

```
import pandas as pd
from sklearn.metrics.pairwise import cosine_similarity
from sklearn.preprocessing import StandardScaler

# Assuming your data is in CSV files, replace 'transactions.csv', 'customers.csv', and 'products.csv'
# with the actual file names.
transactions = pd.read_csv('/Transactions (1).csv')
customers = pd.read_csv('/Customers (1).csv')
products = pd.read_csv('/Products (1).csv')

# Merge datasets for creating user profiles
data = transactions.merge(customers, on="CustomerID").merge(products, on="ProductID")

# Aggregate features for customer profiles
customer_profiles = data.groupby("CustomerID").agg({
    "TotalValue": "sum",
    "Quantity": "sum",
    "TotalValue": "mean"
}).reset_index()

# Normalize features
scaler = StandardScaler()
features = scaler.fit_transform(customer_profiles.iloc[:, 1:])

# Compute similarity matrix
similarity_matrix = cosine_similarity(features)

# Find top 3 lookalikes for first 20 customers
lookalikes = {}
for i, customer in enumerate(customer_profiles['CustomerID'][:20]):
    similar_indices = similarity_matrix[i].argsort()[-4:-1][::-1]
    similar_customers = [(customer_profiles['CustomerID'][j], similarity_matrix[i][j]) for j in similar_indices]
    lookalikes[customer] = similar_customers

# Save lookalikes to a CSV
lookalike_df = pd.DataFrame([
    {"CustomerID": cust, "SimilarCustomers": str(similar)} for cust, similar in lookalikes.items() ])

lookalike_df.to_csv("Lookalike.csv", index=False)
print("Lookalike.csv generated.")

Lookalike.csv generated.

from google.colab import files
files.download("Lookalike.csv")
```