

AIRBNB STORYTELLING CASE STUDY

In this case study, we have used Power BI tool to do the initial data cleaning as well as the analysis.

Dataset provided have 48895 rows and 16 columns:



AB_NYC_2019.csv

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
Dataset Description	

Categorical Variables:

- room_type
- neighbourhood_group
- neighbourhood

Continous Variables(Numerical):

- Price
- minimum_nights
- number_of_reviews
- reviews_per_month
- calculated_host_listings_count
- availability_365
- Continous Variables could be binned in to groups too

Location Variables:

- latitude
- longitude

Time Varibale:

- last_review

Variable Categories

➤ Step 1: Importing data to Power BI:

- We have imported the data using **Get data** in the Home panel.

The screenshot shows the Microsoft Power BI Desktop interface. The 'Home' ribbon is active, displaying various toolbars including 'Clipboard', 'Data', 'Queries', 'Relationships', 'Calculations', 'Security', 'Sensitivity', and 'Share'. Below the ribbon, a data table is displayed with columns: id, name, host_id, host_name, neighbourhood_group, neighbourhood, latitude, longitude, room_type, and price. The table contains 30 rows of data, including property listings like 'Lovely large 1-bdrm apt UWS' and 'Best Central Park/NYC View Apt'. A search bar is visible on the right side of the table.

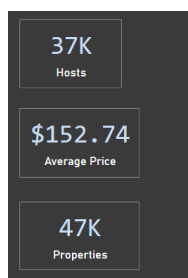
id	name	host_id	host_name	neighbourhood_group	neighbourhood	latitude	longitude	room_type	price
1965712	Lovely large 1-bdrm apt UWS	10153235	Tammie	Manhattan	Upper West Side	40.78122	-73.98473	Entire home/apt	
1979579	Best Central Park/NYC View Apt	8219931	Jason	Manhattan	Upper West Side	40.78468	-73.97137	Entire home/apt	
2003807	UWS MANHATTAN APT FOR SUPERBOWL WE	10317189	Edouard	Manhattan	Upper West Side	40.79321	-73.9694	Entire home/apt	1
2103480	Thanksgiving in Manhattan!	10725397	Andrea	Manhattan	Upper West Side	40.77822	-73.98016	Entire home/apt	
2107372	Best NYC Deal on upper west!	10346860	Shlom And Lyndz	Manhattan	Upper West Side	40.79994	-73.97094	Entire home/apt	
2135766	Super Bowl, 2 bdrm, UWS apartment	4074918	Julie	Manhattan	Upper West Side	40.79478	-73.97474	Entire home/apt	
2219294	Great Studio!	10485182	Jennifer	Manhattan	Upper West Side	40.78835	-73.97257	Entire home/apt	
2224896	NYC SuperBowl Wk 5 Bdrs River View	11353904	Todd	Manhattan	Upper West Side	40.79476	-73.97299	Entire home/apt	4
2243769	Super Bowl New York City Apartment	11460768	Brian	Manhattan	Upper West Side	40.8002	-73.96045	Entire home/apt	1
2249878	Great UWS Apt/Central Park	4708070	Ruben	Manhattan	Upper West Side	40.78531	-73.96988	Entire home/apt	
2249928	Super Bowl Wknd! 3-Bedroom Apt UWS	2675913	Sam	Manhattan	Upper West Side	40.78574	-73.9764	Entire home/apt	1
2273019	Super Bowl Weekend	8105524	Rhett	Manhattan	Upper West Side	40.77018	-73.98643	Entire home/apt	
2285355	SUPER BOWL RENTAL NYC LOW PRICE	11674823	Bruce	Manhattan	Upper West Side	40.77859	-73.97964	Entire home/apt	
229047	Large 2 Bedroom available	9754117	David	Manhattan	Upper West Side	40.77794	-73.97699	Entire home/apt	
3265275	2BR, UWS, Doorman, Balcony, CntrlPK	16512899	Jeremy	Manhattan	Upper West Side	40.7898	-73.9708	Entire home/apt	
3703037	SUBLET in the UWS	18876264	Daniel	Manhattan	Upper West Side	40.77817	-73.97844	Entire home/apt	
4163804	Classic UWS 1 Bed in Doorman Bldg	15573381	Lindsay	Manhattan	Upper West Side	40.78085	-73.98471	Entire home/apt	
4233705	looking for someone to stay 12days	11243583	Joseph	Manhattan	Upper West Side	40.77715	-73.98235	Entire home/apt	
4262120	Columbus Circle and Park Views	6914968	Lisa	Manhattan	Upper West Side	40.7681	-73.98377	Entire home/apt	2
4525475	Manhattan, UWS, Bright, 1BR, & Loft	12149644	Brandon	Manhattan	Upper West Side	40.78304	-73.98146	Entire home/apt	
4620962	Beautiful one bedroom apt 850 sqft!	21685221	Sael	Manhattan	Upper West Side	40.79293	-73.97126	Entire home/apt	
4782522	PENTHOUSE STUDIO WITH LARGE TERRACE	24657274	Daniel	Manhattan	Upper West Side	40.79202	-73.97947	Entire home/apt	
5395116	Studio in UWS- Manhattan	14586567	Paula	Manhattan	Upper West Side	40.79368	-73.97324	Entire home/apt	
5812875	Large studio, view on Central Park	5963594	Geoffroy	Manhattan	Upper West Side	40.77647	-73.98211	Entire home/apt	

➤ Step 2: Data Cleaning:

- Checked datatypes of columns and they are of correct format.
- The columns such as last_review and calculated_host_listings_count are not necessary for our analysis hence removed.
- Columns like name and host_name have null values (very less compared to number of records) which are dropped from dataset.

➤ Step 3: Data Preparation:

- Created 3 new measures: to show the cards in the dashboard
 - 1) Hosts = `DISTINCTCOUNT(AB_NYC_2019[host_id])` - indicating total number of hosts
 - 2) Properties = `DISTINCTCOUNT(AB_NYC_2019[name])` - indicating total number of properties available.
 - 3) Average Price = `AVERAGE(AB_NYC_2019[price])` - indicating the average price of the properties.

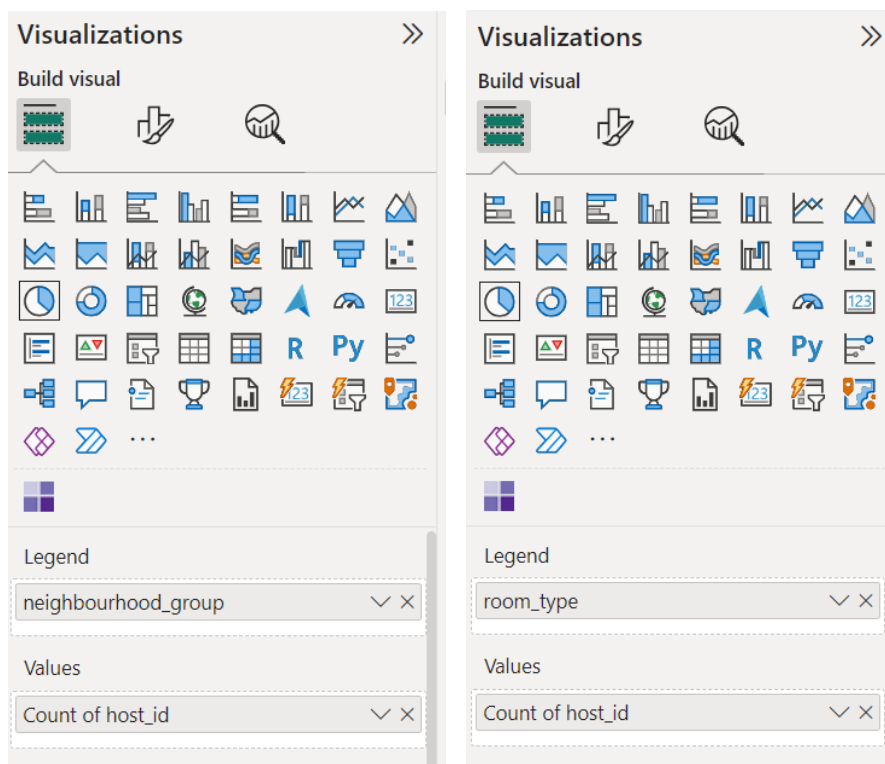


- We have binned the numerical variables `minimum_nights`, `number_of_reviews`, `availability_365` and `price` into different categories.
- 1) `availability_365_categories = SWITCH(True(), AB_NYC_2019[availability_365] <= 1, "Very Low", AB_NYC_2019[availability_365] <= 100, "Low", AB_NYC_2019[availability_365] <= 200, "Medium", AB_NYC_2019[availability_365] <= 300, "High", "Very High")`
 - 2) `minimum_nights_categories = SWITCH(True(), AB_NYC_2019[minimum_nights] <= 1, "Very Low", AB_NYC_2019[minimum_nights] <= 3, "Low", AB_NYC_2019[minimum_nights] <= 5, "Medium", AB_NYC_2019[minimum_nights] <= 7, "High", "Very High")`
 - 3) `number_of_reviews_categories = SWITCH(True(), AB_NYC_2019[number_of_reviews] <= 1, "Very Low", AB_NYC_2019[number_of_reviews] <= 25, "Low", AB_NYC_2019[number_of_reviews] <= 50, "Medium", AB_NYC_2019[number_of_reviews] <= 100, "High", "Very High")`
 - 4) `price_categories = SWITCH(True(), AB_NYC_2019[price] <= 10, "Very Low", AB_NYC_2019[price] <= 100, "Low", AB_NYC_2019[price] <= 500, "Medium", AB_NYC_2019[price] <= 1000, "High", "Very High")`

➤ Step 4: Data Analysis and Visualization:

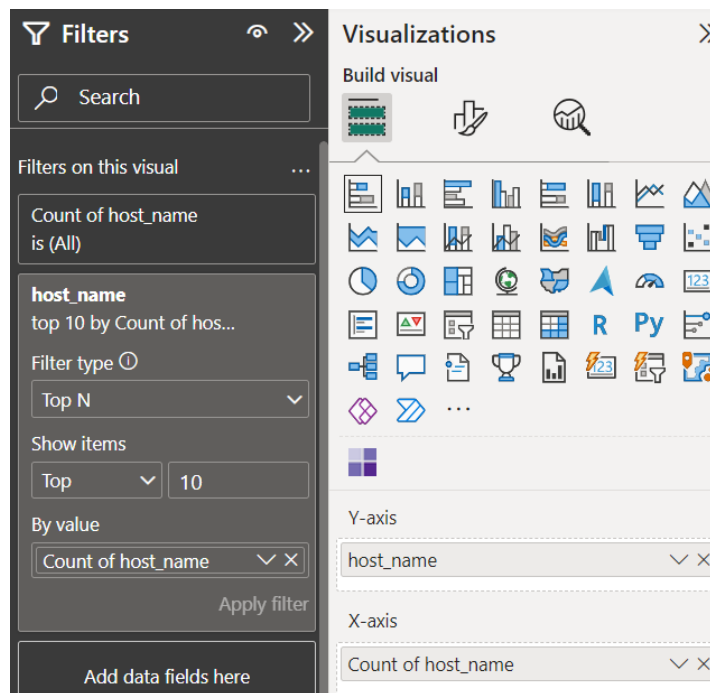
1. Preferred Locations and Preferred Room Types:

We have used pie-chart to identify the preferred neighbourhood groups and room-types with respect to customer bookings (i.e. count of `host_id`)



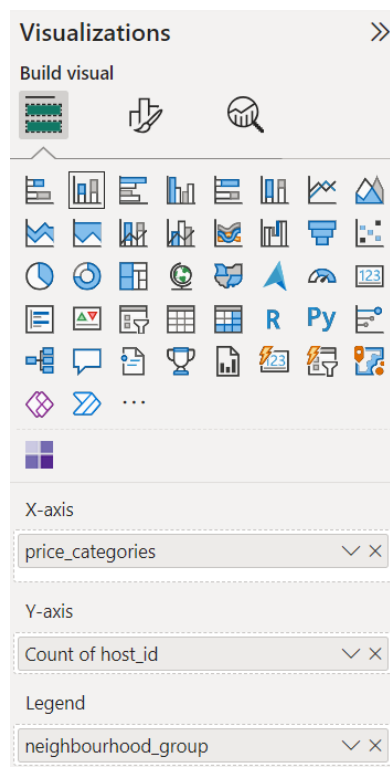
2. Top 10 Hosts:

Identified the top 10 hosts with respect to their count/no: of bookings using the filter panel in horizontal bar chart.



