

# LA Airbnb Analysis

---

Peike Li, Franklin Truong, Jun Ryu, Khang Thai,  
Audrey Huang, Kiv Atmojo





## TABLE OF CONTENTS



**01**

### **HOST ANALYSIS**

ML



**02**

### **DASHBOARD**

VISUALIZATION



**03**

### **LISTING ANALYSIS**

NLP

# DATASET

**Data overview:** Our data is from **Inside Airbnb**, a mission driven project that provides data and advocacy about Airbnb's impact on residential communities. Specifically, we used data for Los Angeles only.

**Data structure:** Our datasets are in .csv format, and we mainly work with two datasets (one about listings and the other containing comments/reviews on the listings).

The dimensions of the two dataframes:

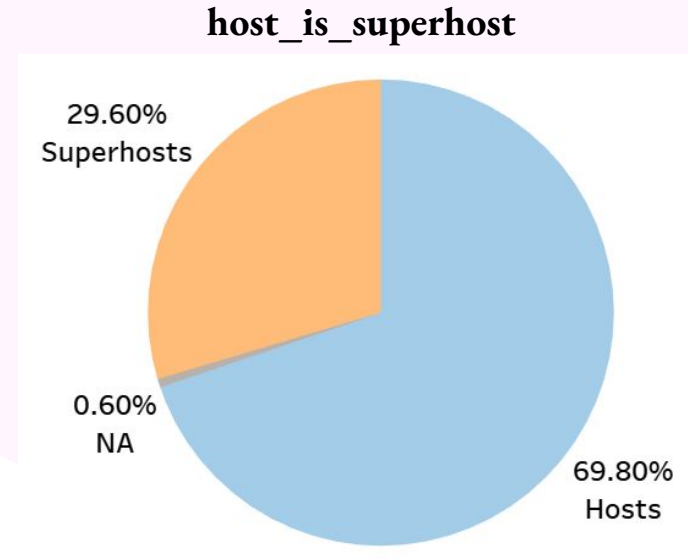
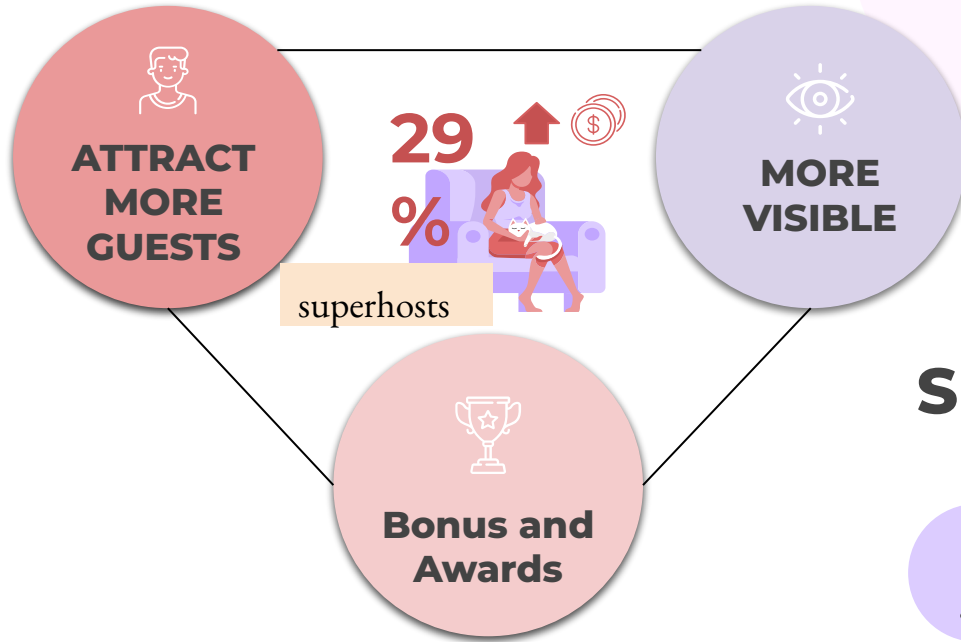
Listings - 45,591 rows

Reviews - 55,240 reviews

## Key variables

Variable Name	Variable Type
host_is_superhost	boolean
host_response_time	numeric
host_response_rate	numeric
host_acceptance_rate	numeric
ratings	numeric
price	numeric
days_since_host	numeric
host_has_profile_pic	boolean
longitude & latitude	numeric
comments	text

# Host Analysis



## Superhosts = More Profit



What **suggestions** can we give to Airbnb hosts to help them **become superhosts**?

