

¿Cómo afecta la localización al éxito de tu negocio?

Gabriel Abilleira Rodríguez

Agosto 2022

Índice

1. Introducción	3
1.1. Motivación	3
1.2. Problema	3
1.3. Objetivo	3
2. Datos	3
2.1. Obtención de los barrios	4
2.2. Obtención de los servicios	5
2.3. Obtención de las reviews	6
2.4. Obtención de las matrices	7
3. Metodología	7
3.1. Obtención de la similitud entre barrios	8
4. Análisis	9
5. Resultados y conclusión	10

1. Introducción

1.1. Motivación

La localización de una empresa es un elemento clave para su correcto desarrollo, y puede constituir uno de los **factores competitivos** más importantes de la misma. En este sentido, una empresa podría decidir colocarse cerca de sus competidores si no puede competir en precios, ya que vende un producto homogéneo, o podría colocarse a cierta distancia si esto le otorga cierto poder de mercado.

1.2. Problema

Por este motivo, los dueños de los restaurantes estarían interesados en acceder a un servicio que pudiese aportar información adicional sobre la localización de los otros restaurantes en la ciudad, de forma que puedan tomar una decisión informada sobre este factor clave. El objetivo de este análisis es, por tanto, **predecir el éxito de un restaurante en función de su localización**.

1.3. Objetivo

Para ello, voy a diseñar un **sistema de recomendación colaborativo** que permita elaborar una lista para un barrio de la ciudad de Los Ángeles que contenga qué restaurantes tendrían más éxito de abrir uno ahí, en función de la **categoría del restaurante**.

2. Datos

Para poder abordar este problema, no solamente es necesario obtener datos sobre la localización de los diferentes restaurantes de una ciudad, sino que también es necesario adquirir **datos que permitan medir el éxito de los restaurantes**. Precisamente por este motivo, además de emplear los datos de la **API de Foursquare** (para la ciudad de Los Ángeles, California), utilizaremos

también datos de **reseñas de Google** que los usuarios dejan en estos negocios, mediante el uso de la **API de Google Maps**. Estos datos de reseña tienen un valor de **1 a 5 estrellas**, donde 1 sería la puntuación más baja que se puede otorgar a un negocio.

La API de Foursquare necesita los datos de *latitud y longitud* de los barrios objetivo. Para obtener esta información, empleamos la **API de Google Geocode**. Una vez hecho esto, se pueden obtener los datos de los diferentes negocios que existen en cada barrio a través de la API de Foursquare y, finalmente, se pueden usar sus nombres para obtener las reseñas que los usuarios de Google hicieron sobre el negocio.

2.1. Obtención de los barrios

Primeramente, es necesario obtener los datos sobre los diferentes vecindarios de la ciudad de California, de forma que se pueda establecer un área de interés. Para ello, es necesario obtener los datos de latitud y longitud. Veamos cómo se realiza este proceso.

Para ello, es necesario obtener una lista con todos los vecindarios, para lo que emplearemos la página de Wikipedia correspondiente y el paquete **BeautifulSoup**. Una vez obtenida esta lista, mediante el empleo de la API de Google Geocode se pueden obtener los datos de latitud y longitud de los barrios, lo que posteriormente nos permitirá obtener los servicios de cada barrio.

Veamos los diferentes barrios en un mapa:

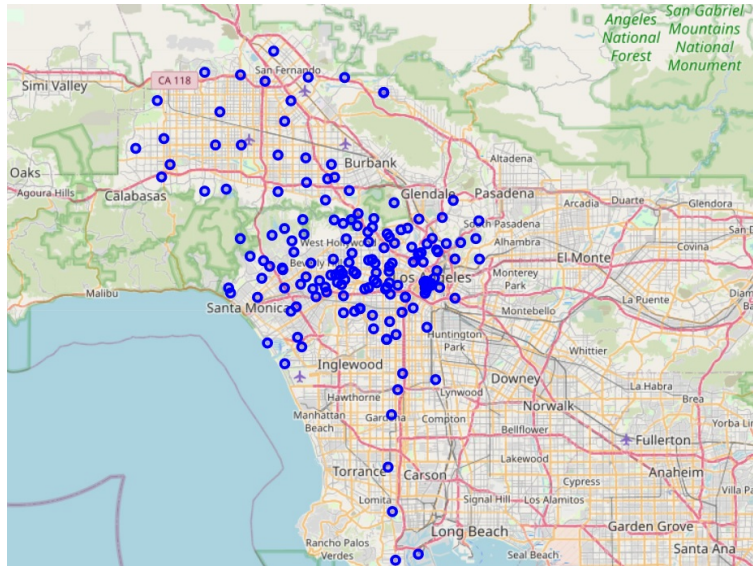


Figura 1: Mapa con los barrios de la ciudad de Los Ángeles

Como se puede ver, cada punto azul sobre el mapa marca un vecindario. Todos ellos componen los barrios de la base de datos, para los que se obtendrá una lista de servicios y sus correspondientes valoraciones de los usuarios.

2.2. Obtención de los servicios

Ahora es el momento de usar estos datos para poder obtener los restaurantes localizados en los diferentes barrios de Los Ángeles. Para ello, empleamos la API de Foursquare.

La API de Foursquare nos permite obtener todos los servicios que se encuentran en cada barrio, simplemente con saber la latitud y la longitud del mismo. Una vez obtengamos estos datos, es necesario filtrar para obtener solamente los restaurantes. El resultado se puede observar en el siguiente *DataFrame*:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Adams-Normandi	34.034344	-118.299314	Orange Door Sushi	34.032550	-118.299437	Sushi Restaurant
4	Adams-Normandi	34.034344	-118.299314	Sushi Delight	34.032483	-118.299436	Sushi Restaurant
9	Adams-Normandi	34.034344	-118.299314	La Estrella Mexican Food	34.033002	-118.301072	Fast Food Restaurant
17	Adams-Normandi	34.034344	-118.299314	Orange Sekai Ramen	34.032519	-118.299366	Ramen Restaurant
20	Adams-Normandi	34.034344	-118.299314	Lago De Guija	34.032538	-118.298800	Fast Food Restaurant

Figura 2: Restaurantes de cada barrio

2.3. Obtención de las reviews

Una vez hecho esto, llega el paso final, que consiste en combinar estos datos con los de la API de Google Maps para obtener las reseñas de los restaurantes. El objetivo es obtener una matriz de reseñas. Esto nos permitirá, en el siguiente apartado, elaborar un sistema de recomendación de filtrado colaborativo. Para ello, es necesario tener una base de datos donde, para cada *usuario*, en este caso, los barrios, se tiene la información acerca de qué restaurantes existen en la zona y la nota que tienen. Dado que las notas van de 1 a 5, un 0 indica que ese tipo de restaurante no se encuentra en la zona. El resultado final es el siguiente:

	Neighborhood	Category	Mean Review
0	Adams-Normandi	Fast Food Restaurant	4.3
1	Adams-Normandi	Latin American Restaurant	4.8
2	Adams-Normandi	Ramen Restaurant	4.2
3	Adams-Normandi	Sushi Restaurant	2.5
4	Angelino Height	Asian Restaurant	3.6
...
819	Yucca Corrido	Indian Restaurant	3.4

Figura 3: Reviews de cada restaurante

2.4. Obtención de las matrices

Finalmente, es necesario separar los datos entre el *usuario activo*, que en este caso va a ser el barrio de Los Ángeles Oeste, y el resto de *usuarios*, que son los que nos permiten elaborar el filtrado colaborativo. Este caso es a modo de ejemplo, y se podría replicar este análisis para cualquier barrio de interés, no solamente para el escogido en esta ocasión.

3. Metodología

Antes de elaborar el sistema de recomendación, sería interesante comprender qué es. En este sentido, podemos definirlos como una colección de algoritmos utilizados para sugerir temas a los usuarios, basados en información tomada desde el punto de vista del usuario.

En este caso concreto, la idea es encontrar *usuarios*, en este caso, los vecindarios, que tengan preferencias y opiniones parecidas, en este caso, sobre los tipos de restaurantes, para entonces recomendar items que se hayan parecido al ingreso anterior. Existen diferentes métodos para encontrar los diferentes vecindarios que son similares entre sí. En este caso, el método que vamos a utilizar estará basado en la **Función de Correlación de Pearson**.

