B. Task 2: Lookalike Model:

- Loading Datasets:

```
from google.colab import files

# Upload CSV files
uploaded = files.upload()

# Load datasets
customers = pd.read_csv('Customers.csv')
products = pd.read_csv('Products.csv')
transactions = pd.read_csv('Transactions.csv')
```

- Verify Data Loading:

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

OUTPUT:

```
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
```

- Merge Datasets:

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

  OUTPUT:

  Index(['TransactionID', 'CustomerID', 'ProductID', 'TransactionDate',
  'Quantity', 'TotalValue', 'Price_x', 'CustomerName', 'Region',
  'SignupDate', 'ProductName', 'Category', 'Price_y'],
  dtype='object')

- Product Preferences Aggregation:

  product_preferences = merged_data.groupby(['CustomerID', 'Category']).size().unstack(fill_value=0)
  print(product_preferences.head())

  OUTPUT:

Category	Books	Clothing	Electronics	Home Decor
CustomerID				
C0001	1	0	3	1
C0002	0	2	0	2
C0003	0	1	1	2
C0004	3	0	2	3
C0005	0	0	2	1

- Normalize Features:

  print(customer_summary.isnull().sum())

  OUTPUT:

  TotalValue 0
  TransactionID 0
  Books 0
  Clothing 0
  Electronics 0
  Home Decor 0
  dtype: int64

  *Handle missing values if any:

  customer_summary.fillna(0, inplace=True)

- Debug Recommendations:

  print(similarity_matrix[:5, :5]) # Check a small part of the matrix

  print(top_similar_customers) # Inspect recommendations

OUTPUT:

[[1. -0.54352611 0.00852079 0.20724726 0.63939904]

[-0.54352611 1. 0.79054587 -0.5776848 0.19310442]

[0.00852079 0.79054587 1. -0.38843671 0.62647879]

[0.20724726 -0.5776848 -0.38843671 1. -0.42275638]

[0.63939904 0.19310442 0.62647879 -0.42275638 1.]]

{'C0001': [('C0069', 0.9474257972151854), ('C0127', 0.8739694001028301), ('C0190', 0.8460722354249515)], 'C0002': [('C0133', 0.9681437939265284), ('C0062', 0.8997910818956721), ('C0134', 0.8968440792176382)], 'C0003': [('C0166', 0.9944603992451134), ('C0031', 0.9746433592752327), ('C0158', 0.9376095644412319)], 'C0004': [('C0090', 0.9178464884605366), ('C0122', 0.9118789079040195), ('C0017', 0.9094137328867413)], 'C0005': [('C0197', 0.9996873537675999), ('C0007', 0.9906572143006201), ('C0140', 0.8991908631524648)], 'C0006': [('C0135', 0.9131662546767203), ('C0187', 0.7746920876470779), ('C0185', 0.7290633873719096)], 'C0007': [('C0005', 0.9906572143006201), ('C0197', 0.9869375698842499), ('C0120', 0.8957528613441336)], 'C0008': [('C0162', 0.9354260196136521), ('C0154', 0.8903984349746906), ('C0113', 0.8845529363450987)], 'C0009': [('C0198', 0.920168728981332), ('C0029', 0.9196461287569269), ('C0033', 0.8764337474094815)], 'C0010': [('C0061', 0.9181948166116802), ('C0176', 0.9141140947390333), ('C0042', 0.9120596258448855)], 'C0011': [('C0126', 0.9539127794003), ('C0171', 0.9061044912926547), ('C0193', 0.8859277991023892)], 'C0012': [('C0065', 0.9732107193964651), ('C0136', 0.931281573213204), ('C0104', 0.9274639598748136)], 'C0013': [('C0067', 0.9489699202294121), ('C0105', 0.9401951306112102), ('C0102', 0.8468491088707529)], 'C0014': [('C0151', 0.9999087464608989), ('C0097', 0.9996715010128533), ('C0060', 0.9994465585577332)], 'C0015': [('C0123', 0.9988694847917271), ('C0014', 0.9573581146035457), ('C0151', 0.9572574682152036)], 'C0016': [('C0183', 0.9999646152266838), ('C0107', 0.9961432432416787), ('C0105', 0.8878019025387205)], 'C0017': [('C0075', 0.9547625456376855), ('C0090', 0.9410347298753767), ('C0194', 0.9295087951122939)], 'C0018': [('C0023', 0.8796611833380044), ('C0168', 0.8689860208984774), ('C0068', 0.7789484717048029)], 'C0019': [('C0191', 0.9549716093424133), ('C0174', 0.8707863705075659), ('C0070', 0.8459063597422131)], 'C0020': [('C0130', 0.9987465475702518), ('C0095', 0.9518201439712922), ('C0198', 0.9470386937374203)]}

Inspecting the output:

1. **Similarity Matrix:**

   print(similarity_matrix[:5, :5]) # View a sample of the similarity matrix

   OUTPUT:

   [[1. -0.54352611 0.00852079 0.20724726 0.63939904]
   [-0.54352611 1. 0.79054587 -0.5776848 0.19310442]
   [0.00852079 0.79054587 1. -0.38843671 0.62647879]
   [0.20724726 -0.5776848 -0.38843671 1. -0.42275638]
   [0.63939904 0.19310442 0.62647879 -0.42275638 1.]]

2. **Recommendations:** Ensure that the top 3 similar customers are correctly identified for each of the first 20 customers

```
for customer, recommendations in top_similar_customers.items():
    print(f"Customer {customer}: {recommendations}")
```

OUTPUT:

```
Customer C0001: [('C0069', 0.9474257972151854), ('C0127', 0.8739694001028301), ('C0190', 0.8460722354249515)]
Customer C0002: [('C0133', 0.9681437939265284), ('C0062', 0.8997910818956721), ('C0134', 0.8968440792176382)]
Customer C0003: [('C0166', 0.9944603992451134), ('C0031', 0.9746433592752327), ('C0158', 0.9376095644412319)]
Customer C0004: [('C0090', 0.9178464884605366), ('C0122', 0.9118789079040195), ('C0017', 0.9094137328867413)]
Customer C0005: [('C0197', 0.9996873537675999), ('C0007', 0.9906572143006201), ('C0140', 0.8991908631524648)]
Customer C0006: [('C0135', 0.9131662546767203), ('C0187', 0.7746920876470779), ('C0185', 0.7290633873719096)]
Customer C0007: [('C0005', 0.9906572143006201), ('C0197', 0.9869375698842499), ('C0120', 0.8957528613441336)]
Customer C0008: [('C0162', 0.9354260196136521), ('C0154', 0.8903984349746906), ('C0113', 0.8845529363450987)]
Customer C0009: [('C0198', 0.920168728981332), ('C0029', 0.9196461287569269), ('C0033', 0.8764337474094815)]
Customer C0010: [('C0061', 0.9181948166116802), ('C0176', 0.9141140947390333), ('C0042', 0.9120596258448855)]
Customer C0011: [('C0126', 0.9539127794003), ('C0171', 0.9061044912926547), ('C0193', 0.8859277991023892)]
Customer C0012: [('C0065', 0.9732107193964651), ('C0136', 0.931281573213204), ('C0104', 0.9274639598748136)]
Customer C0013: [('C0067', 0.9489699202294121), ('C0105', 0.9401951306112102), ('C0102', 0.8468491088707529)]
Customer C0014: [('C0151', 0.9999087464608989), ('C0097', 0.9996715010128533), ('C0060', 0.9994465585577332)]
Customer C0015: [('C0123', 0.9988694847917271), ('C0014', 0.9573581146035457), ('C0151', 0.9572574682152036)]
Customer C0016: [('C0183', 0.9999646152266838), ('C0107', 0.9961432432416787), ('C0105', 0.8878019025387205)]
Customer C0017: [('C0075', 0.9547625456376855), ('C0090', 0.9410347298753767), ('C0194', 0.9295087951122939)]
Customer C0018: [('C0023', 0.8796611833380044), ('C0168', 0.8689860208984774), ('C0068', 0.7789484717048029)]
Customer C0019: [('C0191', 0.9549716093424133), ('C0174', 0.8707863705075659), ('C0070', 0.8459063597422131)]
Customer C0020: [('C0130', 0.9987465475702518), ('C0095', 0.9518201439712922), ('C0198', 0.9470386937374203)]
```

3. **Output File:** Check that the Lookalike.csv file is generated

   !cat Lookalike.csv # View the contents of the CSV

   OUTPUT:

   CustomerID, SimilarCustomers
   C0001,"[('C0069', 0.9474257972151854), ('C0127', 0.8739694001028301), ('C0190', 0.8460722354249515)]"
   C0002,"[('C0133', 0.9681437939265284), ('C0062', 0.8997910818956721), ('C0134', 0.8968440792176382)]"
   C0003,"[('C0166', 0.9944603992451134), ('C0031', 0.9746433592752327), ('C0158', 0.9376095644412319)]"
   C0004,"[('C0090', 0.9178464884605366), ('C0122', 0.9118789079040195), ('C0017', 0.9094137328867413)]"
   C0005,"[('C0197', 0.9996873537675999), ('C0007', 0.9906572143006201), ('C0140', 0.8991908631524648)]"
   C0006,"[('C0135', 0.9131662546767203), ('C0187', 0.7746920876470779), ('C0185', 0.7290633873719096)]"
   C0007,"[('C0005', 0.9906572143006201), ('C0197', 0.9869375698842499), ('C0120', 0.8957528613441336)]"
   C0008,"[('C0162', 0.9354260196136521), ('C0154', 0.8903984349746906), ('C0113', 0.8845529363450987)]"
   C0009,"[('C0198', 0.920168728981332), ('C0029', 0.9196461287569269), ('C0033', 0.8764337474094815)]"
   C0010,"[('C0061', 0.9181948166116802), ('C0176', 0.9141140947390333), ('C0042', 0.9120596258448855)]"
   C0011,"[('C0126', 0.9539127794003), ('C0171', 0.9061044912926547), ('C0193', 0.8859277991023892)]"
   C0012,"[('C0065', 0.9732107193964651), ('C0136', 0.931281573213204), ('C0104', 0.9274639598748136)]"
   C0013,"[('C0067', 0.9489699202294121), ('C0105', 0.9401951306112102), ('C0102', 0.8468491088707529)]"
   C0014,"[('C0151', 0.9999087464608989), ('C0097', 0.9996715010128533), ('C0060', 0.9994465585577332)]"
   C0015,"[('C0123', 0.9988694847917271), ('C0014', 0.9573581146035457), ('C0151', 0.9572574682152036)]"
   C0016,"[('C0183', 0.9999646152266838), ('C0107', 0.9961432432416787), ('C0105', 0.8878019025387205)]"
   C0017,"[('C0075', 0.9547625456376855), ('C0090', 0.9410347298753767), ('C0194', 0.9295087951122939)]"
   C0018,"[('C0023', 0.8796611833380044), ('C0168', 0.8689860208984774), ('C0068', 0.7789484717048029)]"
   C0019,"[('C0191', 0.9549716093424133), ('C0174', 0.8707863705075659), ('C0070', 0.8459063597422131)]"
   C0020,"[('C0130', 0.9987465475702518), ('C0095', 0.9518201439712922), ('C0198', 0.9470386937374203)]"