Task 2: Lookalike Model for Customer Recommendations

Overview:

In Task 2, the goal is to build a **Lookalike Model** that recommends similar customers to a given user based on their profile and transaction history. The idea behind the lookalike model is to find customers who share common characteristics and behaviors with a given customer, making it easier for businesses to target similar customers for marketing campaigns or personalized offers.

We use both customer and product information to create a similarity score for each recommendation. The model is built using a combination of customer attributes (like region and signup date) and transaction data (like product purchases and total transaction value). The similarity scores will allow businesses to identify customers with similar purchasing habits and preferences, providing actionable insights for customer segmentation and targeted marketing.

---

1. Dataset Overview:

To build the lookalike model, we will use the following data:

1. **Customers.csv:**
    - **CustomerID:** Unique identifier for each customer.
    - **CustomerName:** Name of the customer.
    - **Region:** Continent or geographical location where the customer resides.
    - **SignupDate:** Date when the customer signed up.
2. **Transactions.csv:**
    - **TransactionID:** Unique identifier for each transaction.
    - **CustomerID:** Customer who made the purchase.
    - **ProductID:** Product purchased in the transaction.
    - **TransactionDate:** Date of the transaction.
    - **Quantity:** Quantity of the product purchased.
    - **TotalValue:** Total value of the transaction.
    - **Price:** Price of the product sold.
3. **Products.csv (for product information):**

- ProductID: Unique identifier for each product.
- **ProductName:** Name of the product.
- **Category:** Product category.
- **Price:** Price of the product.

The Lookalike Model will recommend similar customers based on shared characteristics and transaction behavior. The output will be a CSV file containing a list of top 3 similar customers for each of the first 20 customers in Customers.csv, along with their similarity scores.

---

2. Approach and Methodology:

The process for building the lookalike model can be broken down into the following key steps:

2.1 Data Preprocessing:

- **Customer Data:** We begin by extracting relevant customer attributes such as Region, SignupDate, and other demographic information from the Customers.csv.
- **Transaction Data:** We then process the Transactions.csv to extract transaction history for each customer, including the total number of transactions, frequency, and total spend.
- **Product Data:** Product information like ProductID, Category, and Price from the Products.csv is merged with transaction data to provide more insight into customer preferences.

We will clean the data by handling any missing or erroneous values, ensuring it is in a format suitable for building the model.

2.2 Feature Engineering:

- **Customer Profile Features:** From the Customers.csv, we can extract features like the Region of the customer and the SignupDate.
- **Transaction Features:** From the Transactions.csv, we will aggregate data to create features like:
    - **Total Spend:** The total amount spent by the customer in the given period.
    - **Product Categories Purchased:** Which categories a customer tends to purchase.
    - **Frequency of Purchases:** How often a customer makes a purchase.

These features are used to define the "profile" of a customer in the model.

2.3 Similarity Calculation:

To find similar customers, we calculate a **similarity score** based on the features mentioned above. A common method for this is using **Cosine Similarity** or **Euclidean Distance** between the feature vectors of customers. The similarity score helps us measure how similar the behaviors and characteristics of customers are, ranging from 0 (completely dissimilar) to 1 (completely similar).

- **Cosine Similarity:** Measures the cosine of the angle between two vectors, where a smaller angle (closer to 1) indicates higher similarity.
- **Euclidean Distance:** Measures the straight-line distance between two points in a multi-dimensional space (feature space).

2.4 Recommendations:

- For each customer, we calculate the similarity score with all other customers.
- We then sort the customers by similarity score and select the top 3 most similar customers for each of the first 20 customers.
- The similarity score will be included in the final recommendation list, which will be saved as a Lookalike.csv file.

2.5 Output:

The output will be a Lookalike.csv file containing the following information:

- **CustomerID (for which the recommendation is made)**
- **Top 3 Lookalike Customers**
- **Similarity Scores for each of the Top 3 Lookalike Customers**

The file will look something like this:

CustomerID	LookalikeCustomer1	Score1	LookalikeCustomer2	Score2	LookalikeCustomer3	Score3
C0001	C0015	0.85	C0032	0.82	C0009	0.79
C0002	C0021	0.88	C0043	0.80	C0031	0.78
...	...	...	...	...	...	...

---

3. Evaluation Criteria:

To evaluate the effectiveness of the lookalike model, we will assess the following:

3.1 Accuracy and Logic of the Model:

- **Model Accuracy:** The quality of the recommendations will depend on the similarity calculations. High similarity scores should reflect real-world similarities in purchasing behaviors.
- **Logic:** The model should logically identify customers with similar product preferences, purchasing habits, and demographic profiles.

3.2 Quality of Recommendations:

- The recommended lookalike customers should be meaningful and useful for businesses. For example, customers from the same region or those with similar purchasing behaviors should be grouped together.

---

4. Deliverables:

1. **Lookalike Model CSV (Lookalike.csv):**
   - This file will contain the top 3 lookalike customers for each of the first 20 customers, along with their similarity scores.
2. **Jupyter Notebook/Python Script:**
   - This notebook will include the code for data preprocessing, similarity calculation, and recommendation generation.

---

5. Conclusion:

The Lookalike Model built for customer recommendations helps businesses target customers who share similar behaviors and characteristics. The model can be applied to marketing strategies, sales campaigns, and customer segmentation efforts. By using similarity scores, the business can refine its efforts and attract customers who are more likely to convert, based on the behaviors of existing customers. This approach ensures personalized recommendations that align with customer preferences, improving engagement and conversion rates.