



Ingeniería de Sistemas y Computación
Maestría en Ingeniería de la información

MINE 4101 – Ciencia de datos aplicada

Taller 1

Nicolás Esteban Niño Vega

202324436

Usuario GitHub: nicolasninov

Link Repositorio:

https://github.com/nicolasninov/Taller_1_Ciencia_de_Datos/tree/main

2023

OBJETIVO

Ayudar a que una serie inversionistas adquieran propiedades para el alquiler en la plataforma Airbnb, para esto se debe determinar los aspectos que generen una mayor satisfacción en los potenciales usuarios de los alquileres, los cuales son aquellos que tienen un impacto significativo en la percepción de calidad de una propiedad y en la calidad en el sector de la ciudad y, por lo tanto, en el potencial de generar ingresos retorno para el inversionista.

SELECCIÓN Y ENTENDIMIENTO DEL DATASET

El dataset almacena conjunto de datos que contiene información de propiedades de Airbnb, la cual es una plataforma en línea que facilita la oferta y reserva de alojamientos temporales en todo el mundo, para este caso se realizará el estudio sobre la ciudad de Barcelona, España.

El dataset ("listings.csv") está compuesto por un total de 75 columnas con 17230 registros, entre los cuales se encuentran diferentes tipos de datos como interger, float, boolean, date, text entre otros. En primera medida se determinó las 5 variables más importantes, a las cuales se hará una definición y una breve descripción de la información que en ellas reposa.

neighbourhood_cleansed (Barrio codificado): Es una variable categórica nominal la cual hace referencia al nombre y ubicación (polígono) del barrio en la ciudad, es un factor determinante para la inversión ya que va a sectorizar la zona más adecuada para invertir.

La variable neighbourhood_cleansed tiene 70 posibles barrios.

Moda: la Dreta de l'Eixample

Frecuencia Absoluta y Relativa (5 más altos):

Barrio	Frecuencia relativa	Frecuencia absoluta
la Dreta de l'Eixample	2213	0.128439
el Raval	1379	0.080035
el Barri Gòtic	1228	0.071271
Sant Pere, Santa Caterina i la Ribera	1118	0.064887
la Vila de Gràcia	997	0.057864

Tabla 1. Información estadística variable neighbourhood_cleansed

review_scores_rating (Puntuación general de las reseñas): Es una variable cuantitativa discreta que muestra la puntuación que da el usuario en líneas generales del servicio adquirido, adicional, posiblemente el puntaje más común puede diligencia un usuario ya que generaliza su estadía.

Moda	5.0
Media	4.5334
Mediana	4.67
Varianza	0.38076922114773104

Desviación Estándar	0.6170650056094018
---------------------	--------------------

Tabla 2. Información estadística variable review_scores_rating

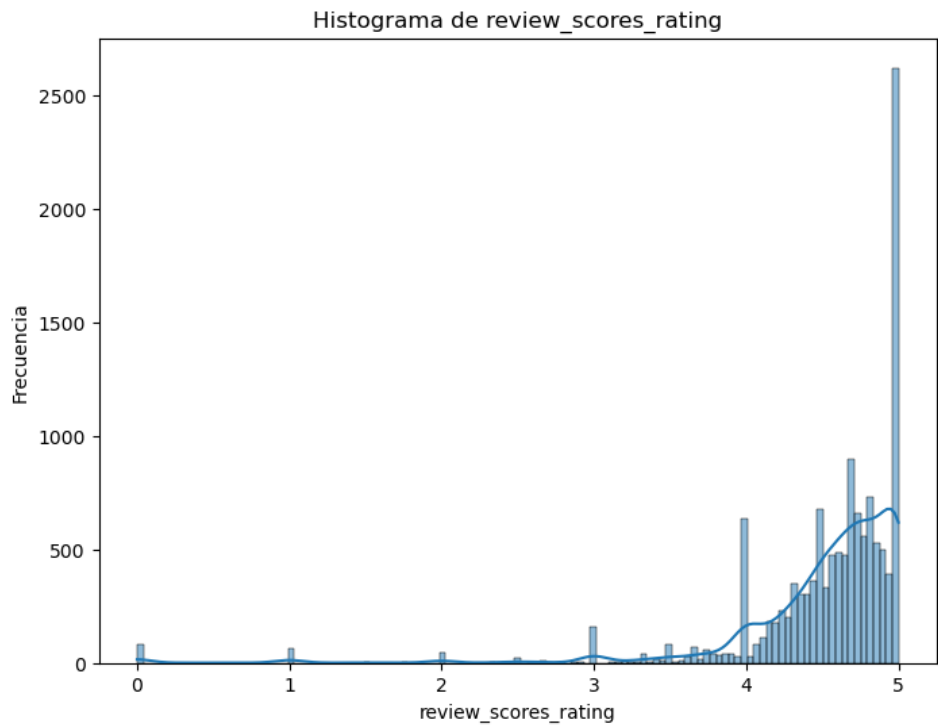


Ilustración 1. Histograma review_scores_rating

review_scores_value (Puntuación del precio): Es una variable cuantitativa discreta que hace referencia al puntaje dado en relación al valor dado pagado por el usuario en su estadía, esta variable tiene su relevancia ya que categoriza en un rango el factor precio según la estadía.

