



airbnb Madrid

Madrid

INTRODUCCIÓN

- Objetivo: Elaborar un modelo de clasificación en base a una tabla de datos de Airbnb extraídos de la web **InsideAirbnb**.
- Las clases para la clasificación se extraerán de un modelo de clustering hecho previamente
- Tabla con **mucha información** (21000 filas y 75 columnas)
↳ *Importante tarea de limpieza*
- Al contrario de lo que pensaba, el precio no ha sido crucial a la hora de definir los clusters. Han tomado más en cuenta características como el número de habitaciones o número de reviews



¿CON QUÉ VALORES NOS QUEDAMOS?

	neighbourhood_group_cleansed	room_type	bathrooms	bedrooms	beds	price	number_of_reviews	availability_365
0	Chamartín	Private room	1.0	1.0	1.0	\$90.00	102	168
1	Latina	Private room	1.0	1.0	1.0	\$31.00	33	226
2	Centro	Entire home/apt	1.0	NaN	1.0	\$92.00	26	126
3	Arganzuela	Private room	1.0	1.0	1.0	\$29.00	185	344
4	Centro	Entire home/apt	1.0	1.0	2.0	\$150.00	172	15

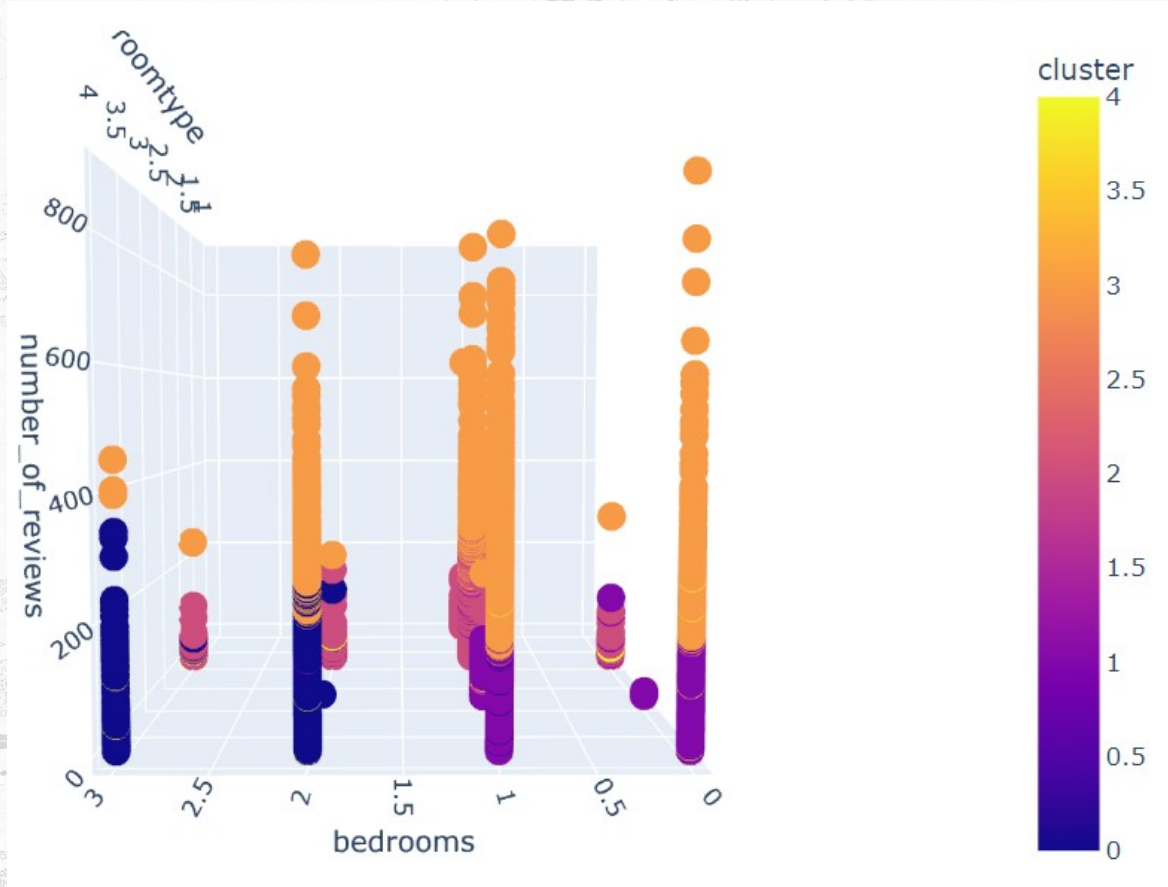
LIMPIEZA DE DATOS

- Pasamos de **72 columnas** a solo **10 columnas**. El **13%** son útiles
- A los valores nulos (*NaN*) fruto de algún fallo en el scraping se les da valor de 0.
- Se extraen columnas numéricas de atributos:
 - 'Pool', que determina si la casa tiene o no piscina.
 - La **mediana** del precio del distrito al que pertenezca cada registro, para comparar.
 - Tipo de alquiler, de más a menos deseable en orden numérico:
 - Casa completa → 4
 - Hab. Hotel → 3
 - Hab. en casa compartida → 2
 - Hab. compartida → 1



