Online Product Recommendation System

```
import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

from sklearn.metrics.pairwise import cosine_similarity


dt = pd.read_excel('OnlineRetail (1) (1).xlsx')

dt.shape

dt = dt.loc[dt['Quantity'] > 0]

dt['CustomerID'].isna().sum()

dt.loc[dt['CustomerID'].isna()].head()

dt = dt.dropna(subset = ['CustomerID'])

dt.shape

customer_product_matrix = dt.pivot_table(index='CustomerID', columns='StockCode',
values='Quantity', aggfunc='sum')

customer_product_matrix

customer_product_matrix = customer_product_matrix.applymap(lambda x: 1 if x>0 else 0)

customer_product_matrix

user_user_similarity_matrix = pd.DataFrame(cosine_similarity(customer_product_matrix))

user_user_similarity_matrix

user_user_similarity_matrix.columns = customer_product_matrix.index

user_user_similarity_matrix

user_user_similarity_matrix['CustomerID'] = customer_product_matrix.index

user_user_similarity_matrix = user_user_similarity_matrix.set_index('CustomerID')

user_user_similarity_matrix

user_user_similarity_matrix.loc[12350.0].sort_values(ascending=False)

product_bought_by_A = set(customer_product_matrix.loc[12350.0].iloc[

   customer_product_matrix.loc[12350.0].to_numpy().nonzero()

].index)

product_bought_by_A

product_bought_by_B = set(customer_product_matrix.loc[17935.0].iloc[
```

```
    customer_product_matrix.loc[17935.0].to_numpy().nonzero()
].index)
product_bought_by_B
recommend_to_B = product_bought_by_A - product_bought_by_B
recommend_to_B
dt.loc[
    dt['StockCode'].isin(recommend_to_B),
    ['StockCode', 'Description']
].drop_duplicates().set_index('StockCode')
item_item_sim_matrix = pd.DataFrame(
    cosine_similarity(customer_product_matrix.T)
)
item_item_sim_matrix
item_item_sim_matrix.columns = customer_product_matrix.T.index

item_item_sim_matrix['StockCode'] = customer_product_matrix.T.index
item_item_sim_matrix = item_item_sim_matrix.set_index('StockCode')

item_item_sim_matrix
top_10_similar_items = list(
  item_item_sim_matrix\
      .loc[23166]\
      .sort_values(ascending=False)\
      .iloc[:10]\
    .index
)
top_10_similar_items
dt.loc[
    dt['StockCode'].isin(top_10_similar_items),
    ['StockCode', 'Description']
].drop_duplicates().set_index('StockCode').loc[top_10_similar_items]
```