p178dsjz3

January 27, 2025

1 Task 2: Lookalike Model

Step 1: Importing Libraries

```
[1]: import pandas as pd
     from sklearn.metrics.pairwise import cosine_similarity
     from sklearn.preprocessing import MinMaxScaler
```

Step 2: Loading Datasets

```
[4]: customers = pd.read_csv(r"C:\Users\JOGESH\Downloads\Customers.csv")
     transactions = pd.read_csv(r"C:\Users\JOGESH\Downloads\Transactions.csv")
     data = pd.merge(transactions, customers, on='CustomerID', how='inner')
```

Step 3: Feature Engineering

```
[10]: customer_features = data.groupby('CustomerID').agg({
          'TotalValue': 'sum',
          'Quantity': 'sum',
          'ProductID': lambda x: x.nunique(),
          'Region': 'first'
      }).reset_index()
      from sklearn.preprocessing import MinMaxScaler
      scaler = MinMaxScaler()
      customer_features[['TotalValue', 'Quantity', 'ProductID']] = scaler.
       ↪fit_transform(
          customer_features[['TotalValue', 'Quantity', 'ProductID']]
      )
```

Step 4: Building Similarity Matrix

```
[6]: similarity_matrix = cosine_similarity(customer_profiles.iloc[:, 1:])
     similarity_df = pd.DataFrame(similarity_matrix,␣
      ↪index=customer_profiles['CustomerID'],␣
      ↪columns=customer_profiles['CustomerID'])
```

Step 5: Generating Top 3 Lookalikes

```
[20]: lookalike_results = {}
      for customer_id in customer_features['CustomerID']:
          similar_customers = similarity_df[customer_id].
       ↪sort_values(ascending=False)[1:4]
          lookalike_results[customer_id] = [
              (idx, round(score, 2)) for idx, score in similar_customers.items()
          ]
      lookalike_data = []
      for customer_id, lookalikes in lookalike_results.items():
          row = [customer_id]
          for lookalike in lookalikes:
              row.extend(lookalike)
          while len(row) < 7:
              row.extend([None, None])
          lookalike_data.append(row)
      lookalike_df = pd.DataFrame(
          lookalike_data,
          columns=['CustomerID', 'Lookalike1', 'Score1', 'Lookalike2', 'Score2',␣
       ↪'Lookalike3', 'Score3']
      )
      lookalike_df.to_csv('Keerthana_Kumbham_Lookalike.csv', index=False)
      lookalike_df.head()
```

```
[20]:   CustomerID Lookalike1  Score1 Lookalike2  Score2 Lookalike3  Score3
      0      C0001      C0107     1.0      C0137     1.0      C0174    1.00
      1      C0002      C0084     1.0      C0106     1.0      C0175    1.00
      2      C0003      C0133     1.0      C0190     1.0      C0174    0.99
      3      C0004      C0113     1.0      C0104     1.0      C0102    1.00
      4      C0005      C0145     1.0      C0022     1.0      C0101    1.00
```