

# AIRBNB Data Review

By Heryanto Jaya

Connect Me on [LinkedIn](#)



# Disclaimer

Data Source: [Kaggle](#)

Assumptions:

1. Data as per 31 October 2017 & short-term stay
2. Objective Customer Review Rate





1

## **Business Overview**

Airbnb connects hosts with guests for short-term & long-term stays



2

## **Problem Scope**

Increasing trend of low rated reviews (under 90 of 100) from 2012 – 2017 in some cities on USA



3

## **Factors of Low Rated**

Unit with bigger capacity & strict cancellation Policy tend to have lower Customer Rating Review Score



4

## **Pricing & Listings**

Averagely \$4.73 per night for 51.838 units



5

## **Recommendation**

Compensation on factors that boost customer satisfaction for units which tend to have low rating



# ***Executive Summary***



# Overview

Industry: Hospitality and Tourism

Founded on 2007 in San Francisco, California, United States of America

## Business Model

Airbnb operates as a platform-based business that earns revenue mainly through service fees from guests and hosts. It follows a B2C model by **connecting travelers with individual hosts**, but also incorporates B2B elements through corporate travel partnerships and its Airbnb for Work offering

### **Platform Model**

Airbnb operates as a digital platform connecting hosts and guests

### **Fee Based Income**

Charged to both Guest and Host

### **Diversity of Options**

Unique stays ranging from apartments to treehouses, accommodating various travelers' needs

### **Customer Relationships**

Focus on building community through reviews, ratings, and continuous support for both hosts and guests

## **Key Process**

Source: Airbnb (2019)



# **Business Problem**

# Project Scope & Analysis

This project analyze customer satisfaction for Airbnb services per accommodation based on what does the unit provides, host information, and policies implemented on unit



## Problem

What causes customer review score **rating less than 90** during 2012-2017?



## Objectives

Aim to have overall review score **rating is at least 90 on Q2 2018**



## Analysis Method

Using [Python](#), create **analysis customer satisfaction score and metrics**

**1**

## **Business Understanding**

Undertand the business  
to effectively adress the problem

**2**

## **Data Collection**

From listings in USA  
across 6 cities

**3**

## **Data Cleaning**

Using phyton to clear  
noises in dataset

**4**

## **Data Analysis**

Insight generations from  
keystats generated using  
Phyton (Google Collab)

**5**

## **Data Visualization**

Using Tableau to deliver  
better insights

**6**

## **Recommendation**

Advice to address  
regarding issue

# **Methodology**





# Focused Metrics

Possible Root Cause	Hypothesis	Metrics	Rationale
Facilities from Accommodation	Customer Dissatisfaction with listing from: <ul style="list-style-type: none"><li>• Less facilities and capacities</li><li>• Certain room type</li></ul>	<ul style="list-style-type: none"><li>• Accommodates</li><li>• # of Amenities</li><li>• Room Type</li></ul>	Those metrics reflect accommodation's value and facilities
Policies	Customer will be more satisfied if the accommodation impose less obligation	<ul style="list-style-type: none"><li>• Cancellation Policy</li><li>• Cleaning Fee</li></ul>	Those two metrics are obligation for customer
Host' Characteristic	Unclear information of Host verification and slow response rate will impair customer's trust on the listing. Impairment of customer's trust may drag the satisfaction rate.	<ul style="list-style-type: none"><li>• Host's Profile Pic &amp; ID Verified</li><li>• % Host Response Rate</li></ul>	<ol style="list-style-type: none"><li>1. Fully Verified mean Host has profile photo and ID verified by Airbnb</li><li>2. Will find the correlation of Response Rate with Cust Rating</li></ol>

# Segregation of Duties

Click link below to access appendix for

[Management Structure \(slide 32\)](#)

ROLE	PIC - DUTIES
Decider	Chief Executive Officer (CEO) - Final approval for Business Initiatives
Accountable	Chief Business Officer (CBO) - Data Analyst guider in technical role
Responsible	<ul style="list-style-type: none"><li>• Chief Technology Officer (CTO) - Ensures Data Analysts deliver accurate customer review tools and models</li><li>• Global Head of Operations (GHoO) - Gathers insights from ops/customer service and integrates customer feedback.</li></ul>
Consulted	<ul style="list-style-type: none"><li>• Chief Financial Officer (CFO) - Considers cost impact of operational insights and actions</li><li>• Global Head of Marketing (GHoM) - Ensures alignment with branding and user perception</li><li>• Chief Strategy Officer (CSO) - Assesses long-term strategic impact of host or policy changes</li></ul>
Informed	<ul style="list-style-type: none"><li>• Global Head Policy &amp; Communication (GHPC) - Stays informed if insights need public communication or PR handling</li><li>• Chief Legal Officer (CLO) - Must be informed of changes affecting legal compliance or host-guest policies</li></ul>



