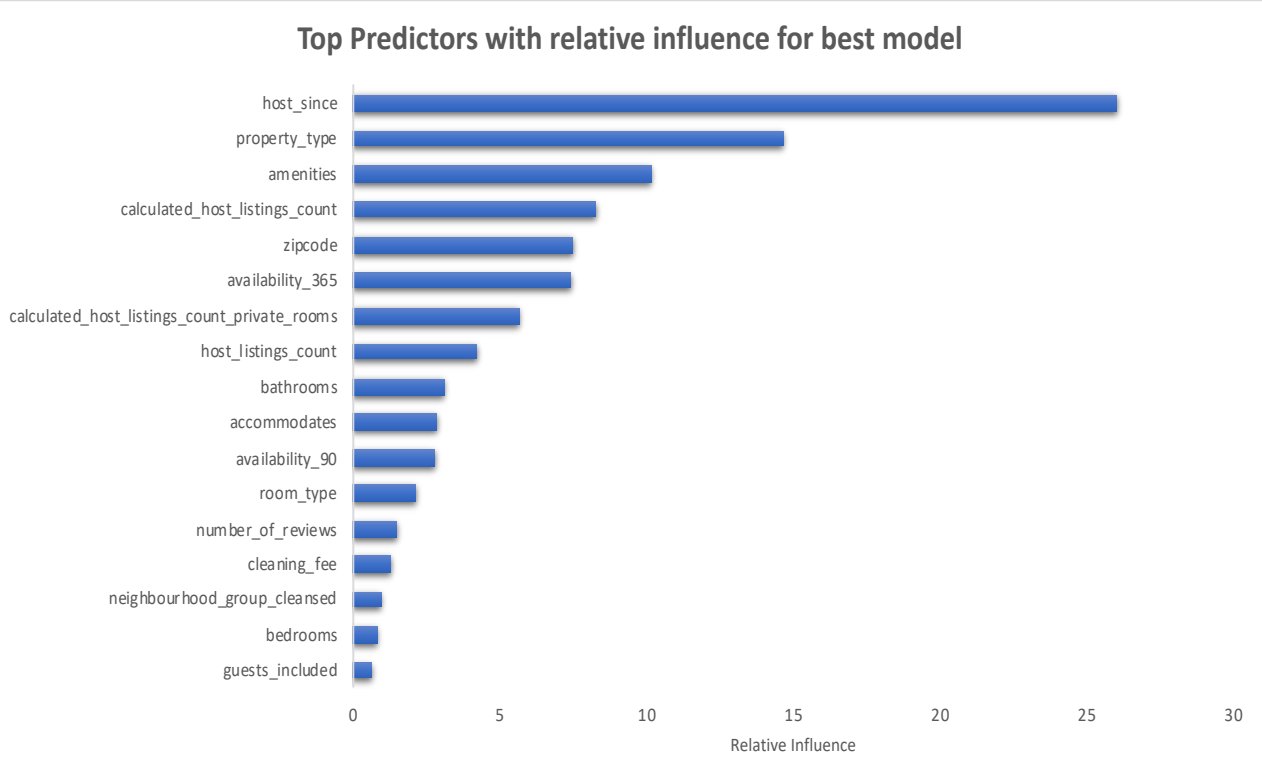
Airbnb Rental Price Prediction



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

What Worked:

- Identifying patterns of data
- Using stepwise selection of features
- Removing repetitive features
- Using minimum number of features with high predictive power

Surprising elements:

- Square foot and reviews has less prediction power
- Other variables fitted better

Improvements:

- Grouping of variables like neighborhood group with neighborhood group cleansed etc.
- Converting categorical to numerical variables

