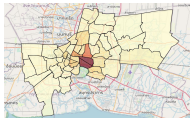
# Airbnb in Bangkok Popularity Analysis

Xue Qin
Georgetown University, MS.

## Introduction



**Goals:**

- Explore the Airbnb Market in Bangkok
- Identify the key factors that influence Airbnb popularity and customer satisfaction
- Predict Airbnb Popularity based on selected features
- Provide insights to help business understand potential success of properties

## Data & EDA

**About the Data:**

This dataset, sourced from Inside Airbnb, is extracted directly from the Airbnb website and is updated regularly. It contains 22,104 Airbnb listings from Bangkok.

**Labels:**

We classify Airbnb listings into four popularity categories—Low, Moderate, High, and Very High—based on the number of reviews they receive as a proxy for their popularity.

**Features:**

The dataset included 75 features, such as location, pricing, number of reviews, and host information. After the data cleaning process, 32 features were retained..

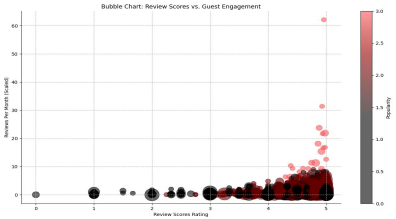
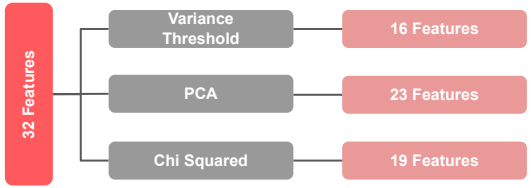



*Figure 1: Review Frequency*

## Feature Selection




**Variance Threshold**: Removing features with low variance, but it may neglect the relationship between variable and target
**PCA**: reduce dimensionality, but it may be hard for making interpretation
**Chi-Squared:** test independence between features and target, but only for categorical data

**Final Choices**: Chi-Squared

1. Best for categorical variables
2. Comparably high accuracy for all selected models

## Model Selection

- We created 7 models to predict the labels Here is the comparison.

Model Comparison

| Metric | LR | SVM | LDA | QDA | RFC | Tree | ANN |
|---|---|---|---|---|---|---|---|
| Accuracy | 0.6483 | 0.7606 | 0.6113 | 0.5596 | 0.9014 | 0.8729 | 0.8463 |
| Precision | 0.6337 | 0.7606 | 0.6003 | 0.5815 | 0.9087 | 0.8728 | 0.8531 |
| Recall | 0.6483 | 0.7606 | 0.6113 | 0.5596 | 0.9066 | 0.8729 | 0.8529 |
| F-1 Score | 0.6362 | 0.7583 | 0.6005 | 0.5550 | 0.9072 | 0.8728 | 0.8531 |
| AUC Score | 0.8726 | 0.8781 | 0.8473 | 0.8130 | 0.9483 | 0.9155 | 0.9065 |

*Figure 2: Model Result Comparison Table*

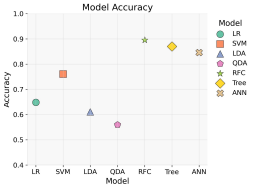
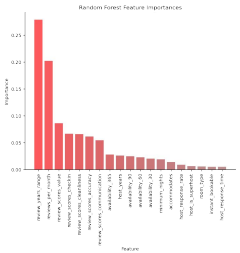



*Figure 3: Model Accuracy*

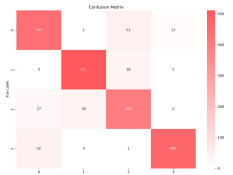
- Initially, we explored seven models for classification work:: Logistic Regression, SVM, LDA, QDA, Random Forest, Decision Tree and Deep Learning ANN model.
- According to **Figure 2**, highest precision of Random Forest model refers fewer false positive, i.e. fewer low quality airbnbs.
- According to **Figure 3**, the Random Forest Classifier achieved nearly 90% accuracy, helping to provide customers with pleasant tourism experience.

Our classification of popularity is highly correlated with **'review_years_range'** and **'review_per_month'**, which underscores the significance of a listing's established presence and ongoing guest engagement.

Beyond these, the importance of **'review_scores_value'**, **'review_score_cleanliness'**, and **'review_scores_checkin'** offers actionable insights.




## Model Analysis




| | precision | recall | f1-score | support |
|---|---|---|---|---|
| 0 | 0.88 | 0.87 | 0.87 | 507 |
| 1 | 0.92 | 0.93 | 0.93 | 549 |
| 2 | 0.82 | 0.86 | 0.84 | 466 |
| 3 | 0.98 | 0.92 | 0.95 | 508 |
| accuracy | | | 0.90 | 2030 |
| macro avg | 0.90 | 0.90 | 0.90 | 2030 |
| weighted avg | 0.90 | 0.90 | 0.90 | 2030 |

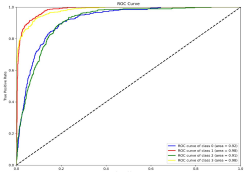



*Figure 4: RFC Heatmap and roc curve*

- Training set Accuracy result: 100%
- Test set Accuracy result: 90%

According to **Figure 4**, ROC curve indicates that Random Forest Classifier outperforms other models in distinguishing the appropriate class labels. It indicates that the model provides effective information for tourists housing choices.

## Conclusion

After an extensive analysis of Airbnb listings in Bangkok, our insights offer some actionable recommendations for hosts looking to increase their popularity of their listing:

- Focus on review scores to distinguish a listing.
- pay meticulous attention to cleanliness to ensure a positive reputation.
- Ensure a hassle-free check-in process to enhance guest satisfaction.
- Prioritize excellent communication between guest and host.