Moda	5.0
Media	4.451830180042721
Mediana	4.57
Varianza	0.29794282398570415
Desviación Estándar	0.5458413908689081

Tabla 3. Información estadística variable review_scores_value

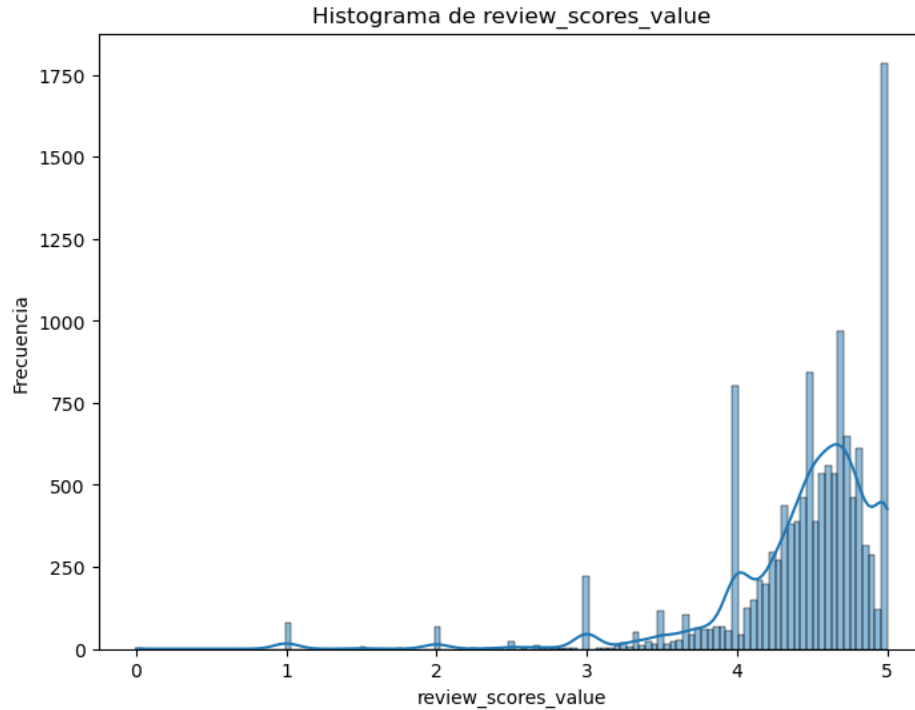


Ilustración 2. Histograma review_scores_value

room_type (Tipo de habitación): Es una variable cualitativa nominales y hace referencia a los cuatro tipos de habitaciones que se encuentra en el lugar de alojamiento.

- Todo el lugar (Entire home/apt): Son aquellos en los cuales se alquila la totalidad de una propiedad e incluye dormitorios, baños, cocina, entradas independientes y exclusivas.
- Habitación privada (Private room): Son aquellas propiedades en las cuales las personas comparten algunas zonas comunes como sala, cocina o baños, pero se dispondrá de una habitación exclusiva para el usuario.
- Habitación compartida (Shared room): Son aquellas propiedades en las que además de compartir zonas como sala, cocina o baños se comparte la habitación.
- Habitación de hotel (Hotel room): Aquellas habitaciones que se encuentren en hoteles, pueden o no compartir zonas como baños.

Moda: Entire home/apt (todo el lugar)

Frecuencia Absoluta y Relativa:

Barrio	Frecuencia relativa	Frecuencia absoluta
Entire home/apt	10262	0.595589
Private room	6675	0.387406
Shared room	150	0.008706
Hotel room	143	0.008299
Total	17230	1

Tabla 4. Información estadística variable room_type

Latitude y longitud (Latitud y longitud): Estas variables dependen la una de la otra ya que dan la ubicación geográfica a una propiedad, son variables categóricas nominales, estas son importantes para determinar la inversión ya que determinan espacialmente la ubicación de la propiedad en relación a lugares de interés de la ciudad.

CALIDAD Y LIMPIEZA DE DATOS

Realizando una revisión general de los datos se pudo observar que en el barrio 'la Clota' solo se tiene solo un registro y no se tiene valores en las variables reseñas, esto puede ser un problema ya que al realizar los análisis posteriores puede influir de forma negativa en los resultados, por lo cual, se decide omitir dicha información.