# Data Cleaning & Analysis

# DATASET

Steps	Before	After
Dropping Columns	36 Columns	19 Columns
Null & Duplicates	74097 Rows	52713 Rows
Data Formatting	Object	<ul style="list-style-type: none"><li>• Float</li><li>• Datetime</li></ul>

- Data cleaning [steps](#)
- Data dictionary on google [sheet](#)
- Data formatting [steps](#)

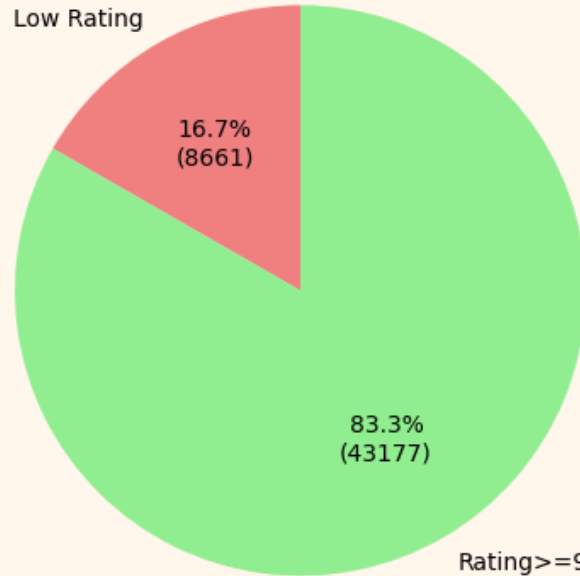
# EXPLORATORY DATA ANALYSIS

Factors - Host	Average Review Rate
Verified Host	94.31
Unverified Host	93.30
Factors - Room	Average Review Rate
Entire Room Type	94.23
Private Room Type	93.85
Shared Room Type	92.10
Factors - Host	Correlation Review Rate
Host Resp. Rate	0.07
Factors - Listing	Correlation Review Rate
No of Amenities	0.13
Accommodates	-0.03

- Customers tend to be **satisfied with** the listing that has **Verified Host (has verified Id & photos)** and **Entire Room Type**
- **Responsive Host** and **decent Provided Amenities boost** customer satisfaction rate
- While **more capacity (Accommodates)** in listing **dragging** satisfaction rate

