Predicting Airbnb Rental Prices

Harsh Dhanuka, hd2457

**90 features
36839 rentals**

RMSE: 56.3614

STEPS, MIS-STEPS & LEARNING…

What did I do right ?

- Removing Outliers & Astronomical Values
- Determining Significant Variables
- Imputing Missing Values
- Grouping and Creating New Variables
- Tuning Model Parameters to get best output

Explore Data

STEP 1

90 columns, identify important variables, ignore repeated variables, understand string columns, identify 3 types of NA's

Clean Data

STEP 2

Price=0, Outliers in 4 columns, Replace NA's with median/imputation, Convert string's to word count, Revalue 't' and 'f' to Logical, Clean Zipcode

What did I do wrong ?

- Initial approach was un-organized
- Kept manipulating variables throughout
- Ignored some important variables initially
- **Spent too much time on conventional models**

Feature Engineering

STEP 3

Uppercase in Name, Keyword count luxury/ious, List of Host Verifications & Amenities, Split Amenities to Wifi, TV, etc., Date – count days till today

Model

STEP 4

Hybrid Stepwise, Lasso, Ridge, Random Forest to understand feature importance
Random Forest, GBM, Ranger, XGBOOST to predict

Doing it all over, what would I change?

- Use organized approach, **NOT GET EXCITED**
- Use high performance models, **SAVE TIME**

"Re-do models with different parameters"

Display Name: **apan_wizard123**
Username: **harshdhanuka1994**

Leaderboard: **#11**
Submissions: **35**

kaggle