TASK-2

Report

Deliverable: Lookalike.csv

The file contains the top 3 lookalike recommendations with their similarity scores for the first 20 customers (`C0001` to `C0020`).

Explanation of the File Format:

- **CustomerID**: The ID of the customer for whom lookalikes are provided.
- **Lookalikes**: A JSON string containing a list of dictionaries, where each dictionary represents:
  - `CustomerID`: ID of the similar customer.
  - `Score`: The similarity score calculated using cosine similarity.

Steps Taken:

1. **Feature Engineering**:
   - Created a customer-product matrix using `pivot_table`.
   - Filled missing values with `0`.
2. **Similarity Calculation**:
   - Used `cosine_similarity` to compute pairwise similarities between customers.
3. **Top 3 Lookalike Extraction**:
   - Sorted customers by similarity scores, excluding the customer themselves.
   - Picked the top 3 most similar customers for each of the first 20 customers.
4. **CSV Generation**:
   - Saved recommendations in the specified JSON-like format in a CSV file.

Example Recommendations:

For `CustomerID: C0001`:

- **C0097** (Score: 0.547723)
- **C0194** (Score: 0.469668)
- **C0199** (Score: 0.438178)

For `CustomerID: C0020`:

- **C0026** (Score: 0.617213)
- **C0007** (Score: 0.588348)
- **C0104** (Score: 0.488678)

Evaluation:

- **Average Overlap Score**: 1.23
  - This score evaluates the similarity between recommended customers based on overlapping purchased products.