IBM CAPSTONE PROJECT

Setting up a Business

Comparing and evaluating neighborhoods in Guadalajara, Mexico to open a Gymnastics Business.

Gustavo Santana Velázquez

Business Problem

- Finding a suitable location for a business is a key factor that will have a great impact on its success.

- Location depends on the target customers, surrounding venues, brand visibility, reliance on traffic, etc.

- For this project, the neighborhoods of metropolitan area of Guadalajara was studied to choose the possible locations for opening a Gymnastics Business.

Background

- Guadalajara Metropolitan Area is the second largest in Mexico and is an international center of business, finance, arts and culture.

- Because of its high population density and key role on Mexico's economy, Guadalajara is a great place for start-ups.
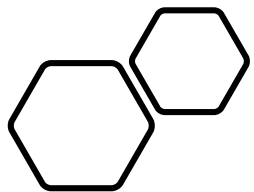
Objectives

The guide for How to Start a Gymnastics Business issued by USA Gymnastics will set the requirements needed for starting a Gymnastics Business, focusing on:

- Spending capacity of the target neighborhood
- Closeness to residential areas and convenient venues (businesses catering parents or families to increase exposition and attract customers)
- Evaluating competitors

Study Approach

- Collect data from the Government of Jalisco Website to get the neighborhoods and their average housing prices to separate the wealthiest.

- Use Geopy and Nominatim API to determine the coordinates for each location

- Use Foursquare API to study the competition and get the nearby venues for each neighborhood

- Visualize the neighborhoods using Folium library

- Cluster the neighborhoods using K-mean Clustering algorithm to determine which neighborhoods are suitable for the business.

Methodology

Data sources: Geographical and Statistical Information Institute of Jalisco (IIEG), Foursquare API, Geopy Libary

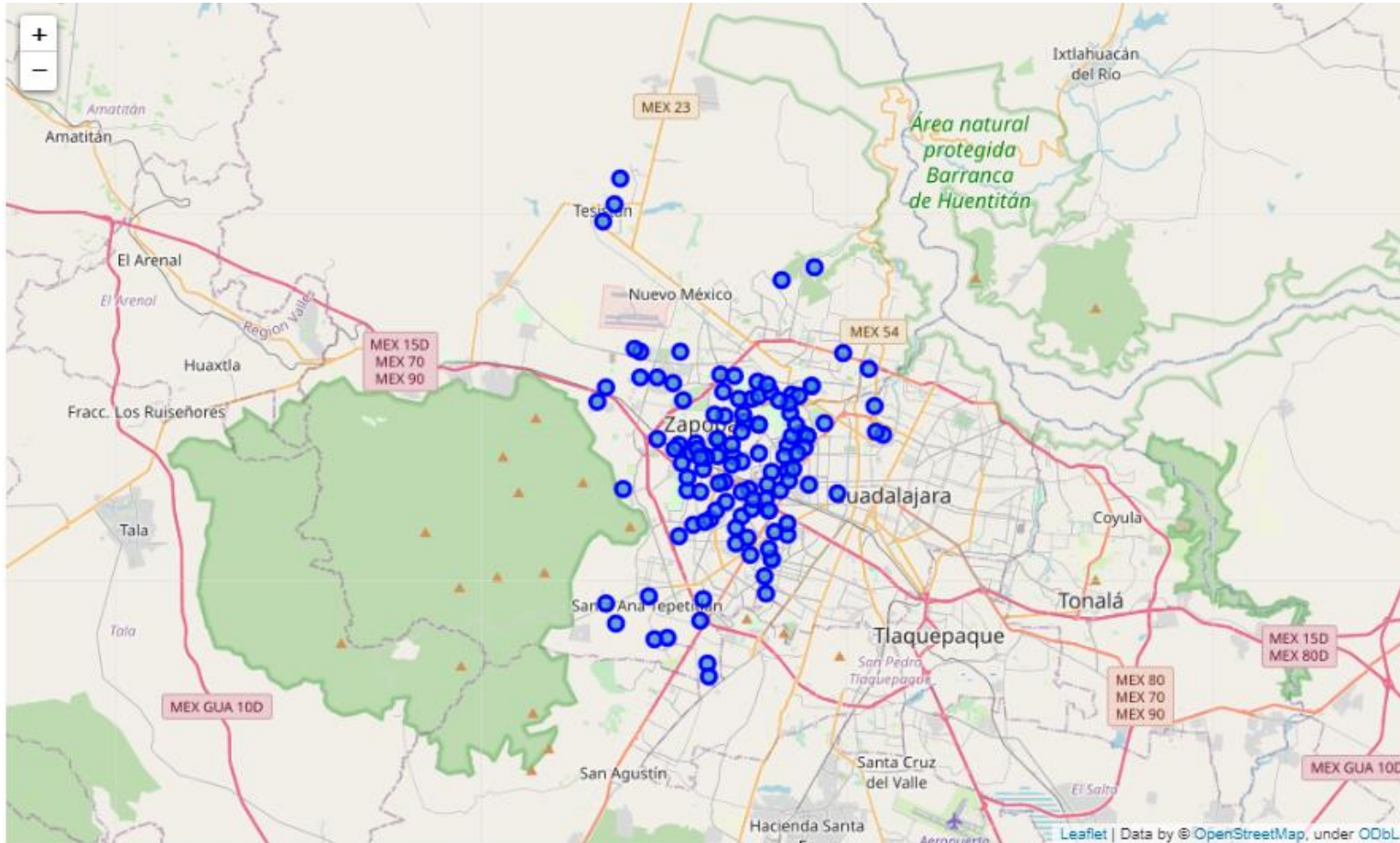RETRIEVE, SORT AND CLEAN A DATABASE TO GET THE 150 WEALTHIEST NEIGHBORHOODS IN GUADALAJARA.