ESTRATEGIA DE ANÁLISIS

Se desea buscar a través de análisis estadísticos y relación entre variables una serie base escenarios donde se de respuestas a las posibles preguntas que pueda tener el inversionista y así, que puedan tomar la mejor decisión, para ello se inicia realizando el cálculo de la tasa de ocupación ya que este es un indicador de las dinámicas de la frecuencia de alquiler que tienen las propiedades en el año asociado con los respectivos barrios.

Adicional se quiere conocer que barrios cuentan con los mejores puntajes en reseñas de valor y puntajes en reseñas en general, ya que esto ayuda a entender el comportamiento del mercado con base a la información del usuario, reduciendo riesgos debido a las buenas referencias con las que ya cuentan estos barrios. Otro análisis que se puede realizar la relación del puntaje en la limpieza de la propiedad con el tipo de habitación, esto a través de un diagrama de cajas (boxplot) el cual indica que relación ofrece el tipo de habitación con su calificación en limpieza, luego relacionaremos dichos resultados con el puntaje en la reseña del valor, con el fin de visualizarlos a través de un gráfico de dispersión el cual indicara como la puntuación de limpieza se relaciona con el tipo de habitación en función del puntaje del precio.

También se va a analizar el precio por noche que los usuarios estarían dispuestos a pagar dependiendo el tipo de habitación y el barrio, esto puede mostrarle al inversionista los rangos de precios que podría cobrar por el alquiler de cada tipo de habitación y en qué sector de la ciudad puede tener una mayor rentabilidad. Por último y a través de los atributos latitud y longitud se va a determinar donde hay una mayor concentración de propiedades ubicándolas espacialmente a través de un mapa de calor, esto podrá asociarse con sitios turísticos que pueden ser de atractivo para los usuarios y que sirva de forma visual para que el inversionista tenga un panorama más concreto del sector.

DESARROLLO DE LA ESTRATEGIA

Se inicia con el calculo de la tasa de ocupación con base a la siguiente ecuación:

$$Tasa\ de\ ocupaci3n = \left(\frac{Noches\ ocupadas}{Noches\ disponibles} \right) \times 100$$

Donde:

Noches ocupadas = (365 – "availability_365") y Noches disponibles = "availability_365"

La variable “availability_365” indica la cantidad de días que una propiedad está disponible para ser reservada durante el año. Posteriormente agruparemos esta información con base a los barrios y así poder determinar cuales tienen una mayor tasa de ocupación, se obtuvo:

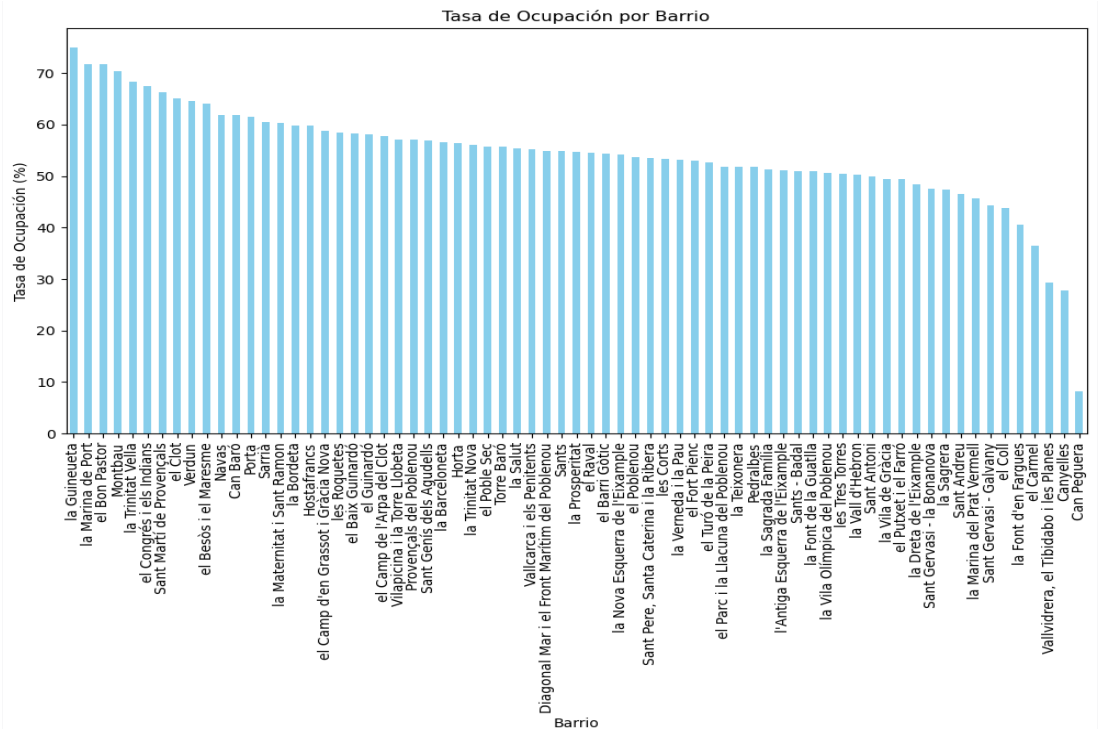


Ilustración 3. Tasa de ocupación por barrio

Los barrios donde se tiene una mejor tasa de ocupación son:

Barrio	Tasa ocupación
la Guineueta	74.916286
la Marina de Port	71.753425
el Bon Pastor	71.735160
Montbau	70.273973
la Trinitat Vella	68.356164

Tabla 5. Barrios con mejor tasa de ocupación

Ahora se evalúa en donde están los mejores puntajes en reseñas en general y puntajes en reseñas en valor (review_scores_rating y review_scores_value) según el barrio (neighbourhood_cleansed), para realizar este proceso se procedió a calcular la media review_scores_rating y la mediana de review_scores_value, luego se relacionará con los barrios y se visualizará en una gráfica para cada comparación.

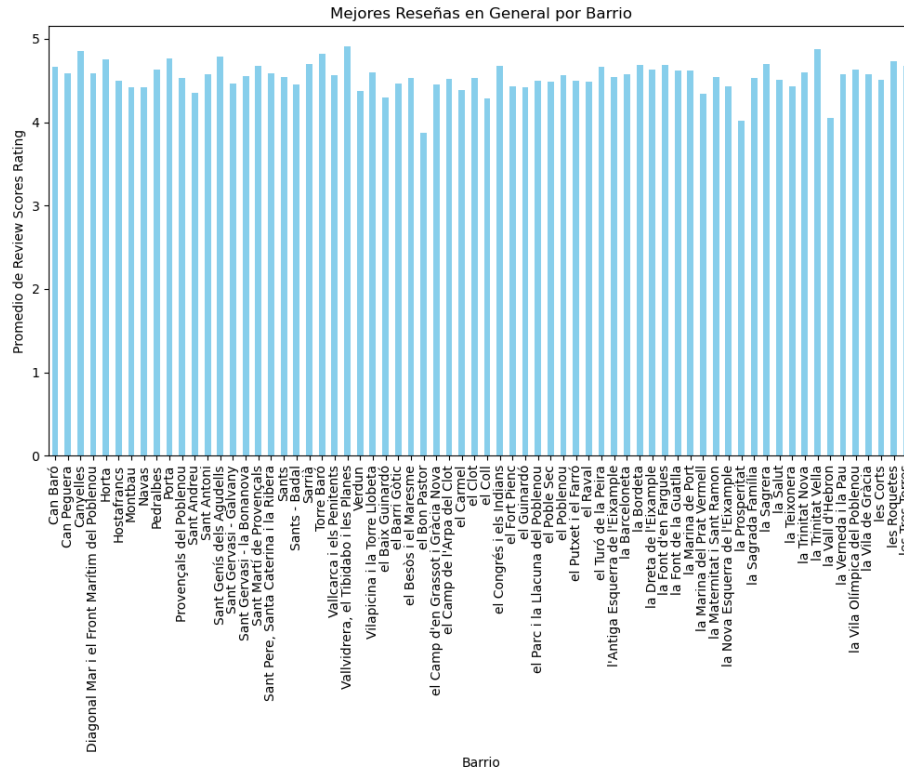


Ilustración 4. Gráfica de relación review_scores_rating con los barrios

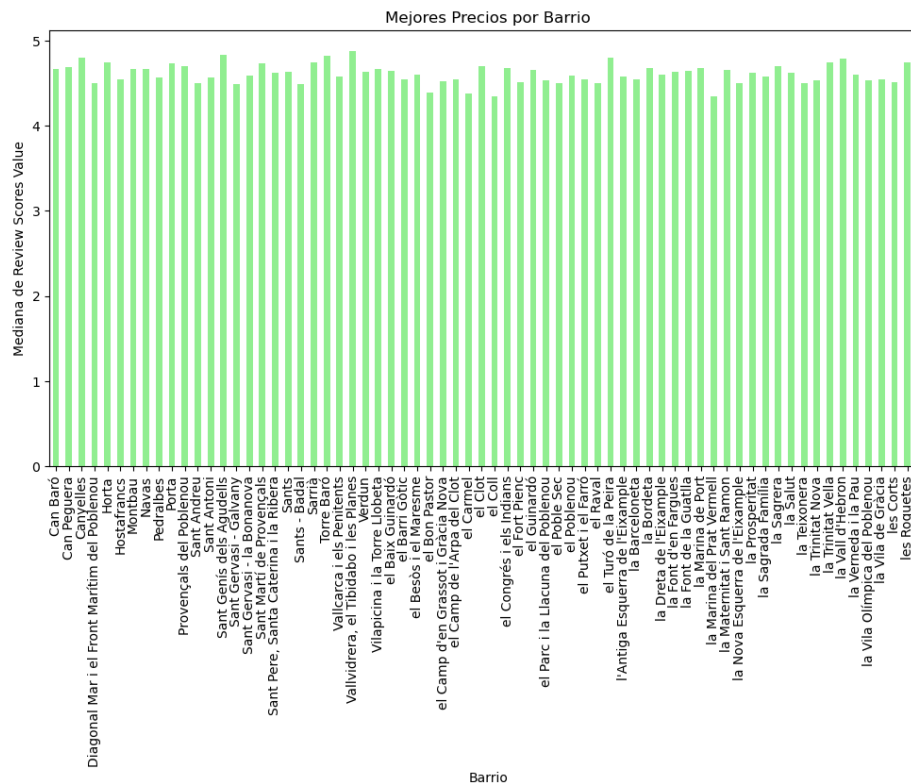


Ilustración 5. Gráfica de relación review_scores_value con los barrios

De las dos ilustraciones anteriores se puede observar que los barrios con mejor calificación en reseñas en general y precios son:

- 'Vallvidrera, el Tibidabo i les Planes'
- 'la Trinitat Vella'
- 'Canyelles'
- 'Torre Baró'
- 'Sant Genís dels Agudells'

Es importante resaltar que como se realizó en la etapa de Calidad y Limpieza no aparece el valor para el barrio 'la Clota' ya que generaba una información errónea.

También se desea saber la relación del tipo de habitación (room_type) con mejor puntaje en la reseña de la limpieza y puntaje con la reseña del valor (review_scores_cleanliness y review_scores_value) para lo cual se realizó un histograma que relacione las variables anteriormente mencionadas y se obtuvo:

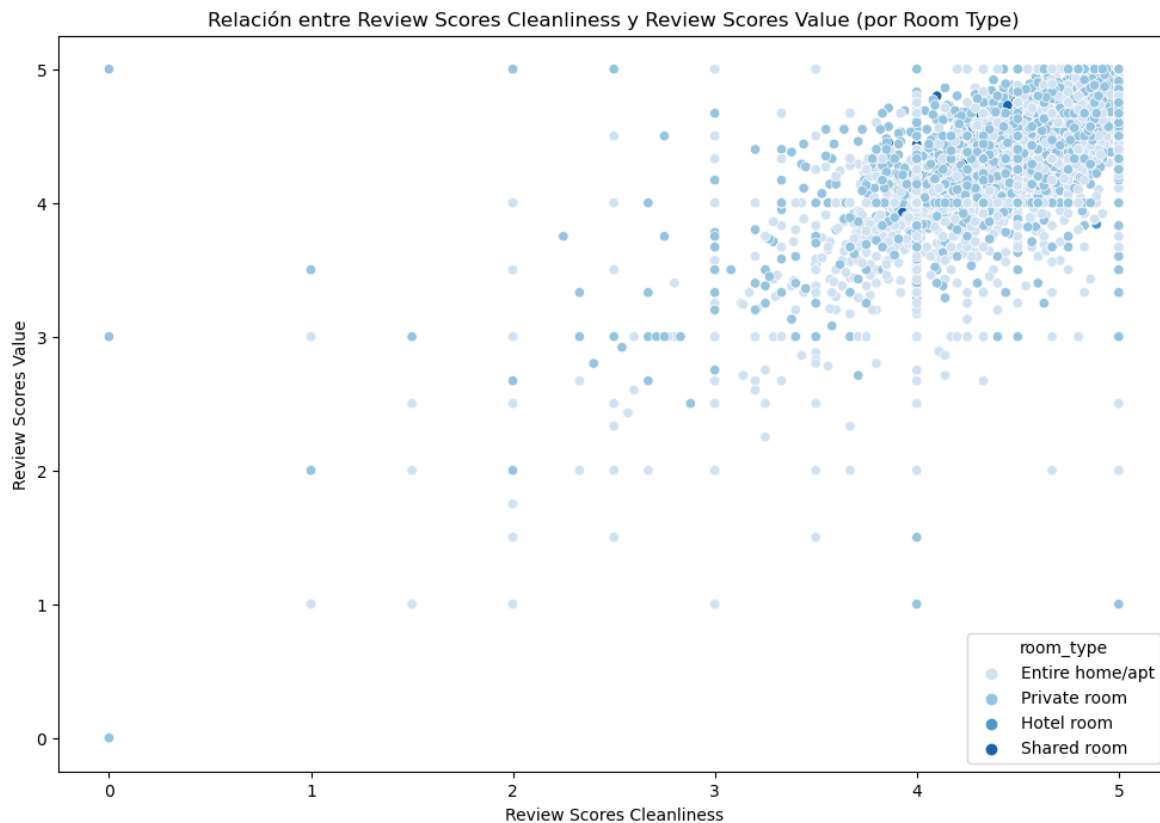


Ilustración 6. Histograma de relación entre review_scores_value y review_scores_cleanliness con los room_type

En la gráfica se observa que la mejor puntuación se obtuvo para las habitaciones tipo hotel y privadas. Esto sirve para que el inversionista sepa que tipo de habitación tiene mejores reseñas en función a la limpieza y valor.

Otro análisis realizado fue el cálculo de los precios (Price) que el usuario está dispuesto a pagar según el barrio (neighbourhood_cleanse) y tipo de habitación (room_type), para ello

primeramente se convirtió el precio a tipo numérico, luego se calculó el rango intercuartílico con el fin de eliminar valores atípicos y se mostraron los rangos mínimos y máximos de los precios para finalizar con una gráfica y se obtuvo:

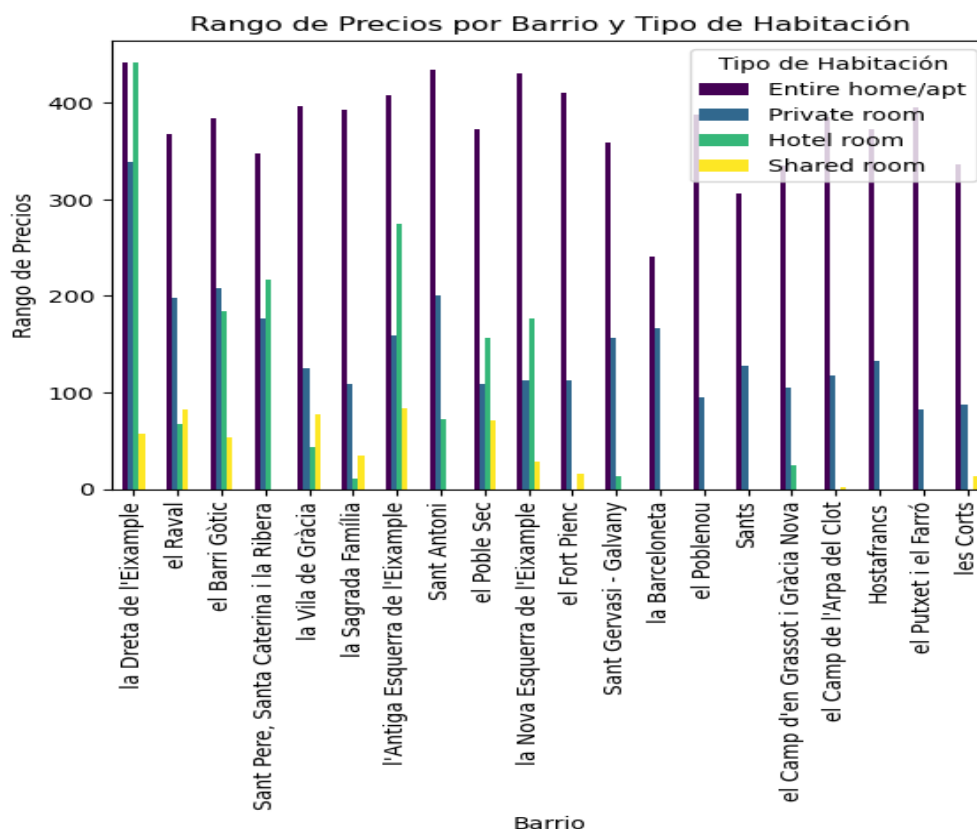


Ilustración 7. Gráfica de rango de precios con base a los barrios y tipo de habitación.

- El barrio con precio más alto para Entire home/apt: la Dreta de l'Eixample (Rango de Precios: 442.0).
- El barrio con precio más alto para Private room: la Dreta de l'Eixample (Rango de Precios: 339.0).
- El barrio con precio más alto para Hotel room: la Dreta de l'Eixample (Rango de Precios: 441.0).
- El barrio con precio más alto para Shared room: l'Antiga Esquerra de l'Eixample (Rango de Precios: 84.0)

Por último, se realizó un mapa de calor el cual concentra las zonas con mayor aglomeración de predio, para ello se eliminó valores atípicos de las coordenadas con base a los rangos intercuartílicos para su posterior visualización.

