NYC AirBnb Data

V&G Analytics

27/11/2024

Business Understanding

Background

Airbnb es una plataforma global que permite a las personas listar, descubrir y reservar alojamientos a corto plazo. La empresa ofrece una amplia gama de propiedades, desde habitaciones privadas hasta casas enteras, atendiendo a diversas necesidades de los clientes. Como parte de sus esfuerzos continuos para optimizar la experiencia del usuario y apoyar a los anfitriones en la maximización de sus ingresos, Airbnb busca obtener una comprensión más profunda de sus anuncios. Específicamente, la empresa está interesada en entender cómo diferentes factores influyen en el éxito de los anuncios en la ciudad de Nueva York, uno de sus mercados más grandes y competitivos.

El dataset proporcionado para este proyecto incluye una gran cantidad de información sobre los anuncios de Airbnb, como detalles de los hosts, características de las propiedades, precios, reseñas de clientes y disponibilidad. Estos datos forman la base para explorar aspectos clave de las operaciones de Airbnb, incluyendo estrategias de precios, tendencias de ocupación y la calidad general de los anuncios.

Objetivo

El objetivo principal de esta investigación es analizar los factores que influyen en los precios, la ocupación, las reseñas y el rendimiento de los anfitriones en los anuncios de Airbnb en Nueva York. El análisis buscará descubrir tendencias, correlaciones y patrones que puedan llevar a obtener información útil. Estos conocimientos ayudarán a Airbnb a formular recomendaciones basadas en datos para los anfitriones, mejorar las estrategias de precios y mejorar la experiencia general del cliente.

Al comprender cómo diversos elementos como el tipo de habitación, el vecindario, la identidad del anfitrión y los patrones de reseñas afectan el éxito de los anuncios, Airbnb puede refinar sus servicios, optimizar las ganancias de los anfitriones y ofrecer una experiencia más personalizada para los usuarios. Además, identificar los factores más importantes que impulsan las reservas y las reseñas puede ayudar a mejorar la oferta de la plataforma y a mantener a Airbnb competitivo en un mercado saturado.

Los interesados en este proyecto incluyen a la dirección de Airbnb, los hosts y los clientes.

- Stakeholders Airbnb: El equipo de liderazgo de Airbnb está interesado en comprender las tendencias del mercado más amplias e identificar áreas donde la empresa pueda mejorar sus servicios. Los conocimientos obtenidos de este análisis podrían informar decisiones estratégicas en torno a la fijación de precios, la promoción de listados y las características de la plataforma.
- Hosts: Los Hosts juegan un papel crítico en el éxito de Airbnb, y se beneficiarán de los conocimientos sobre cómo pueden optimizar sus anuncios. Ya sea ajustando los

- precios, actualizando las reglas de la casa o mejorando sus reseñas, el objetivo es ayudar a los anfitriones a maximizar la ocupación y las ganancias.
- Clientes: El objetivo final es mejorar la experiencia del cliente. Al comprender qué impulsa la satisfacción del cliente, Airbnb puede emparejar mejor a los usuarios con anuncios que cumplan con sus preferencias, asegurando una estancia más agradable.

Preguntas Clave de Negocios

- ¿Cómo afectan los factores de precios como el tipo de habitación, la ubicación y la identidad del anfitrión al precio de los anuncios?
- ¿Cuáles son los principales factores que impulsan la frecuencia de reservas y la disponibilidad?
- ¿Cómo se correlacionan las reseñas de los huéspedes, incluyendo las tasas de reseñas y la frecuencia de reseñas, con el desempeño del anfitrión y el éxito del anuncio?
- ¿Qué impacto tienen las políticas de cancelación, las noches mínimas y las reglas de la casa en las reservas y la satisfacción de los huéspedes?
- ¿Existen patrones geográficos o demográficos en las reseñas de los huéspedes y las tasas de ocupación en diferentes vecindarios de Nueva York?
- ¿Qué factores son los más influyentes en el éxito general de un anuncio, y cómo se pueden utilizar estas ideas para optimizar la plataforma de Airbnb?

Estas preguntas guiarán el análisis y asegurarán que el proyecto se mantenga enfocado en áreas que tienen un impacto comercial.

