

Product launching into the Fitness Industry

Rodrigo Calderón

May 9, 2020

1. Introduction

1.1 Background

The importance of creating innovative & sustainable products and introducing them to the market is growing in Mexico City and falling in the hands of small group of private entrepreneurs. One of the sectors that is growing the most with this product is the health industry, where numerous startups are trying to bring sustainability into the city. Also, there is awareness among the city's society of consuming products that are beneficial to health, are organic & have many nutrition benefits.

1.2 Problem

Okuilli is a new innovative product that substitutes your regular protein shake by a sustainable made, organic, high protein that is made from "tenebror molitor" a type of insect. This protein is 100% natural and there are very few insect protein products in the market.

This product is made outside of Mexico City, but the entrepreneur wants to introduce it into Mexico City's market. So, some of the questions to be solved will be:

1. Where in the city is best to introduce the product?
2. What neighborhoods are most likely to buy the product?
3. Which places that sell this type of products should he call to sell his products?

2. Data acquisition and cleaning

2.1 Data sources

We need to find data about Mexico City's boroughs and neighborhoods and also the type of venues in each one.

To solve this, the data that will be used to solve this problem will be of two sources.

1. First, for the list of boroughs which in Spanish they are called "Alcaldías" and neighborhoods which are called "Colonias" were found in the **government's open data**¹.

¹ <https://datos.cdmx.gob.mx/explore/dataset/alcaldias/>

2. To fulfill our objectives for positioning the product we will use the **Foursquare Location Data**, where we extracted venues to categorize.

2.2 Data cleaning

The dataset for the list of boroughs & neighborhood was downloaded into a csv file from the government's page and had 1811 rows with the following features: ALCALDIA, COLONIA & Latitude & Longitude. Then I looked for missing values of Latitude & Longitude, which there were 4 missing values (they were not relevant neighborhoods) so I decided to drop them.

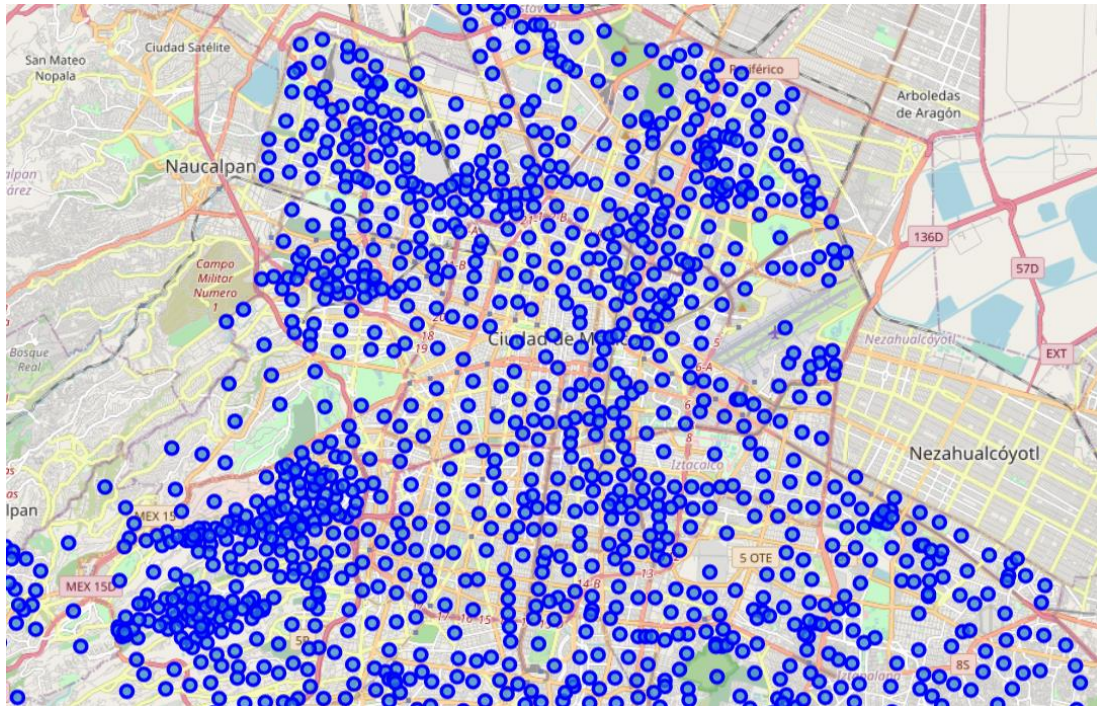


Figure 1: Mexico City's Neighborhoods

Then, I called the Foursquare API with the 1807 neighborhoods left to extract the following Venue features: Venue, Venue Latitude, Venue Longitude & Venue Category. After calling the Foursquare API, the data set contained 38,804 venues with 7 features.

2.3 Feature selection

After cleaning the data set and having all the venues for the neighborhoods I needed to categorize the venues to fulfill the objective: I only needed venues that are related to fitness, health and wellness, so I identified 15 venue categories and discarded the rest. Some of the categories selected were for example gyms, fitness centers, organic markets, parks, health & beauty, supplement stores and more.

After identifying these categories, the data was reduced to contain only these categories in it and only 3,051 venues remained in the data set.

Finally, after having the categorized data set that I needed, I used “one hot” encoding to the “Venue Category” feature, to have a data set that will proportion identification to the neighborhoods that had or hadn’t certain type of venue.

	Neighborhood	Athletics & Sports	Boxing Gym	Climbing Gym	College Gym	Gourmet Shop	Gym	Gym / Fitness Center	Gym Pool	Gymnastics Gym	Health & Beauty Service	Health Food Store	Ji
3	LOMAS DE CHAPULTEPEC	0	0	0	0	1	0	0	0	0	0	0	
4	LOMAS DE CHAPULTEPEC	0	0	0	0	0	0	0	0	0	0	0	
8	LOMAS DE CHAPULTEPEC	0	0	0	0	0	0	0	0	0	0	0	
23	LOMAS DE REFORMA (LOMAS DE CHAPULTEPEC)	0	0	0	0	0	0	1	0	0	0	0	
39	DEL BOSQUE (POLANCO)	0	0	0	0	0	0	1	0	0	0	0	

Figure 2: Data Set Example after One Hot Encoding

3. Exploratory Data Analysis

3.1 Top Boroughs & Neighborhoods

The first task was to identify the top boroughs and neighborhoods that had the most well-ness category venues. We found out the following:

Alcaldia	#
COYOACAN	479
IZTAPALAPA	314
ALVARO OBREGON	304
GUSTAVO A. MADERO	302
BENITO JUAREZ	290
TLALPAN	271
MIGUEL HIDALGO	261

Figure 3: Top Boroughs

Alcaldias	Colonias	Venue Categor
MIGUEL HIDALGO	RINCON DEL BOSQUE	16
BENITO JUAREZ	CENTRO URBANO PRESIDENTE ALEMAN (U HAB)	14
MIGUEL HIDALGO	BOSQUES DE CHAPULTEPEC (POLANCO)	13
ALVARO OBREGON	MIGUEL HIDALGO	13
CUAUHTEMOC	ROMA NORTE	13
TLALPAN	FLORESTA-PRADO-VERGEL COAPA	12
BENITO JUAREZ	DEL VALLE II	12
ALVARO OBREGON	MERCED GOMEZ	12
COYOACAN	LOS SAUCES (FRACC)	12
COYOACAN	LOS GIRASOLES II	11
CUAJIMALPA DE MORELOS	BOSQUES DE LAS LOMAS	11
BENITO JUAREZ	ZACAHUITZCO	10
COYOACAN	LOS GIRASOLES III	10
COYOACAN	HACIENDAS DE COYOACAN (FRACC)	10
COYOACAN	LOS GIRASOLES I	10
BENITO JUAREZ	ACTIPAN	10
CUAUHTEMOC	HIPODROMO I	10
MIGUEL HIDALGO	BOSQUE DE CHAPULTEPEC I, II Y III SECCIONES	10

Figure 4: Top Neighborhoods

The boroughs of Coyoacan, Iztapalapa, Alvaro Obregon, Gustavo Madero, Benito Juarez, Tlalpan & Miguel Hidalgo make more than 60% of the venues. These boroughs also have the neighborhoods with the top venues, which we can see above in figure 4.

3.2 Top Neighborhood with only Supplement Shops

We found out that there were only 13 unique supplement, organic & health shops and are distributed in the following neighborhoods table:

Colonia	Count
BOSQUES DE CHAPULTEPEC (POLANCO)	2
CHAPULTEPEC MORALES (POLANCO)	2
AMPLIACION SAN FRANCISCO CULHUACAN (EJ)	1
CENTRO URBANO PRESIDENTE ALEMAN (U HAB)	1
CHIMALISTAC	1
CONDESA	1
DEL VALLE VII	1
EDUCACION	1
EX EJIDO DE CHURUBUSCO	1
FERROCARRILERA INSURGENTES	1
HEROES DE PADIARNA I	1
INDUSTRIAL I	1
INSURGENTES MIXCOAC	1
JOSE MARIA MORELOS Y PAVON I (U HAB)	1
JOSE MARIA MORELOS Y PAVON II (U HAB)	1
LA NORIA TEPEPAN	1
NARVARTE V	1
PETROLERA TAXQUEA	1
RINCON DEL BOSQUE	1
ROMA NORTE	1
ROMERO DE TERREROS (COND)	1
SAN ANGEL	1
SAN JUAN	1
SANTA CATARINA (BARR)	1
SANTA MARTHA DEL SUR	1
TETENCO (PJE)	1
TORRES DE SAN JUAN 1B (U HAB)	1
VILLA XOCHIMILCO (U HAB)	1

Figure 5: Neighborhoods with well-ness type shops

This gives us a very valuable insight: there is no more than two well-ness shops in a neighborhood. We can deduce that there is a good opportunity for the product to be introduced to this industry, because the market is very big and the offer of this products is limited to only a few shops.

4. Clustering

I used k-means, an unsupervised model to try to find clusters of data and patterns to pick the correct places that have the opportunity for the product to be well positioned. I should choose clusters where there are many gyms, fitness centers & health-supplement stores. Also, the importance of having many of these venues is a good sign to choose the best cluster.

4.1 The clusters

First, I tried using 5 clusters of data. After viewing them in a map and tables, there were no clear distinctions of patterns in those clusters. The clusters were incremented by one until we stopped choosing a k of 8. Here we could find a clear clustering pattern:

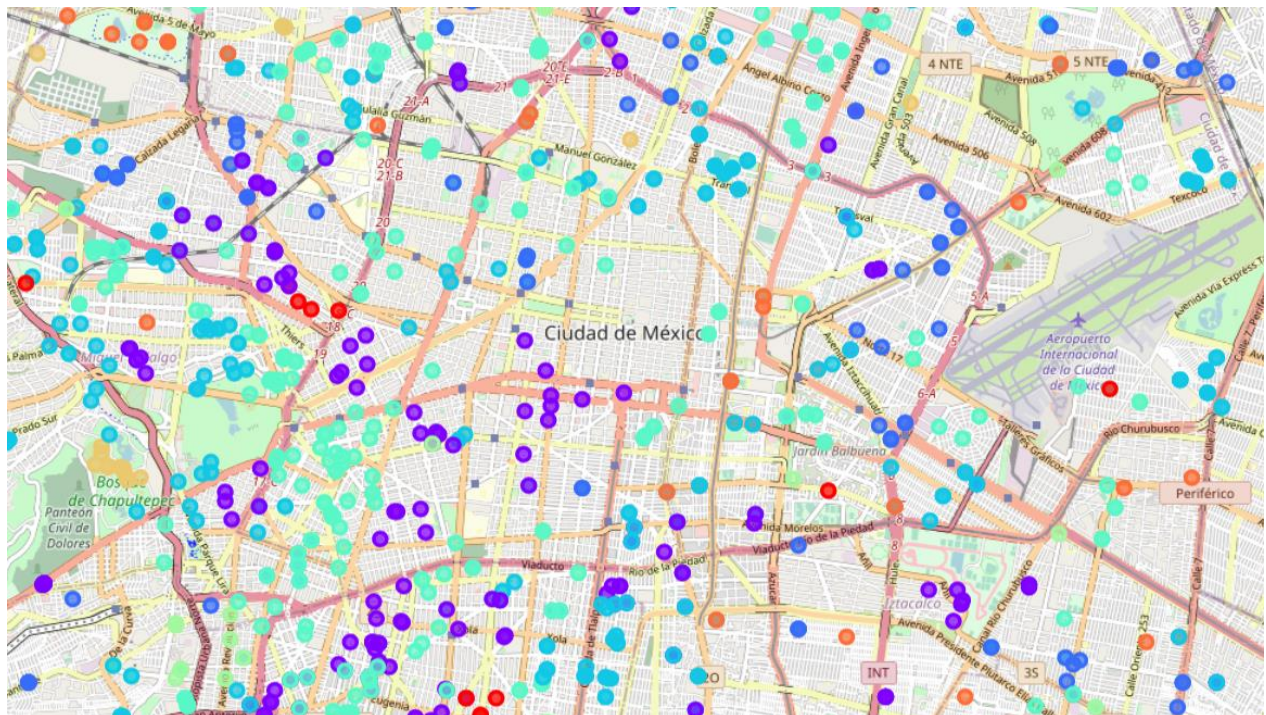


Figure 6: Clusters

4.2 Results & Discussions

The results can be seen in the following figure:

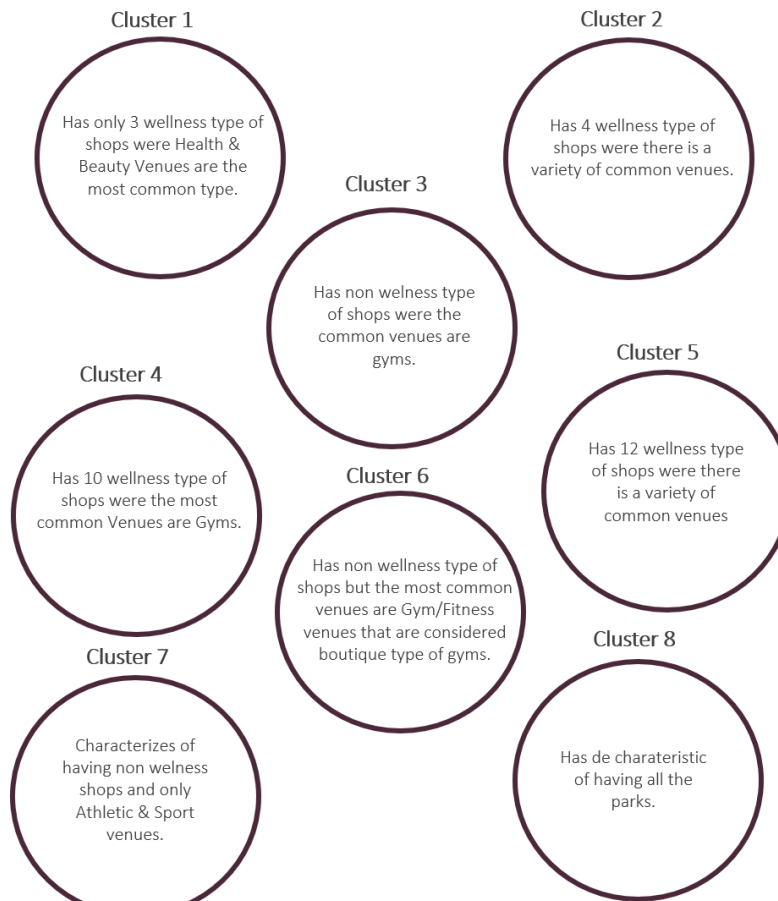


Figure 7: Clusters of Dataset

The results are the following:

- Best cluster to call & sell product: Cluster 5
- Best cluster that have opportunity to open a store: Cluster 6

I recommend the following: Choose to sell in stores in the top neighborhoods of the “Alcadías” of Cluster number 5, were wellness shops represent only 0.87% of the venues. This is an opportunity because there is probably more market share in a place where there is a lot of venues and not so many products offered.

Also, I recommend choosing to open store in Alvaro Obregon or Iztapalapa, were the boutique gyms are the trend (see Cluster 6), which is an opportunity to position an innovative protein product as there are no nearby wellness shops.

5. Conclusions & Future Directions

In this study, we can conclude the following: The product has a good opportunity to be received by consumers because we found out there are little wellness shops compared to the total other fitness venues there are in the city. The clustering technique was an effective technique because it identified patterns in data that delivered opportunities for selling the product or maybe opening a new store with it. We then found out the boroughs and neighborhoods effectively where we should position our product or open a new store with it.

Next steps may be in choosing to sell the product to a store or open one. This analysis requires a more accurate decision with cost and capital variables.