

# Data preparation

```
df = pd.read_excel('Customer Survey.xlsx')
```

```
df.head(2)
```

	Unnamed: 0	ประทับ เวลา	จักรยาน เสือ ภูเขา	คอล ลา เจน	Brownie อวกาศ	แม่ โขง	Botox	กัญชา	กัญ ชง	เครื่อง วัด ความ ดัน	...	มะม่วง เบา	ห้องน้ำ แมว อัดโนมัต	กระเป๋า chanel	ข้าวขา หมูคากิ คาจิก ออนไลน์	ถัง เช่า	หมอน สำหรับ คนนอน ไม่หลับ	อาหาร แมวคลี นเพื่อ สุขภาพ ลดหุ่น	น้ำส้ม ดีได้	Software ถูก ลิขสิทธิ์
0	2021-03-20	11:53:45	ไม่เคย	ไม่ เคย	ไม่เคย	ไม่ เคย	ไม่เคย	ไม่เคย	ไม่ เคย	ไม่เคย	...	ไม่เคย	ไม่เคย	ไม่เคย	ไม่เคย	ไม่ เคย	ไม่เคย	ไม่เคย	เคย	เคย
1	2021-03-20	11:54:01	ไม่เคย	เคย	ไม่เคย	เคย	ไม่เคย	ไม่เคย	ไม่ เคย	เคย	...	เคย	ไม่เคย	ไม่เคย	ไม่เคย	เคย	เคย	ไม่เคย	เคย	ไม่เคย

2 rows x 63 columns

```
df.drop(['Unnamed: 0', 'ประทับเวลา'], axis=1, inplace=True)
```

```
df.replace({'ไม่เคย':0, 'เคย':1}, inplace=True)  
df.fillna(1, inplace=True)
```

# Data preparation

## Most famous item

```
df.iloc[:,1:].sum().sort_values(ascending=False)[:10]
```

กระเป๋	44.0
ชาบู	44.0
ครีมกันแดด	44.0
โรตี่	42.0
กะหรี่(ปั่น)	41.0
เม้ดมะม่วงหิมพานต์	40.0
หมอนข้าง	38.0
หมูแดดเดียว	36.0
แฟน ที่แปลว่าพัดลม	36.0
Software ถูกลิขสิทธิ์	34.0

dtype: float64

## Least famous item

```
df.iloc[:,1:].sum().sort_values(ascending=True)[:10]
```

PS5	2.0
เครื่องย่อยเศษอาหาร	2.0
คอร์ส bikini wax	3.0
เครื่องย่อยเศษอาหาร.1	3.0
Surfskate	3.0
อาหารแมวคลีนเพื่อสุขภาพลดหุ่น	3.0
ห้องน้ำแมวอัตโนมัติ	3.0
ไดร์เป่าผม Dyson	4.0
กัญชง	4.0
Brownie อวกาศ	5.0

dtype: float64

# Association rule

```
from mlxtend.frequent_patterns import apriori
from mlxtend.frequent_patterns import association_rules
```

```
frequent_itemsets = apriori(df, min_support=0.6, use_colnames=True, low_memory=True)
rules = association_rules(frequent_itemsets, metric="lift", min_threshold=1)
rules.head(15)
```

