IST 5520, Fall 2022, Chen

**Project Evaluation Form – Milestone 2: Data Analysis I**

**Submission Due Oct 30, 11:59 PM**

**Instruction:**

1. Cleanse and visualize data. The project report should include:

* Introduction (refined from M1)
* Data Source and Collection (refined from M1)
* Data manipulation (newly developed)
* Data summarization and visualization (newly developed)

Read the evaluation criteria carefully on the next page for the detail.

1. Use Markdown in jupyter notebook to write your project report. You need to use proper Markdown syntax to format your report. Do not use MS Word or other format.
2. Please submit the following documents into Canvas:

* The project report written in .ipynb file;
* The Evaluation form with full project team information (see below table).

**Project Team Information (filled in by students)**

|  |  |  |
| --- | --- | --- |
| **Member name** | **Percent contribution** | **Activities completed by the member** |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

**Evaluation Summary – M2 (filled in by instructor)**

|  |  |  |  |
| --- | --- | --- | --- |
| **Criteria** | **Target %** | **Comments** | **Evaluation** |
| * Refine your report based on M1. * Extract and transform potential variables from the data source(s). Cleanse your dataset(s). | 20 | The second research question seems unsolvable using your current dataset. | 20 |
| * Manipulate and clean your dataset properly. * Properly deal with categorical variables. * Properly detect outliers and deal with missing values in your dataset. | 30 | No need to print out a very long list of prices in line 148. This makes your jupyter notebook hard to read. Similarly, no need to print out all neighbors in line 151. | 30 |
| * Summarize and visualize data by using appropriate methods. * Professionally interpret your data summary and visualization. * Use various dimension reduction techniques (visualization, correlation, principal component analysis, variable selection etc.) to explore your data. * Provide summary statistics, correlation table/plot, and at least 4 professional graphs with detailed and proper interpretations. | 30 | Your data analyses can be more comprehensive and in-depth. The dataset contains rich information. You need to explore more factors, especially those not very straightforward variables, that influence price so that your analyses can be more comprehensive and interesting.  The map visualization section looks not complete. Did you submit all of your work?  It’s misleading to include id and host\_id in the correlation matrix (line 176), since they are categorical variables. | 29 |
| * Format your project report in a professional way. * Professionally organize your contents to show your data management and analysis efficiently and concisely. * Write your project report by using appropriate Markdown syntax. | 10 | The output lines of the jupyter notebook are not sequential. I wonder whether your jupyter notebook can work correctly if we run it again.  Before you submitted the jupyter notebook, you should re-run it. | 9 |
| * Use the repository to manage all your project documents including meeting schedules, meeting minutes, and the proposal (the instructor should be able to verify this). * Your github repository should contain the latest documents for your project deliverables. | 10 |  | 10 |
| The report satisfies all of the following criteria:   * It tells a very interesting story; * The data manipulation methods are professionally applied; * The whole document is well written with no or few grammar or writing issues. | 10 bonus | Data analysis is about telling stories. Delve into the dataset and do a more comprehensive analysis to find more interesting patterns. In terms of writing, use a storyline to organize all data analysis and tell interesting stories.  Another issue is that some members did not contribute significantly to the project. If this continues to be a problem, students without significant contributions will get a reduction in the project points. |  |
| **Total** | **100** |  | **98** |