

USER-BASED PREDICTING

**hệ thống gợi ý dựa trên
ng tương tự**



ASED FILTERING

- Là một thuật toán trong hệ thống gợi ý, sử dụng sự tương đồng giữa người dùng. Giả định chính là: **Những người dùng giống nhau thường thích nhau**
- Gợi ý sản phẩm phù hợp cho từng khách hàng
- Tăng doanh thu và giữ chân khách hàng
- Hiểu rõ mối quan hệ và sở thích giữa các nhóm khách hàng
- Dữ liệu sạch, đã được loại bỏ các giá trị NaN
- Các cột cần thiết: CustomerID, StockCode, Quantity
- Các khách hàng trong bộ dữ liệu

Ví dụ minh họa về hệ thống gợi ý User-based Filtering

cáo. Cô B thích dưa hấu và dứa.

dưa hấu và nho.

Cô C cùng thích dưa hấu nên họ
và những người dùng tương tự

chưa từng biết
hệ thống gợi ý sẽ đề xuất dứa

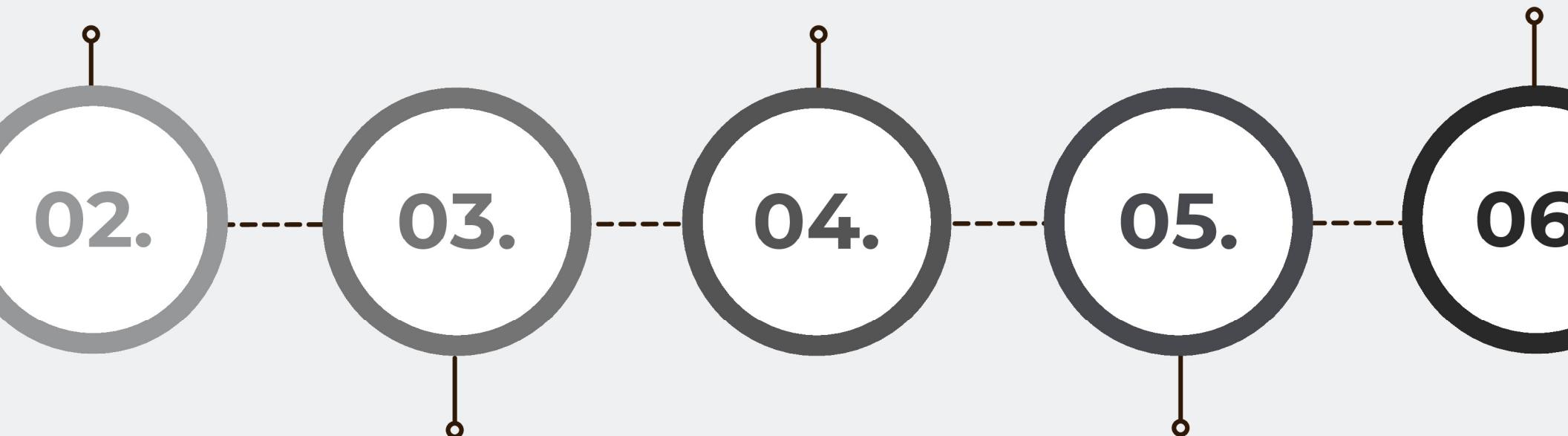


ASED FILTERING

o ma trận
ser x Item

Xác định những người
dùng tương tự

Đề xuất sản phẩm
người dùng



Tính ma trận độ tương
đồng giữa những

Thu hẹp tập hợp
sản phẩm

Tất cả được làm sạch & Tạo ma trận Customer x Item

Độ cẩn thiết: "CustomerID", "StockCode", "Quantity"

`pivot_table()` để chuyển dataset thành dạng ma trận Custom

Điều thành dạng ma trận: Customer x Item

```
rix = df_yc2.pivot_table(index='CustomerID', columns='StockCode', values='Quantity',  
rix.head()
```

Độ tương quan giữa những người dùng

Chúng ta sẽ sử dụng ma trận Pearson để tính ma trận độ tương quan giữa những khách hàng.

```
tương quan giữa những khách hàng  
similarity = user_item_matrix.T.corr()  
similarity.head()
```

