Predicting Prices for Airbnb Listings in Singapore

Group 17
Edgar Tok Jun Jie
Janaki H Nair
Sean Yap Hock Keong,
Dodda Sharon Olivia
Wayne Lim Wei En

About Airbnb

House-sharing platform

191 countries

150 million users worldwide

5 million listings worldwide

Problem Description

53% of travelers use Airbnb because of the low cost compared to traditional hotels

No official pricing tool for setting the price

Our Solution

a strategy that **recommends a price range** to an owner based on their location

by **predicting** the price range based on location

Models used:

- Decision Tree Classifier
- Forest Regression

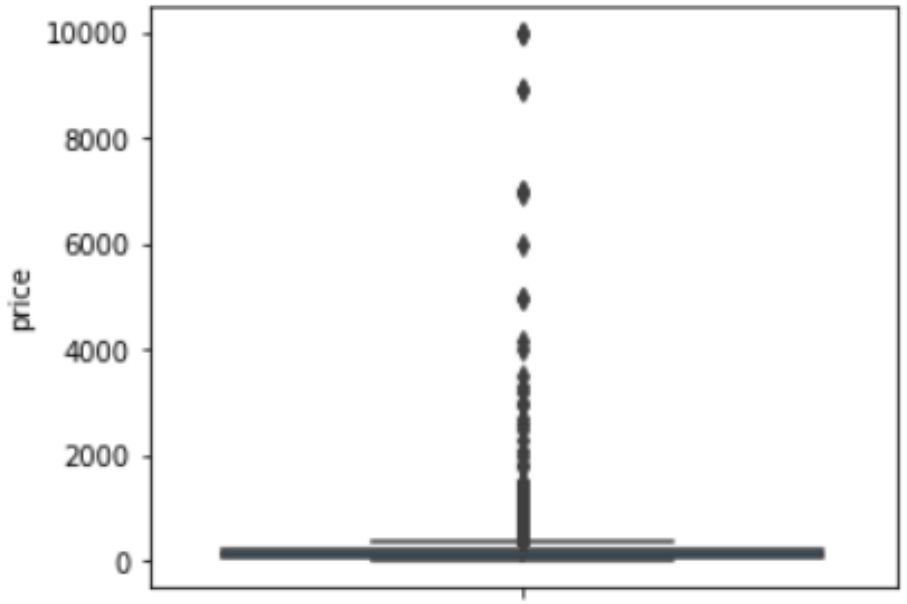
Dataset

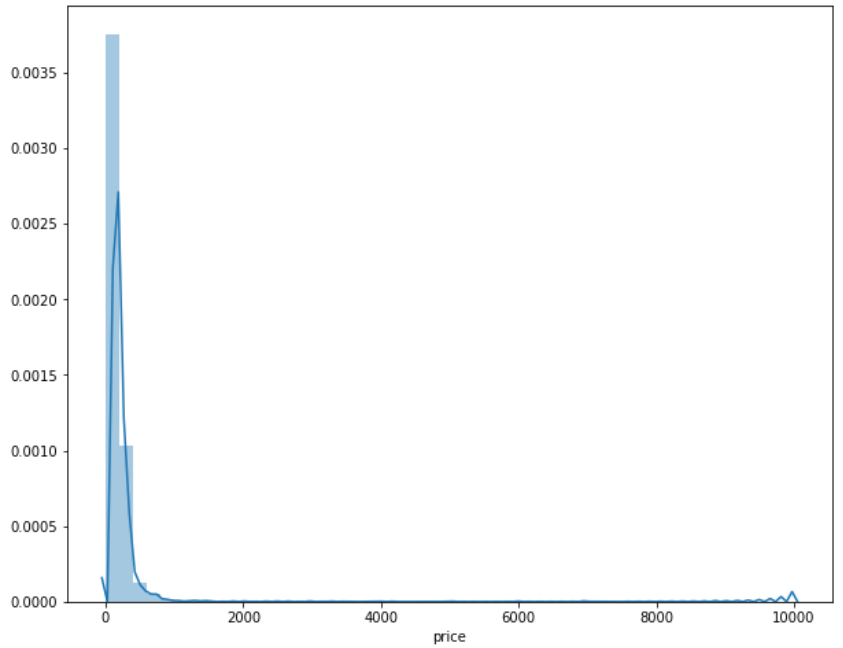
```
id
name
host_id
host_name
neighbourhood_group
neighbourhood
latitude
longitude
room_type
price
minimum_nights
number_of_reviews
last_review
reviews_per_month
calculated_host_listings_count
availability_365
dtype: int64
```

Features

Label

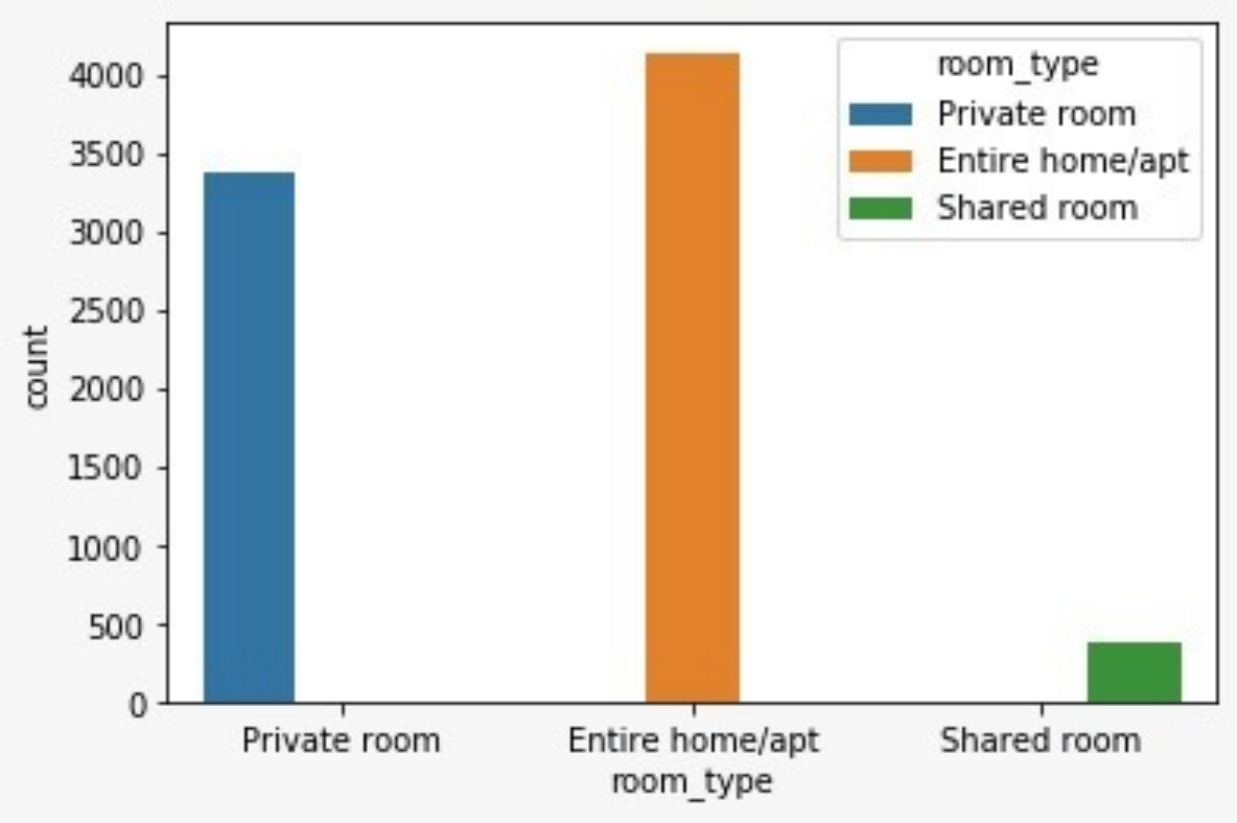
Data Exploration - Price





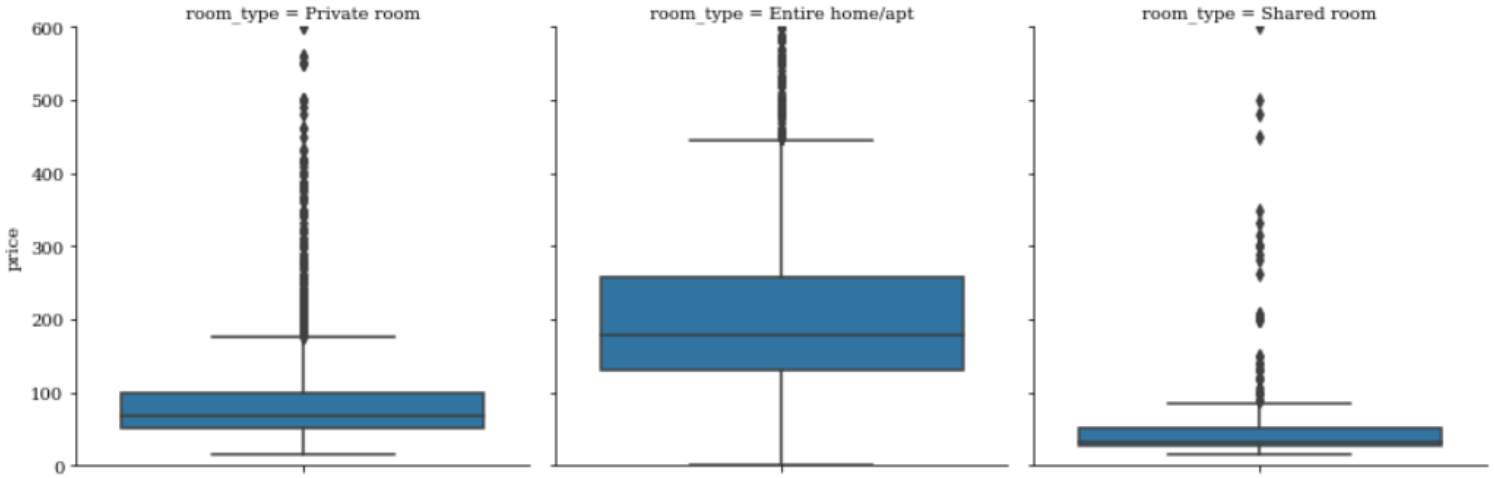
- Outliers (e.g. 10000)
- Most of the prices within price range of 2000

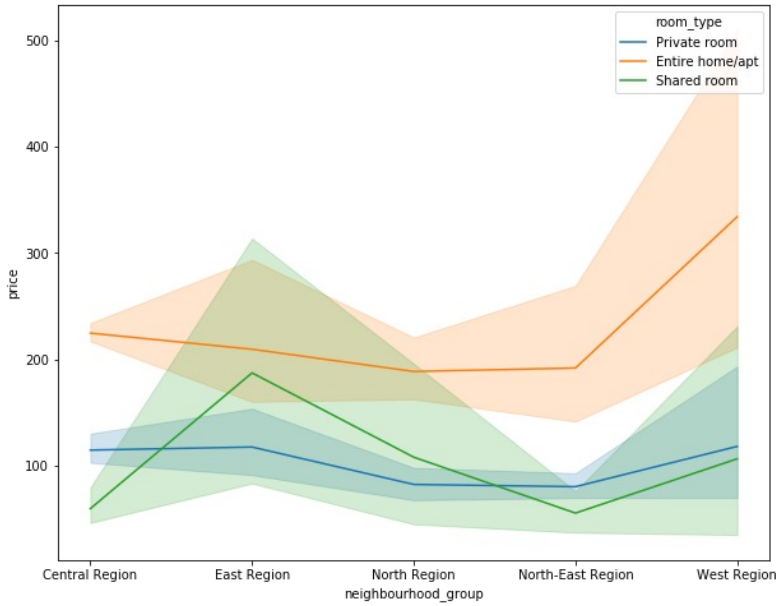
Data Exploration – Room Type



Entire home/apt is the most popular room type

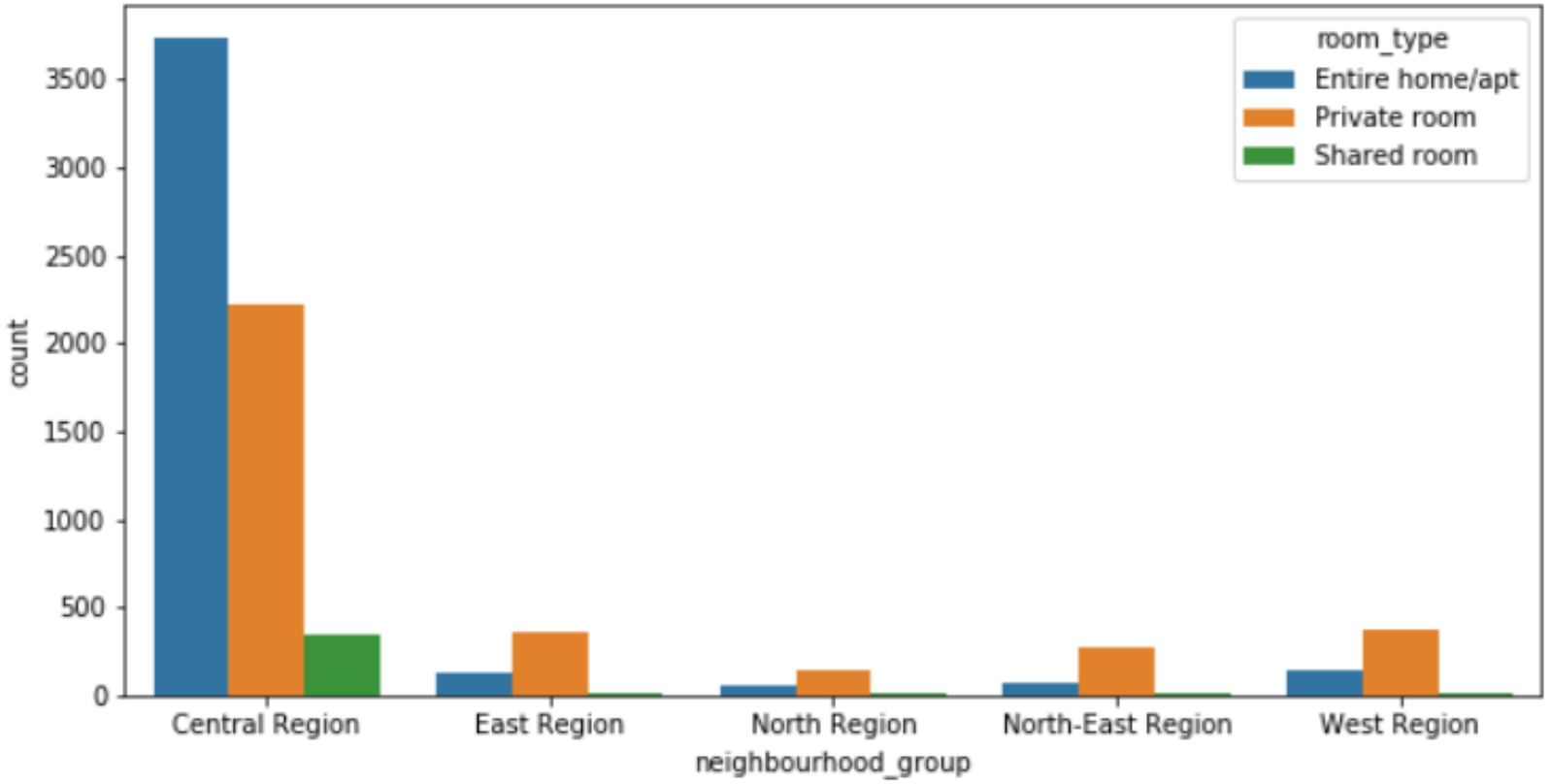
Data Exploration – Room Type





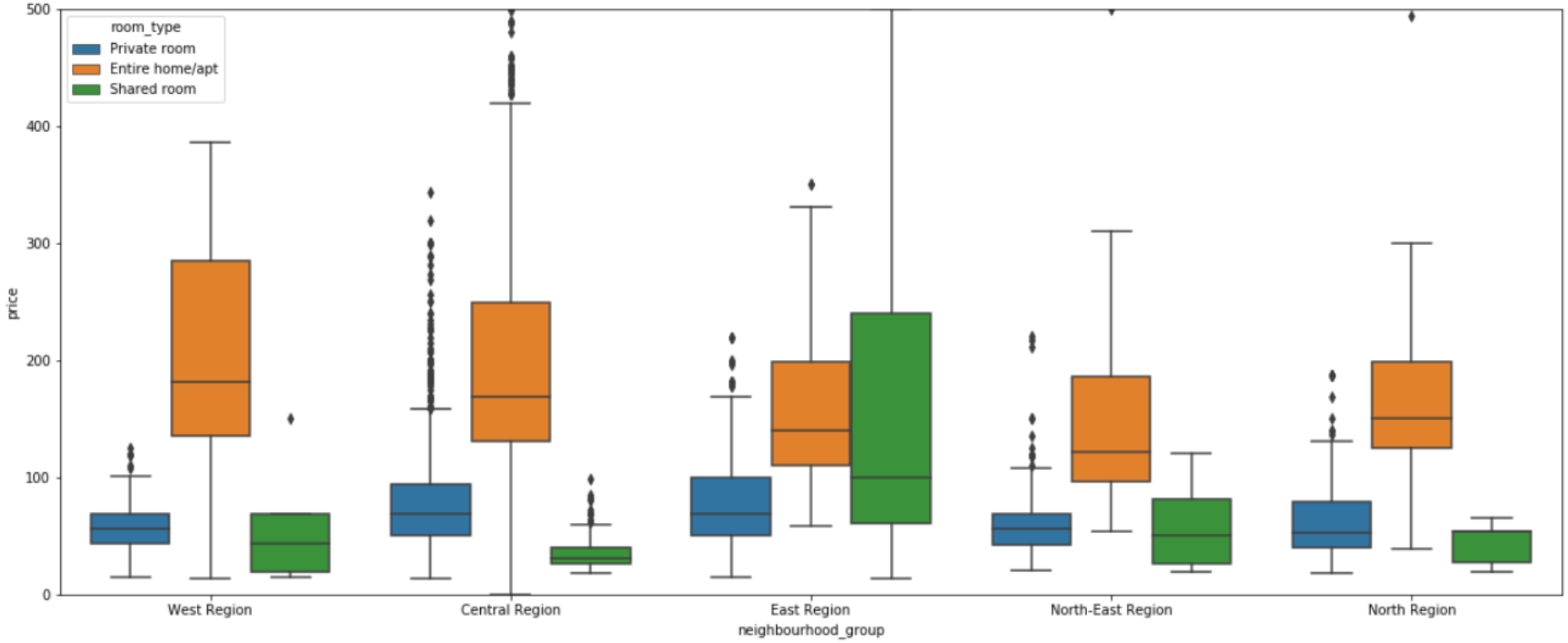
Price range is highest for
Entire home/apt

Data Exploration – Neighbourhood Group



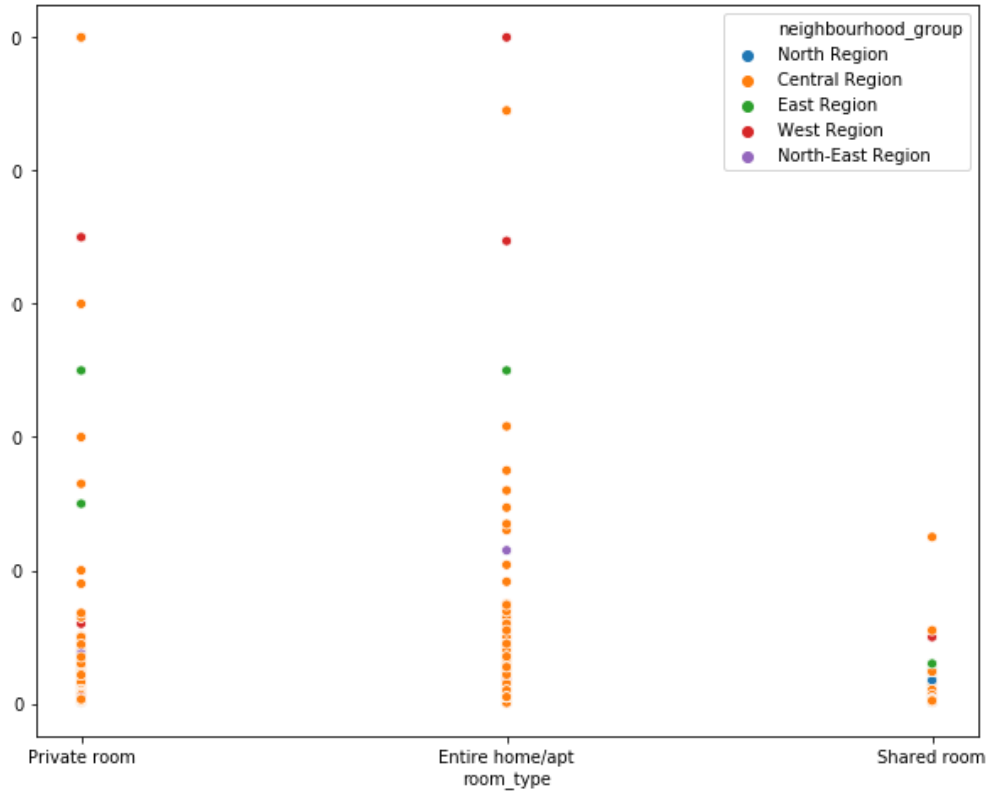
Central Region is the most popular region

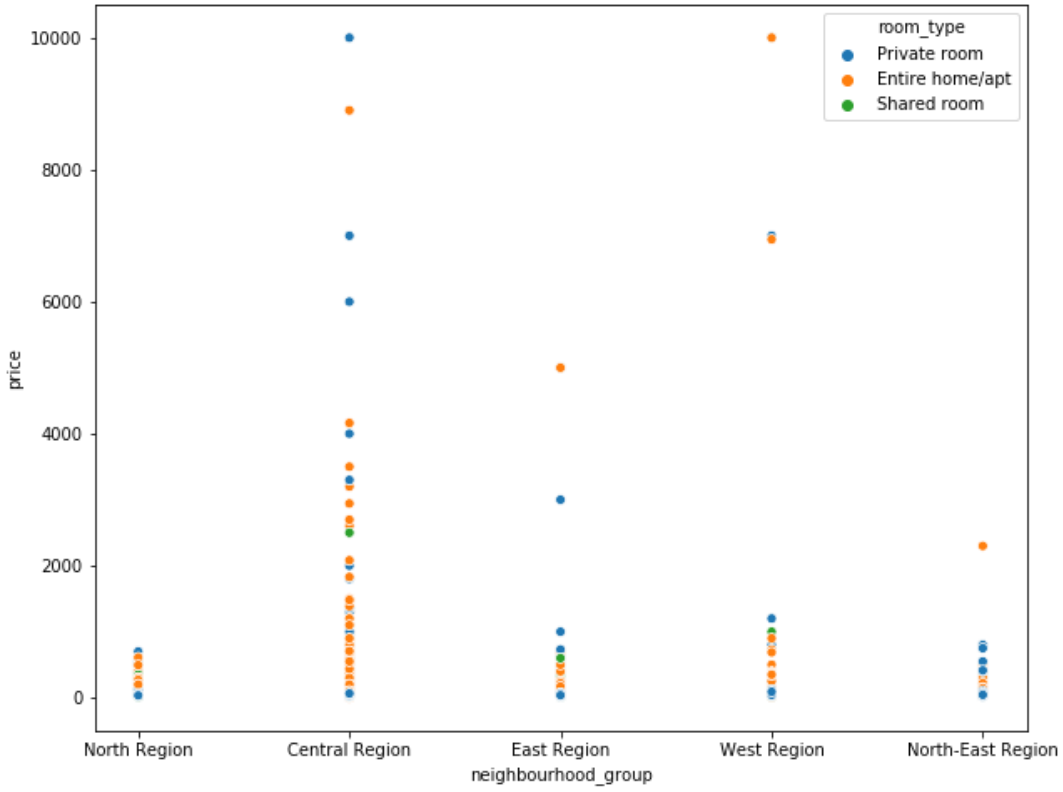
Data Exploration – Neighbourhood Group



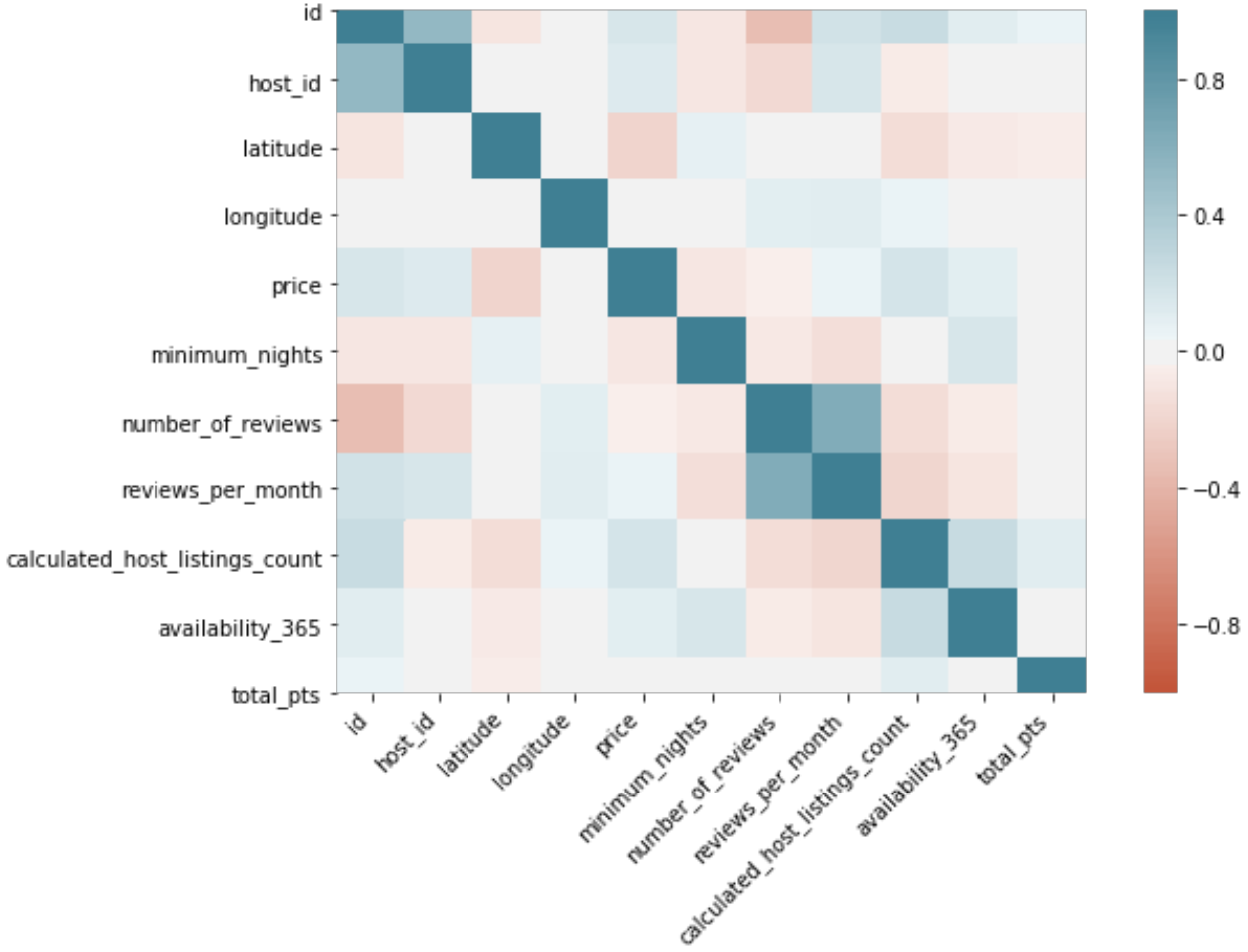
Entire home/apt has a higher price for any region

