

Assignment 2: Exploratory Data Analysis (EDA), Data Insights, and Methodological Foundation

Course: Data Science

Class: BSCS-F22

Instructor: Ghulam Ali

Course Learning Outcome: CLO-02 GA-03

Due Date: Nov 10, 2025 8:00 PM

Objective:

The objective of this assignment is to build on the work done in **Assignment 1** by exploring the data in-depth, extracting meaningful insights, and laying the groundwork for answering the key questions identified earlier. Students will perform Exploratory Data Analysis (EDA), visualize data patterns, and develop an initial methodology to approach their problem as a data scientist.

Assignment Tasks:

1. Exploratory Data Analysis (EDA):

- Conduct a thorough EDA on the dataset collected in Assignment 1.
- Include descriptive statistics (mean, median, mode, standard deviation, etc.).
- Analyze distributions, detect skewness, and assess data balance.
- Identify relationships among variables using correlation matrices and pair plots.
- Visualize data through graphs such as: Histograms, Boxplots, Bar charts, Line graphs, Scatter plots, Heatmaps, Highlight important trends, patterns, and anomalies discovered

2. Data Enhancement (if required):

- Assess whether the current dataset is sufficient to answer all the key questions listed in Assignment 1.
- If necessary, collect additional data to support the analysis and improve the depth of insights.
- Document the source and integration process of any new data added.

3. Foundation for Question Answering:

- Revisit the key questions identified in Assignment 1.
- For each question:
 - Briefly explain how the data can help answer the question.
 - Propose a preliminary methodology (statistical analysis, hypothesis testing, machine learning models, etc.) for answering it.
 - Justify your choice of method.

4. Documentation:

Students must submit a well-structured report covering:

- I. **Introduction:** Brief recap of the problem and dataset.
 - II. **EDA Summary:** Description of analyses conducted and insights derived.
 - III. **Visualizations:** Clear and labeled graphs with brief interpretations.
 - IV. **Data Enhancement:** Explanation and justification for any additional data collection.
 - V. **Key Questions Revisited:** Clear mapping of each question to data insights and proposed methodology.
 - VI. **Methodology:** Description of the approach/algorithms that might be used to answer the questions in subsequent analysis.
 - VII. **Code Snippets:** Include relevant EDA and visualization code.
 - VIII. **Conclusion:** Summary of findings and readiness for the next step.
-

Submission Guidelines:

1. The report should be in PDF format.
2. The code must be submitted separately in a Jupyter Notebook (.ipynb).
3. Proper citations should be included if external sources are used.
4. Submit your assignment on Google Classroom by the due date.