**AirBnb Exploratory Data Analysis**

**Kabeer Pande**

**Abstract:**

Airbnb, Inc. is an American company that operates an online marketplace for lodging, primarily homestays for vacation rentals, and tourism activities. Based in San Francisco, California, the platform is accessible via website and mobile app. Our goal is to analyze the dataset over pricing, neighbourhood and crowd and draw various conclusions from it.

**Use case:**

In my case, the use case is of a solo traveller who is looking for a peaceful time away from the noise of the city. He is travelling on a budget so he prefers a cheap neighbourhood. He has planned a week’s trip so the minimum nights has to be seven or less. In addition the user requires a private room because he wants a peaceful and alone time.

**Description of the dataset:**

The dataset of interest ‘df’ is the New York City Airbnb dataset containing 48895 rows and 16 columns. There are some null values which I have taken care of.

**Data cleaning:**

Before filtering the data according to our requirement, we first removed the null values and outliers from the dataset. Our dataset contains null values which might tend to disturb our accuracy hence I dropped them at the beginning of my project in order to get a better result.

Then I filtered my data for use case. I started with filtering minimum\_nights<7 and <0 because it is a week’s trip. Then I chose a less crowded neighbourhood and filtered the room type by private rooms.

**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?
    - Use case

**Methodology**:

I used python for performing all the steps required to draw conclusions from the dataset. I cleaned and filtered the dataset using built in python functions. Then I visualized the data according to my need using matplotlib and seaborn. My approach was to filter out data according to my use case. Therefore data I used contained inputs of Airbnbs with price<100 and minimum nights <7.

**Conclusions:**

1. People prefer to stay in cheaper accommodations.
2. If you are looking for cheaper stays then it is recommended to not travel to Bronx in the month of February and Queens in the month of November
3. June and July are the busiest months.
4. Private rooms are preferred as compared to other room type
5. Brooklyn and Manhattan experiences highest traffic. This is because they are tourist place. The average price in also Brooklyn and Manhattan are higher as compared to other neighbourhoods.
6. Queens is the ideal destination for our client as a lot of Airbnbs with reviews are available there that fills his criteria.

**This project can be further extended by:**

1. Breaking the data month wise to see who are the top hosts during the busiest months, i.e, June and July.
2. Deeper analysis of Bronx in the month of February because at peak Bronx is even more expensive than Manhattan. Therefore big profit can be made with good management.