New York City Airbnb Rental Analysis

Shelby Steele

Executive Summary

Airbnb is a global community of hosts and travelers that operates as an online marketplace via website and mobile app for lodging, homestays, vacations rentals, and tourism activities. It allows everyday people to share their space and environment by providing guests the feeling of connection and comfortability while exploring a new city. In this analysis, the primary focus will be Airbnb rentals in New York City through an analysis of geographical trends, price predictions, and the impact of bookings at different times of the year.

Data Source

The primary New York City Dataset comes from an external source, Airbnb Inside which explores different descriptive statistics like room type, short-term rentals, and host information based on a selected city.

The following Airbnb activity is included in the primary New York City Dataset:

- Geographical Location: including neighborhoods, latitudes, and longitudes
- Reviews: including average review rate, last review, and overall total reviews
- Listing Information: including price, availability, and listings per host

Data Dictionary

Data Limitations and Ethics

- The data limitations within the New York City dataset are manual entries of Airbnb activity. This can be prone to error due to the nature of how it's recorded and updated. The last update was 4 months ago and spans back to 2012.
- Reviews are a large categorical variable within the dataset and could have collection bias based on motivation behind a certain level of review resulting in less bookings.
- There was Personal Identifying Information (PII) within the primary dataset, such as "host_name" which was removed for data security.

Data Quality and Integrity Checks

- Removed PII issue for host_name column and dropped 8 others due to irrelevance to analysis
- Changed 17 column names to a more understandable heading
- Checked for missing data
  - All were left since the impact per column was minimal.
- There were 541 duplicates removed
- Data types were changed for the following columns
  - Verified: str
  - Borough: str
  - Neighborhood: str
  - Latitude: float64
  - Longitude: float64
  - Instant_booking: str
  - Cancellation_policy: str
  - Room_type: str
  - Build_year: float64
  - Price_per_night: float64
  - Service_fee: float64
  - Minimum_night_stay: int16
  - Total_reviews: int16
  - Last_review_date: datetime64[ns]
  - Reviews_per_month: float64
  - 5_star_rating: int8
  - Listings_per_host: int16
  - Availablity_in_next_365_days: float64

Data Profile:

Original Shape: (102599, 26) : New Shape: (102058, 18)

Dataset contains Airbnb information from 2012 - 2022

Defining Questions

- Main Question:
  - Where in New York City can you expect the most revenue out of your Airbnb ?
- Other Questions to be answered:
  - How many bookings per year does the average place have according to the neighborhood ?
  - How does the minimum stay affect the service fee ?
  - How does the room type affect the price ?
  - Does the cancelation policy affect the price ?
  - Does the ability to instantly book affect the price ?
  - What are the top 10 highest revenue neighborhoods ?