Price Prediction Model

Seattle's Airbnb Open Data

Problem Statement

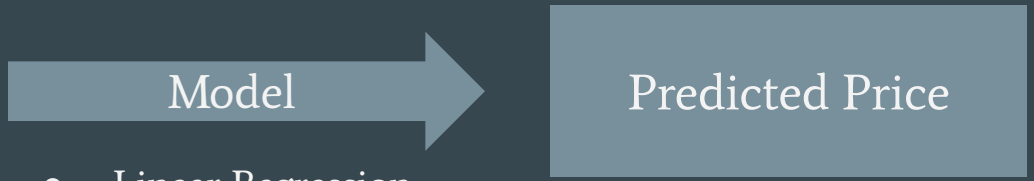
- Which features influence the price for an airbnb listing in Seattle?
- Can these features be used to predict the price?

Modeling Process

INPUTS

- Accommodates (log)
- Room Type
- # of Bedrooms
- # of Bathrooms
- Property Type
- Superhost
- Cancellation Policy
- Availability
- Cleaning Fee
- Extra People
- Security Deposit
- Location
- Minimum Nights
- Maximum Nights
- Instant Bookable
- Bed Type
- # of Reviews

Goal: Predict price for airbnb listings in Seattle.



- Linear Regression
- Decision Trees
- Random Forest
- Generalized Linear Model

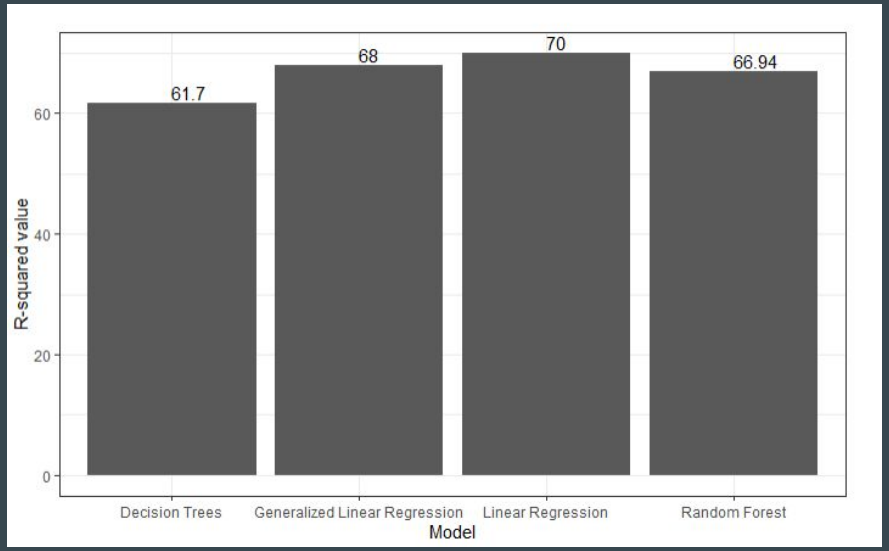
Model Assessments

Features found to be significant based on all models:

- # of bedrooms
- # of bathrooms
- Cleaning Fee
- Latitude / Longitude
- Log_accommodates
- Security Deposit
- Minimum Nights
- Room Type - Shared or Private
- # of days of Availability in a Year