Criterios de Éxito

- Información Accionable: El análisis debe proporcionar información clara y accionable que pueda ayudar a Airbnb a optimizar los precios, las tasas de ocupación y el rendimiento de los anfitriones. Estos conocimientos deben ser relevantes y fáciles de implementar para los interesados.
- Precisión de los Insights: La precisión de los patrones y correlaciones descubiertos en el conjunto de datos debe ser validada a través de un análisis estadístico riguroso y técnicas de modelado.
- Impacto en el Negocio: En última instancia, los hallazgos deberían ser capaces de influir en las decisiones empresariales que mejoren el rendimiento de la plataforma de Airbnb. Esto podría incluir mejores algoritmos de precios, una mayor satisfacción del cliente y un aumento en las reservas para los anfitriones.

• Satisfacción de los Stakeholders: El éxito del proyecto también se medirá por qué tan bien los conocimientos satisfacen las necesidades de los interesados clave, incluyendo la gestión de Airbnb, los anfitriones y los clientes.

Constraints y Assumptions

- Calidad de los Datos: El conjunto de datos puede incluir información ausente, incoherente o parcial. Por ejemplo, algunas características como "fecha de la última reseña" o "reglas de la casa" podrían no estar presentes en algunos registro. Es esencial gestionar estos datos ausentes para la exactitud del análisis.
- Limitación Geográfica: El análisis se centra únicamente en los listados de la ciudad de Nueva York, lo que puede limitar la generalización de los hallazgos a otras ubicaciones.
- Suposición de Mercado Homogéneo: El análisis asume que el mercado de Airbnb en Nueva York es relativamente homogéneo en términos de comportamiento del cliente. Sin embargo, diferentes áreas de la ciudad (por ejemplo, Manhattan vs. Brooklyn) pueden tener características distintas que deben tenerse en cuenta.

Project Plan

Data Exploration: El primer paso será explorar a fondo el conjunto de datos para entender su estructura, identificar cualquier valor faltante y detectar cualquier anomalía. Esto incluirá estadísticas descriptivas básicas y visualización de datos para descubrir patrones iniciales.

Data Cleaning: Basado en la exploración, los datos serán limpiados para manejar valores faltantes, atípicos e inconsistencias. Esto puede implicar técnicas como imputación, filtrado o transformación para asegurar que el conjunto de datos esté listo para el análisis.

Análisis: Después de la limpieza, realizaremos un análisis más profundo para descubrir tendencias clave, correlaciones y patrones relacionados con precios, ocupación y reseñas. Se emplearán visualizaciones y métodos estadísticos para comprender estas relaciones.

Evaluación y Reporte: La fase final implicará evaluar los resultados del analisis y proporcionar un informe completo sobre los insigts. Esto incluirá información práctica y recomendaciones para las partes interesadas de Airbnb.

Data Understanding

El conjunto de datos consta de 102,599 filas y 26 columnas, representando una colección detallada de anuncios de Airbnb en la ciudad de Nueva York. Cada fila corresponde a un anuncio único e incluye información sobre el anfitrión, la ubicación, los precios, la disponibilidad y otras características relevantes. El conjunto de datos incorpora diversos tipos de datos, como numéricos, categóricos y textuales, lo que lo hace adecuado para una

variedad de análisis. Sin embargo, se han identificado problemas iniciales, como datos faltantes y valores atípicos, que deben abordarse para obtener resultados confiables.

Descripción de los Campos

- id: Identificador único para cada anuncio de Airbnb.
- host id: Identificador único para cada anfitrión, que puede manejar varios anuncios.
- host_identity_verified: Indica si la identidad del anfitrión ha sido verificada (verified o unconfirmed).
- host name: Nombre proporcionado por el anfitrión al registrarse.
- neighbourhood group: El distrito donde se encuentra el anuncio (por ejemplo, Manhattan, Brooklyn).
- neighbourhood: Una zona más específica dentro del distrito.
- lat: Coordenada de latitud de la ubicación del anuncio.
- long: Coordenada de longitud de la ubicación del anuncio.
- country: País donde se encuentra el anuncio, predominantemente United States.
- country code: Código ISO del país, típicamente US.
- room type: Especifica si la propiedad es un Entire home/apt, Private room o Shared room.
- Construction year: Año en que se construyó la propiedad.
- price: Costo por noche del anuncio (en USD), almacenado como cadena de texto con un signo de dólar.
- service fee: Cargos adicionales impuestos por Airbnb, también almacenados como cadena de texto.
- instant_bookable: Indica si los huéspedes pueden reservar instantáneamente sin aprobación previa (True o False).
- cancellation policy: Tipo de política de cancelación (strict, moderate, etc.).
- minimum nights: Número mínimo de noches requerido para reservar el anuncio.
- number of reviews: Total de reseñas dejadas por los huéspedes para el anuncio.
- last review: Fecha de la reseña más reciente.

