METHODOLOGY

Problem Statement:

Airbnb, Inc. is a leading American corporation facilitating an online platform connecting travellers with diverse lodging options, including homestays for vacation rentals, and a range of tourism experiences.

Airbnb has seen a major decline in revenue. Now that the restrictions have started lifting and people have started to travel more. Airbnb wants to make sure that it is fully prepared for this change.
The different leaders at Airbnb want to understand some important insights based on various attributes in the dataset so as to increase the revenue such as –

- Which type of hosts to acquire more and where?
- The categorisation of customers based on their preferences.
- What are the neighbourhoods they need to target?
- What is the pricing ranges preferred by customers?
- The various kinds of properties that exist w.r.t. customer preferences.
- Adjustments in the existing properties to make it more customer oriented.
- What are the most popular localities and properties in New York currently?
- How to get unpopular properties more traction? and so on...

Data Used:

The dataset has been used for this case study is Airbnb NYC. This dataset contains information about different Airbnb listings along with their hosts, locations, prices, and other attributes.

Below is the screenshot of the columns in the dataset to get a better idea of what each column signifies.

Column	Description
id	listing ID
name	name of the listing
host_id	host ID
host_name	name of the host
neighbourhood_group	location
neighbourhood	area
latitude	latitude coordinates
longitude	longitude coordinates
room_type	listing space type
price	
minimum_nights	amount of nights minimum
number_of_reviews	number of reviews
last_review	latest review
reviews_per_month	number of reviews per month
calculated_host_listings_count	amount of listing per host
availability_365	number of days when listing is available for booking

Note: The price column contains the price/night.

Tools & Techniques:

1. Microsoft Excel: We have used MS Excel platform for importing, cleaning and data transformation.

2. Tableau: We used Tableau for data visualization and comparing different variables to find key insights.

Below are the screenshots of the process that is applied for Data Analysis in MS Excel:

a) Reading the dataset
b) Finding Null values:

Column	Null Count	% Null
id	0	0%
name	16	0%
host_id	0	0%
host_name	21	0%
neighbourhood_group	0	0%
neighbourhood	0	0%
latitude	0	0%
longitude	0	0%
room_type	0	0%
price	0	0%
minimum_nights	0	0%
number_of_reviews	10052	21%
last_review	10052	21%
reviews_per_month	0	0%
calculated_host_listings_count	0	0%
availability_365	0	0%

c) Replaced the null values of number of reviews & last review columns with zero as they contain dates and replacing dates with mean or mode is not logical.
d) The host name & host id columns have very few null values so we dropped them.
e) Changing the data type and making the correct data structure for all attributes.
f) Started data visualization in Tableau to get key insights.