

```

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

customers = pd.read_csv("/content/Customers.csv")
products = pd.read_csv("/content/Products.csv")
transactions = pd.read_csv("/content/Transactions.csv")

customer_transactions = pd.merge(transactions, customers,
on="CustomerID", how="inner")
data = pd.merge(customer_transactions, products, on="ProductID",
how="inner")

customer_profiles = data.groupby('CustomerID').agg({
    'TotalValue': ['sum', 'mean'],
    'ProductID': 'count',
    'Category': lambda x: x.mode()[0]
}).reset_index()

customer_profiles.columns = ['CustomerID', 'TotalSpent',
'AvgTransactionValue', 'NumPurchases', 'TopCategory']

customer_profiles

{"summary":{"\n  \"name\": \"customer_profiles\",\n  \"rows\": 199,\n  \"fields\": [\n    {\n      \"column\": \"CustomerID\",\n      \"properties\": {\n        \"dtype\": \"string\",\n        \"num_unique_values\": 199,\n        \"samples\": [\n          \"C0083\",\n          \"C0016\",\n          \"C0112\",\n          ],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n      },\n      {\n        \"column\": \"TotalSpent\",\n        \"properties\": {\n          \"dtype\": \"number\",\n          \"std\": 1832.677958387418,\n          \"min\": 82.36,\n          \"max\": 10673.869999999999,\n          \"num_unique_values\": 199,\n          \"samples\": [\n            642.8,\n            3874.42,\n            1959.51,\n            ],\n          \"semantic_type\": \"\",\n          \"description\": \"\"\n        },\n      {\n        \"column\": \"AvgTransactionValue\",\n        \"properties\": {\n          \"dtype\": \"number\",\n          \"std\": 237.9366486597465,\n          \"min\": 82.36,\n          \"max\": 1323.1333333333332,\n          \"num_unique_values\": 199,\n          \"samples\": [\n            214.26666666666665,\n            774.884,\n            653.17,\n            ],\n          \"semantic_type\": \"\",\n          \"description\": \"\"\n        },\n      {\n        \"column\": \"NumPurchases\",\n        \"properties\": {\n          \"dtype\": \"number\",\n          \"std\": 2,\n          \"min\": 1,\n          \"max\": 11,\n          \"num_unique_values\": 11,\n          \"samples\": [\n            7,\n            5,\n            9,\n            ],\n          \"semantic_type\": \"\",\n          \"description\": \"\"\n        },\n      {\n        \"column\": \"TopCategory\",\n        \"properties\": {\n          \"dtype\": \"category\",\n          \"num_unique_values\": 4,\n          \"samples\":

```

```

[\\n          \\\"Clothing\\\",\\n          \\\"Books\\\",\\n
\\\"Electronics\\\"\\n          ],\\n          \\\"semantic_type\\\": \\\"\\\",\\n
\\\"description\\\": \\\"\\\"\\n          }\\n          ]\\
n}\\",\"type\":\"dataframe\",\"variable_name\":\"customer_profiles\"}

customer_profiles = pd.get_dummies(customer_profiles,
columns=['TopCategory'])

scaler = StandardScaler()
numeric_features = ['TotalSpent', 'AvgTransactionValue',
'NumPurchases']
customer_profiles[numeric_features] =
scaler.fit_transform(customer_profiles[numeric_features])

similarity_matrix =
cosine_similarity(customer_profiles.drop(columns=['CustomerID']))

lookalike_map = {}
for idx, customer_id in enumerate(customer_profiles['CustomerID']):
    scores = list(enumerate(similarity_matrix[idx]))
    scores = sorted(scores, key=lambda x: x[1], reverse=True)[1:4]
    lookalike_map[customer_id] = [(customer_profiles['CustomerID']
[s[0]], round(s[1], 2)) for s in scores]

filtered_map = {k: lookalike_map[k] for k in
customer_profiles['CustomerID'][:20]}

filtered_map

{'C0001': [('C0072', 0.95), ('C0190', 0.94), ('C0069', 0.91)],
'C0002': [('C0029', 1.0), ('C0010', 1.0), ('C0009', 0.97)],
'C0003': [('C0178', 1.0), ('C0052', 0.98), ('C0166', 0.96)],
'C0004': [('C0021', 1.0), ('C0101', 1.0), ('C0075', 0.99)],
'C0005': [('C0112', 1.0), ('C0197', 1.0), ('C0095', 0.98)],
'C0006': [('C0117', 1.0), ('C0168', 0.98), ('C0185', 0.96)],
'C0007': [('C0120', 0.99), ('C0140', 0.98), ('C0020', 0.94)],
'C0008': [('C0113', 0.93), ('C0124', 0.9), ('C0109', 0.86)],
'C0009': [('C0077', 1.0), ('C0083', 1.0), ('C0033', 0.98)],
'C0010': [('C0029', 1.0), ('C0002', 1.0), ('C0009', 0.98)],
'C0011': [('C0064', 0.97), ('C0137', 0.92), ('C0135', 0.87)],
'C0012': [('C0104', 0.97), ('C0059', 0.95), ('C0065', 0.94)],
'C0013': [('C0143', 1.0), ('C0099', 0.99), ('C0053', 0.97)],
'C0014': [('C0128', 1.0), ('C0151', 1.0), ('C0097', 0.98)],
'C0015': [('C0132', 0.98), ('C0036', 0.98), ('C0131', 0.98)],
'C0016': [('C0183', 1.0), ('C0107', 0.97), ('C0149', 0.94)],
'C0017': [('C0004', 0.97), ('C0090', 0.97), ('C0075', 0.97)],
'C0018': [('C0187', 1.0), ('C0171', 0.99), ('C0006', 0.94)],
'C0019': [('C0116', 0.98), ('C0047', 0.94), ('C0121', 0.89)],
'C0020': [('C0140', 0.98), ('C0120', 0.95), ('C0007', 0.94)]}
```

```
lookalike_df = pd.DataFrame([{'cust_id': k, 'lookalikes': v} for k, v
in filtered_map.items()])
lookalike_df.to_csv("Tejeshwar_Kathiravan_Lookalike.csv", index=False)

print("Lookalike.csv has been created successfully!")

Lookalike.csv has been created successfully!
```