Airbnb London

Team: B4

airbnb

Meet the Team

Aditya Chopra

Dhruv Gandhi

Sahil Khatnani

David Ueda

Benson Yu

Table of Contents

Introduction

Selecting Our Dataset

Option 1: New York Housing Prices

- Piqued our interest in the real estate market
- Failed because after cleaning we only had 4 useful variables

Option 2: London AirBnB Dataset

- Pivoted from houses to airbnb and changed locations
- Had more features and instances than the previous dataset

What Are We Trying To Achieve?

Build a model to predict the price of an AirBnB in London

<u>Potential Uses</u>

- If someone wanted to stay at an AirBnB in London, what price can they expect to pay?
- What factors affect the price of an AirBnB?
- How much should a host list their AirBnb for?

Data Preprocessing

Our Raw Dataset

9993 Listings

20 Features

6 Categorical, 14 Numerical

Target Variable: Price

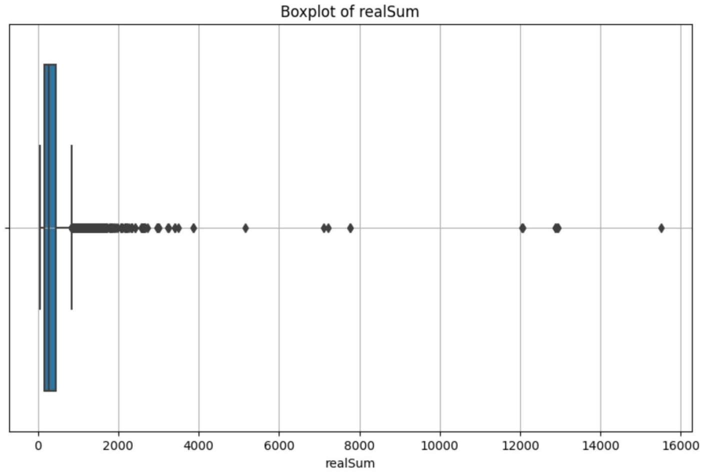
- Unnamed: 0 (id)
- realSum (price)
- room_type
- room_shared
- room_private
- person_capacity
- host_is_superhost
- multi
- biz
- cleanliness_rating
- guest_satisfaction_overall
- bedrooms
- dist
- metro_dist
- attr_index
- attr_index_norm
- rest_index
- rest_index_norm
- lng
- lat

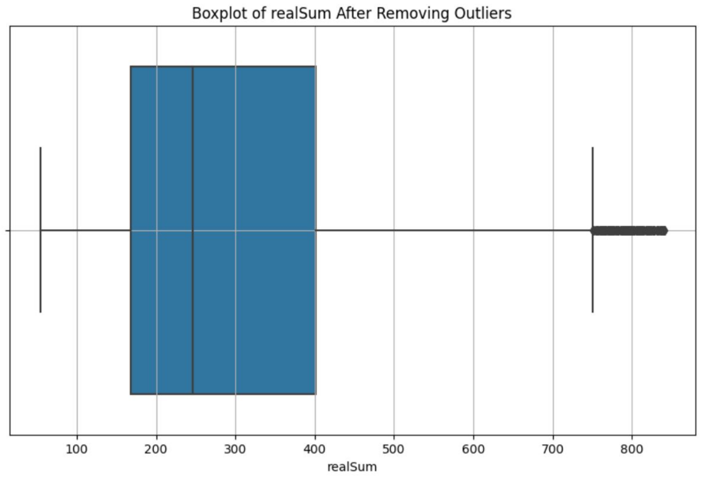
Data Preprocessing

- We combined two datasets from weekdays and weekends data
- Checked for Null and Duplicates
- Checked for Outliers
- Dropped Unnecessary Features
- EDA and PCA

Data Preprocessing - Cont.

Checking for Outliers





Columns Dropped

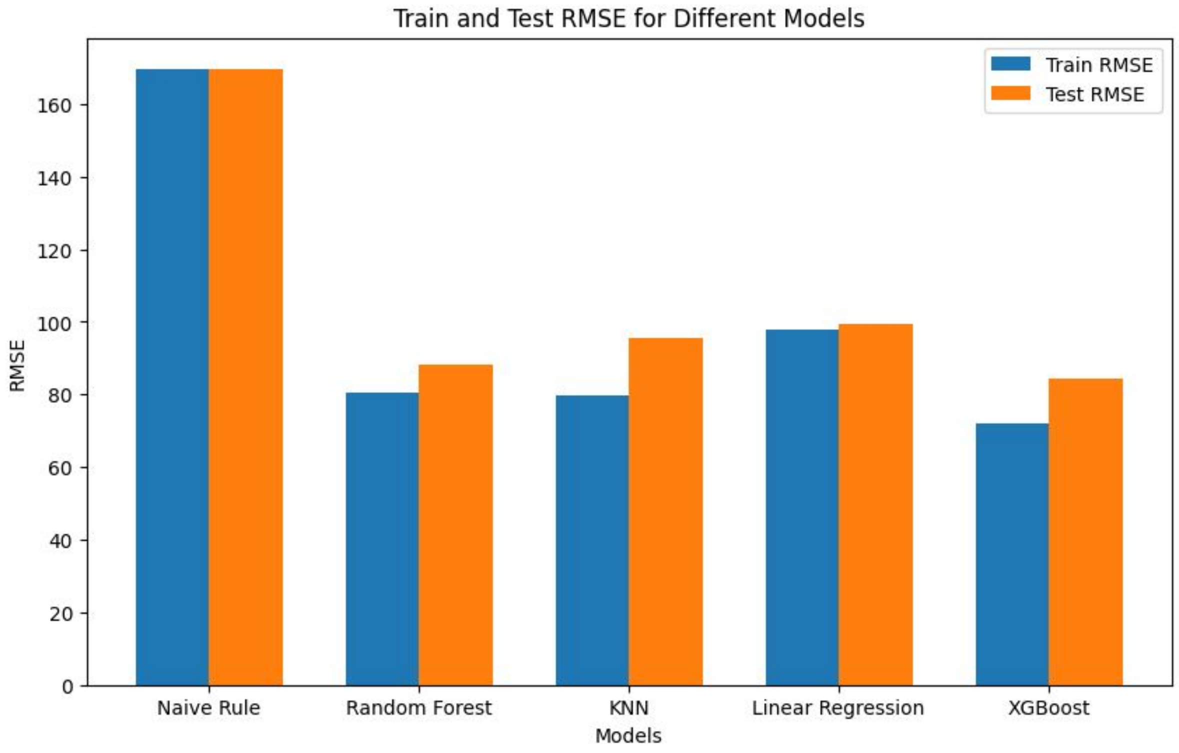
- Unnamed: 0
- attr_index
- rest_index
- room_shared
- room_private
- lng
- lat
- dist

Dummy Variables

- room_type → dummy var
- is_weekend → dummy var

Predictive Models

Models We Tested




```
Naive Rule RMSE 169.71537119259716
Random Forest RMSE: 88.40182692765374
KNN RMSE: 95.63322917905936
Linear Regression RMSE: 99.24230165218675
XGBoost RMSE: 84.27547329939559
```

Top Models: Random Forest and XGBoost

Naive Rule

The Naive Rule provides a baseline for our models to outperform

What is Naive Mean?

- takes the mean of target (price)
- uses avg as prediction for all data points in test set
- independent data points
- does not consider any relationships b/w features or target

