

```
#Import Library
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
from scipy.spatial.distance import cosine
import networkx as nx

#Thai Font
!wget -q https://github.com/Phonbopit/sarabun-webfont/raw/master/fonts/thSarabunNew

# !pip install -U --pre matplotlib
import matplotlib as mpl
mpl.font_manager.fontManager.addfont('thsarabunnew-webfont.ttf')
mpl.rc('font', family='TH Sarabun New')
```

## ▼ Preprocessing

```
df = pd.read_csv('customer_preference.csv')
```

```
df.head(5)
```

	ประทับ เวลา	จักรยาน เสือ ภูเขา	คอล ลา เจน	Brownie อวกาศ	แม่ โขง	Botox	กัญชา	กัญ ชง	เครื่อง วัด ความ ดัน	หลอด ไฟ เปลี่ยน สีได้	คอร์ส bikini wax
0	3/20/2021 11:29:59	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1	3/20/2021 11:33:10	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2	3/20/2021 11:53:45	ไม่เคย	ไม่ เคย	ไม่เคย	ไม่ เคย	ไม่เคย	ไม่ เคย	ไม่ เคย	ไม่เคย	ไม่เคย	ไม่เคย
3	3/20/2021 11:54:01	ไม่เคย	เคย	ไม่เคย	เคย	ไม่เคย	ไม่ เคย	ไม่ เคย	เคย	ไม่เคย	ไม่เคย
4	3/20/2021 11:54:05	เคย	ไม่ เคย	ไม่เคย	ไม่ เคย	ไม่เคย	ไม่ เคย	ไม่ เคย	เคย	เคย	ไม่เคย

```
print(df.shape)
print(df.columns)
print(df.info())
```

```
5    Botox                                45 non-null    object
6    กัญชา                                45 non-null    object
7    กัญชง                                45 non-null    object
8    เครื่องวัดความดัน                    45 non-null    object
```

9	หลอดไฟเปลี่ยนสีได้	45	non-null	object
10	คอร์ส bikini wax	45	non-null	object
11	เก้าอี้นวดไฟฟ้า	45	non-null	object
12	ต้นไม้ประดับห้อง	45	non-null	object
13	หมอนข้าง	45	non-null	object
14	ตาข่าย	45	non-null	object
15	คอร์สโบว์ท็อกกราม	45	non-null	object
16	แสงโสม	45	non-null	object
17	กะทรี(ปั๊)	45	non-null	object
18	Lego	45	non-null	object
19	แฟน ที่แปลว่าพัดลม	45	non-null	object
20	ยาดอง	45	non-null	object
21	ชุดนอนไม่ได้นอน	45	non-null	object
22	ทียกน้ำหนัก (Dumbell)	45	non-null	object
23	เครื่องดูดฝุ่นอัตโนมัติ (robot)	45	non-null	object
24	เนื้อวากิล	45	non-null	object
25	หมูแดดเดียว	44	non-null	object
26	drone	45	non-null	object
27	เครื่องย่อยเศษอาหาร	45	non-null	object
28	เสื้อชีทรู	46	non-null	object
29	ไฟยิปซี	45	non-null	object
30	เนื้อย่าง	45	non-null	object
31	เครื่องทำลายเอกสาร	45	non-null	object
32	กระเป๋	45	non-null	object
33	ชาบู	45	non-null	object
34	Surfskate	45	non-null	object
35	Board game	45	non-null	object
36	เครื่องย่อยเศษอาหาร.1	44	non-null	object
37	ชุดน้องหมา	45	non-null	object
38	Omakase	45	non-null	object
39	เม็ดมะม่วงหิมพานต์	45	non-null	object
40	ไฟอัจฉริยะ	45	non-null	object
41	แบตเตอรี่ที่ตัดผม	45	non-null	object
42	ไดร์เป่าผม Dyson	45	non-null	object
43	แผ่นเสียง	45	non-null	object
44	กรรไกรตัดขนจมูก	45	non-null	object
45	PS5	45	non-null	object
46	เครื่องตรวจน้ำตาลในเลือด	44	non-null	object
47	โรตี่	45	non-null	object
48	กะปิตอง	45	non-null	object
49	น้องหมา	45	non-null	object
50	แพ็คเกจไทยเที่ยวไทย	44	non-null	object
51	เกมส์ FIFA Online	45	non-null	object
52	มะม่วงเบา	45	non-null	object
53	ห้องน้ำแม่อัตโนมัติ	45	non-null	object
54	กระเป๋ chanel	44	non-null	object
55	ข้าวขาหมูคากิคาจิกอนไลน์	45	non-null	object
56	ถังเช่า	44	non-null	object
57	หมอนสุขภาพสำหรับคนนอนไม่หลับ	45	non-null	object
58	อาหารแมวคลีนเพื่อสุขภาพลดหุ่น	45	non-null	object
59	น้ำส้มดีโต้	45	non-null	object
60	Software ถูกลิขสิทธิ์	45	non-null	object
61	ครีมกันแดด	45	non-null	object

dtypes: object(62)

memory usage: 22.9+ KB

None

df.isnull().head(5)

	ประทับ เวลา	จักรยาน เสือ ภูเขา	คอล ลา เจน	Brownie อวกาศ	แม่ โขง	Botox	กัญชา	กัญชง	เครื่อง วัด ความ ดัน	หลอด ไฟ เปลี่ยน สีได้	คอร์ส bikini wax
2	False	False	False	False	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False	False	False	False
5	False	False	False	False	False	False	False	False	False	False	False
8	False	False	False	False	False	False	False	False	False	False	False

```
#Remove Null
df1 = df.dropna().reset_index(drop=True)
df1 = df1.drop(['ประทับเวลา'], axis=1)
df1 = df1.drop(['เครื่องย่อยเศษอาหาร.1'], axis=1)
```

```
df1.head(3)
```

	จักรยาน เสือ ภูเขา	คอล ลา เจน	Brownie อวกาศ	แม่ โขง	Botox	กัญชา	กัญ ชง	เครื่อง วัด ความ ดัน	หลอด ไฟ เปลี่ยน สีได้	คอร์ส bikini wax	เก้าอี้ นวด ไฟฟ้า	ต้นไม้ ประดับ ห้อง
0	ไม่เคย	ไม่ เคย	ไม่เคย	ไม่ เคย	ไม่เคย	ไม่ เคย	ไม่ เคย	ไม่เคย	ไม่เคย	ไม่เคย	ไม่ เคย	ไม่เคย
1	ไม่เคย	เคย	ไม่เคย	เคย	ไม่เคย	ไม่ เคย	ไม่ เคย	เคย	ไม่เคย	ไม่เคย	ไม่ เคย	ไม่เคย
2	เคย	ไม่ เคย	ไม่เคย	ไม่ เคย	ไม่เคย	ไม่ เคย	ไม่ เคย	เคย	เคย	ไม่เคย	ไม่ เคย	เคย

```
#Replace String to Value
for i in df1.columns:
    df1[i] = df1[i].replace(['ไม่เคย'],0)
    df1[i] = df1[i].replace(['เคย'],1)
```

```
df1.head(3)
```

จ้กรยาน เสื้อ คอล ลา Brownie แม่ Botox กัณชา กัญ เครื่อง วัต หลอด ไฟ คอรัส bikini แก้ว นวด ตันไม้ ประดับ

