Price Predictions

Puerto Rico Real Estate

Johnathan Yater

Objective:

Develop a machine learning model to predict property prices in Puerto Rico

Goal:

Assist real estate agents and buys in understanding the market value of properties in Puerto Rico. A predictive model will assist stakeholders to make informed decision when buying or selling properties

Data Insights

Attribute	Description
Total Listings	2,645 listings, all marked as "for sale".
Top Cities	San Juan (406), Carolina (140), Guaynabo (127), Bayamon (121), Humacao (95).
Bedrooms	Range: 1 to 33, Average: ~3.63.
Bathrooms	Range: 1 to 35, Average: ~2.52.
Acre Lot	Range: 0 to 100,000, Average: ~60.34 (Note: Average skewed by large properties).
House Size	Range: 181 to 1,450,112 sqft, Average: ~2,807.72 sqft.
Price	Range: $15,900 to $25,000,000, Average: ~$677,452.

Statistic	Bedrooms	Bathrooms	Acre Lot	House Size (sq ft)	Price (USD)
Count	2,113	2,127	2,116	2,083	2,645
Mean	3.63	2.52	60.34 acres	2,807.72	677,452
Std Dev	1.67	1.67	2,177.20 acres	31,810.86	1,796,442
Min	1	1	0 acres	181	15,900
25%	3	2	0.09 acres	1,012	95,000
50%	3	2	0.21 acres	1,498	170,000
75%	4	3	0.62 acres	2,483.5	465,000
Max	33	35	100,000 acres	1,450,112	25,000,000

Data Preparation

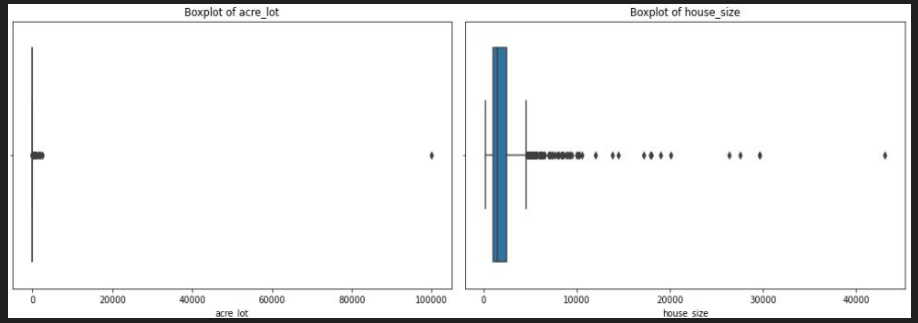
Preprocessing Steps:

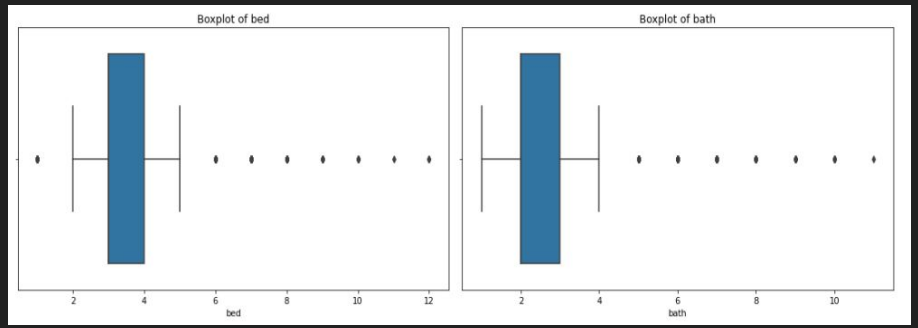
- Handled missing values and outliers
- Encoding categorical variables

