Analysis of Housing Leasing Market and Research on Optimization Strategy

Chenyu Jiang Jinhao Liu Rui Zhang

cjiang232@wisc.edu jliu2449@wisc.edu rzhang534@wisc.edu

Shumeng Fang Siyu Wang

sfang58@wisc.edu swang2379@wisc.edu

1. Introduction

Our research report conducts an exhaustive analysis of the housing leasing market in Hong Kong, tackling two principal inquiries: Firstly, what are the determinants that affect rent levels? Secondly, which predictive models most accurately forecast future rent levels? We systematically mapped rental supply and pricing trends across various districts, and conducted a detailed examination of how variables such as room type, the number of bathrooms, and bedrooms influence rental prices. Enhanced data accuracy was achieved through meticulous data cleaning and the management of missing values. Employing both regression and ensemble methodologies, including Linear Models, Random Forests, and Gradient Boosting Regression Trees (GBRT), we projected future rental levels and pinpointed critical factors influencing prices.

2. Background

2.1 Motivation

With the acceleration of the urban construction process, the tenant group is expanding, which drives the hot housing rental market. Thus, we would like to find out which factors affects rental prices by quantitatively analyzing them. What’s more, we wish to build a suitable model to predict rentals.

We chose Hongkong city as the research area of this paper, since it one of the most densely populated regions in the world with a high number of rental occupants.The rental housing market in Hong Kong features a unique mixed system of public and private housing, thus analyzing rental data in Hong Kong is expected to yield particularly distinctive results.

2.2 Methods

Data imputation: handle missing values
One-hot: encode for non-numeric data
Linear model, random forest and GBRT: find the best model
Feature importance: select suitable features
Grid Search Validation: find the best parameters

3. Data

Data is from Airbnb, a community website for traveler's home rentals. http://insideairbnb.com/get-the-data/

Original data has 5933 huose and 75 features. To ignore the useless information, we only choose the features with more than 250 frequence. And to avoid Multicollinearity, we select the feature whose VIF is less than 10.

4. Result

We attempted three different models: linear regression, random forest, and GBDT, using grid search for hyperparameter tuning. Initially, the performance was not good, with most scores even being negative. After ruling out potential data issues and plotting real data against predictions, We discovered that the main issue was outliers. (Figure 1) We suspected that some rental rates might be set beyond reasonable ranges, so we removed the outliers and refitted the data into the models. (Figure 2)

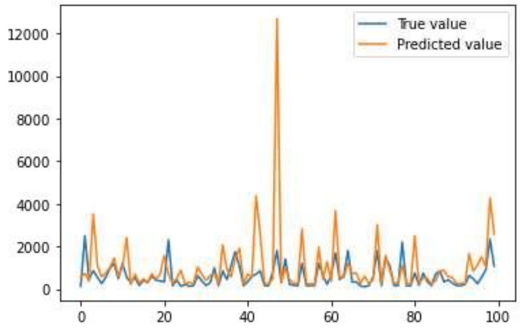

Figure 1

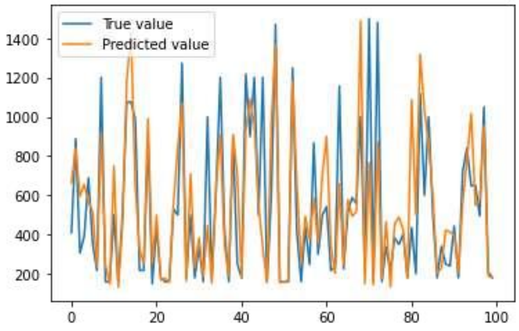

Figure 2

Comparing the RMSE and R^2 scores again, the models' accuracy improved significantly, and the RMSE decreased considerably. From the table 1, it's evident that random forest performed the best among these three models, with the following parameters: max_depth=90, n_estimators=300.

	Linear Regression	Random Forest	GBRT
RMSE	230.40	181.41	193.97
R^2	0.59	0.75	0.72

Table 1

Observing the top five important factors derived from all three models, they were quite similar and mostly aligned with common sense.(Figure 3,Figure 4) However, the interesting finding was that the hairdryer ranked very high among the factors. Therefore, we further explored it by placing it as the dependent variable (y) and fitting other factors to it.(Figure 5,Figure 6)

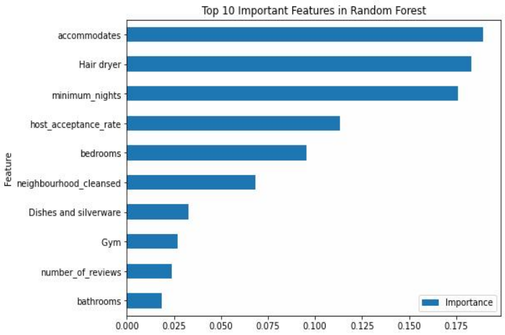

Figure 3

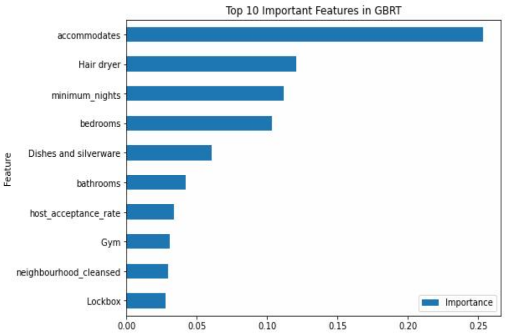

Figure 4

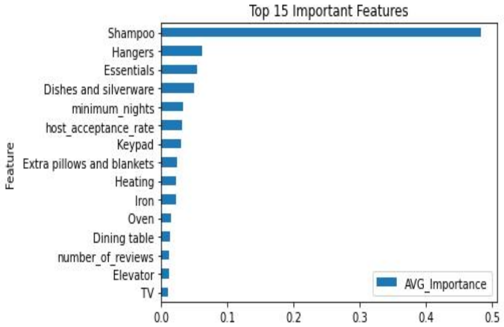

Figure 5

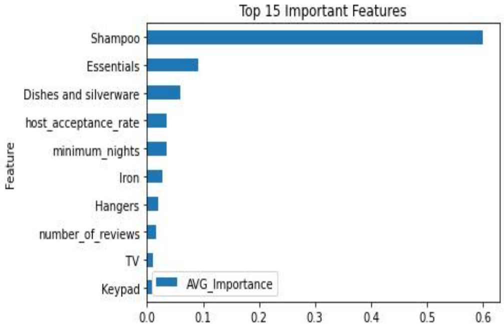

Figure 6

The shortcomings lie in the fact that most of the factors fitted by the model may exhibit correlation, and pricing is a subjectively influenced issue, also related to landlords, which the model cannot fully explain.

5. Conclusion

In general, there are three main conclusions:

1. According to R square, the best model is Random Forest.
2. Based on the models, the top 5 features are: *Accommodates, Hair dryer, Minimum nights, bedrooms, Host acceptance rate or Dishes and silverware.*
3. For the unexpected result *Hair dryer*, we find the correlations between hair dryers and other significant factors are strong. Therefore, having a hair dryer also implies an overall good rental environment.

Based on conclusions, future works are following:

1. Guidance for the housing rental market. For landlords, they could use it for pricing. For tenants, they can find their reference.
2. Maintaining order in the rental market. Management personnel can identify abnormal cases and then take corresponding measures.

3. Incentivize landlords to improve their properties. Landlords could focus on improving the quality of their properties based on primary influencing factors.

6. Contributions

Member	Proposal	Coding	Presentation	Report
Chenyu Jiang	1	0.9	0.7	1
Jinhao Liu	0.6	1	1	1
Rui Zhang	1	0.7	0.9	1
Shumeng Fang	0.7	0.9	1	1
Siyu Wang	1	0.6	1	1