- reviews per month: Promedio de reseñas recibidas mensualmente.
- review rate number: Puntuación promedio de reseñas, típicamente en una escala de 5 puntos.
- calculated host listings count: Número de anuncios gestionados por el anfitrión, lo que refleja su escala de actividad.
- availability 365: Número de días al año que el anuncio está disponible para reservar.
- house_rules: Campo textual que describe las reglas establecidas por el anfitrión para los huéspedes.
- license: Información regulatoria del anuncio, aunque está mayormente ausente en este conjunto de datos.

Data Quality

- Campos como last review y reviews per month tienen datos faltantes en el 15% de las filas. Esto afecta la usabilidad de estos campos para análisis históricos de reseñas.
- El campo house_rules carece de datos en casi el 50% de los anuncios, reduciendo su valor para análisis basados en reglas.
- El campo license contiene datos para solo 2 entradas, haciéndolo irrelevante para la mayoría de los análisis.
- Los campos de precios (price y service fee) están almacenados como cadenas de texto (por ejemplo, \$193) en lugar de valores numéricos, lo que requiere limpieza y conversión.
- El campo Construction year podría incluir valores no realistas, como años futuros o excesivamente antiguos, que deben ser validados.
- Los campos categóricos como host_identity_verified, cancellation_policy y room type deben revisarse para detectar categorías redundantes o inesperadas.
- Los campos geográficos (neighbourhood y neighbourhood group) deben estar correctamente alineados para evitar conflictos en el análisis basado en ubicación.
- Los campos textuales, como NAME y house_rules, pueden proporcionar información valiosa, pero requieren preprocesamiento (por ejemplo, análisis de texto o de sentimiento).

- Los campos numéricos como price, minimum nights y availability 365 pueden contener valores extremos o ilógicos (por ejemplo, precios excesivamente altos o estancias mínimas inusualmente largas) que sesgan los análisis generales.
- El campo last review debe convertirse de texto a un formato de fecha.

Proximos Pasos

- 1. Limpieza de Datos
 - Tratar valores faltantes mediante imputación o eliminación.
 - Validar y limpiar campos como price y Construction year para garantizar precisión.
- 2. Transformación de Datos
 - Convertir campos numéricos basados en texto (price, service fee) a formatos numéricos.
 - Estandarizar campos categóricos para mantener consistencia.
- 3. EDA
 - Examinar correlaciones entre ubicación, precio y tipo de habitación.
 - Identificar tendencias en la actividad de reseñas y disponibilidad por distrito.

Data Preparation & Modelling

La fase de preparación de datos incluyó la limpieza, transformación y segmentación del conjunto de datos para hacerlo adecuado para el análisis y la visualización. El conjunto de datos inicialmente presentaba varios problemas, como valores duplicados, nulos y entradas inconsistentes, los cuales fueron abordados de manera sistemática. Los datos preparados se segmentaron en subconjuntos significativos para análisis específicos, como precios, disponibilidad y noches mínimas.

Instalación y Configuración de Librerías

Se instalaron e importaron las librerías necesarias (numpy, pandas, seaborn, matplotlib y wordcloud) para manejar el análisis de datos y la visualización. Se suprimieron las advertencias para evitar mensajes innecesarios durante la ejecución.

Carga e Inspección Inicial de los Datos

• El conjunto de datos se cargó con pandas y se inspeccionó brevemente con head(), describe() e info() para comprender su estructura, tipos de columnas y posibles problemas como valores nulos o duplicados.

Eliminación de Duplicados

• Se eliminaron las filas duplicadas para garantizar la integridad de los datos y reducir la redundancia.

Eliminación de Columnas

• Se eliminó la columna host name, ya que no era relevante para el análisis.

