

Análisis de Airbnb Listings en Madrid



Proyecto Final
Febrero 2023



KEEPCODING
Tech School



Integrantes:

- Giorgia Scuderi
- Griselda Vilar
- Katerina Margariti
- Yesenia Andraca



Fase 1. Definición del Dataset y validación de datos

- Comprobación y exploración de datos de dataset
 - Columnas
 - Tipo
- Primera limpieza
 - Definición de columnas a analizar



```
#En la columna "City" se seleccionan los datos cuyo valor sea igual a "Madrid"  
df_madrid = df[(df['City'] == 'Madrid')]
```

Fase 1.

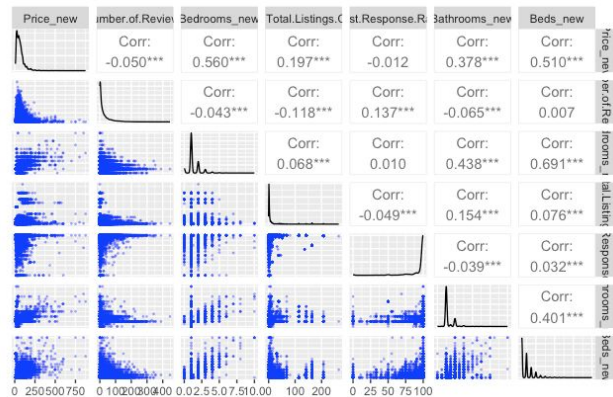
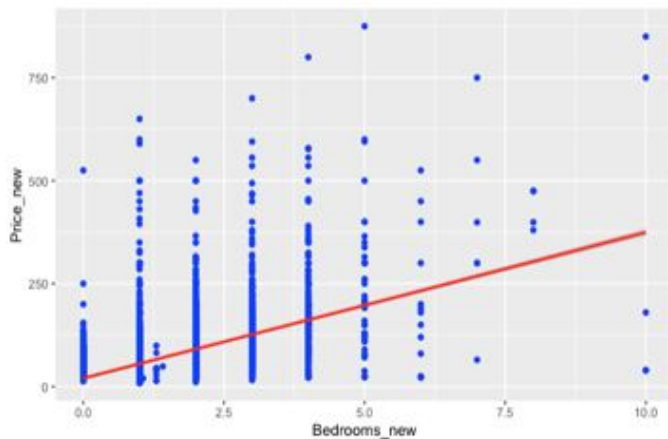
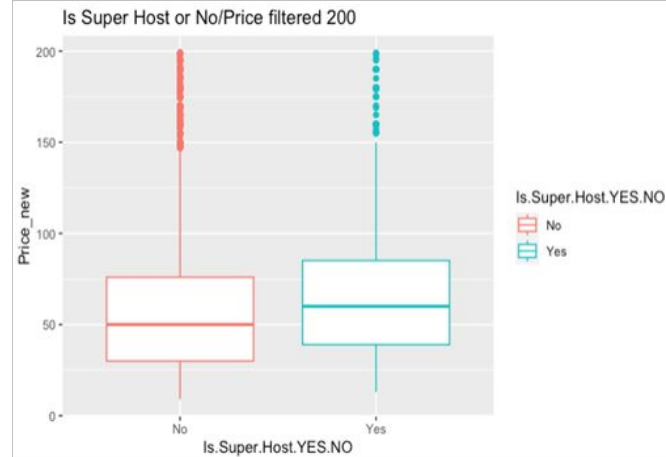
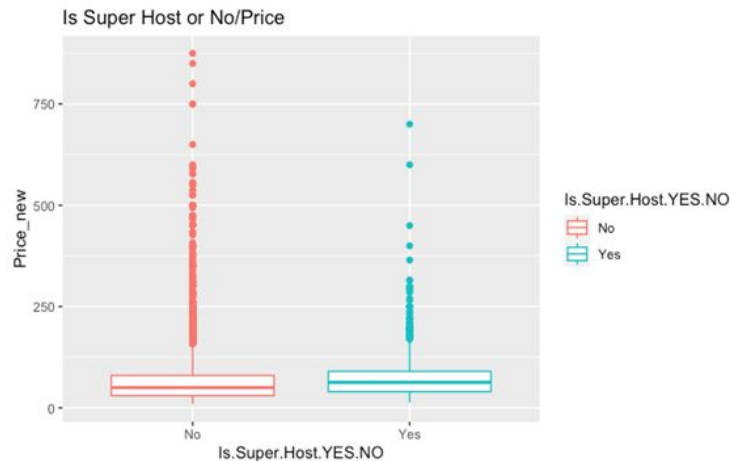
- Comprobación de integridad de los datos existentes
- Normalización de valores
- Obtención de medias para creación de nuevas columnas
- Creación de columnas SUPERHOST

Fase 2. Análisis Exploratorio y modelado

- Análisis exploratorio y calidad del dato en R
- Análisis exploratorio y visualización de datos
- Matrices de correlación
- Regresiones lineales
- Regresiones lineales múltiples



Fase 2.



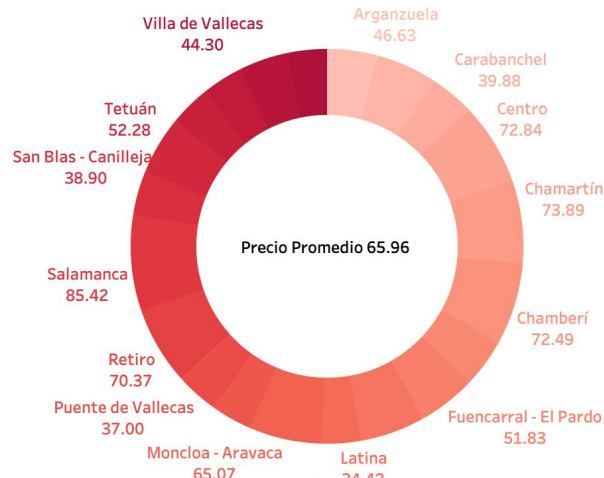
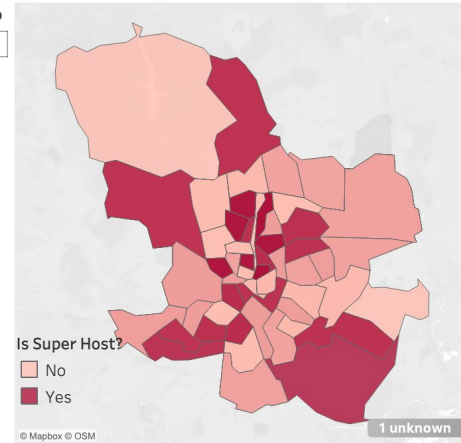
Fase 3. Visualización de las métricas

ANÁLISIS DE LA BASE DE DATOS AIRBNB EN MADRID

Alojamientos	Sum Reseñas
12.768	300.508
Precio Promedio Diario	Super Hosts
65.96	No 11.245
	Yes 1.523

Top N: Barrios con Host/SuperHost y el Promedio de Precio

5		
Sol	91.15	84.89
Cortes	92.03	81.37
Palacio	88.80	75.84
Universidad	75.74	66.16
Embajadores	71.59	58.92



Barrios con el % de listings

Centro	51,69%
Chamberí	7,21%
Salamanca	6,45%
Arganzuela	5,97%
Tetuán	3,39%
Moncloa - Aravaca	3,18%
Retiro	3,17%
Latina	2,86%
Chamartín	2,68%
Carabanchel	2,68%
Ciudad Lineal	2,28%
Puente de Vallecas	
Hortaleza	1,42%
Usera	1,09%
Fuencarral - El Pardo	1,09%
San Blas - Canillejas	0,88%
Villaverde	0,61%
Barajas	0,61%
Moratalaz	0,57%
Villa de Vallecas	0,34%



Conclusiones y “lessons learned”

- **¿Qué nos ha aportado el desarrollar este proyecto?**

Aplicar los conceptos teóricos y prácticos aprendidos durante el curso a un contexto de datos real.

- **¿Qué aprendimos?**

La importancia de la evaluación cualitativa y cuantitativa de los datos.

La imputación de los nulos: para mejorar la calidad de las estimaciones y obtener resultados mejorados.

La exploración y visualización de datos en R Python y Tableau para ver en detalle los indicadores y patrones de los datos.

Conclusiones y “lessons learned”

- **¿Qué es lo que no volveríamos a hacer de la misma manera?**

Intentar otras metodologías para rellenar los valores NaN.

Realizar más pruebas de análisis y consultar otras fuentes para enriquecer los datos.

- **¿Qué cosas seguimos haciendo para mejorar el proyecto?**

Realizar más pruebas de cálculos y visualización de datos con una distribución más uniforme.

Comparar con una base de datos más amplia y de otras zonas y ciudades para comprobar patrones o tendencias en los datos.



Gracias