Train RMSE (Predicting Mean): 169.655

Test RMSE (Predicting Mean): 169.715

Random Forest

Results

Train RMSE: 80.378

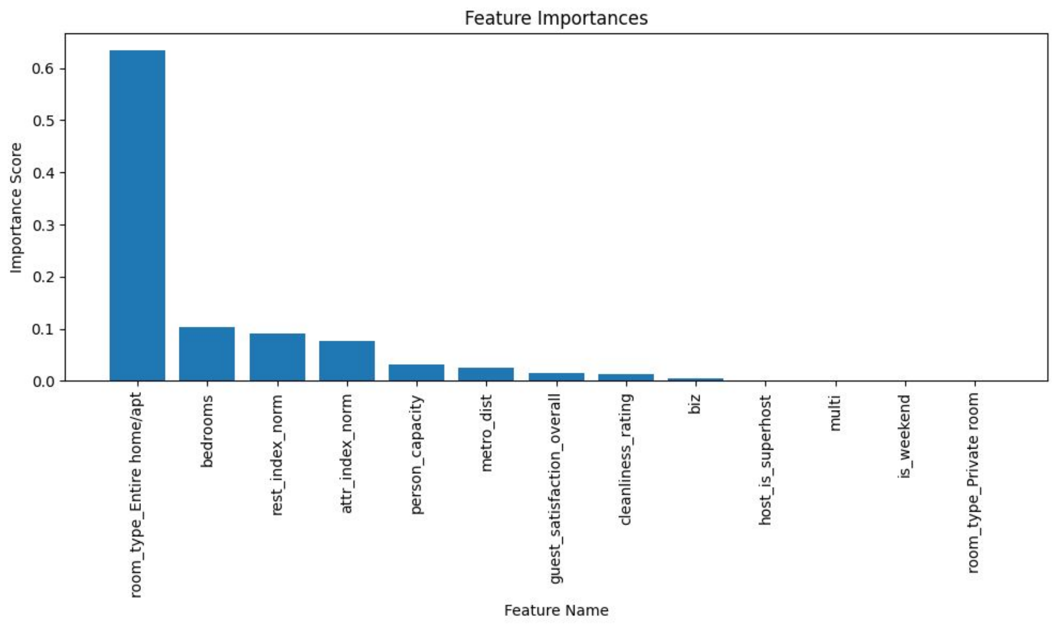
Test RMSE: 88.402

Most Important Features:

1. room_type_Entire home/apt
2. bedrooms
3. rest_index_norm

The price is affected the most by if the room type is the entire home or not

Feature Importance



XGBoost

Leading ML library for classification, **regression**, and ranking problems the industry.

XGBoost is based on the **gradient boosting** framework.

Gradient Boosting **ensemble** learning method that **combines** multiple weak learners (decision trees) to create a **strong** learner.

XGBoost can visualize **feature importances**.

Pop Quiz

What do you guys think is going to be the most important feature that relates to price?

A. Person Capacity

B. Distance From The Metro

C. Cleanliness

D. Guest Satisfaction Rate

The Correct Answer is ...

B ✅

(distance from the metro)

XGBoost cont.

Results

Train RMSE: 71.962

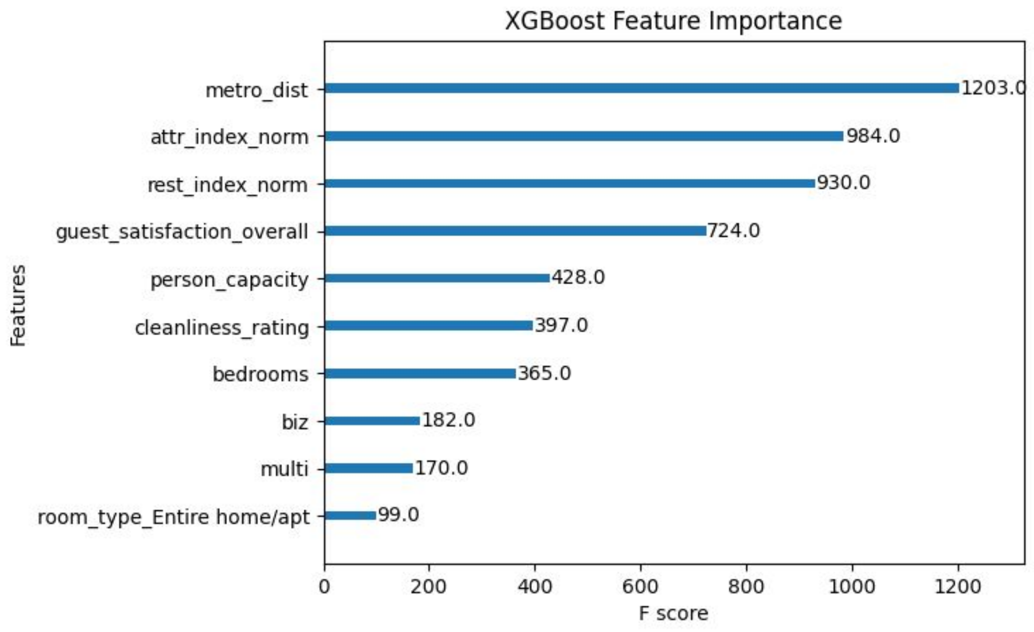
Test RMSE: 84.275

Most Important Features:

1. metro_dist
2. attr_index
3. rest_index

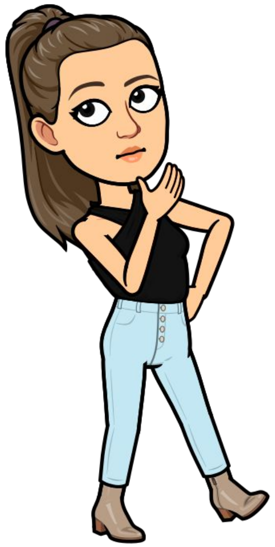
The price depends the most on how convenient the location of the airbnb is

Feature Importance



Case Study

Use Case 1 - Couple's Trip



This is Tom and Ellie.

They are planning a trip to London. They both enjoy exploring and eating good food.