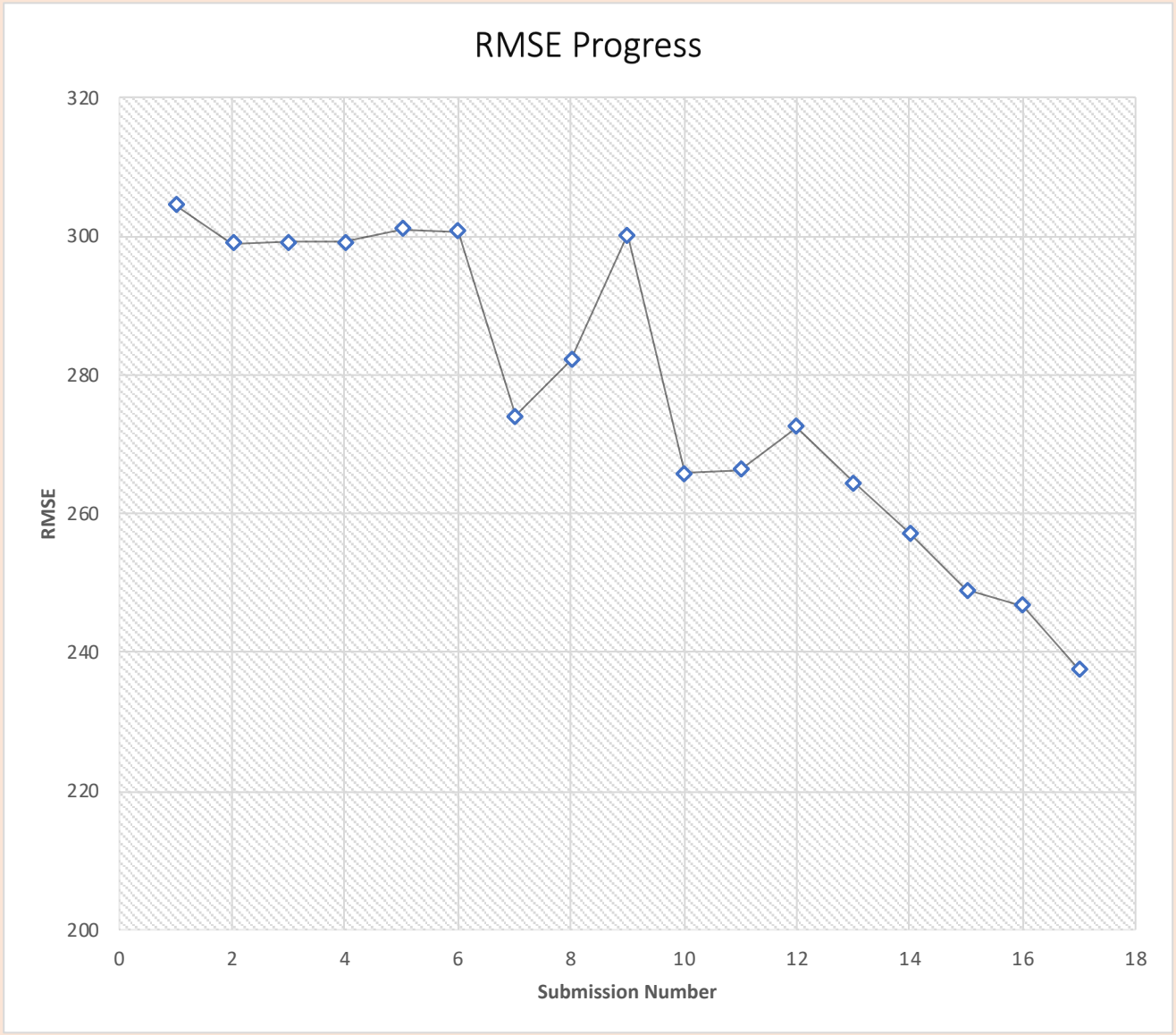
Airbnb Rental Price Prediction

What I did Right?

- Converted text variables as numeric such as amenities
- Evaluated correlations
- Tried different models(Linear,Boosting,RandomForest,Boosting with Cross Validation)
- Boosting model suggested multiple variables with their relative influence which helped to remove redundant variables.

What I did Wrong?

- Treated and cleaned many variables afterwards
- My approach was reverse for selecting features.
- Did not use Lasso and other feature selection methods
- Did not create much of dummy variables which could have resulted in better modelling



What I would do differently?

- Begin by cleaning all variables first
- Explore data variables by feature selection method
- Create more dummy variables like major amenities etc.
- Trying more combination of other variables