Marketing campaigns for FITAHOLIC

(New health drink brand)

Bijula Ratheesh

November 26, 2019

Table of Contents

1 Executive Summary

A newly formed healthy drink brand FITAHOLIC based out of Bronx, New York City, wants to introduce their product in the market and hence looking at a better marketing strategy to help them target the right customers and increase profitability.

Right customers for FITAHOLIC could be the fitness enthusiast, sports freaks and athletes, however they also want to spread the word in the neighborhood. To increase the market capitalization, we need to prioritize and fund the campaigns according to customer segment.

This project aims at creating segments based on venues of Bronx neighborhood for effective marketing. The two main segments considered are potential loyal customers and potential regular customers as below,

1. Fitness freaks – Mostly customers of gym, yoga or fitness center (potential loyal customers)
2. Sports freak – customers who goes to sports conditioning (potential regular customers)

For segment one, the idea is to offer fitness pamphlet and sample of juices every day for a week.
And, for segment two, a fitness pamphlet and sample of juices every alternate day for a week.

2 Data Sources

This Project needs information on New York City's Bronx neighborhood and its venue and these are extracted from the sources below:

1. Dataset containing New York City neighborhood is downloaded from the site, as GeoJson file.

https://data.cityofnewyork.us/City-Government/Neighborhood-Names-GIS/99bc-9p23

2. Venues of the neighborhood of Bronx, New York City is extracted from Foursquare API.

3 Literature Review – K-Means

Clustering is important and essential concept of data mining field used in various applications.

In Clustering, data are divided onto various classes. These classes represents some important features. Means, classes are the container of similar behavior of objects.

The objects which behave or are closer to each other are grouped in one class and who are far or non-similar are grouped in different class.

Clustering is a process of unsupervised learning. Highly superior clusters have high intra-class similarity and low inter-class similarity.

K-means clustering technique is a technique of clustering which is widely used. This algorithm is the most popular clustering tool that is used in scientific and industrial applications.

It is a method of cluster analysis which aims to partition observations into k clusters in which each observation belongs to the cluster with the nearest mean.

K-means clustering: K-Means clustering is unsupervised clustering technique in which data points are given as input and without and predefined result it generate clustering results. It is heavily used in scientific and industrial applications. For. E.g. clustering of similar gene expression, weather data, text classification etc.

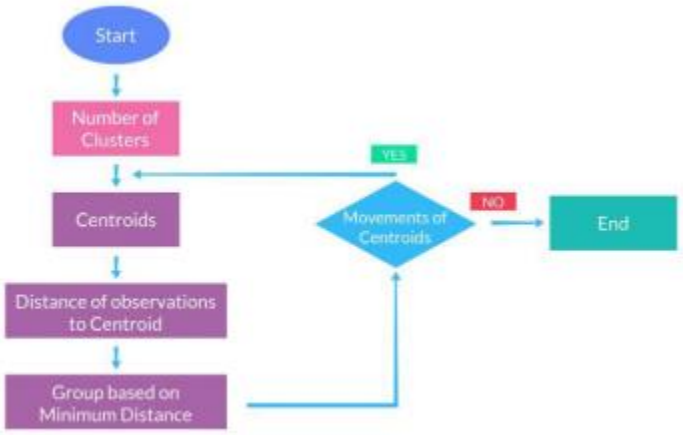


FIGURE COURTESY: HTTP://WWW.IJSRET.ORG/PDF/121888.PDF

The generic algorithm is very simple as presented in figure1.

1. Choose the number of clusters, K
2. Define the centroid ci as indices of the observation in each cluster and satisfies the below criteria

   C1 ∪ C2 ∪ . . . ∪ CK = {1, . . . , n}. In other words, each observation belongs to at least one of the K clusters.

