



# Airbnb's Growth



**6,000,000 +**

Airbnb's active  
listings



**28 Days +**

Long-term stays are  
fastest growing  
category



**23%**

Increased in supply  
of properties

Credit:

<https://investors.airbnb.com/home/default.aspx>

# Listing Example



Start your search



Become a Host

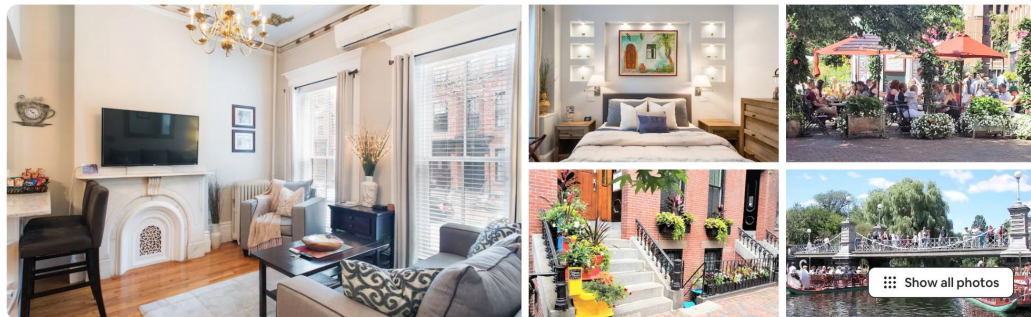


## Astonishing 1 Bed South End Boston's Best Location

★ 4.99 · 160 reviews

Superhost · Boston, Massachusetts, United States

Share Save



Show all photos

Entire rental unit hosted by [REDACTED]

2 guests · 1 bedroom · 2 beds · 1 bath



\$330 night

★ 4.99 · 160 reviews

## About this space

Clean, unique and inviting, 1-bedroom on the 1st floor -parlor with astonishing ceilings, new tall custom windows, new SS appliances and the best of Boston outside your door.

Located ½ a block to Restaurant Row, where you'll find some of the best dining in the city. Famous bakeries, coffee shops are all near. It's a super effortless walk to the very BEST Boston has to offer.

Complimentary Netflix, movie channels, high speed internet, Keurig coffee, snacks. It doesn't get any better than this!

### The space

This well appointed apartment is located on a quiet residential street in the desirable South End with its charming Brownstones, brick sidewalks and tree lined streets. This apartment is on the



Caleb

July 2022

This place is amazing! Stay here!!!



Alyssa

June 2022

This is your sign to book this place!! No need to keep scrolling. Stephanie and Alex's place is perfect. If you're looking for a beautiful area in Boston with a ton of great food options, ver...

[Show more](#) >



Rosa

June 2022

Great time had in Boston! The south end has a lovely charm. Walking distance to shops and quaint restaurants! Stephanie and Alex have a charming place equipped with a full kitchen,...

[Show more](#) >

# New Listing Example

## Sunny private home w/ patio - very convenient!

[Boston, Massachusetts, United States](#)

[Share](#) [Save](#)



Entire condo hosted by [REDACTED]

4 guests · 1 bedroom · 2 beds · 1 bath



\$243 \$207 night

## About this space

You will love staying in vibrant South Boston! This neighborhood is the setting for countless movies and is conveniently host to the Seaport District, Convention Center, and incredible bars and restaurants - many of which are <10 mins walking away! Easily walk to the Red Line subway to reach downtown Boston in minutes and explore the historic city. Once you are back on Telegraph Hill you can relax on the private patio with your favorite beverage!

Pictures from

[https://www.airbnb.com/rooms/689981885807910567?adults=1&children=0&infants=0&check\\_in=2022-09-26&check\\_out=2022-10-03&federated\\_search\\_id=1b80dde1-fda6-4bdc-be5d-e9fc08573758&source\\_impression\\_id=p3\\_1660432800\\_UybjoPHcHr4BsVTZ](https://www.airbnb.com/rooms/689981885807910567?adults=1&children=0&infants=0&check_in=2022-09-26&check_out=2022-10-03&federated_search_id=1b80dde1-fda6-4bdc-be5d-e9fc08573758&source_impression_id=p3_1660432800_UybjoPHcHr4BsVTZ)

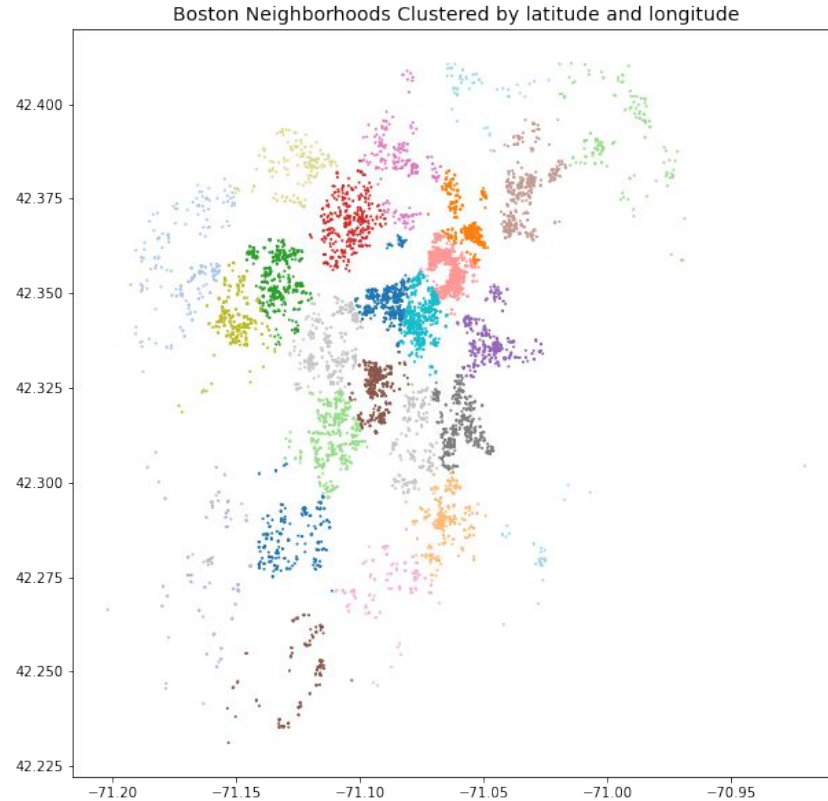
# Problem



Is it possible to predict what new listings on Airbnb would be rated (despite not having reviews/ratings history for reference) to help vacationers make more informed decisions about their stays?



# Boston Airbnb Listings



# Text Data Types



July 2022

Just a warning for anyone not familiar with the Boston area like I wasn't. The listing says free street parking, but that's extremely deceptive. Any street spot close is 2 hours max, if you can find somewhere. It is free between 8pm and 8 am, but you'd better have some other plan during the day.

And if there is more than one person or a couple coming, you should definitely bring an air mattress. There are two couches in the living area, one of which is so broken down it appears to have been found on the side of the road in front of a frat house, and a not terribly comfortable futon.

## User Reviews

## About this space

A one bedroom mezzanine apartment right next to Airport and only a 7min Uber ride from Downtown. This is perfect for a couple or someone who's just looking to save some money on their stay. Hair shampoo conditioner and body wash provided. Clean towels along with a hair dryer, iron and ironing board. Smart TV in the living room/kitchenette comes with Roku, YouTube and Netflix.

### The space

This is a mezzanine apartment so it has a private lit entrance to the back of the house which is shown in the pictures.

## Host Descriptions



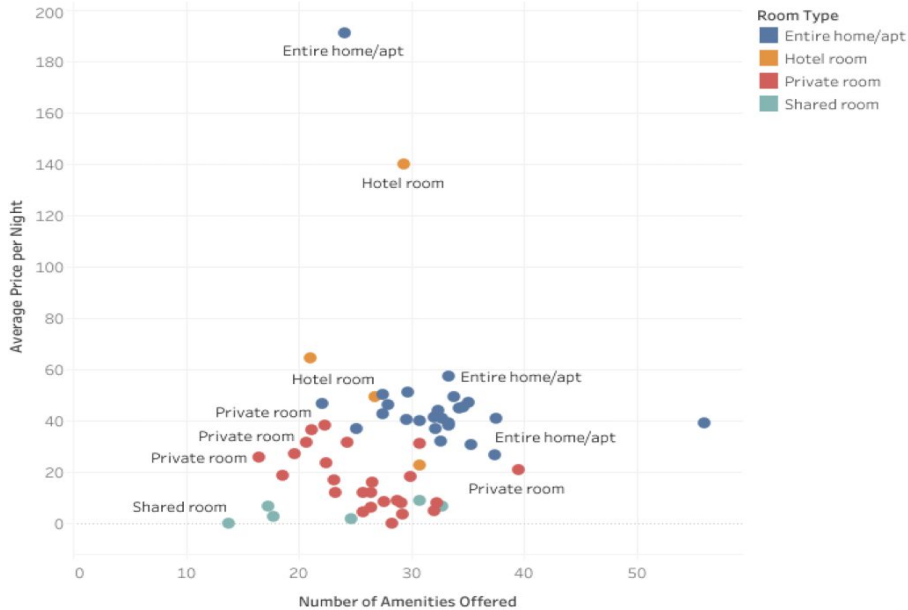
Pictures from:

[https://www.airbnb.com/rooms/23948295?adults=1&children=0&infants=0&check\\_in=2022-09-30&check\\_out=2022-10-07&federated\\_search\\_id=1b80dde1-fda6-4bdc-be5d-e9fc08573758&source\\_impression\\_id=p3\\_1660495387\\_TarpDqtoGo%2BACFIJF](https://www.airbnb.com/rooms/23948295?adults=1&children=0&infants=0&check_in=2022-09-30&check_out=2022-10-07&federated_search_id=1b80dde1-fda6-4bdc-be5d-e9fc08573758&source_impression_id=p3_1660495387_TarpDqtoGo%2BACFIJF)

# Categorical/Numerical Data Types

## Listing Features

Price vs. Amenities



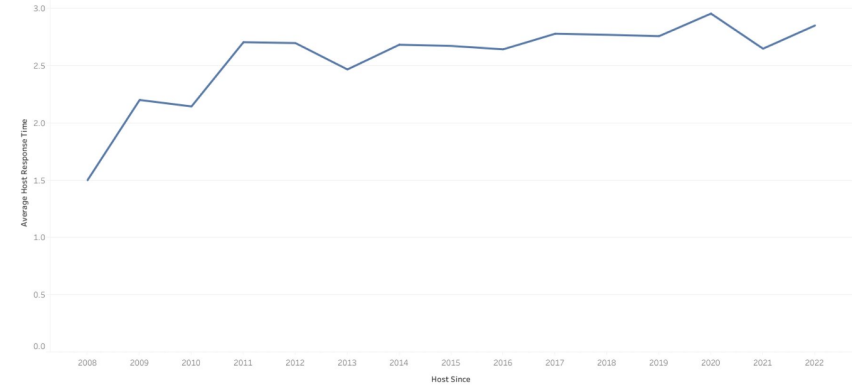
### What this place offers

- Wifi
- TV with Roku, Netflix
- Patio or balcony
- Hair dryer
- Microwave
- Free street parking
- Air conditioning
- Luggage dropoff allowed
- Refrigerator
- Long term stays allowed

Show all 27 amenities

~~\$135~~ \$122 night

Host Response Time vs. Year Started



The trend of average host response time as a function of when the host started on Airbnb.

For Response Times:  
3 indicates the host responds within an hour  
2 indicates within a few hours  
1 indicates within a day  
0 indicates within a few days or more



# NLP Modeling

# Cleaning and EDA on Reviews Data

● **230,000 Reviews (dating back to September 2021)**

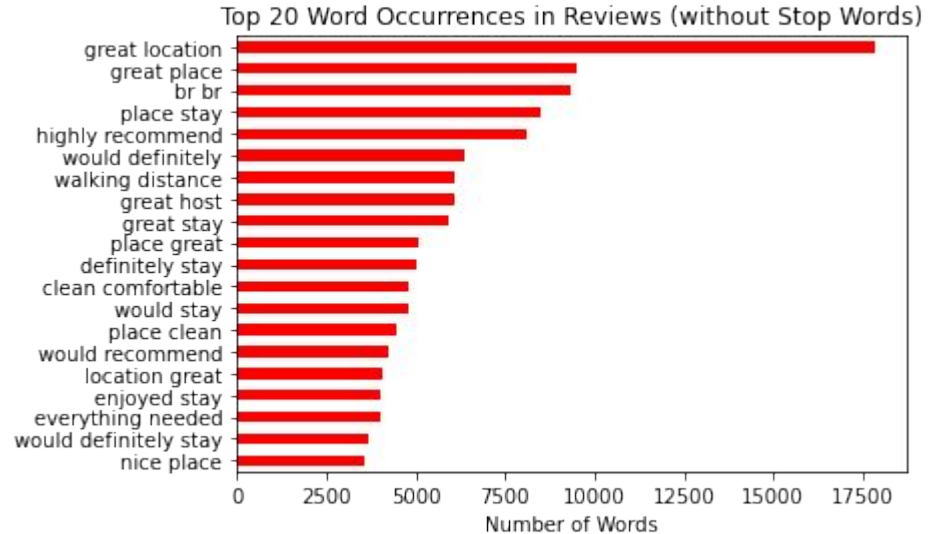
● **208,000 distinct reviewers**

- Top reviewer had 55 reviews

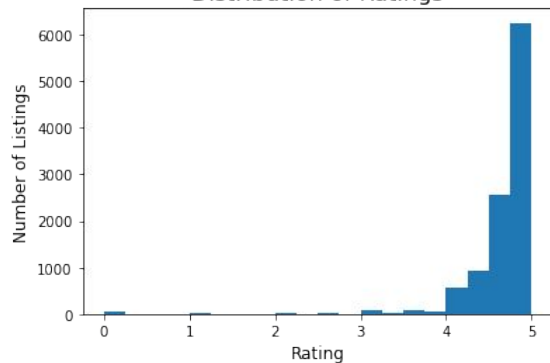
● **Across 6,700 distinct listings**

- Most reviewed listing had 1,071 reviews

● **Average word count = 42**  
**Median word count = 29**



Distribution of Ratings



# Reviews NLP Modeling

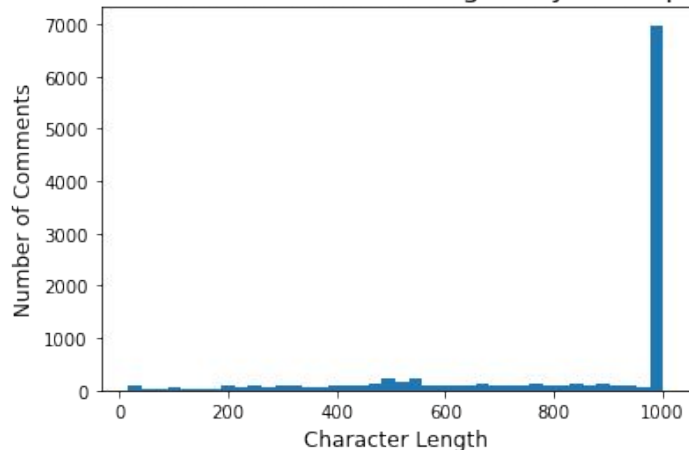
**Baseline Accuracy = 41.80%**

Transformer	Classifier	
CountVectorizer	Multinomial Naive Bayes	
	Train: 0.58	Test: 0.54
	Multinomial Logistic Regression	
	Train: 0.62	Test: 0.56
	Random Forest	
	Train: 0.91	Test: 0.47
	Ada Boosted	
	Train: 0.93	Test: 0.38

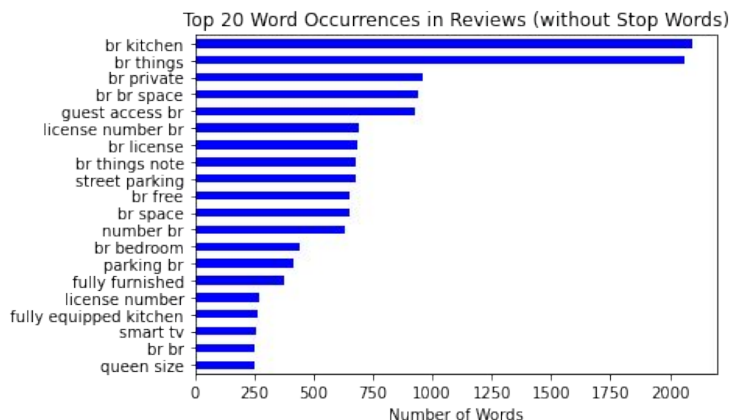
Transformer	Classifier	
Tf-IDF Vectorizer	Multinomial Naive Bayes	
	Train: 0.52	Test: 0.49
	Multinomial Logistic Regression	
	Train: 0.62	Test: 0.57
	Random Forest	
	Train: 0.97	Test: 0.47
Grid Searching	TF-IDF	Logistic Regression
	Train: 0.57	Test: 0.57

# Cleaning and EDA on Descriptions Data

Distribution of Character Lengths by Description



- **Average word count = 137 (compared to 42 for reviews)**  
**Median word count = 156 (compared to 29 for reviews)**
- **10,000 Host Descriptions (dating back to September 2021)**



# Descriptions NLP Modeling

**Baseline Accuracy = 25.44%**

Transformer	Classifier	
CountVectorizer	Multinomial Naive Bayes	
	Train: 0.66	Test: 0.57
	Multinomial Logistic Regression	
	Train: 0.88	Test: 0.75
	Random Forest	
	Train: 0.92	Test: 0.75
	Ada Boosted	
	Train: 0.90	Test: 0.76

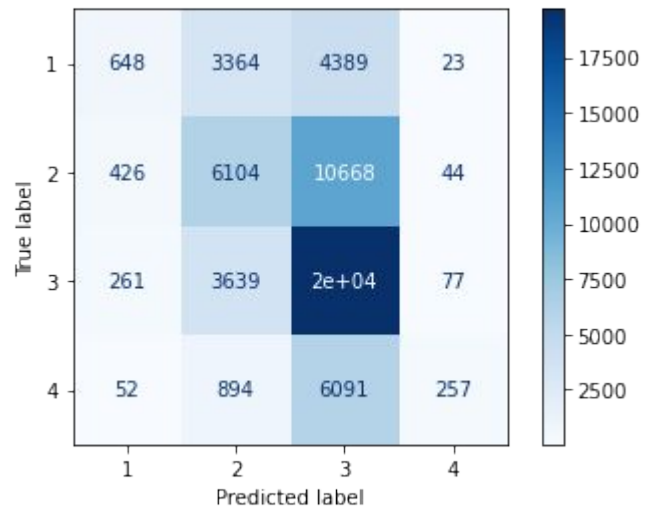
Transformer	Classifier	
TF-IDF Vectorizer	Multinomial Naive Bayes	
	Train: 0.68	Test: 0.58
	Multinomial Logistic Regression	
	Train: 0.77	Test: 0.65
	Random Forest	
	Train: 0.92	Test: 0.78
Grid Searching	Tf-IDF	Random Forest
	Train: 0.92	Test: 0.784

**\*Specificity score of .79**

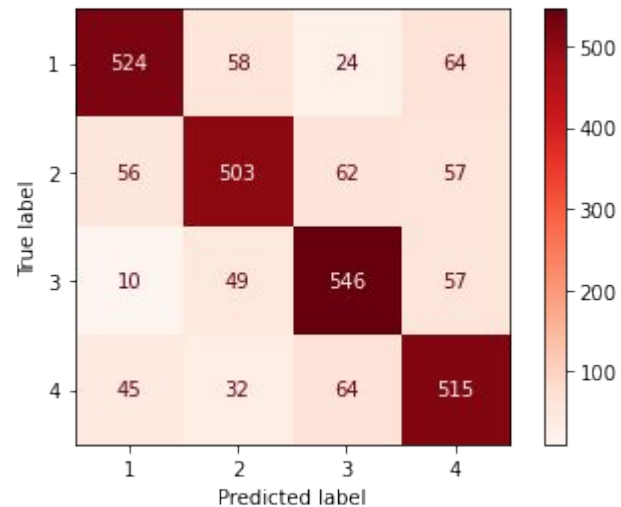
**\*Precision score of 0.78**

# NLP Model Evaluation

Reviews



Descriptions



# Modeling Listing Features

# Cleaning/Preprocessing the Listings Data

**September 2021 - June 2022**  
10,664 Listings (4,200 unique)



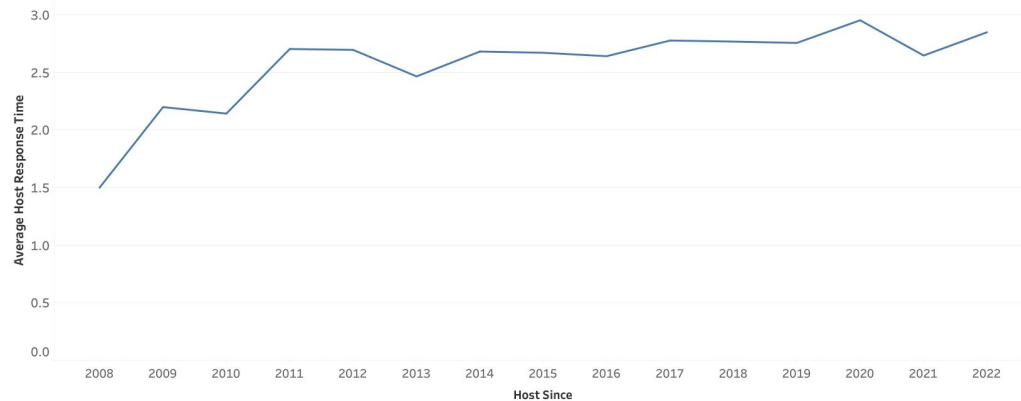
**10%**

of response  
rates missing

**4,800**

Missing  
Neighborhood  
Descriptions

Host Response Time vs. Year Started



The trend of average host response time as a function of when the host started on Airbnb.

For Response Times:  
3 indicates the host responds within an hour  
2 indicates within a few hours  
1 indicates within a day  
0 indicates within a few days or more



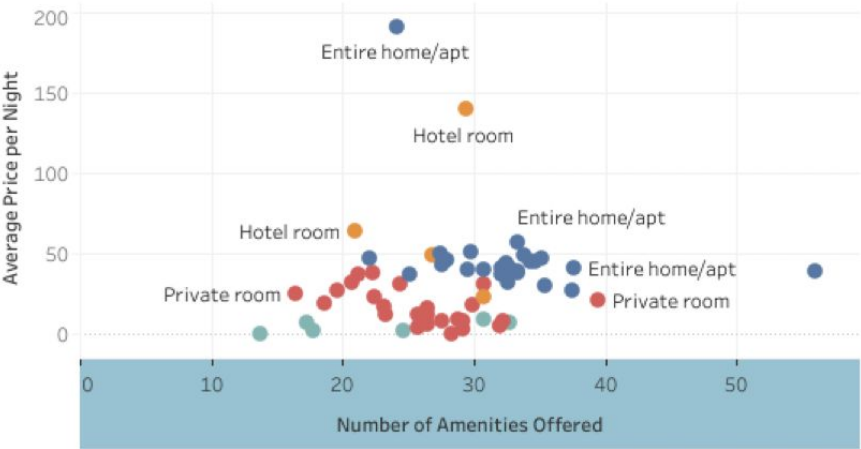
# Prices



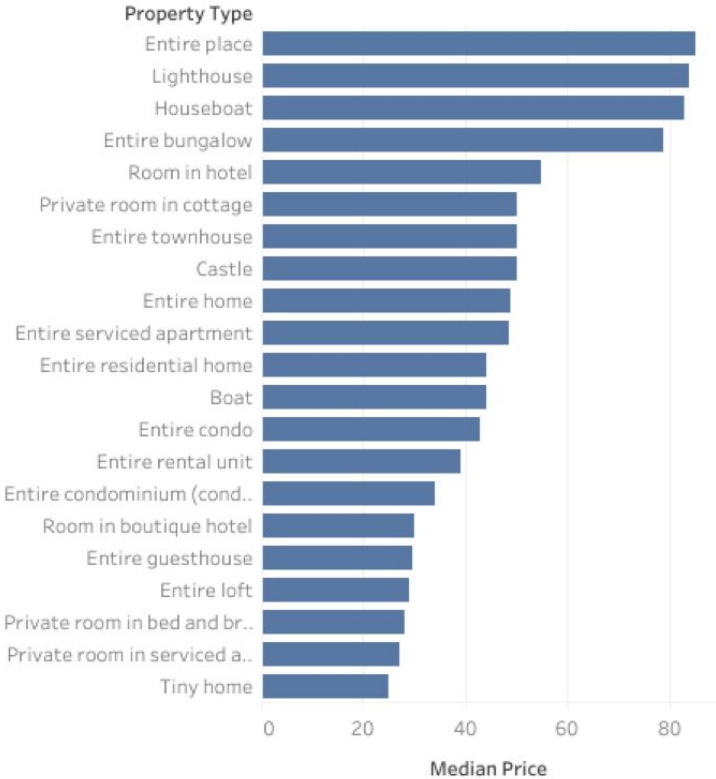
Median Price Across Neighborhoods



Price vs. Amenities



Price by Property Type



Median Ratings Across Neighborhoods



# Prices vs. Ratings

Median Price Across Neighborhoods



Correlation between  
rating and price = 0.037

# Ratings

Median Rating

4.5000

4.9900

Room Type

Entire home/apt

Hotel room

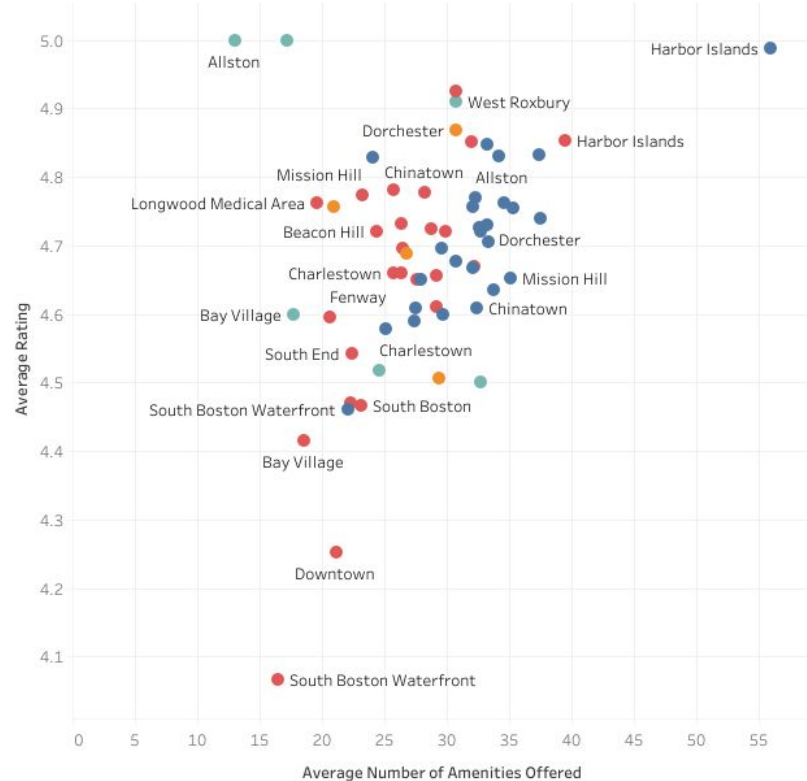
Private room

Shared room

## Median Ratings Across Neighborhoods



## Ratings Vs. Amenities



# Amenities

Amenities Offered

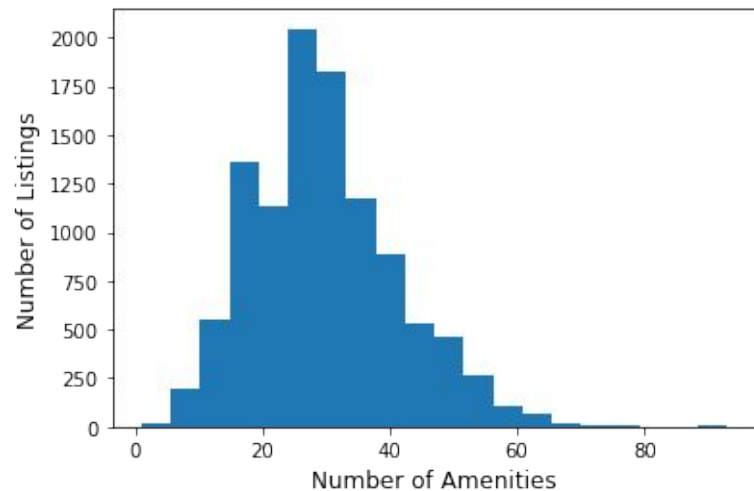


## What this place offers

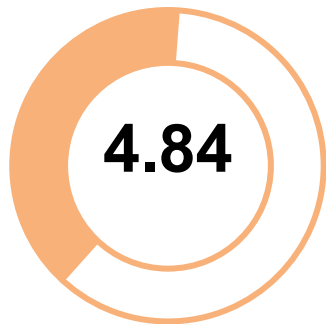
- Mountain view
- Beach access
- Wifi
- Free parking on premises
- Pets allowed
- Lake view
- Kitchen
- Dedicated workspace
- Private hot tub - available all year, open 24 hours
- Security cameras on property

Show all 75 amenities

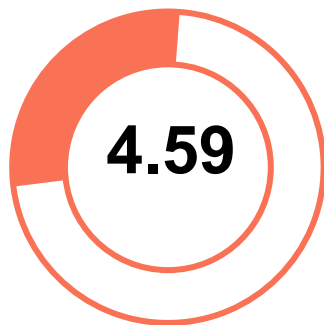
Distribution of Amenities



# Superhost Average Ratings

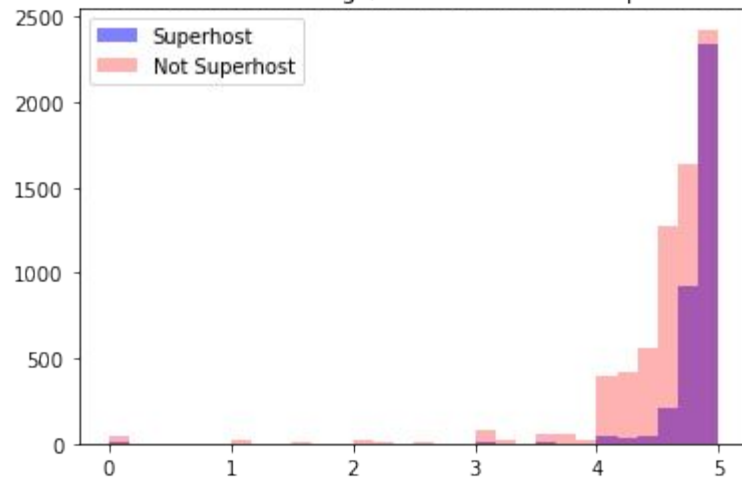


**Superhost**

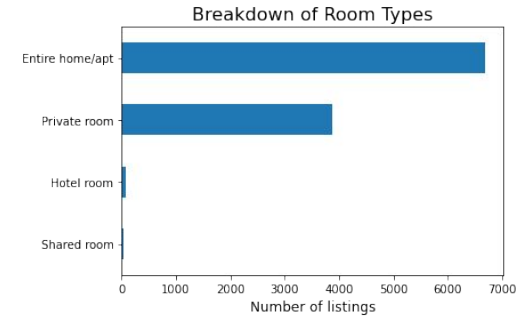
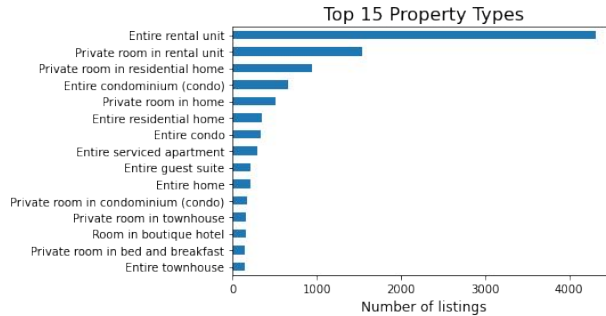


**Not a Superhost**

Distribution of Ratings, with and without Superhosts



# Property and Room Types



## Rating by Room Type

Room Type	
Shared room	4.72
Entire home/..	4.69
Hotel room	4.66
Private room	4.66

