Capstone Project Final Report

1. Data Wrangling/Cleaning Process

- Loaded dataset with 16,278 rows, 48 columns of various data types.

- Removed rows with excessive missing values, reducing dataset to 14,540 rows.

- Standardized 'Created Date' to datetime and removed non-variate 'Location Type'.

- Removed rows with zero 'adr', resulting in a dataset of 4,208 rows.

- Conducted outlier analysis and saved cleaned data for further analysis.

2. Exploratory Data Analysis (EDA) Process

- Visualized distributions for 'adr', 'Occupancy Rate LTM', and 'Overall Rating'.

- Found 'adr' right-skewed and 'Occupancy Rate LTM' bimodal.

- Relationships analyzed between features and key indicators.

- Conducted correlation and significance testing.

- Analyzed yearly trends in property listings and conducted multicollinearity check.

- Proposed feature engineering ideas.

3. Modeling Process

- Converted categorical variables into dummies and standardized numeric features.

- Split dataset into training and testing sets (75% train, 25% test).

- Evaluated Linear Regression, Random Forest, and Gradient Boosting models.

- Conducted hyperparameter tuning using GridSearchCV.

- Found Random Forest model as the best performing model after tuning.

- Assessed business implications and mathematical considerations of model metrics.

4. Recommendations for Airbnb Property Investors

1. Focus on Amenities: Prioritize properties with pools, hot tubs, and other amenities that significantly impact occupancy rates and rental prices.

2. Optimize Listing Types: Invest in properties categorized as 'Entire home/apt' due to their higher average daily rates.

3. Continuous Model Evaluation: Regularly reassess and fine-tune models to adapt to market changes and maintain prediction accuracy.

Future Research:

- Explore alternative modeling approaches for better handling skewed distributions.

- Consider more extensive feature engineering and variable transformation.

- Investigate the impact of geographic and temporal factors on occupancy rates.