   Ck ∩ Ck′ = ∅ for all k 6= k′. In other words, the clusters are non-overlapping no observation belongs to more than one cluster.
3. The idea behind K-means clustering is that a good clustering is one for which the within-cluster variation is as small as possible. The within-cluster variation for cluster Ck is a measure W(Ck) of the amount by which the observations within a cluster differ from each other. Hence we want to solve the problem.

$$\underset{C_1,\ldots,C_K}{\text{minimize}} \left\{ \sum_{k=1}^{K} W(C_k) \right\}$$

4. Within the cluster distance is defined by Euclidian distance and the formula can be written as

$$W(C_k) = \frac{1}{|C_k|} \sum_{i,i' \in C_k} \sum_{j=1}^{p} (x_{ij} - x_{i'j})^2$$

5. Now assign the new centroid as the mean of the points in the cluster.

6. Iterate steps 2-4 until the cluster assignments stop changing.

How to find the K for clustering?

Elbow Method

A. Compute clustering algorithm (e.g., k-means clustering) for different values of k. For instance, by varying k from 1 to 10 clusters.
B. For each k, calculate the total within-cluster sum of square (wss).
C. Plot the curve of wss according to the number of clusters k.
D. The location of a bend (knee) in the plot is generally considered as an indicator of the appropriate number of clusters.

Average silhouette method

A. Compute clustering algorithm (e.g., k-means clustering) for different values of k. For instance, by varying k from 1 to 10 clusters.
B. For each k, calculate the average silhouette of observations (avg.sil).
C. Plot the curve of avg.sil according to the number of clusters k.
D. The location of the maximum is considered as the appropriate number of clusters.

Gap Statistic Method

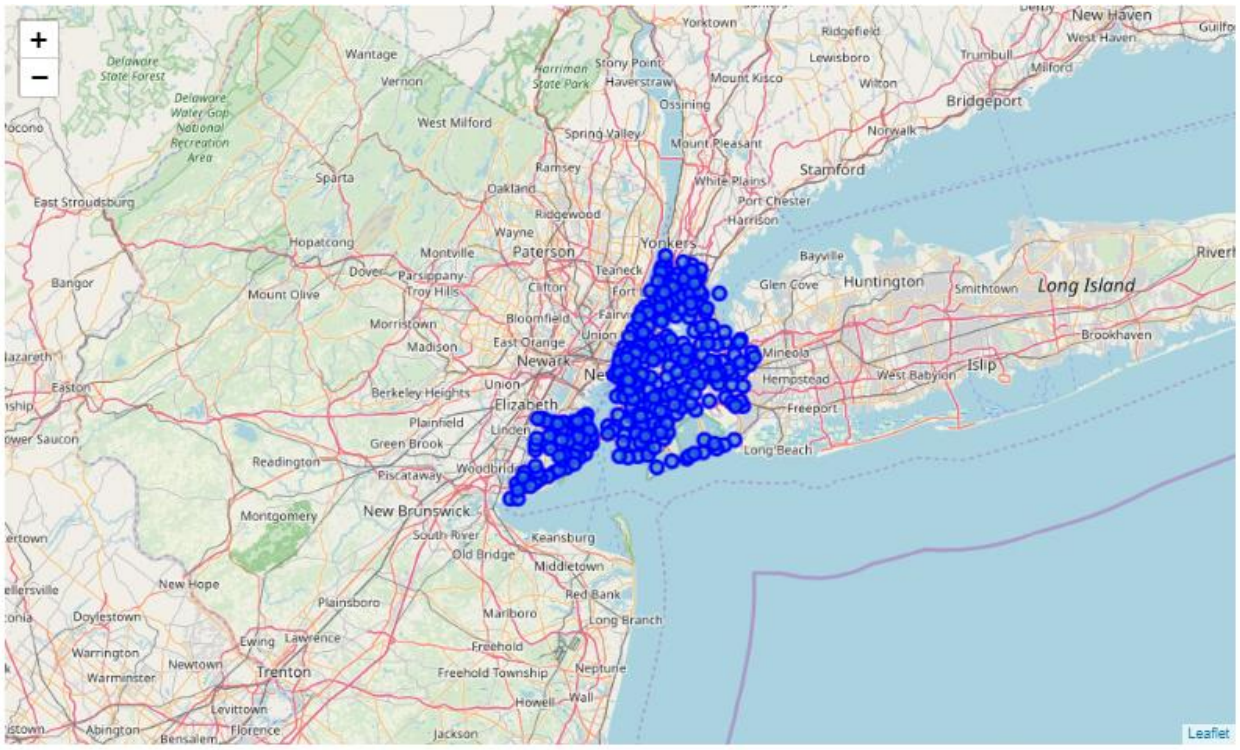
A. Cluster the observed data, varying the number of clusters from k = 1, …, kmax, and compute the corresponding total within intra-cluster variation Wk.
B. Generate B reference data sets with a random uniform distribution. Cluster each of these reference data sets with varying number of clusters k = 1, …, kmax, and compute the corresponding total within intra-cluster variation Wkb.
C. Compute the estimated gap statistic as the deviation of the observed Wk value from its expected value Wkb under the null hypothesis: $Gap(k)=1B\sum b=1Blog(W*kb)-log(Wk)Gap(k)=1B\sum b=1Blog(Wkb*)-log(Wk)$. Compute also the standard deviation of the statistics.
D. Choose the number of clusters as the smallest value of k such that the gap statistic is within one standard deviation of the gap at k+1: $Gap(k)\geq Gap(k+1)-sk+1$.

4 Methodology

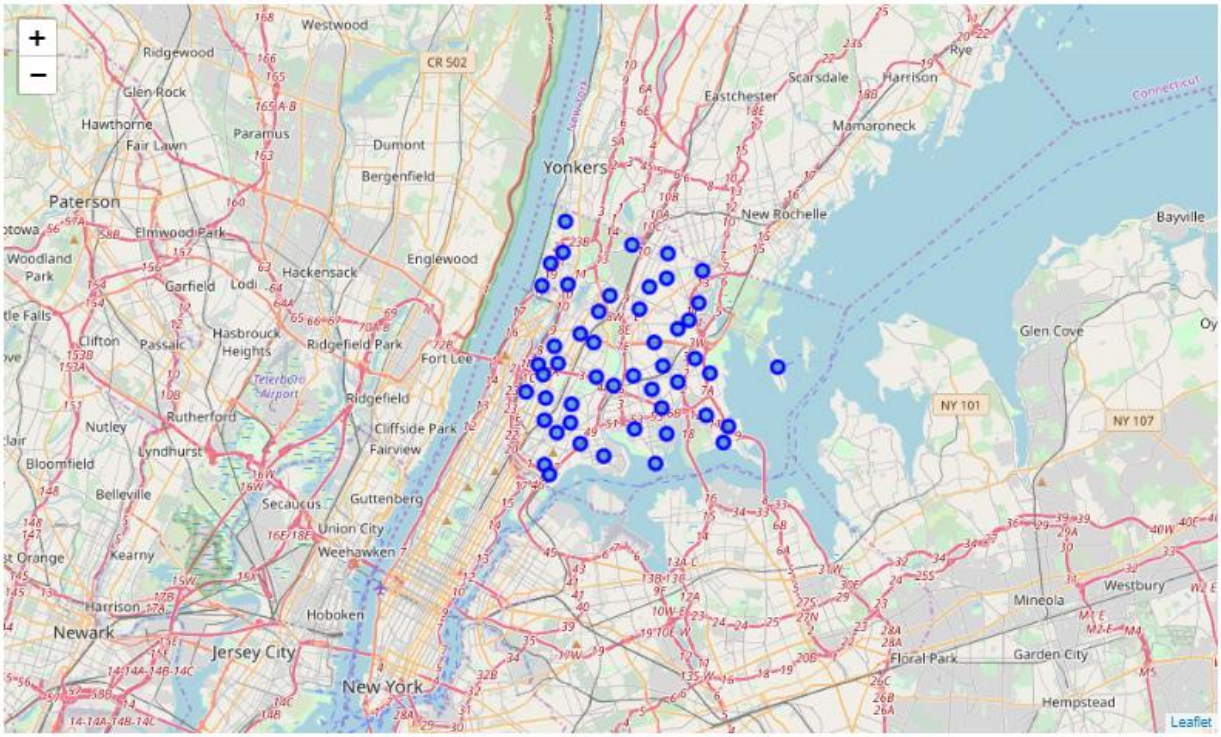
4.1 Exploratory Data Analysis

New York City and its neighborhood geo co-ordinates are available as geojson file, as discussed in data sources section.

The NYC neighborhood is shown in the figure below.



The marketing is targeted at Bronx and hence the co-ordinates are extracted only for Bronx and passed to Foursquare API to return a list of surrounding venues as per the given radius by using the explore end-point.



Health drink are promoted only in the areas near to any of the fitness centers as per the marketing strategy. Hence, venues related to sports and gym fitness activities are then extracted from the Bronx neighborhood venues as given in the table below.

TABLE 1: BRONX NEIGHBORHOOD VENUES

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Wakefield	40.894705	-73.847201	Lollipops Gelato	40.894123	-73.845892	Dessert Shop
1	Wakefield	40.894705	-73.847201	Rite Aid	40.896649	-73.844846	Pharmacy
2	Wakefield	40.894705	-73.847201	Carvel Ice Cream	40.890487	-73.848568	Ice Cream Shop
3	Wakefield	40.894705	-73.847201	Shell	40.894187	-73.845862	Gas Station
4	Wakefield	40.894705	-73.847201	Dunkin'	40.890459	-73.849089	Donut Shop
5	Wakefield	40.894705	-73.847201	SUBWAY	40.890656	-73.849192	Sandwich Place
6	Wakefield	40.894705	-73.847201	Koss Quick Wash	40.891281	-73.849904	Laundromat
7	Co-op City	40.874294	-73.829939	Dollar Tree	40.870125	-73.828989	Discount Store
8	Co-op City	40.874294	-73.829939	Rite Aid	40.870345	-73.828302	Pharmacy
9	Co-op City	40.874294	-73.829939	Mattress Firm	40.872234	-73.828607	Mattress Store
10	Co-op City	40.874294	-73.829939	Baskin Robbins	40.870045	-73.829578	Ice Cream Shop
11	Co-op City	40.874294	-73.829939	Bagels on Bartow	40.870280	-73.828611	Bagel Shop
12	Co-op City	40.874294	-73.829939	Capri II Pizza	40.876374	-73.829940	Pizza Place
13	Co-op City	40.874294	-73.829939	Arby's	40.870280	-73.828611	Fast Food Restaurant
14	Co-op City	40.874294	-73.829939	Food Universe Marketplace	40.876740	-73.828980	Grocery Store
15	Co-op City	40.874294	-73.829939	Townhouse Restaurant	40.876086	-73.828868	Restaurant
16	Co-op City	40.874294	-73.829939	truman track n field	40.874963	-73.830847	Baseball Field
17	Co-op City	40.874294	-73.829939	Guang Hui Chinese Restaurant	40.876651	-73.829092	Chinese Restaurant
18	Co-op City	40.874294	-73.829939	Pure Romance Parties By Amy	40.876792	-73.830476	Gift Shop
19	Co-op City	40.874294	-73.829939	MTA MaBSTOA Bus Bx23 / Bx26 / Bx28 / Q50 at As...	40.871387	-73.830646	Bus Station

The frequency of visits are then calculated by one-hot encoding and grouping by the venues and neighborhood. The table below shows a snapshot of the data.

TABLE 2: FREQUENCY OF VENUES OF BRONX

	Neighborhood	Baseball Field	Basketball Court	Gym	Gym / Fitness Center	Yoga Studio
0	Baychester	0.500000	0.000000	0.000000	0.500000	0.0
1	Bedford Park	1.000000	0.000000	0.000000	0.000000	0.0
2	Castle Hill	1.000000	0.000000	0.000000	0.000000	0.0
3	City Island	1.000000	0.000000	0.000000	0.000000	0.0
4	Claremont Village	0.000000	0.000000	1.000000	0.000000	0.0
5	Co-op City	0.500000	0.500000	0.000000	0.000000	0.0
6	Concourse Village	0.000000	0.000000	0.000000	1.000000	0.0
7	Fordham	0.000000	0.000000	0.333333	0.666667	0.0
8	High Bridge	0.000000	0.000000	1.000000	0.000000	0.0
9	Melrose	0.000000	0.000000	0.333333	0.666667	0.0
10	Morris Park	0.000000	0.000000	0.000000	0.000000	1.0
11	Mott Haven	0.333333	0.000000	0.666667	0.000000	0.0
12	Mount Eden	0.000000	0.333333	0.333333	0.333333	0.0
13	Olinville	0.000000	1.000000	0.000000	0.000000	0.0
14	Parkchester	0.000000	0.000000	0.500000	0.500000	0.0
15	Pelham Bay	0.000000	0.000000	0.333333	0.666667	0.0
16	Pelham Parkway	0.000000	0.000000	0.000000	1.000000	0.0
17	Riverdale	0.500000	0.000000	0.500000	0.000000	0.0
18	Soundview	0.000000	1.000000	0.000000	0.000000	0.0
19	West Farms	0.000000	1.000000	0.000000	0.000000	0.0
20	Westchester Square	0.000000	0.000000	1.000000	0.000000	0.0

This data is then used to cluster the neighborhood according to the frequency of visits. Clustering is performed using K-Means clustering algorithm (more can be read in the literature section).

Once the clusters are ready, they need to be examined according to the feasibility of marketing.

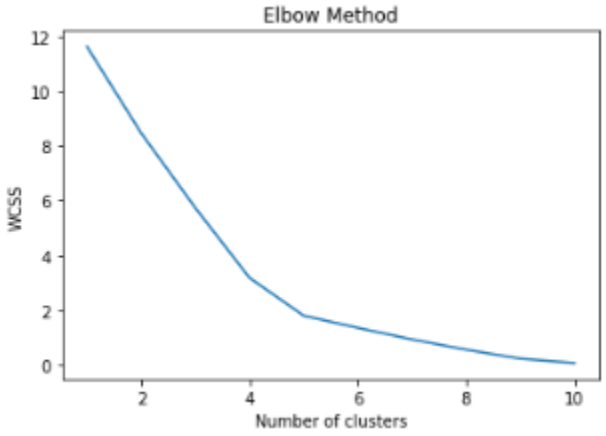
4.2 K-Means Clustering and Segmentation

K-Means algorithm is used to perform the clustering of the final dataset containing the frequency of visits, which was curated after the exploratory data analysis in the section above. The table below gives a glimpse of the frequency dataset for Bronx neighborhood for the 5 top venues.

