

Airbnb Methodology Case Study

IIIT-B

By :- Shruti Thakur, Piyush Rai, Vidyashree Chellapondy

Data Methodology

Method of analysing the code.

Steps

Step 1: Storyboarding

Step 2: Data Wrangling

Step 3: Data Analysis and Visualizations using Tableau:

Important Points ↓

Important Points ↓

Important Points ↓

Step 4: Presentation

Data Methodology



We have used Jupiter notebook to perform initial analysis of the data and Tableau for data analysis and visualization.

Initial Analysis using Jupiter Notebook: Data Set Used:	AB_NYC_2019.csv
Number of Rows:	48895
Number of Columns:	16

Method of analysing the code.

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

<https://gist.github.com/prai85/3b514477e590359ff14c51ec9161bb77>

Steps

Step 1: Storyboarding

- In order to familiarise ourselves with the data, we went through it and made a list of key fields.
- Created a mind map of the presentation's numerous slides.
- Based on this mind map, created a basic template.

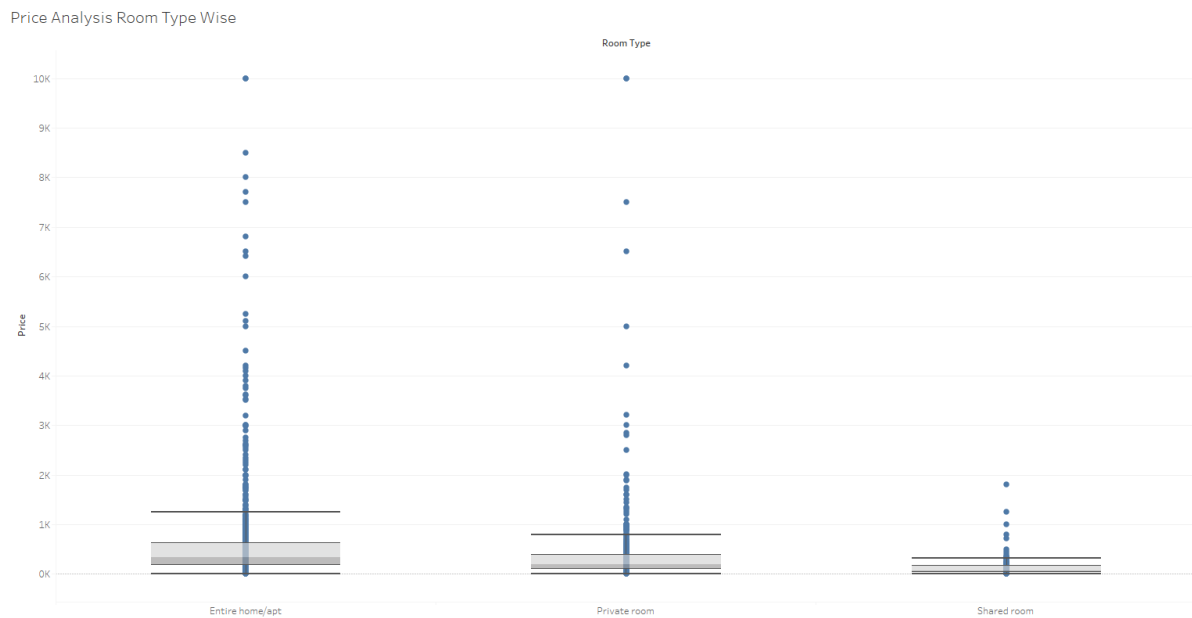
Step 2: Data Wrangling

- No duplicate data was discovered after checking our dataset's duplicate rows.
- Evaluated the dataset's null values. Name, hostname, last review, and review-per-month are examples of columns with null values.
- The column name has been removed because there are fewer missing data and eliminating so won't significantly affect the analysis.
- Evaluated our dataset's formatting.
- Outliers have been found and reviewed.

