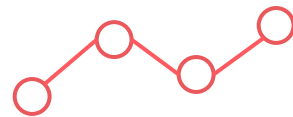


Práctica final de Big Data

Análisis de Datos



KeepCoding

Glovo - Woman in Tech

Glovo

El Equipo: Data Squad



Airbnb

Análisis de Datos de la ciudad de
Madrid

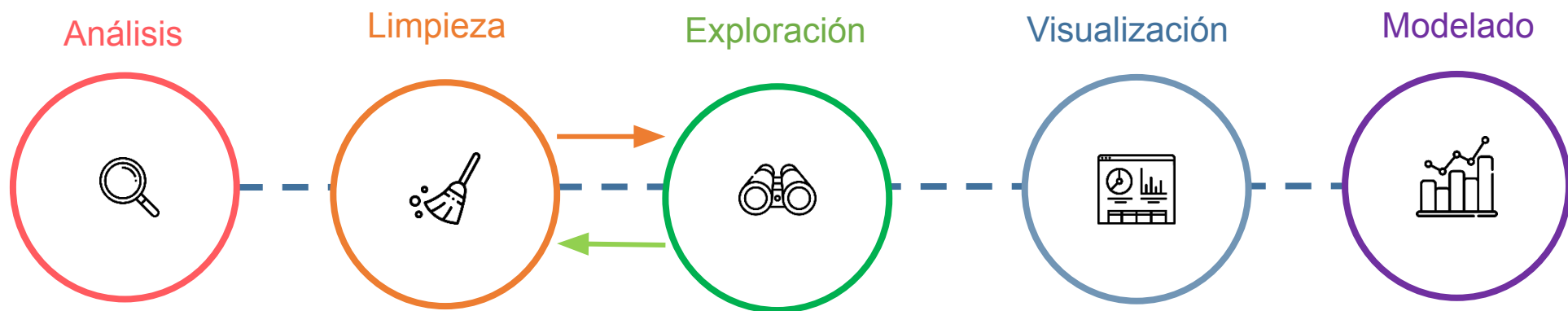


Objetivo:



Pensamos en el desarrollo de un producto digital que **mejorara la experiencia del usuario**, que pudiera mantenerse en el tiempo y que fuese capaz de **adaptarse a las fluctuaciones** del mercado mediante **parámetros dinámicos**

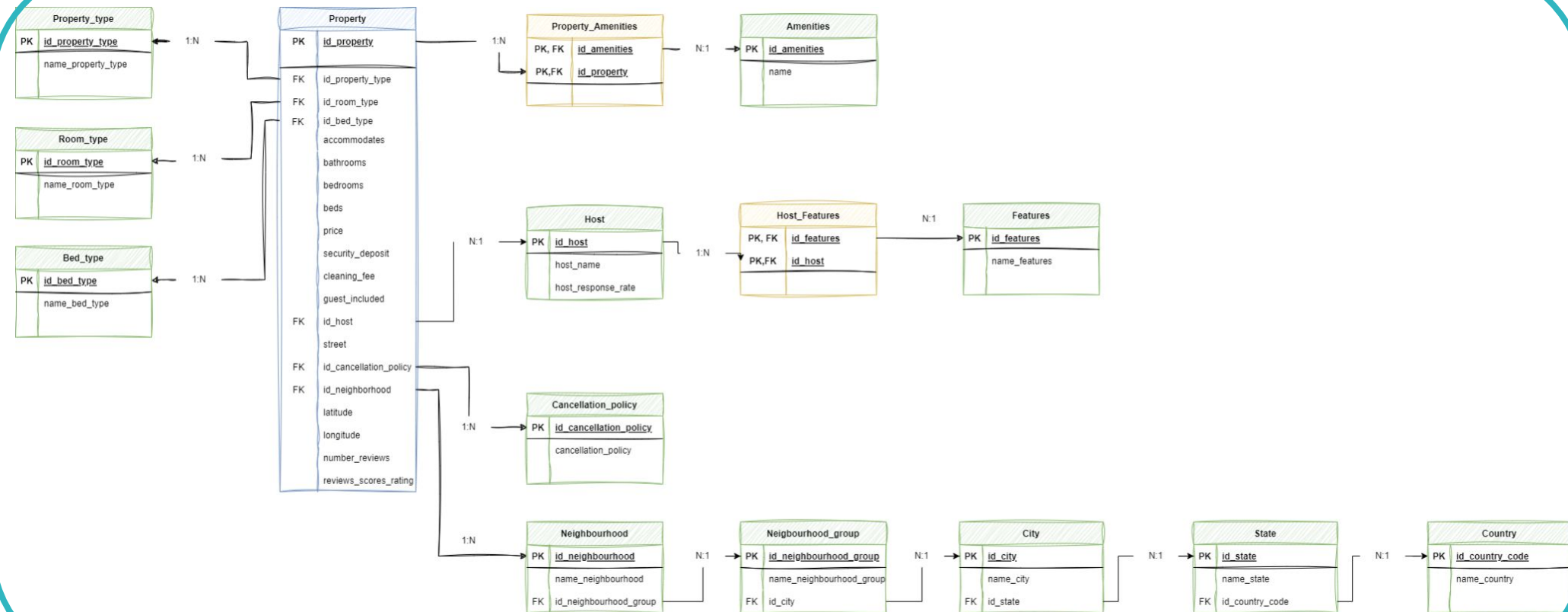
Proceso: Airbnb de Madrid



Análisis: Recolección y Familiarización de los datos



airbnb_madrid_clean	
1/5	column1
12/	id
12/	Host ID
100	Host Name
1/5	Host Response Rate
100	street
100	Neighbourhood Cleansed
100	Neighbourhood Group Cleansed
100	city
100	state
100	zipcode
100	country
12/	latitude
1/5	longitude
100	Property Type
100	Room Type
12/	accommodates
12/	bathrooms
1/5	bedrooms
1/5	beds
100	Bed Type
12/	price
12/	Security Deposit
1/5	Cleaning Fee
1/5	Guests Included
1/5	Number of Reviews
12/	Review Scores Rating
100	Cancellation Policy
100	Host Is Superhost
100	Host Identity Verified
100	kitchen
100	internet
100	Air conditioning
100	heating
100	washer
100	dryer
100	elevator
100	Wheelchair accessible
100	tv
100	pool
100	24-hour check-in
12/	Amenities Score
100	Amenities Rating



Visualización: Exploración y visualización

Informaciones generales

de los Airbnb de Madrid

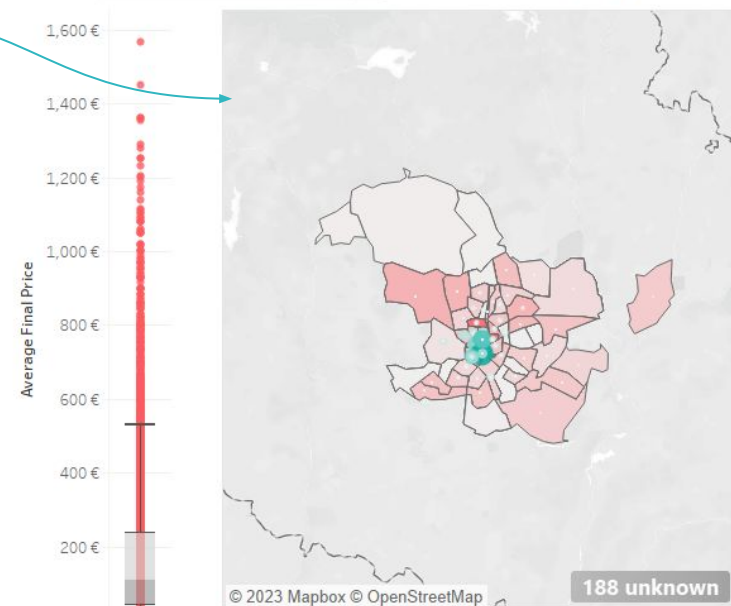
Filtros

Neighbourhod y Map Type

Airbnb: Madrid

Exploring neighborhoods of Madrid in terms of price, availability, popularity and accessibility.

Neighbour.. (All) Map Type Map Availability



Mapas

Availability,
Popularity,
Accessibility

Boxplot

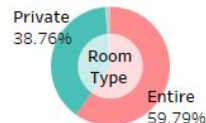
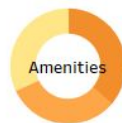
Final Price

Gráficos de distribución

Amenities y
Room Type

Amenities Rating

- A - Excelent
- B - Very Good
- C - Good



Final Price Mean	Total Listing	Total SuperHost	Review Scores	Number of Reviews
278,96 €	13,247	1,556	92	306,099

Top Charts

Top Price

Neighbourhood ..

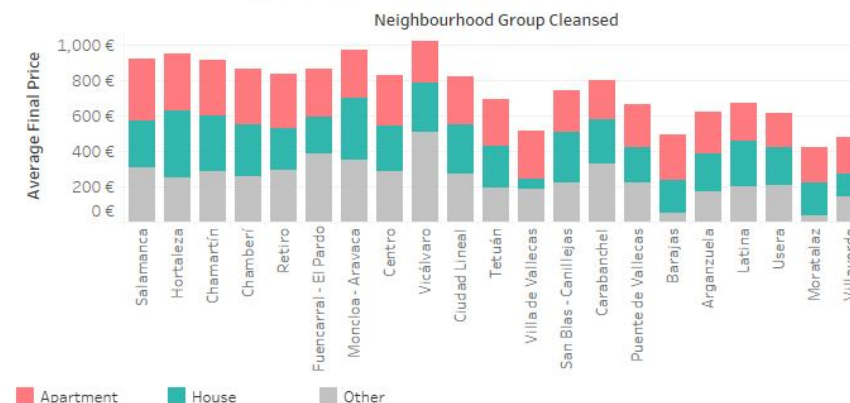


Top 5

Final Price, Availability,
Popularity

Comparison Charts

Property Type x Price

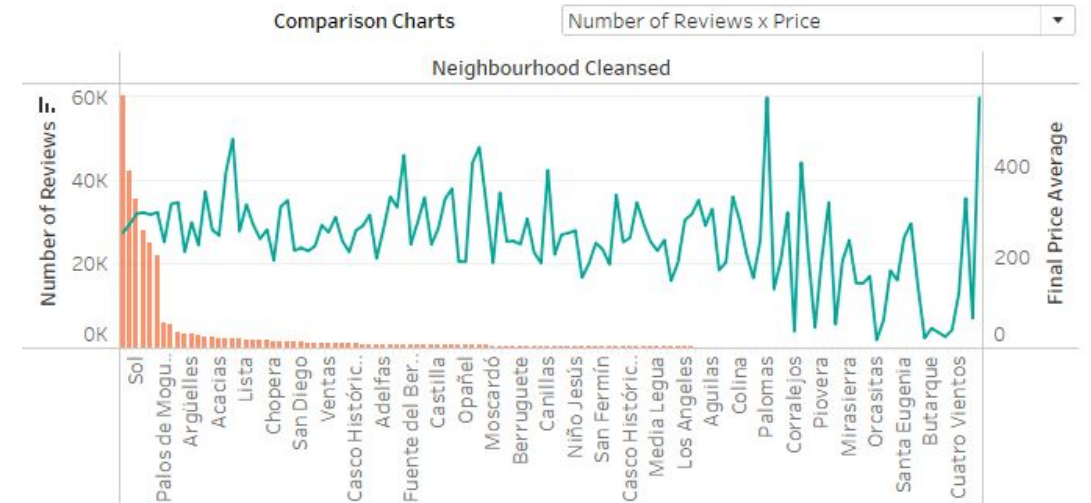
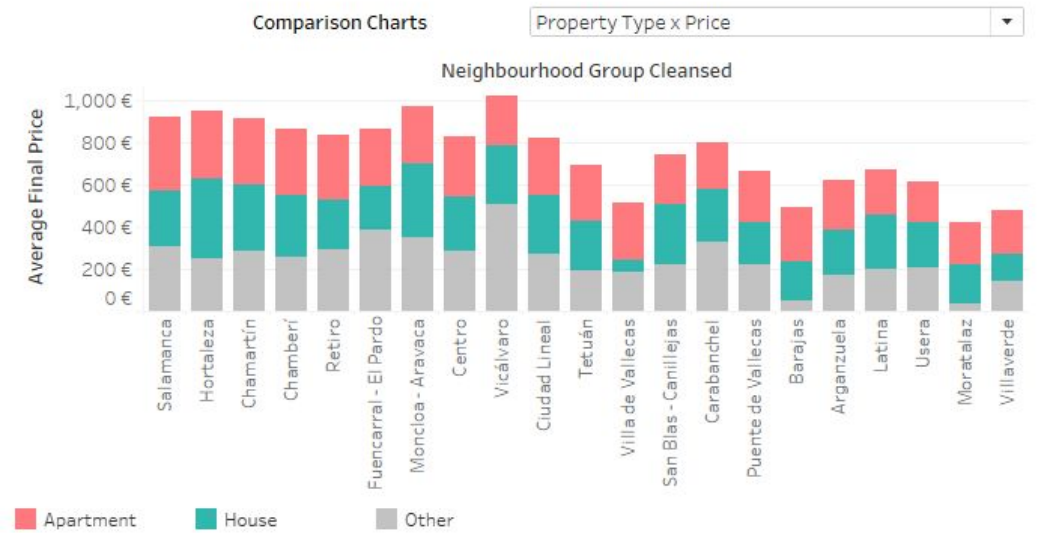
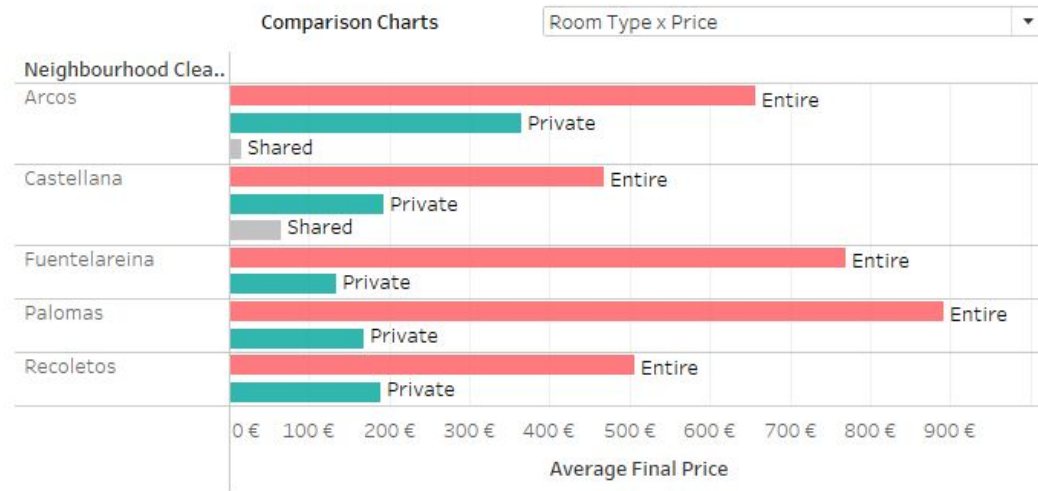


Comparaciones

con Final Price :
Room Type,
Accommodates,
Number of Reviews,
Property Type

Visualización: Exploración y visualización

Final Price Mean	Total Listing	Total SuperHost	Review Scores	Number of Reviews
278,96 €	13,247	1,556	92	306,099



Modelado: Análisis de correlación



Host Response Rate	1	0.062	-0.037	0.012	0.035	-0.008	-0.027	0.023	0.066	0.098	0.046
Accommodates	0.062	1	0.33	0.68	0.82	0.58	0.17	0.38	0.58	-0.061	0.31
Bathrooms	-0.037	0.33	1	0.42	0.39	0.35	0.19	0.32	0.19	0.01	0.1
Bedrooms	0.012	0.68	0.42	1	0.69	0.53	0.21	0.38	0.44	0.01	0.16
Beds	0.035	0.82	0.39	0.69	1	0.49	0.17	0.35	0.46	-0.053	0.2
Price	-0.008	0.58	0.35	0.53	0.49	1	0.36	0.66	0.37	0.056	0.34
Security Deposit	-0.027	0.17	0.19	0.21	0.17	0.36	1	0.48	0.075	-0.071	0.11
Cleaning Fee	0.023	0.38	0.32	0.38	0.35	0.66	0.48	1	0.2	0.0048	0.06
Guests Included	0.066	0.58	0.19	0.44	0.46	0.37	0.075	0.2	1	0.0087	0.23
Review Scores Rating	0.098	-0.061	0.01	0.01	-0.053	0.056	0.071	0.0048	0.0087	1	0.15
Amenities Score	0.046	0.31	0.1	0.16	0.2	0.34	0.11	0.26	0.23	0.15	1

