Final Project - OPIM 606: Programming I – Introduction to Data

Project Description and Guidelines

Airbnb is a company that provides an online marketplace for short-term rentals of homes and apartments. Much of the data from Airbnb's website has been compiled and made publicly available on the website Inside Airbnb. For this assignment, you will analyze a sample of the Airbnb listings from Washington, DC, scraped in September 2022. Each row in this dataset represents a single Airbnb listing.

The main goal of the assignment is to use your R skills in conducting an exploratory analysis of the hosts and the listings. In doing so, you should also build regression models to measure the predictive power of subsets of the variables in predicting the price of a listing. In addition, given this sample, you are asked to test whether the average rating of all listings in the population is more than 4.

In the real world, you would, of course, have very substantial latitude in how to undertake the analysis and present the results. For the sake of setting expectations for this assignment, though, you will need to do the following:

1. Describe each table, plot, or output in your report briefly but clearly.

2. Combine the two data sets ("Listings.csv" and "Reviews.csv".)

3. The final dataset used for the analysis should not have any missing values.

4. Include at least one but no more than four figures. Each figure may have multiple panels.

5. Create at least one function and incorporate it into your analysis.

6. Have at least two chunks of code written in the tidyverse package's piping form.

7. Implement two meaningful regression models for the "price" of the listings. Compared/present the results of the two models as a table. (Note: The point is not to build the best regression models in terms of fit. The point is to simply build and compare two relevant regression models, either simple linear ones or multiple linear ones.)

→

Deliverables:

You should submit your well-commented and fully functional R code and a PDF report displaying and explaining the result of your analysis. Limit the total pages of your PDF document to 5 pages.

In terms of the format of the submission files, you have two options:

- **Option A**: Submit an R Markdown file (.Rmd) and a PDF fie. The R Markdown should contain every step - from loading the data to producing the report. Use R Markdown chunk settings so that the R code is not visible in the final PDF output.
- **Option B**: Submit an R file (.R) and a PDF file. The R file should contain every step - from loading the data to producing the report.

Evaluation:

Each student's grade will be made up of two different components including the report/code and the evaluation submitted by their teammates.

The written report and the code account for 75% of your grade. This is where the overall quality of the project will be evaluated using the following framework. Each student in the team is expected to get the same grade this part.

- Whether the instructions were followed.
- Quality of the R code (including annotation)
- Quality of the analysis
- Clarity of the explanation
- Clarity/visual appeal of the figures

The feedback of your group members about your performance in the team accounts for 25% of your grade. If you fail to submit the form, you will automatically lose the feedback score.

Instructor reserves the right to adjust each individual's grade and/or the grading distribution based on feedbacks received from the team members. In a circumstance that a student does not actively participate in team's activities and does not contribute to the team deliverables, instructor reserves the right to assign 0 for the individual's total project grade.