TABLE 3 : LIST OF COMMON VENUES

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Baychester	Gym / Fitness Center	Baseball Field	Yoga Studio	Gym	Basketball Court
1	Bedford Park	Baseball Field	Yoga Studio	Gym / Fitness Center	Gym	Basketball Court
2	Castle Hill	Baseball Field	Yoga Studio	Gym / Fitness Center	Gym	Basketball Court
3	City Island	Baseball Field	Yoga Studio	Gym / Fitness Center	Gym	Basketball Court
4	Claremont Village	Gym	Yoga Studio	Gym / Fitness Center	Basketball Court	Baseball Field
5	Co-op City	Basketball Court	Baseball Field	Yoga Studio	Gym / Fitness Center	Gym
6	Concourse Village	Gym / Fitness Center	Yoga Studio	Gym	Basketball Court	Baseball Field
7	Fordham	Gym / Fitness Center	Gym	Yoga Studio	Basketball Court	Baseball Field
8	High Bridge	Gym	Yoga Studio	Gym / Fitness Center	Basketball Court	Baseball Field
9	Melrose	Gym / Fitness Center	Gym	Yoga Studio	Basketball Court	Baseball Field
10	Morris Park	Yoga Studio	Gym / Fitness Center	Gym	Basketball Court	Baseball Field
11	Mott Haven	Gym	Baseball Field	Yoga Studio	Gym / Fitness Center	Basketball Court
12	Mount Eden	Gym / Fitness Center	Gym	Basketball Court	Yoga Studio	Baseball Field
13	Olinville	Basketball Court	Yoga Studio	Gym / Fitness Center	Gym	Baseball Field
14	Parkchester	Gym / Fitness Center	Gym	Yoga Studio	Basketball Court	Baseball Field
15	Pelham Bay	Gym / Fitness Center	Gym	Yoga Studio	Basketball Court	Baseball Field
16	Pelham Parkway	Gym / Fitness Center	Yoga Studio	Gym	Basketball Court	Baseball Field
17	Riverdale	Gym	Baseball Field	Yoga Studio	Gym / Fitness Center	Basketball Court
18	Soundview	Basketball Court	Yoga Studio	Gym / Fitness Center	Gym	Baseball Field
19	West Farms	Basketball Court	Yoga Studio	Gym / Fitness Center	Gym	Baseball Field

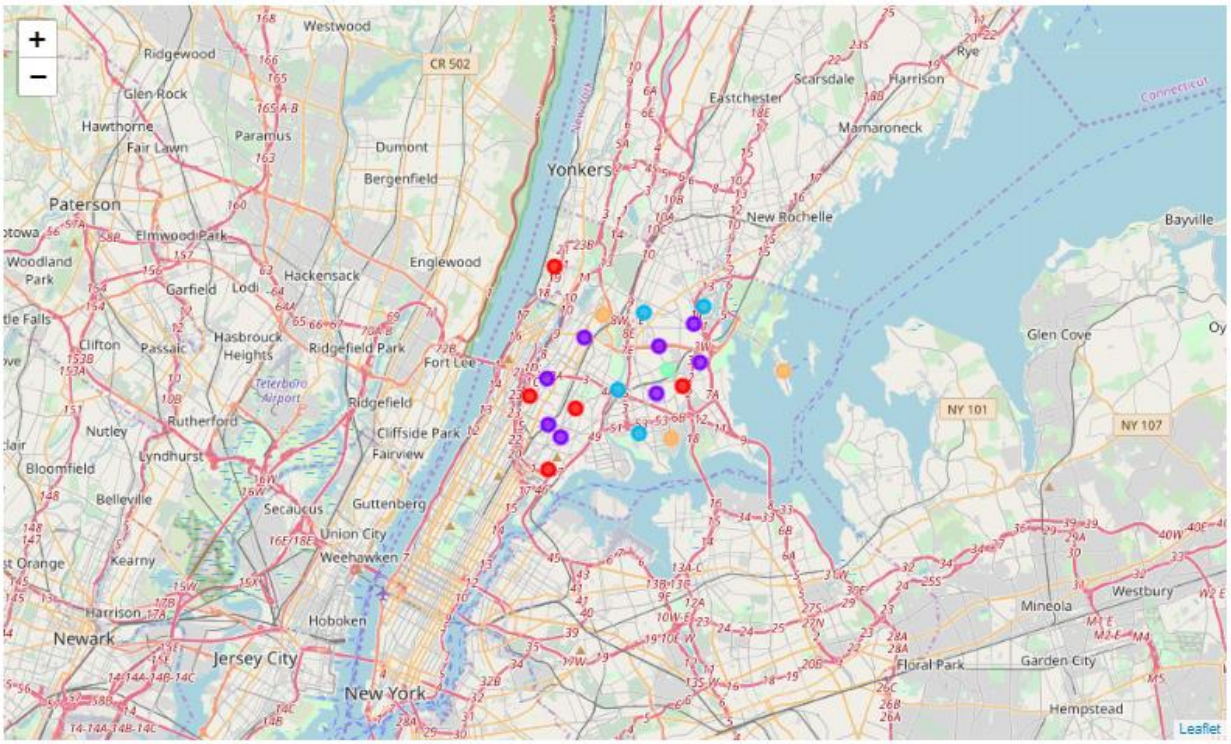
K for the K-Means is derived using the elbow method as mentioned in the literature review section. From the graph below the k value for the clustering is observed to be 5.



Hence we arrive at 5 clusters based on the common venue visit frequency as is given in the table and the map as below:

TABLE 4: CLUSTERS

	Cluster Labels	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	1	Baychester	Gym / Fitness Center	Baseball Field	Yoga Studio	Gym	Basketball Court
1	4	Bedford Park	Baseball Field	Yoga Studio	Gym / Fitness Center	Gym	Basketball Court
2	4	Castle Hill	Baseball Field	Yoga Studio	Gym / Fitness Center	Gym	Basketball Court
3	4	City Island	Baseball Field	Yoga Studio	Gym / Fitness Center	Gym	Basketball Court
4	0	Claremont Village	Gym	Yoga Studio	Gym / Fitness Center	Basketball Court	Baseball Field
5	2	Co-op City	Basketball Court	Baseball Field	Yoga Studio	Gym / Fitness Center	Gym
6	1	Concourse Village	Gym / Fitness Center	Yoga Studio	Gym	Basketball Court	Baseball Field
7	1	Fordham	Gym / Fitness Center	Gym	Yoga Studio	Basketball Court	Baseball Field
8	0	High Bridge	Gym	Yoga Studio	Gym / Fitness Center	Basketball Court	Baseball Field
9	1	Melrose	Gym / Fitness Center	Gym	Yoga Studio	Basketball Court	Baseball Field
10	3	Morris Park	Yoga Studio	Gym / Fitness Center	Gym	Basketball Court	Baseball Field
11	0	Mott Haven	Gym	Baseball Field	Yoga Studio	Gym / Fitness Center	Basketball Court
12	1	Mount Eden	Gym / Fitness Center	Gym	Basketball Court	Yoga Studio	Baseball Field
13	2	Olinville	Basketball Court	Yoga Studio	Gym / Fitness Center	Gym	Baseball Field
14	1	Parkchester	Gym / Fitness Center	Gym	Yoga Studio	Basketball Court	Baseball Field
15	1	Pelham Bay	Gym / Fitness Center	Gym	Yoga Studio	Basketball Court	Baseball Field
16	1	Pelham Parkway	Gym / Fitness Center	Yoga Studio	Gym	Basketball Court	Baseball Field
17	0	Riverdale	Gym	Baseball Field	Yoga Studio	Gym / Fitness Center	Basketball Court
18	2	Soundview	Basketball Court	Yoga Studio	Gym / Fitness Center	Gym	Baseball Field
19	2	West Farms	Basketball Court	Yoga Studio	Gym / Fitness Center	Gym	Baseball Field



5 Results

Now that we have arrived at the number clusters for the marketing campaign, let's explore each of the clusters.

5.1 Cluster 0

This cluster consists of all the neighborhood with common venue as Gym and also based on the commonalities among the second, third, fourth and fifth venue.

TABLE 5: CLUSTER 0

	Cluster Labels	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	Borough	Latitude	Longitude
4	0	Claremont Village	Gym	Yoga Studio	Gym / Fitness Center	Basketball Court	Baseball Field	Bronx	40.831428	-73.901199
8	0	High Bridge	Gym	Yoga Studio	Gym / Fitness Center	Basketball Court	Baseball Field	Bronx	40.836623	-73.926102
11	0	Mott Haven	Gym	Baseball Field	Yoga Studio	Gym / Fitness Center	Basketball Court	Bronx	40.806239	-73.916100
17	0	Riverdale	Gym	Baseball Field	Yoga Studio	Gym / Fitness Center	Basketball Court	Bronx	40.890834	-73.912585
20	0	Westchester Square	Gym	Yoga Studio	Gym / Fitness Center	Basketball Court	Baseball Field	Bronx	40.840619	-73.842194

5.2 Cluster 1

This cluster consists of all the neighborhood with common venue as Gym/Fitness Centre and also based on the commonalities among the second, third, fourth and fifth venue.

TABLE 6: CLUSTER 1

	Cluster Labels	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	Borough	Latitude	Longitude
0	1	Baychester	Gym / Fitness Center	Baseball Field	Yoga Studio	Gym	Basketball Court	Bronx	40.866858	-73.835798
6	1	Concourse Village	Gym / Fitness Center	Yoga Studio	Gym	Basketball Court	Baseball Field	Bronx	40.824780	-73.915847
7	1	Fordham	Gym / Fitness Center	Gym	Yoga Studio	Basketball Court	Baseball Field	Bronx	40.860997	-73.896427
9	1	Melrose	Gym / Fitness Center	Gym	Yoga Studio	Basketball Court	Baseball Field	Bronx	40.819754	-73.909422
12	1	Mount Eden	Gym / Fitness Center	Gym	Basketball Court	Yoga Studio	Baseball Field	Bronx	40.843826	-73.916556
14	1	Parkchester	Gym / Fitness Center	Gym	Yoga Studio	Basketball Court	Baseball Field	Bronx	40.837938	-73.856003
15	1	Pelham Bay	Gym / Fitness Center	Gym	Yoga Studio	Basketball Court	Baseball Field	Bronx	40.850641	-73.832074
16	1	Pelham Parkway	Gym / Fitness Center	Yoga Studio	Gym	Basketball Court	Baseball Field	Bronx	40.857413	-73.854756

5.3 Cluster 2

This cluster consists of all the neighborhood with common venue as sports Centre and also based on the commonalities among the second, third, fourth and fifth venue.

TABLE 7: CLUSTER 2

	Cluster Labels	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	Borough	Latitude	Longitude
5	2	Co-op City	Basketball Court	Baseball Field	Yoga Studio	Gym / Fitness Center	Gym	Bronx	40.874294	-73.829939
13	2	Olinville	Basketball Court	Yoga Studio	Gym / Fitness Center	Gym	Baseball Field	Bronx	40.871371	-73.863324
18	2	Soundview	Basketball Court	Yoga Studio	Gym / Fitness Center	Gym	Baseball Field	Bronx	40.821012	-73.865746
19	2	West Farms	Basketball Court	Yoga Studio	Gym / Fitness Center	Gym	Baseball Field	Bronx	40.839475	-73.877745

5.4 Cluster 3

This cluster consists of all the neighborhood with common venue as Yoga studio and it had only one neighborhood.

TABLE 8: CLUSTER 3

	Cluster Labels	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	Borough	Latitude	Longitude
10	3	Morris Park	Yoga Studio	Gym / Fitness Center	Gym	Basketball Court	Baseball Field	Bronx	40.847549	-73.850402

5.5 Cluster 4

This cluster consists of all the neighborhood with common venue as sports Centre and also based on the commonalities among the second, third, fourth and fifth venue.

TABLE 9: CLUSTER 4

	Cluster Labels	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	Borough	Latitude	Longitude
1	4	Bedford Park	Baseball Field	Yoga Studio	Gym / Fitness Center	Gym	Basketball Court	Bronx	40.870185	-73.885512
2	4	Castle Hill	Baseball Field	Yoga Studio	Gym / Fitness Center	Gym	Basketball Court	Bronx	40.819014	-73.848027
3	4	City Island	Baseball Field	Yoga Studio	Gym / Fitness Center	Gym	Basketball Court	Bronx	40.847247	-73.786488

6 Discussions and Conclusions

There were five clusters identified using k-means based on the frequency of visit.

However, FITOHOLIC has recommended the launch of only campaigns for two key segments of the population based on their loyalty towards fitness.

Hence the decision was to club the clusters according to their first common visit as below.

Potential Loyal Customer Segment – Cluster 0 + Cluster 1 + Cluster 3 and

Potential Regular Customer Segment – Cluster 2 + Cluster 4