Data Exploration (Outliers)





Data Exploration



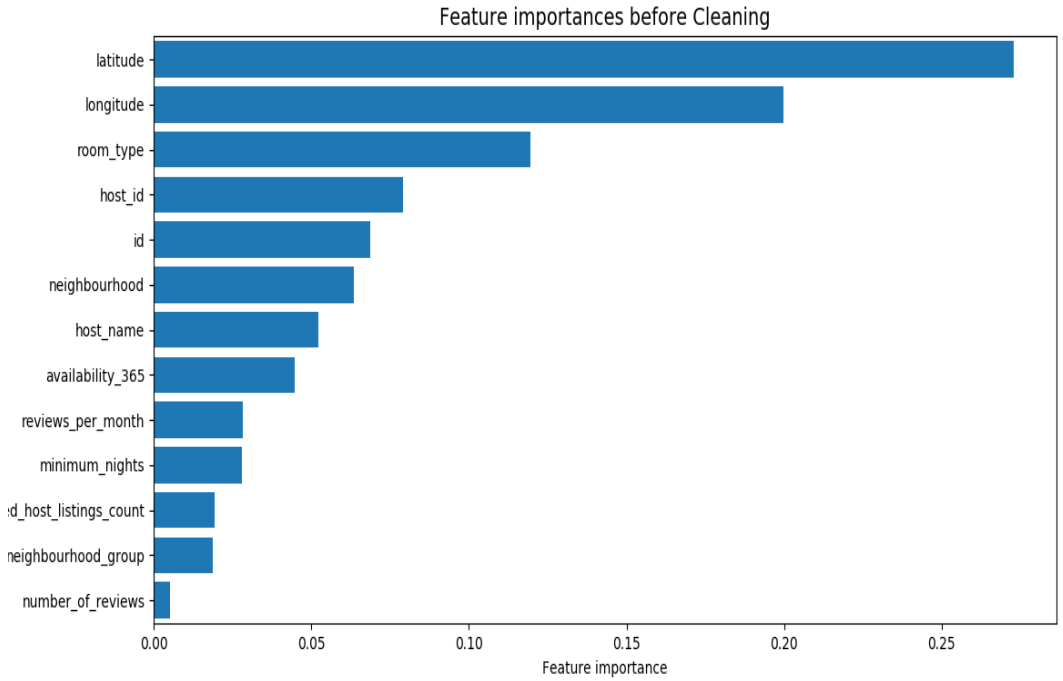
Most correlated variables with price:

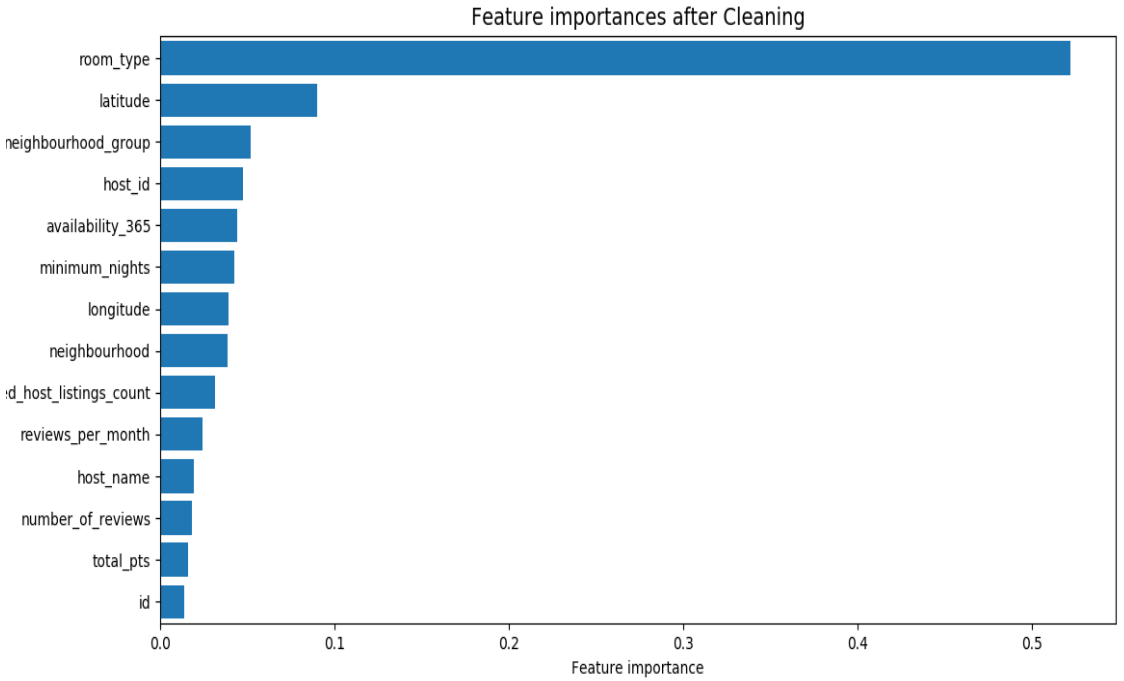
- **id**
- **calculated_host_listings count**

Somewhat correlated variables:

- **host_id**
- **reviews_per_month**
- **availability_365**

Data Exploration (Cleaning)





Data Exploration (Cleaning)

Explain what we removed

Models

Decision Tree Classifier

Uncleaned

Features	Accuracy
Neighbourhood, Room_type	0.4348
Neighbourhood, Room_type, Latitude, Longitude	0.3801
Neighbourhood, Room_type, Number_of_reviews	0.3940
Neighbourhood, Room_type, Latitude, Longitude, Number_of_reviews	0.3843

Cleaned

Features	Accuracy
Neighbourhood, Room_type	0.4552
Neighbourhood, Room_type, Latitude, Longitude	0.4209

Models

Random Forest Regressor

Uncleaned

Features	Accuracy
room_type, longitude, latitude	64.26%
room_type, longitude, latitude, host_id	70.7%
room_type, longitude, latitude, host_id, availability_365	71.44%

Cleaned

Features	Accuracy
room_type, longitude, latitude	36.23%
room_type, longitude, latitude, host_id	43.04%
room_type, longitude, latitude, host_id, availability_365	44.3%