# 

# **IBM - Data Science Specialization**

# 

# Capstone Project – Final Report

# **Gym Location Recommendation**

# **in Istanbul**

# **Rahim RAHIMOV -2019**

Contents

[Description of the problem 2](#_Toc14104267)

[Description of the data 2](#_Toc14104268)

[Foursquare API: 3](#_Toc14104269)

[District- Neighborhood Population: 3](#_Toc14104270)

[Methodology 3](#_Toc14104271)

[Results 10](#_Toc14104272)

[Discussion 10](#_Toc14104273)

[Conclusion 10](#_Toc14104274)

[References 11](#_Toc14104275)

# **Description of the problem**

[Istanbul](http://worldpopulationreview.com/world-cities/istanbul-population/) is the largest city in [Turkey](http://worldpopulationreview.com/countries/turkey-population/) and one of the largest metropolises in the world. Istanbul had [an estimated population of over 15 million and has a population density of 2813 people per square kilometer.](http://www.dailysabah.com/nation/2016/01/28/turkeys-population-expanding-istanbul-still-most-crowded-city) The city is also one of the most favorite city in tourism destination. According to the information compiled from the report, 13.4 million foreign tourists came to Istanbul in 2018, up by 24 percent compared to 2017 when 10.8 million foreign visitors arrived in the city. Istanbul. The ministry of Tourism forecast a record number of tourists in 2019.

Considering population, density and a large number of visitors, the city is highly attractive for venue business.

This report will be beneficial for whom is looking to open a new gym as well as other kind of venue in Istanbul. I will try to recommend some specific neighborhoods of Istanbul based on neighborhood population and Foursquare API data. The main goal of this report will be searching a new gym location in a given district of Istanbul by considering neighborhoods population and Foursquare.

# Description of the data

In this section, we will talk about what data sources will be used to build our model. We will use the following source of data for this business case:

## Foursquare API:

We will need data about different venues in different neighborhoods of that specific district. In order to gain that information we will use **Foursquare API** to get and explore the different categories of venues of given specific district of Istanbul. Istanbul and its districts geographical coordinate’s data will be utilized as input for Foursquare API.

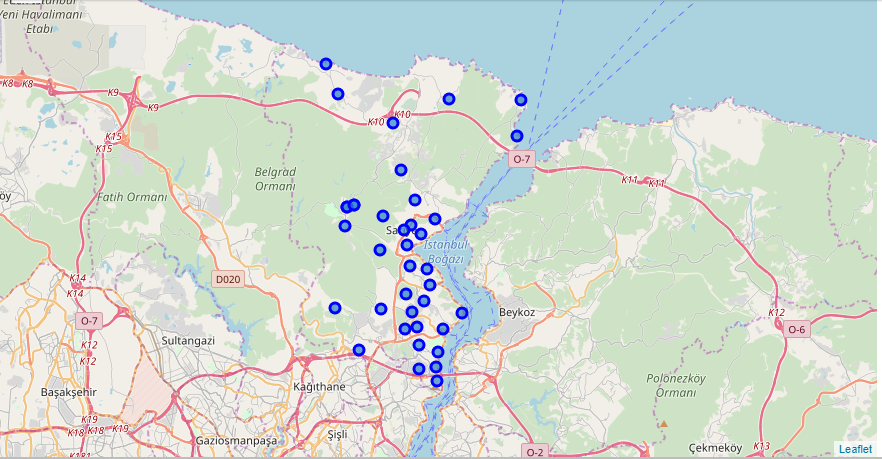
## District- Neighborhood Population

We will use <http://www.nufusune.com/> web site to get the districts, neighborhoods and neighborhood population of Istanbul. The data set contains all the districts, neighborhoods and neighborhoods population of Istanbul.

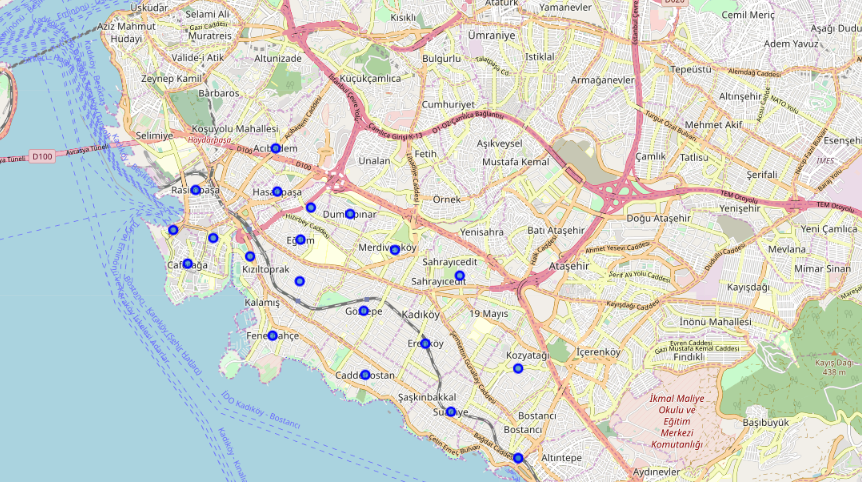
# Methodology

In this part of the report, we will describe the main components of our analysis. The aim was to give a specific neighborhood in a given district for a new gym. The recommendation will be based on the population and the number of gym in a given neighborhood.

