```
#Importing all Libraries
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

Customers_data=pd.read_csv(r"C:\Users\A.Rohith Venkatesh\Downloads\
Customers.csv")
Products_data=pd.read_csv(r"C:\Users\A.Rohith Venkatesh\Downloads\
Products.csv")
Transactions_data=pd.read_csv(r"C:\Users\A.Rohith Venkatesh\Downloads\
Transactions.csv")

import os
print("Current working directory:", os.getcwd())
```

```
Current working directory: C:\Users\A.Rohith Venkatesh
```

```
"C:/Users/YourUsername/Documents/Customers_data.csv"
```

```
'C:/Users/YourUsername/Documents/Customers_data.csv'
```

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

# Load datasets
customers = pd.read_csv(r'C:\Users\A.Rohith Venkatesh\Downloads\
Customers.csv')
products = pd.read_csv(r'C:\Users\A.Rohith Venkatesh\Downloads\
Products.csv')
transactions = pd.read_csv(r'C:\Users\A.Rohith Venkatesh\Downloads\
Transactions.csv')
# Data Preprocessing
transactions_summary = transactions.groupby('CustomerID').agg({
    'TotalValue': 'sum',
    'Quantity': 'sum',
    'ProductID': lambda x: ' '.join(x.astype(str))
}).reset_index()

# Merge customer and transaction data
customer_profiles = pd.merge(customers, transactions_summary,
on='CustomerID', how='left')

# Feature encoding and scaling
customer_profiles.fillna(0, inplace=True)
scaler = StandardScaler()
encoded_features =
pd.get_dummies(customer_profiles.drop(columns=['CustomerID',
'CustomerName']))
scaled_features = scaler.fit_transform(encoded_features)
```

```python
# Cosine similarity computation
similarity_matrix = cosine_similarity(scaled_features)

# Create Lookalike Recommendations
lookalike_dict = {}
for idx, customer_id in enumerate(customer_profiles['CustomerID']
[:20]):
    similar_indices = similarity_matrix[idx].argsort()[::-1][1:4]  #
Top 3 lookalikes
    similar_scores = similarity_matrix[idx][similar_indices]
    similar_customers =
customer_profiles['CustomerID'].iloc[similar_indices]
    lookalike_dict[customer_id] = list(zip(similar_customers,
similar_scores))

# Save results to Lookalike.csv
lookalike_df = pd.DataFrame({
    'CustomerID': lookalike_dict.keys(),
    'Lookalikes': [str(v) for v in lookalike_dict.values()]
})
lookalike_df.to_csv('Lookalike.csv', index=False)

print(customers.head())
print(products.head())
print(transactions.head())
```

```
  CustomerID        CustomerName         Region  SignupDate
0      C0001    Lawrence Carroll  South America  2022-07-10
1      C0002      Elizabeth Lutz           Asia  2022-02-13
2      C0003      Michael Rivera  South America  2024-03-07
3      C0004  Kathleen Rodriguez  South America  2022-10-09
4      C0005         Laura Weber           Asia  2022-08-15
  ProductID              ProductName     Category   Price
0      P001     ActiveWear Biography        Books  169.30
1      P002    ActiveWear Smartwatch  Electronics  346.30
2      P003  ComfortLiving Biography        Books   44.12
3      P004            BookWorld Rug   Home Decor   95.69
4      P005          TechPro T-Shirt     Clothing  429.31
  TransactionID CustomerID ProductID      TransactionDate  Quantity  \
0        T00001      C0199      P067  2024-08-25 12:38:23         1
1        T00112      C0146      P067  2024-05-27 22:23:54         1
2        T00166      C0127      P067  2024-04-25 07:38:55         1
3        T00272      C0087      P067  2024-03-26 22:55:37         2
4        T00363      C0070      P067  2024-03-21 15:10:10         3

   TotalValue   Price
0      300.68  300.68
1      300.68  300.68
2      300.68  300.68
```

```
3      601.36  300.68
4      902.04  300.68

print(transactions_summary.head())

  CustomerID  TotalValue  Quantity
ProductID
0      C0001     3354.52        12                  P054 P022 P096 P083
P029
1      C0002     1862.74        10                       P095 P004 P019
P071
2      C0003     2725.38        14                       P025 P006 P035
P002
3      C0004     5354.88        23  P049 P053 P038 P025 P097 P024 P008
P077
4      C0005     2034.24         7                            P025 P039
P012

print(encoded_features.head())
print(scaled_features[:5])

   TotalValue  Quantity  Region_Asia  Region_Europe  Region_North
America  \
0     3354.52      12.0        False          False
False
1     1862.74      10.0         True          False
False
2     2725.38      14.0        False          False
False
3     5354.88      23.0        False          False
False
4     2034.24       7.0         True          False
False

   Region_South America  SignupDate_2022-01-22  SignupDate_2022-02-02
\
0                  True                  False                  False

1                 False                  False                  False

2                  True                  False                  False

3                  True                  False                  False

4                 False                  False                  False


   SignupDate_2022-02-10  SignupDate_2022-02-13  ...  \
0                  False                  False  ...
1                  False                   True  ...
2                  False                  False  ...
```

```
3                  False                  False  ...
4                  False                  False  ...

   ProductID_P092 P021 P048  \
0                      False
1                      False
2                      False
3                      False
4                      False

   ProductID_P092 P050 P027 P078 P021 P059 P063 P046  \
0                                              False
1                                              False
2                                              False
3                                              False
4                                              False

   ProductID_P092 P076 P048  ProductID_P093 P033 P002 P091  \
0                      False                          False
1                      False                          False
2                      False                          False
3                      False                          False
4                      False                          False

   ProductID_P095 P004 P019 P071  ProductID_P095 P040 P001  \
0                         False                      False
1                          True                      False
2                         False                      False
3                         False                      False
4                         False                      False

   ProductID_P095 P062 P010 P084 P042 P033 P097  ProductID_P096  \
0                                         False           False
1                                         False           False
2                                         False           False
3                                         False           False
4                                         False           False

   ProductID_P096 P075 P045  ProductID_P100 P081 P097
0                      False                     False
1                      False                     False
2                      False                     False
3                      False                     False
4                      False                     False

[5 rows x 385 columns]
[[-0.05188436 -0.1107351  -0.53881591 ... -0.07088812 -0.07088812
  -0.07088812]
 [-0.86271433 -0.43404927  1.85592145 ... -0.07088812 -0.07088812
  -0.07088812]
```

```
[-0.393842    0.21257907 -0.53881591 ... -0.07088812 -0.07088812
 -0.07088812]
[ 1.03537505  1.66749283 -0.53881591 ... -0.07088812 -0.07088812
 -0.07088812]
[-0.76949861 -0.91902053  1.85592145 ... -0.07088812 -0.07088812
 -0.07088812]]
```