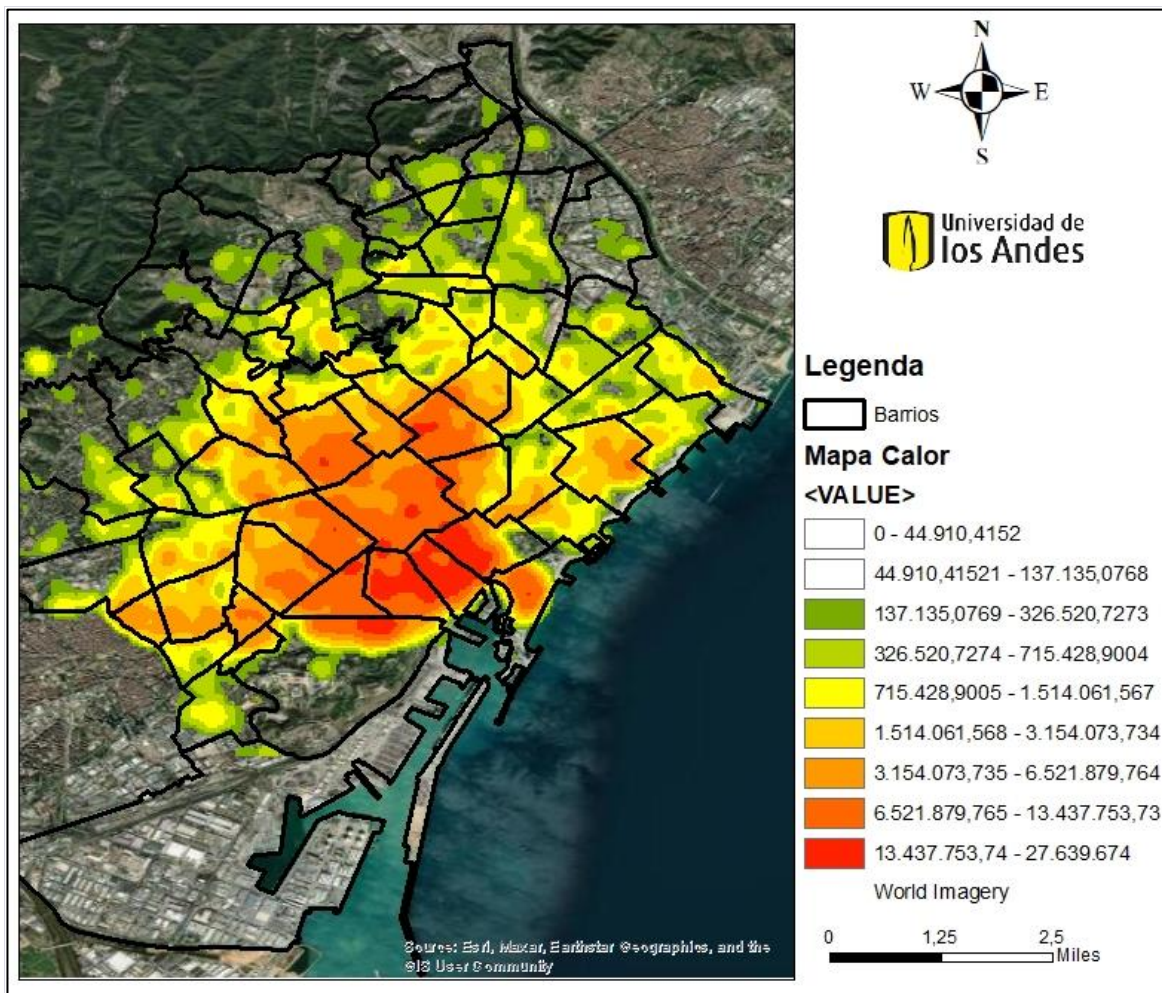


Ilustración 8. Mapa de calor de propiedades en la plataforma Airbnb.

Visualmente se puede determinar que los barrios con mayor aglomeración de propiedades destinadas para la plataforma Airbnb son:

- Sant Pere, Santa Caterina i la Ribera.
- Sant Antoni.
- la Sagrada Família.
- la Dreta de l'Eixample.
- el Raval.
- el Fort Pienc.
- el Barri Gòtic

GENERACIÓN DE RESULTADOS

Teniendo en cuenta los análisis previamente realizados, en los que se evaluaron una serie conjunto variables como precio, reseñas generales de los usuarios, tipo de habitaciones, tasa de ocupación, ubicación. Se determinó que los barrios en donde se recomienda invertir son:

- la Trinitat Vella.

