

Methodology Document

Airbnb Case Study

1. Data Source:

- a. The data used in the analysis was provided by Airbnb.
- b. The data contained 16 columns and 48,896 rows.

2. Data Cleaning:

- a. Removed rows with values such as "Not available", "No longer available", etc. in [name].
- b. Removed rows with [price] of 0, 9999, 10000.
- c. Removed rows with [minimum_nights] more than 365.
- d. Removed rows with [availability_365] of 0 or more than 365.
- e. Removed rows with [number_of_reviews] of 0 (considering them inactive).
- f. Brought [last_review] to one format (mm/dd/yyyy).

3. Data Assumptions:

New columns were derived based on the assumptions made:

- a. [years_inactive] assuming [last_review] the day of the last booking made in a listing (**=DATEDIF([last_review], 12/31/2019, "y")**).
- b. [price_bin] in order to categorize [price]:
 - i. Very Low (< \$50)
 - ii. Low (\$50-\$100)
 - iii. Medium (\$100-\$250)
 - iv. High (\$250-\$1000)
 - v. Very High (> \$1000)

```
IF [Price] < 50 THEN 'Very Low (<50)'  
ELSEIF [Price] >= 50 AND [Price] < 100 THEN 'Low (50-100)'  
ELSEIF [Price] >= 100 AND [Price] < 250 THEN 'Medium (100-250)'  
ELSEIF [Price] >= 250 AND [Price] < 1000 THEN 'High (250-1000)'  
ELSE 'Very High (>1000)'  
END
```

- c. [minimum_nights_bin] in order to categorize customers by the length of their stay:
 - i. Touristic Booking (1-8 nights) assuming that only tourists will stay in a property for such a short time.
 - ii. Monthly Rent (9-31 nights) assuming that people stay in such properties for living in New York short term.

- iii. Longer Stay (32-365 nights) assuming that people rent such properties for living in New York

```
IF [Minimum Nights] <= 8 THEN 'Touristic Booking'  
ELSEIF [Minimum Nights] > 8 AND [Minimum Nights] <= 31 THEN 'Monthly Rent'  
ELSE 'Longer Stay'  
END
```