12347.0	12348.0	12353.0	12354.0	12355.0	12356.0	12358.0	12359.0	12360.0	12361.0
1.000000	0.455205	-0.001861	0.000150	0.754132	0.036851	-0.003116	0.034077	0.054596	0.062000
0.455205	1.000000	-0.001089	-0.003294	0.683745	-0.001937	-0.001403	-0.003417	-0.004001	-0.004596
-0.001861	-0.001089	1.000000	-0.003255	0.005016	-0.001914	0.009373	-0.003377	-0.004200	-0.004800
0.000150	-0.003294	-0.003255	1.000000	-0.003640	0.011057	-0.005451	0.008252	0.133080	0.160000
0.754132	0.683745	0.005016	-0.003640	1.000000	0.002045	-0.002015	0.003350	0.054447	-0.054447

ch hàng tương tự

hàng bất kỳ

ch hàng đấy khỏi danh sách khách hàng tương tự

ng tương tự có độ tương quan lớn hơn 0.3

tương tự

0.3

0.3

ong quan cao nhất

```
similarity[user_similarity[customer_id]>user_similarity_threshold][customer_id].sort_values()
```

tương tự với khách hàng 18282.0 là

782

447

245

256

Sản phẩm

đã mua

```
picked_userid_bought.columns, axis=1, inplace=True, errors='ignore')
```

24 20728 20971

- **Xác định sản phẩm khách hàng mục tiêu**

Loại bỏ những sản phẩm khách hàng chưa mua

- **Xác định sản phẩm khách hàng tương tự**

Loại bỏ những sản phẩm mà khách hàng đã mua

- **Gộp dữ liệu**

Bỏ dataframe những sản phẩm mà khách hàng đã mua khỏi dataframe những sản phẩm khách hàng chưa mua

Tạo cho khách hàng mục tiêu

ctionary
_score

ng lặp tính
ểm dự đoán

ểm số dự
trung bình

```
# Tạo 1 dictionary lưu trữ điểm số dự đoán của từng sản phẩm
item_score = {}

# Tạo vòng lặp sản phẩm
for i in similar_user_items.columns:
    # Số lượng mua của tất cả khách hàng tương tự
    item_quantity = similar_user_items[i]
    # Tạo 1 biến để tổng hợp số lượng mua
    total = 0
    # Tạo 1 biến để đếm số lượng khách hàng đã mua
    count = 0
    # Tạo vòng lặp khách hàng tương tự
    for u in similar_users.index:
        # Nếu đã mua sản phẩm
        if item_quantity[u] != 0:
            # Tổng điểm tương đồng của khách hàng này
            score = similar_users[u] * item_quantity[u]
            # Cộng điểm số vừa tính được vào tổng điểm
            total += score
        # Tăng biến count lên 1
```

