Análisis de Datos Abiertos de Airbnb en Madrid

# Descripción del proyecto

Este proyecto forma parte del curso de Data Analytics y tiene como objetivo analizar los datos abiertos de Airbnb en la ciudad de Madrid. A lo largo del proyecto se realiza una transformación y limpieza de datos, análisis descriptivo, creación de un dashboard interactivo y elaboración de un informe explicativo con las principales conclusiones.

* En qué consiste mi proyecto.
* Objetivo
* Contexto del análisis.
* Qué problema se está resolviendo

# Estructura del Proyecto

📂 Proyecto-Airbnb-Madrid  
├── data/  
│ └── listings.csv # Datos brutos descargados de Inside Airbnb  
├── cleaned\_data/  
│ └── listings\_cleaned.xlsx # Datos transformados y limpios  
├── analysis/  
│ └── analisis\_descriptivo.xlsx # Tablas y gráficos con estadísticas descriptivas  
├── dashboard/  
│ └── dashboard\_airbnb.xlsx # Dashboard interactivo en Excel  
├── informe/  
│ └── informe\_airbnb.pdf # Informe final del análisis  
└── README.md

# Fuente de Datos y herramientas utilizadas

## Fuente de datos

Los datos se han obtenido del portal Inside Airbnb (http://insideairbnb.com/get-the-data/), en su versión actualizada para Madrid.

En concreto, se han utilizado dos archivos llamados “listings.csv” donde se recoge información sobre los distintos anuncios y los alojamientos y sobre los *hosts* de cada anuncio.

## Herramientas utilizadas

* Excel: limpieza, análisis, gráficos y dashboard
* Inside Airbnb: fuente de datos
* Canva / Word: elaboración del informe final (PDF)

# Desarrollo del Proyecto

## Comprensión general de los datos

Se ha hecho el análisis de dos archivos llamados “listings.csv” obtenidos de la web de datos abiertos de Airbnb. Finalmente, solo se utilizó uno de los dos archivos, ya que uno contenía la misma información ampliada que otro.

Se trata de un archivo de 79 columnas y 25.289 filas con información sobre los anfitriones (hosts) y sus anuncios de alojamientos en la ciudad de Madrid. Por lo tanto, se va a proceder a hacer 2 análisis: uno sobre los hosts que tiene alojamientos en la Comunidad de Madrid y otro sobre los propios alojamientos. Se hará un cruce de datos para entender mejor el perfil del host.

## Transformación y Limpieza

### Eliminación de columnas irrelevantes

El archivo cuenta con 79 columnas, de las cuales se van a emplear X. El resto han sido descartadas ya que no aportan valor real al análisis, como las columnas relativas al propio anuncio en la web (*listing\_url*, *name*, *description*, *picture\_url*, *host\_url*, *host\_name*, *host\_thumbnail\_url*, *host\_picture\_url*, *host\_verifications*, *host\_identity\_verified*, etc.).

Finalmente, tras el análisis inicial de los datos, la tabla con los datos a analizar consta de X columnas.

Eliminación de valores nulos y registros duplicados

1. No hay valores duplicados por anuncio. Por ID de Host, en cambio, sí, ya que un mismo host puede tener varios anuncios.
2. Filtro por *host\_since*: existen valores nulos que se van a eliminar, ya que no suponen una muestra de datos representativa frente al total (19 frente a 25.289). Además, se observa que no disponen de información sobre el *host*, solo sobre el alojamiento.
3. Filtro por *host\_location*: unificar opciones. Solo coger el valor del país, hay gran variedad, incluso diferenciados por estados para el caso de Estados Unidos.

Filtros por disponibilidad y tipo de alojamiento

## Análisis Descriptivo

- Distribución de precios por barrio y tipo de alojamiento  
- Número de anuncios por anfitrión  
- Disponibilidad media por zona  
- Relación entre puntuaciones y precios

## Dashboard Interactivo

- Mapa de calor por barrios  
- Filtros por tipo de alojamiento y rango de precios  
- Evolución mensual de disponibilidad  
- Comparativa de precios promedio

## Informe Final

- Metodología de limpieza y análisis  
- Gráficos principales con explicación  
- Conclusiones clave y posibles usos del análisis  
- Recomendaciones para anfitriones, turistas y reguladores

# Resultados y conclusiones

# Próximos pasos

# Autora

Este proyecto ha sido desarrollado por Laura Pomares Bleda.

Fecha de entrega: Abril 2025.