Duy Pham - Weiqi Huang

What is important for a successful listing Post COVID?

Introduction

To understand what makes an Airbnb listing successful, we believe that providing the best service to customers is key. Therefore, we aim to identify the potential factors that contribute to higher composite ratings. By analyzing various listing attributes, we hope to uncover insights that can help hosts enhance guest experiences and improve their overall ratings.

Data Preparation

During the data preparation process, we filtered the dataset to include only listings posted after 2019 to ensure relevance. We then removed columns with more than 50% missing values and dropped rows with missing values in the selected columns. Additionally, we generated a compound sentiment score based on the listing descriptions to capture overall sentiment. To quantify overall guest satisfaction, we combined the six review score columns into a single composite score by averaging them.

Variables

Response Variables:

- Composite Rating Score (by combining review score from six columns: accuracy, cleanliness, checkin, communication, location, value)

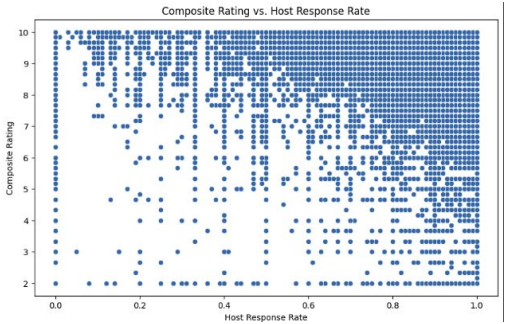
Explanatory Variables:

- Host response rate (change to decimal from percentage)
- Host is a superhost (dummy variable)
- Cleaning fee (in unit of dollars)
- Instant bookable (dummy variable)
- Compound sentiment score (score from sentiment analysing the listing description)

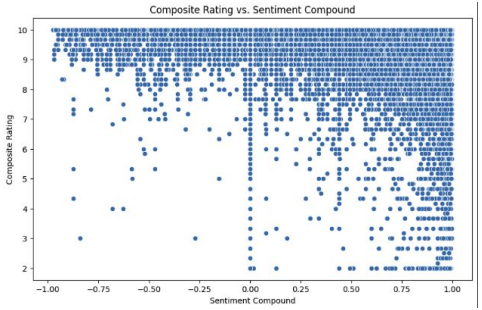
Method

For our methodology, we will first examine the relationship between each explanatory variable and the response variable to identify potential patterns and correlations. Next, we will apply a Multiple Linear Regression (MLR) model to estimate which explanatory variables have a statistically significant impact on the response variable. This approach will help us determine the key factors that contribute to higher composite ratings and assess their relative importance.

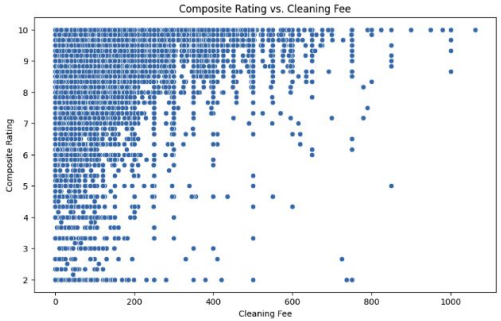
Exploratory Data Analysis (EDA)



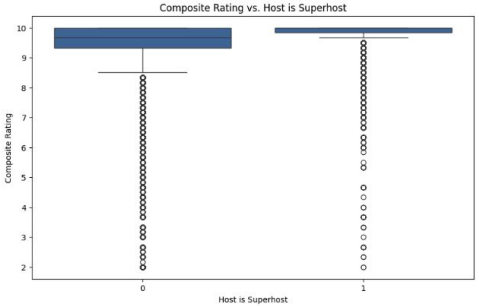
The plot suggests higher response rates correlate with higher composite ratings, highlighting the importance of prompt host communication.



