

AIR BNB Case Study – SANDEEP SUMAN PRADHAN

Methodology Document PPT-1:

In the case study we have used Jupiter notebook to perform initial analysis of the data and Tableau for data analysis and visualization.

Analysis using Jupiter Notebook: Data Set Used: AB_NYC_2019.csv

Number of Rows: 48895

Number of Columns: 16

upGrad | Learning x Storytelling Case x Air BnB Case Study x aboutblank x aboutblank x sandeep12d3/AirBnB Case Study x Home Page - Storytelling Case x what tools were used x

localhost:8888/notebooks/Air%20BnB%20Case%20Study-StoryTelling.ipynb#

Welcome to NATO... Gmail YouTube Maps nm_/ https://cdn.upgrad.com/

jupyter Air BnB Case Study-StoryTelling Last Checkpoint: Yesterday at 9:57 AM (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3 (pykernel)

In [1]: # Import the necessary Libraries
import warnings
warnings.filterwarnings("ignore")
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns

In [2]: # Data Conversion and Understanding
airbnb = pd.read_csv(r"C:\Users\ABHIPSA\Desktop\Upgrad\AB_NYC_2019.csv")
airbnb.head(15)

Out[2]:

	id	name	host_id	host_name	neighbourhood_group	neighbourhood	latitude	longitude	room_type	price	minimum_nights	number_of_reviews
0	2539	Clean & quiet apt home by the park	2787	John	Brooklyn	Kensington	40.64749	-73.97237	Private room	149	1	
1	2595	Skylit Midtown Castle	2845	Jennifer	Manhattan	Midtown	40.75362	-73.98377	Entire home/apt	225	1	
2	3647	THE VILLAGE OF HARLEM - NEW YORK I	4632	Elisabeth	Manhattan	Harlem	40.80902	-73.94190	Private room	150	3	
3	3831	Cozy Entire Floor of Brownstone	4869	LisaRoxanne	Brooklyn	Clinton Hill	40.68514	-73.95976	Entire home/apt	89	1	
4	5022	Entire Apt. Spacious Studio/Loft by central park	7192	Laura	Manhattan	East Harlem	40.79651	-73.94399	Entire home/apt	80	10	

Air BnB Case Study-...py Air BnB Case Study-...py AB_NYC_2019.csv Show all

upGrad | Learning x Storytelling Case x Air BnB Case Study x aboutblank x aboutblank x sandeep12d3/AirBnB Case Study x Home Page - Storytelling Case x what tools were used x

localhost:8888/notebooks/Air%20BnB%20Case%20Study-StoryTelling.ipynb#

Welcome to NATO... Gmail YouTube Maps nm_/ https://cdn.upgrad.com/

jupyter Air BnB Case Study-StoryTelling Last Checkpoint: Yesterday at 9:57 AM (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3 (pykernel)

In [3]: # Analyse the dataset
airbnb.shape

6	5121	BlissArtsSpacel	7356	Garon	Brooklyn	Bedford-Stuyvesant	40.68688	-73.95596	Private room	60	45	
7	5178	Large Furnished Room Near B'way	8967	Shunichi	Manhattan	Hell's Kitchen	40.76489	-73.98493	Private room	79	2	
8	5203	Cozy Clean Guest Room - Family Apt	7490	MaryEllen	Manhattan	Upper West Side	40.80178	-73.96723	Private room	79	2	
9	5238	Cute & Cozy Lower East Side 1 bdrm	7549	Ben	Manhattan	Chinatown	40.71344	-73.99037	Entire home/apt	150	1	
10	5295	Beautiful 1br on Upper West Side	7702	Lena	Manhattan	Upper West Side	40.80316	-73.96545	Entire home/apt	135	5	
11	5441	Central Manhattan/near Broadway	7989	Kate	Manhattan	Hell's Kitchen	40.76076	-73.98867	Private room	85	2	
12	5803	Lovely Room 1, Garden, Best Area, Legal rental	9744	Laurie	Brooklyn	South Slope	40.66829	-73.98779	Private room	89	4	
13	6021	Wonderful Guest Bedroom in Manhattan for SINGLES	11528	Claudio	Manhattan	Upper West Side	40.79626	-73.96113	Private room	85	2	
14	6090	West Village Nest - Superhost	11975	Alina	Manhattan	West Village	40.73530	-74.00525	Entire home/apt	120	90	

Air BnB Case Study-...py Air BnB Case Study-...py AB_NYC_2019.csv Show all

upGrad | Learning | Storytelling Case | Air BnB Case Study | aboutblank | aboutblank | sandeep12d3/AirBnB | Home Page - Sandeep | what tools were | +

localhost:8888/notebooks/Air%20BnB%20Case%20Study-StoryTelling.ipynb#

Welcome to NATO... | Gmail | YouTube | Maps | nm_/ | https://cdn.upgrad...

Jupyter Air BnB Case Study-StoryTelling Last Checkpoint: Yesterday at 9:57 AM (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3 (pykernel)

SINGLES

14	6090	West Village	Nest - Superhost	11975	Alina	Manhattan	West Village	40.73530	-74.00525	Entire home/apt	120	90
----	------	--------------	------------------	-------	-------	-----------	--------------	----------	-----------	-----------------	-----	----

In [3]: # Analyse the dataset
airbnb.shape

Out[3]: (48895, 16)

The dataset contains 48895 rows and 16 columns

We have to check for missing values

In [4]: # Missing Values
airbnb.isnull().sum()

Out[4]:

id	0
name	16
host_id	0
host_name	21
neighbourhood_group	0
neighbourhood	0
latitude	0
longitude	0
room_type	0
price	0
minimum_nights	0
number_of_reviews	0
last_review	10052

upGrad | Learning | Storytelling Case | Air BnB Case Study | aboutblank | aboutblank | sandeep12d3/AirBnB | Home Page - Sandeep | what tools were | +

localhost:8888/notebooks/Air%20BnB%20Case%20Study-StoryTelling.ipynb#

Welcome to NATO... | Gmail | YouTube | Maps | nm_/ | https://cdn.upgrad...

Jupyter Air BnB Case Study-StoryTelling Last Checkpoint: Yesterday at 9:57 AM (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3 (pykernel)

airbnb.isnull().sum()

Out[4]:

id	0
name	16
host_id	0
host_name	21
neighbourhood_group	0
neighbourhood	0
latitude	0
longitude	0
room_type	0
price	0
minimum_nights	0
number_of_reviews	0
last_review	10052
reviews_per_month	10052
calculated_host_listings_count	0
availability_365	0
dtype:	int64

In [5]: # Drop columns which are not essential
airbnb.drop(['id', 'name', 'last_review'], axis = 1, inplace = True)

In [6]: # check if the columns are dropped
airbnb.head(15)

Out[6]:

neighbourhood	latitude	longitude	room_type	price	minimum_nights	number_of_reviews	reviews_per_month	calculated_host_listings_count	availability_365
Kensington	40.64749	-73.97237	Private room	149	1	9	0.21	6	365
Midtown	40.75693	-73.98277	Entire	225	1	15	0.38	9	365

```
The dataset contains 48895 rows and 16 columns

We have to check for missing values

In [4]: # Missing Values
airbnb.isnull().sum()

Out[4]: id                0
        name              16
        host_id           0
        host_name        21
        neighbourhood_group 0
        neighbourhood     0
        latitude          0
        longitude         0
        room_type          0
        price              0
        minimum_nights    0
        number_of_reviews 0
        last_review       10052
        reviews_per_month 10052
        calculated_host_listings_count 0
        availability_365   0
        dtype: int64

In [5]: # Drop columns which are not essential
airbnb.drop(['id', 'name', 'last_review'], axis = 1, inplace = True)

In [6]: # check if the columns are dropped
airbnb.head(15)
```

We removed the columns like Id, Name, Last Review which was not giving much information.

Step 2: Data Wrangling:

- Checked for Duplicate rows in the dataset
- Checked for Null Values in the dataset. Columns like name, host-name, last review and review-per-month had null values.
- We dropped the columns where missing values were less and which would not impact our analysis in the dataset
- Identified and reviewed outliers.

Data Analysis and Visualizations using Tableau:

We have used tableau to visualize the data for the assignment. Below are the detailed steps used for each visualization.

- 1) Top 10 Host:
 - We identified the top 10 Host Ids, Host Name with count of Host Ids using the tree map.
- 2) Preferred Room type with respect to Neighbourhood group:
 - We created a pie chart for understanding the percentage of room type preferred w r t neighbourhood group
 - We added Room Type to the colours Marks card to highlight the different Room Type in different colours and count of Host Id to the size
- 3) For Variance of price with Neighbourhood Groups:
 - We used a box and whisker's plot with Neighbourhood Groups in Columns and Price in Rows.

- We changed the Price from a Sum Measure to the median measure.
- 4) Average price of Neighbourhood groups:
 - We created a bubble chart with Neighbourhood Groups in Columns and Price column in Rows.
 - We added the Neighbourhood Groups to the colors Marks card to highlight the different neighbourhood Groups in different colors. Also Put Avg price in Label.
 - 5) Popular Neighborhoods:
 - We took neighbourhood in rows and sum of reviews in column and took neighbourhood groups in colour. • We used filter to show Top 20 neighbours as per the sum of reviews.
 - 6) Neighbourhood vs Availability:
 - We created a dual axis chart using bar chart for availability 365 and line chart for price for top 10 neighbourhood group sorted by price.

Methodology Document PPT 2:

- 1) Room type with respect to Neighbourhood group:
 - We created a pie chart for understanding the percentage of room type preferred w r t neighbourhood group
 - We added Room Type to the colours Marks card to highlight the different Room Type in different colours and count of Host Id to the size
- 2) Neighbourhood vs Availability:
 - We created a dual axis chart using bar chart for availability 365 and line chart for price for top 10 neighbourhood group sorted by price.
- 3) Price range preferred by Customers:
 - We have taken pricing preference based on volume of bookings done in a price range and no of Ids to create a bar chart. We have created bin for Price column with interval of \$20)
- 4) Understanding Price variation w.r.t Room Type & Neighbourhood:
 - We created Highlights Table chat by taking Room Type in rows & Neighbourhood Group in column.
 - We took the average price in colour Marks card to highlight the different Room Type in different colours.
- 5) Price variation w r t Geography:
 - We used Geo location chart to plot neighbourhood, neighbourhood Group in map to show case the variation of prices across.
- 6) Popular Neighborhoods:
 - We took neighbourhood in rows and sum of reviews in column and took neighbourhood groups in colour.
 - We used filter to show Top 20 neighbours as per the sum of reviews.

7) Tools Used:

Data cleaning and preparation: Jupyter Notebook-Python

Visualization: Tableau

Data story Telling and Presentation : Microsoft Powerpoint