Laycee Glass

NKM2 Task 3

Data Analytics Graduate Capstone

6/4/25

Western Governors University

## NKM2 Performance Assessment, Task 3

### Student Information

Student Name: Laycee Glass

Student ID: 0011896026

Date: 6/4/25

### A. Executive Summary

**Problem and Hypothesis**

This study aimed to analyze a synthetic dataset of Airbnb listings across Pennsylvania to identify key factors influencing nightly rental prices and to develop a predictive model. The hypotheses were as follows:

* Null Hypothesis (H0): There is no statistically significant relationship between any of the independent variables and nightly Airbnb rental prices in Pennsylvania, and a multiple linear regression (MLR) model cannot predict price with acceptable accuracy (R2 ≤ 0.60).
* Alternative Hypothesis (H1): At least one independent variable is statistically significant in predicting nightly Airbnb rental prices in Pennsylvania, and a MLR model can predict price with acceptable accuracy (R2 > 0.60).

**Data Analysis Process**

A synthetic dataset was generated using Python’s Faker package, emulating realistic Airbnb listings across Pennsylvania. The dataset comprised 21 variables, from listing city to number of bathrooms to WiFi availability.

Data preparation involved checking for missing values (none found), removing duplicates, and assessing outliers using the interquartile range method. Outliers were retained as they represented plausible values.

Exploratory data analysis (EDA) was conducted using matplotlib and seaborn to examine variable distributions and relationships. Categorical variables were one-hot encoded, with dummy variables renamed for clarity, and converted to float data type for modeling compatibility.

After EDA, a MLR model was fitted using the statsmodels library. Initial diagnostics indicated potential multicollinearity, and as such, variance inflation factor scores were calculated to iteratively remove variables with high multicollinearity. Backward elimination based on p-values further refined the model, resulting in a final model with 16 variables explaining over 93% of the variance in nightly Airbnb prices.

**Key Findings:**

* R2 = 0.931, indicating strong explanatory power
* RMSE = $18.12, MAE = $13.95, reflecting strong predictive accuracy
* Top influencing predictors included:
  + Accommodates: +$10.07 per additional guest
  + Bathrooms: +$29.07 per additional bathroom
  + Superhost: +$10.47
  + Private room: –$41.14 vs. entire home (baseline)
  + Shared room: –$28.96 vs. entire home (baseline)
  + Philadelphia listings: +$20.85 vs. Allentown (baseline)
  + Erie listings: –$21.29 vs. Allentown (baseline)
* Diagnostic plots confirmed linear regression assumptions: normality of residuals and homoscedasticity

Based on these results, the null hypothesis was rejected, supporting the conclusion that at least one independent variable was statistically significant, and a MLR model can predict prices with acceptable accuracy (R2 > 0.60).

**Limitations**

A limitation for this study was the use of synthetic data. This limited the generalizability of the findings to real-world Airbnb markets. Along the same vein, the model also could not account for temporal factors such as seasonality, special events, and market fluctuations.

**Proposed Actions**

1. Validate with real-world Airbnb data to assess external validity.
2. Develop dynamic pricing tools using model outputs to guide new/inexperienced hosts.
3. Use insights for strategic listing decisions (e.g., offering WiFi or becoming a superhost).

**Expected Benefits**

1. A model that explains over 93% of price variance, providing a reliable pricing framework.
2. Actionable recommendations to help Airbnb hosts optimize listings and increase revenue.

### B. Presentation of Findings

A slide-based presentation summarizing the study’s visualizations, diagnostics, and recommendations is attached alongside this executive summary.

### C. Sources

No external sources were used specifically for this executive summary. All sources for the study itself can be found in Task 2 documentation.