Tasks on Job posting dataset

Data Cleaning & Preprocessing Tasks

1 Handling Missing Values

- Check for **NaN (missing values)** in all columns.
- Fill missing values using **mode (for categorical)** and **median (for numeric)** where appropriate.
- Drop irrelevant columns (e.g., `Attach`, `Notes` if they contain too much noise).

2 String Operations & Text Processing

- Clean `JobDescription` and `JobRequirment` by:
    - Removing **HTML tags, special characters, and excessive whitespace**.
    - **Tokenizing and lowercasing** for NLP tasks.
    - Extracting **key skills** (Python, Java, SQL, etc.).
- Normalize `Title` (e.g., "Software Eng." → "Software Engineer").
- Standardize `Term` (e.g., "FT", "full time" → "Full-time").

3 Date Operations

- Convert date, `StartDate`, `OpeningDate`, `Deadline` into **datetime format**.
- Calculate:
    - **Days left to apply** → `Deadline - OpeningDate`.
    - **Job start delay** → `StartDate - OpeningDate`.
    - Extract **year & month** from dates (Year, Month).

4 Splitting & Extracting Data

- Split `Location` into **City & Country**.
- Extract **numeric salary range** from Salary (e.g., "50K-60K USD" → `min_salary=50000`, `max_salary=60000`).

5 *Filtering & Categorization*

- Identify **IT-related** jobs from the IT column.
- Categorize `Eligibility` and `Audience` into structured **Education Level / Experience Ranges**.