RESUMEN EJECUTIVO

# INTRODUCCIÓN

El proyecto surge de la necesidad de comprender el mercado de alojamientos vacacionales, específicamente a través de la plataforma Airbnb y los datos proporcionados por la empresa White Hosting. Se busca proporcionar información estratégica a inversores y actores del sector para tomar decisiones informadas.

# ANÁLISIS DATASET

Lectura de los Datos

Se realizó la lectura de los datos provenientes del dataset proporcionado por White Hosting y para el estudio del mercado de alojamientos vacacionales. Contiene datos hasta el año 2021.

EDA: Exploratory Data Analysis

Se llevó a cabo un análisis exploratorio de datos para entender la naturaleza y características de los datos. Esto incluyó la visualización de variables, distribución de datos, identificación de tendencias y patrones iniciales.

Comprensión de la Estructura de los Datos:

Se profundizó en la estructura de los datos para comprender la relación entre las diferentes variables, identificar variables clave y entender la lógica subyacente de los datos.

Búsqueda de Valores Nulos:

Se realizó una búsqueda exhaustiva de valores nulos o faltantes en el conjunto de datos para determinar su impacto en el análisis y tomar decisiones sobre cómo manejar estos valores ausentes.

Revisión y Eliminación de Duplicados:

Se revisaron los datos en busca de registros duplicados y se tomaron medidas para eliminarlos, asegurando la integridad y precisión de los datos.

Análisis Estadístico de las Variables:

Se aplicaron técnicas estadísticas para analizar la distribución de las variables, calcular medidas de tendencia central y dispersión, y comprender la variabilidad de los datos en el conjunto.

Valores Atípicos:

Se identificaron y analizaron valores atípicos o outliers que podrían afectar el análisis, evaluando su impacto y considerando posibles acciones para manejarlos.

Correlaciones:

Se exploraron las relaciones entre las variables mediante el cálculo de correlaciones, lo que permitió identificar posibles asociaciones o dependencias entre ellas.

Variables Creadas:

Se crearon nuevas variables o características a partir de las existentes, utilizando transformaciones o combinaciones de datos para generar información adicional relevante.

Rankings:

Se generaron rankings o clasificaciones de acuerdo a ciertos criterios como precio, ocupación, rentabilidad, etc., para identificar los inmuebles más relevantes o destacados en el mercado.

Hipótesis:

Se formularon hipótesis sobre el comportamiento de ciertas variables o relaciones entre ellas, las cuales fueron evaluadas y validadas durante el análisis de datos.

Volatilidad:

Se examinó la volatilidad de los precios u otras variables a lo largo del tiempo, identificando períodos de fluctuaciones significativas y posibles factores causales.

Estacionalidad:

Se investigó la presencia de patrones estacionales en los datos, especialmente en variables como precios, ocupación, rentabilidad, etc., para comprender mejor la dinámica del mercado en diferentes momentos del año.

# WEB SCRAPING

Se realizó el proceso de web scraping para obtener información relevante de páginas web específicas, en este caso, relacionadas con el mercado de alojamientos vacacionales.

Importación de Librerías

Se importaron las librerías necesarias en el entorno de desarrollo para ejecutar el proceso de web scraping de manera efectiva, incluyendo herramientas como BeautifulSoup y Selenium.

Configuración de Opciones de Chrome y Creación de WebDriver

Se configuraron las opciones de Chrome y se creó un WebDriver para permitir la interacción automatizada con el navegador, facilitando la extracción de datos de forma programática.

* Definir la URL: (ejemplo Madrid):

Se definió la URL de la página web a escrapear, especificando detalles como la ciudad de interés (en este caso, Madrid) y otros parámetros relevantes para la búsqueda de datos.

* Conexión:

Se estableció la conexión con la URL definida para iniciar el proceso de web scraping y acceder al contenido de la página web.

* Obtener el HTML de la Página:

Se obtuvo el HTML completo de la página web, lo que permitió acceder a la estructura y contenido de la página para extraer la información requerida.

* Parsear el HTML:

Se utilizó BeautifulSoup u otras herramientas de análisis HTML para parsear el código HTML y extraer los elementos específicos que contenían los datos de interés.

* Imprimir Información sobre la Respuesta HTTP:

Se imprimió información relevante sobre la respuesta HTTP obtenida al acceder a la página web, incluyendo detalles como el estado de la respuesta, códigos de error, etc.

Adquisición de HTML con Espera Explícita:

Se implementó un mecanismo de espera explícita para la adquisición del HTML, asegurando que se obtuviera el contenido completo de la página web antes de proceder con la extracción de datos.

Extracción de Elementos de Navegación por Paginación y Código Fuente:

Se extrajeron elementos de navegación como botones de paginación para acceder a múltiples páginas de resultados y se analizó el código fuente para identificar la estructura de la información a extraer.

Crear DataFrame y Guardar en CSV en Formato Texto:

Se creó un DataFrame con los datos extraídos y se guardó en formato CSV para su posterior análisis y manipulación, permitiendo trabajar con los datos obtenidos de manera organizada y estructurada.

# PREGUNTAS CLIENTE

## ¿Cuántos inmuebles únicos hay en el dataset vs AirBnb?

* Inmuebles únicos en el df original White Hosting: 8796
* Inmuebles únicos en el df AirBnb 2023: 66181
* Inmuebles únicos en el df AirBnb 2024: 2837

TOTAL INMUEBLES ÚNICOS ENTRE LOS 4: 77.814

## ¿Cuál es la diferencia en número de inmuebles posteados entre las dos fuentes de datos?

La diferencia entre el número de inmuebles únicos posteados entre el dataset original y los extraídos de AirBnb es de 60.222

## ¿Cuáles son los parámetros que conforman el precio?¿Son iguales los parámetros en ambas fuentes de datos?

Para comprobar qué parámetros conforman el precio se ha llevado a cabo el cálculo de la correlación existente entre el precio y las demás variables, obteniendo como resultado lo siguiente:

* DATASET ORIGINAL: En el dataset original las variables con un coeficiente de correlación mayor con el precio son, en orden de mayor a menor:

‘acommodates’, ‘bedrooms’, ‘bathrooms’, ‘bed’.

* AIRBNB: En los dataset de AirBnb las variables con un coeficiente de correlación mayor con el precio son, en orden de mayor a menor:

‘acommodates’, ‘bedrooms’, ‘bathrooms’, ‘bed’.

Por lo tanto, en ambas fuentes de datos, aun dando cifras diferentes, debido, seguramente, a diferencias en los datos recopilados, los parámetros que conforman el precio son los mismos.

En el primer análisis de la correlación se observó que en las valoraciones de los huéspedes no tenían correlación con el precio.

## ¿Has podido ver algún inmueble que esté en ambas fuentes de datos?¿Tienen la misma valoración?

* **Presencia de Inmuebles Comunes**: Se confirmó la existencia de inmuebles que aparecen en más de uno de los DataFrames analizados. Esto indica que algunos inmuebles han sido listados en Airbnb en múltiples ocasiones a lo largo de los años considerados.
* **Desafíos y Limitaciones**: La investigación enfrentó desafíos inherentes a la calidad y consistencia de los datos, como la ausencia de la columna 'license' en algunos DataFrames, en 2024, y las variaciones en la cantidad de reseñas por año. Estos aspectos subrayan la complejidad de realizar análisis longitudinales en conjuntos de datos dinámicos como los de Airbnb.
* **Variabilidad en las Valoraciones**: El análisis detallado de los inmuebles en Airbnb revela que, si bien hay consistencia en las valoraciones para algunos apartamentos, existe una variabilidad significativa para otros. Esta dinámica resalta la naturaleza volátil del mercado de alquileres vacacionales, donde las percepciones y experiencias de los huéspedes son susceptibles a una amplia gama de factores. Estos pueden incluir desde mejoras en los inmuebles y cambios en la gestión, hasta variaciones en las tendencias del mercado y las expectativas de los clientes.

## 5. ¿En qué fuente de datos hay más volatilidad en cuanto al precio?¿Hay algún tipo de estacionalidad?

El análisis se basa en la medición de la volatilidad mediante la desviación estándar de los precios en diferentes ciudades para cada año. Este indicador proporciona una medida de la variabilidad de precios en relación con su media. La volatilidad se midió calculando la desviación estándar de los precios de los alojamientos en diferentes ciudades para cada año. Este índice proporciona una medida de la dispersión de precios en relación con su media.

* Volatilidad por Ciudad en 2021: Se observó una amplia variabilidad en la volatilidad entre las ciudades, con Menorca, Mallorca y Sevilla mostrando la mayor volatilidad, mientras que Valencia, Málaga y Madrid exhibieron una volatilidad más baja.
* Aumento General en 2023: La volatilidad experimentó un aumento notable en todas las ciudades en 2023 en comparación con 2021, lo que puede atribuirse a factores como la recuperación post-COVID-19 o impactos económicos globales.
* Estabilización en 2024: Aunque la volatilidad se mantuvo alta en ciudades como Barcelona y Madrid en 2024, se observó una tendencia hacia la estabilización en general.

La interpretación de la Volatilidad y Estacionalidad que obtuvimos fue:

* Factores Influenciadores: La estacionalidad turística, eventos locales y nacionales, y las fluctuaciones económicas pueden influir en la volatilidad de los precios.
* Consideración de Variables Complementarias: Es esencial examinar la volatilidad junto con otras variables como el precio promedio y la tasa de ocupación para obtener una comprensión más completa del mercado.
* Uso Estratégico del Análisis: La comprensión de la volatilidad puede beneficiar a los anfitriones de Airbnb para ajustar estratégicamente los precios y a los viajeros para planificar sus viajes de manera más informada.

Con las siguientes limitaciones y consideraciones adicionales

* Cobertura de Datos: El análisis está limitado a los datos disponibles para los años específicos y no incluye información para el año 2022, lo que puede afectar la precisión de la interpretación a largo plazo.
* Exclusión del DF Scraping: La ausencia de datos del DF Scrapping debido a la falta de información sobre la variable "Last\_review" limita el alcance del análisis y podría proporcionar insights adicionales sobre la volatilidad del mercado.

## 6. Crea un dataset único

Explicado en la memoria.

## 7. ¿Hay alguna variable que podríamos añadir para mejorar el análisis?

### VARIABLE OCUPACIÓN

VARIABLE RENTABILIDAD

VARIABLE ESTACIONALIDAD

\*Explicadas en la memoria.

## 8. ¿Puedes hacer un ranking de los inmuebles más caros?

Para responder a la pregunta de cuáles son los inmuebles más caros, se realizó un análisis en dos partes: primero, se creó un ranking de los alojamientos más caros en general, y segundo, se hizo un ranking por ciudad para identificar el alojamiento más caro en cada una.

Ranking General de Inmuebles Más Caros

Se ordenó el conjunto de datos de forma descendente por precio sin distinguir por ciudad. Esto nos permitió identificar los 10 alojamientos con el precio más alto en el conjunto de datos. Curiosamente, vemos que el precio máximo en esta muestra es de 500 euros, y estos alojamientos están distribuidos en diferentes ciudades y barrios, mostrando que hay una diversidad en la ubicación de los alojamientos más caros. La mayoría son del tipo "Entire home/apt", indicando que los alojamientos completos tienden a tener precios más altos.

Ranking de Inmuebles Más Caros por Ciudad

Después, se ordenó el conjunto de datos primero por ciudad y luego por precio de forma descendente, agrupando por ciudad para seleccionar el alojamiento más caro de cada una. Esto nos permitió ver cuál es el alojamiento más caro en cada ciudad, todos con precios cercanos o igual a 500 euros, excepto en Menorca, donde el precio más alto es ligeramente inferior. De nuevo, la mayoría son del tipo "Entire home/apt", reafirmando que este tipo de alojamiento suele ser el más costoso.

Análisis Gráfico de la Distribución de Precios

Por último, se creó un gráfico de caja (boxplot) para visualizar la distribución de precios de los alojamientos por ciudad. Este gráfico es útil para identificar no solo los alojamientos con precios máximos (como se hizo en los análisis previos), sino también para observar la mediana, los cuartiles y la presencia de valores atípicos en cada ciudad. La distribución de precios nos ayuda a entender en qué ciudades se concentra un mayor número de alojamientos caros y cómo varía el rango de precios entre diferentes ubicaciones.

## 9. ¿Cuáles son los inmuebles más rentables?

De nuevo, tenemos que tener en cuenta que el análisis de rentabilidad, se tuvo que realizar teniendo en cuenta sólo los ingresos por apartamento, pues no disponíamos de más información.

Para identificar los inmuebles más rentables, se llevó a cabo un análisis considerando tanto la ocupación como la rentabilidad de los alojamientos. Primero, se creó un ranking de los alojamientos más ocupados por ciudad y tipo de habitación, teniendo en cuenta la cantidad de días reservados en promedio y el precio. Este ranking muestra los alojamientos que han tenido una alta ocupación, lo que puede indicar su popularidad y demanda entre los usuarios.

Luego, se calculó la rentabilidad de los alojamientos multiplicando los días reservados por el precio. Posteriormente, se obtuvo un ranking de los alojamientos con la rentabilidad más alta, lo que indica aquellos inmuebles que han generado mayores ingresos.

Además, se crearon gráficos para visualizar mejor esta información. Por ejemplo, se mostró un gráfico de dispersión que representa la rentabilidad frente a la ubicación de los alojamientos más rentables, permitiendo ver claramente qué ciudades y barrios concentran los inmuebles más lucrativos. Asimismo, se elaboró un gráfico de barras que presenta la rentabilidad promedio de los alojamientos por ciudad, brindando una visión general de la rentabilidad en cada ubicación. Estos análisis y gráficos ofrecen una comprensión más completa de los inmuebles más rentables, tanto a nivel general como por ciudad y tipo de alojamiento.