



Proyecto Capstone

La Batalla de los Vecindarios

Sebastián Larrea

28 de marzo del 2021

Contenido

1. Introducción	3
1.1 Problema comercial	3
1.2 Condiciones del cliente	3
2. Datos	3
2.1 Minería de datos	3
2.2 Tratamiento de la información	4
2.3 Limpieza de los datos	6
3. Metodología	7
4. Resultados y discusión	7
4.1 Aplicación del modelo de aprendizaje automático	9
5. Conclusión	10

Tema: Selección del mejor sector para la construcción o implementación de un hotel en la ciudad de Los Angeles, California

1. Introducción

1.1 Problema comercial

Una industria hotelera desea abrir una nueva cadena de hoteles de 4 y 5 estrellas en la ciudad de Los Ángeles- California, Estados Unidos

Para ello requieren realizar un estudio completo donde se encuentre, la minería, la limpieza, el tratamiento y el análisis de datos para determinar el mejor sector, barrio o calle en la que se deben situar los hoteles, con la ayuda de algún modelo de aprendizaje automático que facilite la selección de lugar, y sobre todo buscar que el hotel represente una inversión admisible para el cliente, donde las ganancias económicas crezcan exponencialmente, y la satisfacción de los usuarios sea la mejor posible, provocando las recomendaciones hacia más personas para que visiten el hotel, algunas de las condiciones que planteo el cliente se las puede ver en la siguiente sección.

1.2 Condiciones del cliente

Las condiciones o recomendaciones que el cliente a propuesto para la implementación del hotel son:

- Debe estar situado cerca de lugares turísticos.
- Debe estar situado cerca de restaurantes y sitios de comida.
- Debe estar situada cerca de playas y lugares de descanso.
- Se debe evitar zonas peligrosas en donde los clientes puedan sufrir de robos o agresiones.

Bajo estas consideraciones se debe realizar el estudio de datos y determinar el mejor lugar disponible para la construcción del hotel.

2. Datos

2.1 Minería de datos

Para la obtención de los datos, se realizó una minería de información, donde se recopilaron todos los vecindarios existentes en la ciudad de Los Ángeles a través de sitios web, con un total de 194 vecindarios pertenecientes a esta ciudad y su distrito, toda esta información fue almacenada en una hoja de cálculo como se observa en la Figura 1.

	A	B	C	D	E	F	G
1	Barrio	Ciudad	Full Address				
2	Angelino Heights	Los Angeles	Angelino Heights Los Angeles				
3	Angeles Mesa	Los Angeles	Angeles Mesa Los Angeles				
4	Angelus Vista	Los Angeles	Angelus Vista Los Angeles				
5	Arleta	Los Angeles	Arleta Los Angeles				
6	Arlington Heights	Los Angeles	Arlington Heights Los Angeles				
7	Arts District	Los Angeles	Arts District Los Angeles				
8	Atwater Village	Los Angeles	Atwater Village Los Angeles				
9	Baldwin Hills	Los Angeles	Baldwin Hills Los Angeles				
10	Baldwin Hills/Crenshaw	Los Angeles	Baldwin Hills/Crenshaw Los Angeles				
11	Baldwin Village	Los Angeles	Baldwin Village Los Angeles				
12	Baldwin Vista	Los Angeles	Baldwin Vista Los Angeles				
13	Beachwood Canyon	Los Angeles	Beachwood Canyon Los Angeles				
14	Bel Air	Los Angeles	Bel Air				
15	Benedict Canyon	Los Angeles	Benedict Canyon Los Angeles				
16	Beverly Crest	Los Angeles	Beverly Crest Los Angeles				
17	Beverly Glen	Los Angeles	Beverly Glen Los Angeles				
18	Beverly Glen	Los Angeles	Beverly Glen Los Angeles				
19	Beverly Grove	Los Angeles	Beverly Grove Los Angeles				
20	Beverly Park	Los Angeles	Beverly Park Los Angeles				
21	Beverlywood	Los Angeles	Beverlywood Los Angeles				
22	Boyle Heights	Los Angeles	Boyle Heights Los Angeles				
23	Brentwood	Los Angeles	Brentwood Los Angeles				
24	Brentwood Circle	Los Angeles	Brentwood Circle Los Angeles				
25	Brentwood Glen	Los Angeles	Brentwood Glen Los Angeles				
26	Broadway-Manchester	Los Angeles	Broadway-Manchester Los Angeles				
27	Brookside	Los Angeles	Brookside Los Angeles				
28	Bunker Hill	Los Angeles	Bunker Hill Los Angeles				
29	Canoga Park	Los Angeles	Canoga Park Los Angeles				
30	Canterbury Knolls	Los Angeles	Canterbury Knolls Los Angeles				
31	Carthay	Los Angeles	Carthay Los Angeles				
32	Castle Heights	Los Angeles	Castle Heights Los Angeles				
33	Central-Alameda	Los Angeles	Central-Alameda Los Angeles				