Tratamiento de Valores Nulos

- Los valores nulos en la columna price se reemplazaron con el precio promedio calculado según el tipo de habitación (room type) y el vecindario (neighbourhood).
- Los valores nulos en la columna service fee se llenaron con 0.

Manejo de Datos Inconsistentes

• Se eliminaron entradas incorrectas en la columna neighbourhood group. Se excluyeron filas con valores como brookln y manhatan.

3. Segmentación de Datos y Visualizaciones

Análisis de Precios

- Se creó un subconjunto (prices_view) con datos completos para las columnas relacionadas con el precio, incluyendo price, neighbourhood group, room type y service fee.
- Se generó un diagrama de caja (boxplot) para detectar valores atípicos y visualizar la distribución de precios.

Análisis de Disponibilidad

- Se creó un subconjunto (availability_view) para analizar la disponibilidad anual.
- Se filtró el subconjunto para incluir solo valores válidos de disponibilidad (1–365 días).
- Se generó un histograma para examinar la distribución de la disponibilidad, y se registraron los valores mínimo y máximo.

Análisis de Reseñas

• Se creó un subconjunto (reviews_num_view) para analizar el número de reseñas en diferentes vecindarios y políticas de cancelación.

Reglas de la Casa

• Se creó un subconjunto (house_rules_view) con propiedades que tienen reglas de la casa (house_rules) válidas.

Tipos de Habitaciones

 Se creó un subconjunto (room_types_view) para analizar la distribución de los tipos de habitación (Entire home/apt, Private room, Shared room) en diferentes vecindarios.

Análisis de Noches Mínimas

- Se creó un subconjunto (min_nights_view) para propiedades con valores válidos de noches mínimas (1–365 días).
- Se registraron los valores mínimo y máximo para esta columna.

Propiedades Reservables de Inmediato

 Se creó un subconjunto (instant_bookable_view) para analizar propiedades con la opción de reserva instantánea (instant bookable).

Datos Geoespaciales

• Se creó un subconjunto (Map_view) que incluye información geográfica (lat, long) junto con detalles de precios y habitaciones, para análisis de mapeo.

4. Exportación de Datos Preparados

• Los subconjuntos clave (prices_view, availability_view, min_nights_view, room_types_view, house_rules_view, reviews_num_view, instant_bookable_view y Map_view) se exportaron como archivos CSV para análisis y visualización posteriores.

Analysis

Precio Promedio Por Neighbourhood Group

El precio promedio de un Airbnb se refiere al costo típico de reservar una propiedad publicada en la plataforma. Esta métrica puede ser útil en los análisis porque proporciona información sobre las tendencias y patrones de precios de los alquileres de Airbnb en una ubicación específica.

Aquí, el precio promedio de Airbnb en el área de Nueva York es de \$625.06.

Al analizar el precio promedio de los anuncios de Airbnb, podemos identificar fluctuaciones en la demanda y la oferta y ajustar las estrategias de precios en consecuencia para maximizar la rentabilidad.

Además, conocer el precio promedio de un Airbnb puede ayudar a las empresas y a las personas a comprender la competencia y la dinámica del mercado de la industria del alquiler.

Aquí, podemos observar que el precio promedio de un Airbnb está entre \$619 y \$629, con el Bronx a la cabeza.

Esta información puede orientar la toma de decisiones en torno a la adquisición de propiedades, las estrategias de precios y las iniciativas de marketing.

En general, el precio promedio de Airbb proporciona un punto de referencia útil para comprender las tendencias y patrones de precios del mercado de alquileres a corto plazo y puede ser una herramienta valiosa para tomar decisiones informadas en la industria.

Disponibilidad promedio por Neighbourhood group y tipo de habitación

La métrica "Disponibilidad promedio por Neighbourhood group y tipo de habitación" en las estadísticas de Airbnb se refiere a la cantidad promedio de días por año que un tipo de habitación específico en un grupo de vecindario en particular está disponible para reservar en la plataforma Airbnb.

Al analizar la disponibilidad promedio, podemos identificar que Queens tiene una mayor disponibilidad durante todo el año. Seguido de Manhattan, Brooklyn, Bronx y Staten Island.

Además, podemos identificar las áreas de alta demanda y los tipos de habitación más populares que generan las tasas de ocupación más altas.