## ▼ EDA

```
df1.describe()
```

น้องหมา	แพ็คเกจ ไทยเที่ยว ไทย	เกมส์ FIFA Online	มะม่วงเบา	ห้องน้ำแมว อัตโนมัติ	กระเป๋า chanel	ข้าวขาหมู คาเฟ่จก ออนไลน์	ถังเช่า
40.000000	40.000000	40.000000	40.000000	40.000000	40.000000	40.000000	40.000000
0.200000	0.475000	0.125000	0.425000	0.075000	0.075000	0.100000	0.175000
0.405096	0.505736	0.334932	0.500641	0.266747	0.266747	0.303822	0.384808
0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
0.000000	1.000000	0.000000	1.000000	0.000000	0.000000	0.000000	0.000000
1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000

```
#Check distinct Value
#อะไรบ้างที่ทุกคนก็เคยซื้อ
list_all_purc = list()
print('อะไรบ้างที่ทุกคนก็เคยซื้อ?')
for i in df1.columns:
    if len(df1[i].unique()) == 1 :
        list_all_purc.append(i)
    print(i)
```

```
อะไรบ้างที่ทุกคนก็เคยซื้อ?
กระเป๋า
ชาบู
ครีมกันแดด
```

```
df2 = df1.sum().to_frame().transpose()
```

```
df2
```

```
df2 = df1.sum().sort_values(ascending=True).to_frame().transpose()
```

	จักรยานเสือภูเขา	คอลลาเจน	Brownie อวกาศ	แม่โขง	Botox	กัญชา	กัญชง	เครื่องวัดความ	หลอดไฟเปลี่ยน	คอร์ส bikini wax	เก้าอี้ นวดไฟฟ้า	ต้นไม้ประดับ	
	0	14	22	4	12	7	6	3	15	10	3	5	27

df2

	เครื่อง ย่อย เศษ อาหาร	PS5	อาหาร แมวคลี นเพื่อ สุขภาพ ลดหุ่น	กระเป๋า chanel	ไดร์ เป่าผม Dyson	ห้องน้ำ แมว อัตโนมัติ	คอร์ส bikini wax	Surfskate	กัญ ชง	ข้าวขา หมูคากิ คาจิก ออนไลน์	คอร์ส โบว์ อก การม
0	2	2	3	3	3	3	3	3	3	4	4

```
df1.shape[0]

40

# Bar Chart
y = df2.columns
x = df2.sum()

plt.barh(y,x)
plt.title('Preferences')
plt.ylabel('Product')
plt.xlabel('Quantity')
plt.rcParams["figure.figsize"] = (25,25)
plt.show()
```



```
list_product.reverse()
```

$$n = 10$$

```
print('สินค้าที่ใครๆ ก็เคยซื้อ',n,'อันดับ : \n',top_product)
```

['ครีมกันแดด', 'กระเป๋', 'ชาบู', 'โรตี่', 'กะหล่ำ(ปล้ำ)', 'เม็ดมะม่วงหิมพานต์', 'หมอนข้าง',

```
df11 = df1.drop(top_product, axis=1)
```

[https://colab.research.google.com/drive/1foghvecA2pDEsgpOyDVpwANtqUXwmY\\_t#scrollTo=CB\\_xHdGBaKPV&printMode=true](https://colab.research.google.com/drive/1foghvecA2pDEsgpOyDVpwANtqUXwmY_t#scrollTo=CB_xHdGBaKPV&printMode=true)

	จักรยาน เสือ ภูเขา	คอล ลา เจน	Brownie อวกาศ	แม่ โขง	Botox	กัญชา	กัญ ชง	เครื่อง วัด ความ ดัน	หลอด ไฟ เปลี่ยน สีได้	คอร์ส bikini wax	เก้าอี้ นวด ไฟฟ้า	ต้นไม้ ประดับ ห้อง
0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	1	0	1	0	0	0	1	0	0	0	0
2	1	0	0	0	0	0	0	1	1	0	0	1
3	0	0	0	1	0	1	0	0	0	0	1	1

## ▼ Implement Collaborative Filtering

```
#Item to Item Matrix
item_item_matrix = pd.DataFrame(index=df11.columns,columns=df11.columns)
```

```
item_item_matrix.head(5)
```

	จักรยาน เสือ ภูเขา	คอล ลา เจน	Brownie อวกาศ	แม่ โขง	Botox	กัญชา	กัญ ชง	เครื่อง วัด ความ ดัน	หลอด ไฟ เปลี่ยน สีได้	คอร์ส bikini wax	เก้า อี้ นวด ไฟฟ้า
จักรยาน เสือภูเขา	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
คอลลา เจน	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Brownie อวกาศ	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
แม่โขง	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Botox	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

```
#Calculate cosine similarity - Consine between Item to Item Profile
for i in range(0,len(item_item_matrix.columns)):
    for j in range(0,len(item_item_matrix.columns)):
        item_item_matrix.iloc[i,j] = 1 - cosine(df11.iloc[:,i],df11.iloc[:,j])
```

```
item_item_matrix.head(5)
```

	จักรยาน เสือภูเขา	คอลลา เจน	Brownie อวกาศ	แม่โขง	Botox	กัญชา	กัญชง	เครื่องวัด ความดัน
จักรยาน เสือภูเขา	1	0.569803	0.267261	0.385758	0.404061	0.218218	0.154303	0.483046
คอลลา เจน	0.569803	1	0.319801	0.369274	0.564076	0.174078	0.246183	0.440386

Brownie

```
links = item_item_matrix.rename_axis('related_item',axis='columns').stack().reset_  
แมโขง 0.385758 0.369274 0.433013 1 0.218218 0.471405 0.333333 0.372678  
links
```

	level_0	related_item	0
0	จักรยานเสือภูเขา	จักรยานเสือภูเขา	1
1	จักรยานเสือภูเขา	คอลลาเจน	0.569803
2	จักรยานเสือภูเขา	Brownie อวกาศ	0.267261
3	จักรยานเสือภูเขา	แม่โขง	0.385758
4	จักรยานเสือภูเขา	Botox	0.404061
...	...	...	...
2495	น้ำส้มดีโต้	ข้าวขาหมูคากิคาจิกอนไลน์	0.188982
2496	น้ำส้มดีโต้	ถั่งเช่า	0.357143
2497	น้ำส้มดีโต้	หมอนสุขภาพสำหรับคนนอนไม่หลับ	0.634335
2498	น้ำส้มดีโต้	อาหารแมวคลีนเพื่อสุขภาพลดหุ่น	0.327327
2499	น้ำส้มดีโต้	น้ำส้มดีโต้	1