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
0	(หมอนข้าง)	(ต้นไม้ประดับห้อง)	0.863636	0.704545	0.636364	0.736842	1.045840	0.027893	1.122727
1	(ต้นไม้ประดับห้อง)	(หมอนข้าง)	0.704545	0.863636	0.636364	0.903226	1.045840	0.027893	1.409091
2	(ต้นไม้ประดับห้อง)	(กะหล่ำปลี)	0.704545	0.931818	0.659091	0.935484	1.003934	0.002583	1.056818
3	(กะหล่ำปลี)	(ต้นไม้ประดับห้อง)	0.931818	0.704545	0.659091	0.707317	1.003934	0.002583	1.009470
4	(แฟน ที่แปลว่าพัลลภ)	(ต้นไม้ประดับห้อง)	0.818182	0.704545	0.613636	0.750000	1.064516	0.037190	1.181818
5	(ต้นไม้ประดับห้อง)	(แฟน ที่แปลว่าพัลลภ)	0.704545	0.818182	0.613636	0.870968	1.064516	0.037190	1.409091
6	(หมูแดดเดียว)	(ต้นไม้ประดับห้อง)	0.818182	0.704545	0.613636	0.750000	1.064516	0.037190	1.181818
7	(ต้นไม้ประดับห้อง)	(หมูแดดเดียว)	0.704545	0.818182	0.613636	0.870968	1.064516	0.037190	1.409091
8	(กระเป๋)	(ต้นไม้ประดับห้อง)	1.000000	0.704545	0.704545	0.704545	1.000000	0.000000	1.000000
9	(ต้นไม้ประดับห้อง)	(กระเป๋)	0.704545	1.000000	0.704545	1.000000	1.000000	0.000000	inf
10	(ต้นไม้ประดับห้อง)	(ข้าว)	0.704545	1.000000	0.704545	1.000000	1.000000	0.000000	inf

# Association rule

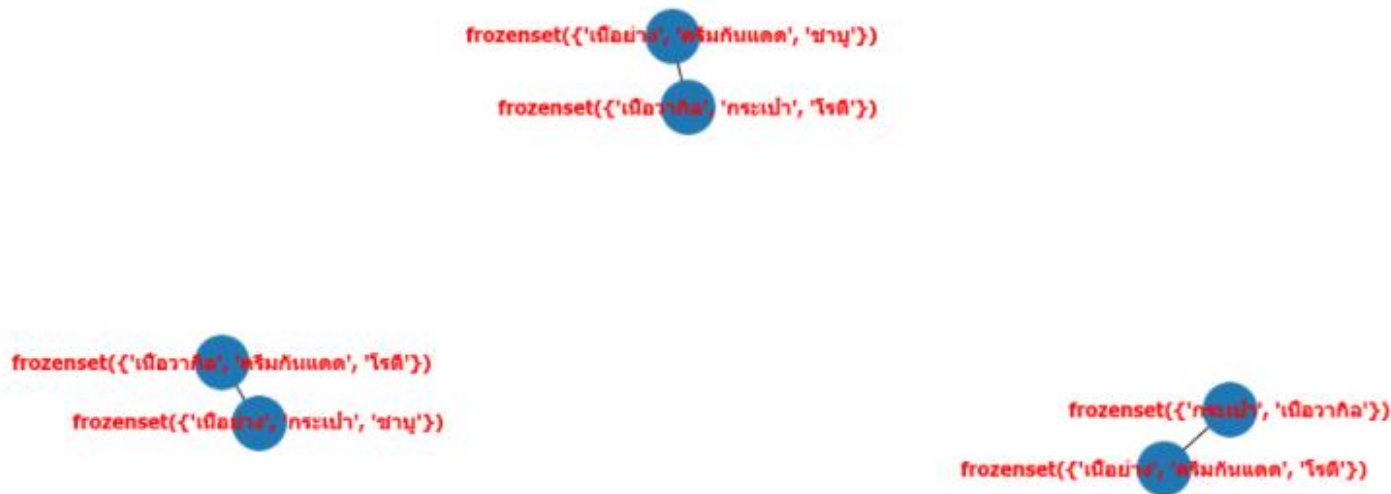
Filter for only association with lift >1 and confidence>0.5 to get only items with high association and occurrence

