Revenue Prediction

Nayana Yeshlur, Korosh Moosavi, Michelle Dea

Outliers

- Revenue data uncovered outliers
  - Minimum nights stay
  - Price per night
  - Calculated revenue

Icons from Freepik on flat icons

Data Refinement

- Removing arbitrary data fields
  - Mainly identifiers
- Iterating on methods to deal with missing data
  - Imputation: Ffill
- Refining predicted data (eliminating negative values)
  - Replace negative values with 0
- Removing additional Outliers
  - Filtered out listings with no revenue data
- Feature selection refinement
  - Removing features with no data

Model Refinement

- Grid Search to determine optimal parameters for XGBoost
- Testing different boosting models (XGBoost, CatBoost, LightGBM)
- Cross Validation vs train_test_split
- Measuring MAE, RMSE, and MAPE
- Testing Decision Tree Regressor

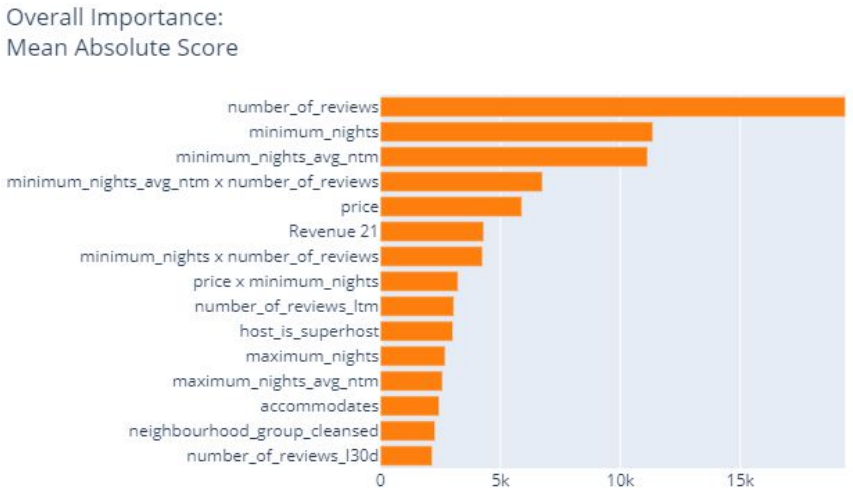
Experiments

SelectKBest

- Optimal K = 3
- Target variable is calculated from other feature values
- Gradual decline in performance past 3 features

EBM

- Top features are those related to the values used to calculate target feature
- Most valuable information is the unrelated features
- Shows why K=3 is optimal



Results

XGBoost

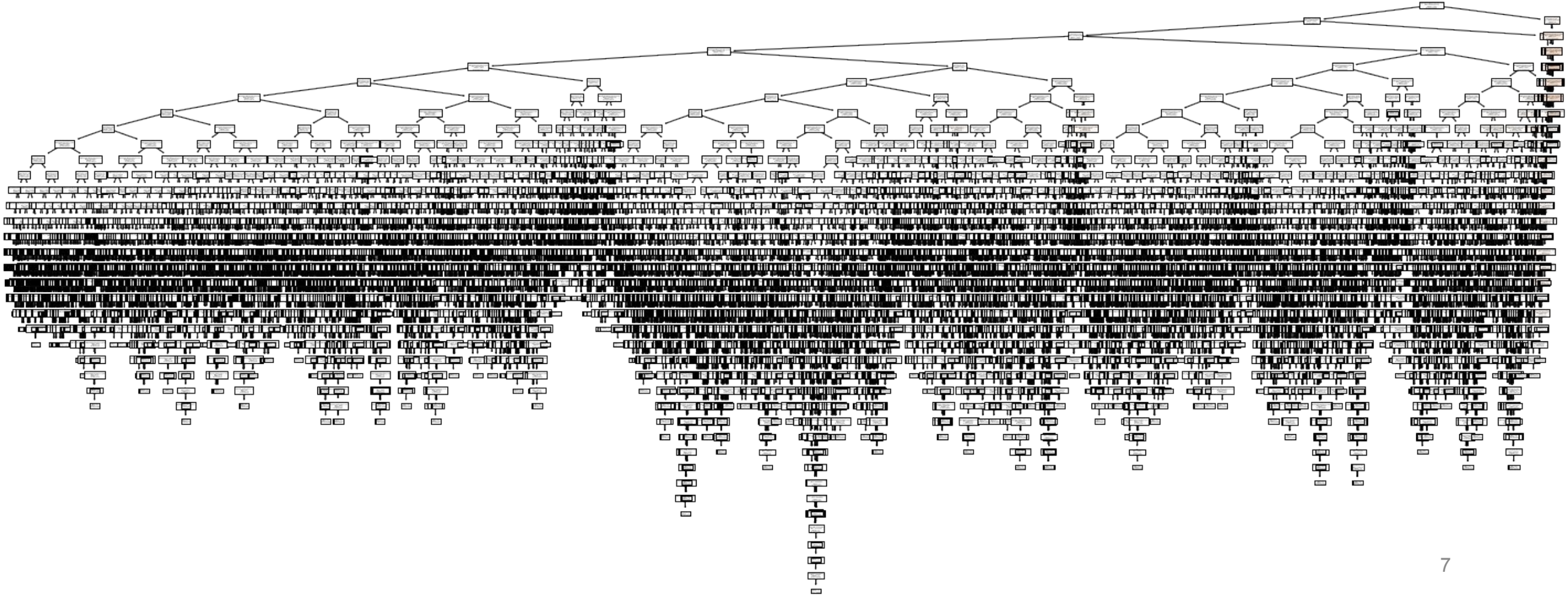
- Before filtering outliers:
  - MAE: 9207.66574080737
    MAPE: 7.97036518365e+18
    RMSE: 35191.29309886487

- After filtering outliers:
  - MAE: 14446.24281806402
    MAPE: 1.08290997858
    RMSE: 75282.93038002329

10 Fold Cross-Validation

Model	MAE	MAPE
Linear Regression	30347.676085727533	15.242610796470808
Naive Bayes	25186.611076990906	2.4381818399073247
Decision Tree	17199.31910313236	0.8261949540654229
SVM	24423.217896518207	4.350827790997944
XGBoost	14902.110080716859	1.6023081726915318
CatBoost	13972.93044822585	1.1835342076432054
LightGBM	15119.178272410409	2.242189803561159

The Decision Tree



Next Steps

- Make feature set more robust and expand feature set
  - add cost/value of property
  - add socioeconomic data (census data)
- Analysis in more cities
- Expansion to other apps (such as Vrbo)

Albert Einstein: Insanity Is Doing the Same Thing Over and Over Again and Expecting Different Results

Machine learning: