Occupancy Level Classification, Revenue Management

Gulay Samatli Pac

2021 Airbnb statistics from CapitalCounselor

- **Hosts** earn approximately **$924** per month.
- The average Airbnb **stay duration** in 2020 in the US grew by **74%**.
- **New hosts** that joined the platform after the pandemic outburst earned **$1 billion**.
- The average Airbnb **occupancy rate** is estimated to be **45.5%** in April 2021.
- Airbnb has **2.9 million hosts** from all parts of the world.
- The US has the highest average Airbnb **host earnings** in the world — **$18,285**.

GOAL

Help hosts in optimize their listings to increase occupancy

IMPACT

★ Improve listings based on the most critical features

★ Achieve higher occupancy

★ Increase host and platform revenues

Methodology

Data	Data Cleaning & EDA	Baseline Models	Model Selection & Tuning
NYC Listings Dataset, InsideAirbnb >30K listings >40 features	**Target** : Occupancy Level -average number of reviews per month -minimum required nights.	SVC, KNN, Bagging, RandomForest with RepeatedStratifiedKFold CV Scoring : fbeta_weighted(beta =0.35), balanced_accuracy	KNN, RandomForest, XGBoost With RepeatedStratifiedKFold CV GridSearch for tuning Scoring: fbeta_weighted, balanced_accuracy

Features

01	Host	• Host response time • Host since • Host superhost ...
02	Location	• Neighborhood • Latitude, Longitude • Neighborhood overview...
03	Listing	• Property type • Rooms, bathrooms • Accommodates ...
04	Rental	• Availability 30, 90 , 365 days • Min/ Max nights • Instant Bookable • Price ...
05	Guest Experience	• Reviews • Scores • First/Last review ...

Target Variable

Estimated Occupancy Rate

Target : Occupancy Level

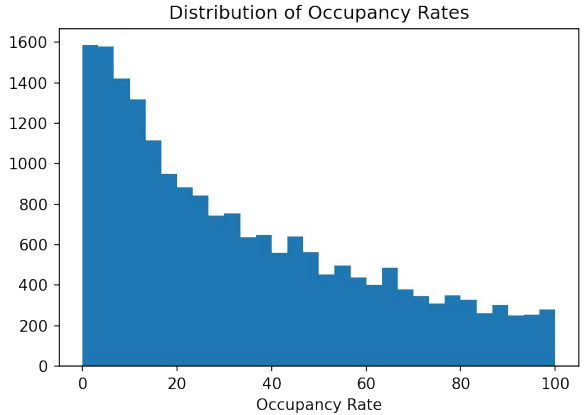
Occupancy/month =

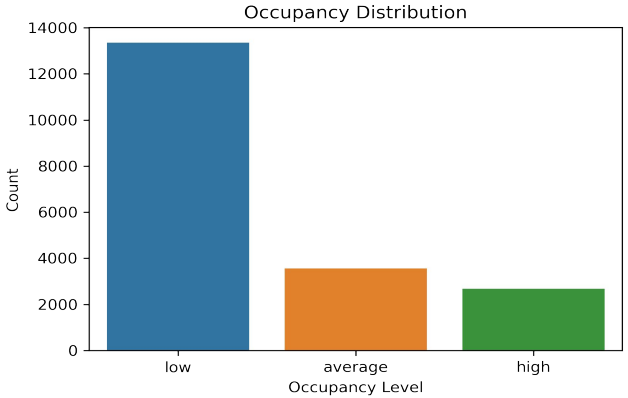
(Reviews/month) * (% stays with reviews) * (Minimum night/stay)

Occupancy Rate	Occupancy Level
<47%	Low
48% - 70%	Average
>70%	High

Relevant statistics used to categorize occupancy rates:

- The average Airbnb occupancy rate in LA are 63.4 % in 2018, 65.3% in 2019.
- The U.S. average occupancy rate is about 46.% by April, 2021.
- An ideal occupancy rate for hotels is between 70% and 95%.





