

# El Proyecto Capstone - La Batalla de los Vecindarios

Oscar Didier Sánchez

June 2022

## 1. Introducción

Guadalajara es la ciudad capital de Jalisco conocida principalmente por el tequila y la música mariachi, ambos originarios de Jalisco. En el centro histórico de Guadalajara hay plazas coloniales y sitios emblemáticos de los cuales podemos destacar al Teatro Degollado, la catedral y la catedral Expiatorio, además del Palacio del Gobierno que alberga murales del famoso pintor José Clemente Orozco. Esta y otras razones hacen del centro de Guadalajara uno de los atractivos turísticos más importantes de la región. Es por ello que diversos comerciantes y empresarios se acercan para invertir en negocios en la zona.

Por lo tanto, es ventajoso para los empresarios detectar las zonas donde existe la mayor probabilidad de éxito, ya sea para restaurantes u otros giros. Por ejemplo, esta información se puede utilizar para proponer los lugares más transitados o donde los negocios tienen más éxito.

## 2. Panteamiento del Problema

En este proyecto intentaremos encontrar una ubicación óptima para un restaurante. Es de interés encontrar el tipo de restaurantes más comunes en la Ciudad de Guadalajara. En base a la información obtenida definir el restaurante con mayores probabilidades de éxito. Debido a su frecuencia de aparición alrededor del centro de Guadalajara, ya que implicaría el éxito de este giro de comida.

Dado que hay muchos restaurantes en la ciudad de Guadalajara, intentaremos detectar ubicaciones que aún no estén saturadas de restaurantes pero lo más cerca posible del centro de la ciudad.

Usaremos la ciencia de datos para generar zonas prometedoras de acuerdo a los criterios anteriores.

## 3. Objetivo

Según el planteamiento del problema:

- \* Encontrar los restaurantes con más frecuencia para proponerlo como posible negocio.

- \* número de restaurantes existentes al rededor de zonas.

- \* distancia del barrio al centro de la ciudad

Se selecciona una zona alrededor de la ciudad. Esa zona se divide en sub zonas para investigar cuáles son las áreas más factibles para el negocio.

Se necesitarán las siguientes fuentes de datos para extraer/generar la información requerida:

- \* Los centros de las áreas candidatas se generarán algorítmicamente.

- \* la cantidad de restaurantes y su tipo y ubicación en cada zona se obtendrá usando API de Foursquare.

- \* La coordenada del centro de Guadalajara se obtiene por ‘geolocator’.

- \* Las distancias son calculadas por ‘geopy.distance’.

## 4. Interés

Este proyecto esta dirigido a personas que esten interesados en abrir un restaurante pero no esten seguros del tema de comida en Guadalajara, Jalisco Mexico.

## 5. Adquisición y limpieza de datos.

Se usa la API de Foursquare para obtener información sobre restaurantes en cada vecindario. Estamos interesados en conocer cuáles son los comercios más comunes, por lo que nos aseguraremos de detectar e incluir todas las subcategorías de la categoría específica de ‘comida’, ya que necesitamos información sobre los restaurantes en la zona.

Los datos obtenidos consisten de lugar, ubicación (latitud y longitu) y categoría (Fig. 1).

## 6. Metodología

En este proyecto se propuso detectar un area de oportunidad de negocio, especificamente restaurantes. En áreas de la zona centro de Guadañajara que tengan una baja densidad de restaurantes del mismo giro. Limitaremos nuestro análisis al área de 6 km alrededor del centro de la ciudad.

En el primer paso, se recopiló datos probenientes de consultas a Fopresquare. Se realizó la ubicacion de restarunates y negocios alrededor de cada zona propuesta en pasos anteriores. Los datos obtenidos en general son ubicación y tipo (categoría) de cada restaurante.

El siguiente paso es obtener los restaurantes con mayor frecuencia por la zona cenro de laciudad con la finalidad de proponer uno o dos temas de comida para un restaurante.

Figura 1: Muestra de datos obtenidos del API deFoursquare.

Unnamed: 0	Zone	Zone Latitude	Zone Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	0 zone 0	20.679719	-103.32516	Tortas danny	20.679181	-103.323992	Sandwich Place
1	1 zone 0	20.679719	-103.32516	taqueria los charros (Taqueria Los Charros)	20.680708	-103.327542	Restaurant
2	2 zone 0	20.679719	-103.32516	Panadería La Perla de Occidente	20.678966	-103.328051	Bakery
3	3 zone 0	20.679719	-103.32516	El Siete Y Medio De Paco	20.678159	-103.322078	Seafood Restaurant
4	4 zone 0	20.679719	-103.32516	Carnes Asadas Fili	20.680322	-103.321226	Mexican Restaurant
5	5 zone 0	20.679719	-103.32516	Mascusia	20.678008	-103.327853	Bar
6	6 zone 0	20.679719	-103.32516	Mariscos Los Amigos	20.676075	-103.323003	Seafood Restaurant
7	7 zone 0	20.679719	-103.32516	Vinos giralda	20.680670	-103.326489	Liquor Store
8	8 zone 0	20.679719	-103.32516	Tacos Don Mario	20.681504	-103.328743	Taco Place
9	9 zone 0	20.679719	-103.32516	Tortas y tacos El güero y su cazo de fritangas	20.681867	-103.327289	Taco Place
10	10 zone 0	20.679719	-103.32516	Carnes Asadas Mode	20.681058	-103.321752	Taco Place
11	11 zone 0	20.679719	-103.32516	Menuería Chipres	20.682759	-103.326833	Mexican Restaurant
12	12 zone 0	20.679719	-103.32516	Farmacia Guadalajara	20.678984	-103.328084	Pharmacy
13	13 zone 0	20.679719	-103.32516	Farmacia El Globo	20.682699	-103.327938	Pharmacy
14	14 zone 0	20.679719	-103.32516	Mama Lola	20.681580	-103.325984	Bakery
15	15 zone 0	20.679719	-103.32516	Armadillo Restaurante	20.680555	-103.320792	Mexican Restaurant
16	16 zone 0	20.679719	-103.32516	OXXO	20.681403	-103.321909	Convenience Store
17	17 zone 0	20.679719	-103.32516	Mariscos El Cocodrilo	20.675645	-103.323384	Seafood Restaurant
18	18 zone 0	20.679719	-103.32516	Oxxo	20.676822	-103.328215	Convenience Store

Una vez seleccionado los posibles negocios se filtraran todos los restaurantes con mayor frecuencia de la zona, descartando así la mayoría de locales por la zona y solamente observar los de interés.

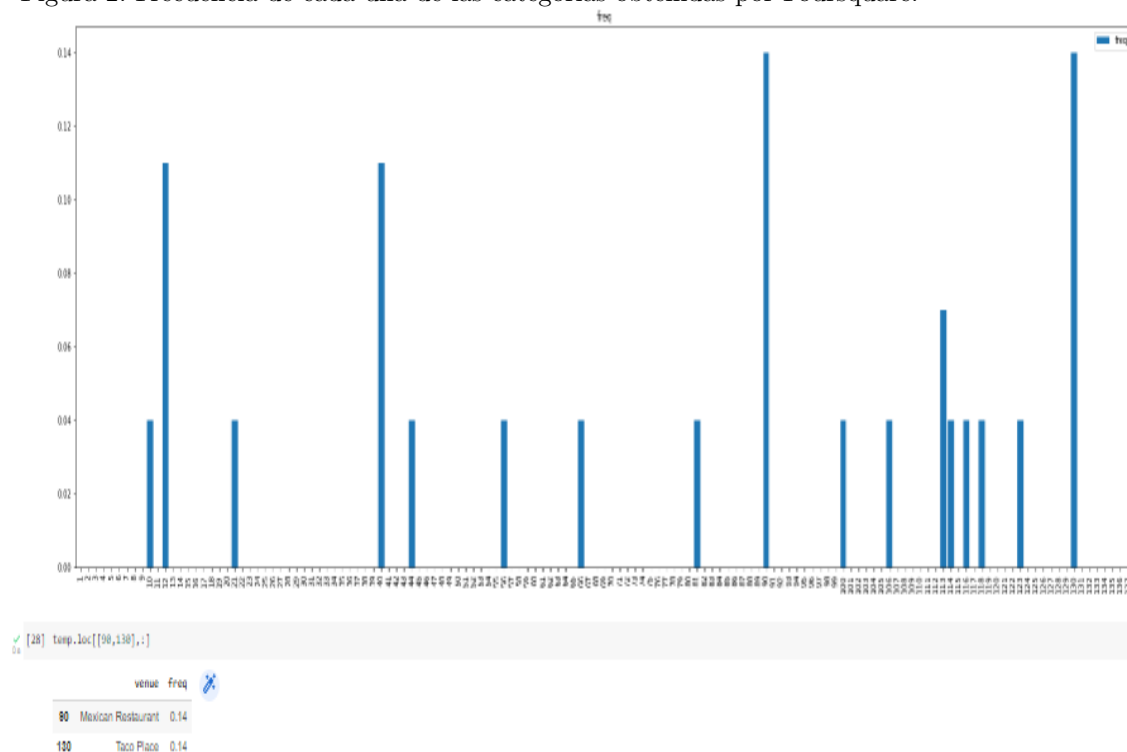
Se calculara la distancia que tiene cada uno de los restaurantes a la zona centro y la cercanía a otros restaurantes del mismo tipo en un zona alrededor de 250m a la redonda.

Finalmente, nos centraremos en las áreas más prometedoras y dentro de ellas crearemos grupos de ubicaciones, tomaremos en consideración ubicaciones con no demasiados restaurantes en un radio de 250 metros. Esta clasificación la realizaremos por medio de k-means clustering y separar en tres categorías. Zonas muy densas en restaurantes. Zonas muy desoladas y zonas con numeros de restaurantes moderos. Siendo esta ultima clase la de mayor interés para nuestra investigación.

## 7. Resultados

Se revisan las frecuencias de aparición de cada una de las categorías. Obtenemos Los comercios con mayor frecuencia en la zona. Se puede observar que las categorías "Taco Placez "Mexican Restaurant"son los mas habituales. Por lo que podemos proponer un restaurante para Tacos o Comida mexicana.

Figura 2: Frecuencia de cada una de las categorías obtenidas por Foursquare.



Una vez seleccionado, las dos categorías mas comunes se filtran los datos y son descartados todos los demás.

Figura 3: Datos filtrados con las categorías "Taco Placez "Mexican Restaurant".

2676	zone 81	20.66665	-103.350206	Tacos El Güero	20.669007	-103.350188	Taco Place
2715	zone 83	20.657573	-103.340757	los gueros barbacoa y algo mas	20.656714	-103.338269	Taco Place
2777	zone 86	20.662656	-103.348911	Tacos De Barbacoa	20.662189	-103.352865	Taco Place
2782	zone 86	20.662656	-103.348911	Gorditas Dona Mago	20.666014	-103.349874	Taco Place
2812	zone 86	20.662656	-103.348911	Tacos Del Pri	20.660812	-103.351703	Taco Place
2827	zone 87	20.66435	-103.35163	Tacos De Barbacoa	20.662189	-103.352865	Taco Place
2832	zone 87	20.66435	-103.35163	Gorditas Dona Mago	20.666014	-103.349874	Taco Place
2865	zone 87	20.66435	-103.35163	Tacos Del Pri	20.660812	-103.351703	Taco Place
4	zone 0	20.679719	-103.32516	Carnes Asadas Fili	20.680322	-103.321226	Mexican Restaurant
11	zone 0	20.679719	-103.32516	Menudería Chipres	20.682759	-103.326833	Mexican Restaurant
15	zone 0	20.679719	-103.32516	Armadillo Restaurante	20.680555	-103.320792	Mexican Restaurant
29	zone 1	20.681415	-103.327877	Menudería Chipres	20.682759	-103.326833	Mexican Restaurant
44	zone 2	20.683111	-103.330594	Gorditas Las Rancheritas	20.68511	-103.333837	Mexican Restaurant
47	zone 2	20.683111	-103.330594	Menudería Chipres	20.682759	-103.326833	Mexican Restaurant
53	zone 3	20.684807	-103.333311	Los Chilaquiles Del Güero	20.687995	-103.332934	Mexican Restaurant
56	zone 3	20.684807	-103.333311	Gorditas Las Rancheritas	20.68511	-103.333837	Mexican Restaurant
59	zone 3	20.684807	-103.333311	Carnes Asadas Los Troncos	20.687319	-103.334239	Mexican Restaurant
60	zone 3	20.684807	-103.333311	Menudo Mely	20.685549	-103.336382	Mexican Restaurant
82	zone 4	20.686503	-103.336029	Menudo Mely	20.685549	-103.336382	Mexican Restaurant
84	zone 4	20.686503	-103.336029	Carnes Asadas Los Troncos	20.687319	-103.334239	Mexican Restaurant

Se muestran la ubicación de los locales en las zonas del centro de Guadalajara. Se puede observar la concentración de restaurantes con categoría "Taco Placez "Mexican Restaurant".

Figura 4: Ubicación de locales con categoría "Taco Placez "Mexican Restaurant".



Se calculó las distancias de cada uno de los locales al centro, además se obtiene la cantidad de restaurantes de cada uno de los locales en un radio de 250 m.

Figura 5: Distancia y restaurantes cercanos.

index	zone	zone latitud	zone longitud	venue	venue latitud	venue longitud	categoría	restaurantes cercanos	distancia centro
0	8	zone 0	20.679719 -103.32516	Tacos Don Mario	20.681504	-103.328743	Taco Place	0	1.4526084972674773 km
1	9	zone 0	20.679719 -103.32516	Tortas y tacos El güero y su cazo de fritangas	20.681867	-103.327289	Taco Place	0	1.5885417652297769 km
2	10	zone 0	20.679719 -103.32516	Carnes Asadas Mode	20.681058	-103.321752	Taco Place	0	2.001189758382035 km
3	26	zone 1	20.681415 -103.327877	Tacos Don Mario	20.681504	-103.328743	Taco Place	15	1.4526084972674773 km
4	30	zone 1	20.681415 -103.327877	Tortas y tacos El güero y su cazo de fritangas	20.681867	-103.327289	Taco Place	15	1.5885417652297769 km
5	41	zone 2	20.683111 -103.330594	Tacos Don Mario	20.681504	-103.328743	Taco Place	0	1.4526084972674773 km
6	43	zone 2	20.683111 -103.330594	Tacos Don Luis	20.682093	-103.333854	Taco Place	0	1.2096810703467684 km
7	46	zone 2	20.683111 -103.330594	Tortas y tacos El güero y su cazo de fritangas	20.681867	-103.327289	Taco Place	0	1.5885417652297769 km
8	63	zone 3	20.684807 -103.333311	Tacos Don Luis	20.682093	-103.333854	Taco Place	6	1.2096810703467684 km
9	64	zone 3	20.684807 -103.333311	Bimía de Res "Los Cuates"	20.684807	-103.337026	Taco Place	6	1.4298022950903397 km

La métricas "Distancia Centro" y restaurantes cercanos" son utilizados para el algoritmo  $k - mean$  y clasificar los restaurantes en 3 grupos.

Según los resultados obtenidos se puede observar que las categorías se dividen en:

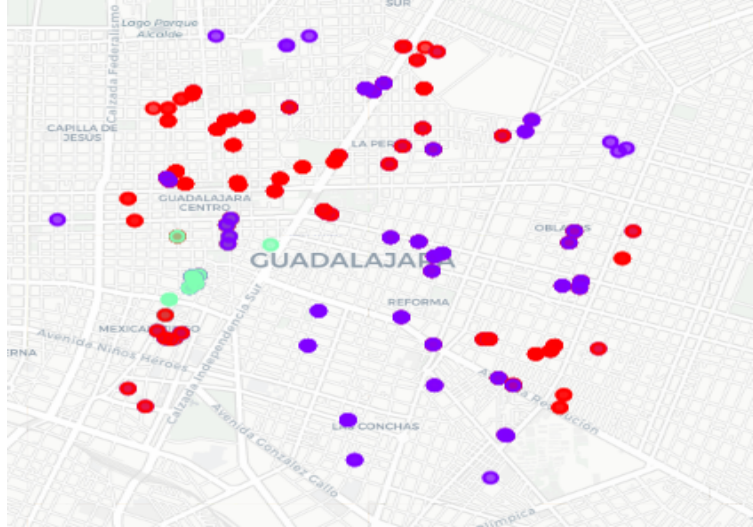
Cluster 0: Restaurantes cercanos al centro de la ciudad con cantidad de restaurantes cercanos moderados.

Cluster 2: Restaurantes alejados al centro de la ciudad con poca cantidad de restaurantes cercanos.

Cluster 3: Restaurantes cercanos al centro con cantidad de bastantes restaurantes cercanos.

La Fig. 6 muestra los diferentes clusters:

Figura 6: Diferentes Clusters obtenidos del algoritmo  $K - mean$ .



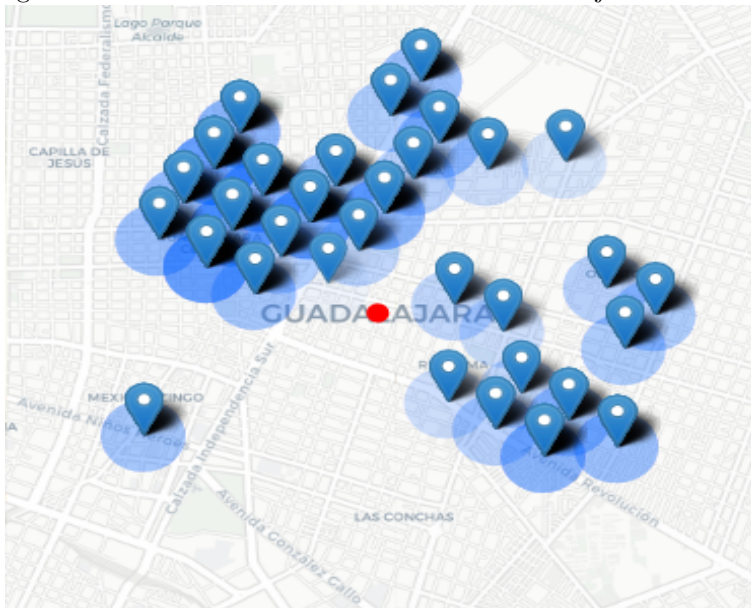
Por lo tanto las zonas seleccionadas son aquellas que tienen cantidad de restaurantes moderados. Ya que una gran densidad, significaría una alta competencia. Pero zonas muy desoladas implican que los negocios no prosperan. Por lo tanto seleccionamos el cluster 0

Figura 7: Zonas del Cluster con mejores resultados.



Finalmente se muestran las zonas con mejor ubicados para restaurantes de categoría "Taco Place" y "Mexican Restaurant" (8).

Figura 8: Ubicación de las zonas del Cluster con mejores resultados.



## 8. Conclusión y Discución

En este trabajo, se analizó los tipos de negocios que prosperan alrededor del centro de la ciudad de Guadalajara. Se extrajo la información para encontrar aquellos que tienen una categoría similar y encontrar su frecuencia. Lo que nos indicaría que estos negocios son lo que más oportunidades tienen de éxito. Entre las diferentes opciones fueron seleccionados dos cuya frecuencia fue la más alta según el análisis. Estos son "Taco Place" y "Mexican Restaurant". Estas categorías nos sirvieron como objetivo de estudio para observar como están dispersados alrededor de la ciudad. De esta manera, encontrar zonas que no tenga una alta competencias o estén demasiado solo, lo que indicaría que ese tipo de negocios no prosperan.

Utilizando el algoritmo de *k - mean* se clasificaron los restaurantes en tres categorías según las distancia que tienen al centro de la ciudad y la cantidad de restaurantes a su alrededor.

Una vez encontrados los restaurantes con mejores características, fueron ubicadas sus zonas. Las cuales representan, aquellas zonas que no han sido sobre pobladas por restaurantes en nuestra categorías objetivos.

Estos resultados son de utilidad para empresarios que quieren invertir en la zona centro de la ciudad, pero no tienen información suficiente de lo más



consumido por la zona. Además de evitar una alta competencias por un sobrepoblación de tiendas con el mismo giro.

## **9. Trabajo futuro**

Este mismo estudio puede ser realizado, con un enfoque de concentración de personas en zonas comerciales, para proponer negocios según la demanda de los consumidores de la zona.