KAGGLE PROJECT REPORT

By Shiyu Hua

- Data Preparation

  - Filter price > 0

  - Impute missing value with median method

  - Extract popular amenities using grepl() function

- Best technique

  - Tuned random forest

    - Best tune hyperparameters with 10 fold cross validation

    - Ranger method and generate 500 tress

  - Top 3 impactful variables: room type, neightborhood_group_cleansed (5 boroughs of NYC), accommodates