# BIA-652 Final Project Description

Airbnb is a travel app that allows people to find rooms to rent in locations all over the world. The service can be convenient and often more affordable than a hotel while you are traveling. Airbnb has become popular in recent years, and there are now many listings available in most major cities.

You are recently hired as a data analyst to look into the **determinants of the number of reviews each listing receives per month (reviews\_per\_month**). If this variable is missing, assume that the listing **received 0 reviews**. The dataset you are working with contains **24886 Airbnb listings in Rio de Janeiro**. The data can be downloaded from <https://drive.google.com/drive/folders/1z_3I_3k6E6CBWXwCYQpKE0KZabZOkWa9?usp=share_link>

The business team hope that your analysis can help them confirm or refute some earlier hypotheses. For example, **they believe that larger properties should receive more reviews** **because larger properties can accommodate more guests and therefore generate more traffic.**

In addition, they believe **that whether a property is overpriced is one of the most important factors in determining how many reviews it will receive.** In general, **they believe listings that are priced higher than listings of similar sizes and/or locations will receive fewer reviews than those that are priced lower.** This is because people are more likely to leave a review if they feel like **they got good value for their money**.

Finally, the management asks you to **build a predictive model for reviews\_per\_month**. You should **compare different models or conduct variable selections**. Other than the factors **mentioned by the business team, what else are important predictors for reviews\_per\_month**?

Write a report of 3 to 5 pages summarizing your results. Your code or notebook should be attached in a separate file and well-documented. Both the report and the code/notebook file should be submitted on Canvas. The next page contains the grading rubric.

# Grading Rubric

|  |  |  |  |
| --- | --- | --- | --- |
|  | **C+ and Below** | **B- to B+** | **A- to A** |
| **Analysis** | Analysis is overly simplistic or incomplete | Analysis is adequate | Complete, accurate, and informative analysis |
| **Results** | Conclusions are missing, incorrect, or not based on analysis. | Conclusions relevant, but partially correct or partially complete | Relevant conclusions explicitly tied to analysis and to context |
| **Writing** | Explanation is illogical, incorrect, or incoherent. There are few complete sentences, only bullet points | Explanation is largely correct. Complete sentences are used. | Explanation is correct, complete, convincing, and elegant. Complete sentences are used. |
| **Organization** | The report is not eﬀectively organized. There may be parts of the report that has no explanation or no clear explanation. Graphics and tables are of poor quality | The report is overall clear but occasionally difficult to follow. Graphics and tables are included at appropriate points. | The report has a logical structure. The chain of reasoning or progression of ideas is clear. Tables and graphs are easy to follow and clearly presented |
| **Code Readability** | Code is messy and poorly organized; unused or irrelevant code distracts when reading code. Variables and functions names do not helpful to understand code. | Code is reasonably well organized. Variable and function names generally meaningful and helpful for understanding. | Code very well organized. No irrelevant or distracting code. Variable and function names have clear relationship to their purpose in the code. Code is easy to read and understand. |
| **Code Reproducibility** | Most of the code failed to run | A few parts of the code failed to run | Code correctly loads data and generate all results and figures in the report |