

Assignment 1: Basic Data Exploitation

Total Marks 100

Objective

The goal of this assignment is to practice conducting basic data exploitation by calculating summary statistics, creating histograms, and generating scatter plots using three datasets of your choice.

Instructions

1. Dataset Selection

- Download the following datasets and complete all the tasks specified for each dataset.
- Ensure the dataset has both numerical and categorical variables for analysis.
 1. <https://archive.ics.uci.edu/dataset/19/car+evaluation>
 2. <https://archive.ics.uci.edu/dataset/1/abalone>
 3. <https://archive.ics.uci.edu/dataset/14/breast+cancer>

2. Summary Statistics

Marks (2 x 5) 10

- Calculate the following summary statistics for at least three numerical variables:
 - Mean
 - Median
 - Standard deviation
 - Minimum and maximum
 - Count of missing values (if any)

3. Histogram Creation

Marks (10+10) 20

- Create histograms for all numerical variables in the dataset.
- Interpret the shape of the distribution (e.g., normal, skewed, bimodal).

4. Scatterplot Generation

Marks 20

- Generate every possible scatterplots to visualize relationships between pairs of numerical variables.
- Clearly label the axes and provide a title for the scatterplots.
- Identify and describe any trends, correlations, or outliers.
- Use color coding in scatterplots to distinguish categories in a third variable (if applicable).

Deliverables

- Submit a short report (2-3 pages) including: (Marks 50)
 - The dataset description (source, variables included).
 - Summary statistics table with key observations.
 - Histograms with interpretations of the distributions.
 - Scatterplots with an explanation of any patterns or trends identified.
- Save your code in a Python notebook (e.g., Jupyter Notebook) or share the script file if applicable.

Evaluation Criteria

- Completeness: All required tasks are completed.
- Accuracy: Correct calculation of summary statistics and properly plotted graphs.
- Presentation: Clear labeling, titles, and interpretations in visualizations and report.
- Insights: Quality of analysis and interpretation of results.

Additional Notes

- Feel free to ask questions if you need help or clarification.
- Ensure your submission is original and properly referenced if you use external data sources.