



# Airbnb – London Rental Predictors

---

**Capstone Project – Sprint 2**

Benedict Medlam





# Problem Statement

*Using machine learning, how might we predict what property characteristics make for an attractive investment and beneficial home improvement, such that rental earning potential can be maximised?*

# Project Overview

## *How to invest?*

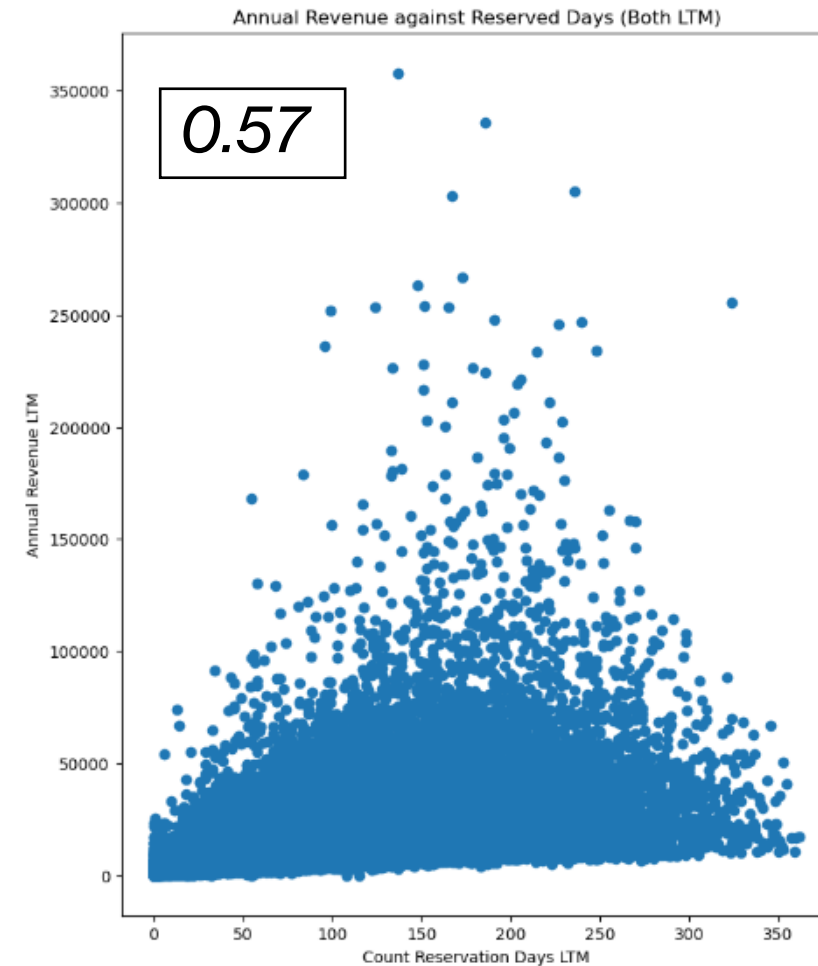
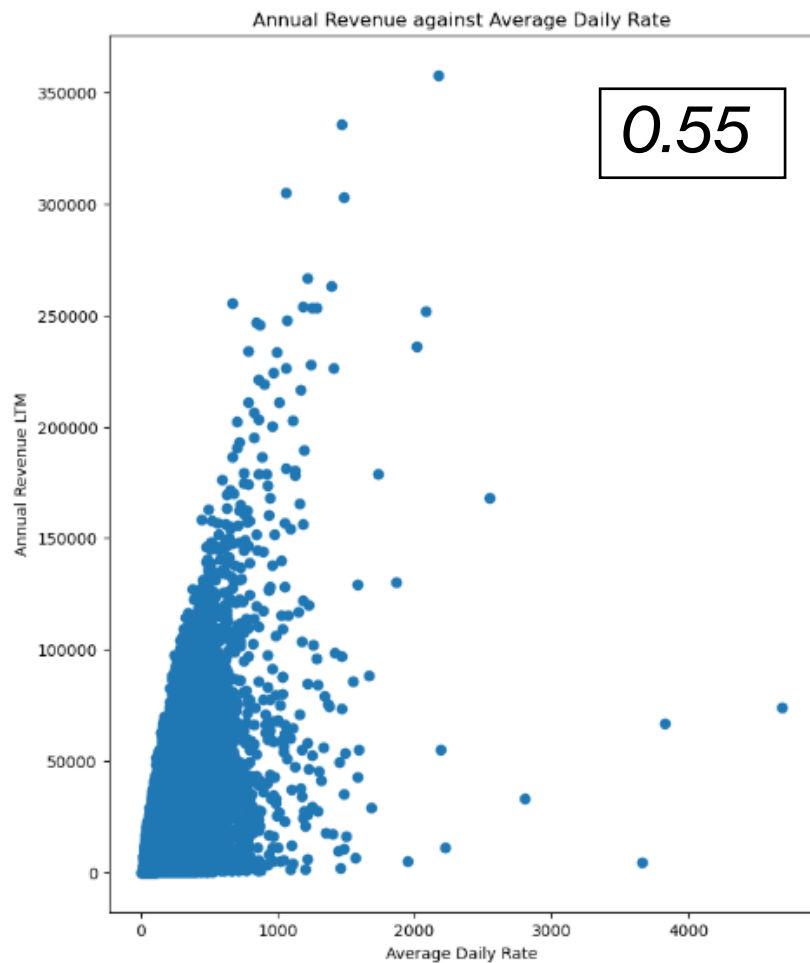
## *How to Improve?*