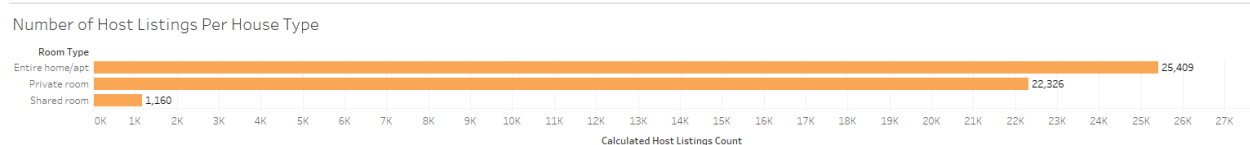
Step 3: Data Analysis and Visualizations using Tableau:

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

- Used Tableau's univariate analysis to look at the fields' distributions, the unique values they contain, the missing values, and any outliers that might exist.
- Price was strongly positively skewed, so the median was quite close to the lower quartile with some outliers as seen in the boxplot below.
 - There was a minor proportion of null values, but they would not have changed the outcome of my study, so I left them alone.



- The maximum number of hosts listed for an entire flat or private room is quite a few, and the median rather than the mean is utilized as the measure of price because price outliers exist. As seen below for a shared room.



- Created a grouped field for Minimum Number of Days assuming null values belonged to the category

Minimum Nights Grouped

```

IF [Minimum Nights]=1 THEN "1"
ELSEIF [Minimum Nights]=2 THEN "2"
ELSEIF [Minimum Nights]=3 THEN "3"
ELSEIF 4<=[Minimum Nights] AND [Minimum Nights]<=5 THEN "4-5"
ELSEIF 6<=[Minimum Nights] AND [Minimum Nights]<=7 THEN "6-7"
ELSEIF 8<=[Minimum Nights] AND [Minimum Nights]<=29 THEN "8-29"
ELSEIF 30<=[Minimum Nights] AND [Minimum Nights]<=31 THEN "30-31"
ELSE ">31" END

```

The calculation is valid.
5 Dependencies
Apply
OK

- Created a calculated field of number of reviews per listing

No. of Reviews per Listing

```

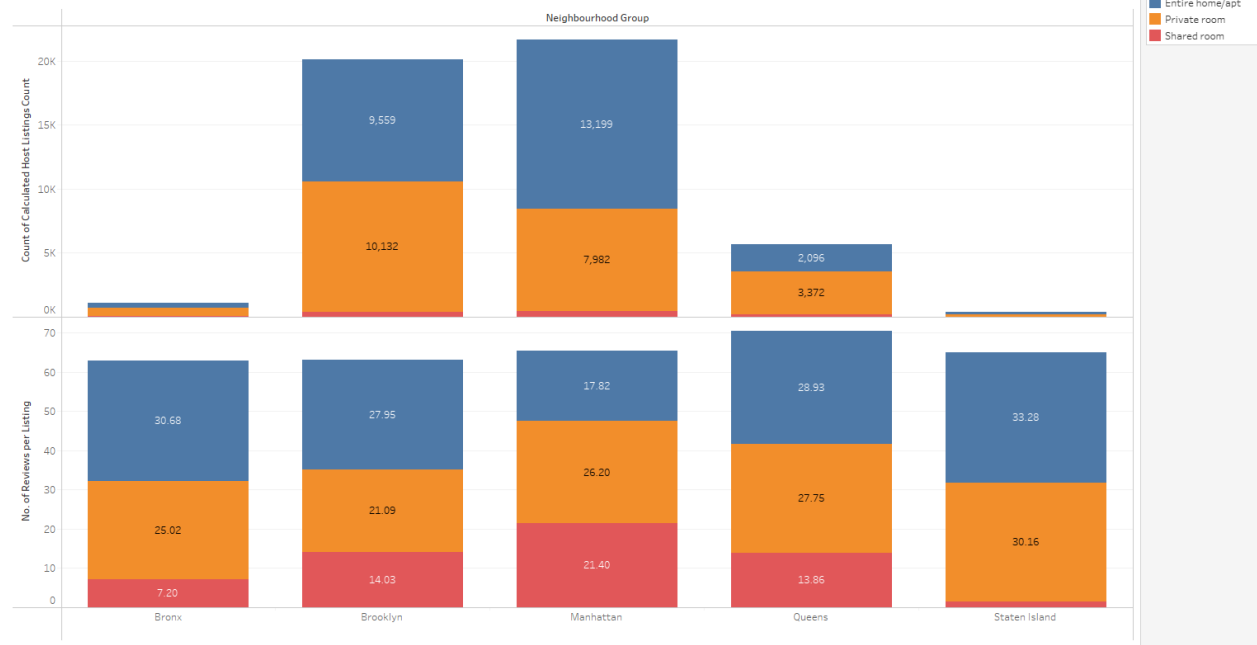
SUM([Number Of Reviews])/COUNT([Calculated Host Listings Count])

```

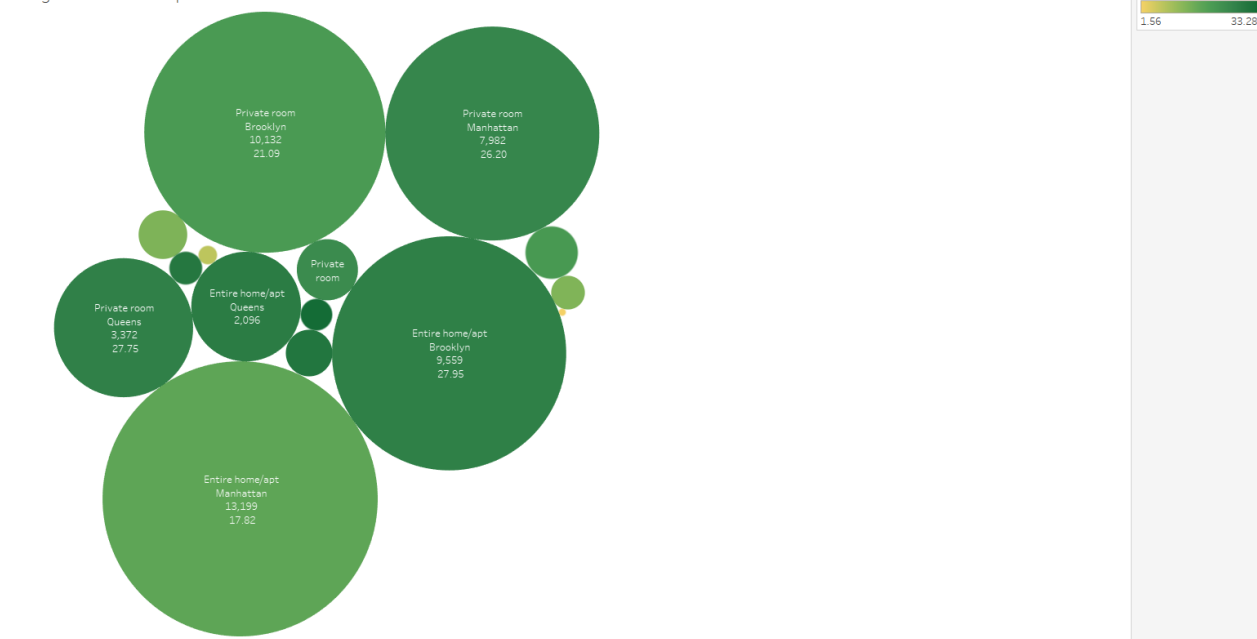
The calculation is valid.
5 Dependencies
Apply
OK

- Checked neighbourhood grouped wise distribution of price and room type

Neighbourhood Group Wise Stats



Neighbourhood Group Bubble Chart



Important Points



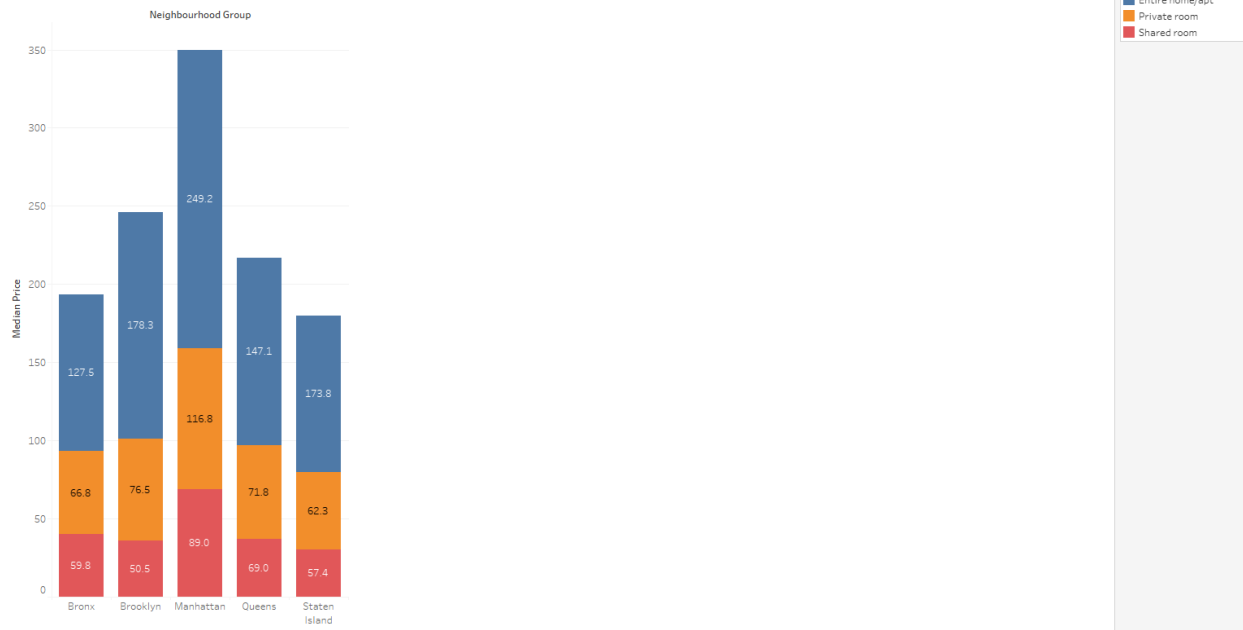


Entire apartments are popular and so are private rooms

Except in Manhattan, shared rooms are very unpopular

Entire room's reviews per listing for Manhattan is 35% lower than the overall average and number of listings is the highest

Price Brackets in Neighbourhoods



Important Points



Manhattan is costliest overall and the cheapest are:

- o Entire apt: Bronx
- o Private room: Staten Islands
- o Shared room: Brooklyn

Found difference in prices between shared rooms, private rooms and entire apartments for each neighbourhood as shown below

I4		fx		=C3/C4-1									
	A	B	C	D	E	F	G	H	I	J	K	L	M
1													
2			Bronx	Brooklyn	Manhattan	Queens	Staten Island		Bronx	Brooklyn	Manhattan	Queens	Staten Island
3			127.5	178.3	249.2	147.1	173.8						
4			66.8	76.5	116.8	71.8	62.3		91%	133%	113%	105%	179%
5			59.8	50.5	89	69	57.4		12%	51%	31%	4%	9%
6													
7													

Important Points



Entire home/apt's price is ~100% more than private rooms except Brooklyn and Staten Island with ~150%

Private room's price is ~10% more than shared rooms except Brooklyn and Manhattan with ~40%

Step 4: Presentation

- Created the presentation by following the pyramid idea and best practices.
- Further guidance for the respective departments.