2500 rows x 3 columns

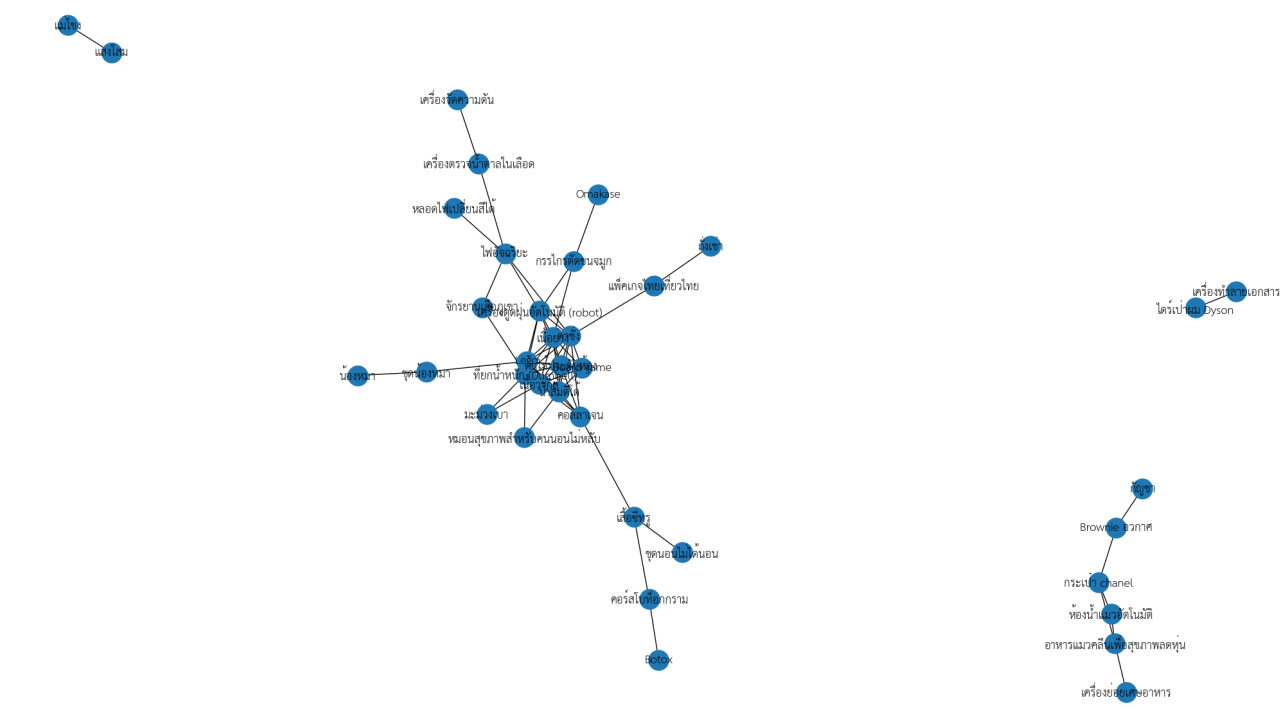
```
links.columns = ['item','related_item','เคยซื้อร่วมกัน']  
  
links_filered = links.loc[(links['เคยซื้อร่วมกัน'] > 0.6)& (links['item']!=links['relat  
  
links_filered.sort_values(by=['เคยซื้อร่วมกัน'],ascending=False)
```



	item	related_item	เคยซื้อร่วมกัน
1270	เนื้อย่าง	เนื้อวากิว	0.929164
1025	เนื้อวากิว	เนื้อย่าง	0.929164
703	แสง โสม	แม่โขง	0.880705
164	แม่โขง	แสง โสม	0.880705
2202	กระเป๋ា chanel	Brownie อวกาศ	0.866025
...	...	...	...
2468	น้ำส้มดีโต้	ที่ยกน้ำหนัก (Dumbell)	0.604743
1525	...	...	...

## ▼ Graph Visualization

```
G = nx.from_pandas_edgelist(links_filered,source='item',target='related_item')
nx.draw(G, node_size=500,with_labels=True,font_family='TH Sarabun New')
plt.axis('equal')
plt.rcParams["figure.figsize"] = (20,20)
plt.show()
```



```
links_filered.sort_values(by=[ 'เคยซื้อร่วมกัน' ],ascending=False).head(15)
```

	item	related_item	เคยซื้อร่วมกัน
1270	เนื้อย่าง	เนื้อวากิล	0.929164
1025	เนื้อวากิล	เนื้อย่าง	0.929164
703	แสง โสเม	แม่โขง	0.880705
164	แม่โขง	แสง โสเม	0.880705
2202	กระเป๋ា chanel	Brownie อวกาศ	0.866025
144	Brownie อวกาศ	กระเป๋า chanel	0.866025
565	ต้นไม้ประดับห้อง	Lego	0.830336
761	Lego	ต้นไม้ประดับห้อง	0.830336
1148	เครื่องย่อยเศษอาหาร	อาหารแมวคลีนเพื่อสุขภาพลดหุ่น	0.816497
2422	อาหารแมวคลีนเพื่อสุขภาพลดหุ่น	เครื่องย่อยเศษอาหาร	0.816497
1261	เนื้อย่าง	ต้นไม้ประดับห้อง	0.786216
575	ต้นไม้ประดับห้อง	เนื้อย่าง	0.786216
1412	Board game	ตาชั่ง	0.769231
628	ตาชั่ง	Board game	0.769231
1265	เนื้อย่าง	Lego	0.764775

```
links_filered.loc[(links['item'] == 'ไฟฉายฉายแสง')]
```

	item	related_item	เคยซื้อร่วมกัน
1587	ไฟอัจฉริยะ	เครื่องตรวจน้ำตาลในเลือด	0.64715
1562	ไฟอัจฉริยะ	ตาชั่ง	0.652714
1569	ไฟอัจฉริยะ	เครื่องดูดฝุ่นอัตโนมัติ (robot)	0.66575
1550	ไฟอัจฉริยะ	จักรยานเสือภูเขา	0.667124
1558	ไฟอัจฉริยะ	หลอดไฟเปลี่ยนสีได้	0.701646

## ▼ Interpretation

หลังจาก Run Model Collaborative Filtering ด้วยการหา Cosine similarity ของ Item to Item พบว่า Item ที่เคยซื้อเหมือนกัน โดย ลอง Filter Item ที่ Similarity > 0.6 มีหลากหลายรูปแบบทั้ง ของชนิดที่ไม่ทางเดียวกัน แต่ของที่ไปคนละทางกัน

- 86% ของคนที่เคยซื้อ "Brownie อวกาศ" เคยซื้อ "กระเป๋าสตางค์ Channel" โดยทั้ง 2 สิ่งนี้เป็นสิ่งดูคนละเรื่องกัน แต่ถ้าเมื่อลอง สืบหา Graph แล้วพบว่า คนที่เคยซื้อกระเป๋าสตางค์ channel ยังเคยซื้อของเกี่ยวกับแมวด้วย เช่น ห้องน้ำแมวอัตโนมัติ อาหารแมวคลีนลดหุ่น ซึ่งถ้าดูจาก ในวงนี้น่าจะเป็นกลุ่มที่สนใจแมว สนใจกระเป๋าสตางค์ Channel และสนใจสุขภาพ สิ่งที่เราควรแนะนำกลุ่มที่ชอบซื้อ Brownie อวกาศ ให้แนะนำกระเป๋าสตางค์ Brandname , สิ่งของสำหรับเลี้ยงแมว
- 83% ของคนที่เคยซื้อ "ต้นไม้ประดับ" เคยซื้อ "เลโก้" ซึ่งจริงๆ แล้วคนที่เคยซื้อเลโก้ จะเคยซื้ออีกหลายๆ อย่าง ซึ่งสามารถ Recommend ให้ซื้อต่อได้ เช่น ชุดน้องหมา, น้ำส้มดีโต้ , ที่ยกน้ำหนัก, ตาชั่ง, Boardgame , Collagen , วาคิว, เครื่องดูดฝุ่นอัตโนมัติ 60%-70%)
- 71% ของคนที่เคยซื้อเสื้อชีทู เคยซื้อชุดนอนไม่ได้นอน จะอยู่ในกลุ่มที่สนใจความสวยงาม สามารถแนะนำผลิตภัณฑ์อื่นๆต่อได้ เช่น Botox , Collagen
- 70% ของคนเคยซื้อไฟอัจฉริยะ เคยซื้อหลอดไฟเปลี่ยนสีได้ ซึ่งจริงๆ แล้วหลอดไฟเปลี่ยนสีอาจจะเป็นส่วนหนึ่งของหลอดไฟอัจฉริยะ แต่มี Gadget อื่นที่สามารถ Recommend ได้เช่น เครื่องตรวจน้ำตาลในเลือด, ตาชั่ง เครื่องดูดฝุ่นอัตโนมัติ และจักรยานเสือภูเขา (มากกว่า 60%)
- สุดท้าย แม้ว่า แม่โขง Vs แสงโสม , เนื้อวากิว Vs เนื้อย่าง จะมีค่า Similarity ที่สูงอยู่แล้ว (มากกว่า 88%) แต่เป็นสิ่งที่เรารู้อยู่แล้ว จึงไม่ควรจะนำมาแนะนำใหม่

---

✓

1s

completed at 21:39

●

×