OPEN A GYM BUSINESS

Abstract

This study has the goal to use Data Analysis techniques to support a Gym Company to identify the best place to open a new Gym.

Antonella Vaccina

Section 1- Introduction/Business Problem

Health and Fitness never goes out of style. Over the past few years people have discovered the importance of Physical Fitness at all ages: young guys can grow stronger, working mature men can take advantage to fight effects of sedentariness/stress and aged people can benefit and have a better quality of life.

A person who is fit is capable of living life to its fullest extent and is less prone to medical conditions as well. Lots of studies prove that Fitness is important to maintain optimal weight, to reduce the risk of cardiac and other health problems.

With this idea in mind, you can understand the importance to incorporate a regular exercise routine in everyone life and.... the Business around Gyms!

Recently we have seen the birth of lots of new Gyms and of a new growing trend: the low cost fitness sector. Gyms are managed in the form a franchising activity and are centrally opened by people that doesn't know the city, it's neighborhoods, population and venues in each neighborhood!

This brings us to the Business problem I was asked to solve. A big Fitness Company (the stakeholder) wants to open a new Gym in my city (Bari – South of Italy) but it needs to choose where to open it. I was reached out from a Manager in the Company that asked to work on complete report answering to the questions here:

- What is the best place to open a Gym?
- Should the place have some specific characteristics?
- Are there some appropriate tenants that it is better to have around it?
- Are there some neighborhoods preferred than others?
- Are there discriminant factors among Neighborhoods?
- Where are current Gyms located in Bari? Are they closed to some preferred venues? Do they follow a logic?

There is a combination of factors that will result in one particular location being the right fit for the new Gym!

Section 2 - Data

The request from the Fitness Company Manager was basically related to Exploring the city. After some investigation I decided to use the FourSquare data. They are a good data source because they provide access to over 75 million short tips from local experts all around the world and they offer lots of API to run queries against their data.

In order to validate the quality of Foursquare data as main data source, as first, I ran preliminary analysis.

Starting point is: do we have enough data about existing Gyms in Bari? To answer to this question I ran Search queries using FourSquare APIs to find data about Gyms in Bari. Bari is in Italy and some Gyms are named in Italian "Palestra". Then I had to run queries by using more than one search term (Gyms/Palestra/Fitness) in Italian and in English, merge results and remove some duplicates. I decided to consider the city without around district. The area that I explored is a circle having radius=3Km (Bari is a medium city having a population of 350,000 people).

As result of my preliminary analysis I found that in the Bari City Area there are 24 distinct Gyms distributed in different Neighborhoods of the city.

Palestra Beauty Line Palestra Palestra S.M.S Carlo Levi Palestra Mondo Fitness Palestra Lei Fit Palestra Per Donne Palestra Eagles Volley Palestra Ist. Margherita gymtonic palestra Palestra McFIT Palestra New Wave Palestra mea My Gym Wellness'N Gym Splendour Of The Seas Gym Iron Gym Empire Gym Wakeup fitness benessere estetica Fitness Lab Millenum Fitness Bari Eleven Fitness Club Eleven Health & Fitness Movida Dance & Fitness Club

I've also ran some exploratory queries to get evidence of venues around each Gym and verify that there are enough location data to create a map of the territory and match it with Gym positions.

After searching for venues around each gym within a radius of 500 meters I was able to find 210 distinct Venues divided in 69 different venue categories.

I have also run some queries to search for some specific venues that seem to be missed in the Foursquare API result: an example is "Offices". I understand that nobody write tips about offices but, an ad-hoc query returned 9 results. At the beginning I thought to add this information but, after having a look at the list I discovered they are not interesting.

It definitely seems that I have all the ingredients for my analysis!!! Let's get started!

Section 3 – Methodology

The methodology I used is focused on exploring areas around each existing gym. What I wanted to discover is if there is any discriminating factor that help to identify the best place.

After downloading the data and preparing them in the right format, I did some preliminary analysis to have a clear understanding of the data.

Then I used K-means clustering to organize existing Gyms in 5 different clusters based on Venue Categories and the frequency of each Venue in an area that have a radius of 500m centered on each Gym.

The K-means allowed to find hidden similarity among entities (Gyms) in the same cluster and differences among entities (Gyms) in different clusters.

For each cluster I identified top 5 most frequent venues. Finally, plotting clusters on a Folium map helped to visually identify Gyms in each cluster.

3.1 - Exploratory Data Analysis

As first, I had to review and prepare the data downloaded through FourSquare API. Then I ran some preliminary analysis to have an high level view of the data.

I created a pivot to count number of venues around each Gym. Results (Figure 1) are that some Gyms seem to have more venues than other. They are placed close each other and in a central area of Bari (Figure 2). I explored those Gyms more in detail later to verify if this is an indicator of a good place.

			Venue
Gym Name	Gym Latitude	Gym Longitude	
Palestra Beauty Line	41.123178	16.866368	30
Wakeup fitness benessere estetica	41.121942	16.867776	30
Wellness'N Gym	41.118334	16.864823	27
Palestra Mondo Fitness	41.118272	16.878119	21
Lei Fit Palestra Per Donne	41.115800	16.864400	20
Palestra	41.101277	16.854608	19
Palestra New Wave	41.103424	16.880884	19
Eleven Health & Fitness	41.099781	16.884221	19
gymtonic palestra	41.115307	16.867434	17
Millenum Fitness Bari	41.119448	16.881772	15
Palestra McFIT	41.105661	16.882551	13
Eleven Fitness Club	41.106880	16.881040	13
Palestra Ist. Margherita	41.107377	16.869998	12
Palestra mea	41.093096	16.863415	12
Splendour Of The Seas Gym	41.136918	16.868350	12
My Gym	41.108081	16.873876	11
Fitness Lab	41.113225	16.871750	11
Palestra Eagles Volley	41.107207	16.868638	10
Iron Gym	41.112778	16.870287	10
Gym	41.112899	16.869554	8
Palestra	41.136836	16.845952	7
Palestra S.M.S Carlo Levi	41.120552	16.847974	5
Empire Gym	41.100854	16.867457	5
Movida Dance & Fitness Club	41.107273	16.903282	4

Figure 1 - Number of Venues around each Gym



Figure 2 - Top 3 Gyms having highest number of Venues. They are in the City Center.

Then I analyzed Venue Frequencies by Venue Categories around each Gym. Then I used this table for K-means clustering.

3.2 - Machine Learning Techniques

To find similarity among Gyms in different Neighborhoods, I used K-Means clustering on the table below. Gyms are Entities to be clustered, Venue Categories are attributes and in each cell there is the frequency of each Category related to each Gym.

	Gym Name	Gym Latitude	Gym Longitude	Art Museum	Arts & Entertainment	Athletics & Sports	Bakery	Bar	Bed & Breakfast	Beer Garden	 Seafood Restaurant	Shopping Mall	Smoke Shop	
0	Eleven Fitness Club	41.106880	16.881040	0.0	0.0	0.0	0.0	0.0	0.0	0.0	 0.0	1.0	0.0	1.0
1	Eleven Health & Fitness	41.099781	16.884221	0.0	0.0	0.0	1.0	2.0	0.0	0.0	 0.0	0.0	0.0	0.0
2	Empire Gym	41.100854	16.867457	0.0	0.0	0.0	0.0	0.0	0.0	0.0	 0.0	0.0	0.0	0.0
3	Fitness Lab	41.113225	16.871750	1.0	0.0	0.0	1.0	0.0	0.0	0.0	 0.0	0.0	0.0	0.0
4	Gym	41.112899	16.869554	0.0	0.0	0.0	0.0	0.0	0.0	0.0	 0.0	0.0	0.0	0.0
5	Iron Gym	41.112778	16.870287	1.0	0.0	0.0	1.0	0.0	0.0	0.0	 0.0	0.0	0.0	0.0
6	Lei Fit Palestra Per Donne	41.115800	16.864400	0.0	0.0	0.0	0.0	0.0	0.0	1.0	 0.0	0.0	0.0	0.0
7	Millenum Fitness Bari	41.119448	16.881772	0.0	0.0	0.0	0.0	0.0	0.0	0.0	 0.0	0.0	0.0	0.0
8	Movida Dance & Fitness Club	41.107273	16.903282	0.0	0.0	0.0	0.0	0.0	0.0	0.0	 0.0	0.0	0.0	0.0
9	My Gym	41.108081	16.873876	1.0	0.0	0.0	0.0	0.0	0.0	0.0	 0.0	1.0	0.0	1.0
10	Palestra	41.101277	16.854608	0.0	0.0	0.0	0.0	1.0	0.0	0.0	 0.0	0.0	0.0	0.0