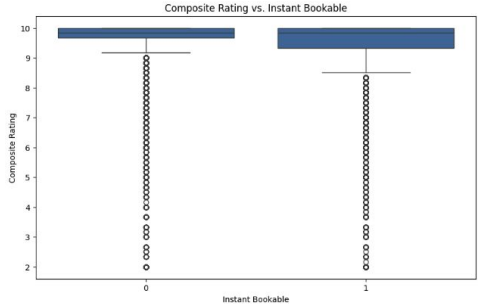
The plot suggests that while most highly-rated listings have moderate cleaning fees, higher fees do not necessarily lead to lower ratings.



The plot suggests that listings with more positive sentiment in descriptions tend to have higher composite ratings, indicating a potential link between descriptive tone and guest perception.



The box plot suggests that Superhosts generally receive higher composite ratings with less variation, indicating that Airbnb's Superhost designation may be associated with better guest experiences.



The box plot suggests little difference in composite ratings between listings with and without instant booking, indicating that booking flexibility may not significantly impact guest ratings.

Multiple Linear Regression (MLR)

```
F-statistic: 13020.657353726368
P-value: 0.0
```

F-test

We conducted an F-test to determine whether at least one predictor variable can explain the variation in the response variable. The test evaluates the overall significance of the regression model by comparing it to a model with no predictors. A statistically significant result (p-value = 0.0) indicates that at least one explanatory variable contributes to predicting the response variable, justifying further analysis of individual predictors.

	Feature	VIF
0	const	68.696850
1	host_response_rate	1.041061
2	cleaning_fee	1.012460
3	sentiment_compound	1.007793
4	host_is_superhost	1.044330
5	instant_bookable	1.022296

VIF

VIF detects multicollinearity, ensuring independent variables are not highly correlated. Here, all predictors have VIF values near 1, indicating low multicollinearity, meaning each contributes uniquely to the model. The high VIF for "const" is expected and does not affect collinearity. Running VIF confirms our MLR model is stable and reliable.

OLS Regression Results

Dep. Variable:	composite_rating	R-squared:	0.120
Model:	OLS	Adj. R-squared:	0.120
Method:	Least Squares	F-statistic:	1.302e+04
Date:	Sat, 01 Mar 2025	Prob (F-statistic):	0.00
Time:	07:40:20	Log-Likelihood:	-4.0869e+05
No. Observations:	475581	AIC:	8.174e+05
Df Residuals:	475575	BIC:	8.175e+05
Df Model:	5		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	9.1823	0.007	1336.962	0.000	9.169	9.196
host_response_rate	0.3013	0.007	45.967	0.000	0.288	0.314
cleaning_fee	-8.231e-05	9.88e-06	-8.335	0.000	-0.000	-6.3e-05
sentiment_compound	0.0838	0.003	25.622	0.000	0.077	0.090
host_is_superhost	0.3762	0.002	219.172	0.000	0.373	0.380
instant_bookable	-0.1240	0.002	-73.868	0.000	-0.127	-0.121

Omnibus:	479999.529	Durbin-Watson:	1.939
Prob(Omnibus):	0.000	Jarque-Bera (JB):	39180428.632
Skew:	-4.917	Prob(JB):	0.00
Kurtosis:	46.365	Cond. No.	1.35e+03

MLR

A successful Airbnb listing post-COVID is influenced by factors such as host response rate, sentiment in listing descriptions, Superhost status, and booking flexibility. The positive coefficients for these variables suggest that higher host responsiveness, positive sentiment, and Superhost status contribute to higher ratings. Conversely, instant booking has a slight negative impact, indicating that more personalized booking processes might enhance guest satisfaction. Cleaning fees have a small but statistically significant negative effect, suggesting that guests may be sensitive to additional costs.

Running a Multiple Linear Regression (MLR) helps by quantifying the impact of each predictor on composite ratings while controlling for other factors. This approach allows us to identify the most influential variables, helping hosts make data-driven decisions to optimize their listings and improve guest experiences.