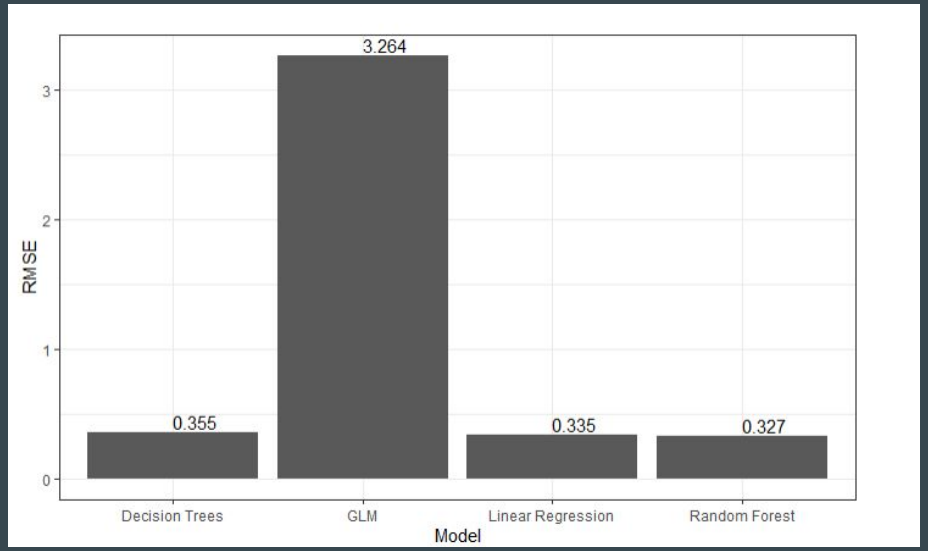
Model Results -

- R-squared Value



Linear Regression wins

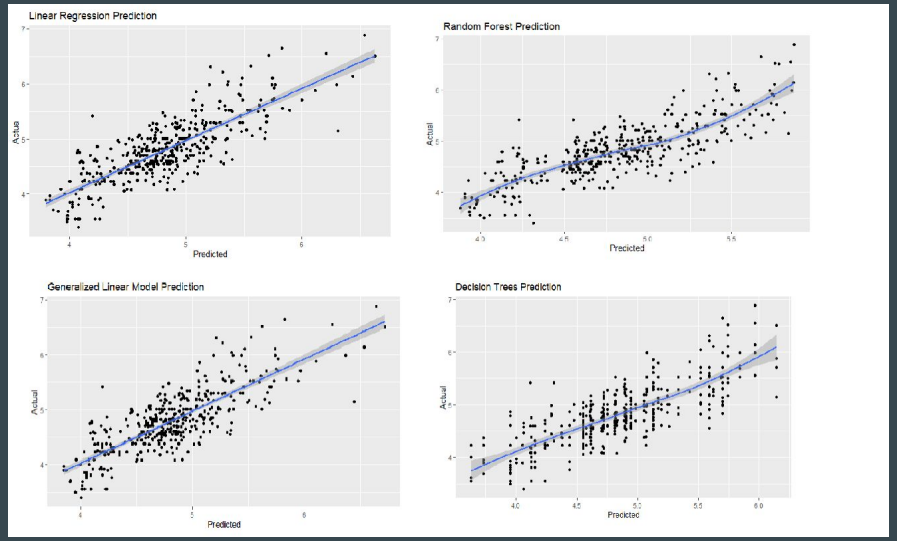
- RMSE - prediction error



Random Forest wins

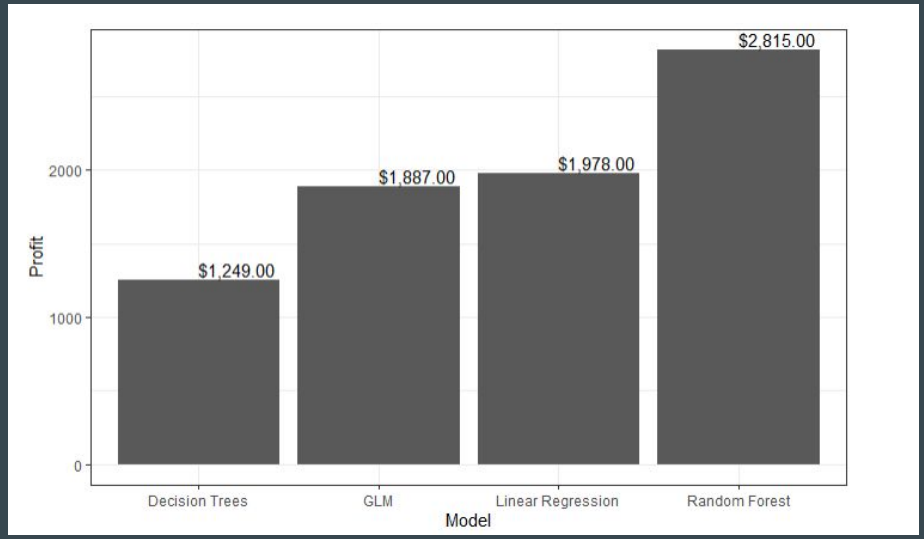
Model Results

- Actual vs Predicted Plots



Random Forest wins

- Utility Function - Profit $$$$



Random Forest wins

Next Steps

- Look at how Amenities, Reviews (NLP) and Seasonality (Time Series) affect price
- Hyper parameter tuning for Decision Trees to improve accuracy
- Some more feature engineering - like calculating proximity to Airport, Bus stops/ Light Rail
- Outlier Detection and capping at thresholds
- Try other models - XGBoost, Support Vector Regression, Neural Networks