Figure 3 - Table used to run K-Means clustering

Clusters are distributed as below:

Cluster Labels 0 6 1 1 2 5 3 10 4 2

There is a big one, 3 medium clusters and one single Gym cluster. Plotting them on the map we see that they are distributed in different areas of the City.

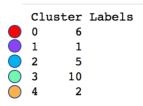




Figure 4 - Clusters distribution in different area of the City

The biggest cluster is the "green cluster" and it is placed close to the center of the city (but not in the center!). Cluster in yellow contains top 2 Gyms having the highest number of venues in their area (the City Center).

A detailed analysis of each cluster was done to explore main characteristics of each of them. What I did is to find top 5 Venues Categories for all Gyms in each cluster. Relevant ones are in red in Figures below.

Cluster 0					
Eleven Health & Fitness	Palestra	Palestra New Wave			
venue freq	venue freq	venue freq			
0 Pizza Place 6.0	0 Pizza Place 5.0	0 Café 4.0			
1 Café 3.0	1 Café 2.0	1 Gym 3.0			
2 Bar 2.0	2 Pub 2.0	2 Pizza Place 2.0			
3 Steakhouse 1.0	3 Hotel 1.0	3 Bar 2.0			
4 Sandwich Place 1.0	4 Bar 1.0	4 Brewery 1.0			
Empire Gym	Palestra McFIT	Palestra mea			
venue freq	venue freq	venue			
0 Pizza Place 3.0	0 Pizza Place 2.0	0 Pizza Place			
1 Café 1.0	1 Gym 2.0	1 Bar			
2 Electronics Store 1.0	2 Café 2.0	2 Hotel			
3 Platform 0.0	3 Breakfast Spot 1.0	3 Italian Restaurant			
4 Pier 0.0	4 Bar 1.0	4 Café			

Cluster 1 ----Splendour Of The Seas Gym---venue freq Boat or Ferry 9.0 Cruise 1.0

2 Harbor / Marina 1.0

Noodle House 0.0

Pier 1.0

3

```
Cluster 2

----Lei Fit Palestra Per Donne----
venue freq
Venue fre
```

```
Cluster 3
                                                                     ----Gym----
                                                                                                                                        ----Movida Dance & Fitness Club----
 ----Eleven Fitness Club----
                                       Venue freq venue rreq
Café 3.0 0 Pub 2.0 0 Pizza Place 1.0
Gym 2.0 1 Gastropub 1.0 1 Home Service 1.0
aurant 1.0 2 Italian Restaurant 1.0 2 Pet Store 1.0
Pub 1.0 3 Steakhouse 1.0 3 Discount Store 1.0
v Shop 1.0 4 Café 1.0 4 Noodle House 0.0
                                                                                                 venue freq
 0
 2 Gluten-free Restaurant 1.0
3 Pub 1.0
                          Pastry Shop 1.0
                                                                     ----Iron Gym----
                                                                                                                                        ----My Gym----
 ----Fitness Lab----
                                                                  venue freq venue

0 Art Museum 1.0 0 Art Museum

1 Steakhouse 1.0 1 Shopping Mall

2 Mobile Phone Shop 1.0 2 Japanese Restaurant

3 Movie Theater 1.0 3 Mobile Phone Shop

4 Outdoors & Recreation 1.0 4 Cluten-free Restaurant
                                                                                                                                                                                venue freq
                 venue freq
                                                                                                                                                                      Art Museum 1.0
 0 Art Museum 1.0
                                                                                                                                                                                                 1.0
         Steakhouse 1.0
2 Dessert Shop 1.0
3 Pub 1.0
                                                                                                                                                                                                 1.0
                                                                                                                                                                                                 1.0
           Sake Bar 1.0
 4
 ----Palestra----
                                                                     ----Palestra Ist. Margherita----

        venue
        freq
        venue
        freq
        venue
        freq

        0 Athletics & Sports
        1.0
        0 Japanese Restaurant
        3.0

        1 Italian Restaurant
        1.0
        1 Outdoors & Recreation
        1.0

        2 Bar
        1.0
        2 Shopping Mall
        1.0

        3 Pizza Place
        1.0
        3 Pizza Place
        1.0

        4 Lighthouse
        1.0
        4 Chinese Restaurant
        1.0

 ----Palestra Eagles Volley----
                                                                        ----Palestra S.M.S Carlo Levi----
venue freq
0 Japanese Restaurant 2.0
1 Pub 1.0
2 Market 1.0
3 Pizza Place 1.0
4 Mobile Phone Shop 1.0
                                                                                           venue freq
                                                                       0
                                                                                       Platform
                                                                                                             1.0
                                                                       Train Station .0
                                                                       2
                                                                                  Steakhouse
                                                                                                             1.0
                                                                            Pizza Place 1.0
                                                                      4
                                                                                    Food 1.0
```

```
Cluster 4
                          ----Palestra Beauty Line----
                                                   venue freq
                                                    Café 6.0
                          n
                                              Pizza Place
                          2 Vegetarian / Vegan Restaurant 2.0
                                      Italian Restaurant 2.0
                                       Seafood Restaurant
                          ----Wakeup fitness benessere estetica----
                                                    venue freq
                                       Café 5.0
Italian Restaurant 2.0
Plaza 2.0
                          0
                          1
                          2
                                              Coffee Shop 2.0
                          4 Vegetarian / Vegan Restaurant 2.0
```

Main differences around Gyms in each cluster are highlighted and they will be discussed in the Results section.

Section 4 - Results

Cluster 0 is in a not central area of the city. Venues around each Gym are "general" venues (Bar, Restaurants...) that we find in all Neighborhoods of the city! Conclusion is that Gym in this cluster are not informative because nothing differentiate this area from other areas of the city.

Cluster 1 is composed by one Gym that is special case. It is the only Gym close to the Ferry Boat.

Cluster 2 is more interesting. All Gyms are closed to the Train Station and Central Hotels.

But Cluster 3 is the most interesting. It is in a central area of the City that is different from the others. There are Venues not related to food-coffe-drink, but to cultural activities. There is Art Museum, Movie Theater, Shopping Mall. There are also venues related to food that are more "health fanatic" like Gluten Free Restaurants.

Finally Cluster 4 is the most central one. We find the highest number of Venues around Gyms but all of them are related to food. That's due to the central position of the are and is not a discriminating factor.

Section 5 - Discussion and Recommendations

How to read and interpret results described above? Different Venue Categories around cluster 3 is the factor to analyze and to interpret.

Have you never thought that a Gym doesn't need fast-food options staring your clients in the eye each and every time they leave the Gym? For this reason it is important to avoid setting up your Gym Business adjacent to Venues that conflict with your message!

For this reason, we recommend that Cluster 3 Area is the preferred one to open a new Gym. And the reason is in the particular venue categories: less categories related to food/drinks and more cultural categories are important to communicate the right message to Gym clients.

Section 6 - Conclusions

Data Analysis allows to discover hidden information that can help to solve Business problems. In this study we faced the problem to find the best place to open a new Gym in an unknown City. It is a common problem for franchising Companies that don't know the Cities.

My approach leverages FourSquare API power to explore the city; identify existing Gyms and where they are placed. I have identified differences in Venues and suggested the best place to open a new Gym based on the Venue Categories that is better to have around a new Gym in order to communicate an image of Health and tempt people to get a Gym member!