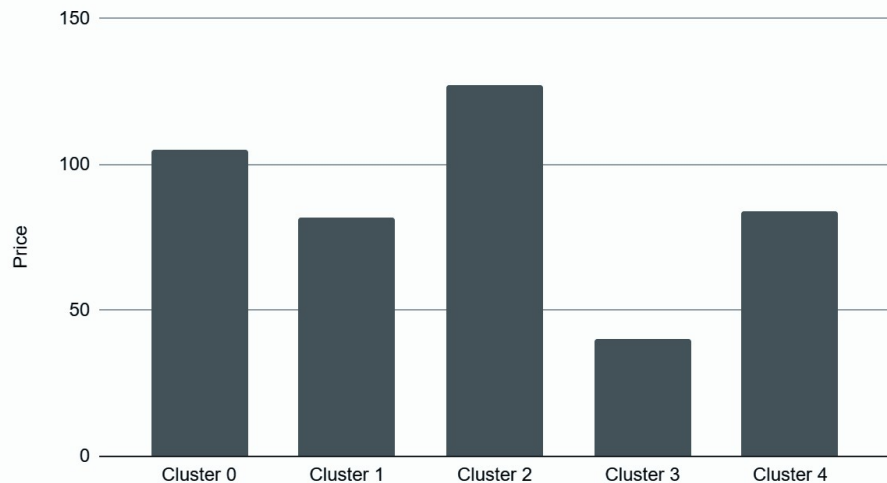
MODELO DE CLUSTERING

- Número de clusters: 5

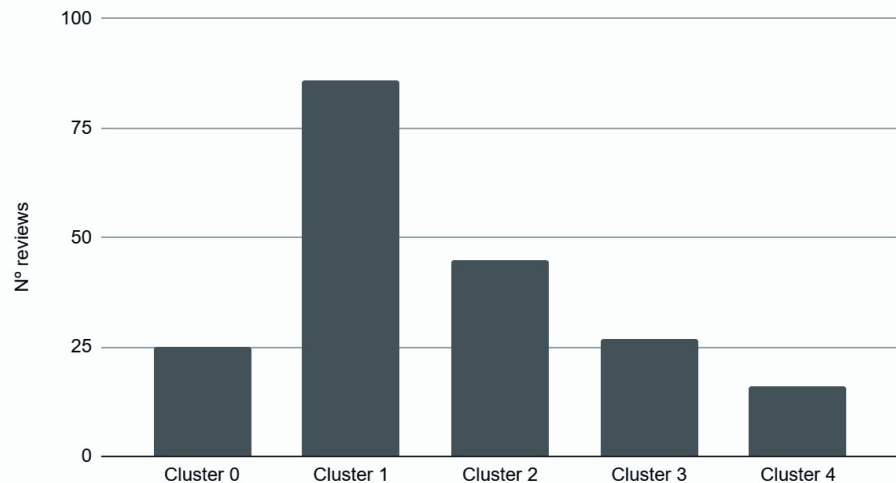


MODELO DE CLUSTERING

Clusters según precio



Clusters según nº reviews



MODELO DE CLUSTERING (tras asignarle nombres)

Clusters	Price	Nº reviews	Bedroom	Bathroom	Availability
Pisos de 1º Clase	127\$	45	2	1,37	165
Pisos de 2º Clase	105\$	25	0,85	1	287
Pisos Medios (- reviews)	84\$	16	1,2	1	158
Pisos Medios (+ reviews)	82\$	86	0,8	1	69
Pisos Baratos	40\$	27	1	0,1	140

IMPLICACIONES PRÁCTICAS



Segmentación de clientes:

El análisis de clustering ha identificado cinco grupos de clientes con características similares.

Personalización de ofertas:

Con la segmentación de clientes, es posible personalizar las ofertas y promociones para cada grupo, satisfaciendo las demandas específicas de cada segmento.

Estrategias de precios diferenciados:

Los diferentes clusters revelan relaciones variables entre el número de reseñas y el precio, lo que sugiere ajustar los precios para atraer más clientes y fomentar la generación de reseñas según el cluster.





Se han encontrado varios desafíos:

- **Dificultad para construir un modelo de regresión fiable del precio**, debida a diferentes factores y características:
 - **Características no contempladas en las tablas:**
La decoración, el mobiliario, etc... pueden influir en la variabilidad de los precios y no han sido considerados en nuestros datos originales, lo que dificulta la construcción de un modelo de regresión fiable.
 - **Limitaciones temporales del conjunto de datos** dado que nuestro conjunto de datos solo contiene información para la fecha actual y no se han podido comparar los precios con datos históricos.

A pesar de estos desafíos:

- Se logra identificar **5 clusters distintos** que pueden ayudar a plantear estrategias de negocio diferenciadas, lo cual podría incrementar beneficios y abordar problemas específicos.
- Se desarrolla un **modelo de clasificación** con una precisión del **90%** lo que nos permite clasificar futuros registros, de gran utilidad para tomar decisiones informadas con próximas actualizaciones de los datos.