# Logistic Regression

Possible predictors that might correlate to whether a host is superhost:

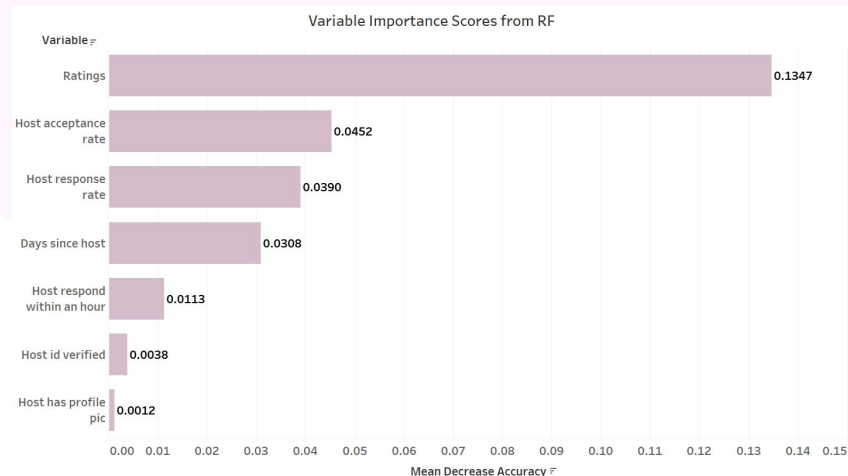
- Response rate
- Acceptance rate
- Has a profile picture
- Has a verified identity
- Days since host
- Response time
- Ratings
- Listing count



Variable	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-29.80	0.74	-40.02	0.00
host_response_rate	0.05	0.00	12.75	0.00
host_acceptance_rate	0.03	0.00	21.69	0.00
host_has_profile_picTRUE	0.82	0.15	5.52	0.00
host_identity_verifiedTRUE	-0.01	0.05	-0.18	0.86
ratings	4.55	0.10	45.67	0.00
days_since_host	0.00	0.00	13.86	0.00
host_response_time_within_an_hour	0.16	0.56	0.28	0.78
host_response_time_within_a_day	-0.42	0.55	-0.76	0.44
calculated_host_listings_count_entire_homes	0.06	0.00	12.99	0.00
calculated_host_listings_count	-0.07	0.01	-12.93	0.00
host_response_time_within_a_few_hours	-0.12	0.55	-0.22	0.83
host_total_listings_count	0.00	0.00	-7.50	0.00

