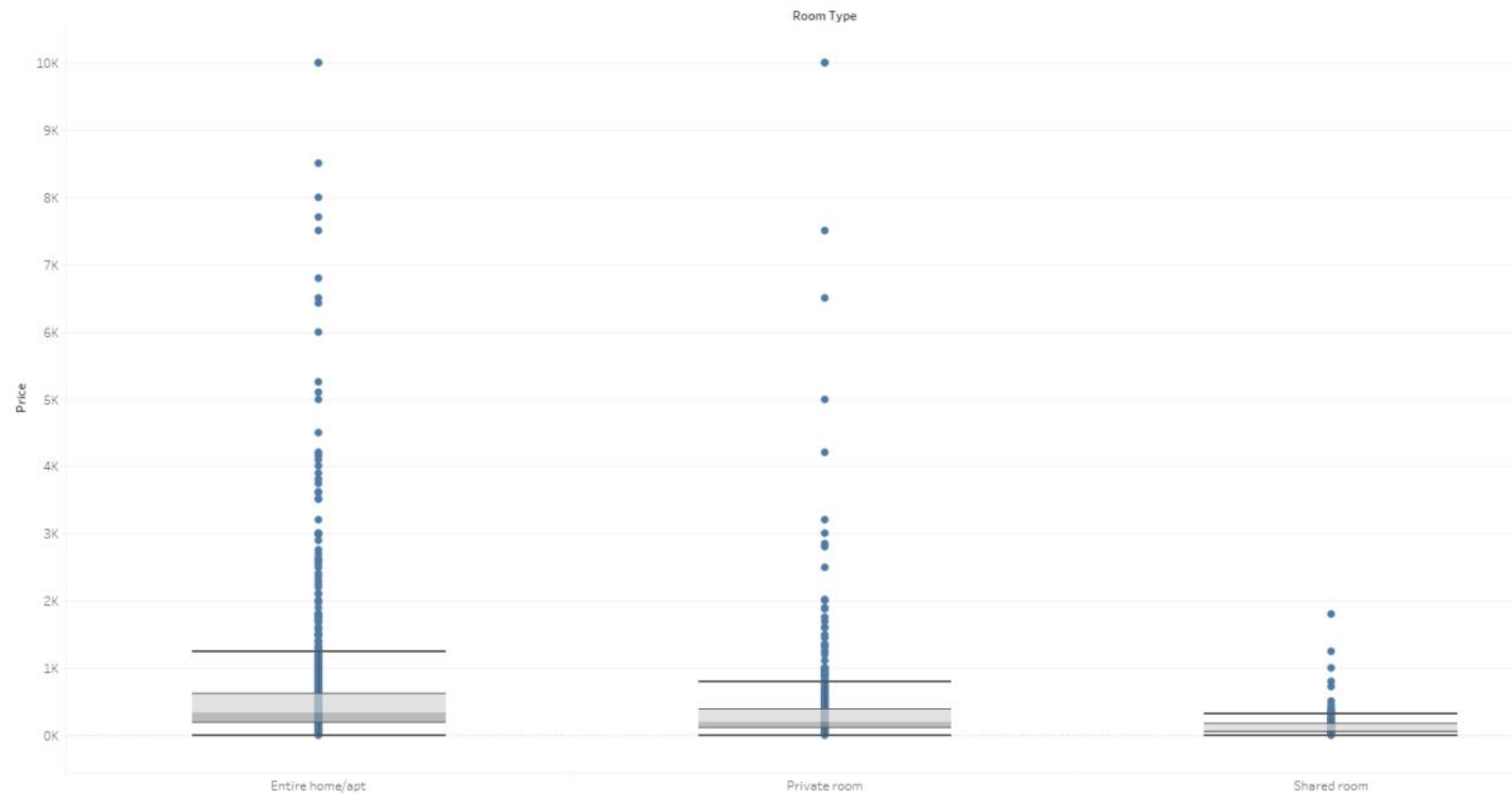# Data Methodology

By : Virgil Kalluvely
Shreyas Chiplunkar
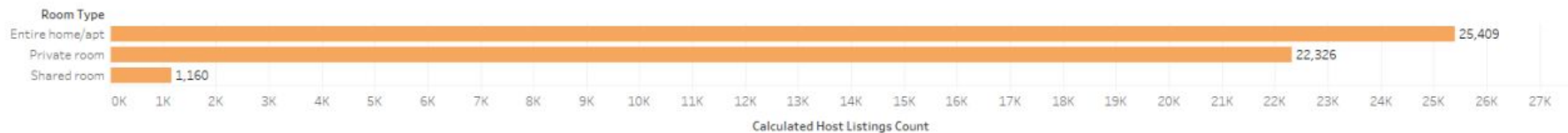& Arun Singh Chauhan

Step 1: Storyboarding

- Went through the data to get familiarized with it and noted down important fields.