Figura 1. Barrios de la ciudad de Los Ángeles- California

2.2 Tratamiento de la información

Con esta base de datos, se procedió a implementar la aplicación Geocode, (Complemento de Google Sheets) ver Figura 2, se procedió a obtener la información geográfica de latitud y longitud de cada uno de los 194 vecindarios, información necesaria que se utilizara más adelante en conjunto con Foursquare para adquirir toda la información de cada vecindario como se ve en la Figura 3.

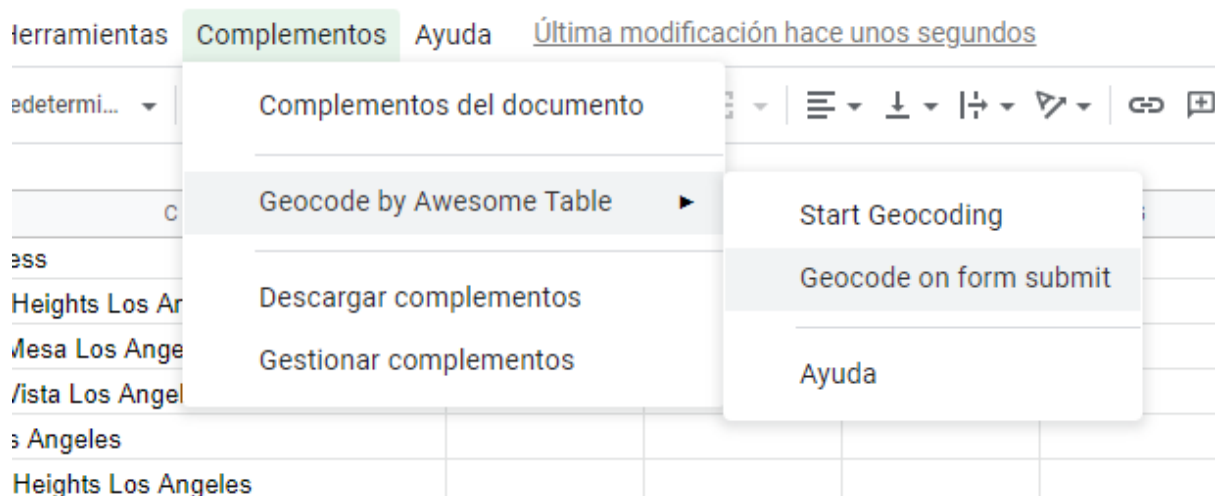


Figura 2. Complemento Geocode

1

Barrio

2

Angelino Heights

3

Angelus Mesa

4

Angelus Vista

5

Arlota

6

Arlington Heights

7

Arts District

8

Atwater Village

9

Baldwin Hills

10

Baldwin Hills/Crenshaw

11

Baldwin Village

12

Baldwin Vista

13

Beachwood Canyon

14

Bel Air

15

Benedict Canyon

16

Beverly Crest

17

Beverly Glen

18

Beverly Grove

19

Beverly Park

20

Beverlywood

21

Boyle Heights

22

Brentwood

23

Brentwood Circle

24

Brentwood Glen

25

Broadway-Manchester

26

Brookside

27

Bunker Hill

28

Canoga Park

29

Carthay

30

Castle Heights

31

Central-Alameda

32

Century City

Ciudad

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Los Angeles

Figura 3. Latitud y Longitud de cada barrio de Los Ángeles.

Esta es la base de datos que se utilizó para el estudio del caso propuesto, posteriormente se procedió a cargar la base de datos en un Data Frame en un Notebook de trabajo para realizar todo el estudio previsto, se realizó además una superposición de cada barrio sobre un mapa de Los Ángeles con la ayuda de sus coordenadas, para determinar la correcta implementación de los datos en el Data Frame, como se visualiza en la Figura 4.

