

G
G

RESEARCH CAMPAIGN: SEATTLE AIRBNB

Rakamin Academy
Data Science Batch 49

Mentor : Muhammad Hanif Fajari



OUTLINE OF CONTENT

**1. BUSINESS
UNDERSTANDING**



**2. DATA
ANALYSIS**



**3. DATA
PREPROCESSING**



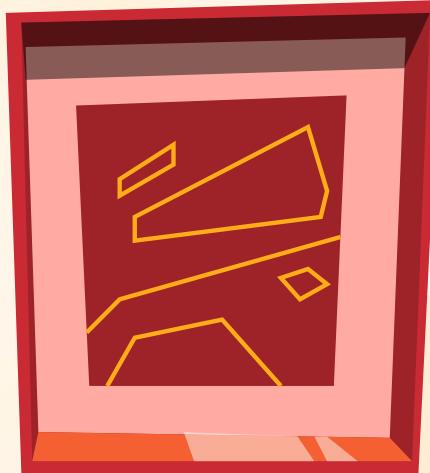
**4. MODELLING &
EVALUATION**



**5. BUSINESS
RECOMMENDATION**



**6. LIMITATIONS &
DEPLOYMENT**





BUSINESS UNDERSTANDING

Problem Statement

Goals, Objective & Business Metrics

COMPANY BACKGROUND

Airbnb is a platform that connects property owners (hosts) with guests seeking a unique and unforgettable stay experience.

To enhance guest satisfaction and strengthen the platform's reputation, the Marketing Team focuses on developing campaigns targeting hosts with the potential to become Superhosts.

Airbnb has hired The Golden Gate team to "Predict" and "Identify" which hosts have the potential to become the target of their flagship host campaign (Superhost), minimizing campaign costs and maximizing profit from the target campaign.



MEET THE TEAM



M Rizdky Maulady
Project Manager



Daniel Andrew S
Data Engineer

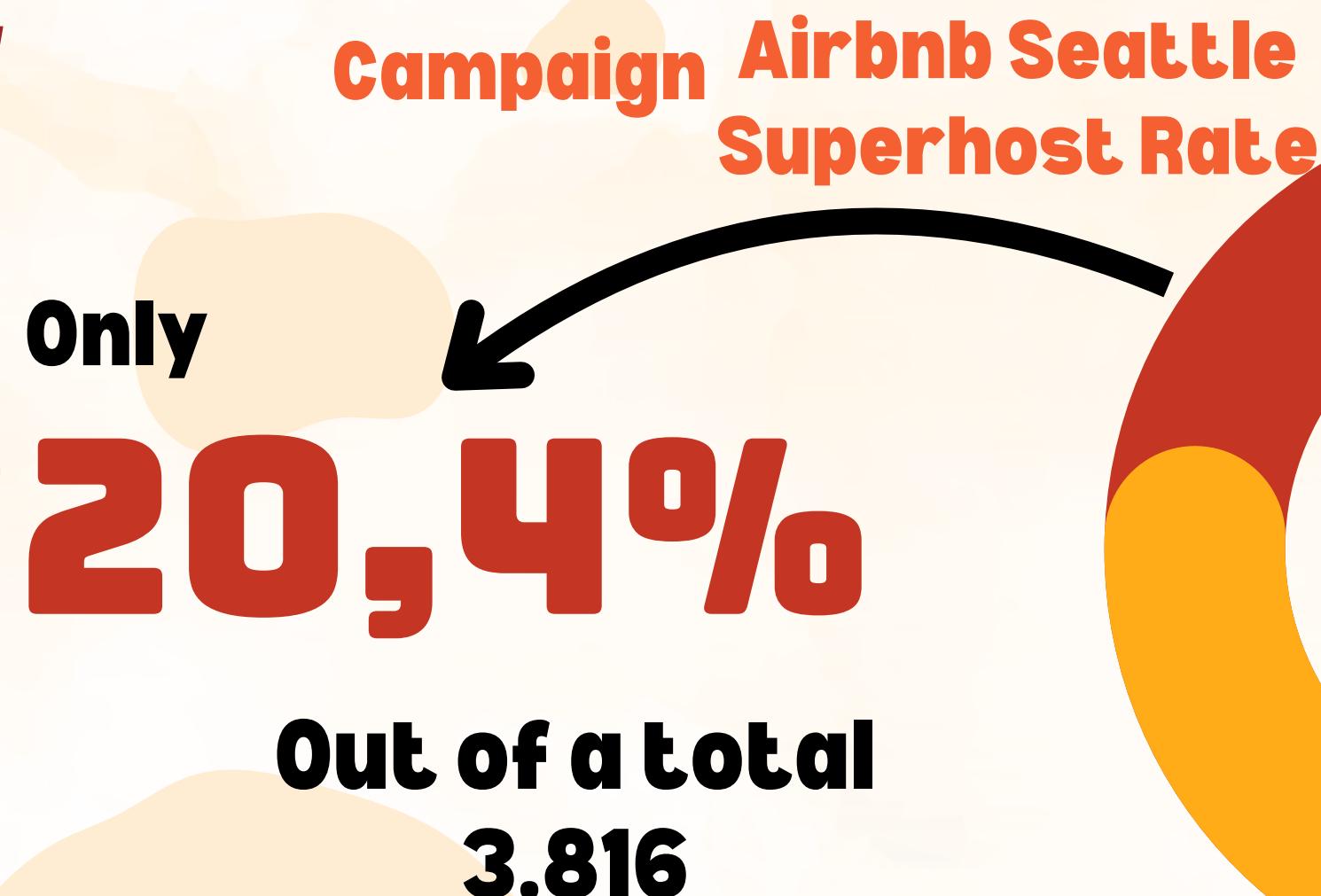
Josua Ricardo S
Data Analyst

Habib Septrian P
Data Scientist



PROBLEM STATEMENT

1. Low Superhost Rate



2. High Campaign Costs



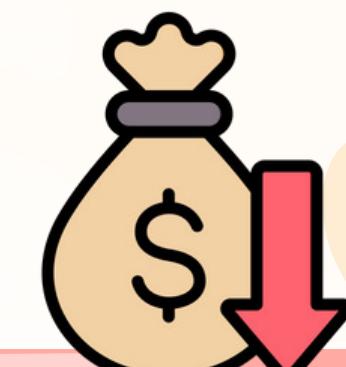
Cost / host
\$12

\$45.792

3. Unoptimized Profit

Revenue / Host
\$44

NPM
-33.76%



GOALS, OBJECTIVE & BUSINESS METRICS



GOALS

Improve the accuracy of Superhost candidate predictions, thereby increasing profitability for future campaigns.



OBJECTIVE

Develop a prediction model for Superhost candidates for future campaigns.



BUSINESS METRICS

- Superhost rate (%)
- Net Profit Margin (%NPM)
- Return On Investment (ROI)



EXPLORATORY DATA ANALYSIS

Dataset information

Business Insight

DATA UNDERSTANDING

MERGE DATASET



Merge the aggregated results (reviews_y) into the listings dataset as the Master dataset

Contains 84,849 samples, 6 features
Agregasi 'reviews' perlistings by 'listings_id'

Dataset Calendar

Contains 1,393,570 samples, 4 features
Latest price from the calendar dataset by listings_id & date

Merge the latest price (price_y), listings_id and date into the Master dataset.



Rename
'price_x' : 'price_listings'
'price_y' : 'price_calendar'
'number_of_reviews_x': 'number_of_reviews'



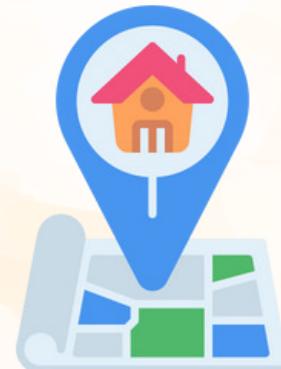
Remove duplicate features
'number_of_reviews_y'

After merging,
the Master dataset contains 3,818 samples, covering 95 features

DATA UNDERSTANDING

DATASET MASTER INFORMATION

Contains 3,818 samples and includes 95 features



Location and Property Type
(17 Features)



Price and Availability
(16 Features)



Capacity and Facilities
(10 Features)



Reviews and Ratings
(12 Features)



Host Information
(15 Features)



Policies and Renter Verification
(12 Features)

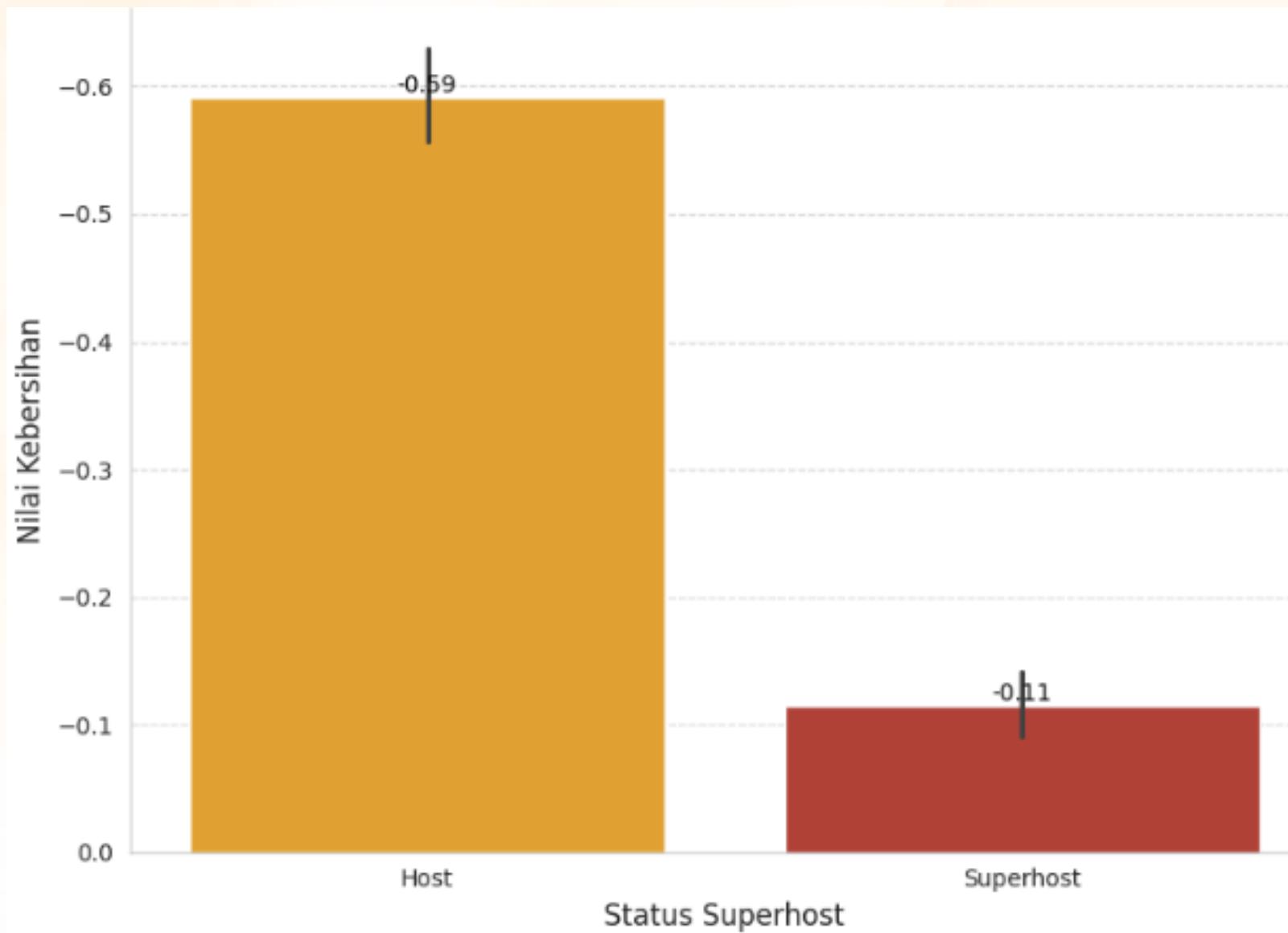


ID and Listing Information
(13 Features)



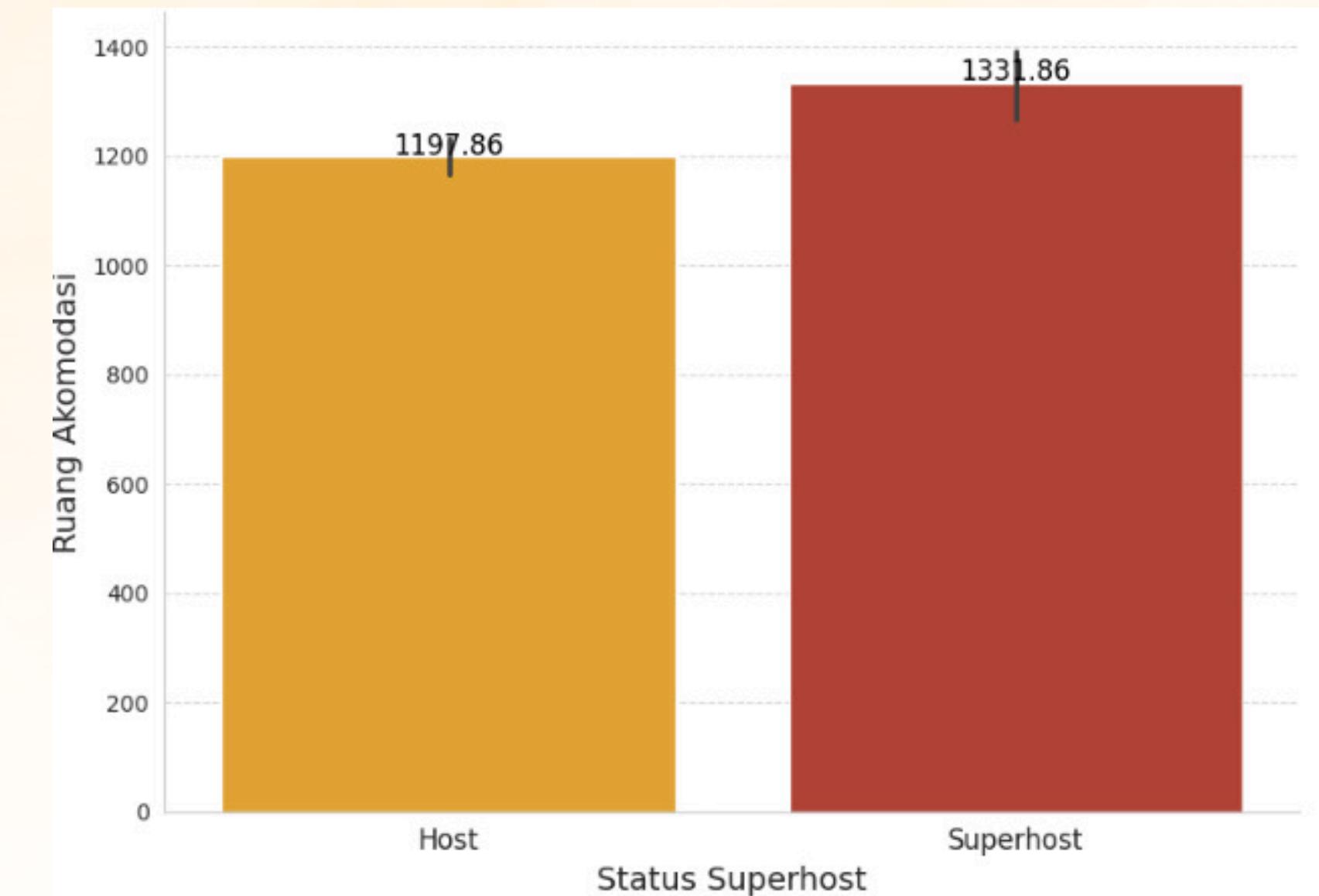
BUSINESS INSIGHTS

Review Scores Cleanliness



Superhost properties receive better cleanliness reviews. Hosts can improve by providing staff training and building communication with customers by requesting constructive reviews.

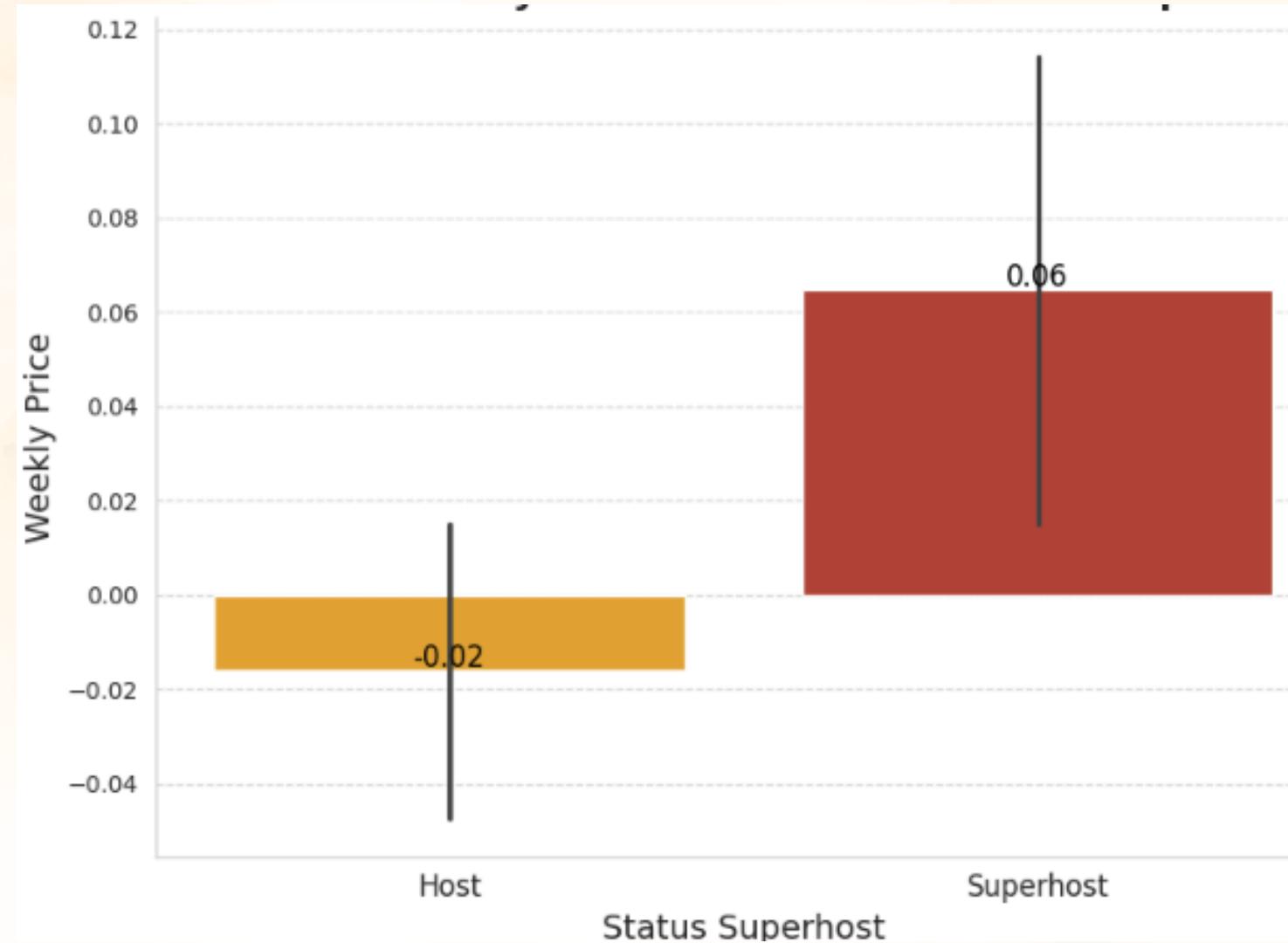
Space



Indicates that Superhosts offer larger accommodation spaces. Optimize the property portfolio by adding more accommodations with larger spaces.

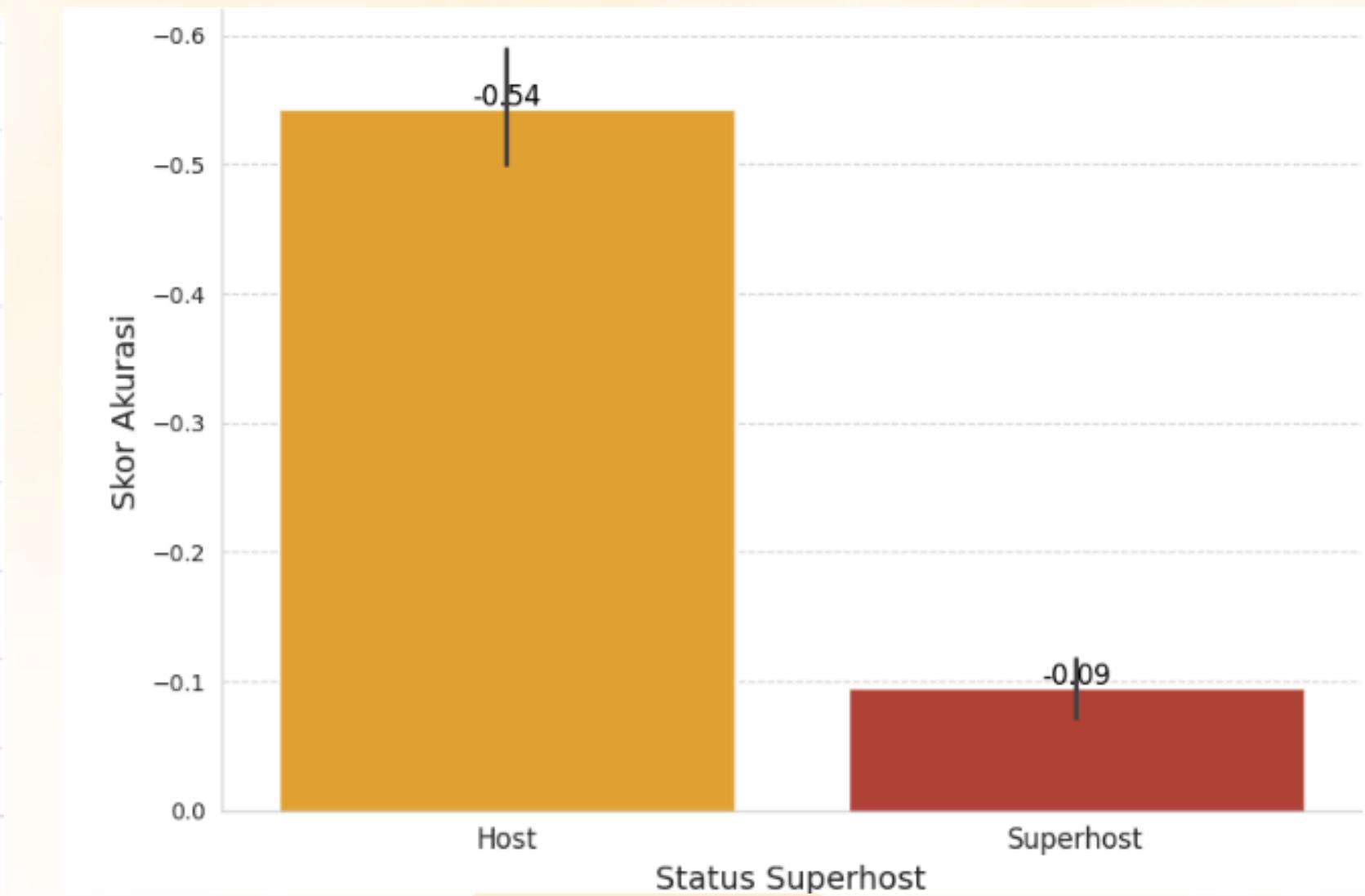
BUSINESS INSIGHTS

Weekly Price



Superhosts have higher weekly rates. They offer trust, visibility, quality, and high demand, making guests willing to pay more for guaranteed quality and a better experience.

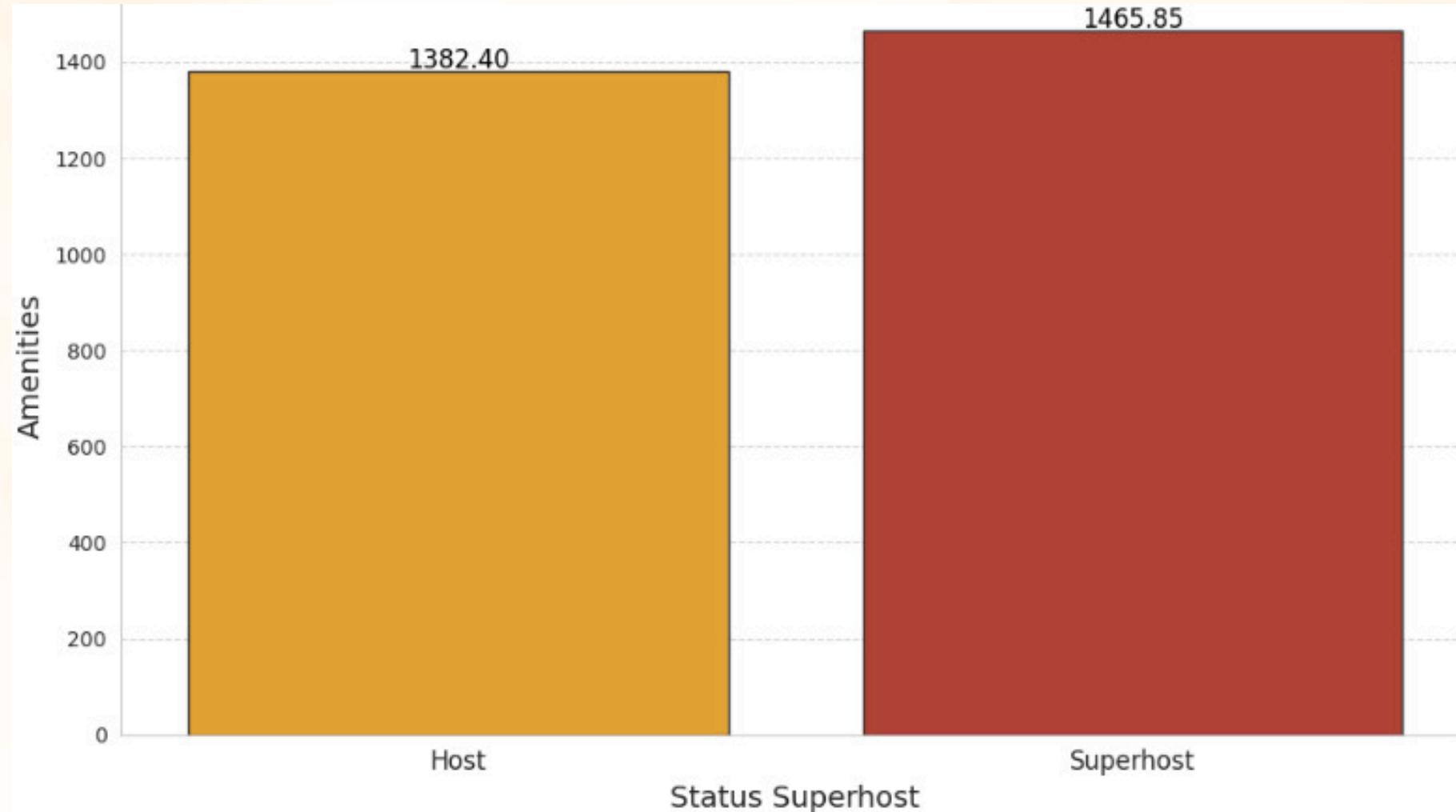
Review Scores Accuracy



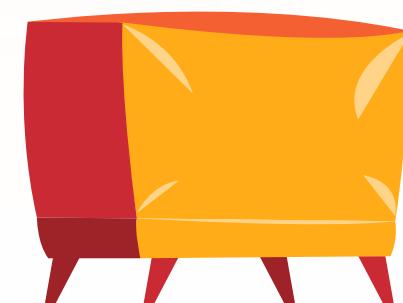
Superhosts have higher listing description accuracy scores (closer to 0). They are likely more reliable in providing accurate information, helping guests make more informed decisions when choosing accommodations.

BUSINESS INSIGHTS

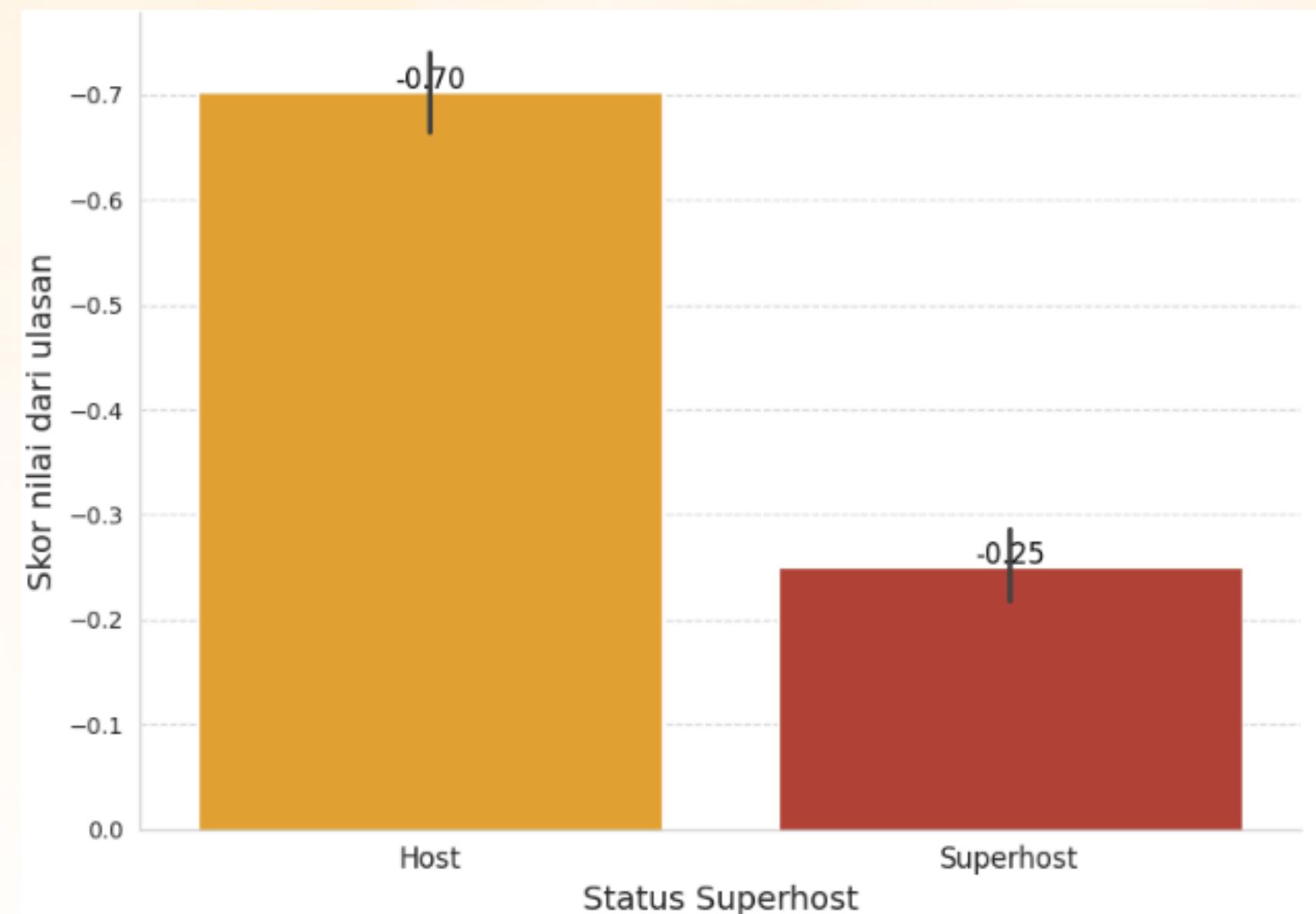
Amenities



Superhost properties have more complete and higher-quality amenities. This can be observed from guest reviews, highlighting the importance of communication with customers.



Review Scores Value

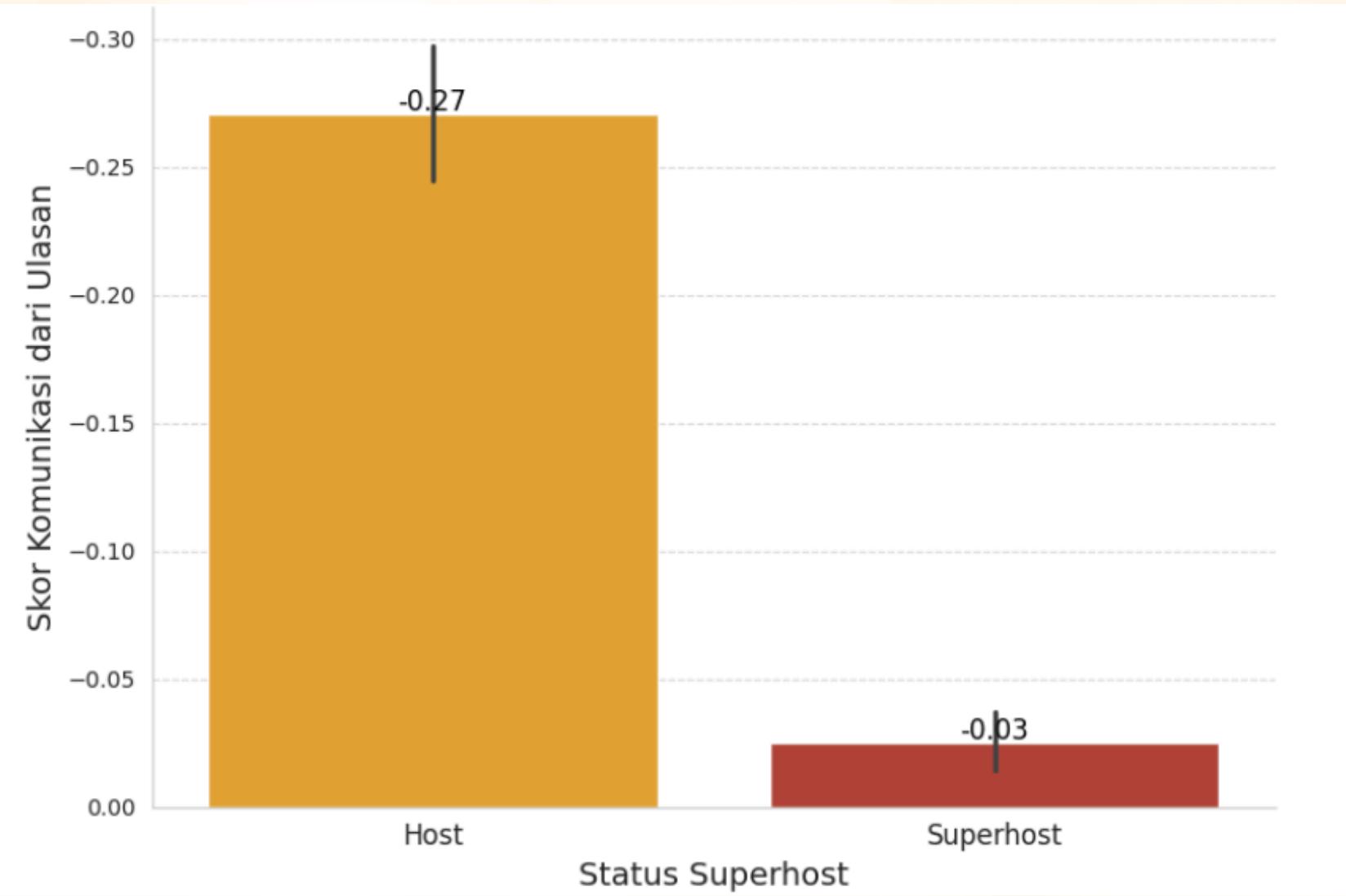


Superhosts have higher average review scores compared to regular hosts. They tend to provide a better stay experience for their guests.

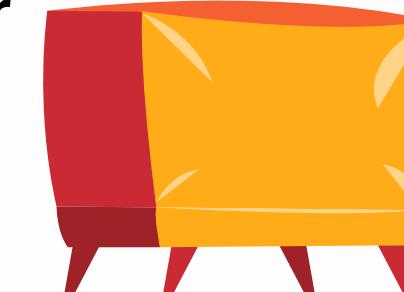


BUSINESS INSIGHTS

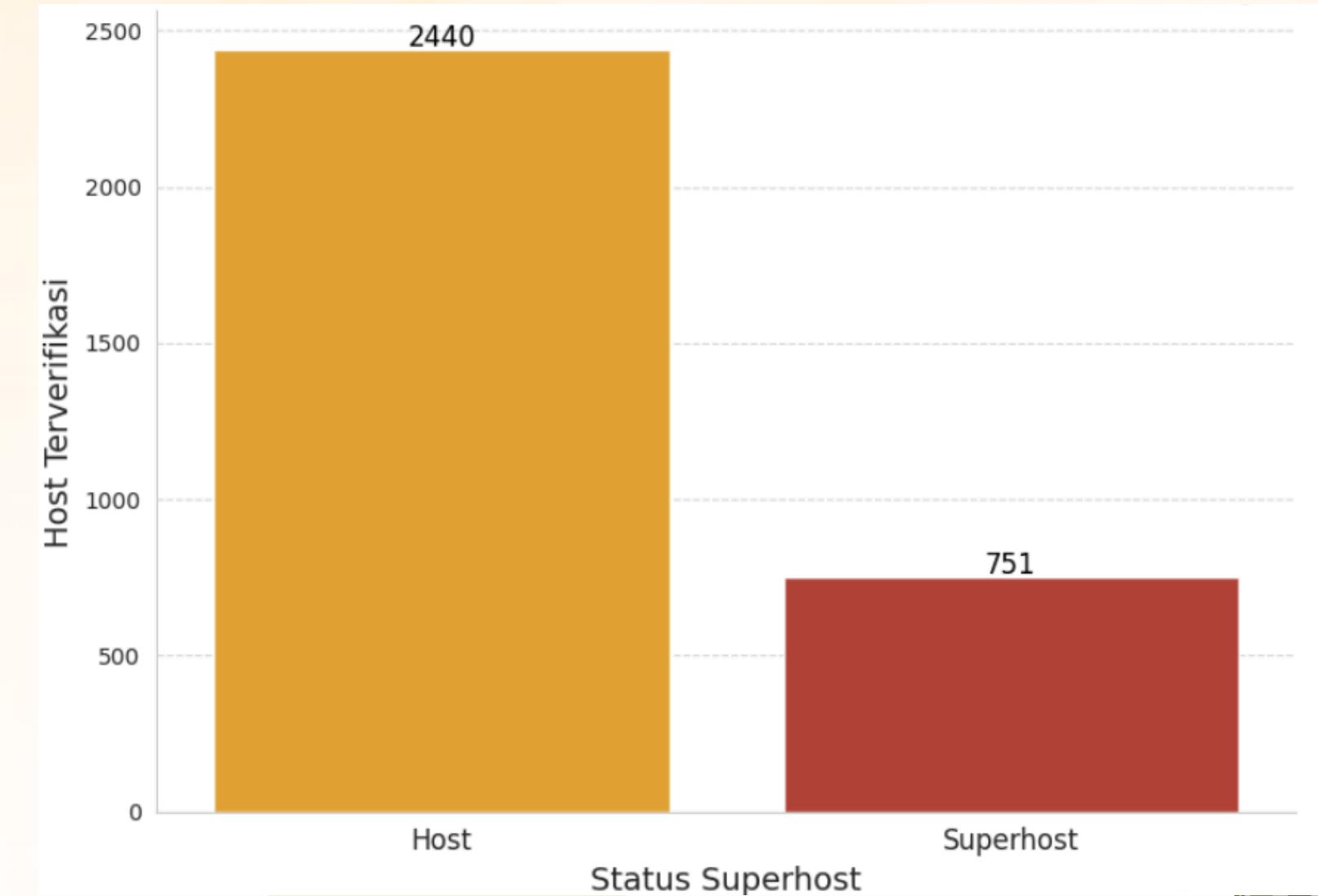
Review Scores Communication



Superhosts have better communication compared to regular hosts, which is likely one of the factors in achieving Superhost status. Reviews play a crucial role in the development or improvement of a property's host status.



Host Identity Verified



Property owner identity verification is mandatory for security reasons and to reassure guests. Host status has a higher value as there are more properties with Host status overall.





DATA PRE-PROCESSING

Handling Missing, Invalid

Feature Transform & Engineering

Feature Selection

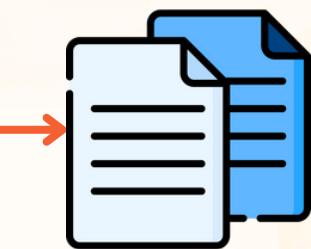
Handling Imbalance Data

DATA PRE-PROCESSING



Handling Invalid Values

- 6 Object Features => Datetime
- 7 Object Features => Float & Symbol Removal
- 8 Object Features => Bool dari t/f to true/false
- 10 Object Features Containing Numbers => int
- 17 Object Features Containing Numbers => float



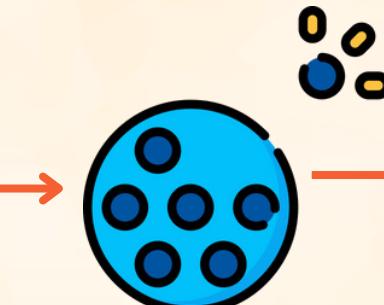
Handling Duplicated Data

0 Duplicated Data



Handling Missing Value

- Drop 3 Features <3000 & Double
- Imputation Numeric
`rate_columns` = Previous Value
`price_columns` = Median Value
`score_columns` = 0
- Imputation Categoric
`frequent_text` = Mode Value
`not_available` = "Tidak Tersedia"
`no_image` = "No Image"



Handling Outliers

Winsorization



Feature Engineering / Extraction

6 Features

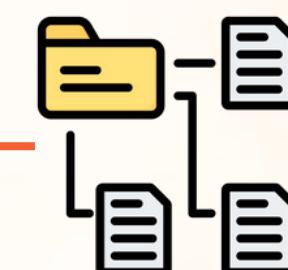


Handling Imbalanced Data

RandomOverSampler
(Oversampling)

Data Train = 2941

**Label 1 before 591,
after 980**



Data Splitting

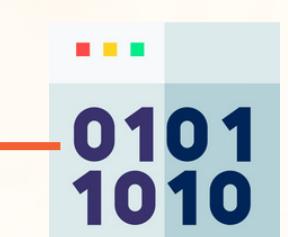
Data Train 80% = (2941)

Data Test 20% = (639)



Feature Selection

- Anova
Variance Threshold
Mutual Information
Selected-Kbest
Multicollinearity



Feature Encoding (Categoric)

Label Encoding



Feature Transformation (Numeric)

Yeo-Johnson Transformation
(Increase the accuracy 2%)

RobustScaler
(For Outlier)

DATA PRE-PROCESSING

SELECTED 23 FEATURE FROM 29 FEATURE

GUEST EXPERIENCE QUALITY

- review_scores_cleanliness
- review_scores_communication
- review_scores_location
- review_scores_value
- review_scores_accuracy
- review_scores_checkin

HOST MANAGEMENT

- host_response_time
- host_response_rate
- host_identity_verified
- require_guest_phone_verification

PROPERTY CHARACTERISTICS

- monthly_price
- weekly_price
- cleaning_fee
- amenities
- property_type
- neighbourhood_cleansed
- space
- zipcode
- bedrooms
- calendar_updated
- cancellation_policy
- extra_people
- guests_included



MODELLING & EVALUATION

Model Evaluation Parameters

Machine Learning Techniques

Model Comparison

Feature Importance

MODELLING & EVALUATION

MODEL EVALUATION PARAMETERS

25 FEATURES

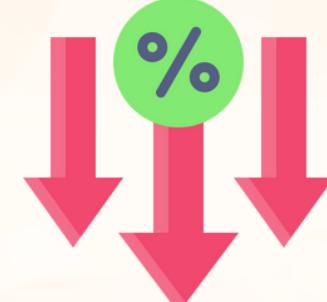
Positive = Superhost Campaign
 Negative = No Superhost Campaign



Model Prediction



Happened



Impact

FALSE POSITIVE

Primary Targets to Reduce

Superhost

No Superhost

 Inefficient Use of
Campaign Resources

FALSE NEGATIVE

Secondary Targets to Reduce

No Superhost

Superhost

 Loss of Potential
Superhosts

FALSE

TRUE

MODELLING & EVALUATION

MODEL EVALUATION PARAMETERS

1. PRECISION

Primary Evaluation Metrics

Reduce False Positives



Increase Superhost Rate



Minimize Marketing Costs

2. RECALL

Secondary Evaluation Metrics

Reduce False Negatives

Increase Revenue



3. F1 SCORE

Combined Evaluation Metrics

Required in this case as the harmonic mean of Precision and Recall.

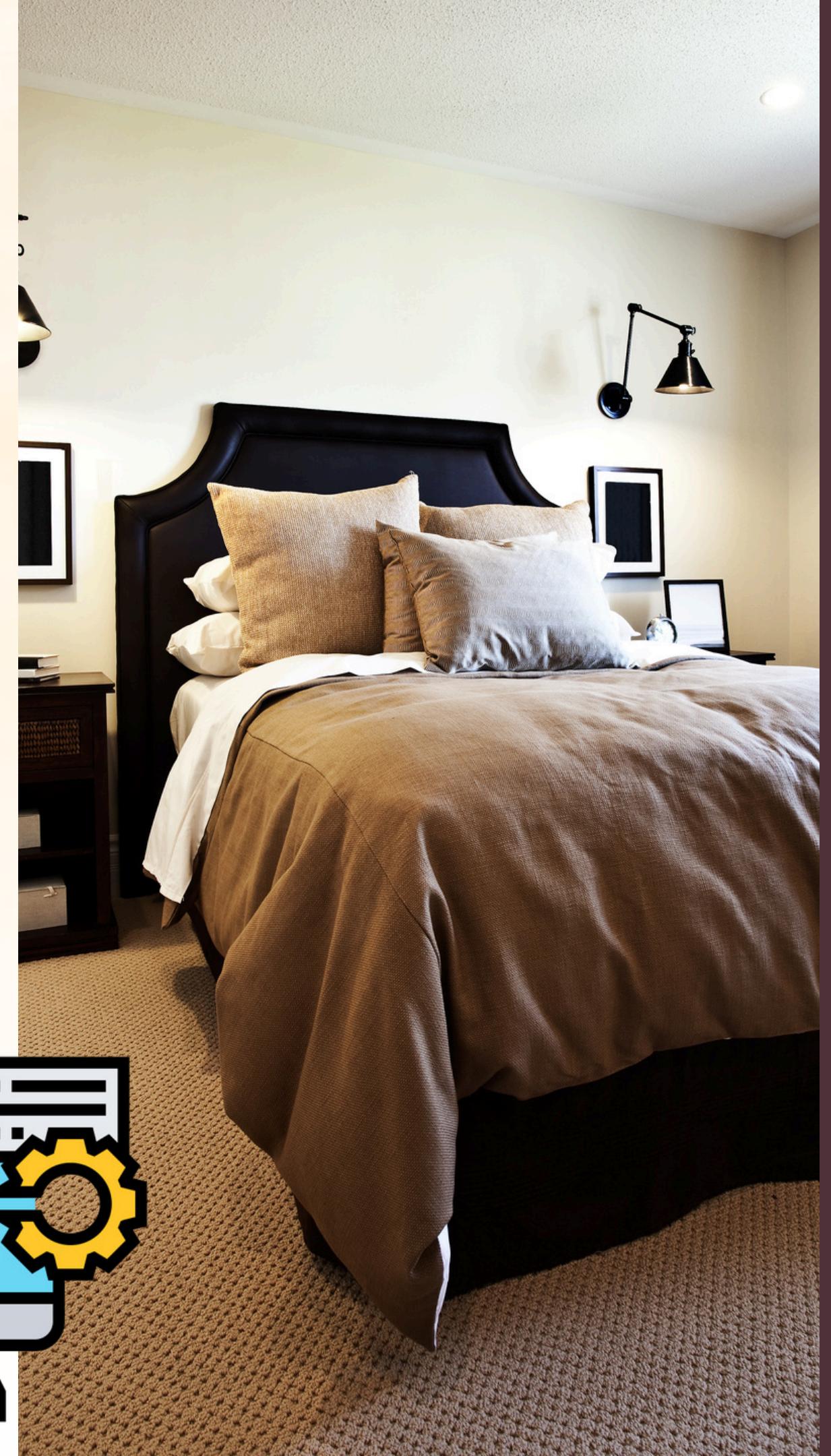
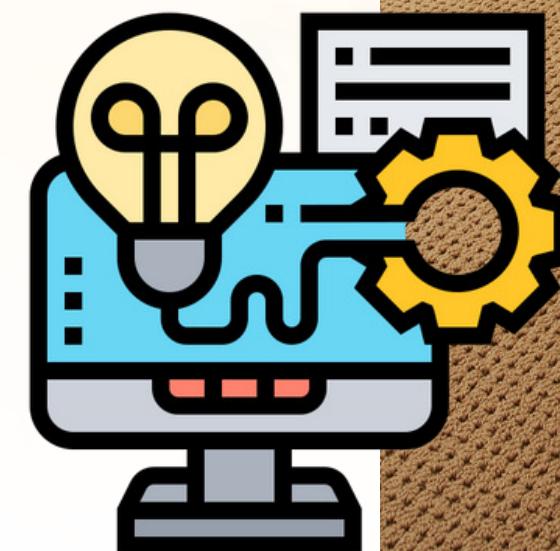
MODELLING & EVALUATION MACHHINE LEARNING TECHNIQUES

1. Decision Tree
2. Random Forest
3. Logistic Regression
4. Gaussian Naive Bayes
5. K-Nearest Neighbor
6. MLP Classifier (Neural Network)
7. Adaboost Classifier
8. Gradient Boosting Classifier
9. XGBoost Classifier
10. LGBM Classifier

GRADIENT BOOSTING CLASSIFIER



Best Fit Models



MODELLING & EVALUATION

MODEL COMPARISON (SMOTE)

Training shape: (2941, 23)
Testing shape: (639, 23)

Model	Train Accuracy	Test Accuracy	Train Precision	Test Precision	Train Recall	Test Recall	Train F1 Score	Test F1 Score
Decision Tree	1.00	0.73	1.00	0.72	1.00	0.73	1.00	0.72
Random Forest	1.00	0.78	1.00	0.76	1.00	0.78	1.00	0.75
Logistic Regression	0.70	0.73	0.68	0.72	0.70	0.73	0.68	0.73
Gaussian Naive Bayes	0.62	0.59	0.76	0.78	0.62	0.59	0.62	0.62
K-Nearest Neighbor	0.81	0.66	0.81	0.64	0.81	0.66	0.81	0.65
MLP Classifier	0.68	0.75	0.69	0.68	0.68	0.75	0.56	0.66
Adaboost Classifier	0.78	0.74	0.78	0.73	0.78	0.74	0.78	0.73
Gradient Boosting Classifier	0.86	0.78	0.86	0.76	0.86	0.78	0.86	0.76
XGBoost Classifier	1.00	0.78	1.00	0.76	1.00	0.78	1.00	0.76
LGBM Classifier	0.99	0.79	0.99	0.77	0.99	0.79	0.99	0.77
Support Vector Machine	0.67	0.75	0.44	0.56	0.67	0.75	0.53	0.64

Precision, Recall, and F1 Score achieve the **best evaluation with SMOTE**.
Gradient Boosting results: Precision 0.76, Recall 0.78, and F1 Score 0.76.



MODELLING & EVALUATION

SMOTE & HYPERPARAMETER TUNING

Model: Gradient Boosting Classifier

Train Accuracy: 0.8623

Test Accuracy: 0.7778

Train Precision: 0.8641

Test Precision: 0.7557

Train Recall: 0.8623

Test Recall: 0.7778

Train F1 Score: 0.8572

Test F1 Score: 0.7561

Training Time: 10.71 seconds

Classification Report:

	precision	recall	f1-score	support
0	0.81	0.92	0.86	479
1	0.60	0.35	0.44	160
accuracy			0.78	639
macro avg	0.70	0.64	0.65	639
weighted avg	0.76	0.78	0.76	639

DEFAULT PARAMETER

The model performs better without
Hyperparameter Tuning.

TUNED (OVERFITTING)

Parameter:

'n_estimators': [50, 100, 200],

'learning_rate': [0.01, 0.1, 0.2],

'max_depth': [3, 5, 7],

Data train 100%

Model: Gradient Boosting Classifier

Train Accuracy: 1.0000

Test Accuracy: 0.8258

Train Precision: 1.0000

Test Precision: 0.8161

Train Recall: 1.0000

Test Recall: 0.8258

Train F1 Score: 1.0000

Test F1 Score: 0.8198

Training Time: 405.24 seconds

Classification Report:

	precision	recall	f1-score	support
0	0.88	0.91	0.89	408
1	0.58	0.49	0.53	103
accuracy			0.83	511
macro avg	0.73	0.70	0.71	511
weighted avg	0.82	0.83	0.82	511

MODELLING & EVALUATION

MODEL COMPARISON (OVERSAMPLING)

Training shape: (2941, 23)
 Testing shape: (639, 23)

Model	Train Accuracy	Test Accuracy	Train Precision	Test Precision	Train Recall	Test Recall	Train F1 Score	Test F1 Score
Decision Tree	1.00	0.72	1.00	0.71	1.00	0.72	1.00	0.72
Random Forest	1.00	0.82	1.00	0.81	1.00	0.82	1.00	0.79
Logistic Regression	0.69	0.71	0.68	0.71	0.69	0.71	0.68	0.71
Gaussian Naive Bayes	0.61	0.58	0.77	0.79	0.61	0.58	0.61	0.60
K-Nearest Neighbor	0.79	0.66	0.79	0.65	0.79	0.66	0.79	0.65
MLP Classifier	0.71	0.73	0.71	0.66	0.71	0.73	0.65	0.67
Adaboost Classifier	0.74	0.75	0.73	0.74	0.74	0.75	0.73	0.75
Gradient Boosting Classifier	0.84	0.79	0.84	0.78	0.84	0.79	0.84	0.78
XGBoost Classifier	1.00	0.79	1.00	0.78	1.00	0.79	1.00	0.78
LGBM Classifier	0.99	0.80	0.99	0.79	0.99	0.80	0.99	0.79
Support Vector Machine	0.67	0.75	0.44	0.56	0.67	0.75	0.53	0.64

Precision, Recall, and F1 Score achieve the best evaluation with RandomOverSampler.

Gradient Boosting Precision 0.78 Recall 0.79 & F1score 0.78



MODELLING & EVALUATION

OVERSAMPLING & HYPERPARAMETER TUNING

DEFAULT PARAMETER

```
Model: Gradient Boosting Classifier
Train Accuracy: 0.8443
Test Accuracy: 0.7887
Train Precision: 0.8420
Test Precision: 0.7765
Train Recall: 0.8443
Test Recall: 0.7887
Train F1 Score: 0.8412
Test F1 Score: 0.7801
Training Time: 9.80 seconds
Classification Report:
```

	precision	recall	f1-score	support
0	0.84	0.89	0.86	479
1	0.60	0.47	0.53	160
accuracy			0.79	639
macro avg	0.72	0.68	0.70	639
weighted avg	0.78	0.79	0.78	639

'n_estimators': [50, 100, 200, 500],
 'learning_rate': [0.01, 0.1, 0.2, 0.9],
 'max_depth': [3, 5, 7, 9],
 'min_samples_split': [2, 10],
 'min_samples_leaf': [1, 6],

Train Acc +26%
Test Acc +1%

ResultOverfitting

```
Model: Gradient Boosting Classifier
Train Accuracy: 1.0000
Test Accuracy: 0.7981
Train Precision: 1.0000
Test Precision: 0.7824
Train Recall: 1.0000
Test Recall: 0.7981
Train F1 Score: 1.0000
Test F1 Score: 0.7746
Training Time: 6761.16 seconds
Classification Report:
```

	precision	recall	f1-score	support
0	0.82	0.94	0.88	479
1	0.68	0.36	0.47	160
accuracy				639
macro avg	0.75	0.65	0.67	639
weighted avg	0.78	0.80	0.77	639

The model performs better **without Hyperparameter Tuning**

MODELLING & EVALUATION

MODEL COMPARISON (WITHOUT METHOD)

Training shape: (2552, 23)
 Testing shape: (639, 23)

Model	Train Accuracy	Test Accuracy	Train Precision	Test Precision	Train Recall	Test Recall	Train F1 Score	Test F1 Score
Decision Tree	1.00	0.70	1.00	0.71	1.00	0.70	1.00	0.71
Random Forest	1.00	0.79	1.00	0.79	1.00	0.79	1.00	0.75
Logistic Regression	0.76	0.74	0.70	0.67	0.76	0.74	0.70	0.67
Gaussian Naive Bayes	0.57	0.59	0.80	0.79	0.57	0.59	0.60	0.62
K-Nearest Neighbor	0.81	0.71	0.79	0.63	0.81	0.71	0.78	0.65
MLP Classifier	0.77	0.72	0.75	0.69	0.77	0.72	0.75	0.70
Adaboost Classifier	0.78	0.76	0.75	0.72	0.78	0.76	0.75	0.71
Gradient Boosting Classifier	0.84	0.78	0.84	0.76	0.84	0.78	0.81	0.74
XGBoost Classifier	1.00	0.79	1.00	0.77	1.00	0.79	1.00	0.77
LGBM Classifier	0.99	0.79	0.99	0.77	0.99	0.79	0.99	0.77
Support Vector Machine	0.77	0.75	0.59	0.56	0.77	0.75	0.67	0.64

Precision, Recall, and F1 Score achieve the best evaluation **without** Oversampling/SMOTE Method. Gradient Boosting results: Precision 0.76, Recall 0.78, and F1 Score 0.74.

MODELLING & EVALUATION COMPARISON

WITHOUT METHOD

Model: Gradient Boosting Classifier

Train Accuracy: 0.8443

Test Accuracy: 0.7887

Train Precision: 0.8420

Test Precision: 0.7765

Train Recall: 0.8443

Test Recall: 0.7887

Train F1 Score: 0.8412

Test F1 Score: 0.7801

Training Time: 9.80 seconds

Classification Report:

	precision	recall	f1-score	support
0	0.84	0.89	0.86	479
1	0.60	0.47	0.53	160
accuracy			0.79	639
macro avg	0.72	0.68	0.70	639
weighted avg	0.78	0.79	0.78	639

WITH OVERSAMPLING

Precision 0.78 Recall 0.79 & F1score 0.78

The model performs better using Random Oversampler. Although the accuracy is lower, this method is more suitable for situations with class imbalance, allowing the model to learn more effectively from minority class data

Precision 0.76 Recall 0.78 & F1 Score 0.74

Model: Gradient Boosting Classifier

Train Accuracy: 0.8464

Test Accuracy: 0.7856

Train Precision: 0.8497

Test Precision: 0.7708

Train Recall: 0.8464

Test Recall: 0.7856

Train F1 Score: 0.8235

Test F1 Score: 0.7451

Training Time: 11.98 seconds

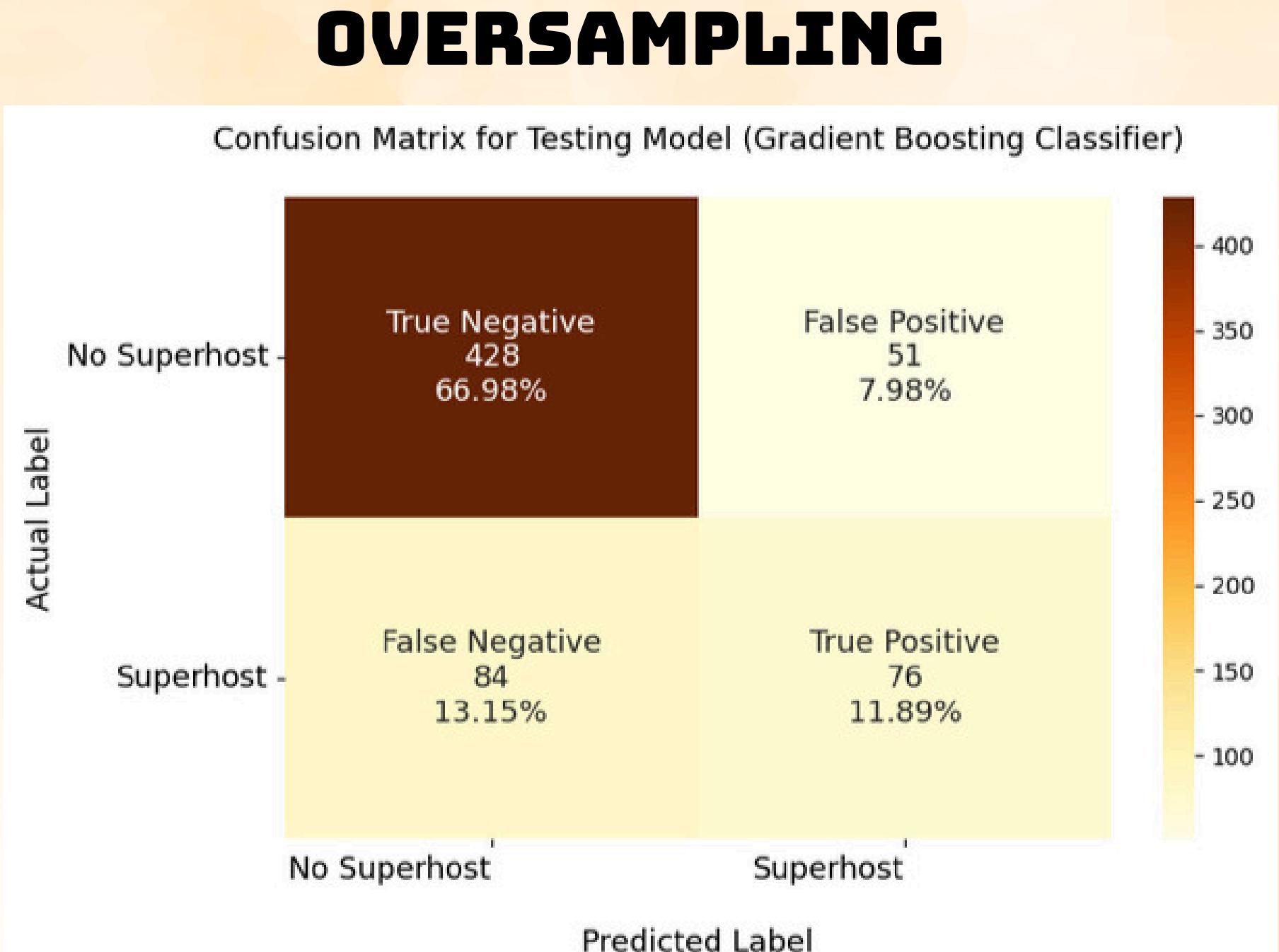
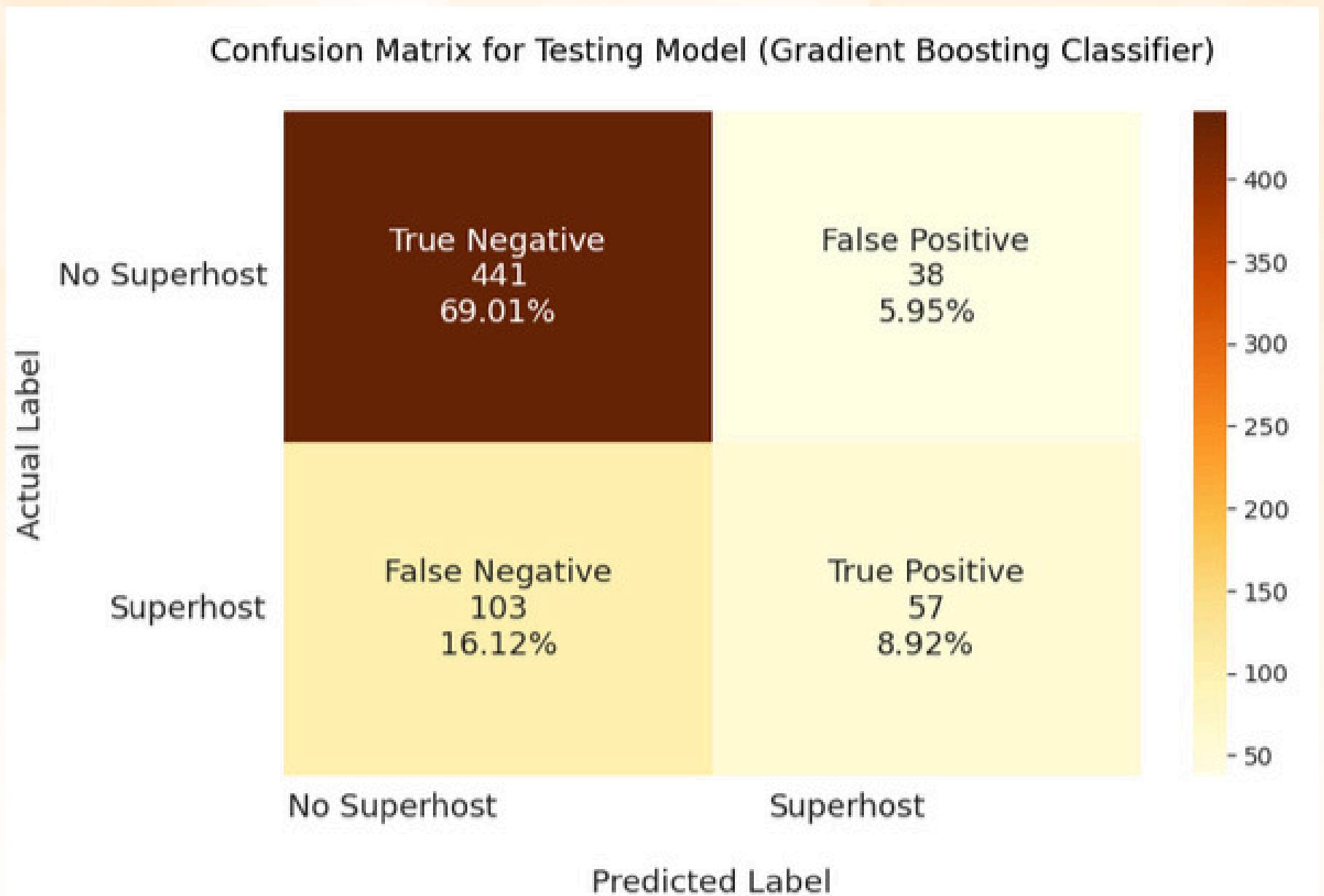
Classification Report:

	precision	recall	f1-score	support
0	0.79	0.96	0.87	479
1	0.70	0.25	0.37	160
accuracy			0.79	639
macro avg	0.75	0.61	0.62	639
weighted avg	0.77	0.79	0.75	639

MODELLING & EVALUATION

COMPARISON CONFUSION MATRIX

SMOTE

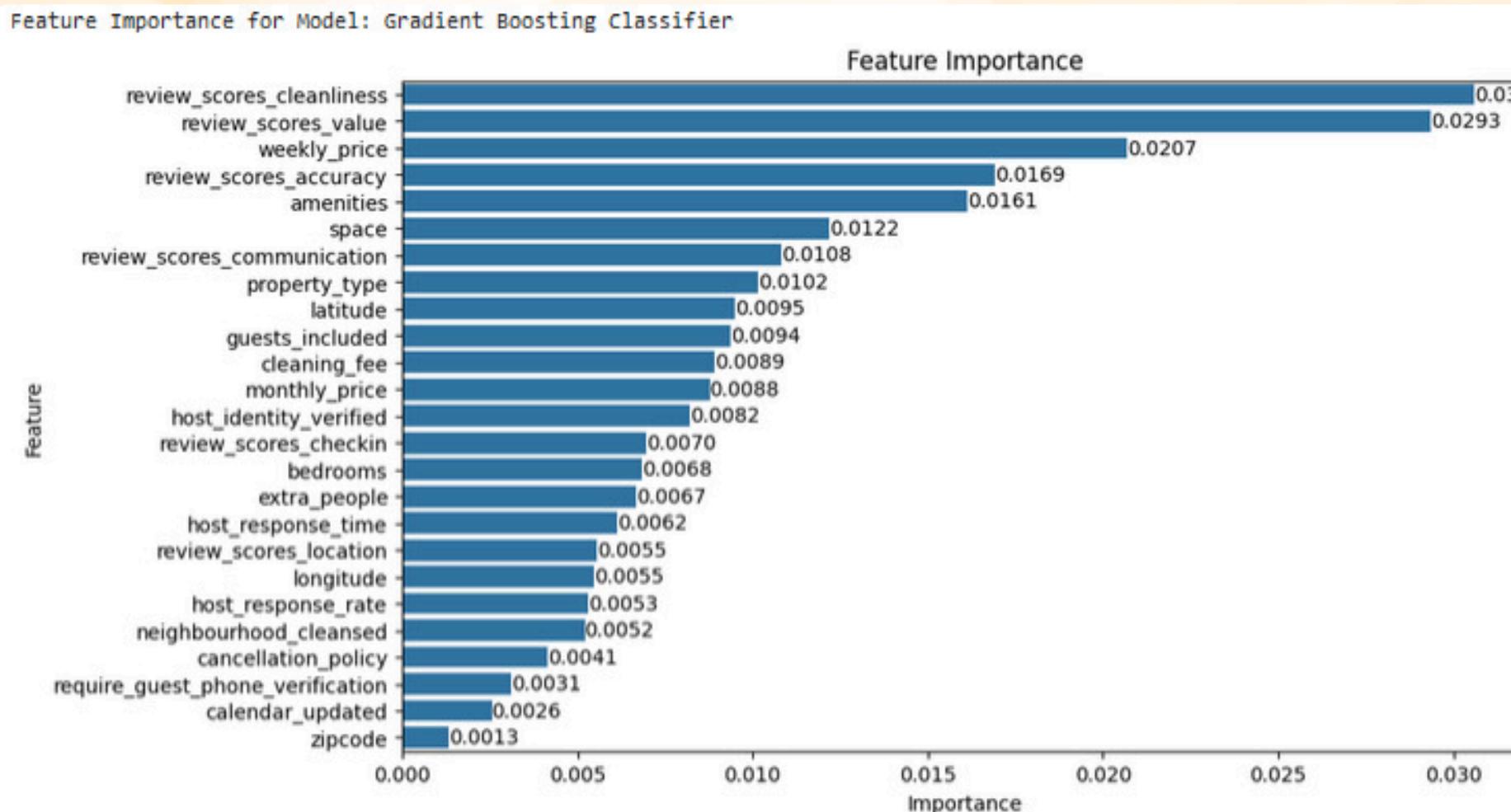


Random Oversampler (oversampling) outperforms SMOTE, with a higher True Positive (TP) value and a lower False Negative (FN) value.