- Made a mind map of the various slides of the presentation

- Made a rough template based on this mind map

Step 2: Data Wrangling

- Did univariate analysis using Tableau on the fields to see their distributions, the unique values in a field, the missing values and to check for outliers if any

- There was a small proportion of null values which would not affect my analysis so let them stay as it is

- Price was highly positively skewed so median was very close the lower quartile with some outliers as seen in the boxplot below

Price for each Room Type.

Room Type

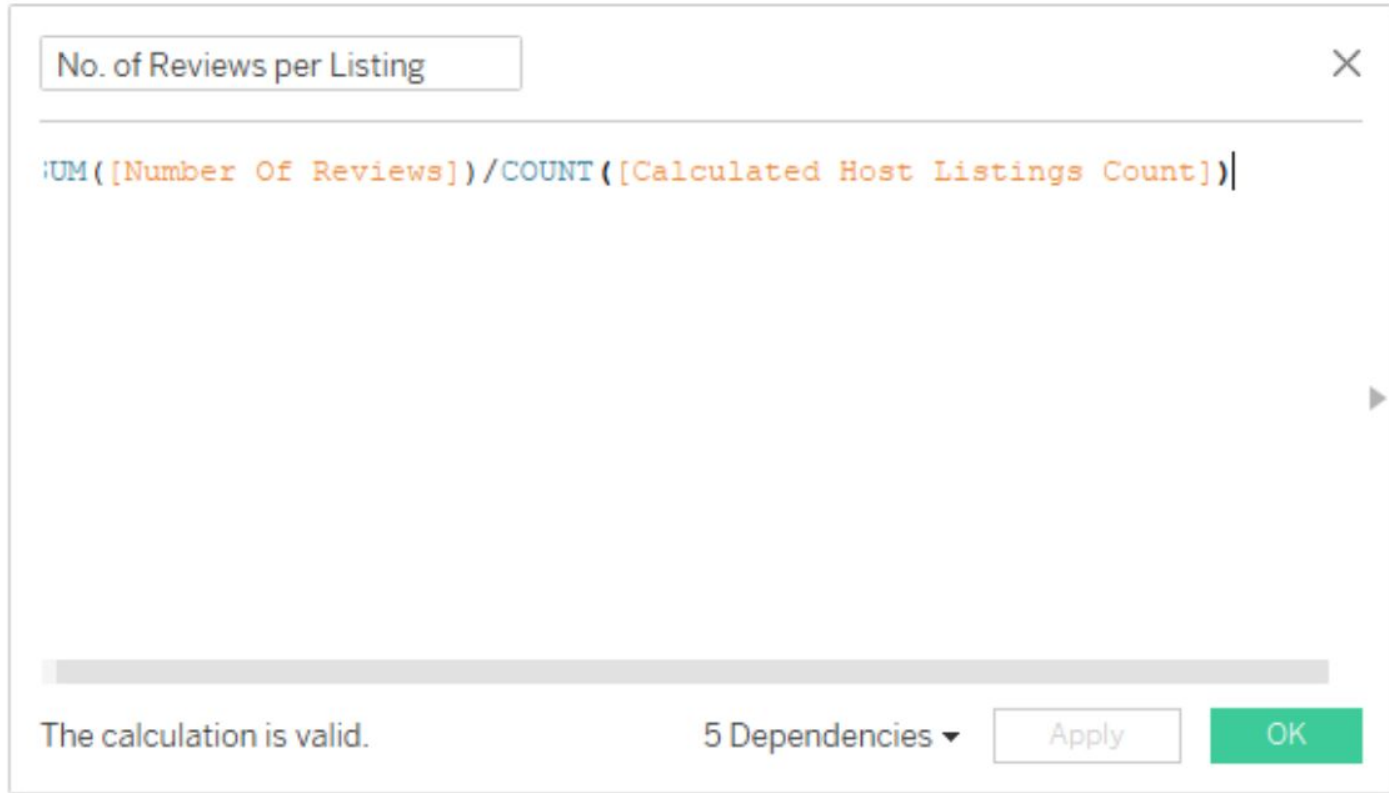| Room Type	Calculated Host Listings Count
Entire home/apt	25,409
Private room	22,326
Shared room	1,160

Calculated Host Listings Count

- Since price has outliers, used median instead of mean as the measure for price
- Host Listings count is maximum for entire apartment and private room and is very small for shared room as seen below

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

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
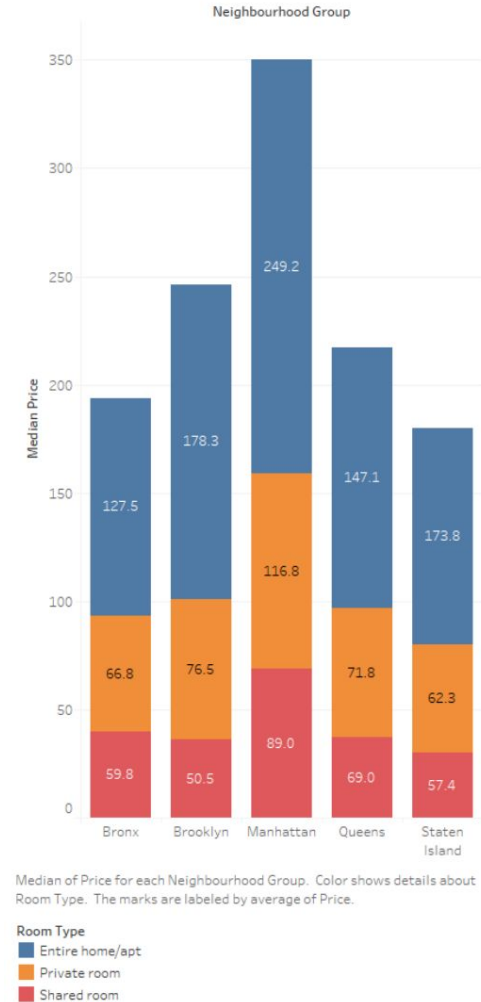
Step 3: Data Analysis

Checked neighbourhood grouped wise distribution of price and room type
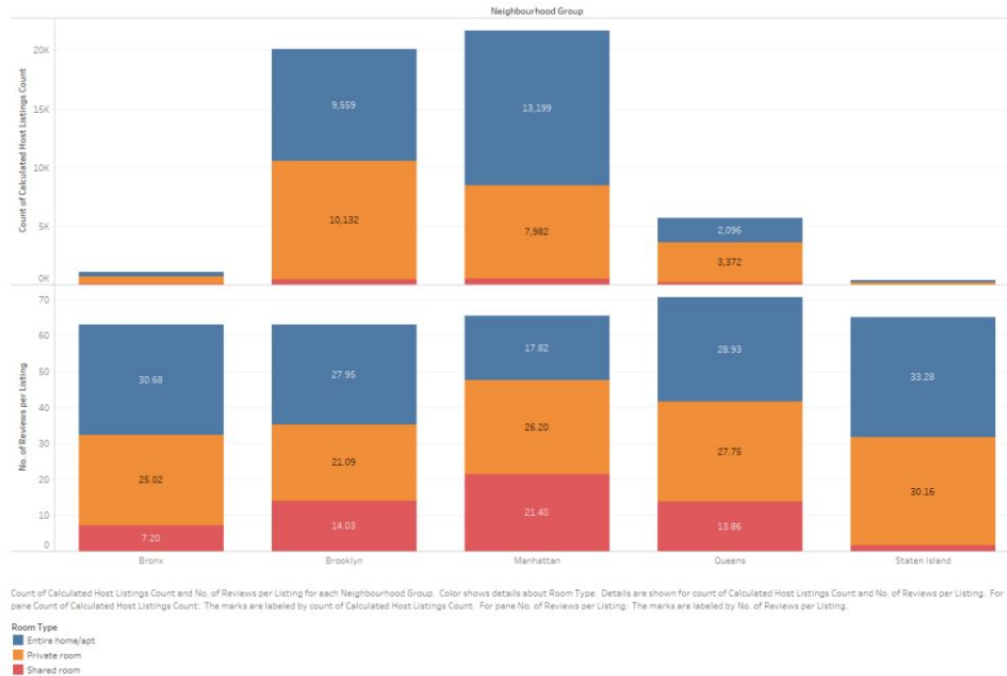
- Manhattan is costliest overall and the cheapest are
    1. Entire apt: Bronx
    2. Private room: Staten Islands
    3. Shared room: Brooklyn

- Found difference in prices between shared rooms,
  private rooms and extire apartments for each
  neighbourhood as shown beside.

Neighbourhood Group

Median of Price for each Neighbourhood Group. Color shows details about Room Type. The marks are labeled by average of Price.

Room Type
- Entire home/apt
- Private room
- Shared room

	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													

I4   fx =C3/C4-1

Through this clearly we can say that:

- 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

- Made the presentation adhering to best practices and pyramid principle

- Added recommendations for the respective departments