# Takeaways from EDA

## Highest Correlations to rating:



Whether the host is  
a superhost



Number of amenities



Number of reviews



Availability



Latitude and longitude

**\*No correlations over 0.2 or -0.2**

# Features Modeling: Classification



## Features:

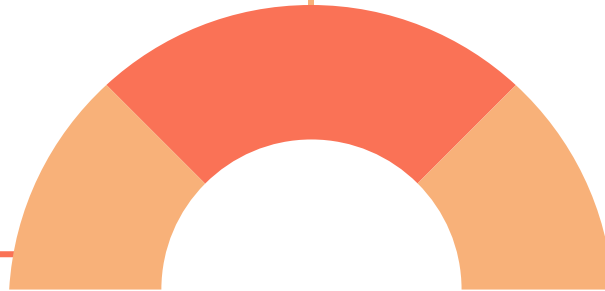
- Amenities Count
- Host Response Time
- Host Response Rate
- Number of Guests Accommodated
- Reviews per Month
- Host Verifications Count



**\*Baseline Accuracy**  
**25.44%**

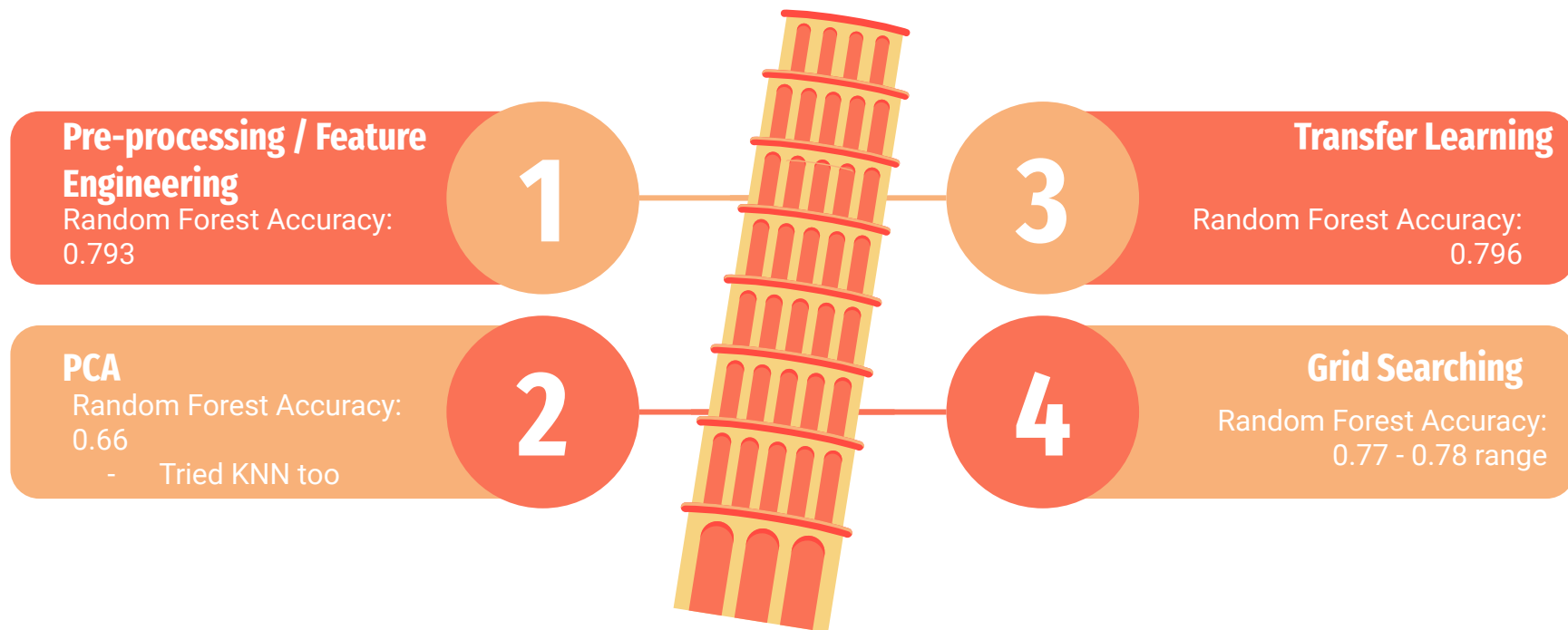
## Classifier

KNN	Multinomial Logistic Regression
Train Accuracy: 0.64	Train Accuracy: 0.37
Test Accuracy: 0.46	Test Accuracy: 0.37





# Improve the Model



# Features Modeling: Regression?



## Features:

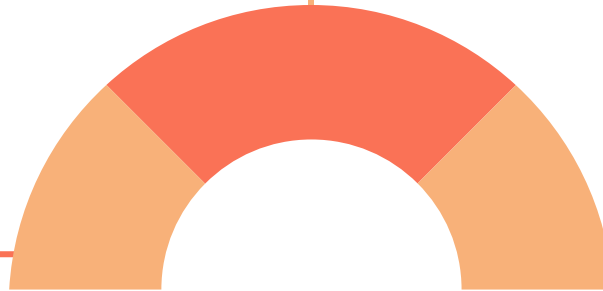
- Amenities Count
- Host Response Time
- Host Response Rate
- Number of Guests Accommodated
- Reviews per Month
- Host Verifications Count



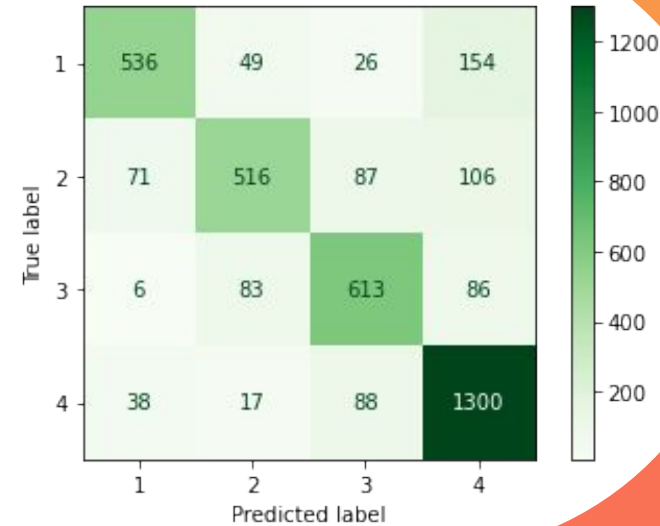
**\*Baseline MSE:**  
**0.25**

## Regressor

PCA	Transfer Learning
Best Model: Random Forest	MSE: 0.10
Training R-squared: 0.95	Test R-squared: 0.58



