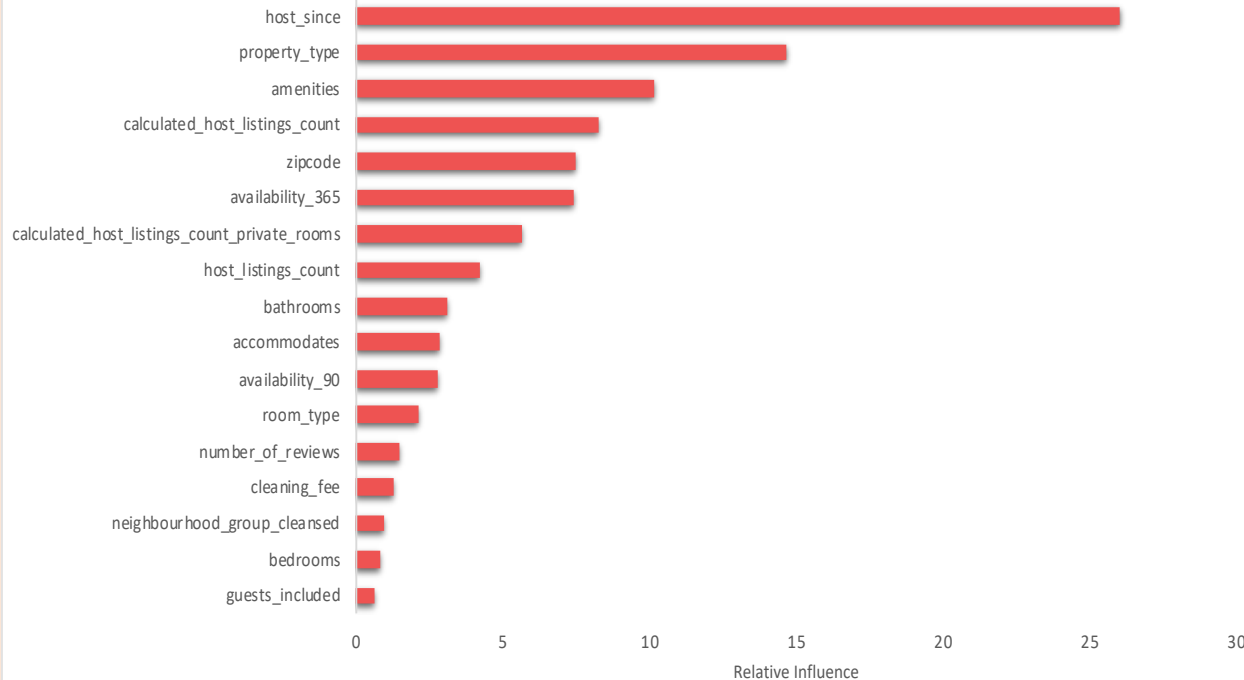




Airbnb Rental Price Prediction

Priyanka Nagpal
PN2357

Top Predictors with relative influence for best model



Data Tidying

- Deleted outliers: Negative price
- Dropped columns with major missing values
- Dealt with missing value
 - Numeric: replaced with mean values
 - Categorical: added a level to replace NA
 - Text: Few converted to numeric & replaced with mean values
 - Logical: Assumed missing values to be False

Feature Selection

- Used linear regression to check relevance of multiple variables
- Concluded with summary of gbm(boosting model) to find relative influence of multiple variables

Model & Prediction

- Finally used Random Forest model with all variables identified from boosting model

Wrong & Different

- Initially used many variables
- Should have created more dummy variables
- Try LASSO and other feature selection methods

Methods used for Predictions	RMSE on public Board	RMSE on private board
Linear Regression(different combinations of variables)	339.48753	300.95064
Own estimate of important factors	274.09158	247.74478
Boosting Model	265.70942	238.97654
Random Forest	237.48946	217.13461