

# Data Scientist - Gen AI Assessment

## Project: Using AI to process a Financial Statement

### Objective:

Build a prototype system to automatically extract key financial metrics from financial statements and generate a summary report using GenAI.

### Instructions:

#### 1. Problem Understanding:

- You are required to process a set of financial statements to extract key metrics such as revenue, net income, operating expenses, and cash flow.
- Based on the extracted data, generate a summary report highlighting the financial health of the company.

#### 2. Data:

- You will be provided with a financial statement in PDF format, which includes Balance Sheets, Income Statements, and Cash Flow Statements for multiple companies.

#### 3. Tasks:

- As below

#### Task 1: Data Extraction

- Write a script to parse the provided PDFs and extract text data. You can use libraries or any tool of your choice.

#### Task 2: Data Preprocessing

- Clean the extracted text data to remove any unnecessary information (e.g., headers, footers, page numbers).
- Segment the text into meaningful sections (e.g., identify sections containing revenue, expenses, and cash flow information).
- Validation of accurate data processing.

#### Task 3: Build GenAI Model

- Use GenAI to build a model that can:
  1. Identify and extract key financial metrics from the pre-processed text.
  2. Understand the context and handle long documents efficiently.
- Provide code snippets or documentation describing how you implemented the GenAI model.

#### Task 4: Generate Summary Report

- Based on the extracted data, generate a summary report that highlights the financial health of the company. The report should include:
  - Key financial metrics (revenue, net income, etc.)
  - Any notable trends or observations
  - A short narrative summary in natural language

#### Task 5: Scalability and Measurement

- Discuss and implement strategies for scaling the model to handle a large volume of financial statements.
- Optimize the model for performance and accuracy.

#### 4. Deliverables:

- **Code:** Provide code.
- **Reports:**
  - Extracted financial metrics in a tabular format (e.g., CSV or Excel).
  - Generated summary reports for 2-3 financial statements.
- **Documentation:** A concise report describing your approach, challenges faced, and solutions implemented.
- **Readme File:** A README file with clear instructions on how to run your solution.

#### 5. Evaluation Criteria:

- **Correctness:** Accuracy in extracting relevant financial metrics and generating summaries.
- **Completeness:** All requested deliverables provided.
- **Code Quality:** Readability, modularity, and documentation of your code.
- **Innovation:** Creative solutions to tackle the problem.
- **Scalability:** Consideration of how the solution could be scaled and tested for production use.

#### 6. Submission Guidelines:

- Provide a link to your GitHub repository or share a ZIP file containing your project.
- Ensure your submission is organized and easy to navigate.