They want to stay at an Airbnb that is clean and wants a place that people have stayed at and enjoyed before them.

They want a listing that has multiple rooms and prefers to be in a residence to themselves.

They are looking for a place to stay for 4 nights and are going from Monday to Friday. They want to know how much to budget for housing for this trip.

Use Case 1 Cont…

Here's how he would use our algorithm:

```
Enter person capacity: 2
Is host superhost? (1 for Yes, 0 for No): 1
Enter multi value: 1
Enter biz value: 0
Enter cleanliness rating: 10
Enter guest satisfaction overall rating: 98
Enter number of bedrooms: 1
Enter metro distance: 0.1
Enter attraction index: 90
Enter restaurant index: 90
Is it a weekend? (1 for Yes, 0 for No): 0
Is room type entire home? (1 for Yes, 0 for No): 1
Is room type private room? (1 for Yes, 0 for No): 0
Predicted Price: $ 419.0226
```

Example Listing:

Entire condo in London, United Kingdom

2 guests · 1 bedroom · 1 bed · 1 bath

$432 night

CHECK-IN	CHECKOUT
6/3/2024	6/7/2024

GUESTS
1 guest

Reserve

Guest favorite — One of the most loved homes on Airbnb, according to guests — 4.92 ★★★★★

Hosted by Stephen
Superhost · 9 years hosting

Self check-in
Check yourself in with the lockbox.

Stephen is a Superhost
Superhosts are experienced, highly rated Hosts.

Free cancellation for 48 hours
Get a full refund if you change your mind.

Where you'll be

Neighborhood highlights

Located in the heart of west end, London. Within a few minutes walk, you will find an abundant choice of Restaurants, pubs, shopping at Oxford st, Regent Street and Bond Street. The British Museum, Theatreland, Chinatown, Soho.

Use Case 2 - Cindy and Jeff

This is Cindy and Jeff.

They are co-superhosts looking to put one of their newly built apartments for sale. They want customers to be business people traveling on weekdays, as they like to spend the weekends cleaning up their place.

Their place has two bedrooms and is kept extremely clean.

Their apartment is not close to the city and does not have highly rated restaurants near it, but it is close to the metro, so they are unsure of how to price their apartment on Airbnb.

What should they do?

Use Case 2 Cont...

Here's how they would use our algorithm:

```
Enter person capacity: 2
Is host superhost? (1 for Yes, 0 for No): 1
Enter multi value: 0
Enter biz value: 1
Enter cleanliness rating: 90
Enter guest satisfaction overall rating: 80
Enter number of bedrooms: 2
Enter metro distance: 0.3
Enter attraction index: 25
Enter restaurant index: 25
Is it a weekend? (1 for Yes, 0 for No): 0
Is room type entire home? (1 for Yes, 0 for No): 1
Is room type private room? (1 for Yes, 0 for No): 0
Predicted Price: $ 651.59094
```

Real-Life Implication: Dhruv Getting Scammed

I recently planned a trip to London for spring break and booked an AirBnB for myself and 3 friends.

We found a listing that was cheaper than others in the area, so we thought it was a really good deal.

However, the listing was not as advertised, so we all had to find another place to stay for our trip.