**74%**  
test accuracy

# Random Forest



**76%** test accuracy

# Insights and Strategies for Hosts

1

## **Maintain high ratings for listings**

How?  
See later analysis on listings w/ bad ratings

2

## **Maintain high acceptance rates**

Specify requirements  
+  
Use instant bookable  
+  
Update calendar

3

## **Maintain high response rates**

app notification  
+  
automated messages  
+  
Pre-written responses

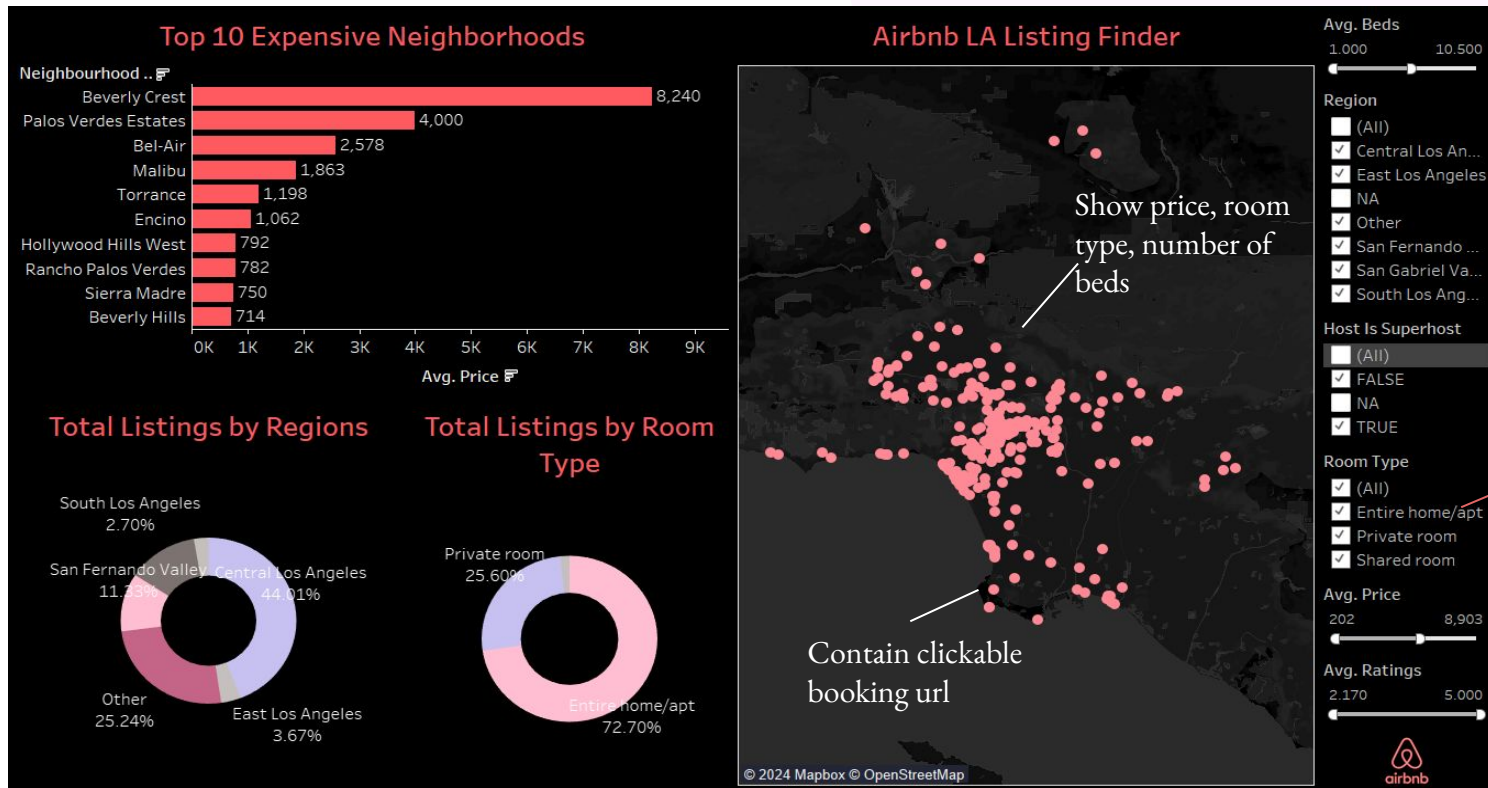
4

## **Keep on hosting**

The longer you host, the more opportunity for

- More reviews
- More occupancy
- Demonstration of hospitality

# Dashboard for Airbnb guests

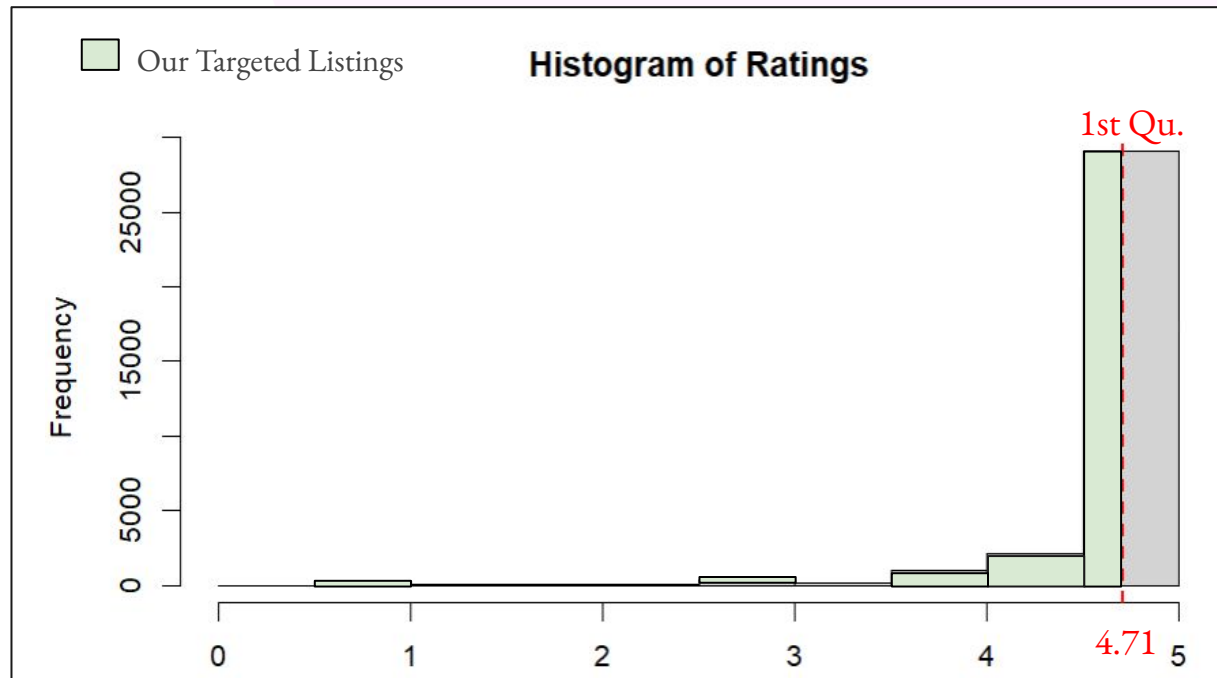


Scan for an interactive dashboard!