- l'Antiga Esquerra de l'Eixample.
- Sant Antoni.
- la Dreta de l'Eixample.
- el Raval.
- Sant Andreu.

El tipo de habitación y los rangos de precios en las mismas son:

- Private_room con rangos entre los 75 a 135 unidades monetarias
- Hotel_room con rangos entre los 150 a 250 unidades monetarias

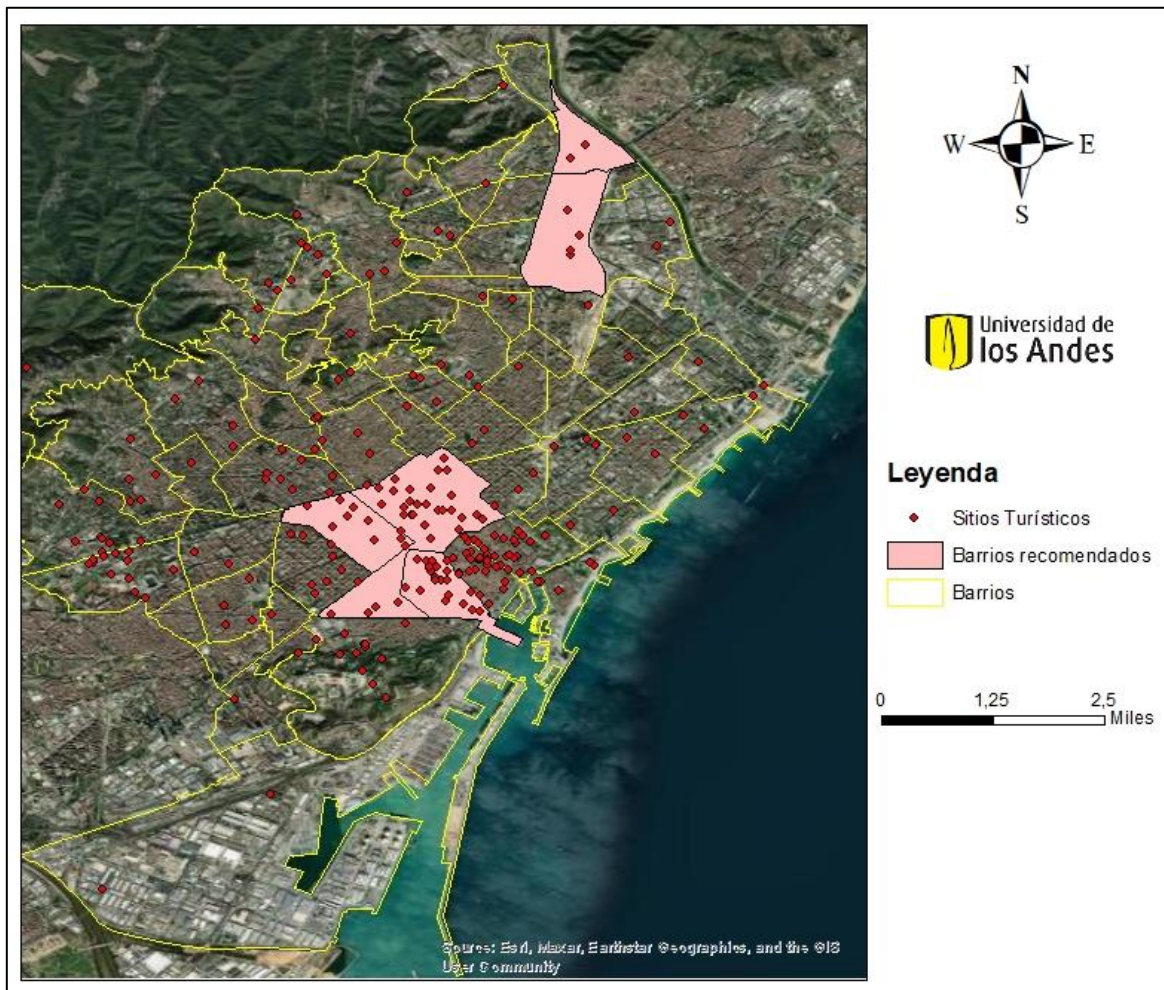


Ilustración 9. Mapa de barrios recomendados para invertir.

Como se puede en la imagen cuatro de los seis barrios que se recomiendan quedan a cortas distancias de la zona de la playa, adicional son barrios donde convergen varios sitios turísticos de la ciudad, adicional es importante tener en cuenta a los otros dos barrios que cuentan con una buena tasa de ocupación y, que al estar en la periferia de la ciudad es posible que las propiedades tengan un menor costo.