

```

#Lookalike Model

# Import necessary libraries

from sklearn.neighbors import NearestNeighbors
from sklearn.preprocessing import StandardScaler
import pandas as pd

# Merge datasets (Customers and Transactions)

merged_data = pd.merge(transactions, customers, on='CustomerID', how='left')

# Feature Engineering: Aggregate transaction information (TotalValue and Quantity) and merge with
customer profile information

customer_profile = merged_data.groupby('CustomerID').agg({
    'TotalValue': 'sum',    # Sum of transaction values
    'Quantity': 'sum',     # Sum of product quantities bought
}).reset_index()

# Merge the aggregated transaction information back with the customer profile data

customer_profile = pd.merge(customer_profile, customers[['CustomerID', 'Region']],
on='CustomerID', how='left')

# Encode the categorical variable 'Region' (if you have more profile columns, include them here)

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

# Normalize the data for lookalike model (excluding the 'CustomerID' column)

scaler = StandardScaler()

features_for_lookalike = customer_profile.drop('CustomerID', axis=1)

customer_profile_scaled = scaler.fit_transform(features_for_lookalike)

# Build the lookalike model using K-Nearest Neighbors (KNN)

knn = NearestNeighbors(n_neighbors=4) # 3 lookalikes + the customer itself

knn.fit(customer_profile_scaled)

```

```

# Find top 3 lookalikes for the first 20 customers (CustomerID: C0001 to C0020)

lookalikes = {}

for i in range(20): # First 20 customers
    customer_id = customer_profile.iloc[i]['CustomerID']
    distances, indices = knn.kneighbors([customer_profile_scaled[i]])

    # Exclude the customer itself (first index)
    similar_customers = customer_profile.iloc[indices[0][1:], 0]
    similarity_scores = 1 - distances[0][1:] # Use 1 - distance as the similarity score

    # Store lookalikes for the customer
    lookalikes[customer_id] = list(zip(similar_customers, similarity_scores))

# Save lookalikes to a CSV file
with open('FirstName_LastName_Lookalike.csv', 'w', newline='') as f:
    writer = csv.writer(f)
    writer.writerow(['CustomerID', 'Lookalikes'])
    for customer, similar in lookalikes.items():
        writer.writerow([customer, similar])

```