IT2011 Group Assignment – Progress Review I: Viva on Data Preprocessing & EDA

you are required to attend the **Progress Review I Viva**, which focuses on your **Data Cleaning, Preprocessing, and Exploratory Data Analysis (EDA)**.

Progress Review I – Viva Details

- **Mode**: In-class (during your regular lab session)
- **Duration**: Approximately **20 minutes per group**
- **Format**: Viva-style Q&A + brief walkthrough of your notebook/script

Individual Requirement (20 marks per student):

- Each group member **must handle and present at least one preprocessing technique**, such as:
    - Handling missing data
    - Encoding categorical variables
    - Outlier removal
    - Normalization / scaling
    - Feature engineering (Feature selection and dimension reduction) etc.
- Each member must:
    - Clearly explain the selected technique
    - Justify why it was needed for your dataset
    - Show the implementation (code/output)
    - Present at least one EDA visualization (e.g., histogram, boxplot, correlation matrix) and interpret it

Students who do not present or fail to show individual contribution will receive zero for the individual component.

Group Requirement (5 marks shared):

- Submit a **combined preprocessing pipeline** that reflects proper integration of each member's work
- Demonstrate logical flow and collaboration
- Clearly commented and organized code

Please ensure each team member is prepared with their section of the code and visual output. This is a critical milestone in your assignment and contributes directly to your final evaluation.

Evaluation Details

Format: Viva

Duration: 15 minutes per group

Mode: In-class during Lab Sessions

Date and Time: During Lab sessions of 4th Week of August **(starting week- 22nd September 2025)**

<u>Repository Layout for Group Deliverable folder (Upload a Zip file) which need to be upload to Courseweb</u>

```
Group_ID/
│
├── README.md
│   └── Overview of the project, dataset details, group member roles, how to run code
│
├── data/
│   ├── raw/            # Original dataset(s)
│   └── external/       # Any external reference datasets (if used)
│
├── notebooks/
│   ├── IT_Number_Preprocessing_technique.ipynb     # Member 1 – IT Number & Preprocessing_technique
│   ├── IT_Number_Preprocessing_technique.ipynb # Member 2 – IT Number & Preprocessing_technique
│   ├── IT_Number_Preprocessing_technique.ipynb         # Member 3 – IT Number &
Preprocessing_technique
│   ├── IT_Number_Preprocessing_technique.ipynb # Member 4 – IT Number & Preprocessing_technique
│   ├── IT_Number_Preprocessing_technique.ipynb # Member 5– IT Number & Preprocessing_technique
│   ├── IT_Number_Preprocessing_technique.ipynb # Member 6 – IT Number & Preprocessing_technique
└── group_pipeline.ipynb          # Integrated pipeline (combined work)
│
└── results/
    ├── eda_visualizations/          # Plots & charts (PNG/JPEG)
    ├── logs/                  # Any logs from execution (Optinal)
    └── outputs/                # Final processed dataset / features
```

Individual Evaluation (20 marks per member)

Criteria	Max Marks	Mark Gained	Excellent (Full Marks)	Good (75%)	Satisfactory (50%)	Poor (25% or 0%)
1. Preprocessing Technique	6		Advanced and appropriate technique(s) used; clear explanation and justification	Suitable technique used with minor gaps in execution or clarity	Basic or partially correct preprocessing with minimal reasoning	Incorrect or missing preprocessing
2. Code Accuracy & Execution	6		Code runs without error; meets all requirements with efficiency and readability	Minor errors; mostly meets functional goals	Partially functional code; inefficient or lacking clarity	Code does not run or is incomplete
3. EDA Output & Insight	4		Well-structured EDA; clear insights with visualizations and interpretation	Adequate EDA with reasonable insight	Basic EDA; limited insight or interpretation	EDA missing or lacks insight
4. Communication Clarity	4		Clear, confident presentation with structured flow; all points well-explained	Mostly clear with minor gaps in explanation or delivery	Adequate explanation with some disorganization or delivery issues	Incoherent or poorly delivered presentation

Group Evaluation (5 marks shared)

Criteria	Max Marks	Mark Gained	Excellent (Full Marks)	Good (75%)	Satisfactory (50%)	Poor (25% or 0%)
5. Integrated Preprocessing Pipeline	2		Seamless integration of individual components; clear collaboration in execution	Mostly integrated with minor issues	Some evidence of integration; loosely connected modules	Components disjointed or inconsistent
6. Logical Flow & Completeness	2		All steps followed logically from problem to results; complete implementation	Mostly logical; minor gaps in process	Partially complete; some inconsistencies	Illogical flow or incomplete work
7. Collaboration Evidence	1		Clear distribution of work, peer evidence submitted, strong teamwork shown	Partial evidence of collaboration	Limited signs of teamwork	No or poor evidence of collaboration