'Sarıyer', 'Silivri', 'Sultanbeyli', 'Sultangazi', 'Şile', 'Şişli', 'Tuzla, Ist
anbul', 'Ümraniye', 'Üsküdar', 'Zeytinburnu', None, None, None, None, None, Non
e]
Out[11]: 45
```

Out[13]:

	Districts
0	Adalar
1	Arnavutköy
2	Ataşehir
3	Avcılar
4	Bağcılar
5	Bahçelievler
6	Bakırköy
7	Başakşehir
8	Bayrampaşa
9	Beşiktaş
10	Beykoz

Figure 1.5: Data Cleaning and Formatting

Now that I have the district names in a clear format, it is time to obtain the coordinate values for each district of Istanbul. To get the exact coordinates of a district, we employ the Open Street Map database. After retrieving these coordinate values, I simply append them to the dataframe given in Fig1.5.

Out[16]:

	Districts	Latitude	Longitude
0	Adalar	40.87625945	29.091027262109563
1	Arnavutköy	41.184182	28.7407289
2	Ataşehir	40.9847487	29.1067199
3	Avcılar	40.9801353	28.7175465
4	Bağcılar	41.0338992	28.8578982
5	Bahçelievler	41.0002895	28.8637451
6	Bakırköy	40.9835414	28.8679735
7	Başakşehir	41.0976935	28.8061626
8	Bayrampaşa	41.0357375	28.9122605

Figure 1.6: District Data with Coordinate values

It is always a good idea to visualize the data you're working with to have a better and clear understanding on it to come up with the right approach. The districts of Istanbul is shown in the map below:

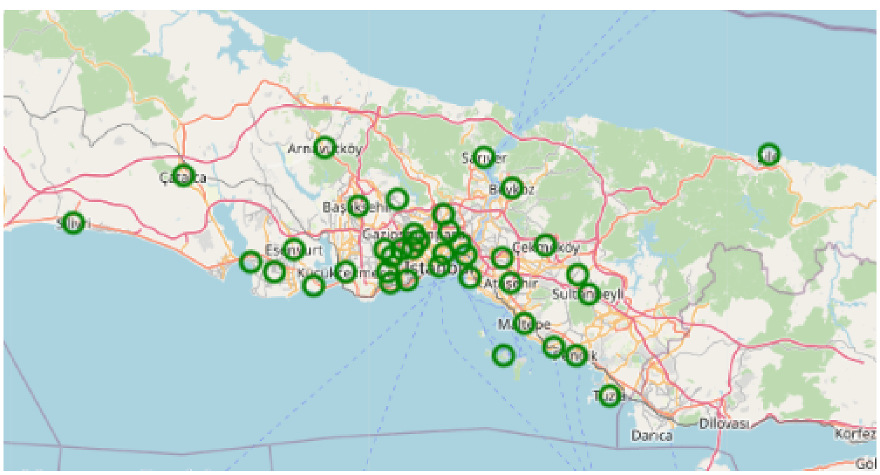

Figure 1.7: Districts of Istanbul (in total 39)

After obtaining the districts in a dataframe format with their latitudes & longitudes, it is time to get what kind of venues each district has (i.e. Universities, High Schools, Museums, etc.). To do so, I use the foursquare database as it is extensive and developer-friendly.

In this part of the project, I search for the following elements in each district since I believe that the amount of the listed places has a direct effect on the success ratio of a fitness center planned to be opened up in that particular district. In the following table, I try to explain the reasoning behind why we do data querying for each of these places for all of the districts.

* Universities

- If there are more universities around, this means that more interest will come from youth people

* High Schools

-If there are more high schools around, this means that more interest will come from youth people.

* Supplement Shops

- This directly shows that there is a high interest for sporting activities in a district.

* Shopping Malls

- A good indicator of people living close by, who have sufficient financial status.

* Sporting Goods Shops

- This directly shows that there is a high interest for sporting activities in a district.

* History Museums

- More history museums means more tourists, less locals

* Gyms / Fitness Centers

- More rivals implies less success ratio.

In the figure 1.8, I give the results of each query on the districts map formed by Folium.

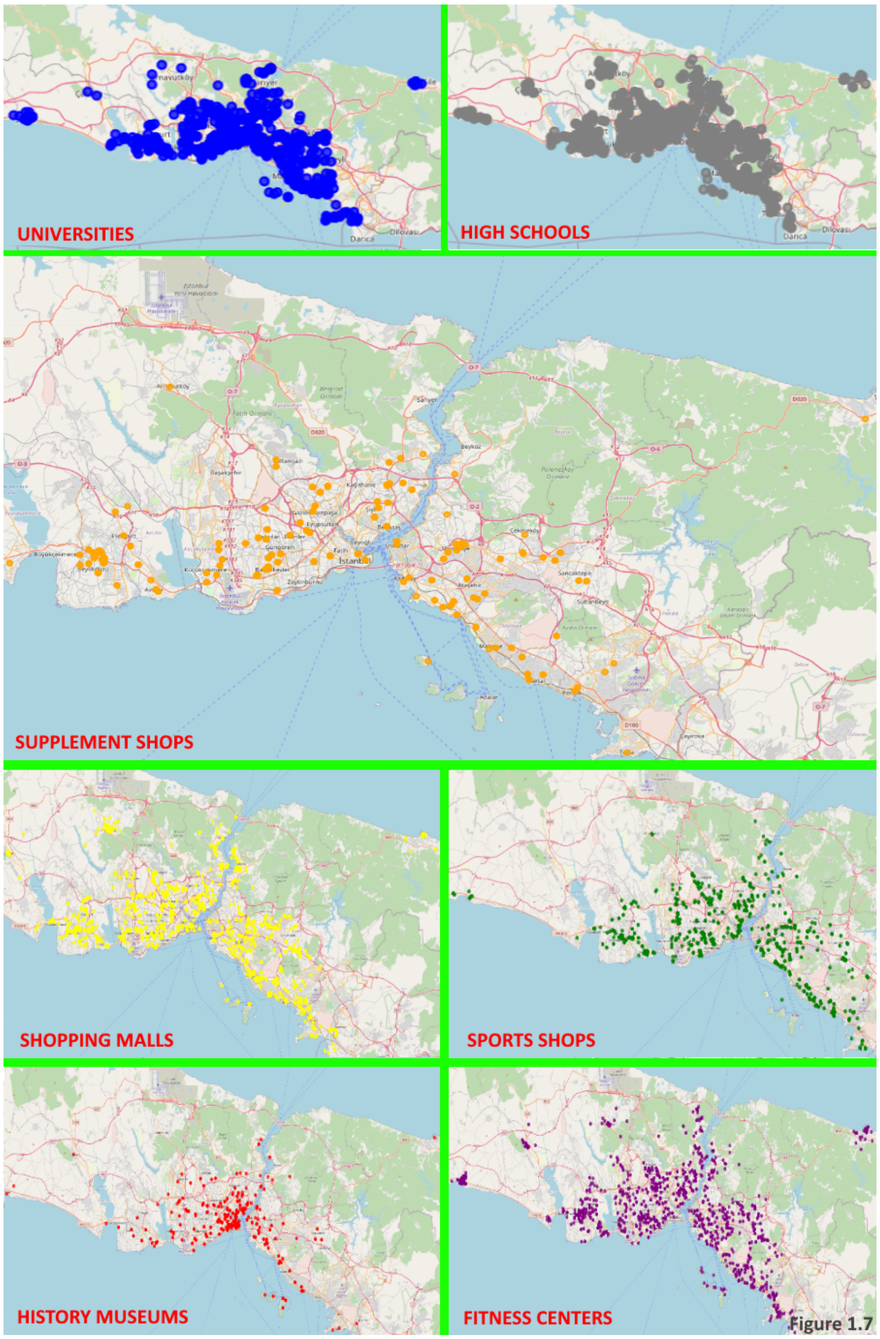

Figure 1.8: All the Foursquare queries shown in separate Folium maps.

To finalize the dataset, I have to add the results of the above queries to the districts of Istanbul dataframe in such a way that one can observe the frequency of each venue in a particular district with ease. By making use of Panda's basic commands, the dataframe given in figure 1.8 can be obtained without hustle.

Out[61]:

	District	Latitude	Longitude	Universities	High Schools	Supplement Shops	Shopping Malls	Sporting Goods Shops	History Museums	Fitness Centers
0	Adalar	40.87625945	29.091027262109563	6.0	13.0	1.0	7.0	1.0	7.0	13.0
1	Arnavutköy	41.184182	28.7407289	4.0	27.0	1.0	47.0	2.0	0.0	29.0
2	Ataşehir	40.9847487	29.1067199	49.0	50.0	20.0	50.0	48.0	21.0	50.0
3	Avcılar	40.9801353	28.7175465	50.0	50.0	10.0	50.0	50.0	11.0	50.0
4	Bağcılar	41.0338992	28.8578982	47.0	50.0	25.0	50.0	49.0	32.0	50.0
5	Bahçelievler	41.0002895	28.8637451	46.0	50.0	22.0	50.0	49.0	30.0	50.0
6	Bakırköy	40.9835414	28.8679735	47.0	50.0	17.0	50.0	49.0	17.0	50.0
7	Başakşehir	41.0976935	28.8061626	44.0	50.0	3.0	50.0	43.0	10.0	50.0
8	Bayrampaşa	41.0357375	28.9122605	50.0	50.0	21.0	50.0	49.0	49.0	50.0
9	Beşiktaş	41.0428465	29.0075283	50.0	50.0	15.0	50.0	49.0	50.0	50.0
10	Beykoz	41.1239355	29.1083151	46.0	47.0	2.0	47.0	14.0	7.0	48.0
11	Beylikdüzü	41.0010788	28.642054	49.0	50.0	24.0	50.0	49.0	8.0	50.0
12	Beyoğlu	41.0284233	28.9736808	50.0	50.0	15.0	50.0	49.0	50.0	50.0
13	Büyükçekmece	41.0156913	28.5955238	49.0	50.0	20.0	50.0	44.0	5.0	50.0
14	Çatalca	41.1435632	28.4619692	2.0	26.0	0.0	16.0	0.0	1.0	26.0
15	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

Figure 1.9: The finalized dataset that shows all the venue details for the districts

4) RESULTS

The results are obtained by basic scoring of each district. Essentially each kind of venue has a different score that is determined by the strength of effect each has on the success ratio. In figure 1.10, predetermined score values for the venues are given.

```
#
# University +3
scoreUni = 3
# High School +2
scoreHigh = 2
# Supplement Shop +2
scoreSupp = 2
# Shopping Malls +2
scoreMall = 2
# Sporting Goods Shop +2
scoreSportsShop = 2
# History Museums -3
scoreMsm = -2
# Gym/ Fitness Center -4
scoreGym = -4
```

Figure 1.10: Scoring of each venue

After employing the formula given in 1.11, we obtain the final table given in 1.12.

```
In [62]: dfScore = dfIstanbulDist[['District']]
         dfScore['Score'] = dfIstanbulDist['Universities'] * scoreUni + dfIstanbulDist['High Schools'] * scoreHigh + dfIstanbulDist['Supplement Shops'] * scoreSupp +
         dfScore = dfScore.sort_values(by=['Score'], ascending = False)
         dfScore
```

Figure 1.11: Scoring of formula

	District	Score
17	Esenyurt	283.0
11	Beylikdüzü	277.0
13	Büyükçekmece	265.0
27	Pendik	250.0
3	Avcılar	248.0
15	Çekmeköy	242.0
2	Ataşehir	241.0
26	Maltepe	241.0
24	Kartal	241.0
6	Bakırköy	239.0
25	Küçükçekmece	233.0
21	Güngören	227.0
4	Bağcılar	225.0
36	Ümraniye	222.0

Figure 1.12: The final result

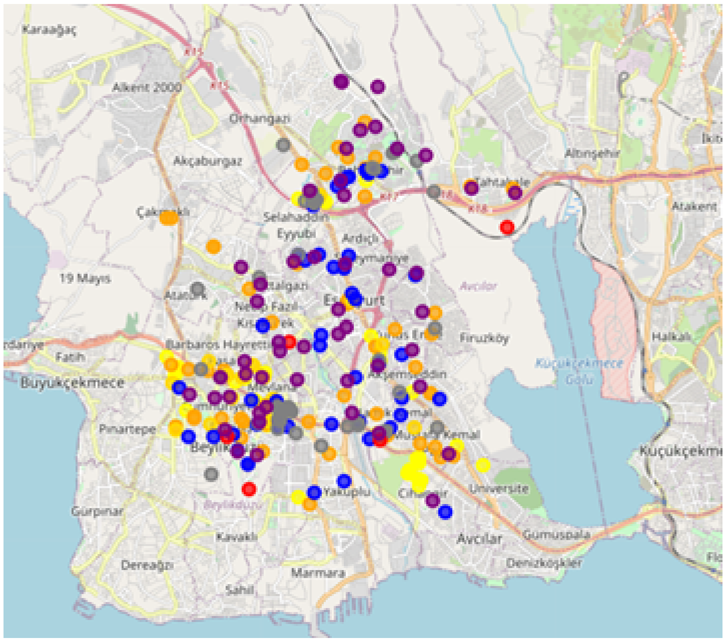

Figure 1.13: Map of Esenyurt with venues pointed

5) DISCUSSIONS

From the results, it is observed that Esenyurt, Beylikdüzü and Büyükçekmece are the best districts to open up a new fitness center. As a matter of fact, this result makes sense since all three of these districts are neither located in historical areas nor in financial centers. Therefore, it can be inferred that each of these districts are mostly residential areas with many high schools and universities.

The reason why Esenyurt emerged as the top choice from the data analysis is the lack of gyms in Esenyurt. So there is a huge incentive for me to go with Esenyurt to start my gym business there.

6) CONCLUSION

Even tough the results reflect logical scores, there is still a room for improvement. For instance, I didn’t take sports parks that the Istanbul Metropolitan Municipality opens up in each district into consideration. This may have a huge effect on the score as white collars might wanna do their activities in outdoor places rather than in a gym.