Assignment Tasks:

Task 1: Exploratory Data Analysis (EDA) and Business Insights

1. Perform EDA on the provided dataset.

2. Derive at least 5 business insights from the EDA.

○ Write these insights in short point-wise sentences (maximum 100 words per

insight).

Deliverables:

● A Jupyter Notebook/Python script containing your EDA code.

● A PDF report with business insights (maximum 500 words).

Task 2: Lookalike Model

Build a Lookalike Model that takes a user's information as input and recommends 3 similar

customers based on their profile and transaction history. The model should:

● Use both customer and product information.

● Assign a similarity score to each recommended customer.

Deliverables:

● Give the top 3 lookalikes with there similarity scores for the first 20 customers

(CustomerID: C0001 - C0020) in Customers.csv. Form an "Lookalike.csv" which has

just one map: Map<cust_id, List<cust_id, score>>

● A Jupyter Notebook/Python script explaining your model development.

Evaluation Criteria:

● Model accuracy and logic.

● Quality of recommendations and similarity scores.

Task 3: Customer Segmentation / Clustering

Perform customer segmentation using clustering techniques. Use both profile information

(from Customers.csv) and transaction information (from Transactions.csv).

● You have the flexibility to choose any clustering algorithm and any number of clusters in

between(2 and 10)

● Calculate clustering metrics, including the DB Index(Evaluation will be done on this).

● Visualise your clusters using relevant plots.

Deliverables:

● A report on your clustering results, including:

○ The number of clusters formed.

○ DB Index value.

○ Other relevant clustering metrics.

● A Jupyter Notebook/Python script containing your clustering code.

Evaluation Criteria:

● Clustering logic and metrics.

● Visual representation of clusters.

Submission Instructions:

1. GitHub Link

○ Upload all the PDF and code files in a public GitHub repository.

2. File Naming Convention:

○ Use the following naming convention for all your files:

■ FirstName_LastName_EDA.pdf

■ FirstName_LastName_EDA.ipynb

■ FirstName_LastName_Lookalike.csv

■ FirstName_LastName_Lookalike.ipynb

■ FirstName_LastName_Clustering.pdf

■ FirstName_LastName_Clustering.ipynb