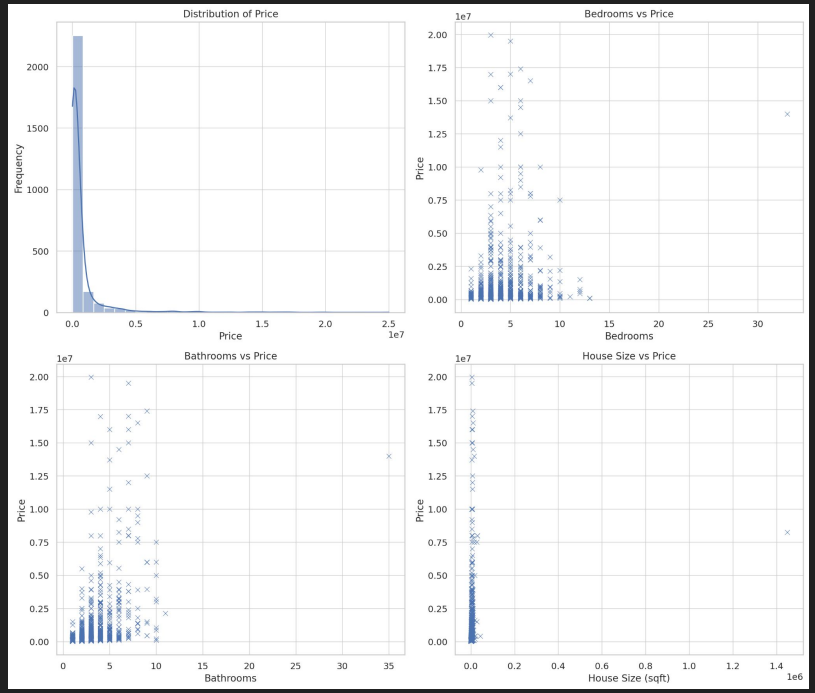
Data Exploration

- Identified features correlated with property prices
- Visualize distributions and relationship between features and target variables

Data Preparation







Model Process

Ensemble Methods

- Random Forest
- AdaBoost
- Gradient Boosting
- XGBoost

Tuning

Hyperparameter

GridsearchCV

Evaluation

- MAE
- RSME
- R2

Random Forest:

A forest symbolizing many decision trees coming together to make a more accurate prediction

Gradient Boosting:

Think of a climber ascending a mountain and improving its predictions by learning from previous mistakes/missteps incrementally.

AdaBoost:

Combine several weak learners to create a strong learner.

XGBoost:

A robot carefully building and creating, indicating an optimized and advanced version of Gradient Boosting that's known for its performance. Highlight advanced, efficient machine learning for high performance.

Tuning

Cross validation:

Ensure model accuracy
through training and testing on
different data segments.

GridSearchCV:

A magnifying glass
inspecting a grid of dials,
representing the search for
the best model settings.

Evaluation

Mean Absolute Error(MAE):

- Measures the average absolute value difference between the predicted values and actual values. Straight forward indication of the model of the models' accuracy in predicting the target price.

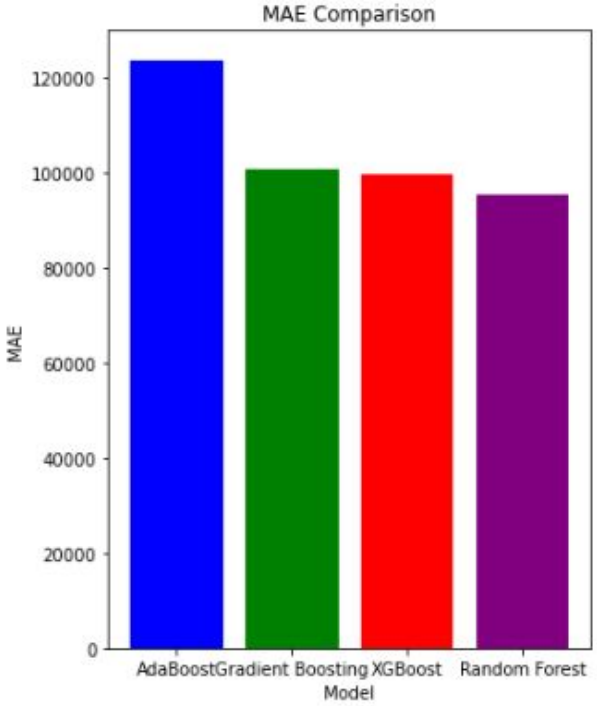
Root mean Squared Error(RSME):

