DOCUMENTATION:

Task 1: EDA on AB_NYC_2019

The EDA has been performed on a dataset named AB_NYC_2019. This dataset contains the information of Airbnb listings in New York City in 2019.

Firstly to explore and analyze the data, some required libraries pandas , numpy , seaborn and pyplot are imported. This dataset contains 48895 rows and 16 columns. The columns are host_id, host_name, neighbourhood_group, neighbourhood,latitude, longitude, room_type,price,minimum_nights, number_of_reviews,reviews_per_month,calculated_host_listings_count and availability_365

No duplicate rows were found in this dataset but there were some missing values in name, host_ name, last_review and reviews_per_month columns. It can be easily observed that some columns are not required to perform EDA. So, the id, name and last_review columns are dropped so it was not required to deal with their missing values. But , the missing values of host_name and reviews_per_month are handled.

INSIGHTS:

1. "Micheal" is the top host and had 417 bookings during the whole year.
2. Most people liked to stay in the neighbourhood_group "Manhattan" than the other neighbourhood groups which becomes 44.3%of the total.
3. The maximum rental price found out is 10,000 but it is an outlier.
4. After finding out 16 as the minimum and 334 as the maximum price, a new dataframe is made which only has rows having prices between these ranges.
5. With the help of new dataframe, it is observed that the neighbourhood Bronx mostly has rental prices between the ranges of 25 and 100, Brooklyn between 30 to 200, Manhattan between 50 to 300, Queens between 30 to 100 and Staten island between 30 to 100.
6. 52% people booked the entire home or apartment, 45.7% booked the private room and only 2.7% booked the shared room.
7. The entire home/apartment had higher prices and shared rooms had least on average.
8. The higher numbers of entire home/apartments are available in Manhattan but the most private and shared rooms are available in Brooklyn.