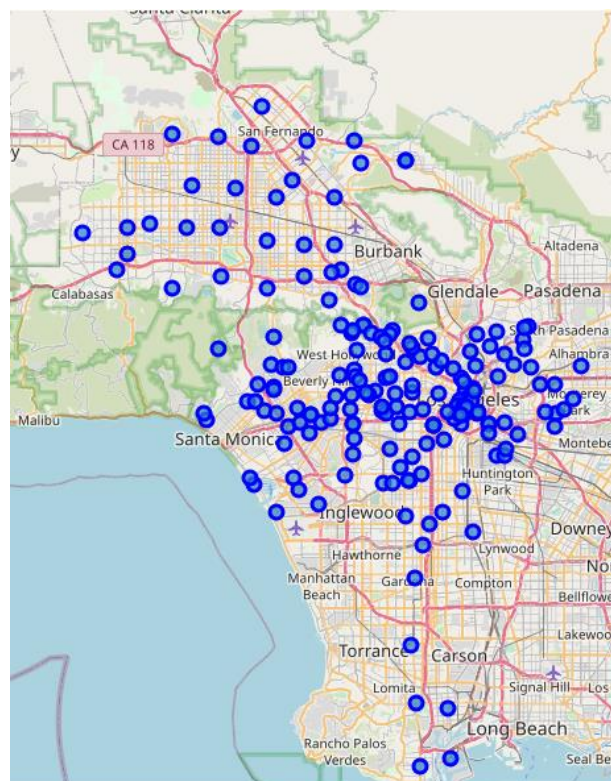


Figura 4. Barrios superpuestos en la ciudad de Los Ángeles.

2.3 Limpieza de los datos

Para un mejor manejo del data frame, se realizó la eliminación de columnas como "ciudad", ya está entendido que se realiza en el estudio en la ciudad de Los Ángeles, porque esta redundante su permanencia en este.

Por otro lado, la columna "Full Address", fue utilizada para la obtención de la latitud y longitud de los vecindarios, por lo que su propósito fue cumplido y no se la necesita más para el estudio.

	Vecindario	Latitud	Longitud
0	Angelino Heights	34.0413	-118.1517
1	Angeles Mesa	33.9942	-118.3136
2	Angelus Vista	34.0469	-118.3176
3	Arleta	34.2419	-118.4255
4	Arlington Heights	34.0232	-118.1908
5	Arts District	34.0417	-118.2329
6	Atwater Village	34.0659	-118.1523
7	Baldwin Hills	34.0047	-118.2125
8	Baldwin Hills/Crenshaw	34.1121	-118.2123
9	Baldwin Village	34.0054	-118.2052
10	Baldwin Vista	34.0061	-118.3624

Figura 5. Limpieza del Data Frame.

Además, con la ayuda del mapa superpuesto con todos los vecindarios de la ciudad, aplicamos una de las condiciones primordiales para la implementación de este nuevo hotel:

- Debe estar situada cerca de playas y lugares de descanso.

Ya que no todos los vecindarios poseen una salida costera, no será necesario realizar un análisis a profundidad de cada uno, y solo se lo realizará en los vecindarios de Pacific Palisades, Venice, Marina Península, Playa del Rey, San Pedro y Wilmington, que cumplen con esta importante condición.

Obteniendo como resultado un Data Frame, listo para aplicar la metodología de análisis adecuada, como se ve en la Figura 6.

	Vecindario	Latitud	Longitud
119	Pacific Palisades	34.035633	-118.515590
165	Venice	33.985047	-118.469483
99	Marina Peninsula	33.979698	-118.465657
127	Playa del Rey	33.956419	-118.442232
135	San Pedro	33.736062	-118.292246
188	Wilmington	33.785795	-118.264357

Figura 6. Data Frame utilizado para el estudio con los 6 barrios seleccionados.

3. Metodología

En este proyecto se realizará un enfoque estadístico para determinar cuál de los 6 vecindarios propuestos de la ciudad de Los Ángeles (Pacific Palisades, Venice, Marina Península, Playa del Rey, San Pedro y Wilmington), es el más adecuado para la implementación del nuevo hotel, para esto utilizaremos la herramienta de Foursquare, que nos proporcionara la información de cada uno de los 6 vecindarios, en un radio de 500 metros.

Una vez con esta información y realizado el análisis estadístico, se determinará cual vecindario cumple con las condiciones establecidas en un principio, una vez escogido, a este vecindario se implementará un modelo de aprendizaje automático, denominado K-means, que *agrupará la información turística y sitios de interés, y nos permitirá visualizar el mejor sector turístico del vecindario mediante un mapa de Folium.

Para terminar, se realizará el análisis y discusión del modelo implementado en el vecindario y se tomará la decisión de que sector cumple con todas las condiciones establecidas, y además se recomendará el mejor sitio turístico y de interés del vecindario seleccionado.

4. Resultados y discusión

Para determinar cuál fue el vecindario con mejor rendimiento o más apto, en base a las condiciones establecidas en un principio, se utilizaron métodos estadísticos, que permitieron ir descartando vecindarios según se realiza los distintos análisis, como:

Con el gráfico de barras donde se visualizó el número de sitios de interés de cada vecindario, podemos deducir que los barrios de Welmington y Pacific Palisades, con 7 y 6 sitios de interés respectivamente tienen un numero bajo en relación al resto como se visualiza en la Figura 7, por lo que se los considera de baja atracción turística y son descartados.

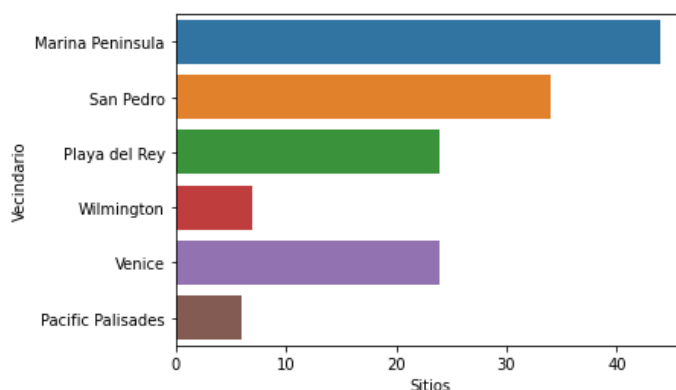


Figura 7. Número de sitios de interés por cada vecindario.

Con las gráficas de frecuencia de sitios de interés por categoría de cada vecindario también se observa que Playa del Rey, cuenta con varios lugares de entretenimiento,

pero pocos sitios de comida o restaurante, como se ve en la Figura 8, de la misma manera San Pedro, cuenta con algunos sitios de restaurantes mexicanos, pero no posee muchos más en variedad, como se ve en la Figura 9, por lo que son descartados debido a que no cumplen con el requisito de tener varios restaurantes de comida.

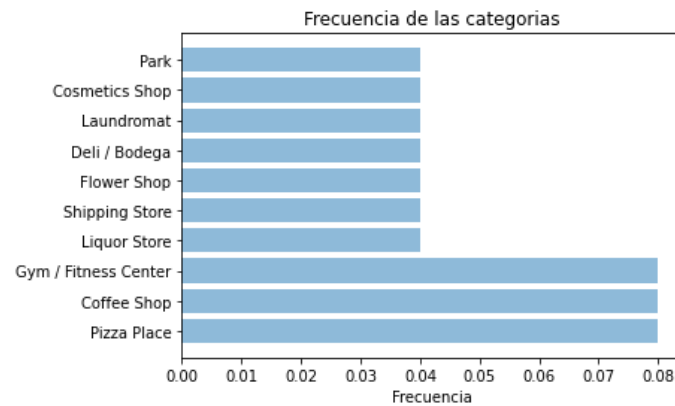


Figura 8. Frecuencia de categorías del vecindario de Playa de Rey

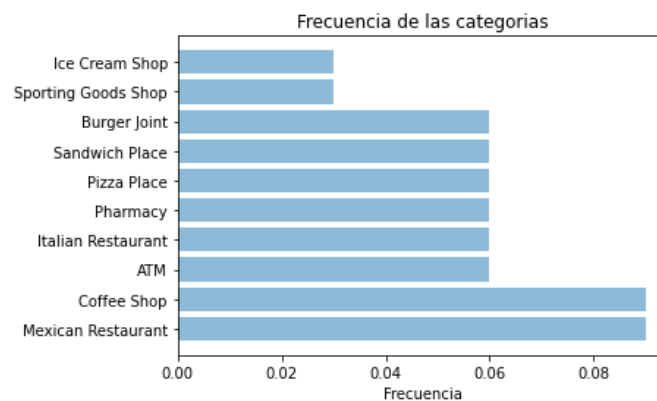


Figura 9. Frecuencia de categorías del vecindario de San Pedro

Con la tabla de los 10 sitios más populares o comunes, como se ve en la Figura 10, se realizó una comparación entre Venice y Marina Península, que tienen en total 24 y 34 de sitios de interés y con buenas frecuencias de lugares turísticos y principalmente variedad en sus restaurantes, sin embargo, Marina Península cuenta con un factor importante, el Muelle, un sitio turístico muy atractivo para los turistas, y la gran cantidad de restaurantes que posee, lo hacen el lugar recomendando para la construcción del hotel.