rules = rules[(rules['lift'] >= 1) & (rules['confidence'] >= 0.5) ]									
rules = rules.sort_values('lift', ascending=False)[:20]									
rules.head(50)									
	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
13674	(เนื้อย่าง, ครีมหั่นแฉลบ, โรตีส)	(กระเป๋, เนื้อวากิล)	0.704545	0.659091	0.636364	0.903226	1.370412	0.172004	3.522727
5010	(ขานู, ครีมหั่นแฉลบ, เนื้อวากิล)	(เนื้อย่าง)	0.659091	0.704545	0.636364	0.965517	1.370412	0.172004	8.568182
13689	(เนื้อย่าง)	(โรตีส, กระเป๋, ครีมหั่นแฉลบ, เนื้อวากิล)	0.704545	0.659091	0.636364	0.903226	1.370412	0.172004	3.522727
24621	(เนื้อวากิล, โรตีส)	(เนื้อย่าง, กระเป๋, ครีมหั่นแฉลบ, ขานู)	0.659091	0.704545	0.636364	0.965517	1.370412	0.172004	8.568182
5006	(เนื้อวากิล)	(เนื้อย่าง, ขานู, โรตีส)	0.659091	0.704545	0.636364	0.965517	1.370412	0.172004	8.568182
5005	(เนื้อย่าง)	(เนื้อวากิล, ขานู, โรตีส)	0.704545	0.659091	0.636364	0.903226	1.370412	0.172004	3.522727
24583	(เนื้อวากิล, โรตีส, กระเป๋, ครีมหั่นแฉลบ, ขานู)	(เนื้อย่าง)	0.659091	0.704545	0.636364	0.965517	1.370412	0.172004	8.568182
5003	(ขานู, เนื้อวากิล)	(เนื้อย่าง, โรตีส)	0.659091	0.704545	0.636364	0.965517	1.370412	0.172004	8.568182
5002	(เนื้อย่าง, ขานู)	(เนื้อวากิล, โรตีส)	0.704545	0.659091	0.636364	0.903226	1.370412	0.172004	3.522727
24619	(เนื้อย่าง, ครีมหั่นแฉลบ, ขานู)	(เนื้อวากิล, กระเป๋, โรตีส)	0.704545	0.659091	0.636364	0.903226	1.370412	0.172004	3.522727
4999	(เนื้อวากิล, โรตีส)	(เนื้อย่าง, ขานู)	0.659091	0.704545	0.636364	0.965517	1.370412	0.172004	8.568182
4998	(เนื้อย่าง, โรตีส)	(ขานู, เนื้อวากิล)	0.704545	0.659091	0.636364	0.903226	1.370412	0.172004	3.522727
4996	(เนื้อวากิล, ขานู, โรตีส)	(เนื้อย่าง)	0.659091	0.704545	0.636364	0.965517	1.370412	0.172004	8.568182

# Association rule

- คนที่เคยกินเนื้ออย่างมักจะเคยซื้อกระเป๋, ครีมกันแดดและเคยกินโรตีสี,เนื้อวากิว
- คนที่เคยกินเนื้อวากิวมักจะเคยกินเนื้ออย่าง,ชาบู และโรตีสี ในขณะที่คนที่เคยกินเนื้ออย่างมักจะเคยกินวากิว,ชาบู และโรตีสีเช่นกัน

```
import matplotlib.pyplot as plt
import networkx as nx
from decorator import decorator
fig, ax=plt.subplots(figsize=(20,20))
GA=nx.from_pandas_edgelist(rules,source='antecedents',target='consequents')
nx.draw(GA, with_labels=True, node_size=1000, font_color="red", font_weight="bold",font_family='Tahoma')
plt.show()
```



[illegible]

# Item to Item

```
item_item_matrix = pd.DataFrame(index=df.columns,columns=df.columns)|
```

```
item_item_matrix
```

	จักรยาน เสือ ภูเขา	คอล ลา เจน	Brownie อวกาศ	แม วีน	Botox	กัญชา	กัญ ชง	เครื่อง วัด ความ ดัน	หลอด ไฟ เปลี่ยน สีได้	คอร์ส bikini wax	...	มะม่วง เบา	ห ัด
จักรยานเสือ ภูเขา	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	
คอลลาเจน	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	
Brownie อวกาศ	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	
แมวีน	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	
Botox	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	
กัญชา	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	
กัญชง	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	

```
from scipy.spatial.distance import cosine
```

```
for i in range(0,len(item_item_matrix.columns)) :  
    # Loop through the columns for each column  
    for j in range(0,len(item_item_matrix.columns)) :  
        # Fill in placeholder with cosine similarities  
        item_item_matrix.iloc[i,j] = 1 -cosine(df.iloc[:,i],df.iloc[:,j])
```

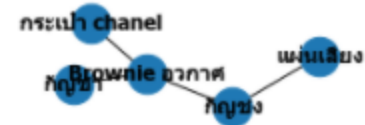
```
links = item_item_matrix.rename_axis('related item',  
                                     axis='columns').stack().reset_index()  
links.columns = ['item', 'related item','value']
```

# Item to Item

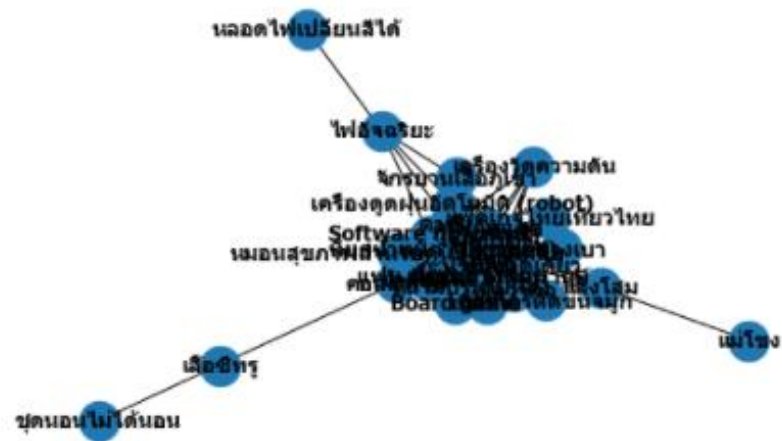
```
links_filtered=links.loc[ (links['value'] > 0.6) &  
                          (links['item'] != links['related item']) ]
```

```
import matplotlib.pyplot as plt  
import networkx as nx
```

```
fig, ax=plt.subplots(figsize=(20,20))  
GA=nx.from_pandas_edgelist(links_filtered,source='item',target='related item')  
nx.draw(GA, with_labels=True, node_size=500, font_color="black", font_weight="bold",font_family='Tahoma')  
plt.show()
```







คอร์สโมโยกการาม  
Botox

# Interpretation

## Strong association (Not surprised)

- คนที่เคยซื้อหมามักเคยซื้อชุดหมา
- คนที่เคยซื้อคอร์ส **Botox** มักเคยฉีด **Botox** ที่กราม
- คนที่เคยซื้อชุดนอนไม่ได้นอนมักเคยซื้อชุดชีทู
- คนที่เคยซื้อแม่โขงมักเคยซื้อแสงโสม
- คนที่เคยซื้อไฟเปลี่ยนสีได้มักเคยซื้อไฟอัจฉริยะ

## Strong association (Surprised)

- คนที่เคยซื้อกระเป๋า **channel** มักเคยซื้อบราวนี้อวกาศ
- คนที่เคยซื้อบราวนี้อวกาศมักเคยซื้อกัญชาและกัญชง
- คนที่เคยซื้อกัญชงมักเคยซื้อแผ่นเสียง
- คนที่เคยซื้อเครื่องย่อยเศษอาหารมักเคยซื้ออาหารแมวเพื่อสุขภาพ
- คนที่ซื้อชาบูมักซื้อตราซัง หมอนข้าง ต้นไม้ประดับห้อง คอลลาเจน (น่าจะเป็นผู้หญิงส่วนใหญ่)

example

#Sort by Item links_filtered			
121	คอลลาเจน	ครีมกันแดด	0.722999
127	Brownie อวกาศ	กัญชา	0.632456
128	Brownie อวกาศ	กัญชง	0.67082
176	Brownie อวกาศ	กระเป๋า chanel	0.730297
198	แม่โขง	แสงโสม	0.868599
268	Botox	คอร์สโบท็อกกราม	0.745356
307	กัญชา	Brownie อวกาศ	0.632456
368	กัญชง	Brownie อวกาศ	0.67082
408	กัญชง	แผ่นเสียง	0.666667
439	เครื่องวัดความดัน	หมอนข้าง	0.629512
440	เครื่องวัดความดัน	ตาชั่ง	0.61993
468	เครื่องวัดความดัน	กระเป๋า	0.621582