Best Hosts

To discuss what the best hosts are doing, the criteria by which they are judged must first be defined. In the absence of information concerning hosting costs, the success of a host can ultimately be determined by its overall level of revenue. The level of revenue can be expressed as a function of the property's price and occupation rate. However, while proxies for the occupancy rate of the properties, including their availability rates, were considered, none of the features in the dataset were determined to have a substantially similar definition to the occupancy rate of a property. Given these definitions and the lack of critical information, the analysis for best practices for hosts has been conducted primarily regarding property price.

Of the three best performing non-ensemble models (XGBoost, LightGBM, random forest), three variables were identified as having the most importance: the cleaning fee of the listing as a percentage of the price, if the listing was for a private room, and the number of bedrooms within the listing (Appendix N). Of these three variables, the number of bedrooms had a feature importance rating of almost 100 in both the random forest and the XGBoost and is hence likely the most significant factor for hosts aiming to increase their listing price.

Table 16 Coefficients of the Identified Features

	coef	Std err	t	P>|t|	[0.025	0.975]
Intercept	97.635	10.212	9.560	0.000	77.617	117.653
Bedrooms	307.934	5.646	54.551	0.000	296.907	319.040
Cleaning_fee_perc	-35.304	2.161	-16.336	0.000	-39.541	-31.068
Room_type_Private_room	-96.687	4.766	-20.287	0.000	-106.029	-87.345

Table 16 shows the coefficients of the identified features estimated through linear regression. This finding is highly intuitive as the number of bedrooms within a listing is typically associated with the number of available occupants and the size of the property. However, it should be noted that it may be difficult for hosts to increase the number of bedrooms within their listings without incurring significant renovation costs.

The second most important feature was room_type_private_room, i.e., if the listing was for a private room. The estimated beta coefficient for this feature from a simple linear regression was approximately -96.687, suggesting that the price for listings of private rooms was $96.687 lower than other listings on average, with all other factors held equal. This finding is unsurprising as private rooms often use a lower price to differentiate from traditional hotel rooms.

Finally, the third most important feature was the cleaning fee as a percentage of the listing price. This is also intuitive as the cleaning fee is not included in the listing price and effectively increases the listing price. Hence, when the cleaning fee was at a lower percentage of the listing price, customers were willing to pay for a higher price. Although this feature is simple for hosts to adjust, further research can be done into the psychological effects of higher listing prices and higher cleaning fee percentages on the perceived value proposition of the property.

However, it is essential to recognise the limitations of the above analysis. It should be noted that the importance of the features was calculated based only on the data from the training set, a limited sample of the broader population. Hence, as more data becomes available, the importance of these features may change as sampling bias is reduced. Therefore, the analysis in this section should be considered suggestions for hosts and other stakeholders to consider when determining the rental prices of their listings, rather than rules. Furthermore, due to the

lack of availability of crucial data, the occupancy rates of the listings have gone unexplored. Intuitively, the demand for a property is likely to decrease when its listing price increases. Hence, the maximisation of listing price may not always be desirable toward the goal of maximising revenue. As data regarding occupancy rates become available, the models should be re-evaluated and feature importance re-calculated to generate a more accurate and comprehensive understanding of the relationship between specific features and the revenue potential of a property and the balance between price and occupancy.