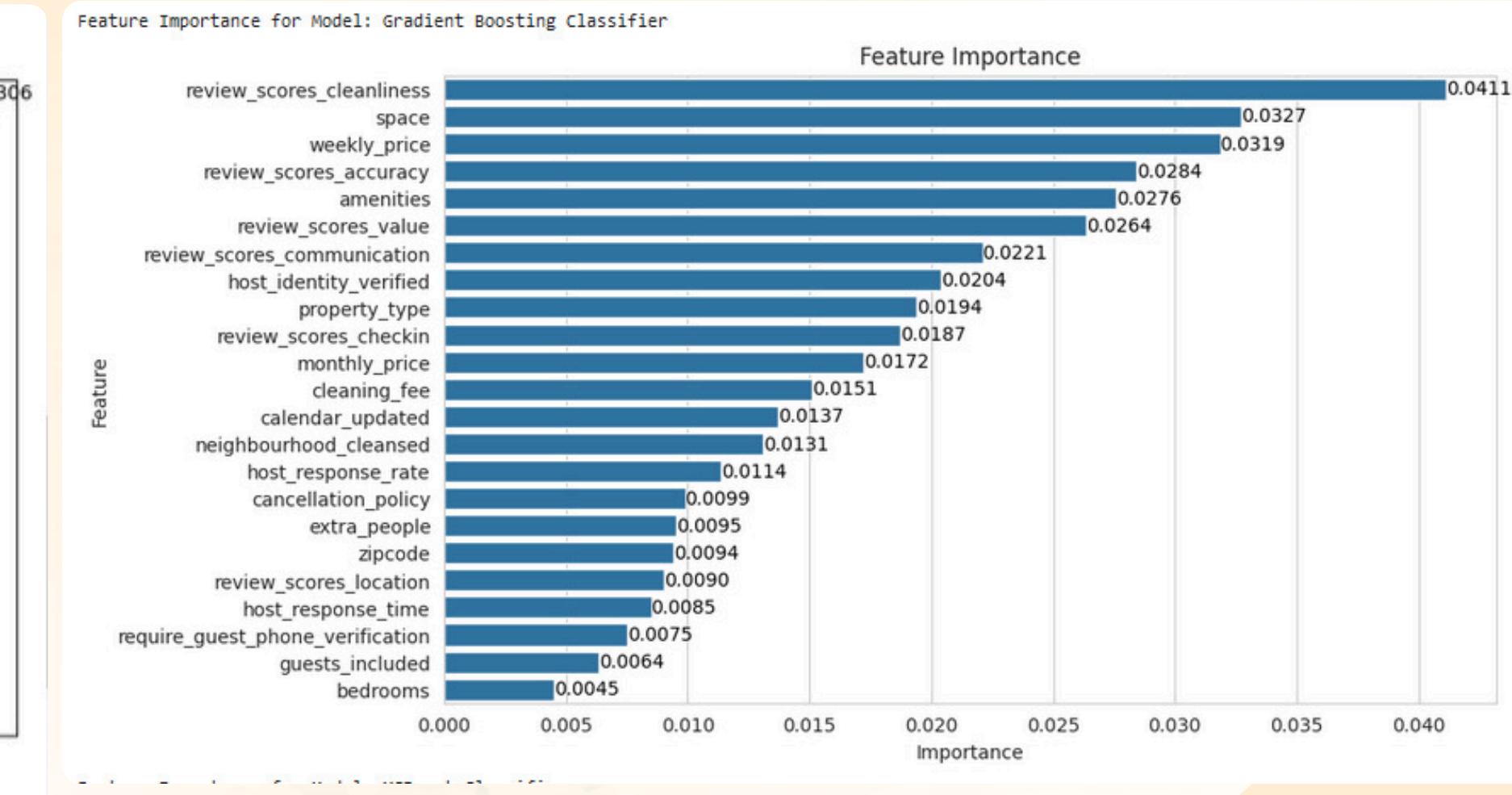
MODELLING & EVALUATION

COMPARISON FEATURE IMPORTANCE

SMOTE



OVERSAMPLING

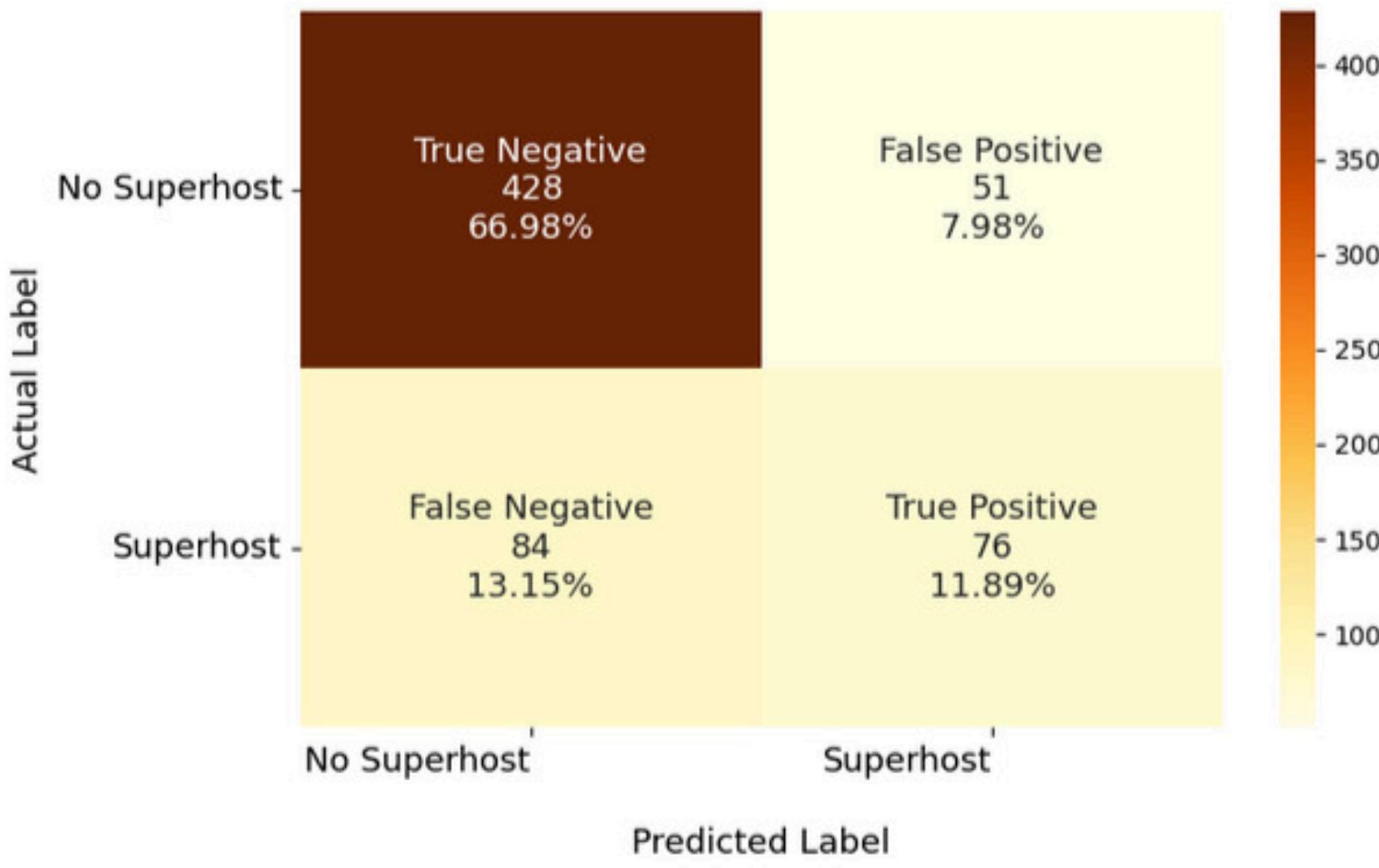


Oversampling is better, as SMOTE causes underfitting. In our case, the synthetic data generated does not accurately reflect the patterns present in the original data.

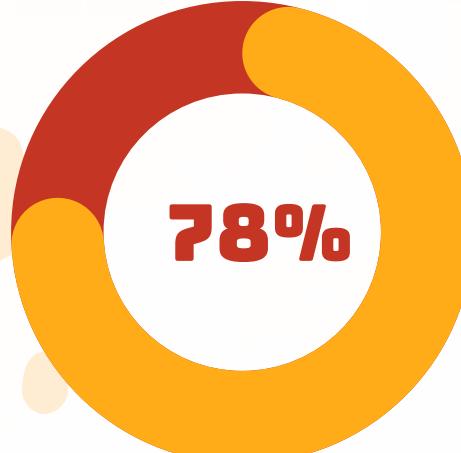
MODELLING & EVALUATION

MODEL SELECTION

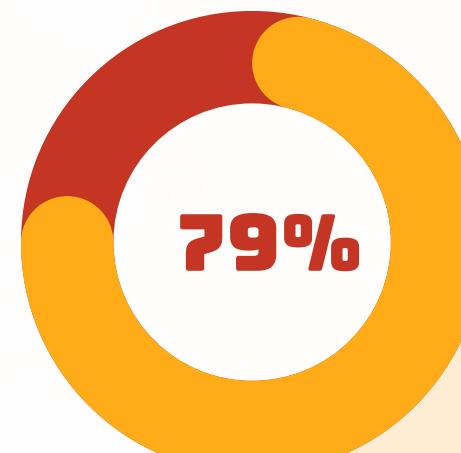
Confusion Matrix for Testing Model (Gradient Boosting Classifier)



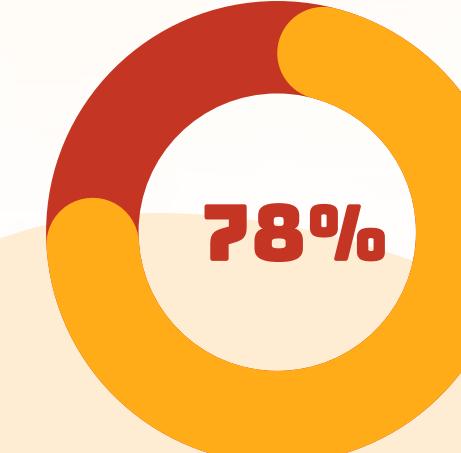
PRECISION



RECALL



F1 SCORE



With Random Oversampling

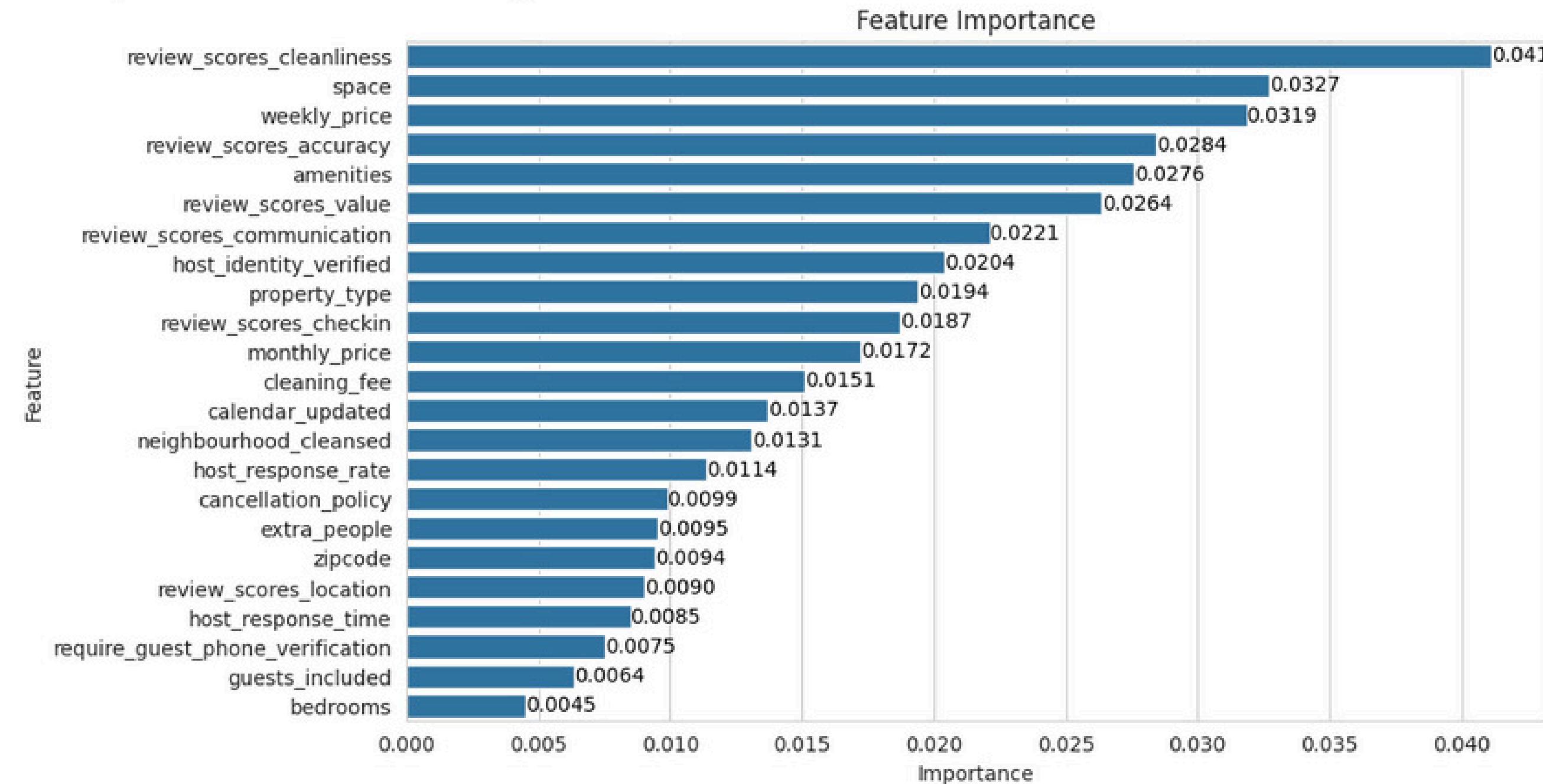
True Positive	Predicted Superhost 	True, Superhost 
True negative	Predicted Not Superhost 	True, No Superhost 
False Positive	Predicted Superhost 	False, Superhost 
False Negative	Predicted Not Superhost 	False, No Superhost 

RandomOverSampler is better at capturing existing patterns in the data, improving accuracy and generalization while reducing underfitting compared to SMOTE..

MODELLING & EVALUATION

FEATURE IMPORTANCE

Feature Importance for Model: Gradient Boosting Classifier



TOP 8 Feature :

1. **review_scores_cleanliness**
 - Property cleanliness score
2. **space**
 - Description of the accommodated space
3. **weekly_price**
 - Weekly rental price
4. **review_scores_accuracy**
 - Listing description accuracy score
5. **Amenities**
 - Complete and high-quality facilities
6. **review_score_value**
 - Review value score
7. **review_scores_communication**
 - Communication score from reviews
8. **host_identity_verified**
 - Host identity verification



BUSINESS & RECOMMENDATION

Recommendation

Business Impact

Simulation

BUSINESS RECOMMENDATIONS

STRATEGIC ACTION

RECOMMENDATIONS

FOCUS ON IMPROVING PROPERTY QUALITY & GUEST EXPERIENCE

Cleanliness scores have the highest value and significantly impact Superhost status. While many properties are already clean, there is an opportunity to enhance guest trust through accurate descriptions and by providing complete amenities for a more enjoyable experience.

1. **Implement Regular Cleaning Protocols:** Conduct regular cleanliness audits to ensure high standards and consistency.
2. **Enhance Listing Descriptions:** Update property descriptions to be more accurate and reflect all available amenities.
3. **Upgrade Amenities:** Add attractive and high-quality amenities to enhance the property's appeal and guest experience.



LEVERAGE ENHANCED COMMUNICATION AND GUEST TRUST

The host's communication score and identity verification are crucial for building guest trust. There is an opportunity to enhance trust by strengthening interactive communication and ensuring the host's identity is verified.

1. **Implement Proactive Communication Strategies:** Encourage hosts to communicate proactively with guests before and during their stay.
2. **Enhance Identity Verification Process:** Ensure that all hosts undergo strict identity verification to enhance trust.
3. **Collect Feedback on Communication:** Gather guest feedback on their communication experience to identify areas for improvement.

Weekly rental prices influence guest decisions. Competitive pricing can attract more bookings.

1. **Gather Guest Feedback for Pricing:** Collect guest feedback on pricing and the value they received, and discuss it with hosts to adjust weekly rental rates for better competitiveness and alignment with guest expectations.
2. **Promote Special Events:** Collaborate with hosts to promote local or seasonal events and adjust rental prices to attract guests during these periods.
3. **Conduct Joint Promotions:** Encourage hosts to collaborate on special promotions, such as discounts for weekend bookings or holiday seasons.

ACTIONABLE ITEMS

BUSINESS RECOMMENDATIONS

BUSINESS IMPACT

SUPERHOST RATE

Predicted	
Actual	No
Superhost	428
No Superhost	51
Superhost	76
	84

True Positive (TP) = 76

False Positive (FP)= 51

$$\begin{aligned}\text{Superhost} &= \text{TP}/(\text{TP} + \text{FP}) \\ &= 76/(76 + 51)\end{aligned}$$

60 %

Before Model

20.4 %

194.12%



After Model

60%

BUSINESS RECOMMENDATIONS

BUSINESS IMPACT NET PROFIT MARGIN / NPM

Before
Modelling

-33.76%

261,06%

After
Modelling

54.42%

Total Superhosts (1 = becoming a Superhost)	778
Total campaign (all host)	3.816
Total Cost (\$)	= total campaign*cost = 3.816×12 45.792
Total Revenue (\$)	= total superhost*revenue = 778×44 34.232
Total Profit (\$)	= total revenue - total cost = $34.232 - 45.792$ -11.560
NPM (%)	= (total profit/ total revenue)*100 = $(-11.560/34.232) \times 100$ -33.76



**Host
Revenue (\$)**
44
Cost (\$)
12

Total Campaign (TP+FP)	127
Total Cost (\$)	= total campaign*cost = 127×12 1.524
Total Revenue (\$)	= TP *revenue = 76×44 3.344
Total Profit (\$)	= total revenue - total cost = $3.344 - 1.524$ 1.820
NPM (%)	= (total profit/ total revenue)*100 = $(1.820/3.344) \times 100$ 54.42

BUSINESS RECOMMENDATIONS

BUSINESS IMPACT RETURN ON INVESTMENT

Before Modelling

Average Order Size (AOS)	= total revenue/total order = $3816*44/3816$ = $167904/3816$	44
Average Order Frequency (AOF)	= total order/total customer = $3816/3816$	1
Average Customer Value (ACV)	= AOS/AOF = $44/1$	44
Average Customer Lifetime (ACL) (year)	= first order date-last order date	1
Customer Lifetime Value (CLV) (\$)	= ACL*ACV	44

Number of new Superhosts	778	
CLV	44	
CAC	= total campaign cost/Jumlah superhost baru = $(3816*44)/778$	215,81
ROI	= CLV / CAC = $44/215,81$	0,20

After Modelling

Number of new Superhosts	76	
CLV	44	
CAC	= total campaign cost/Jumlah superhost baru = $(TP+FP)*cost/57$ = $(76+51)*44/76$	73.52
ROI	= CLV / CAC = $44/73.52$	0.60

KK Ramachandran (2023)





LIMITATIONS & DEPLOYMENT

Challenges and Limitations

Deployment

Conclusion

DEPLOYMENT LIVE DEMO

Superhost prediction by The Golden Gate

AirBnb adalah sebuah platform yang menyediakan pemilik properti (host) untuk menyewakan property kepada tamu atau penyewa.

Aplikasi ini dapat memprediksi pemilik properti dengan masukan atau input yang sudah sediakan dengan 25 kolom Feature penting.

User bisa melihat contoh pada tab bagian 'Example' untuk melihat contoh data dan hasil prediksinya

Form CSV Example

Form Input

Ubah data sesuai yang diinginkan

	review_scores_cleanliness	review_scores_communication	host_response_time	review_scores_accuracy
0	9.5	8.9	within an hour	9.5
1	7	8.5	within a few hours	9.5
2	7.5	8	within an hour	9.5
3	9.5	9.7	within an hour	9.5
4	3.4	3.2	more than a week	3.4
5	4.2	4	more than a week	4.2
6	2.5	2	more than a week	2.5
7	5.5	5.2	within a week	5.5
8	5.2	5	within a week	5.2
9	3.8	3.5	more than a week	3.8

Simpan Perubahan

Hasil Prediksi:

Prediksi

LIMITATIONS

- **Data Input Limitations**

Deployment only supports data in CSV format; manual input is not yet available.

IMPLEMENTASI

- Open Deployment URL (streamlit),
- Click the CSV menu to upload data,
- After uploading, users can view a data preview in the input form to ensure the data is correct,
- Click the "Prediksi" button to start the prediction process and display the results,
- The results will be displayed in a table, showing the prediction status (Superhost or Not Superhost),
- If needed, users can save changes or prediction results for future reference.

LINK: SUPERHOST PREDICTION

DEPLOYMENT TOOLS



The Python-based backend framework used to build the API.

Create a Model API to be publicly accessible.

For website development

CONCLUSION



KEY FINDINGS AND OUTCOMES

This project successfully developed a Streamlit-based prediction website to identify potential Superhosts based on specific characteristics.



MAIN IMPACT & VALUE

This solution helps Airbnb target marketing campaigns more effectively, minimize costs, and potentially increase revenue.



—
Thank You
—