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


customers = pd.read_csv(r'C:\Users\kandu\Desktop\Downloads\
Customers.csv')
products = pd.read_csv(r'C:\Users\kandu\Desktop\Downloads\
Products.csv')
transactions = pd.read_csv(r'C:\Users\kandu\Desktop\Downloads\
Transactions.csv')


merged_data = transactions.merge(customers,
on='CustomerID').merge(products, on='ProductID')


customer_summary = merged_data.groupby('CustomerID').agg({
    'TotalValue': 'sum',
    'Quantity': 'sum',
    'Price_x': 'mean'
}).reset_index()


customer_profile = customers.merge(customer_summary, on='CustomerID',
how='left').fillna(0)


customer_profile = pd.get_dummies(customer_profile,
columns=['Region'], drop_first=True)


scaler = StandardScaler()
customer_features =
scaler.fit_transform(customer_profile.drop(columns=['CustomerID',
'CustomerName', 'SignupDate']))


similarity_matrix = cosine_similarity(customer_features)


lookalike_map = {}
customer_ids = customer_profile['CustomerID'].tolist()

for idx in range(20):  # First 20 customers
    customer_id = customer_ids[idx]
    similarities = list(enumerate(similarity_matrix[idx]))
    similarities = sorted(similarities, key=lambda x: x[1],
reverse=True)
    top_3 = [(customer_ids[i], score) for i, score in
similarities[1:4]]  # Exclude self-similarity
```

```
    lookalike_map[customer_id] = top_3


lookalike_df = pd.DataFrame({
    'CustomerID': list(lookalike_map.keys()),
    'Lookalikes': [str(lst) for lst in lookalike_map.values()]
})

lookalike_df.to_csv('Lookalike.csv', index=False)

print("Lookalike model completed. Results saved to Lookalike.csv.")
```

```
Lookalike model completed. Results saved to Lookalike.csv.
```