Por ejemplo, podemos observar que Queens tiene una alta disponibilidad promedio para el tipo de habitación de hotel, lo que podría indicar una baja demanda o una falta de popularidad para ese tipo de habitación. Los anfitriones o administradores de propiedades pueden usar esta información para ajustar su estrategia de precios o esfuerzos de marketing para atraer más reservas.

Por otro lado, el tipo de habitación Casa/apartamento completo tiene baja disponibilidad en Queens y puede indicar una alta demanda para ese tipo de habitación.

Los anfitriones o administradores de propiedades pueden ajustar su estrategia de precios para maximizar los ingresos durante las temporadas altas.

Popularidad De Neighbourhoob basado en el numero de reseñas y cancellation policy

Al analizar los datos de Nueva York, se puede ver que la mayor cantidad de reseñas se reciben en Manhattan, seguida de Brooklyn, Queens, Bronx y, por último, Staten Island.

Invertir en áreas como Brooklyn, Bronx y Staten Island con el objetivo de brindar mejores instalaciones, lo que también generaría una mayor cantidad de reseñas excelentes en esas áreas, podría ser una buena oportunidad de negocio.

Por otro lado, dado que hay una mayor demanda de Airbnb en Manhattan, también podría generar una gran oportunidad de negocio.

Reglas de la Casa

Permitir mascotas:

Aumenta el grupo potencial de huéspedes: algunos huéspedes viajan con sus mascotas y pueden preferir alojarse en un Airbnb que les permita llevar a sus amigos peludos.

Tarifas por noche más altas: al permitir mascotas, los anfitriones pueden cobrar tarifas por noche más altas por sus alojamientos, ya que los huéspedes están dispuestos a pagar más por el privilegio de llevar a sus mascotas.

Mejores reseñas y calificaciones: los huéspedes que pueden llevar a sus mascotas pueden dejar mejores reseñas y calificaciones, lo que puede atraer más reservas en el futuro.

Permitir fiestas:

Tarifas por noche más altas: si un anfitrión está dispuesto a permitir fiestas, puede cobrar una prima por su alojamiento. Esto se debe a que muchos huéspedes que desean organizar fiestas pueden no poder encontrar lugares adecuados y pueden estar dispuestos a pagar más por la comodidad de tener un Airbnb.

Clientes frecuentes: si un huésped organiza una fiesta exitosa en un Airbnb, es más probable que reserve el mismo alojamiento nuevamente en el futuro. Reseñas y valoraciones positivas: si una fiesta sale bien y los invitados están contentos con la experiencia, pueden dejar reseñas y valoraciones positivas, lo que puede atraer más reservas en el futuro.

Permitir fumar:

Tarifas más altas por noche: de manera similar a permitir fiestas, los anfitriones que permiten fumar pueden cobrar una prima por su alojamiento, ya que muchos huéspedes que fuman pueden no encontrar alojamiento adecuado en otro lugar. **Clientes frecuentes**: los huéspedes que fuman pueden tener más probabilidades de reservar el mismo alojamiento nuevamente en el futuro si saben que se permite fumar. **Menos competencia**: puede haber menos alojamientos que permitan fumar, lo que significa menos competencia para los anfitriones que están dispuestos a permitirlo. Sin embargo, es importante tener en cuenta que también existen posibles desventajas asociadas. Por lo tanto, la aceptación de tales reglas de la casa debe ser moderada y tal vez tener una zona para fumadores, permitir fiestas y ruidos fuertes con un límite de tiempo podría ser beneficioso.

Tipos De Habitacion por Neighbourhood

Si analizamos las estadísticas de los anuncios de Airbnb en el área de Nueva York, la mayor cantidad de Airbnb en Manhattan y Brooklyn son casas/apartamentos completos, seguidos de habitaciones privadas y, por último, habitaciones compartidas. En Queens, tenemos la mayor cantidad de Airbnb en habitaciones privadas, seguidas de casas/apartamentos completos y, por último, habitaciones compartidas. Dado que buscamos una oportunidad de inversión, Queens, Bronx y Staten Island tienen una cantidad significativamente baja de anuncios y, por lo tanto, una gran oportunidad de inversión. Sin embargo, la razón de las cifras bajas podría ser la demanda, ya que hay más demanda en las áreas de Manhattan y Brooklyn, la inclusión de más Airbnb brindaría a los clientes una variedad de opciones para elegir y, por lo tanto, mejoraría los ingresos.

Airbnb's Construidos despues del 2020

Tenemos el mayor número de construcciones en el área de Manhattan y el menor en Staten Island. A partir de estas cifras, podemos analizar lo siguiente

- a. Aumento de la demanda en las áreas de Manhattan y Brooklyn
- b. Aumento de la competencia entre Airbbs
- c. Aumento de las comodidades y atracciones en las zonas populares.

De todas nuestras representaciones podemos concluir que las áreas de Manhattan y Brooklyn tienen una alta demanda y competencia, y por lo tanto invertir en estas áreas podría traer más ingresos

Time Series Analysis por tipo de Habitacion

Este gráfico muestra que la cantidad de anuncios ha aumentado con el tiempo, siendo los años más recientes los que han tenido los recuentos más altos.

Por ejemplo, en 2015, hubo 1817 anuncios de casas o departamentos completos y 1554 anuncios de habitaciones privadas. En cambio, en 2013, solo hubo 1785 anuncios de casas o departamentos completos y 74 anuncios de habitaciones compartidas.

Esta tendencia podría deberse a una variedad de factores, como un aumento en la demanda de alquileres a corto plazo en la ciudad de Nueva York, un aumento en la cantidad de anfitriones que anuncian sus propiedades en Airbnb o un aumento en la construcción de propiedades adecuadas para alquileres a corto plazo.

También podemos ver que el tipo de habitación más común varía según el año de construcción. Por ejemplo, en 2013, los anuncios de casas o departamentos completos fueron los más comunes, mientras que en 2015, los anuncios de habitaciones privadas fueron los más comunes. Esta información podría ser útil para que los anfitriones e inversores de Airbnb comprendan mejor las preferencias de los huéspedes y tomen decisiones informadas sobre qué tipos de anuncios ofrecer.

Top 10 Barrios De NYC

Minimo de noches

Podemos observar que la mayoría de los anuncios tienen un requisito de estadía mínima de 1 a 5 noches, y la mayoría de ellos están disponibles por un mínimo de 1 noche y 3975 anuncios requieren un mínimo de 5 noches. A medida que aumenta el requisito de estadía mínima, la cantidad de anuncios disminuye, con una caída abrupta en 13 noches, donde 12928 anuncios tienen un requisito de estadía mínima de 13 noches.

Esta información puede ser útil tanto para anfitriones como para huéspedes de anuncios de Airbnb en la ciudad de Nueva York. Para los anfitriones, puede proporcionar información sobre el requisito de estadía mínima promedio de sus competidores y ayudarlos a

determinar si necesitan ajustar sus propios requisitos de estadía mínima para ser más competitivos. Para los huéspedes, puede ayudarlos a identificar anuncios que coincidan con la duración de estadía deseada y evitar perder tiempo en anuncios con requisitos que no satisfacen sus necesidades.

También es interesante notar que hay muy pocos listados con requisitos de estadía mínima de 13 noches, lo que podría indicar que este tipo de alquileres a largo plazo pueden ser comunes en el mercado de Airbnb en la ciudad de Nueva York.

Airbnb's Mas Costosos

- Es interesante notar que cada anuncio está ubicado en un barrio diferente, con dos en el Bronx, dos en Queens y uno en Manhattan. Esto indica que los anuncios de lujo se pueden encontrar en toda la ciudad, no solo en barrios o distritos específicos.
- El precio de cada anuncio es de \$1,200.00, que es significativamente más alto que el precio promedio de un anuncio de Airbnb en la ciudad de Nueva York. Esto sugiere que los anfitriones pueden estar apuntando a viajeros de alto nivel que están dispuestos a pagar un precio superior por alojamientos de lujo.
- Las descripciones de los anuncios también sugieren que ofrecen experiencias únicas y lujosas. Por ejemplo, un anuncio cuenta con un dormitorio lleno de sol en un barrio de moda, mientras que otro ofrece un gran apartamento de 3 habitaciones con mucho espacio.
- En general, estos datos pueden ser útiles para los viajeros que buscan alojamientos de alto nivel en la ciudad de Nueva York. Muestra que los anuncios de lujo se pueden encontrar en toda la ciudad y que es posible encontrar experiencias únicas y lujosas en una variedad de barrios y distritos.