Filters to help guests find the right listing

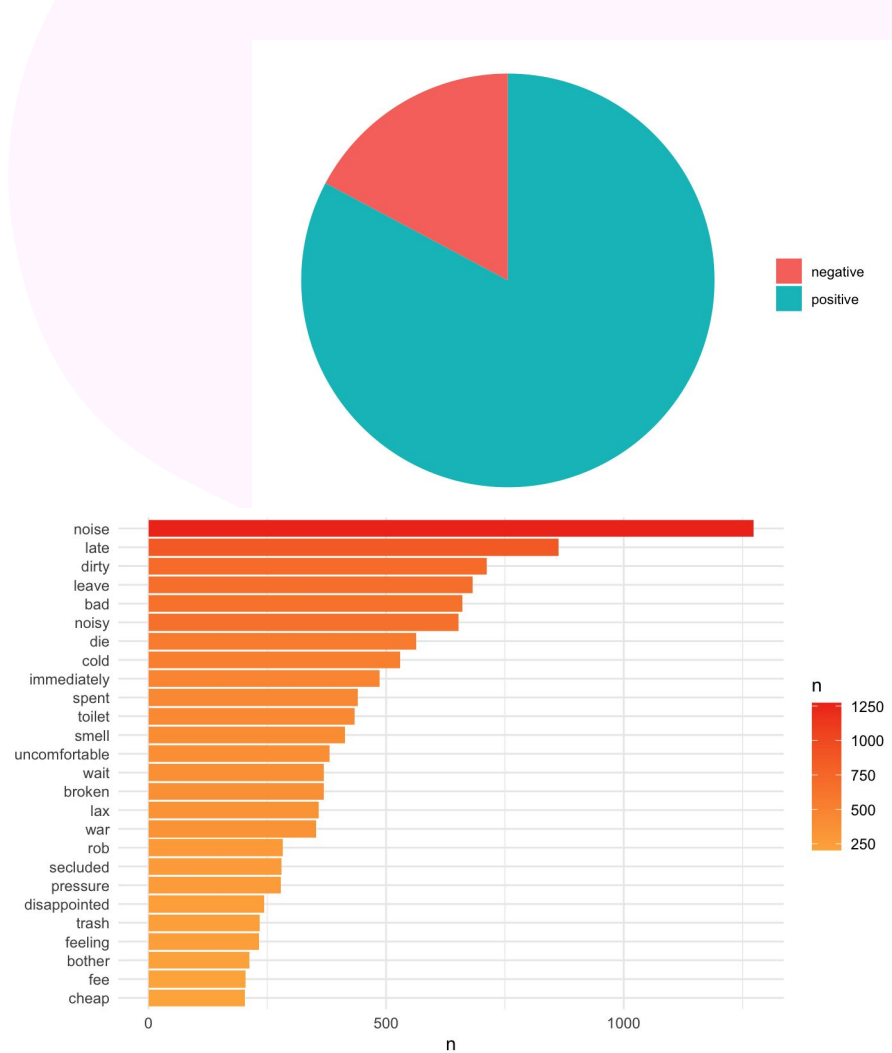
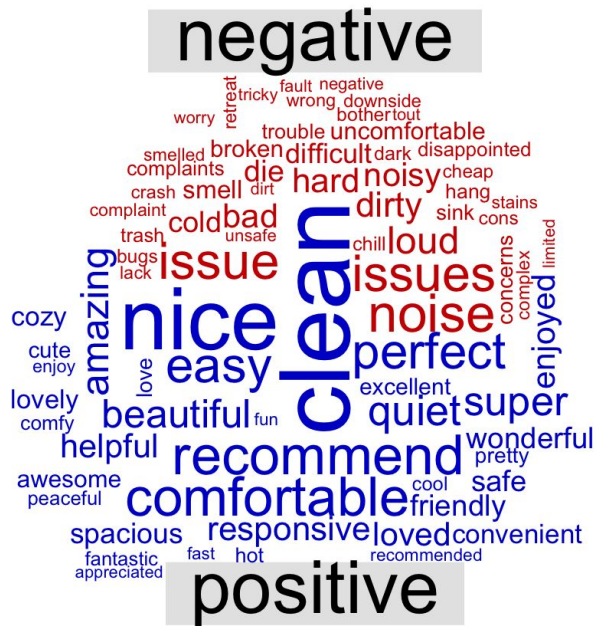
# Sentiment Analysis

Analyze  
Negative Reviews  
On  
Low-rated Listings





# Basic Analysis



# Topic Modeling - LDA

