Assignment: Data Wrangling and Visualization

Value: 10% of Final Grade

Description   This assignment focuses on foundational data wrangling techniques and the ability to effectively clean, analyze, and visualize data. You will work with real-world datasets to practice essential skills in data preparation, exploration, and integration. Your goal is to transform raw datasets into a format ready for analysis, draw meaningful insights using descriptive statistics, and present findings visually.

Assignment Objectives   By completing this assignment, you will:
1. Develop skills in cleaning and preprocessing raw data, including handling inconsistent data.
2. Conduct exploratory data analysis using descriptive statistics.
3. Integrate multiple datasets by performing joins and ensuring consistency in the merged data.
4. Create meaningful visualizations to summarize and communicate your findings.

Assignment Tasks

1. **Data Cleaning**
   - Load the provided datasets and examine their structure (e.g., columns, data types, unique values).

   - Identify and correct issues such as outliers, duplicates, and inconsistent formatting.
2. **Joining Datasets**
   - Combine tne two related datasets by performing appropriate join operations (e.g., inner join, left join).

   - Ensure the resulting dataset maintains consistency and completeness.
3. **Descriptive Statistics**
   - Calculate key statistics (e.g., mean, median, mode, variance, and standard deviation) for relevant variables.

   - Identify trends or anomalies based on these statistics.
4. **Data Visualization**
   - Create at least three visualizations to highlight key findings.

   - Use appropriate chart types (e.g., bar charts, scatterplots, heatmaps) and ensure clarity in labels, legends, and formatting.

Deliverables   Submit your .Rmd and your knit .html file for the analysis. I will be running your .Rmd file so please make sure your submission is executable and replicable. Remember to document your process thoroughly as with descriptions in the text of the .Rmd file.

Assessment Criteria   Your submission will be graded on the following:

1. **Data Cleaning and Preprocessing (25%)**

- Effectiveness in identifying and resolving data quality issues.

2. **Descriptive Statistics and Analysis (20%)**

- Appropriateness and accuracy of the statistical analysis.

4. **Dataset Integration (20%)**

- Correctness and clarity in joining datasets.

5. **Visualizations (20%)**

- Relevance, clarity, and aesthetic quality of the visualizations.

6. **Documentation and Presentation (15%)**

- Clear explanations, well-commented code, and professional formatting of the submission.