Task 2: Lookalike Model

A **Lookalike Model** identifies and recommends customers similar to a given customer based on their characteristics, behavior, or transaction history. It is often used in marketing and customer analytics to find audiences with similar profiles to target existing or potential customers.

Key Components of a Lookalike Model

1. **Input Data**:
    - **Customer Profile**: Demographic details such as age, gender, location, etc.
    - **Transaction History**: Past purchases, product preferences, spending patterns, etc.
    - **Product Information**: Attributes such as category, price, and features of purchased items.
2. **Similarity Metrics**:
    - Measures the similarity between customers using mathematical methods such as:
        - **Cosine Similarity** (for high-dimensional data)
        - **Euclidean Distance** (for numerical features)
        - **Manhattan Distance**
        - **Jaccard Index** (for categorical or binary data)
        - **Pearson Correlation** (for linear relationships)
3. **Recommendation Output**:
    - For a given customer, the model outputs the **top N similar customers** with their **similarity scores**.

Attributes in the Lookalike Dataset

1. CustomerID

- **Description**: This is a unique identifier for a customer in your dataset.
- **Purpose**: Helps in distinguishing one customer from another. It serves as the primary entity for which we are finding similar customers.
- **Example**: C0001 refers to a specific customer in your database.

2. SimilarCustomerID

- **Description**: The unique identifier of another customer who is considered "similar" to the primary customer (CustomerID).
- **Purpose**: Lists the most relevant or similar customers to the one in question, based on the computed similarity from their profiles, behaviors, or transaction patterns.
- **Example**: For CustomerID C0001, C0050 and C0121 are two customers identified as similar.

3. SimilarityScore

- **Description**: A numeric representation of how closely the SimilarCustomerID matches the CustomerID based on specific attributes (e.g., demographics, transaction behavior, purchase history).
- **Purpose**: Indicates the level of similarity between the two customers. A higher score means greater similarity.
- **Example**: A SimilarityScore helps rank similar customers, where C0050 (score 0.5103) is more similar to C0001 than C0121 (score 0.4593).