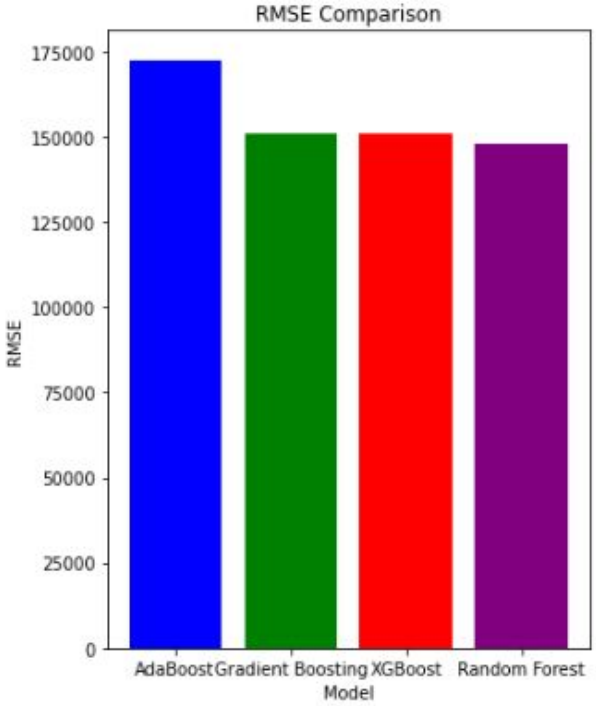
- Calculates the square root of the average squared differences between predicted values and actual values, providing a measure of the model's performance in terms of the scale of the target

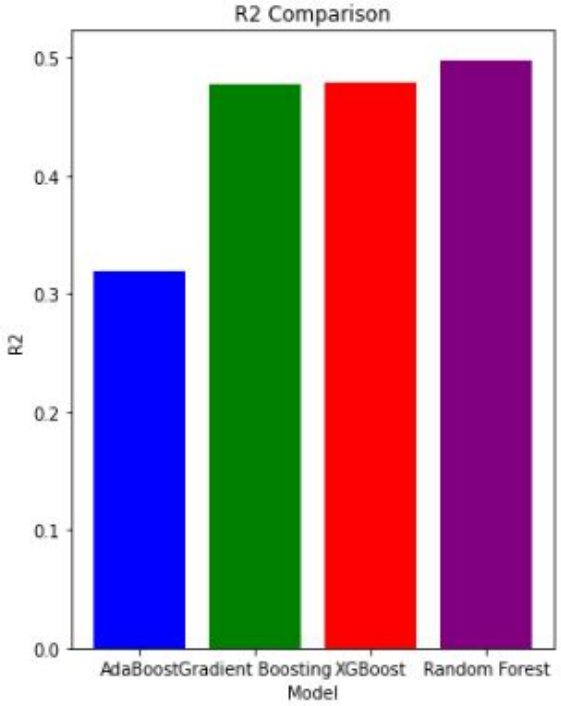
Coefficient of Determination(R2):

- Indicates a goodness of fit from a scale of 0 - 1

Test Results







Recommended Next Steps

Alternate Techniques	Experiment with deep learning models or other regression models
Feature Engineering	Identify and incorporate new features or transformations. Get more data. Explore Regions
Deploy Model	Integrate into a web application or decision system for real estate professionals and buyers

Johnathan Yater

Data Scientist

Email: korean.yater@gmail.com

Github: github.com/yaterjo