



ENTREGA FINAL
Lucía Milagros Torre
Data Science - CODERHOUSE

AGENDA

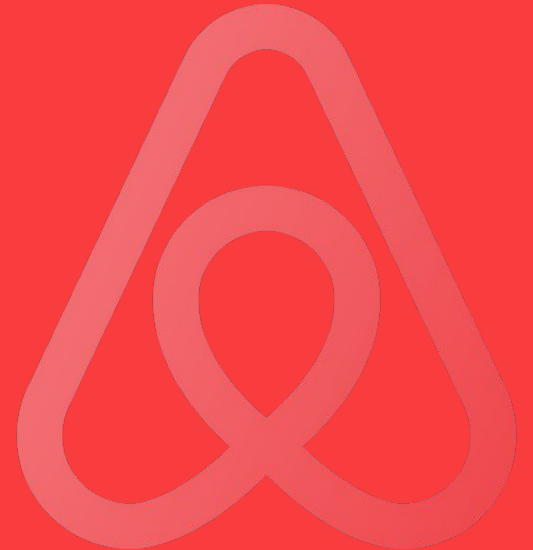
01.CONTEXTO Y AUDIENCIA

02.OBJETIVO/ PREGUNTA DE INTERÉS

03.ANÁLISIS EXPLORATORIO

04.APLICACIÓN DE ALGORITMOS DE ML

05.CONCLUSIONES





01.CONTEXTO Y AUDIENCIA

Airbnb (Air Bed And Breakfast), aplicación utilizada para alojamiento en vacaciones y actividades turísticas.

La base de datos describe los anfitriones (host), los alquileres que poseen, la ubicación del alquiler (describiendo país, latitud, longitud, zona y demás datos), su política de cancelación, tipo de habitación, en qué año fue construída, precio, tarifa de servicio, mínimo de noches requeridas, evaluaciones realizadas (reviews) y reglas de alquiler. Siendo esta información perteneciente a los alquileres situados en la ciudad de Nueva York, Estados Unidos.

Se busca informar a los altos ejecutivos de Airbnb en Nueva York.

CONTEXTO COMERCIAL

Debido a la diferencia en la preferencia de los clientes de Airbnb en Nueva York se desea conocer cuáles son las variables más favorables para un cliente para que los altos ejecutivos puedan orientar inversiones futuras y potenciar su crecimiento.



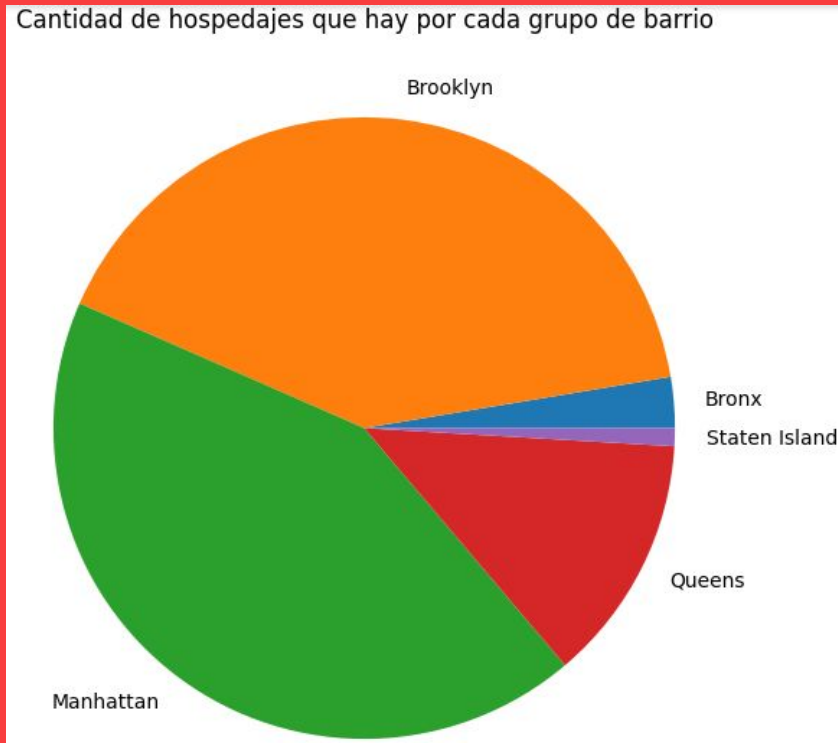
02.OBJETIVO/PREGUNTA DE INTERÉS

Determinar qué lugares son los más elegidos para hospedarse, en la ciudad de Nueva York y por qué, qué variables pueden influir en la elección del cliente.

¿Existen características o variables que influyen en las elecciones de los clientes a la hora de realizar una reserva?



03.ANÁLISIS EXPLORATORIO



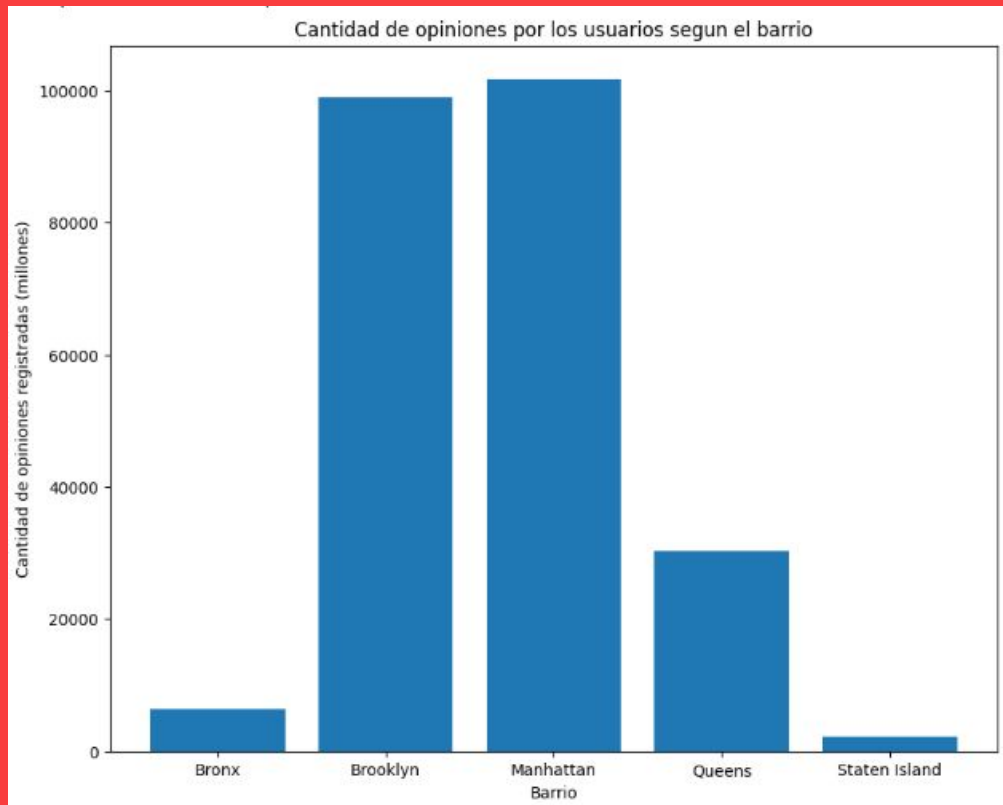
Se observa la cantidad de hospedajes que hay en cada grupo de barrio, siendo los más predominantes **Brooklyn** y **Manhattan**.



03.ANÁLISIS EXPLORATORIO

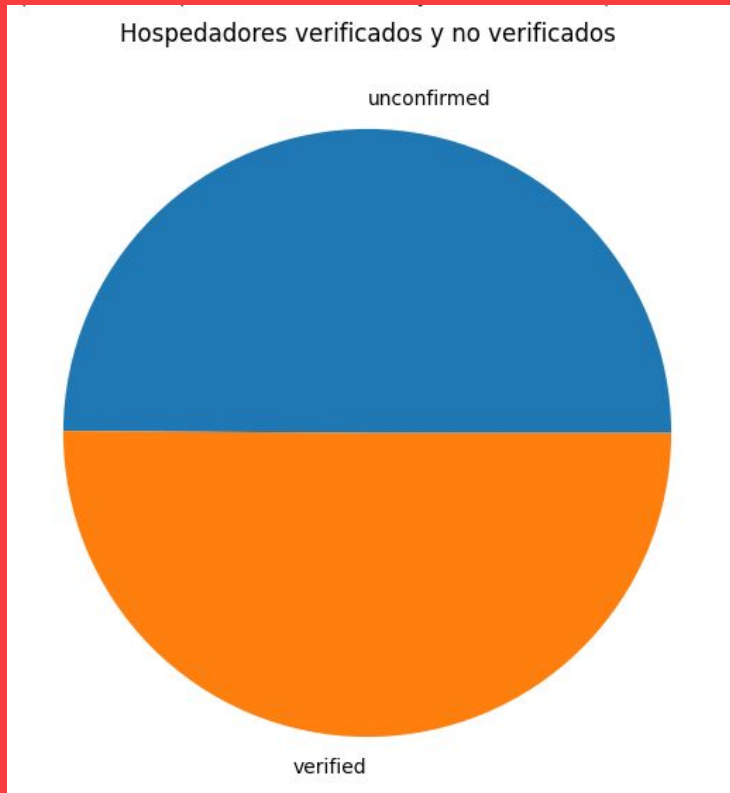
Relación entre la cantidad de opiniones/evaluaciones (reviews) y los distintos tipos de barrio.

-**Manhattan** y **Brooklyn** son los grupos de barrios más evaluados y por lo tanto, los **más visitados**.





03. ANÁLISIS EXPLORATORIO



Se observa la cantidad de hospedadores con cuentas de identidad verificada y no verificada.

-No existen preferencias a la hora de hospedarse de alquilar a un hospedador de cuenta verificada o no.

-Es innecesario verificar la cuenta del hospedador.

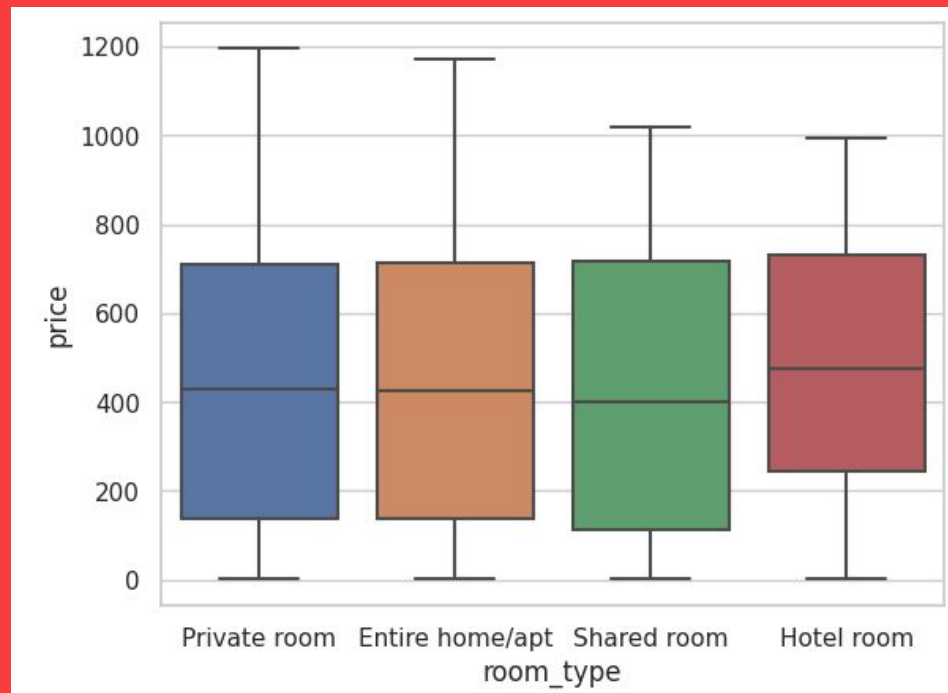


03. ANÁLISIS EXPLORATORIO

Se observa la variación de precio dependiendo el tipo de habitación.

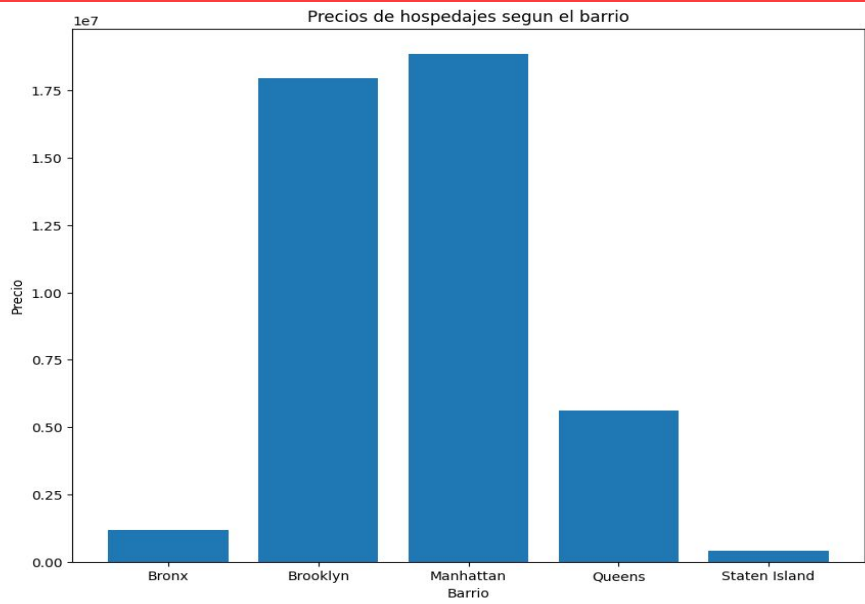
-Las **habitaciones de hotel** son las que poseen los precios **más elevados** y las **habitaciones compartidas** las que poseen los **menores precios**.

-Sin embargo, **no se observan notorias diferencias** en los distintos tipos de habitaciones existentes.





03. ANÁLISIS EXPLORATORIO

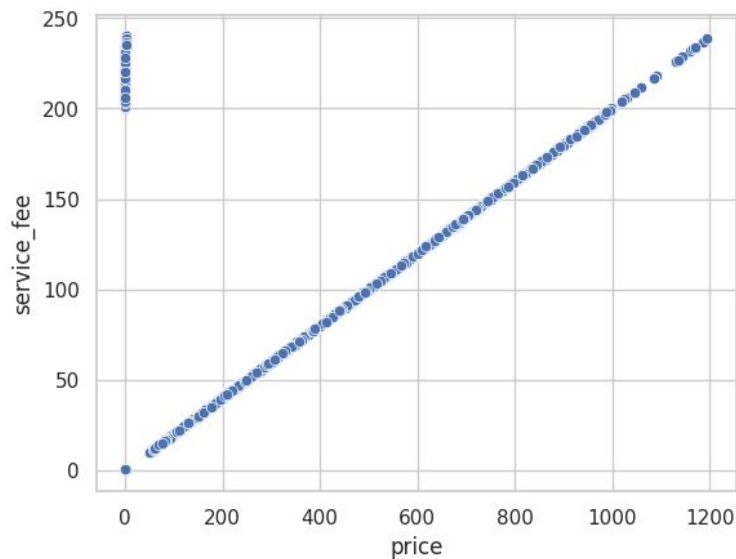


Diferencia de precios de hospedaje en los distintos grupos de barrios.

Manhattan y Brooklyn poseen los **precios más elevados**.

A mayor precio, mayor es la tarifa recaudada por Airbnb.

Resulta conveniente invertir en Brooklyn y Manhattan.





04.APLICACIÓN DE ALGORITMOS DE ML

- Se aplicaron algoritmos de regresión
 - Regresión Lineal
 - Regresión Polinómica (grado 4)
 - Árbol de decisión

Se buscó predecir la escala en el número de evaluaciones realizadas del 1 al 5 (`review_rate_number`), teniendo en cuenta distintas variables numéricas como son: el año de construcción, el precio, la tarifa de servicio, el mínimo de noches requeridas y la cantidad de evaluaciones realizadas.



05.CONCLUSIONES

- Según los valores de MSE obtenidos y luego de aplicar GridSearchCV se concluye que el mejor modelo es el de regresión lineal.
- Aún así en los distintos análisis exploratorios realizados no se observa una buena asociación entre las variables usadas y la variable a predecir.
- Además de esto, en la limpieza de datos, el dataset tenía muchos outliers que fueron eliminados y reemplazados por los sus correspondientes valores de la mediana, lo cual perjudica en su análisis.
- No se observa diferencia alguna entre cuentas verificadas o no verificadas.
- No se observa diferencia de precios notoria en los distintos tipos de habitaciones.
- Si se remarca la preferencia de los grupos de barrio de Brooklyn y Manhattan, debido a que son los más evaluados y los que poseen mayor precio y por lo tanto mayor tarifa de servicio.
- Se recomienda orientar las inversiones futuras a los grupos de barrio Brooklyn y Manhattan.

