EDA PROJECT ON AIRBNB NYC

Team:Artistic Algorithm

Chandan Baraliya

Team Member's Name, Email and Contribution:

1-Chandan Baraliya (chandan.baraliya@gmail.com)

a) Upload dataset to google colab and explain dataset to team member.

b) Analyze null values and filter them.

c) Do Data wrangling.

d) Data visualization on neighborhood group, room types, prices at various locations, price and room type with latitude and longitude, availability_365, Minimum night stay in different rooms and Correlation heat map. Overall, 8 observations

e) Removing outliner from price.

f) Make PPT

g) Project summary and Technical write up.

2-Bhupendra Singh (theduston156@gmail.com)

a) Upload Data set

b) Data wrangling self

c) Data visualization for host, number of reviews, locations.

d) Make technical write up

e) Make PPT

f) Project summary and Technical write up.

Please paste the GitHub Repo link.

Github Link: - https://github.com/ChandanBar/EDA_on_airbnb.git

Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches, and your conclusions. (200-400 words)

Project Content:

Since 2008, guests and hosts have used Airbnb to expand on traveling possibilities and present more unique, personalized way of experiencing the world. This dataset describes the listing activity and metrics in NYC, for 2019.This data file includes all needed information to find out more about hosts, geographical availability, necessary metrics to make predictions and draw conclusions. We in our project tried to provide answers to the following states of problems.

1.Problem statements:

- What can we learn about different hosts and areas?
- What can we learn from predictions? (ex: locations, prices, reviews, etc)
- Which hosts are the busiest and why?
- Is there any noticeable difference of traffic among different areas and what could be the reason for it?

Approach-

In this project, we had performed a descriptive and exploratory analysis of the data, to understand how the phenomena of each variable behave individually and transversely, in addition to generate hypotheses useful for future decision-making. To complete the project, we divided the task in several categories as in below.

Steps involved:

- **Libraries and data loading**

In the first we import some useful pandas' libraries like NumPy, pandas, seaborn, matplotlib. Image.

- **Checking our Dataset**

In these we load our dataset, and we checked the data we are going to work with.

- **Null values Treatment**

Our Dataset contained few numbers of null values for which we checked analyze null values and treated it for cleaning and filtering.

- **Duplicates values Treatment**

Since our data set contain duplicate values so we removed duplicate values from our data set.

- **Exploring and Visualizing Data**

We explored the data using various techniques and found some useful information from it from which we could have few good conclusions.

Conclusion

Various Conclusions were observed looking to our data analysis which may surely help to expand the business of Airbnb and we could derive a lot of useful information from it. Few are stated in as below.

1. Our distribution chart after data wrangling showed us our top busiest host and the top performing host from the list of counts and people review, we had in our data. We came to now our top busiest host were Micheal, David, Sonder (NYC) and john.

2. We Also found that our top hosts to have good reviews from customer were Dona, Jj, Maya, Carol, Danielle.

3. We then found that Manhattan remains the top of tourist attraction place as the number of people count were higher. Manhattan has the highest number of listings of about 44.3 % followed by Brooklyn of 41%. State Island stands the least number of listings less the 1 %.

4. We did price analysis and got as usual Manhattan the costliest followed by Brooklyn and other three. As usual Manhattan being the costliest place to live in NYC, having average price more than 140 USD followed by Brooklyn with around 80 USD on an average for the listings. Queens and Staten Island has nearly the same of 75 USD. The highest price range could go just above 360 USD.

5. We further found a useful insight that Manhattan has the highest number of Entire apartment room type and Brooklyn has the highest number of Private room types.

6. We also got to know that shared rooms remain the least choice of people.

7. The minimum number of nights to stay in an apartment is highest as compared to private room types.

8. The number of reviews and availability 365 had a co- relation which says people like to stay less where reviews are less and stay more where review are more.

9. We also got to know that the avg. cost of each room type is as below

Entire Home may averagely cost 150 USD.

Private room average cost around 65 USD.

Shared rooms cost the least price of less than 50 USD.

10. We find out the availability of rooms at neighbourhood group:

Staten Island has the most availability of rooms all the year.

Brooklyn having the least availability of rooms through a year with the second most count listing gives an opportunity to have an increase in number of rooms. Manhattan, Bronx, and Queens have nearly an equal availability of rooms throughout the year.

11. The people who prefer to stay in Entire home or Apartment they are going to stay bit longer in that Neighborhood only.

12. The people who prefer to stay in Private room they won't stay longer as compared to Home or Apartment.

13. Most people prefer to pay less price.

14. If there are a greater number of Reviews for Neighborhood group that means that place is a tourist place.

15. If people are not staying more than one night means they are travelers.