àò cho khách hàng mục tiêu

onary sang DataFrame và sắp xếp

re sang DataFrame cho dễ thao
p xếp theo thứ tự giảm dần

sách 10 sản phẩm đầu tiên

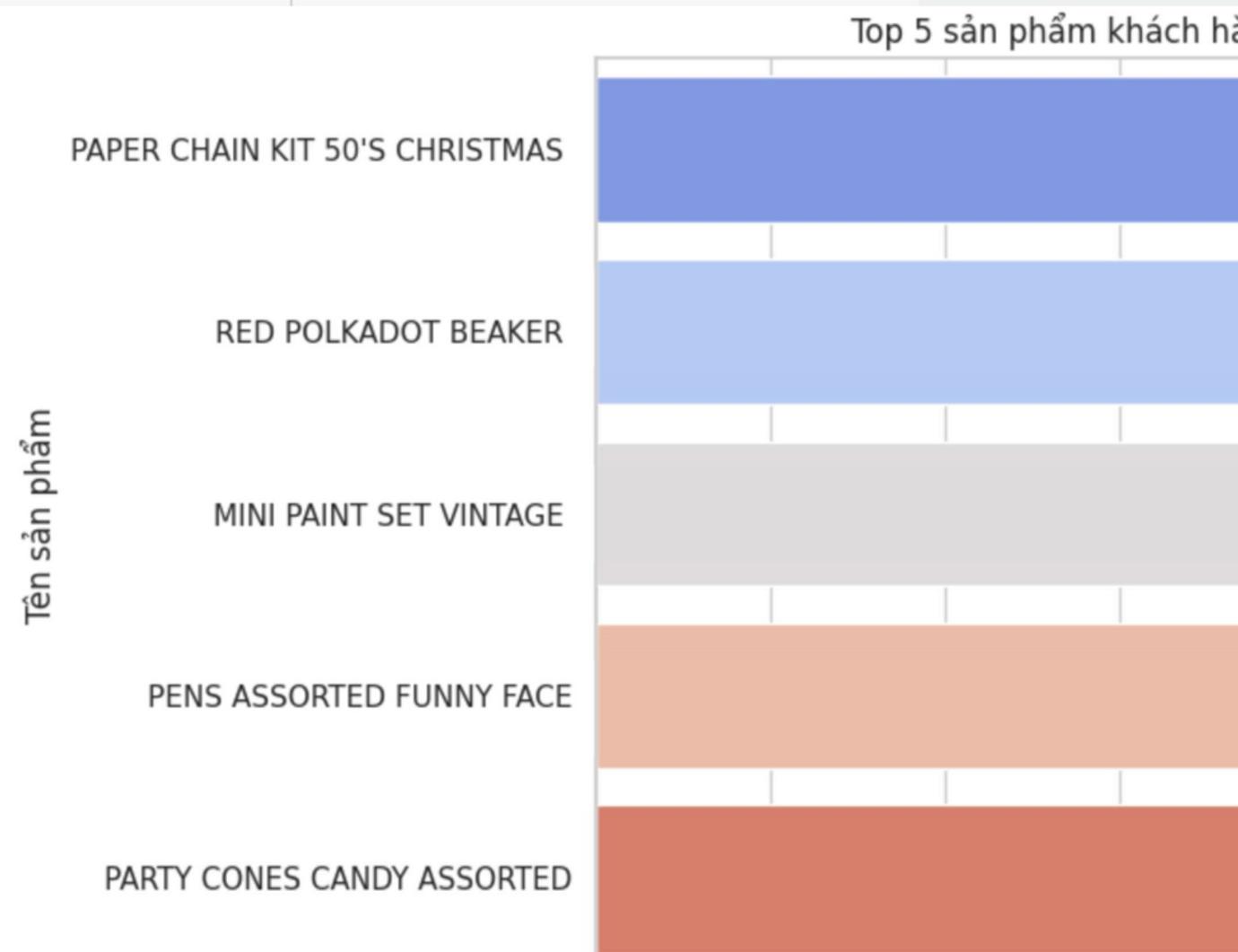
m_score, đưa ra danh sách tên
nhà sản phẩm (StockCode) của 10
nhà sản phẩm có điểm số cao nhất

Desc
PAPER CHAIN KIT 50'S CHR
RED POLKADOT
MINI PAINT SET V
PENS ASSORTED FUN
PARTY CONES CANDY A
PACK OF 6 SWEETIE GIF
RIDGED GLASS FING
PAPER CHAIN KIT
PACK OF 6 PANNETONE GIF
GINGHAM RO

5 sản phẩm cho khách hàng

```
score.head(5)
```

```
['item_score'].values, y=top_5_ranked_items['Description'], palette='coolwarm')  
{customer_id} nên mua')
```



USER-BASED FILTERING		
		ITEM-BASED FILTERING
<ul style="list-style-type: none">Gợi ý dựa trên sự tương đồng giữa người dùngTìm người dùng tương đương và gợi ý các mục họ thíchMa trận người dùng – sản phẩmCustomerID x CustomerIDGiữa các người dùng (user-user similarity)		<ul style="list-style-type: none">Gợi ý dựa trên sự tương đồng mặt hàngTìm các mục giống với sản phẩm đã dùng mục tiêu đã thíMa trận người dùng – sản phẩmStockCode x StockCodeGiữa các sản phẩm (item-item similarity)Loại sản phẩm mà nó