

```

In [1]: import pandas as pd
import numpy as np
from sklearn.metrics.pairwise import cosine_similarity

# Load datasets
customers = pd.read_csv(r"C:\Users\USER\Desktop\Datasets\Customers.csv")
products = pd.read_csv(r"C:\Users\USER\Desktop\Datasets\Products.csv")
transactions = pd.read_csv(r"C:\Users\USER\Desktop\Datasets\Transactions.csv")

# Data Preprocessing
# Merge transactions with product details (e.g., Category, Price)
transaction_data = pd.merge(transactions, products[['ProductID', 'Category', 'Price']])

# Feature engineering: Aggregate data at the customer level
customer_profile = transaction_data.groupby('CustomerID').agg(
    total_spend=('TotalValue', 'sum'),
    transaction_count=('TransactionID', 'count'),
    average_spend=('TotalValue', 'mean')
).reset_index()

# Optional: You can add other features related to products (Category, Price) if needed
# For simplicity, we'll proceed with the basic features first.

# Feature Engineering: Create a feature matrix for customer profiles
profile_features = customer_profile[['CustomerID', 'total_spend', 'transaction_count',

# Create a similarity matrix using cosine similarity based on customer profiles
similarity_matrix = cosine_similarity(profile_features.drop('CustomerID', axis=1))

# Convert the similarity matrix into a DataFrame for better readability
similarity_df = pd.DataFrame(similarity_matrix, index=profile_features['CustomerID'],

# Get the top 3 most similar customers for each of the first 20 customers (C0001 - C00
lookalikes = {}
for customer_id in profile_features['CustomerID'].iloc[:20]:
    # Sort the customers based on similarity score and get the top 3
    top_3_similar_customers = similarity_df[customer_id].sort_values(ascending=False).
    lookalikes[customer_id] = top_3_similar_customers.index.tolist()

# Create the 'Lookalike.csv' file containing the customer ID and top 3 Lookalikes with
lookalike_data = []
for customer_id, similar_customers in lookalikes.items():
    for similar_customer in similar_customers:
        similarity_score = similarity_df.loc[customer_id, similar_customer]
        lookalike_data.append([customer_id, similar_customer, similarity_score])

lookalike_df = pd.DataFrame(lookalike_data, columns=['CustomerID', 'LookalikeCustomerID

# Save the Lookalike data to 'Lookalike.csv'
lookalike_df.to_csv(r"C:\Users\USER\Desktop\Datasets\Lookalike.csv", index=False)

print("Lookalike model created successfully. 'Lookalike.csv' has been saved.")

Lookalike model created successfully. 'Lookalike.csv' has been saved.

```

In []: