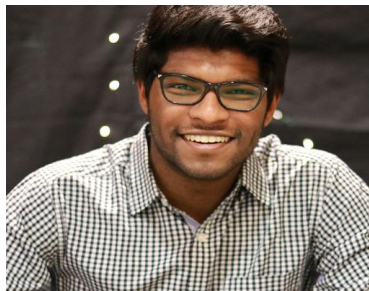# Group 1

AMII Students
Spring
2023 Session

Brett Adams

Tanvir Hossain

Ram Dershan

Cailenys Leslie

Clinton Boyda

AMII Machine Learning Technician

Group 1 Capstone Presentation

Landlord Smart Pricing

March 31, 2023

Project Objective

Find the **best machine learning model** that **predicts** the optimal **price of a rental property** to maximize revenue.

Business Case

Attract more Airbnb hosts and help them decide whether to invest in a property.

Success Criteria

A Coefficient of determination (R^2) $\cong 1$

B Mean squared error (MSE) $\cong 0$

Data Science Life Cycle

1 Data Collection

3 Data Visualization

5 Model Evaluation

Exploratory Data Analysis

2 Data Preprocessing

4 Model Building

Data Collection

1.0 Merging Data

Original Data
49k records
Limited
features

+

Additional Data
42k records
+ new features

Merged Data
16k records
8 features

Data Science Life Cycle

1 Data Collection

3 Data Visualization

5 Model Evaluation

Feature Engineering

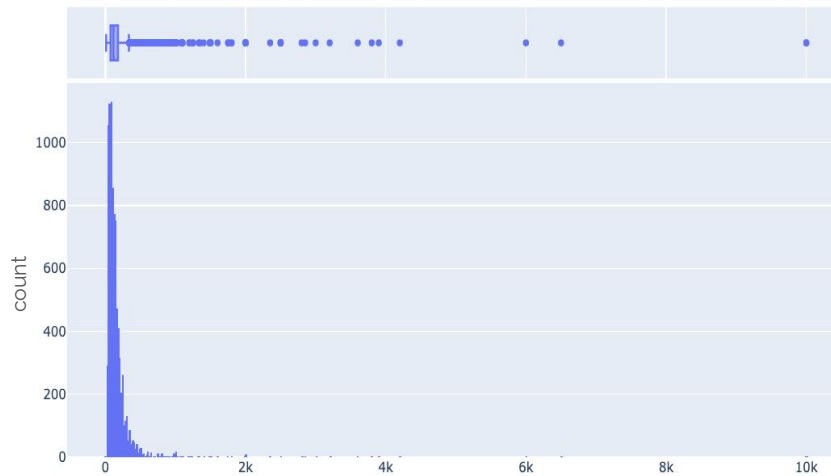**2 Data Preprocessing**

4 Model Building

Data Preprocessing
2.0 Keep only Valid Data

x Zero Price
x Zero Availability
x Extreme Values

Merged Data
16k records

InValid Data
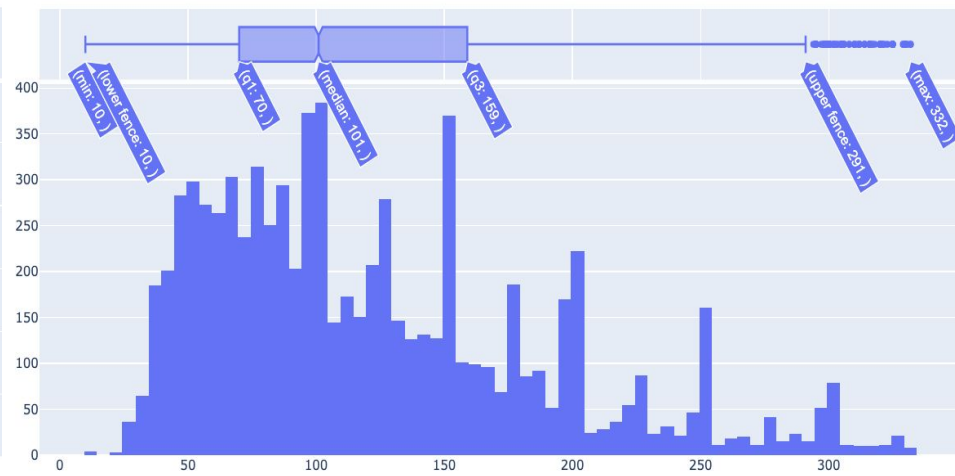8k records

Valid Data
8k records
8 features

Data Cleaning

2.1 Extreme Values

Original Distribution by Price

Cleaned Distribution by Price Comparison

Data Science Life Cycle

1 Data Collection

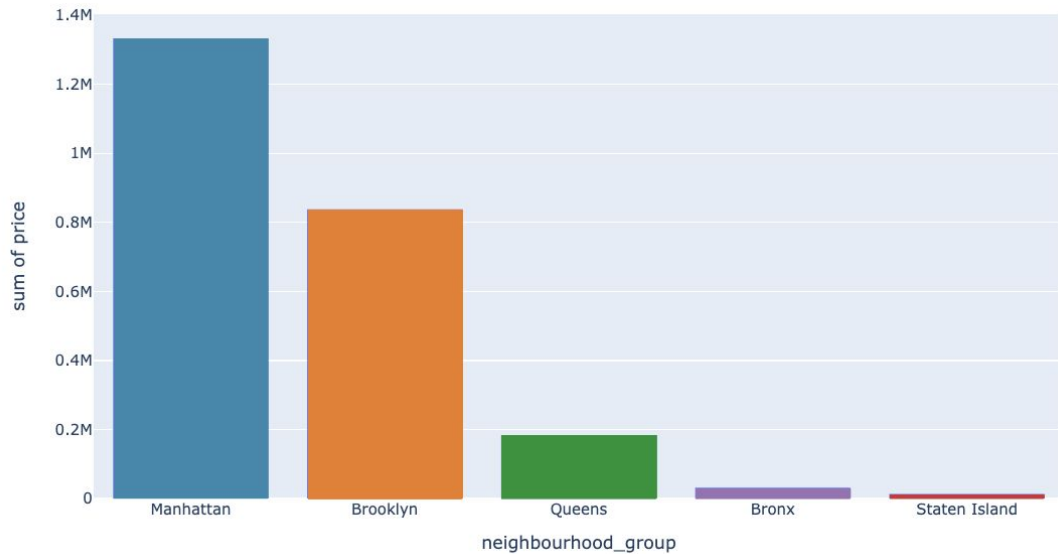3 Data Visualization

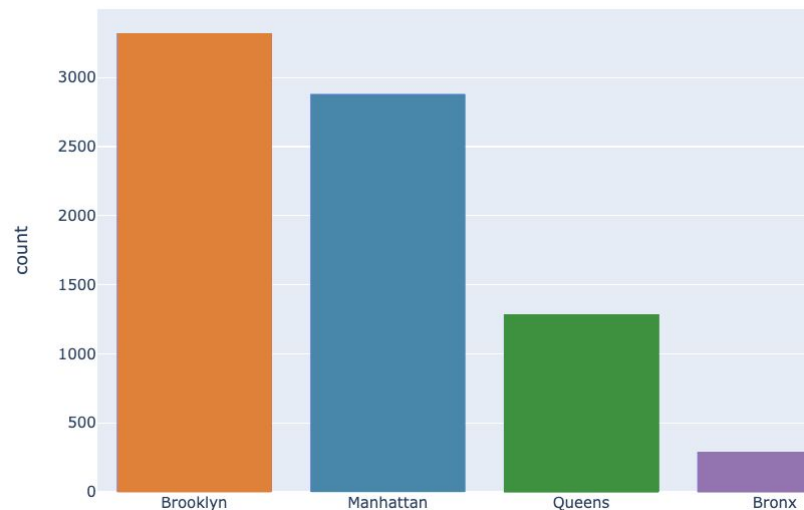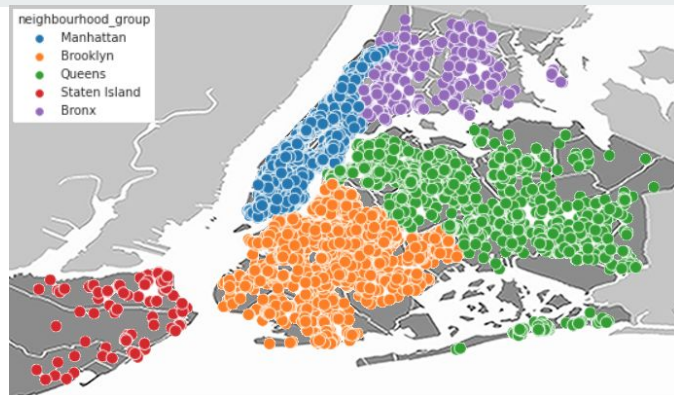5 Model Evaluation

Feature Correlation

2 Data Preprocessing

4 Model Building

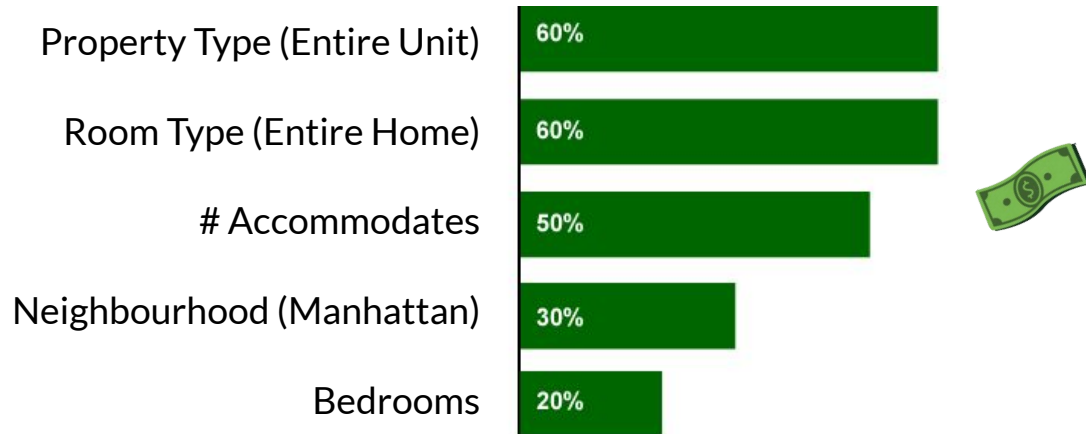# Data Exploration
3.0 Neighbourhood Locations

Data Visualizations
3.1 Correlations with Price

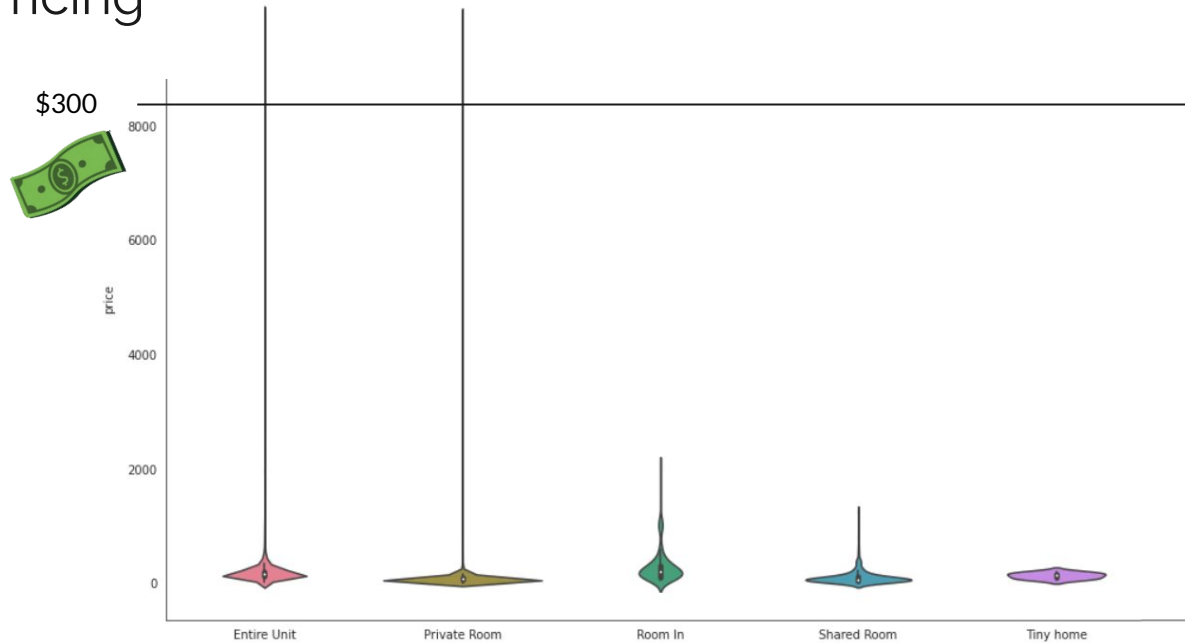|	
Property Type (Entire Unit)	60%
Room Type (Entire Home)	60%
# Accommodates	50%
Neighbourhood (Manhattan)	30%
Bedrooms	20%

Data Visualization

3.2 Property vs Pricing

Data Science Life Cycle

1 Data
Collection

3 Data
Visualization

5 Model
Evaluation

Regression

2 Data
Preprocessing

**4 Model
Building**

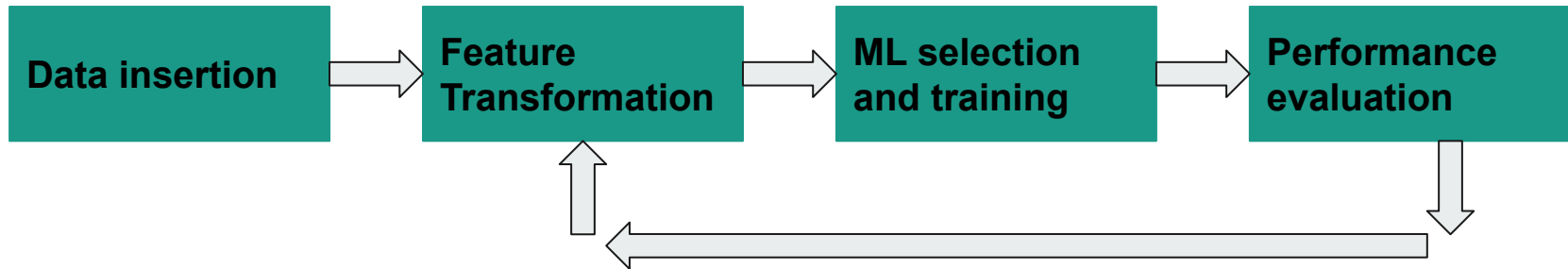# 4.0 Model Building

4.1

Model Pipeline

Feature selection

4.2

4.3

Performance evaluation

4.1 Model Pipeline

4.2 Feature selection

	Model_Name	y_train	y_test
1	Ridge()	64.2%	63.7%
4	SVR()	68.1%	63.5%
7	LinearRegression()	64.3%	63.1%
5	LinearSVR()	63.4%	62.5%
6	GradientBoostingRegressor()	63.4%	62.0%
3	RandomForestRegressor(random_state=0)	76.9%	61.3%
2	DecisionTreeRegressor(random_state=0)	78.2%	56.0%
0	KNeighborsRegressor()	65.9%	55.1%

Before feature selection

	Model_Name	y_train	y_test
6	GradientBoostingRegressor()	57.8%	56.6%
1	Ridge()	54.5%	56.0%
7	LinearRegression()	54.5%	56.0%
4	SVR()	56.6%	55.8%
5	LinearSVR()	54.1%	55.1%
3	RandomForestRegressor(random_state=0)	60.8%	55.0%
2	DecisionTreeRegressor(random_state=0)	61.2%	53.4%
0	KNeighborsRegressor()	55.0%	50.4%

After feature selection

4.3 Performance evaluation

Comparison of Models by R2 Score

4.3 Performance evaluation

Model Comparison
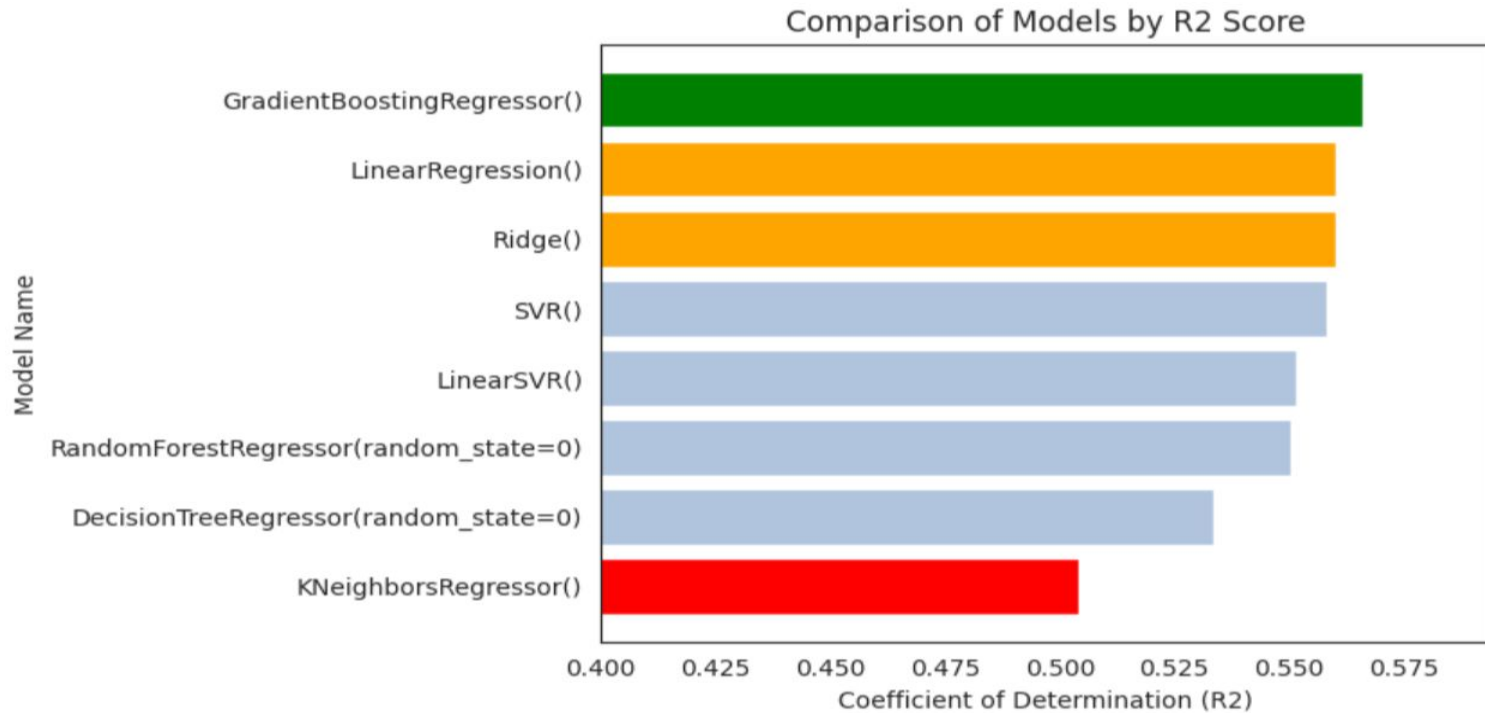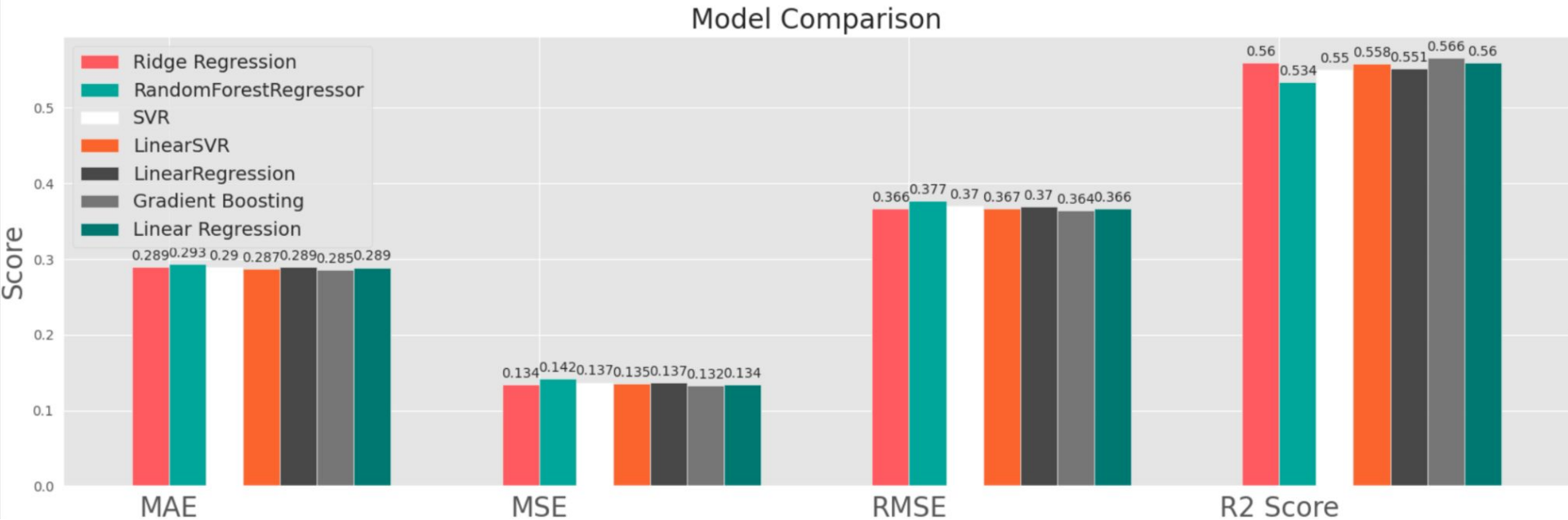
4.3 Performance evaluation

	mean_test_score	std_test_score	rank_test_score
Linear regression	0.542763	0.013050	3
Ridge regression	0.542815	0.013000	2
Gradient boosting regressor	0.562364	0.012964	1

- Minimal improvements with hyperparameter tuning.

Data Science Life Cycle

1 Data Collection

3 Data Visualization

5 Model Evaluation

Accuracy

2 Data Preprocessing

4 Model Building

Model Evaluation

5.0 Performance Results
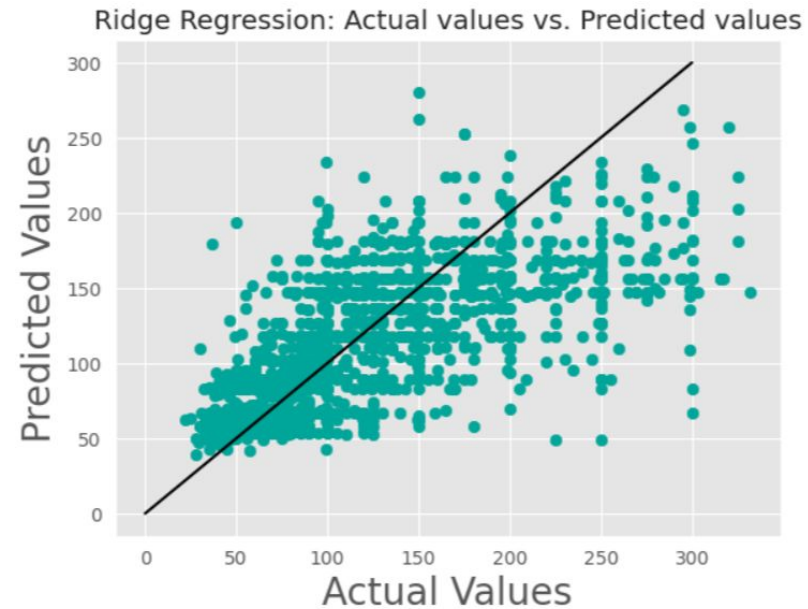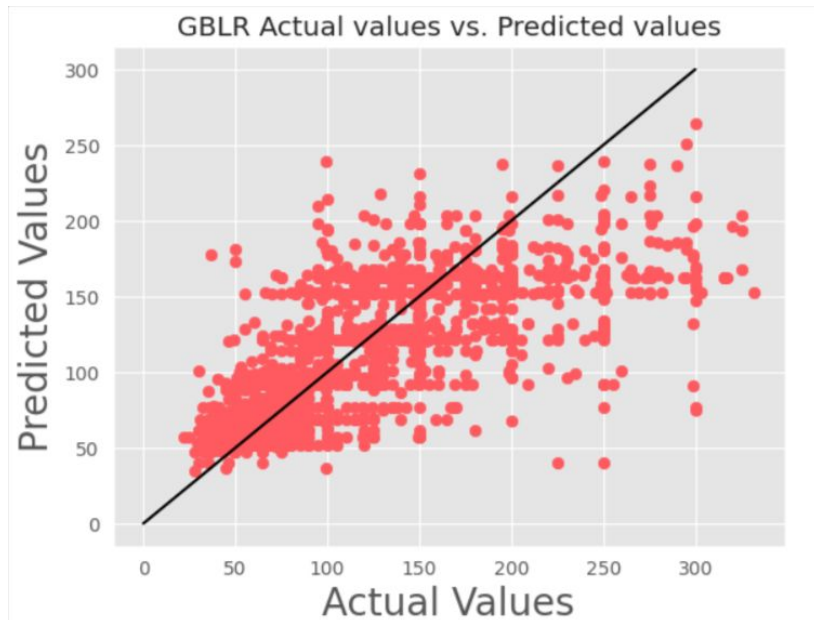
Model Name

Ridge Regression

Accuracy

56%

Model Evaluation

5.1 Actual vs Predicted

Wrap-up: Analysis Limitations

1 Regionally Specific

3 Sampling bias

2 Trendline have limitation

4 Limited variables

Potential Enhancements

1 Include additional Features

2 Pursue Additional ML Models

3 Expand Profit Prediction Potential

airbnb

Anywhere | Any week | Add guests

Airbnb your home

Cabins | Amazing views | Amazing pools | Play | Earth homes | Mansions | OMG! | Tree houses | Luxe | Domes | Boats | Lakefront

Business Summary:

1. Smart Pricing Improvement

2. Pilot Program

3. Uber Surge Pricing Equivalencies
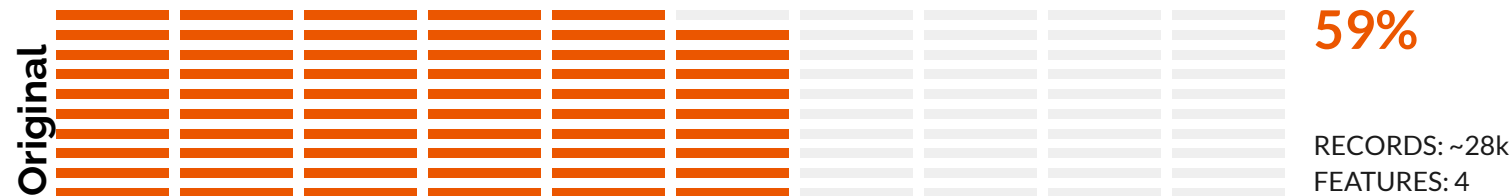
airbnb

Landlord Smart Pricing

Questions?

Optimal Model Selection
ML Insights

Merged

65%

RECORDS: ~8k
FEATURES: 8

**Merged
Features**

Original

59%

RECORDS: ~28k
FEATURES: 4

Original
Features

Pipeline Result with Neighbourhood feature

	Model_Name	y_train	y_test
4	SVR()	68.1%	63.5%
7	LinearRegression()	64.3%	63.1%
6	GradientBoostingRegressor()	63.4%	61.9%
3	RandomForestRegressor(random_state=0)	76.9%	61.3%
2	DecisionTreeRegressor(random_state=0)	78.2%	56.0%
0	KNeighborsRegressor()	65.9%	55.1%
1	KNeighborsRegressor(weights='distance')	74.7%	54.2%
5	AdaBoostRegressor()	47.8%	46.9%