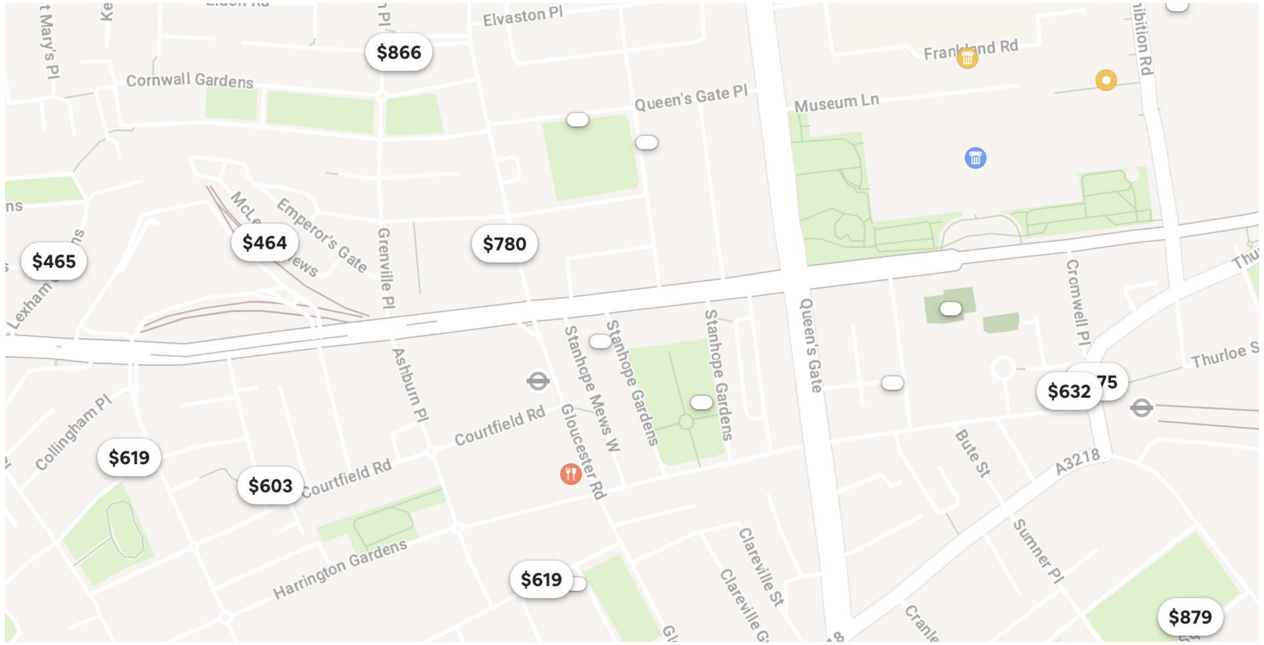
Let's see if our model would have saved me from this disastrous event.

Dhruv Getting Scammed Cont...

Listing Price: $387/night

Our model's predicted price for the listing: $635/night

```
Enter person capacity:  5
Is host superhost? (1 for Yes, 0 for No):  0
Enter multi value:  1
Enter biz value:  1
Enter cleanliness rating:  9
Enter guest satisfaction overall rating:  80
Enter number of bedrooms:  2
Enter metro distance:  0.1
Enter attraction index:  70
Enter restaurant index:  80
Is it a weekend? (1 for Yes, 0 for No):  0
Is room type entire home? (1 for Yes, 0 for No):  1
Is room type private room? (1 for Yes, 0 for No):  0
Predicted Price: $ 634.948
```



Findings and Implications

Can predict a baseline price that a user can expect to pay

Can predict a price that hosts can estimate for their listings

Seasonality and demand plays a large role in the pricing of a listing

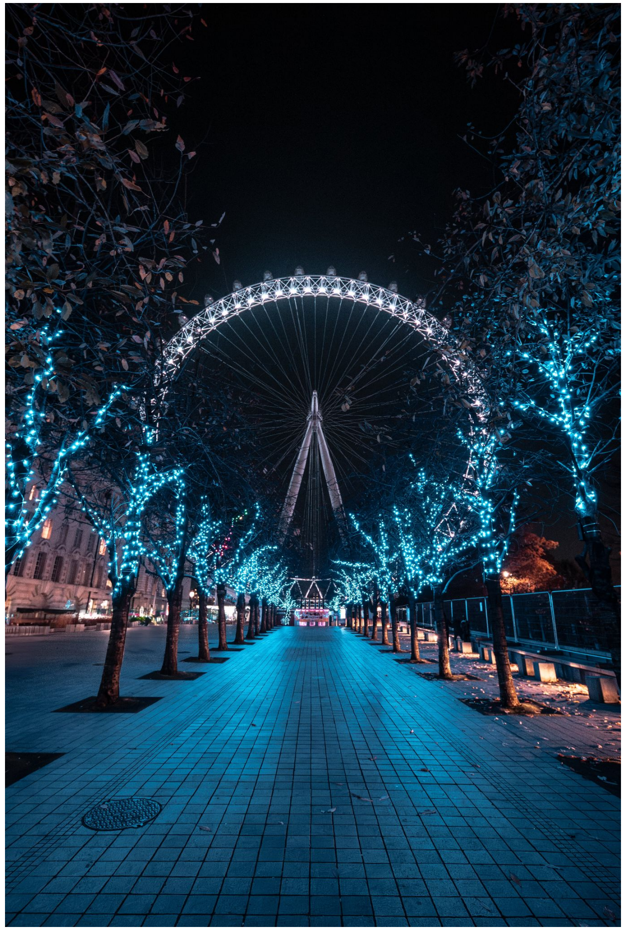
Improvements

Our dataset was missing key features that may play a important role in pricing:

- number of bathrooms
- dates of stay

Our model requires all features to be known for functionality.

The model should be able to predict a fair price even without knowing all variables.



Thank You!

Any Questions?