RMSE Results of Model

Legend: RMSE (Private Score), RMSE (Public Score)

Model	RMSE (Private Score)	RMSE (Public Score)
Random Forest	89.09	80.85
CV Boost	66.45	73.63
GBM Boosting	62.09	69.93

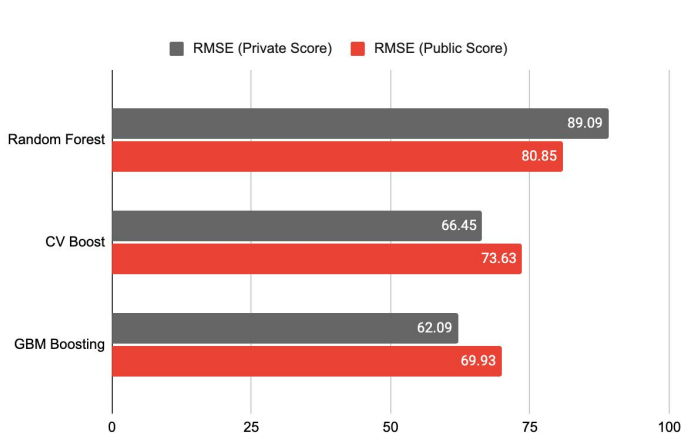## Data Transformation

Converted char & num variables into factors
Created dummy variables for amenities variables
Imputed missing integer values with the mean value
Transformed date variables into length

Variables Selection

Select only variables with the highest coefficient values to reduce overfitting
15 out of 22 variables are correlated to price
Best variables: neighbourhood_group_cleansed, property_type, zipcode

Best Machine Learning Technique

GBM Boosting with 15 variables
N.trees = 10000, interaction.depth = 5, shrinkage = 0.005