GET THE EXISTING COMPETITION AND EVALUATE IF IT REPRESENTS A RISK

GET THE NEARBY VENUES FOR EACH NEIGHBORHOOD

TRANSFORM THE DATA TO CLUSTER THEM BASED ON THE VENUE CATEGORY

Wealthiest Neighborhoods in Guadalajara

Finding and Reviewing the Competition

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue ID	Venue	Venue Latitude	Venue Longitude	Venue Category
0	JARDINES DE LOS ARCOS	20.67	-103.40	4fb1b3aee4b0b5bcd2c46d3e	Gimnasia Artistica Ling	20.67	-103.40	Athletics & Sports
1	LOMAS PROVIDENCIA	20.69	-103.39	55db9d82498e77847671f6a8	Volaré Academia de Gimnasia	20.70	-103.39	Sports Club
2	PROVIDENCIA 4a SECCION	20.70	-103.39	55db9d82498e77847671f6a8	Volaré Academia de Gimnasia	20.70	-103.39	Sports Club
3	JARDINES DE SAN IGNACIO	20.67	-103.40	4fb1b3aee4b0b5bcd2c46d3e	Gimnasia Artistica Ling	20.67	-103.40	Athletics & Sports
4	RESIDENCIAL CHAPALITA	20.66	-103.43	4c05034473a8c9b6f9ca96e0	Club Atlas Chapalita	20.66	-103.43	Athletics & Sports
5	GUADALUPE JARDIN	20.66	-103.43	4c05034473a8c9b6f9ca96e0	Club Atlas Chapalita	20.66	-103.43	Athletics & Sports

```
Gimnasia Artistica Ling
        Number of Tips: 0
        Likes: 6
        Dislikes: False
        This venue has not been rated yet.

Volaré Academia de Gimnasia
        Number of Tips: 0
        Likes: 2
        Dislikes: False
        This venue has not been rated yet.

Club Atlas Chapalita
        Number of Tips: 23
        Likes: 288
        Dislikes: False
        Rating: 8.3
```
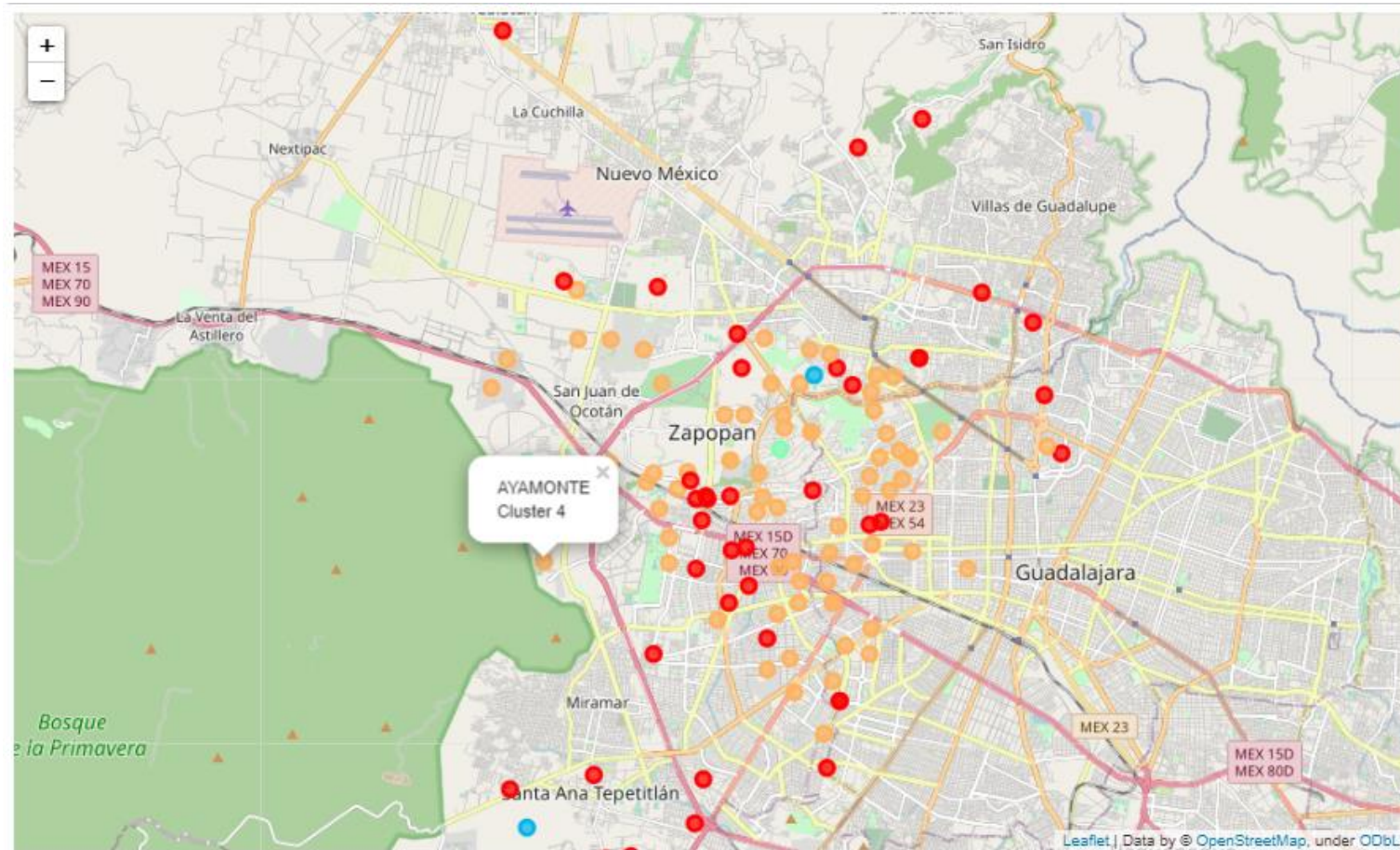
Getting Nearby Venues for each Neighborhood and Preparing for Clustering

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	PALMIRA	20.60	-103.43	Sabra Dios Donde	20.59	-103.43	Farm
1	PALMIRA	20.60	-103.43	Agua Blanca Country Club Centro De Espectaculo...	20.60	-103.43	Speakeasy
2	PALMIRA	20.60	-103.43	Spa Can Camp	20.60	-103.43	Spa
3	LOS FRAILES	20.72	-103.41	Decathlon	20.72	-103.42	Sporting Goods Shop
4	LOS FRAILES	20.72	-103.41	Hiperlumen	20.72	-103.41	Paper / Office Supplies Store
5	LOS FRAILES	20.72	-103.41	The Home Depot	20.72	-103.41	Hardware Store
6	LOS FRAILES	20.72	-103.41	Los Arcos	20.72	-103.41	Seafood Restaurant

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	AGRICOLA	Convenience Store	Restaurant	Lounge	Liquor Store	Construction & Landscaping	Salad Place	Taco Place	Butcher	Food Court	Mexican Restaurant
1	ALTAMIRA	Dog Run	Coffee Shop	Lounge	Burrito Place	Boutique	Food	Pizza Place	Gym / Fitness Center	Medical Center	Vegetarian / Vegan Restaurant
2	AMERICANA	Café	Coffee Shop	Mexican Restaurant	Tea Room	Breakfast Spot	Gastropub	Beer Bar	Bakery	Japanese Restaurant	Pizza Place
3	ARAUCA I	Sports Club	Food Truck	Pool	Yucatecan Restaurant	Flea Market	Farmers Market	Fast Food Restaurant	Film Studio	Fish & Chips Shop	Food
4	ATLAS COLOMOS	Bar	Athletics & Sports	Gym / Fitness Center	Shoe Store	Farm	Farmers Market	Fast Food Restaurant	Film Studio	Fish & Chips Shop	Yucatecan Restaurant

Clustered Neighborhoods in Guadalajara

Most of the neighborhoods were included on either cluster 0 or 4 out of five total generated clusters.

Cluster analysis (K-Mean Custering)

- Cluster 0 (Red)
  - 40 neighborhoods.
  - Common venues: Restaurants, Parks, Athletics & Sports, museums, music venues, shopping malls, spas, café. Family-oriented venues.
- Cluster 1 (Purple)
  - 1 neighborhood
  - Common Venue: Farm
- Cluster 2 (Blue)
  - 2 neighborhoods
  - Common venues: Athletics & Sports and Restaurants
- Cluster 3 (Green)
  - 1 neighborhood
  - Common venues: Pool, Yucatecan Restaurant, Flea Market
- Cluster 4 (Orange)
  - 74 neighborhoods
  - Food related venues, with some exceptions that offer a wide range of activities.

Results

- Based on the clustering results, neighborhoods in cluster 0 are the most suitable for opening the Gymnastics Business, it has the wider range of activities which is what we were looking for: businesses to cater the parents while the children are taking the lessons and venues that give the gym exposure.

- Cluster 4 should also be considered as a second option, since some of the neighborhoods include similar venues to cluster 0, nevertheless most of them are different kind of restaurants, which would provide considerable exposure but may not meet the surrounding venues requirement.

Limitations and Suggestions for Future Research

Foursquare Insufficient Data . – A lot of businesses aren't registered on Foursquare's database and for others the reviews are scares, making it impossible to conduct a complete research.

Using a Different Search Engine. – Using a more complete database such as Google Maps would help getting most of the venues in the area, as well as more reviews.

Future Research. – The characteristics of the possible facilities should also be considered, as they have a key role on the customer's preference

Conclusion

- After this study we can get a general idea of how the vanues in Guadalajara are distributed and which neighborhoods would suit a specific type of business, based on the spending capacity and surrounding shops.

- It is important to highlight that Foursquare's API isn't comlplete for this city in particular and is missing crucial information for a complete research.