Firstly, we need to get a list of districts, neighborhoods and neighborhood population in Istanbul. Fortunately, the list is available in Turkey District – Neighborhood Population website - <http://www.nufusune.com/> . l downloaded the data in csv format. However, this is just a list of names of districts, neighborhoods and neighborhood population. We need to get the geographical coordinates of each neighborhood in the form of latitude and longitude in order to be able to use Foursquare API. To do so, we will use the Geocoder package that will allow us to convert address into geographical coordinates in the form of latitude and longitude. After gathering the data, we will populate the data into pandas DataFrame and then visualize the neighborhoods in a map. This allow us to make sure that the neighborhoods are correctly plotted in the map. As there are 976 neighborhoods in Istanbul, it make very unreadable to plot all of them in a map. Therefore, we will select a few districts to check whether its respective neighborhoods are correctly plotted. Considering Sarıyer District which is one of the most populated district, we plot neighborhoods of Sarıyer district. The Python library called Folium is used to visualize geographically neighborhoods of Sarıyer:



We do the same visualization for Kadıköy district and neighborhoods.



Once we made sure that neighborhoods are correctly plotted in the map, we will explore gyms in some specific neighborhood. We will try to figure out what kind of gym exist in neighborhoods. We will address Foursquare API to gather this information. Foursquare API provides location data of venues and places as well as users visited these places and tips.For each neighborhood, we have chosen the radius to be 1000 meter. It means that we have asked Foursquare to find venues that are at most 1000 meter far from the center of the neighborhood. When the data is completely gathered, we will perform processing on that raw data to find our desirable features for each venue. Our main feature is the category of that venue. We will filter category data in order to get only gyms.

For example, considering selections below, we get this list from Foursquare API

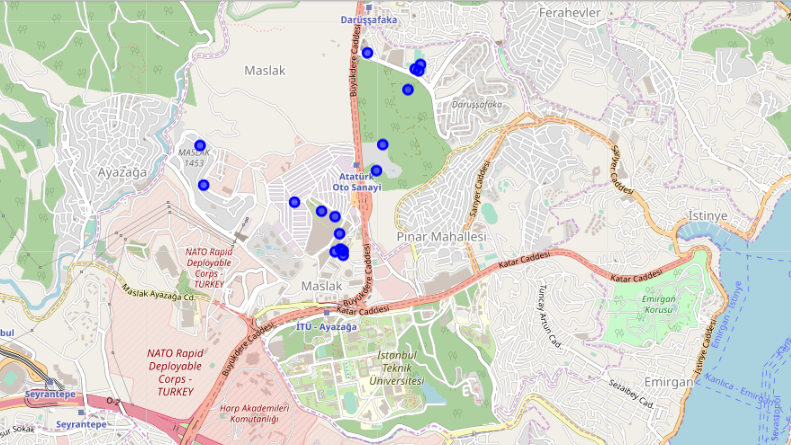
**District** = Sarıyer

**Neighborhood** = Maslak

**Radius** = 1 km – center of Maslak



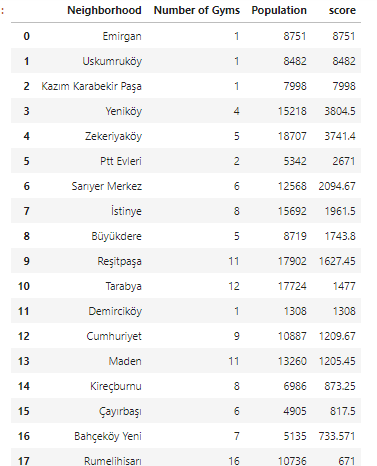
We plot the gyms in the map.



Considering data above and population of Maslak, we can calculate the person per gyms score.. The population of Maslak is 9567 as of 2018.

**Score Maslak = 9567 / 18 = 531**. There is one gym for every 531 people in Maslak neighborhood of Sarıyer. We expect this score would be high for a better location.

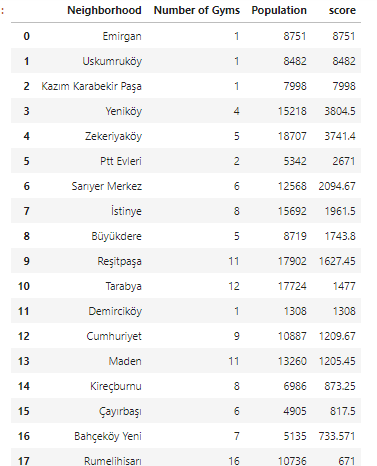
We can repeat the same process for all neighborhoods of Sarıyer district and we get dataset in descending order by score.



# 

# Results

In this section, we can discuss some results that we have got from the analysis. Considering Sarıyer district, we get dataset below:



Our main goal is to get optimum location-neighborhood for a new gym. As we can see, Emirgan has the highest score and there is only one gym in this area. So Emirgan is recommended for a new gym. But we should also consider Uskumruköy and Kazım Karabekir Paşa neighborhoods since scores are closed each other.

We are not limited to perform this model for only “Gym” case. We can perfectly apply this model for any other venues.

# 

# Discussion

When we use GeoCoder package for neighborhoods of Istanbul, it does not give perfectly fair location data to inform our model. If we have exact location data, we could improve our model. After picking the best neighborhood, there is a lot more work that needs to be done to actually find the best solution. However, this project gives a nice start to this process.

# Conclusion

The Foursquare API has been used to collect the information about the different venues present in the different neighborhoods of the city. Foursquare API sends only max 100 venue for a given neighborhood. If we have more than 100 venues for a neighborhood, we can get non-accurate results. Therefore, this analysis is performed on limited data. More amount of data is available; much better results could be gained in this model. We should also consider young population, age intervals of population and GDP per capita of each neighborhood.

In conclusion, this project is mainly focuses to help the businessmen, franchise owners as well as small businessmen looking to open a profitable gym as well as any other kind of venue in their city. We can perfectly to switch from “gym“ case to “Italian Pizza” case.

# References

IBM Data Science Capstone labs

The Foursqure API documentation

<http://www.nufusune.com/> - Population Data

Google Maps