Modeling

- Total 27 Features :
  - original features: accommodates, bathrooms, bedrooms, room type, host since, first review since
  - created features: amenities count, host response rate category, description length, Kitchen, Parking
- KNN, RandomForest, XGBoost
- Baseline models:
  - with selected promising variables
  - with all variables
- Backward and Forward Selection Methods, various combination of features depending scores and feature importances
- Kfold cross validation --- GridSearch
- Final Model: XGBoost Model with 22 features
  - Fbeta = 0.59
  - Balanced Accuracy = 0.46

Final Model Confusion Matrix

XGBoost Confusion Matrix

True Class

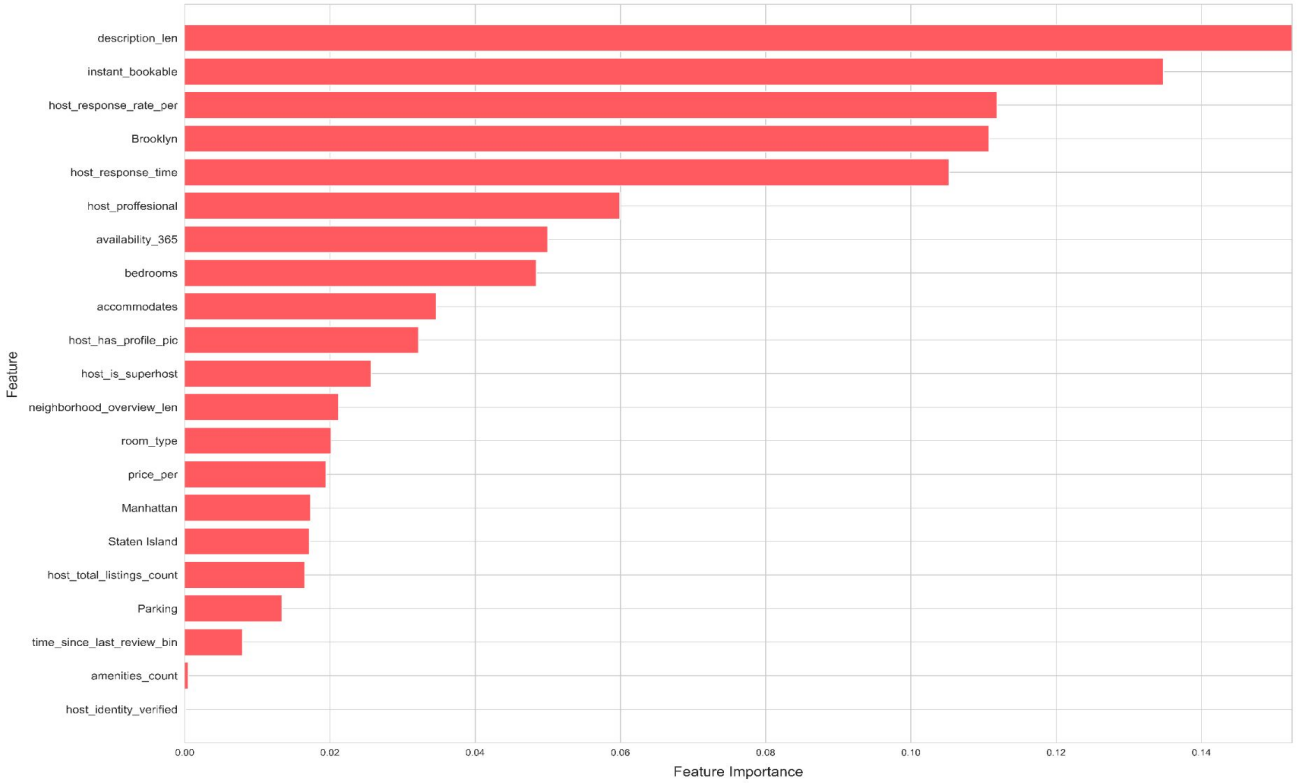
low	**1680**	529	466
average	252	**220**	212
high	178	141	**237**
	low	**average**	**high**

Predicted Class

	Low	Average	High
Precision	**0.80**	0.25	0.26
Recall	**0.63**	0.32	0.43
Fbeta	0.73	0.27	0.30

Feature Importance and Recommendations

- **Listings related:**
  - details description
  - instant bookable
- **Host related:**
  - response time
  - response rate
  - superhost
  - total listings of host
- **Rental related:**
  - amenities



Next Steps and Questions?

- Incorporating more relevant features
  - Guests reviews
  - Calendar of events
  - Neighborhood scores
  - Competitive market analysis
- Implementing different methods for imbalanced data

Thank you!