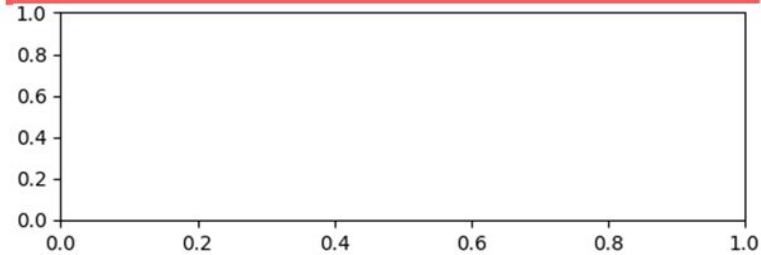
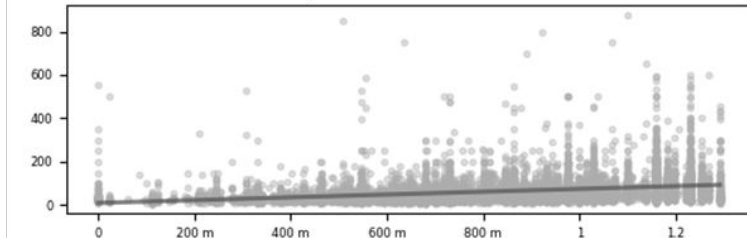
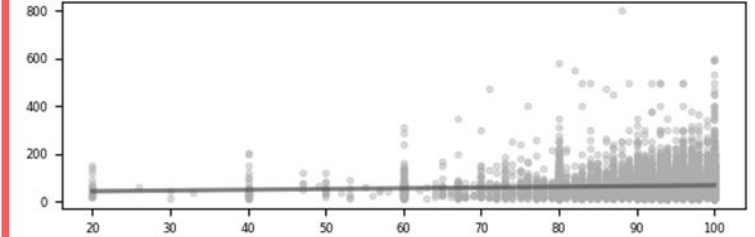
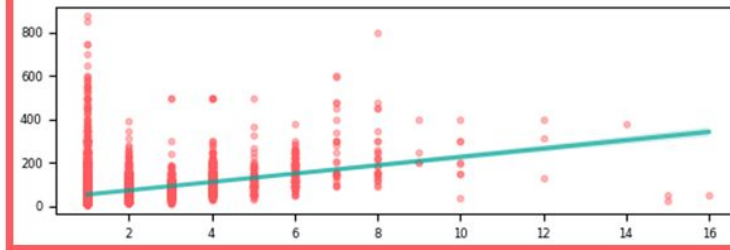
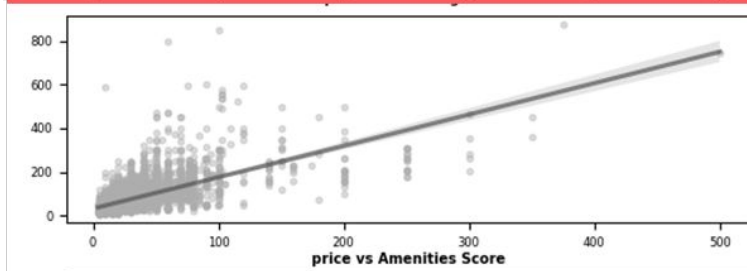
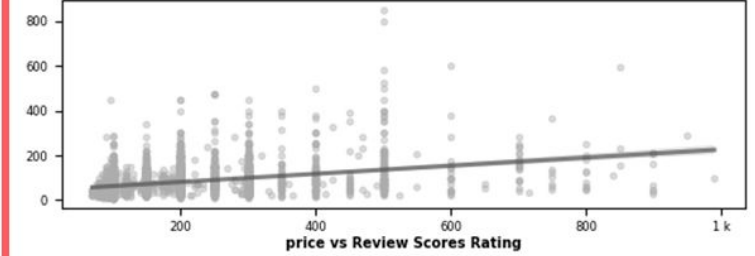
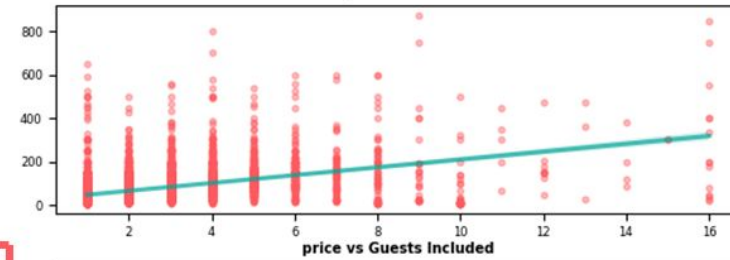
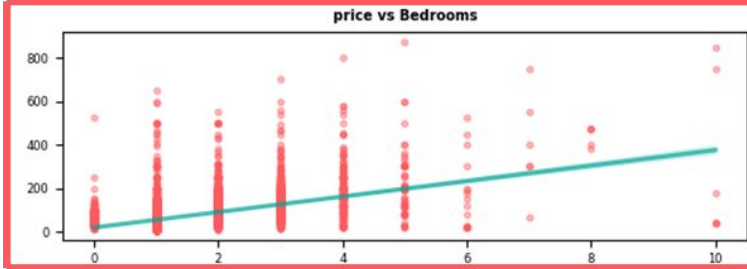
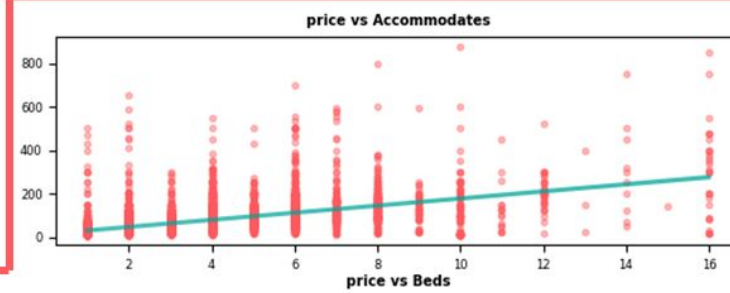
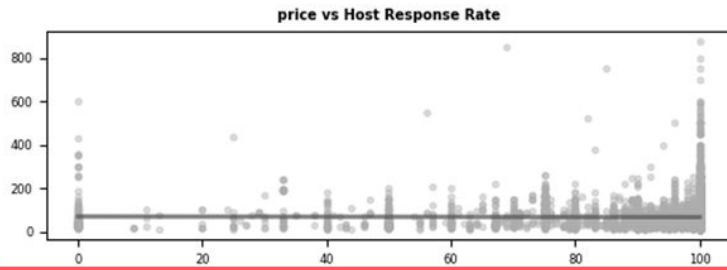
La correlación que demuestra la columna Price con las demás no revela un dato que sobresalga o sea más significativo que otros para poder predecirse.

Por el contrario de podríamos pensar que son varios factores en conjunto que afectan al precio final

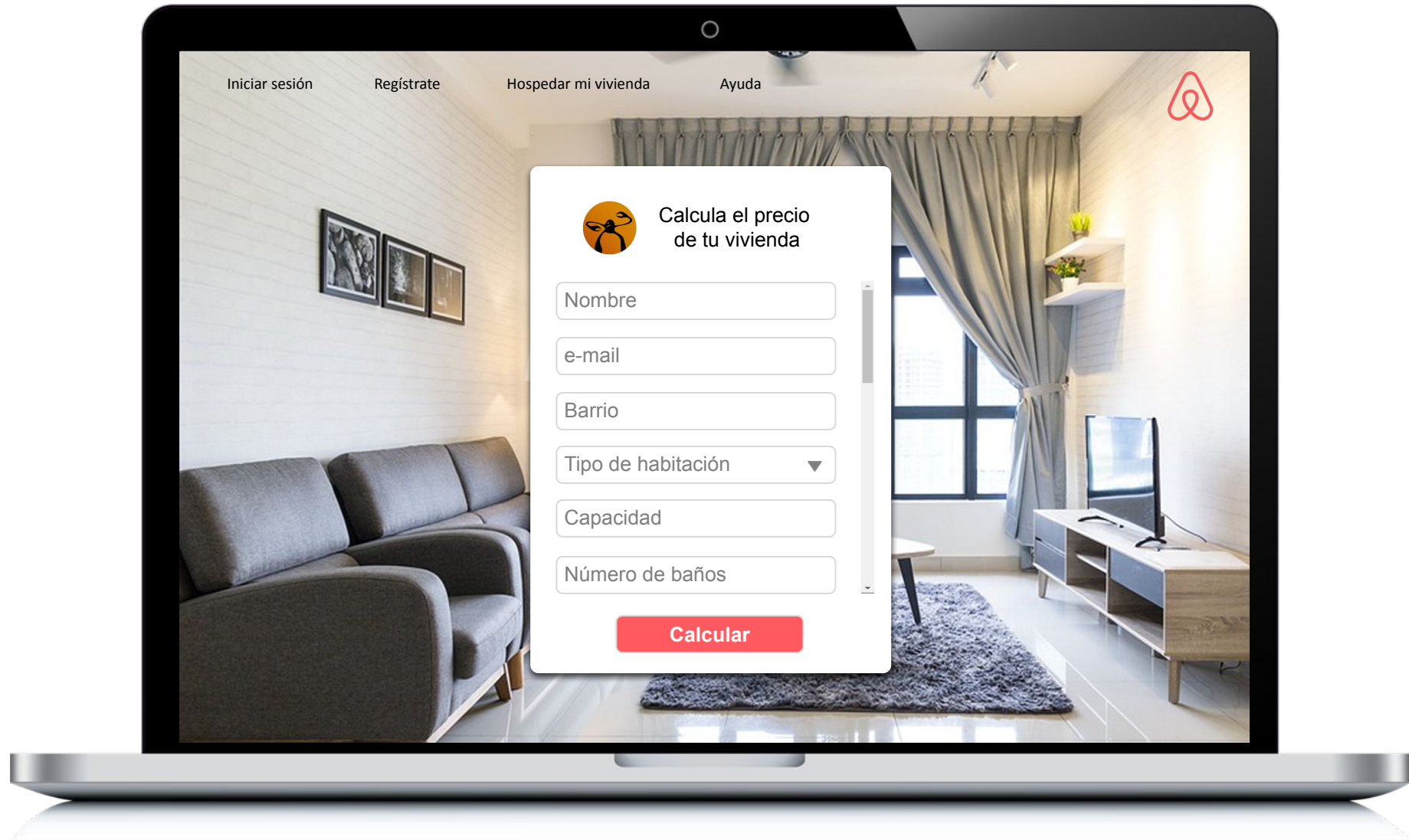
Modelado: Análisis de correlación



Correlation with price



Entregable: Simulador

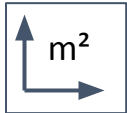




Hacia donde planeamos evolucionar el sistema:



Tener en cuenta la estacionalidad según la ciudad.



Considerar también incluir el dato de m2 en el modelo



Series de datos en el tiempo



Segundo release del producto: considerar asistir en la selección de la propiedad a quien quiera comprar para luego alquilar



No asumir. La búsqueda de información no siempre será tan amigable como el ideal.



Aumentar el nivel de análisis sobre todos los campos.



Hacer una prueba piloto con un número reducido de datos del csv.



Poder utilizar dataset externos que nos permitieran revalidar la información, ejemplo para zipcodes

Gracias

