Capstone Project Submission

Instruction:

1) Please fill in all the required information.
2) Avoid grammatical errors.

Team leader Name, Email and Contribution:
Team leader: Ade suchit shrimant **Contribution Role:** • **Ade suchit shrimant (suchitaade@gmail.com)** • Data Warngling. • What can we learn about different hosts and areas? • Room type based on neighborhoods groups • Which Appartment have been more used by travellers? • Which hosts are the busiest and why? • Is there any noticeable difference of traffic among different areas and w hat could be the reason for it? • How many properties are available for more than 100 days? • check how the average price varies for different room types? • Finding total count of each room types? • Which neighbourhood group has highest number of Airbnb? • Which room available per Neighbourhood Group? • Conclusion.
Please paste the GitHub repo link. GitHub Link: https://github.com/aadesuchit/Airbnb_booking_analysis
Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-400 words)

This Airbnb ('AB_NYC_2019') dataset for the 2019 year appeared to be a very rich dataset with a variety of columns that allowed us to do deep data exploration on each significant column presented. At the beginning we did some data cleaning, which included replacing null values and replacing some values which seems to be wrong. Then we took insights from the data as follows. First, we have found hosts that take good advantage of the Airbnb platform and provide the most listings.

We found that our top host has 327 listings. After that, we proceeded with analysing how prices differed with different room types and in different neighbourhood groups. We also analysed which type of rooms did people preffered the most to stay and found out that people preffered mostly entire houses. Next, we put good use of our latitude and longitude columns and used to create a graph color-coded by the neighbourhood groups and shape coded by their availability. And we also checked which rooms are available more than 100 days. We even checked about how each column were co related with one another. Overall, we discovered a very good number of interesting relationships between features and explained each step of the process.

1. Most number of listing from Manhattan by host name Sonder(NYC) a nd then blureground and Michael.

2. 1250 rooms are always available in the year and 17500 rooms are not available in the year.

3. In neighbourhood group maximum coordinates is Queens and Brooklyn and minimum coordinates is staten island and Bronx

4. Manhattan has highest number in Airbnb spread across the area of nei ghbourhood group. Then Brooklyn then queens.

5. The entire home and private rooms has the maximum used by travelle r. and shared rooms minimum used by traveller.

6. Sonder(NYC) is the most bussiest host and then Michael. Because the se host listed room type as entire room and private room which is prefer red by most number of people.

7. among the heavy traffic areas are Manhattan and Brooklyn nei ghbourhood group.

8. There are only 39.19% of properties are available for more than 100 days.

9. The most expensive room types are the Entire home and then private r ooms.

10. From the above histplot shows the relation between the price of ea ch room type that's is private room vs Entire room/apt vs shared room .From the first figure we are showing that Price less than 600 & From the 2nd data we are showing the prices of Private room , Entire room/apt and shared room less than $600.

11. From the above line chart we are showing the relationship betwee n the price and updated price . Hence, we are observing that there is bit vartion between price and updated price sometimes it is high and someti mes its low. It will vary.