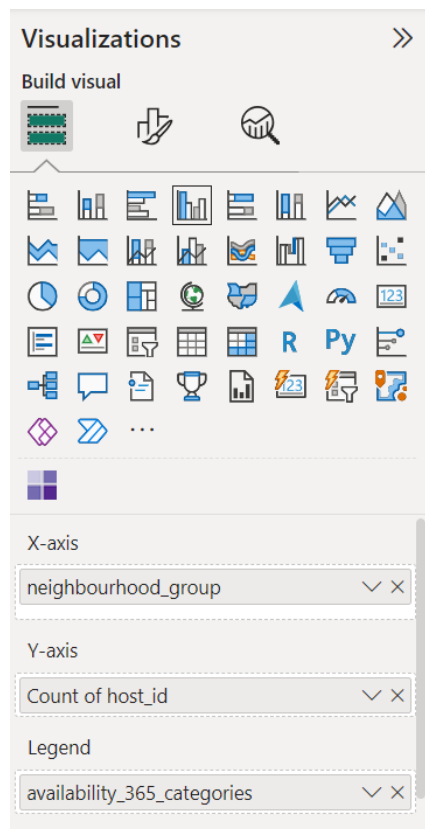
3. Preferred price categories over different neighbourhood groups:

We used a stacked bar chart to represent this with price_categories in x-axis, count of host_id in y-axis and neighbourhood_group in legend.



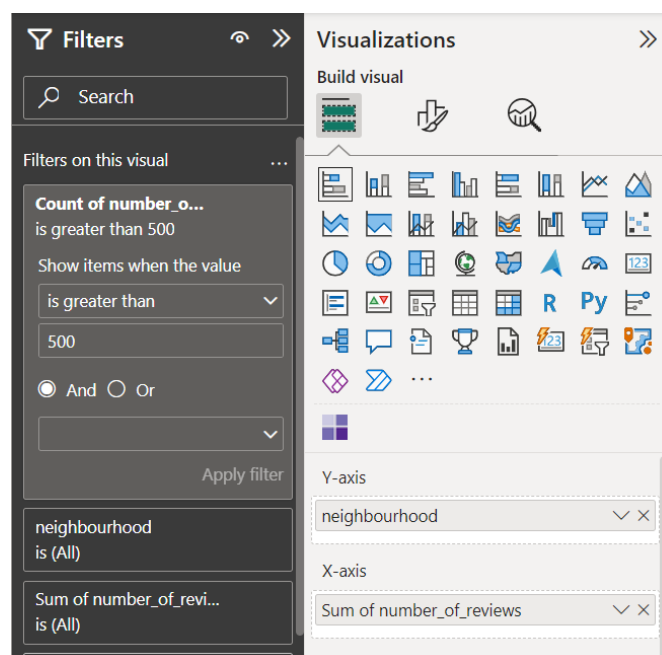
4. Availability Rates over neighbourhood groups:

We used a clustered column chart with neighbourhood groups as x-axis, count of host-id as y-axis and availability_365_categories as legend



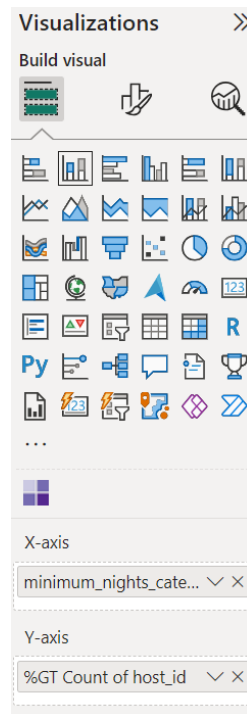
5. Popular Neighbourhoods:

Identified the preferred neighbourhoods with respect to number of reviews.



6. Preferred minimum nights:

We have used a bar chart with minimum_nights_categories on x-axis and count of host_id as percentage in y-axis.



Dashboard:

