# **Executive Summary**

## Statement of Problem and Hypotheses

The research question asked which independent variables from the Airbnb data set affect the dependent variable “price” and to what magnitude. The proposed hypotheses are below:

**Null hypothesis**- H0: The independent variables ‘accomodates’, ‘beds’, ‘bathrooms\_text’, ‘bedrooms’, ‘amenities’, ‘room\_type’, and ‘city’ do not affect the dependent variable ‘price’ in a statistically significant way based on an alpha value of 0.05.

**Alternate Hypothesis**- HA: The independent variables ‘accomodates’, ‘beds’, ‘bathrooms\_text’, ‘bedrooms’, ‘amenities’, ‘room\_type’, and ‘city’ affect the dependent variable ‘price’ in a statistically significant way based on an alpha value of 0.05.

## Summary of the Data Analysis Process

The data-analysis process used was to clean the data and prepare it for stepwise linear regression, LinearRegression() from sklearn.linear\_model, and a random forest regression. After data exploration, cleaning, and preparation, there were 207,968 rows and 26 columns of data used in the analysis. Ordinary Least Squares (OLS) was then used with a stepwise method of removing columns based on multicollinearity and p-value. Then, using the same data, LinearRegression() from sklearn.linear\_model was then utilized. Thirdly, the same data was used and ran through a random forest regression after determining ideal values. Finally, these results were all compared.

## Outline of the Findings

For the three findings, the results of the evaluation scores are shown below:

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Stepwise** | **LinearRegression** | **Random Forest** |
| RMSE | 0.42 | 0.42 | 0.41 |
| MSE | 0.18 | 0.18 | 0.17 |
| R-squared | 0.49 | 0.49 | 0.52 |
| Mean Score | 0.0 | 0.0018 | 0.0023 |

The random forest regression was shown to have the best values for RMSE, MSE, and R-squared. For RMSE and MSE, lower values show a superior fit, and for R-squared, a higher value shows a greater explanation of the dependent variable. Random forest was the best model of the three, but it still is not accurate enough to be used as a model. The R-squared value only explains 52% of the values and predictive data compared to test data proved that to be accurate.

## Explanation of Limitations of the Techniques and Tools Used

Linear regression has a set of assumptions that are required which can create issues. Those include a linear relationship between the dependent and an independent variable, lack of autocorrection, lack of multicollinearity, and especially outliers can prove problematic. (GeeksforGeeks, 2017). Some columns were difficult to translate into usable data due to the way the original data set was created.

## A Summary of the Proposed Actions

Since the null hypothesis had to be accepted in this case due to not all variables listed having a statistically significant correlation with “price.” Furthermore, the models created were not accurate enough to recommend usage for determining “price” based on the variables given within this data set. Proposed actions are to compile more data points, additional variables, and to repeat the process once again and to also use similar Airbnb data from other areas to attempt to see how those models perform in an attempt to gain greater insight.

## Expected Benefits of the Study

Being able to determine “price” before entering the rental property market would be greatly beneficial for mitigating lower rental rates or for maximizing profits based on properties in the area. Mispricing a property by 20 dollars would cost $4,800 a year in opportunity cost based on 20 days of rental per month, for instance. The number of people a property can accommodate was shown within this study to be the most important feature when predicting price based on the data set.

## Sources

GeeksforGeeks. (2017, March 19). *Linear Regression (Python Implementation) - GeeksforGeeks*. GeeksforGeeks. https://www.geeksforgeeks.org/linear-regression-python-implementation/

## Panopto Video Link

https://wgu.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=12e51fde-c887-4136-82e8-b1bf01104c04