



Sebastian Larrea

Batalla de Vecindarios Capstone



Problema

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.

Condiciones

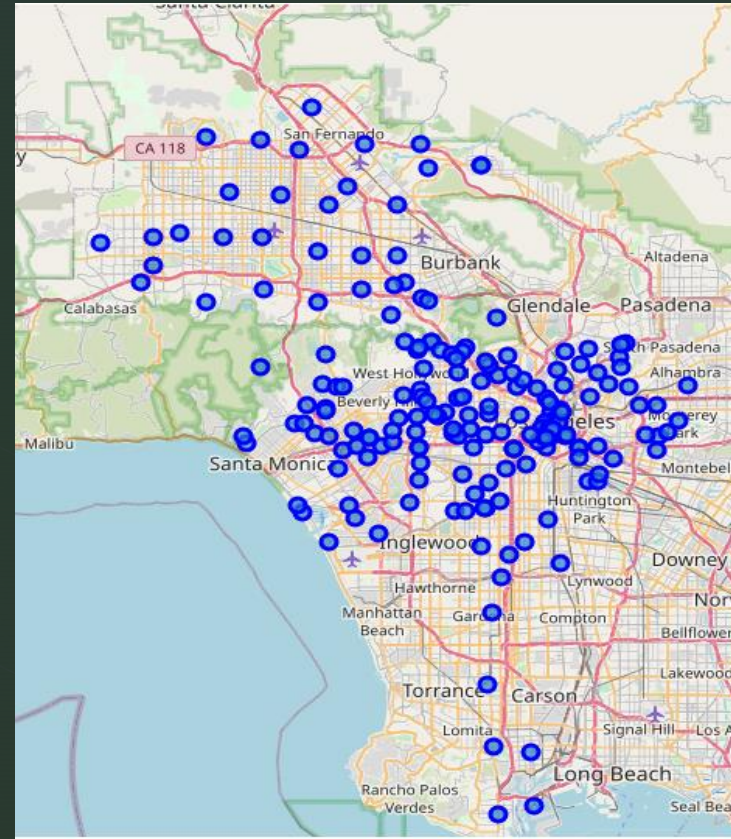
- Debe estar situado cerca de lugares turísticos.
- Debe estar situado cerca de restaurantes y sitios de comida.
- Debe estar situada cerca de playas y lugares de descanso.
- Se debe evitar zonas peligrosas en donde los clientes puedan sufrir de robos o agresiones.

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, con un total de 194 vecindarios pertenecientes a esta ciudad y su distrito, toda esta información fue almacenada en una hoja de cálculo.
- Con esta base de datos, se procedió a implementar la aplicación Geocode, (Complemento de Google Sheets), para 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.

Datos

- 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



Limpieza de los datos

	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

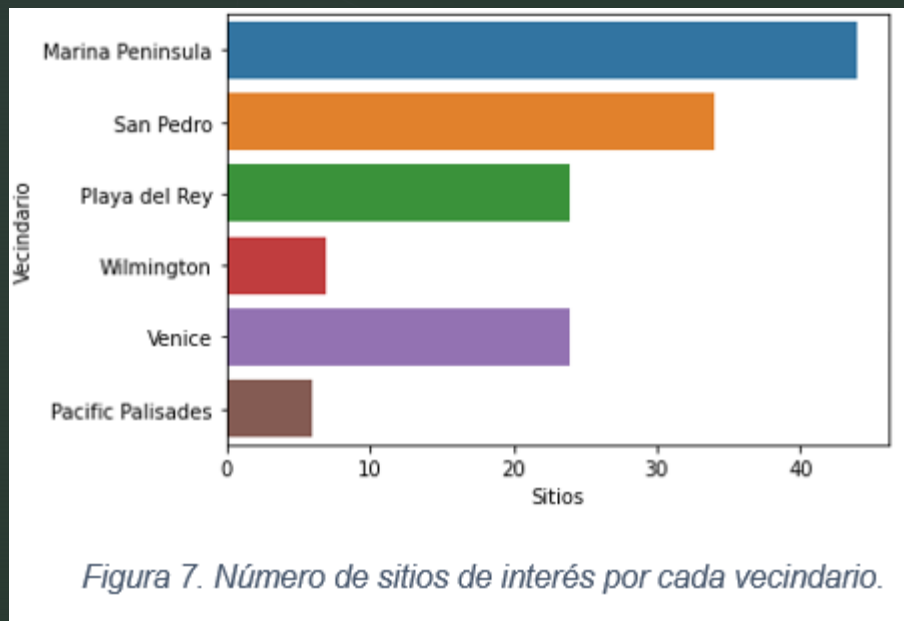
- Para un mejor manejo del data frame, se realizó la eliminación de columnas como "ciudad", ya está entendido que se realiza en el estudio en la ciudad de Los Ángeles, porque esta redundante su permanencia en este.
- Por otro lado, la columna "Full Address", fue utilizada para la obtención de la latitud y longitud de los vecindarios, por lo que su propósito fue cumplido y no se la necesita más para el estudio.

Limpieza de los datos

- Además, con la con la ayuda del mapa superpuesto con los 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 realizara en los vecindarios de Pacific Palisades, Venice, Marina Península, Playa del Rey, San Pedro y Wilmington, que cumplen con esta importante condición.

	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

Análisis



- 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 Wilmington y Pacific Palisades, con 7 y 6 sitios de interés respectivamente tienen un número 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.

Análisis

- 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.

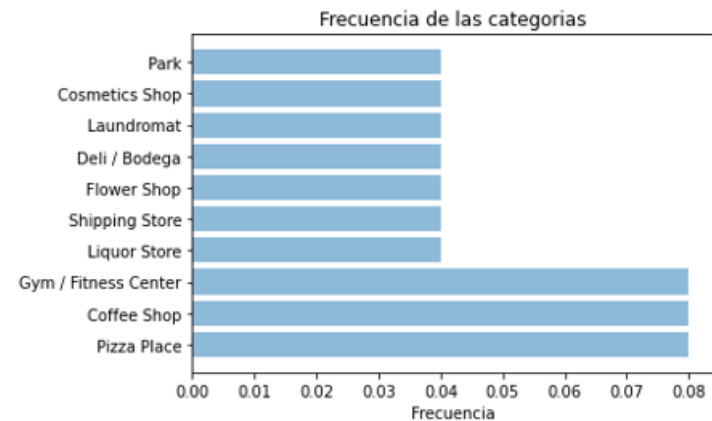


Figura 8. Frecuencia de categorías del vecindario de Playa de Rey

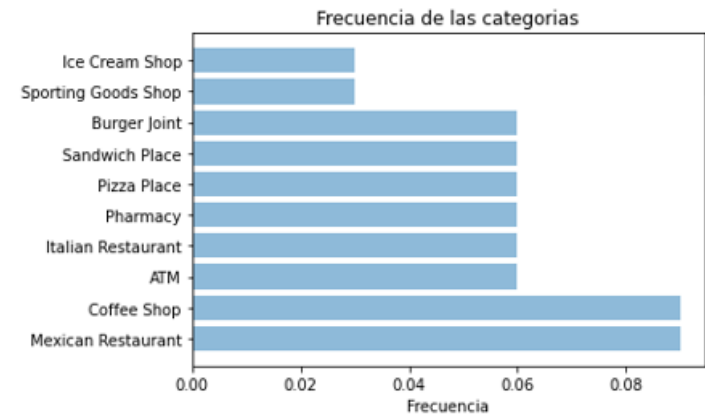


Figura 9. Frecuencia de categorías del vecindario de San Pedro

Análisis

- 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 Peninsula, 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 Peninsula 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

Modelo K-means

- 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.

Modelo K-Means

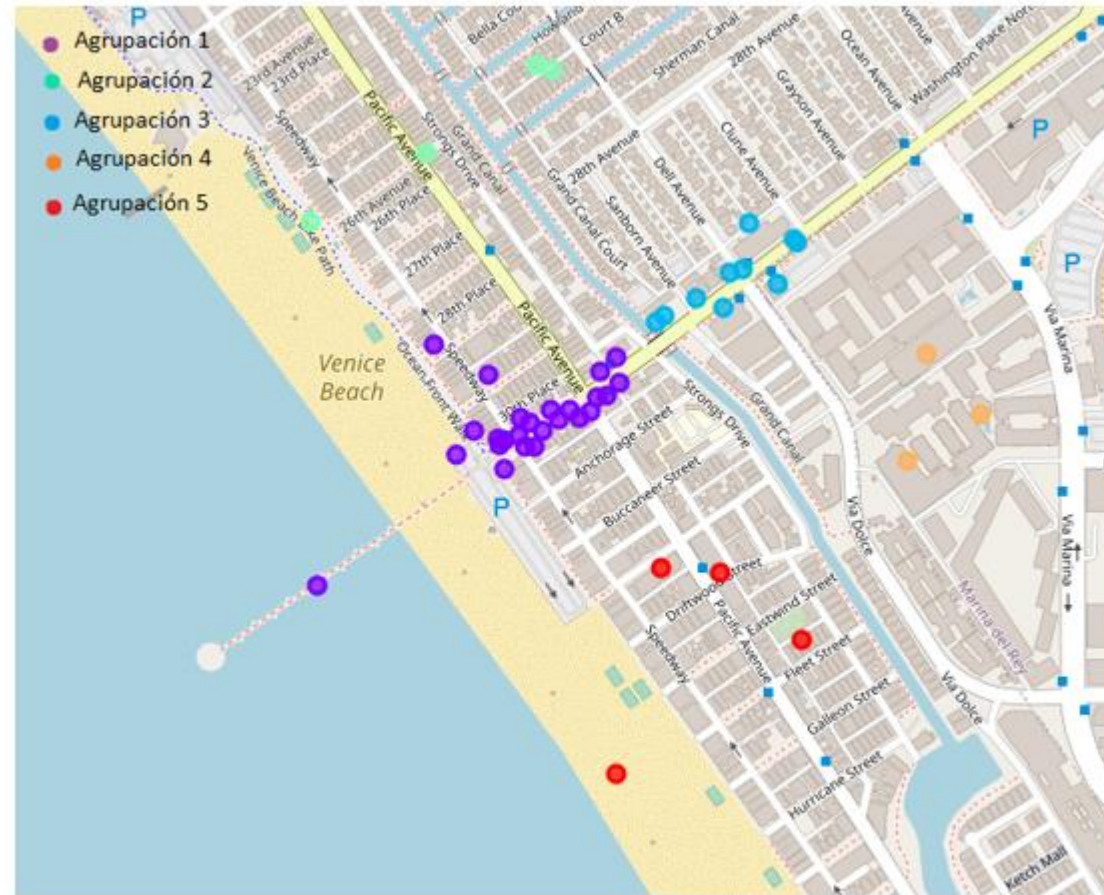


Figura 11 Representación gráfica de todas las agrupaciones obtenidas con el modelo.

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 Peninsula, 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.