# Best Model: Random Forest Classifier



## Clustering Feature

KMeans - 26 clusters



136 features



Train Accuracy: 1.00:

Test Accuracy: 0.796

Specificity: 0.80

Precision: 0.79 (Macro and Micro averaging)



# Validating the Results

## San Francisco Listings

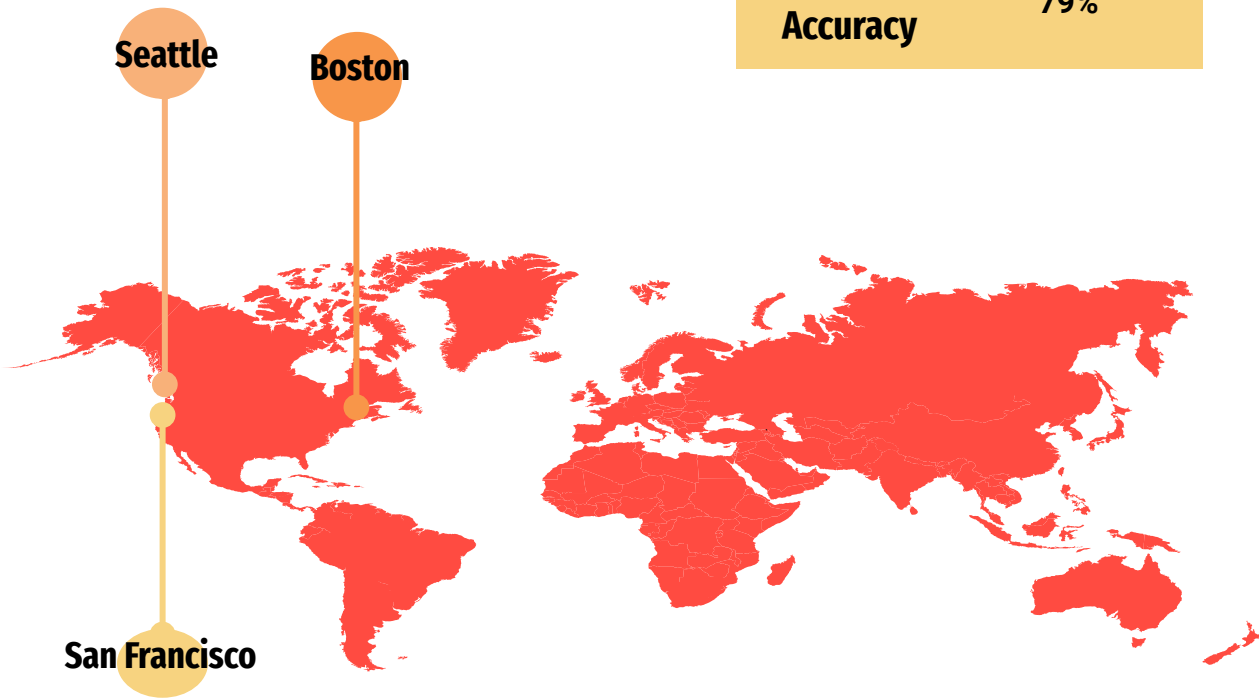
15,000 listings  
Same Random Forest  
Model, Same Features

**Test  
Accuracy** 79%

## Seattle Listings

16,000 Listings  
Same Random Forest  
Model, Same Features

**Test  
Accuracy** 82%



# Findings



Modeling Recap

**1) Listing  
Features**

**2) Host  
Descriptions**

**3) User  
Reviews**



Limitations

**NLP  
Modeling**

**Diverse  
Listing  
Types**

**Validation  
Cities**



Applicability

**79%**

**SEA**

**SF**

# International Expansion?



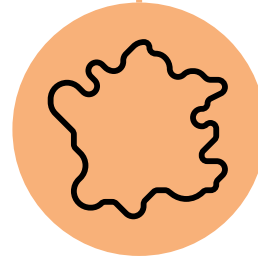
## ITALY

Mercury is the  
closest planet to  
the Sun



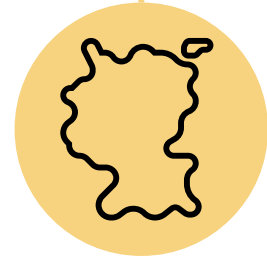
## GREECE

Despite being  
red, Mars is a  
cold place



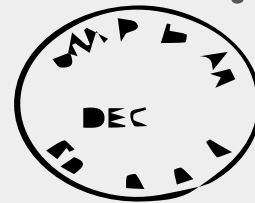
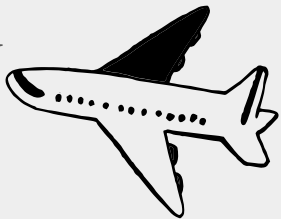
## FRANCE

Jupiter is a gas  
giant and the  
biggest planet



## GERMANY

Neptune is the  
farthest planet  
from the Sun



# Thanks!

Do you have any questions?



CREDITS: This presentation template was created by Slidesgo,  
including icons by Flaticon and infographics & images by Freepik