

Proyecto 3: Analisis de Datos de Airbnb en una Ciudad

Nota: Recuerda que este proyecto es un ejercicio académico, ante la duda pregunta a tus profesores.

1. Descripción.

En el contexto actual del turismo digital, las plataformas de alquiler vacacional juegan un papel clave en la experiencia de viaje de millones de personas. Entender el comportamiento de los alojamientos, los precios, la disponibilidad y las valoraciones de los usuarios es fundamental para mejorar la experiencia tanto de huéspedes como de anfitriones.

Con el objetivo de analizar el comportamiento de los alojamientos de Airbnb en una ciudad concreta, se nos ha encargado el desarrollo de un proyecto de análisis de datos que permita identificar patrones, tendencias y oportunidades de mejora. Este análisis ayudará a comprender mejor cómo se distribuyen los alojamientos, qué factores influyen en el precio, qué características están asociadas a mejores valoraciones y cómo se comporta la oferta en distintos barrios o zonas de la ciudad.

A lo largo del proyecto, realizaremos un análisis exploratorio de datos, transformaciones, visualizaciones y el diseño de una base de datos que permita estructurar la información de forma eficiente. Finalmente, presentaremos conclusiones basadas en datos que puedan servir como apoyo a la toma de decisiones estratégicas.

Los datos que se utilizarán en este proyecto provienen de <https://insideairbnb.com/get-the-data>, una fuente que recopila datos públicos de Airbnb y los pone a disposición para su análisis. El conjunto de datos incluye información sobre los alojamientos, como su ubicación, precio, disponibilidad, número de reseñas, entre otros.

¿Qué es Airbnb?

Airbnb es una plataforma digital fundada en 2008 que permite a particulares ofrecer alojamientos turísticos —como habitaciones, apartamentos o viviendas completas— a viajeros de todo el mundo. La plataforma se ha convertido en un actor clave dentro del sector turístico, generando grandes volúmenes de datos relacionados con precios, localización, disponibilidad, valoraciones y comportamiento de los usuarios.

El análisis de estos datos resulta especialmente relevante para:

- Entender dinámicas del mercado turístico urbano.
- Detectar patrones de precios y demanda.
- Analizar la distribución geográfica de los alojamientos.

Fases del proyecto:

Fase 1: Análisis Exploratorio de Datos(EDA).

Antes de comenzar con cualquier transformación o modelado, es fundamental comprender el conjunto de datos y sus características. En esta fase deberás realizar un análisis exploratorio detallado para familiarizarte con los datos y responder preguntas como:

- ¿Cuántos alojamientos únicos contiene el dataset?
- ¿Qué tipos de alojamiento existen y cuál es el más común?
- ¿Cómo se distribuyen los alojamientos por barrio o zona?
- ¿Existen barrios con una concentración especialmente alta de alojamientos?
- ¿Qué variables presentan valores nulos y en qué proporción?
- ¿Hay columnas cuyo contenido no aporta valor al análisis y podrían eliminarse?
- ¿Existen valores atípicos en el precio de los alojamientos? ¿Cómo afectan a la media?

El objetivo de esta fase es entender el contexto general de los datos y detectar posibles problemas de calidad.

Fase 2: Transformación de los datos.

En esta fase se llevarán a cabo las transformaciones necesarias para garantizar la calidad y coherencia del conjunto de datos. Las transformaciones se realizarán mediante funciones de Python aplicadas al DataFrame.

Algunas de las transformaciones que podrías realizar son:

- Conversión de columnas numéricas que estén almacenadas como texto (price, reviews_per_month, etc.).
- Limpieza de símbolos o caracteres especiales en columnas como el precio.
- Tratamiento de valores nulos y evaluación de si deben eliminarse, imputarse o mantenerse.
- Detección y análisis de valores duplicados.
- Normalización o estandarización de variables si es necesario.
- Corrección de valores inconsistentes o errores tipográficos en columnas categóricas.
- Eliminación de columnas irrelevantes para el análisis.

Fase 3: Análisis y Visualización.

El objetivo de esta fase es generar un informe visual que permita comprender mejor el comportamiento de los alojamientos de Airbnb en la ciudad analizada.

3.1 💰 Análisis de precios

1. ¿Cuál es el precio medio y mediano de los alojamientos?
2. ¿Cómo varía el precio según el tipo de alojamiento?
3. ¿Existen diferencias significativas de precio entre barrios?
4. ¿Qué barrios presentan los precios más altos y cuáles los más bajos?
5. ¿Qué tipo de alojamiento ofrece la mejor relación entre precio y valoración?
6. ¿Existen alojamientos con precios extremadamente altos o bajos? ¿A qué podrían deberse?

3.2 ★ Valoraciones y reseñas

1. ¿Cuál es la valoración media de los alojamientos?
2. ¿Existen diferencias en las valoraciones según el tipo de alojamiento?
3. ¿Los alojamientos con más reseñas tienen mejores valoraciones?
4. ¿Qué barrios concentran los alojamientos mejor valorados?
5. ¿Se observa alguna relación entre precio y valoración?

3.3 17 Disponibilidad y comportamiento de la oferta

1. ¿Cómo se distribuye la disponibilidad anual de los alojamientos?
2. ¿Existen barrios con mayor disponibilidad media?
3. ¿Qué tipo de alojamiento tiende a tener mayor disponibilidad?
4. ¿Puede la disponibilidad estar relacionada con el precio o la demanda?

Importante: Las visualizaciones deberán ir acompañadas de análisis descriptivos claros.

Fase 4: Diseño de bases de datos e inserción de los datos. (Bonus)

En esta fase el objetivo es diseñar la estructura de una base de datos que permita almacenar la información de los alojamientos de manera organizada y escalable.

Los aspectos principales de esta fase son:

1. Diseño de la estructura de la base de datos:
 - Identificar las tablas necesarias (alojamientos, anfitriones, localización, etc.).
 - Definir claves primarias y foráneas.
 - Establecer relaciones entre las tablas.
2. Creación de la base de datos:
 - Implementar la base de datos utilizando SQL y las herramientas vistas en el módulo.
3. Inserción de datos:
 - Insertar los datos transformados en la base de datos diseñada.

Los Datos.

Los datos que se utilizarán en este proyecto provienen de <https://insideairbnb.com/get-the-data>, una fuente que recopila datos públicos de Airbnb y los pone a disposición para su análisis. El conjunto de datos incluye información sobre los alojamientos, como su ubicación, precio, disponibilidad, número de reseñas, entre otros. La ciudad concretamente a analizar será proporcionada por tu profesora.

Criterios de aceptación.

- Crear la infraestructura necesaria: repositorio en GitHub y con acceso para todos los miembros del equipo.
- Extraer datos de distintas fuentes de datos y creación de una Base de Datos.
- Tener la extracción de datos, creación de la base de datos e inserción de datos automatizada en funciones y clases.
- Tener en el repositorio de GitHub todo el código del desarrollo del proyecto.

Presentación y Entrega

Debe presentar los resultados del análisis de datos de manera clara y efectiva, para que las partes interesadas puedan entender y utilizar la información para la toma de decisiones.

- Preparar un informe que incluya una visión general del proyecto, metodologías utilizadas, y resultados clave.
- El formato de entrega de este proyecto será mediante la carga del mismo a la plataforma de GitHub. Se creará un repositorio.