

EDA PROJECT ON AIRBNB NYC

By:-Kartik Kumar

- **Kartik Kumar(kartikbhatoye777@gmail.com)**
 - a. Upload dataset to google collab and understand dataset .
 - b. Analyse null values and filter them.
 - c. Do Data exploration
 - d. Data visualization on neighbourhood 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. Make PPT
 - f. Project summary and Technical write up.

- **GitHub Repo link.**

<https://github.com/KartikKumar7070/Airbnb-Dataset.git>

Project Content:

Since 2008, visitors and hosts have used airbnb to expand on traveling possibilities and use more precise, customized way of experiencing the world. This dataset describes the listing pastime and metrics in nyc, for 2019. This data record includes all wanted records to discover more about hosts, geographical availability, vital metrics to make predictions and draw conclusions. We in our project tried to provide answers to the following states of problems.

1. Problem statements:

- a. What can we learn about different hosts and areas?
- b. What can we learn from predictions? (ex: locations, prices, reviews, etc)
- c. Which hosts are the busiest and why?
- d. Is there any noticeable difference of traffic among different areas and what could be the reason for it?
- e. What is Neighbourhood Group vs. Availability Room?
- f. What are Properties in every neighbourhood group?
- g. What is Map of Neighbourhood group?
- h. What are Price distribution in different types of rooms?

Approach-

On this venture, we had finished a descriptive and exploratory analysis of the data, to apprehend how the phenomena of every variable behave in my opinion and transversely, further to generate hypotheses beneficial for future selection-making. To finish the venture, we divided the challenge in several classes 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:

Diverse conclusions had been located trying to our data analysis which may virtually help to make bigger the enterprise of airbnb and we may want to derive numerous useful statistics from it. Few are stated in as underneath.

1.Our distribution chart after information wrangling showed us our pinnacle busiest host and the top acting host from the listing of counts and those assessment, we had in our information. We came to now our top busiest host had been micheal, david, sonder (nyc) and john. We Also found that our top hosts to have good reviews from customer were Dona, Jj, Maya, Carol, Danielle.

2.We then discovered that manhattan remains the top of visitor attraction vicinity because the variety of people remember had been better. Manhattan has the very best wide variety of listings of about forty four. 3 % followed via brooklyn of 41%. Country island stands the least wide variety of listings less the 1 %.

3.We further observed a beneficial insight that new york has the best variety of entire condominium room kind and brooklyn has the very best range of personal room sorts.

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

5.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.

6.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.

7.We find out the availability of rooms at neighbourhood organization: staten island has the maximum availability of rooms all the year. Brooklyn having the least availability of rooms thru a year with the second maximum count listing gives an opportunity to have an increase in range of rooms. New york, bronx, and queens have nearly an same availability of rooms during the year.

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

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

10.Most people prefer to pay less price.

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

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

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

