# Property Rentals Case Study

**Manuel Urbano Rodriguez**

Property Rentals Analysis and Estimation

21 April 2022

# Introduction

## Business Problem

- Inn the Neighborhood allows people to rent out their properties for short stays.

- 2% of the people visiting the webpage leave it without even signing up.

- A new application needs to be developed to attract people. This application estimates the money that they would get for their property.

- Different features of current rental properties need to be analysed.

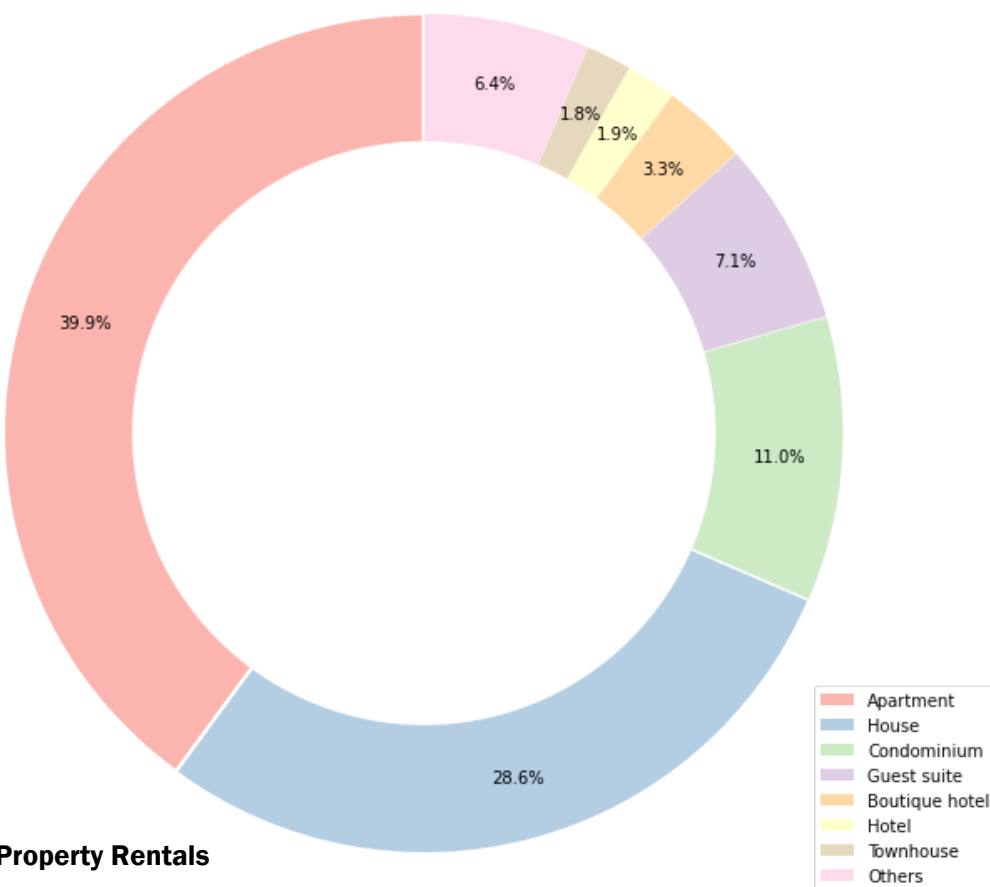- We are interested in avoiding estimating prices that are more than 25 USD off the actual price.

## Methodology

- We have access to the data of 8111 rental properties of various types, sizes, prices and locations.

- An initial Exploratory Data Analysis (EDA) will be performed to determine the best approach to build the model.

- The model will cover how the different characteristics of properties relate to each other and contribute to the final price of the property.

- The model will target to be less than 25 USD away from the actual price of the property.

- The language used for the analysis is Python with the libraries Pandas, Seaborn and Matplotlib and scikit-learn.
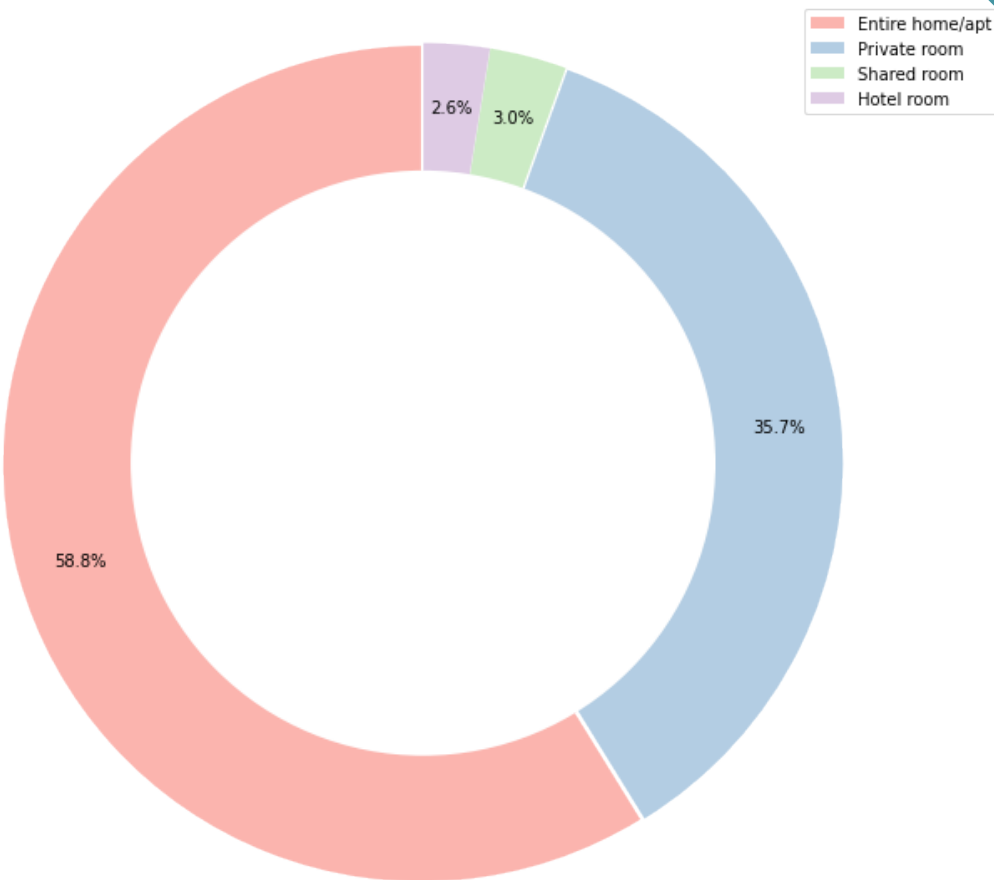
# Exploratory Data Analysis

## What is the Distribution of Property and Room Types?
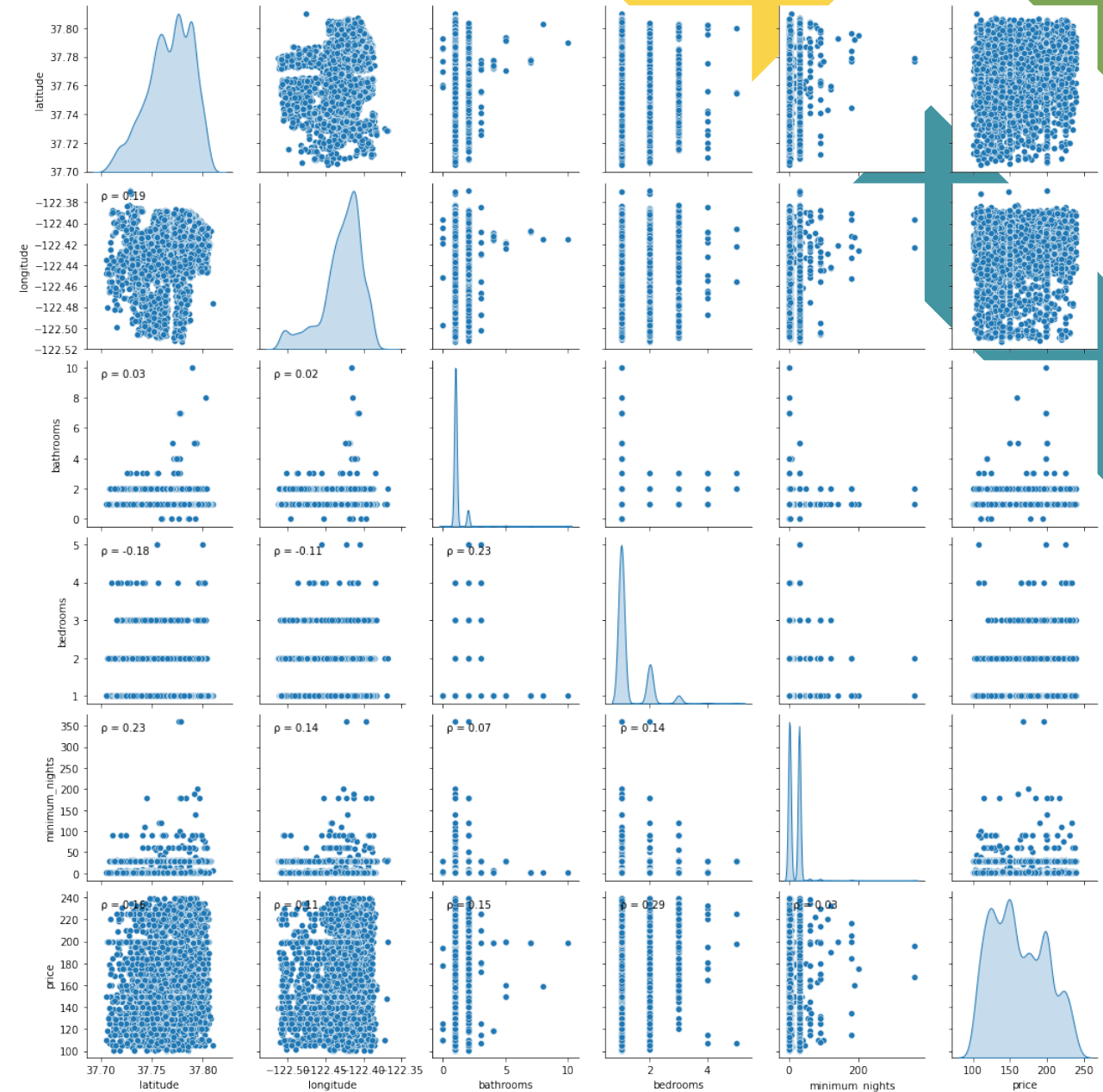


### Distribution of Property Types

- 6.4%
- 1.8%
- 1.9%
- 3.3%
- 39.9%
- 7.1%
- 11.0%
- 28.6%

Legend:
- Apartment
- House
- Condominium
- Guest suite
- Boutique hotel
- Hotel
- Townhouse
- Others

### Distribution of Room Types

- 2.6%
- 3.0%
- 35.7%
- 58.8%

Legend:
- Entire home/apt
- Private room
- Shared room
- Hotel room

**Property Rentals**

# Exploratory Data Analysis

## What is the Correlation between variables?

- Removing outliers using a 0.25 and 0.75 quantile and maintaining all types of property types.

- Plotting each parameter against each other and computing their Pearson correlation coefficient, the ones with **highest correlation** with price are:

  - Bedrooms and Price -> p = 0.29

  - Latitude and Price -> p = 0.16

  - Longitude and Price -> p = 0.11

# Exploratory Data Analysis

## What is the Price Distribution of Property Types and Room Types?



Distribution of Prices per Property Type



Distribution of Prices per Room Type

- Cottage and Villas are the most expensive property types followed closely by Hotels.
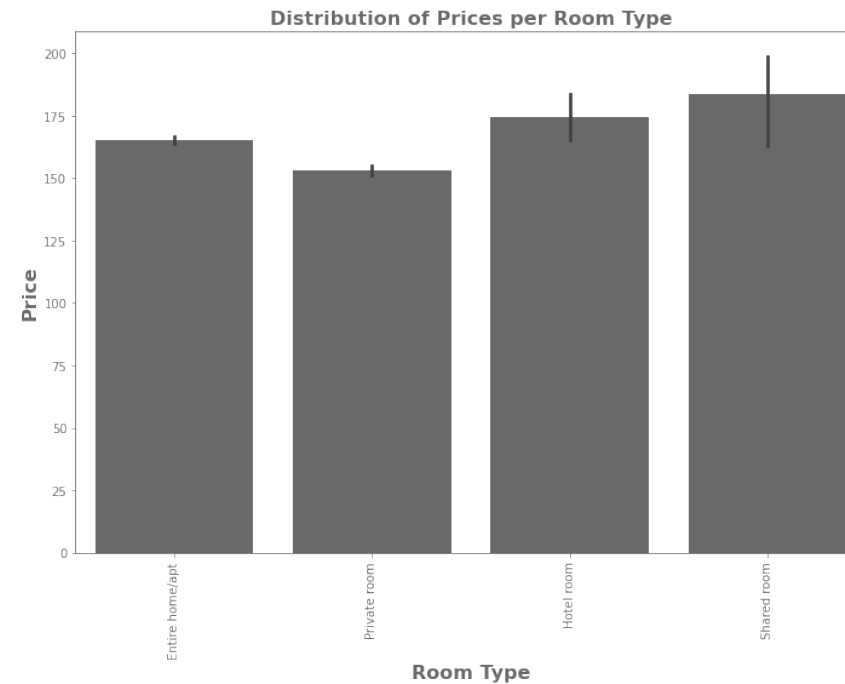
- Shared rooms and Hotel rooms tend to be the most expensive ones.

# Exploratory Data Analysis

## What is the Price Distribution depending on Location?

- All of the properties are based in San Francisco.

- Bigger circles account for more expensive prices.

- More properties are available at the city centre.

- Properties tend to be more expensive at the city centre.



Price Distribution of Properties on a Map
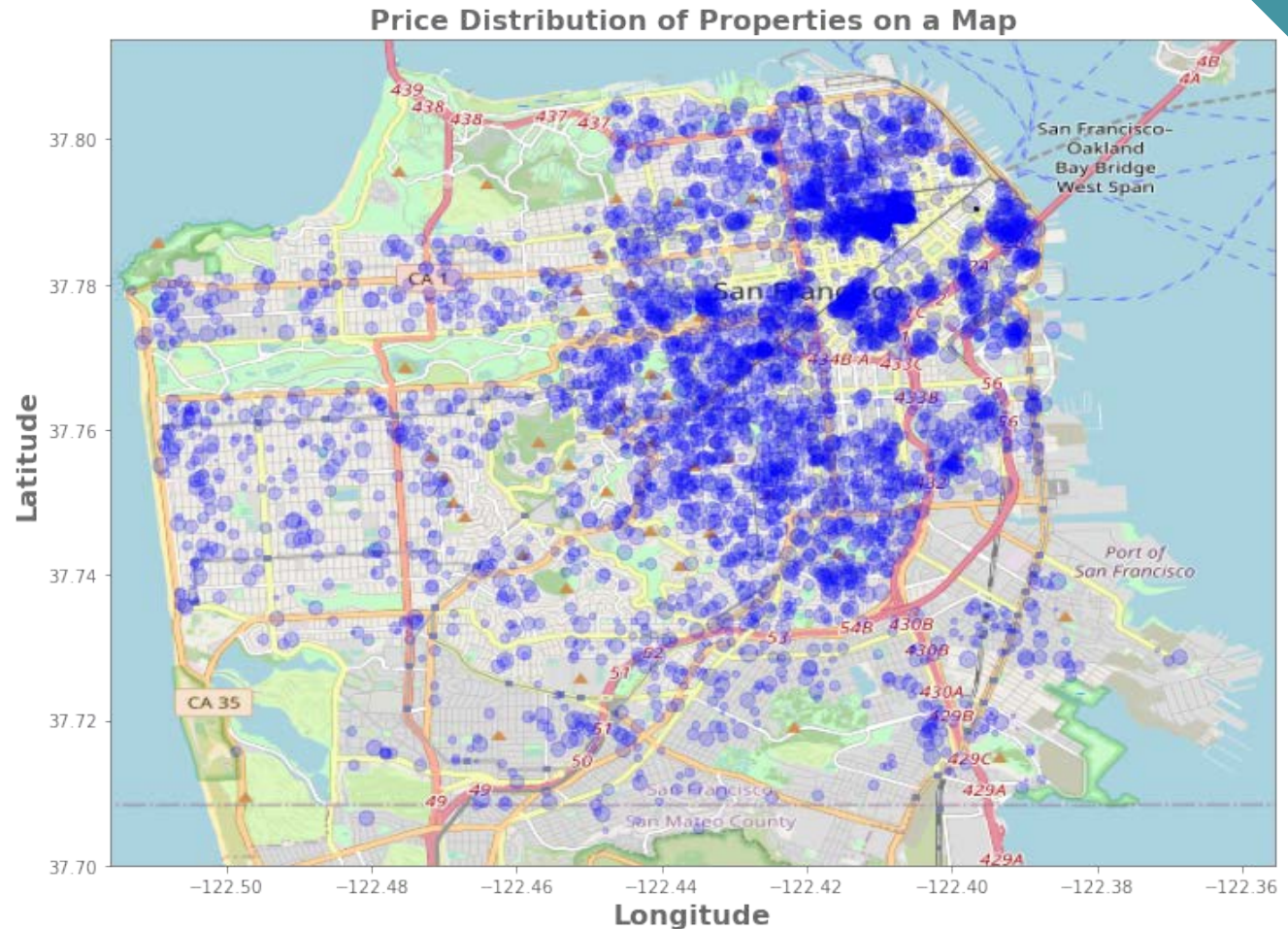
# Modelling

## Methodology and Deployment

- Data with One-Hot Encoding is used to train Regression models as the price of the properties is a continuous numeric variable.

- Four different algorithms are trained and tested to select the best one.

- The error already meets the target for the Random Forest without performing any further tuning.

| Model Type | Mean Absolute Error (USD) | Cross Validation – MAE (USD) |
|---|---|---|
| Decision Tree | 30.82 | 31.68 |
| Random Forest | 24.22 | 23.29 |
| Gradient Boosting | 25.53 | 23.31 |
| Support Vector Machine | 26.03 | 24.35 |

# Modelling

## Fine Tuning and Features Importance

- Performing a Randomized Search, the best model parameters are:

  | n_estimators = 600 | min_samples_leaf = 1 |
  | --- | --- |
  | min_samples_split = 7 | max_features = sqrt |

- The final MAE achieved is 23.91 USD.

- Latitude, longitude and the number of bedrooms are the features found to have the greatest importance for the model.



Feature Importances
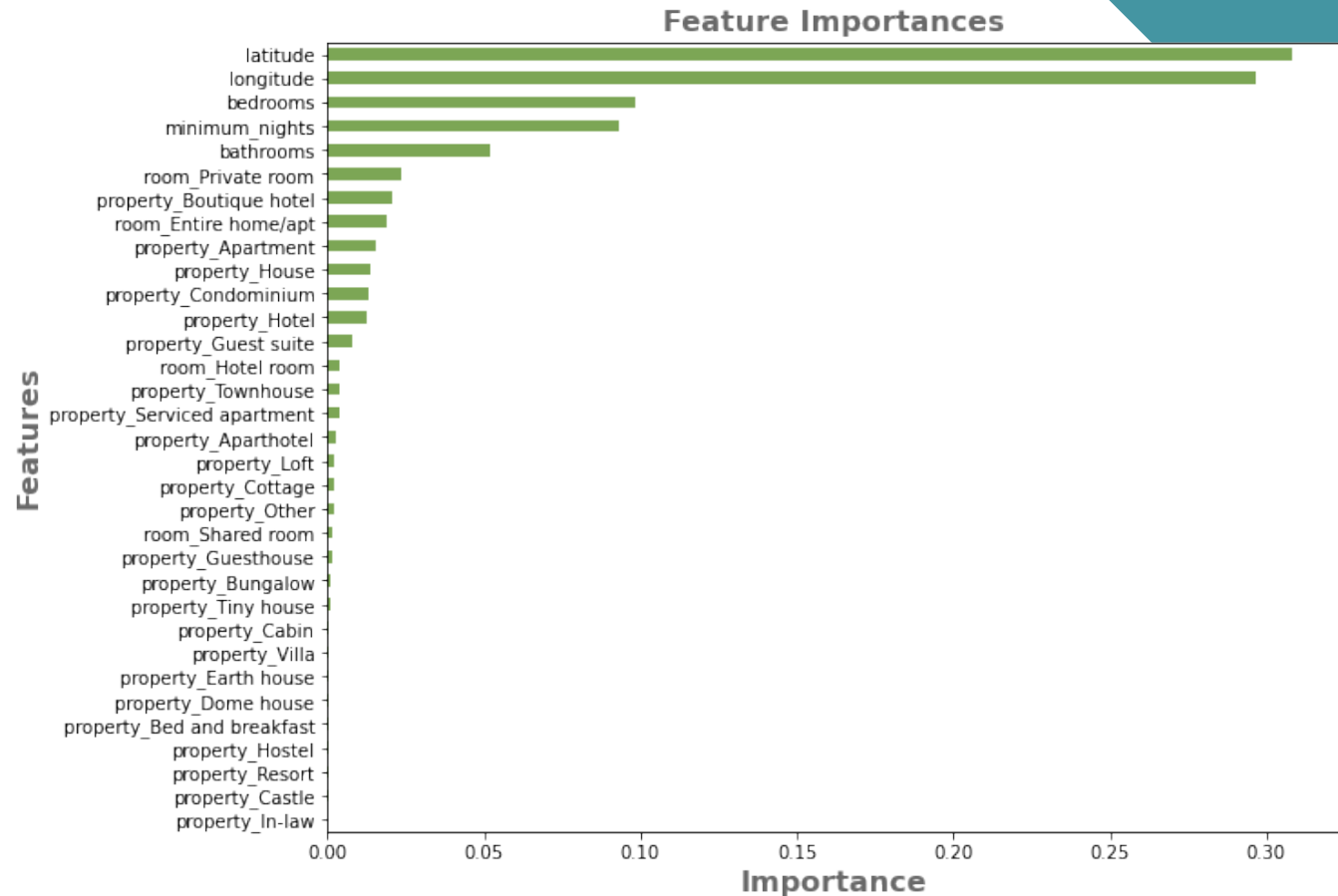
# Conclusions and Future Work

## Conclusions

### Location is Key

The most relevant parameter to calculate the price of a property is its location.

### More Bedrooms means Greater Rental Price

Properties with more bedrooms are more expensive. Bedrooms have a good contribution also to the price of the property.

### Short-stay Properties Lead the Market

Most of the listed properties have a minimum night stay of up to 10 nights.

## Future Work

### Refine the Data Cleaning Process

Refine the Data Cleaning for properties with outliers. Calculating an estimated price per $ft^2$ to improve the price estimation per location.

### Explore Overpriced Properties

Determine whether expensive properties have special features or if they are overpriced. This would help improving the models performance.

### Hyper-Parameter Tuning Refinement

To explore a broader range of hyper-parameters to train and test the model to improve its final accuracy.