## 16.7% of listings across 6 cities in USA from 2012 – 2017 are low rated

Percentage of Listings by Rating Category

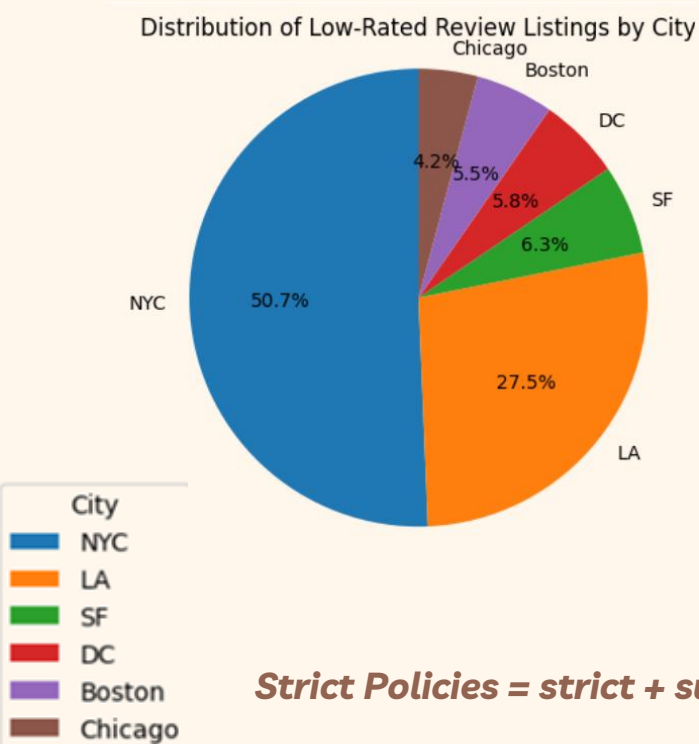


Rating >= 90

Total Listings: 51838

*Stats on customer review score  
please find on [slide 35](#)*

# NYC has the most Low-Rated Listings, which has 50.7% contribution from 8661 low-rated listings

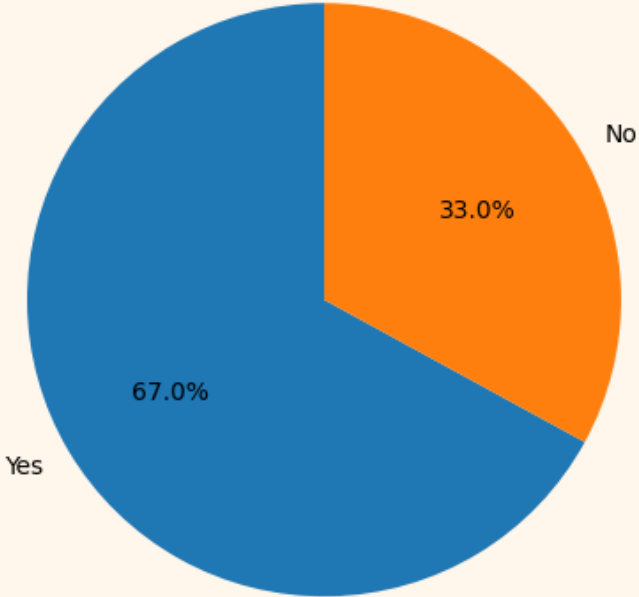


City	Average Resp Rate	Unverified Host Count	Average No Amenities	No Strict Policies	Total Listing
NYC	74%	7,171	17	11,403	22,824
LA	86%	3,739	18	7,416	15,454
SF	77%	1,094	19	2,033	4,490
DC	83%	873	19	1,421	3,638
Boston	85%	887	20	1,329	2,900
Chicago	92%	529	20	1,311	2,532
Total Listing from 6 Cities					51,838

*Strict Policies = strict + super\_strict\_30 + super\_strict\_60* (Source of Airbnb [Policies](#))

# 33% of low-rated listings have unverified hosts, suggesting host verification may impact customer satisfaction

Distribution of Fully Verified Hosts in Low-Rated Listings

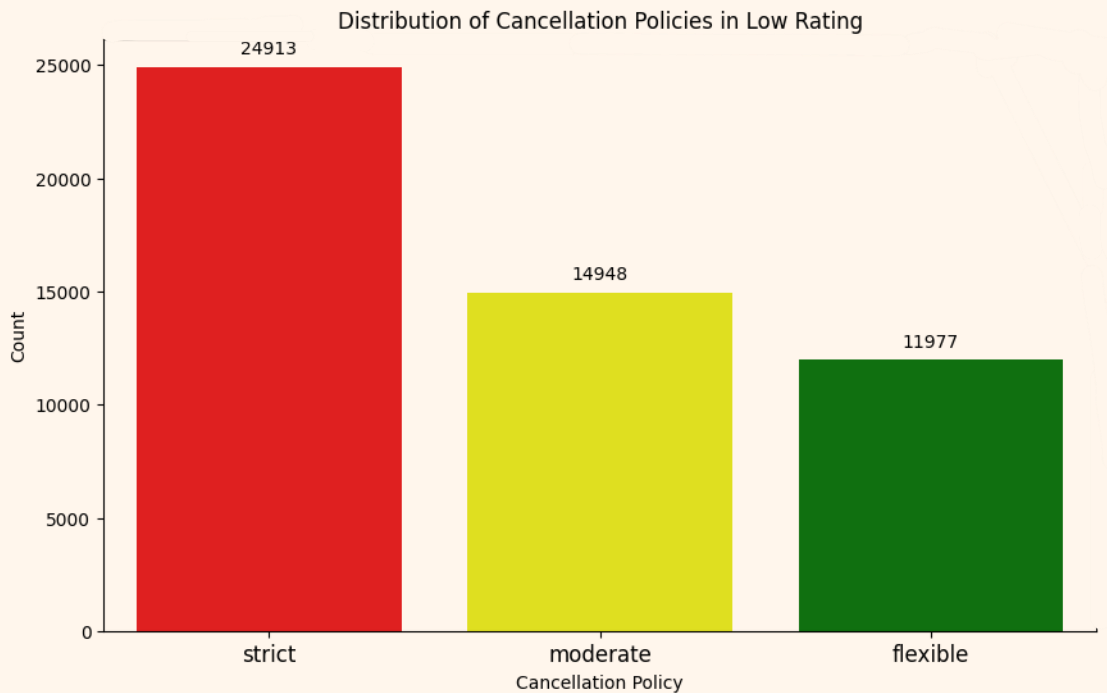




## AVERAGE RESPONSE RATE ACROSS LISTING CATEGORIES

Listing Category	Average Host Response Rate
Low Rated	74%
Overall	79%
Rating >= 90	81%

# Strict cancellation policies are more common in low-rated listings, hinting at a link between inflexibility and poor customer experience



This statistics, support the correlation before, higher capacity, lower the average review rating

### Average Capacity per Room Type



Entire Unit

**4.25**

Low Rated Listing

---

**3.92**

Rated  $\geq 90$



Entire Room

**2.12**

Low Rated Listing

---

**2.01**

Rated  $\geq 90$



Shared Room

**2.35**

Low Rated Listing

---

**1.94**

Rated  $\geq 90$

More amenities lead to higher customer satisfaction, support that both satisfaction review score and number of amenities are causative related