Como ya hemos definido anteriormente cuál será el *usuario activo*, quedan por realizar los siguientes pasos:

1. Encontrar los primeros X barrios similares basándonos en la puntuación de las categorías de los restaurantes.
2. Obtener el registro de la categoría del restaurante que "valoró" el vecindario para cada barrio.
3. Calcular un puntaje de similitud utilizando alguna fórmula.
4. Recomendar las categorías de restaurantes con los puntajes más altos.

3.1. Obtención de la similitud entre barrios

Para encontrar la similitud entre el barrio escogido y los que forman el grupo, como ya quedó señalado, se emplea el **Coefficiente de Correlación de Pearson**. Este coeficiente se utiliza para medir la fuerza de una asociación lineal entre dos variables. La fórmula para encontrar este coeficiente entre los conjuntos X e Y con los valores de N se puede ver en la imagen debajo:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$

Figura 4: Correlación de Pearson

Los valores obtenidos pueden variar de -1 a 1, donde 1 se correlaciona directamente entre las dos entidades (esto sería una correlación positiva perfecta) y -1 forma una correlación negativa perfecta. En nuestro caso, un 1 se refiere a que dos usuarios tiene gustos parecidos, mientras que -1 es lo opuesto.

El resultado de todo este proceso se puede observar aquí:

	similarityIndex	Neighborhood
124	1.0	El Seren
129	1.0	Venic
115	1.0	Vermont Squar
125	1.0	Harbor Gatewa
127	1.0	Lafayette Squar

Figura 5: Similitud entre barrios

4. Análisis

Ahora, es el momento de recomendar qué restaurantes deberían abrir en función de la categoría. Para ello, tomaremos el peso promedio de los ratings de las categorías de los restaurantes utilizando la Correlación Pearson. Pero para hacer esto, primero necesitamos que los barrios vean las categorías en nuestro *DataFrame* a partir del *DataFrame* de puntajes y luego guardar su correlación en una nueva columna.

Posteriormente, todo lo que se necesita hacer es multiplicar el puntaje de la categoría por su peso (El índice de similitud), luego se suman los nuevos puntajes y dividen por la suma de los pesos. Esto se logra sencillamente multiplicando dos columnas, luego agrupando el *DataFrame* por la columna correspondiente y luego dividiendo dos columnas. Este es el resultado:

weighted average recommendation score		Category
Category		
African Restaurant	4.558442	African Restaurant
American Restaurant	2.477823	American Restaurant
Argentinian Restaurant	4.489855	Argentinian Restaurant
Armenian Restaurant	4.200000	Armenian Restaurant
Asian Restaurant	0.053375	Asian Restaurant

Figura 6: Matriz final

5. Resultados y conclusión

Reordenando la matriz, se consigue la siguiente lista:

weighted average recommendation score		Category
Category		
Burger Joint	22.134043	Burger Joint
Indian Restaurant	8.182609	Indian Restaurant
Sushi Restaurant	6.417497	Sushi Restaurant
Taco Restaurant	5.820370	Taco Restaurant
Thai Restaurant	5.456878	Thai Restaurant
Sandwich Restaurant	5.369811	Sandwich Restaurant
Mexican Restaurant	5.291489	Mexican Restaurant
French Restaurant	5.185185	French Restaurant
Restaurant	5.175276	Restaurant
Cuban Restaurant	5.000000	Cuban Restaurant

Figura 7: Resultado

Como se puede ver, gracias a este sistema de recomendación, una persona interesada en abrir un restaurante en el distrito de Los Ángeles Oeste podría llegar a saber qué tipos de restaurante podrían funcionar mejor en esa ubicación, en este caso, una cadena de hamburguesas sería la primera opción.

Este análisis se puede extender a cualquier otro barrio de la ciudad de Los Ángeles, permitiendo así a las personas que quieran abrir un restaurante entender no solamente qué tipos de restaurantes podrían funcionar mejor en función de la localización, sino que también podrían llegar a entender qué localización es la mejor para abrir un tipo determinado de restaurante.