	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	Marina Peninsula	American Restaurant	Hotel	Bar	Beach	Italian Restaurant	Burger Joint	Mexican Restaurant	Food & Drink Shop	Food	Japanese Restaurant
1	Pacific Palisades	Trail	Home Service	Arcade	Athletics & Sports	Food Truck	Yoga Studio	Fast Food Restaurant	Cuban Restaurant	Deli / Bodega	Dessert Shop
2	Playa del Rey	Coffee Shop	Pizza Place	Gym / Fitness Center	Mexican Restaurant	Mediterranean Restaurant	Massage Studio	Cosmetics Shop	Liquor Store	Laundromat	Deli / Bodega
3	San Pedro	Coffee Shop	Mexican Restaurant	Sandwich Place	Burger Joint	Pharmacy	ATM	Italian Restaurant	Pizza Place	Liquor Store	Convenience Store
4	Venice	American Restaurant	Coffee Shop	Hotel	Board Shop	Hotel Bar	Farmers Market	Dessert Shop	Marijuana Dispensary	Mexican Restaurant	Chinese Restaurant
5	Wilmington	Deli / Bodega	Burger Joint	Mexican Restaurant	Doctor's Office	Tea Room	Donut Shop	Market	Yoga Studio	Farmers Market	Cuban Restaurant

Figura 10. Tabla comparativa de los 10 sitios mas populares de cada vecindario

4.1 Aplicación del modelo de aprendizaje automático

Una vez aplicado distintos métodos estadísticos para realizar un análisis exploratorio de los datos, (ver la sección de resultados y discusión), se puede determinar y eliminar vecindarios, que en comparación con otros se encuentran con desventajas basándonos en las condiciones establecidas previamente, por lo que el mejor vecindario elegido es el de Marina Peninsula.

Ahora aplicamos el modelo de K-means a este vecindario, para determinar la ubicación exacta, que mejor se ajuste a las condiciones del cliente, con los resultados del modelo, se puede determinar este sector exacto para la construcción o implementación, para ello se estudió las 5 agrupaciones obtenidas, donde la agrupación 1, que se encuentra en la calle Washington Boulevard con intersección en Pacific Avenue, cuenta con alrededor de 20 sitios de interés, como se observa en la Figura 11, de color morado, se representan estos lugares, además la ubicación está situada en los alrededores de la playa y el hotel.

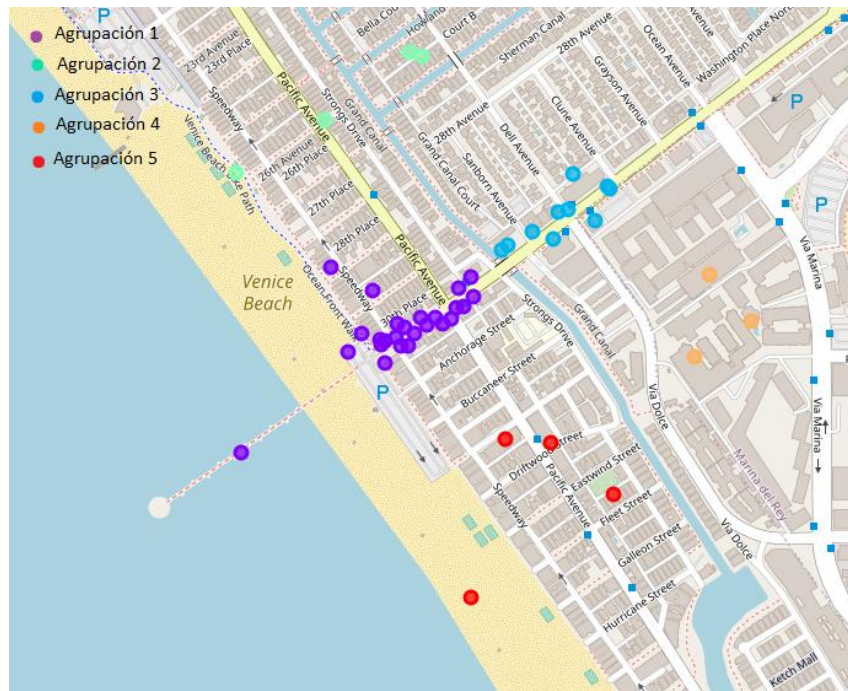


Figura 11 Representación gráfica de todas las agrupaciones obtenidas con el modelo.

5. Conclusión

Como conclusión final de este proyecto, se recomienda totalmente la implementación o construcción de un nuevo hotel 4 o 5 estrellas, en la calle de Washington Boulevard intersección en Pacific Avenue, ubicada dentro del vecindario de Marina Península, perteneciente a Los Ángeles, California, que cuenta con más de 20 sitios de interés, entre los que se encuentra una variedad de restaurantes, sitios turísticos (34 en todo el vecindario), cercanía a la playa, zonas seguras, además de acceso al muelle, lo que lo hace una zona totalmente atractiva a los turistas y potenciales clientes.