INFERENCE REPORT

**Data Understanding**

* The dataset has 1054 rows and 16 columns. There are both numerical and categorical variables. Numerical variables include features like price, minimum nights, latitude, longitude, etc.
* Categorical variables include name, host name, neighborhood group, neighborhood, room type, and last review.

**Missing Values**

* There is one missing value in the 'name' column.
* 'last\_review' and 'reviews\_per\_month' columns have 210 missing values each.

**Outlier Detection**

* Outliers are detected using boxplots and z-scores.
* Variables like 'price,' 'minimum\_nights,' 'number\_of\_reviews,' 'reviews\_per\_month,' 'calculated\_host\_listings\_count,' and 'availability\_365' have outliers.

**Checking Normality**

The data is not normally distributed, which is not necessarily a defect in this context. Some variables like 'price' may require transformation.

**Linear Regression Model**

* A simple linear regression model with 'reviews\_per\_month' as the predictor variable doesn't show a strong relationship with the target variable 'price' (R-squared: 0.0003).
* The p-value for 'reviews\_per\_month' is 0.606, suggesting it is not a significant predictor.

**Improvements**

* Further exploratory data analysis (EDA) and feature engineering may be performed.
* Handling missing values through imputation or removal.
* Transformation or removal of outliers based on the context.
* Transformation or removal of outliers based on the context.
* Trying different regression models and evaluating their performance.
* So, the goal is to refine the dataset and build a predictive model that accurately estimates the price based on the given features.