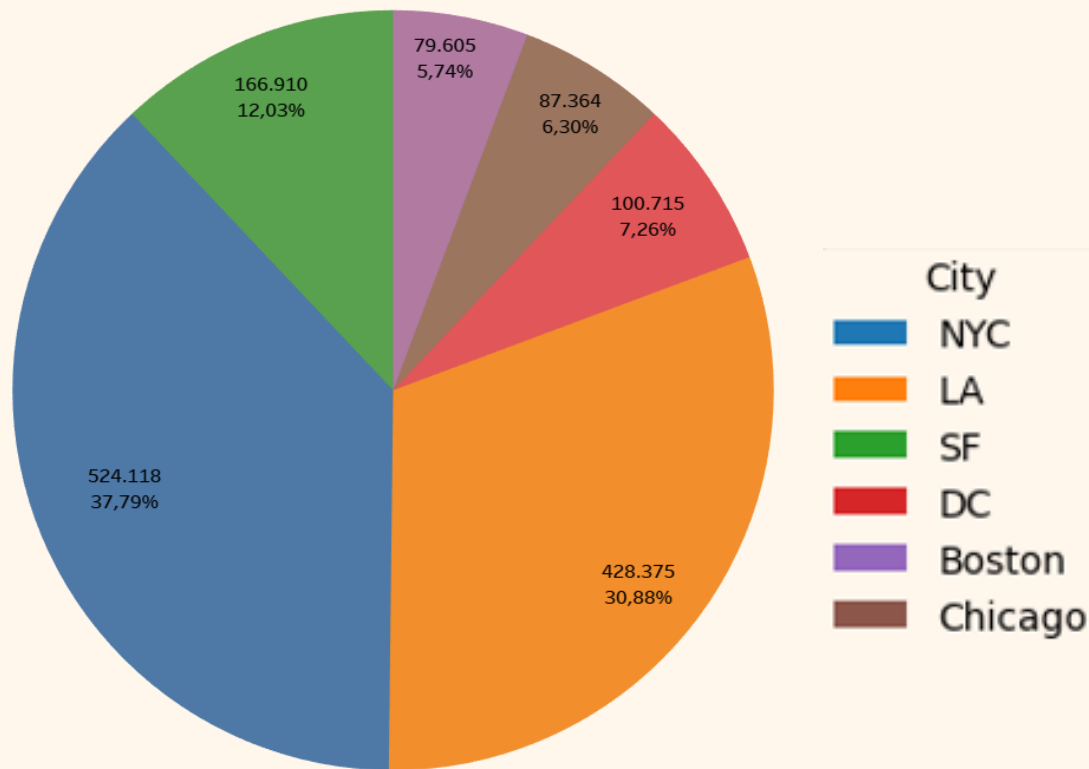
Listing Category	Average No of Amenities
Low Rated	16
Overall	18
Rating >= 90	19



# Data Visualization

Using [tableau](#)

# NYC has the most active guests who are giving customer review rate (524118 reviews)



# NYC has the most Airbnb listings across all room types, followed by LA

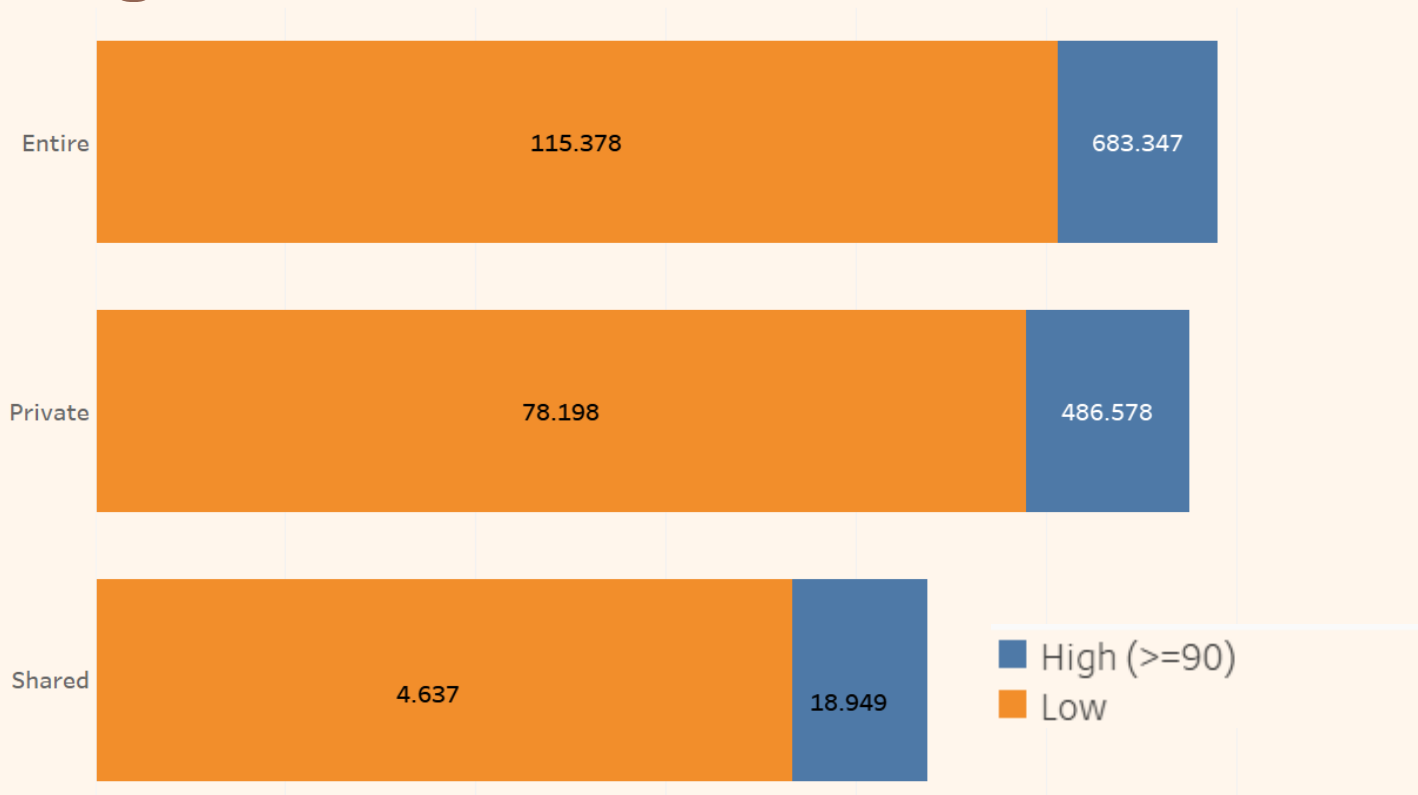
City	Entire Unit	Private Room	Shared Room
<b>Boston</b>	1,567	932	33
<b>Chicago</b>	1,744	1,091	65
<b>DC</b>	2,446	1,112	80
<b>LA</b>	9,433	5,569	452
<b>NYC</b>	11,749	10,549	526
<b>SF</b>	2,655	1,780	55

# Boston's shared rooms have the lowest rating, while NYC shows generally lower ratings across all types despite high listings

	Entire	Private	Shared
Boston	93,858	93,215	88,030
Chicago	94,958	95,297	94,077
DC	94,999	94,566	91,388
LA	94,234	94,570	91,936
NYC	93,784	93,210	91,998
SF	95,254	94,406	95,509



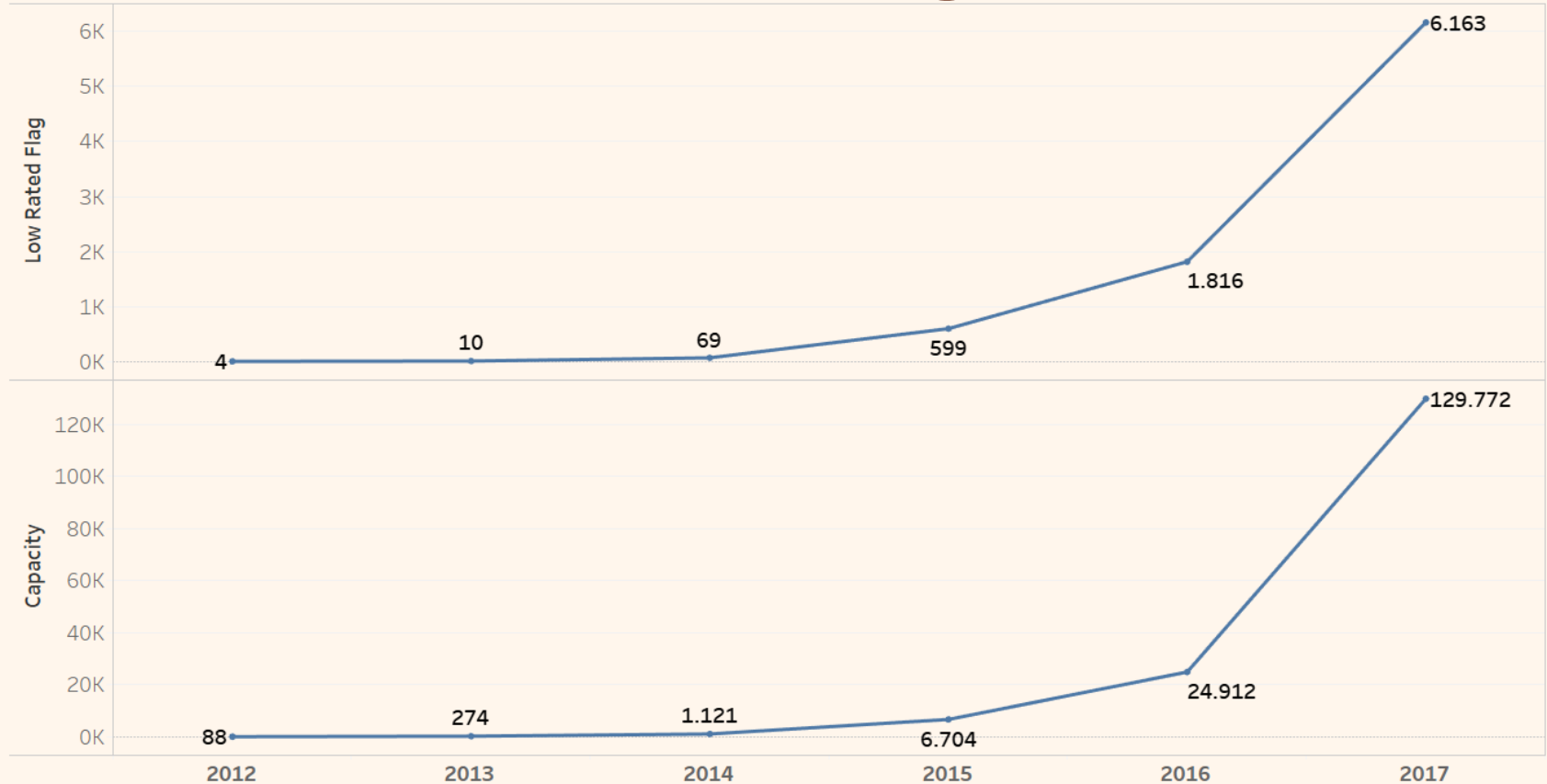
# Entire Unit Type Has the biggest Number Listings with Low Rated Review Score



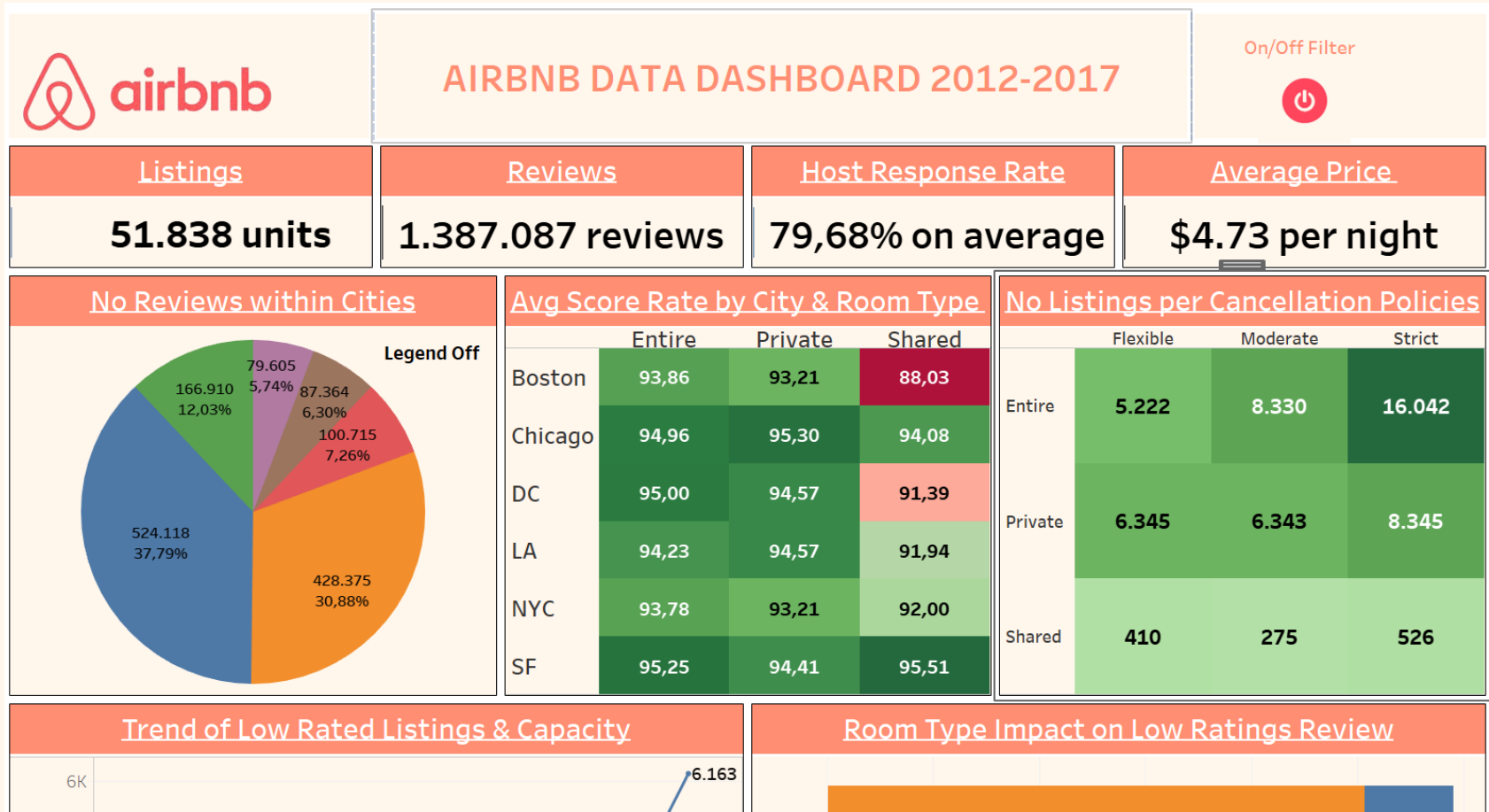
# Entire Unit Type has the most Number of Strict Policy Implementations

	Flexible	Moderate	Strict
Entire	5.222	8.330	16.042
Private	6.345	6.343	8.345
Shared	410	275	526

# Increasing Trend of Low Rated Listing consistent with Increasing Accomodates



# Full Dashboard on Tableau Public



# Insight Summary



## *Host Verification & Responsiveness Drive Satisfaction*

Listings with **verified hosts** and **responsive communication** consistently receive higher customer ratings.



## *Strict Policies & High-Capacity Hurt Reviews*

Units with **strict cancellation policies** and **higher guest capacity** tend to get lower review scores, indicating these are potential pain points



## *More Amenities = Happier Guests*

A **positive correlation** exists between the **number of amenities provided** and **customer satisfaction**, suggesting added value leads to better experiences.

Context	Recommendations
Cities	<u>NYC</u> <ul style="list-style-type: none"><li>• <b>Targeted host education and quality control</b> for low-rated listings</li><li>• <b>Incentives or support</b> for highly rated hosts to maintain standards</li></ul> <u>Boston</u> <ul style="list-style-type: none"><li>• <b>Investigate</b> shared quality room</li></ul>
Host	<ul style="list-style-type: none"><li>• <b>Strengthen Host Verification</b> by implementing stricter verification measures on identity and property</li><li>• <b>Improve Host Responsiveness</b> by introducing incentives for quick responses, automated messaging tools, or host training</li></ul>
Policies	<b>Adjust cancellation rules</b> based on <b>demand, listing type, and host reliability</b> to balance flexibility and protection.
Room Type	<ul style="list-style-type: none"><li>• <b>Reassess strict</b> cancellation policies in <b>Entire Unit</b> listings</li><li>• <b>Targeted quality assurance</b> such as <b>Cleanliness &amp; Enhanced Host Support</b></li></ul>
Capacities	<ul style="list-style-type: none"><li>• <b>Evaluate the impact</b> of larger accommodations on guest experience</li><li>• <b>Enhance service quality</b> for high-capacity listings</li></ul>
Amenities	<ul style="list-style-type: none"><li>• <b>Promote key amenities</b> like Wi-Fi and kitchen essentials</li><li>• <b>Give hosts data-driven insights on amenities that improve ratings</b></li></ul>



***Thanks***



# Appendix



# Airbnb Management Structure

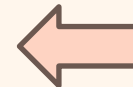


As of April 6, 2025, Airbnb's executive management team is structured as follows:

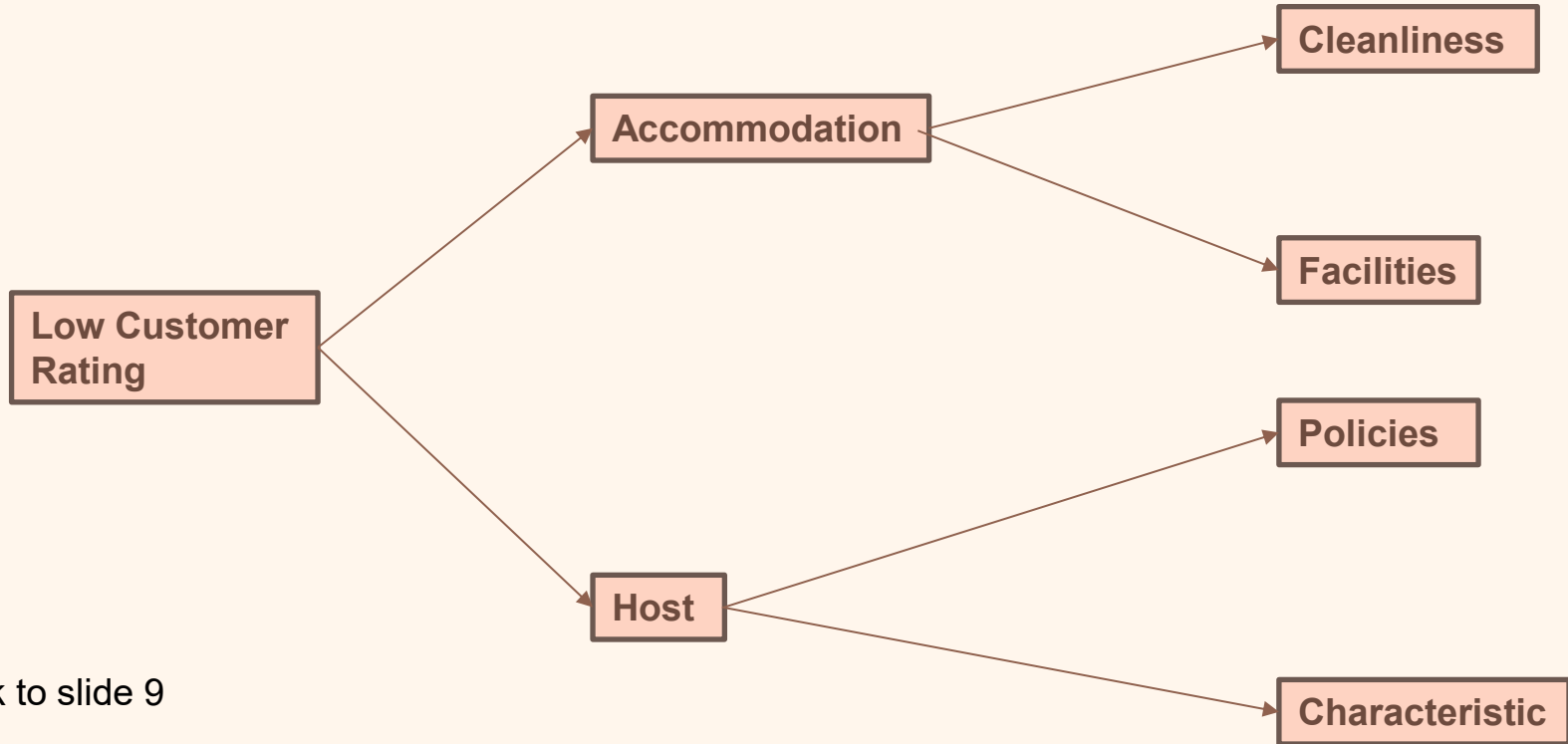
- **Brian Chesky:** Co-founder and Chief Executive Officer (CEO).
- **Dave Stephenson:** Chief Business Officer (CBO).
- **Ellie Mertz:** Chief Financial Officer (CFO).
- **Ari Balogh:** Chief Technology Officer (CTO).
- **Hiroki Asai:** Global Head of Marketing.
- **Tara Bunch:** Global Head of Operations.
- **Jay Carney:** Global Head of Policy and Communications.
- **Ronald A. Klain:** Chief Legal Officer

Source [1](#), [2](#)

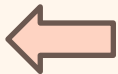
Back to slide 10



# Possible Root Cause



Back to slide 9

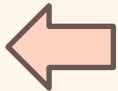


# Problem Prioritization



Possible Root Cause	Hypothesis	Priority
Cleanliness	Dirty accommodation obviously drive away customer's satisfaction	Low (Relative Opinion)
Facilities from Accommodation	Less Facilities and capacities, certain room type, and high price per night will drive away customer satisfaction score.	High
Policies	Customer will be more satisfied if the accommodation impose flexible policies	High
Host' Characteristic	Unclear information of Host verification and slow response rate will impair customer's trust on the listing. Impairment of customer's trust may drag the satisfaction rate.	High

Back to slide 9



# STATISTICAL ANALYSIS



This analysis focus on Customer Rating Score, which have the details below

`review_scores_rating`

**count** 51838.00

**mean** 94.03

**std** 7.87

**min** 20.00

**25%** 92.00

**50%** 96.00

**75%** 100.00

**max** 100.00

[Back to slide 14](#)

`dtype: float64`

# DATA CLEANING – DROPPING COLUMN



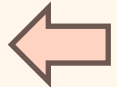
The columns ***first\_review, name, description, host\_since, thumbnail\_url, instant\_bookable, latitude, longitude, zipcode, Neighbourhood and Unnamed: 29 to Unnamed: 35*** were dropped from the dataset to focus on analysis

***No Duplicated Data*** found in ***Listing ID***, which must be unique

***Null data in 'last\_review', 'number\_of\_reviews'*** are initially ***removed*** as Review Score will be invalid if both of them are 0, while null in ***'bathrooms', 'bedrooms' and 'beds'*** are ***removed*** subsequently

However, ***null data*** in ***Host Verification dan Response Rate*** can be ***imputed by 0***

Back to slide 12



# DATA ADJUSTMENT



Adjusting data type for:

## ***Host Response Rate (from Object to Float):***

```
df['host_response_rate'] =  
df['host_response_rate'].str.rstrip('%').astype('float64') / 100
```

## ***Last Review (from Object to Datetime):***

```
df['host_response_rate'] =  
df['host_response_rate'].str.rstrip('%').astype('float64') / 100
```

Back