*Identify key characteristics that influence property annual revenue:*

- Number of bedrooms/bathrooms
- Pet friendly?
- Cancellation policy... stricter the better?

# EDA

## Feature Correlation



# Preprocessing

## *Data Transformation*

- *Checkout Time*

## *One-hot Encoding*

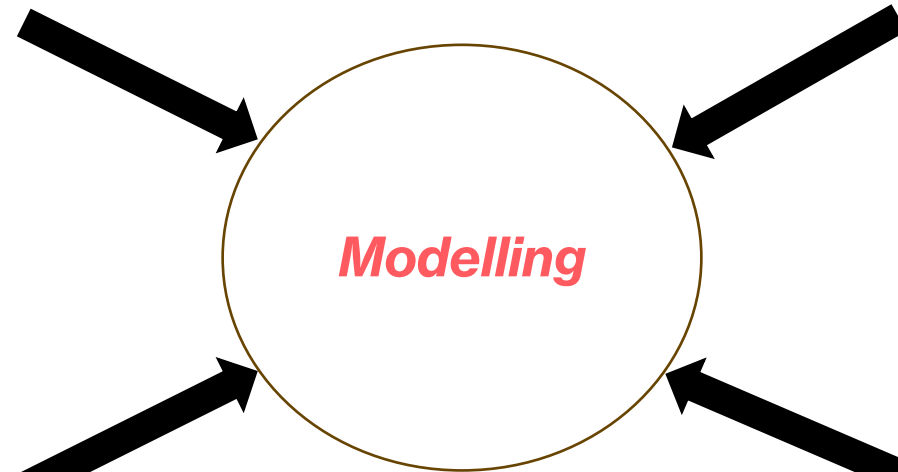
- *Property type*
- *Granular*

## *Data Granularity*

- *Cancellation Policy*
- *Checkout Time*

## *Boolean Conversion*

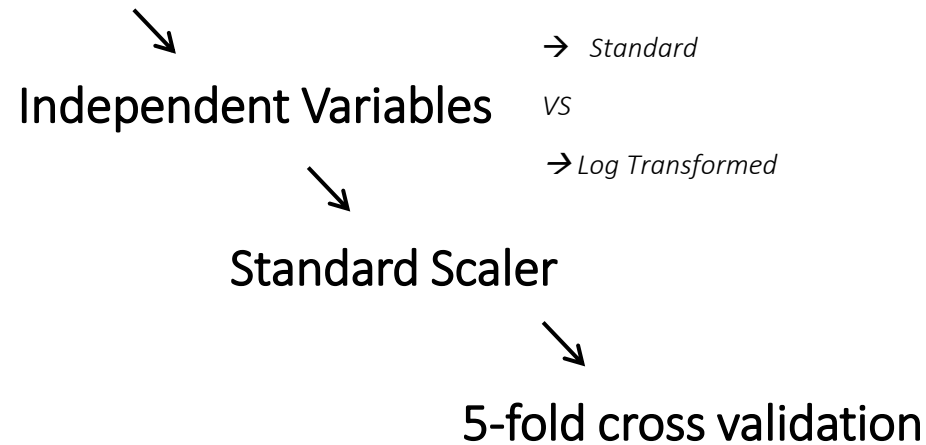
- *Checkout Time*



# Initial Modelling

## Linear Regression Model

*Continuous Target Variable*



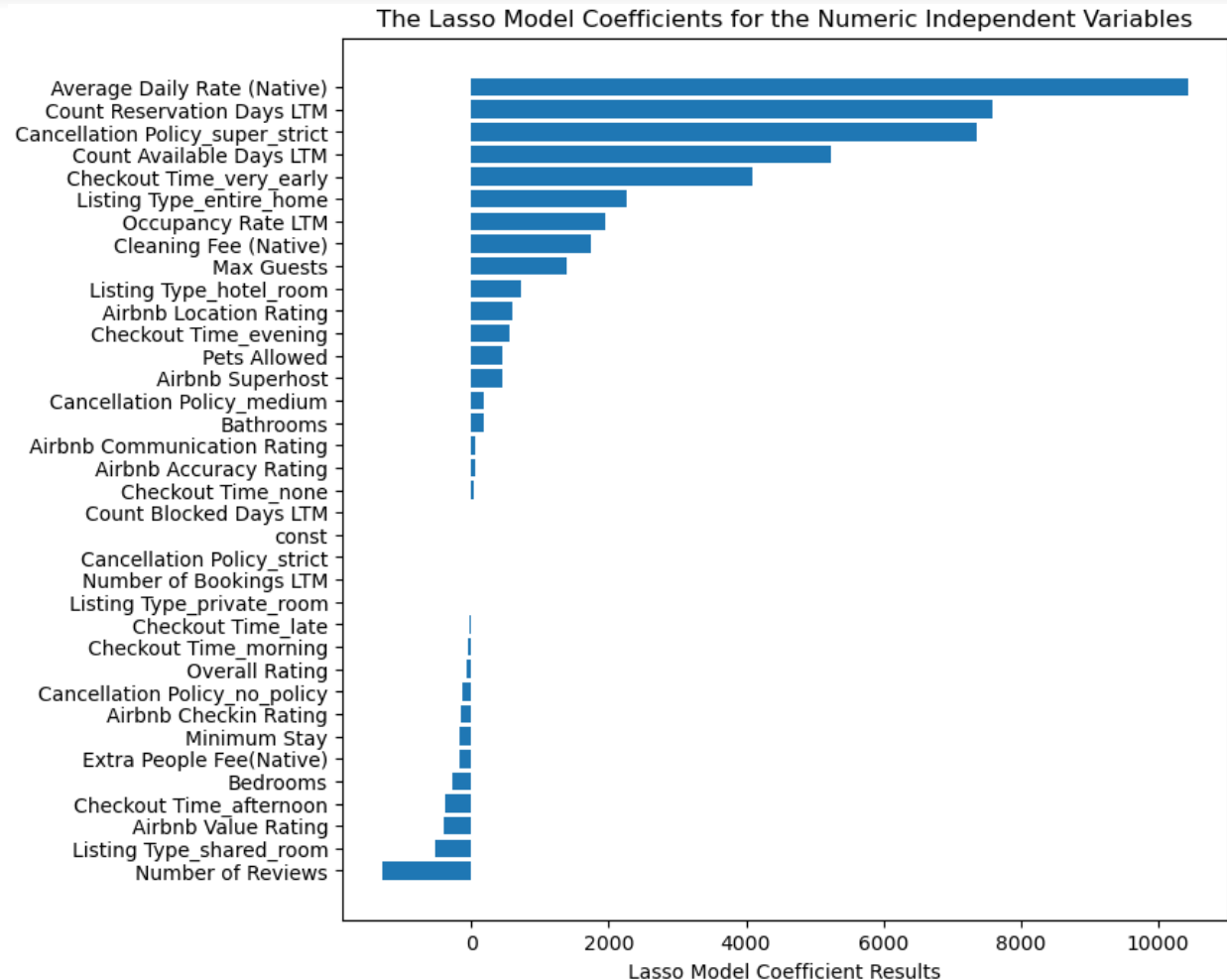
Average R-squared value:

0.7

# Lasso Model

## Linear Regression Model

- *L1 Penalty*
- *Feature Reduction*



# Decision Tree Model

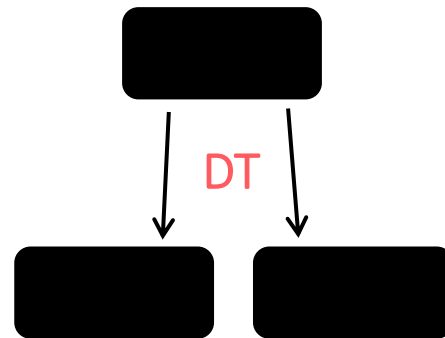
## Standard Scaled Independent Variables

### Hyperparameters:

- Max depth
- Min samples split
- Min samples per leaf
- Max features

## Grid Search Cross Validation

- 5 folds



### Model 1:

- Entire train dataset Score: 0.94
- Cross validation Score: 0.75

**OVERFITTING!**

### Model 2:

- Entire train dataset Score: 0.85
- Cross validation Score: 0.81





# Next Steps

- Baseline Modelling
  - *More feature engineering and optimisation.*
- Advanced Models
  - *Count vectorizing the numerous text categorical columns.*
  - *Input these features into the various models.*