

```
In [64]: import numpy as np
import pandas as pd
import glob
import time
import re
import seaborn as sb
import matplotlib.pyplot as plt
import matplotlib.font_manager as fm
```

## Data loading

### ทำการ load data จาก excel sheet

```
In [66]: crm_survey = pd.read_excel('Customer_Behaviors_Responses.xlsx', sheet_name = 'Form Respo
```

## Data Cleansing and Data Encoding

### ทำการ encoding data จาก category to numerical range และ ทำการเปลี่ยนวันเดือนปีเกิด ให้อยู่ใน format คศ และทำการหาอายุ

```
In [67]: list_grade_inter = {'สนใจอย่างมากที่สุด':7, 'สนใจอย่างมาก':6, 'สนใจ':5, 'เฉยๆ':4, 'ไม่สนใจ':3, 'ไม่สนใจอย่างมาก':2}
consume = {'แทบทุกวัน':6, 'อาทิตย์ละครั้ง':5, 'เดือนละ 2-3 ครั้ง':4, 'หลายเดือนครั้ง':2, 'แทบไม่ได้บริโภค':1, 'เดือนละครั้ง':3}
```

```
In [68]: ## clean data
day_brith = ['Year', 'Month', 'Day']
# 'คุณเกิดวันที่เท่าไร'
crm_survey['คุณเกิดวันที่เท่าไร_s'] = crm_survey['คุณเกิดวันที่เท่าไร'].astype(str)
crm_survey[day_brith] = crm_survey['คุณเกิดวันที่เท่าไร_s'].str.split('-', expand = True)
crm_survey['Year'] = crm_survey['Year'].apply(lambda x : int(x)-543 if int(x) > 2400 else int(x))
crm_survey['Day'] = crm_survey['Day'].apply(lambda x : x.split()[0])
crm_survey['Age'] = crm_survey['Year'].apply(lambda x : 2021- int(x))
##### Mapping value ##
def get_value(x) :
    if x in consume:
        y = consume.get(x)
    else :
        y = -1
    return y

def map_value(x) :
    if x == 'สนใจอย่างมากที่สุด' :
        y=7
    elif x == 'สนใจอย่างมาก' :
        y = 6
    elif x == 'สนใจ' :
        y = 5
    elif x == 'เฉยๆ' :
        y = 4
    elif x == 'ไม่สนใจ' :
        y = 3
    elif x == 'ไม่สนใจอย่างมาก' :
        y = 2
    elif x == 'ไม่สนใจอย่างมากที่สุด' :
```

[illegible]



```

crm_survey["คุณบริโภคสิ่งเหล่านี้บ่อยขนาดไหน [การท่องเที่ยว]"] =crm_survey["คุณบริโภคสิ่งเหล่านี้บ่อยขนาดไหน [การท่องเที่ยว]"]
crm_survey["คุณบริโภคสิ่งเหล่านี้บ่อยขนาดไหน [การอ่านหนังสือ]"] =crm_survey["คุณบริโภคสิ่งเหล่านี้บ่อยขนาดไหน [การอ่านหนังสือ]"]
crm_survey["คุณบริโภคสิ่งเหล่านี้บ่อยขนาดไหน [การอ่านบทความ Online]"] =crm_survey["คุณบริโภคสิ่งเหล่านี้บ่อยขนาดไหน [การอ่านบทความ Online]"]
crm_survey["คุณบริโภคสิ่งเหล่านี้บ่อยขนาดไหน [การเรียนรู้ e-learning]"] =crm_survey["คุณบริโภคสิ่งเหล่านี้บ่อยขนาดไหน [การเรียนรู้ e-learning]"]
crm_survey["คุณบริโภคสิ่งเหล่านี้บ่อยขนาดไหน [ดู Series]"] =crm_survey["คุณบริโภคสิ่งเหล่านี้บ่อยขนาดไหน [ดู Series]"]
crm_survey["คุณบริโภคสิ่งเหล่านี้บ่อยขนาดไหน [ดูหนัง]"] =crm_survey["คุณบริโภคสิ่งเหล่านี้บ่อยขนาดไหน [ดูหนัง]"]
crm_survey["คุณบริโภคสิ่งเหล่านี้บ่อยขนาดไหน [ดู Youtube]"] =crm_survey["คุณบริโภคสิ่งเหล่านี้บ่อยขนาดไหน [ดู Youtube]"]
crm_survey["คุณบริโภคสิ่งเหล่านี้บ่อยขนาดไหน [ดู Netflix]"] =crm_survey["คุณบริโภคสิ่งเหล่านี้บ่อยขนาดไหน [ดู Netflix]"]
crm_survey["คุณบริโภคสิ่งเหล่านี้บ่อยขนาดไหน [Content ด้านบันเทิง]"] =crm_survey["คุณบริโภคสิ่งเหล่านี้บ่อยขนาดไหน [Content ด้านบันเทิง]"]
crm_survey["คุณบริโภคสิ่งเหล่านี้บ่อยขนาดไหน [Content เกี่ยวกับเกมส์]"] =crm_survey["คุณบริโภคสิ่งเหล่านี้บ่อยขนาดไหน [Content เกี่ยวกับเกมส์]"]
crm_survey["คุณบริโภคสิ่งเหล่านี้บ่อยขนาดไหน [Content เกี่ยวกับการ review ร้านอาหาร]"] =crm_survey["คุณบริโภคสิ่งเหล่านี้บ่อยขนาดไหน [Content เกี่ยวกับการ review ร้านอาหาร]"]
crm_survey["คุณบริโภคสิ่งเหล่านี้บ่อยขนาดไหน [Content เกี่ยวกับการท่องเที่ยว]"] =crm_survey["คุณบริโภคสิ่งเหล่านี้บ่อยขนาดไหน [Content เกี่ยวกับการท่องเที่ยว]"]
crm_survey["คุณบริโภคสิ่งเหล่านี้บ่อยขนาดไหน [Content เกี่ยวกับการแต่งงาน]"] =crm_survey["คุณบริโภคสิ่งเหล่านี้บ่อยขนาดไหน [Content เกี่ยวกับการแต่งงาน]"]
crm_survey["คุณบริโภคสิ่งเหล่านี้บ่อยขนาดไหน [Content เกี่ยวกับการโทรศัพท์]"] =crm_survey["คุณบริโภคสิ่งเหล่านี้บ่อยขนาดไหน [Content เกี่ยวกับการโทรศัพท์]"]
crm_survey["คุณบริโภคสิ่งเหล่านี้บ่อยขนาดไหน [Content เกี่ยวกับละคร]"] =crm_survey["คุณบริโภคสิ่งเหล่านี้บ่อยขนาดไหน [Content เกี่ยวกับละคร]"]
crm_survey["คุณบริโภคสิ่งเหล่านี้บ่อยขนาดไหน [Content เกี่ยวกับการเมือง]"] =crm_survey["คุณบริโภคสิ่งเหล่านี้บ่อยขนาดไหน [Content เกี่ยวกับการเมือง]"]
#####

```

## Impute Missing Value with Age

Found some data put wrong birth date use mean age to impute missing value

```

In [69]: temp_survey = crm_survey.copy()
temp_survey2= crm_survey.copy()
#### replace missig value
temp_survey['Age'] = temp_survey[['Age']].replace(0, np.nan )
temp_survey['Age1'] = temp_survey['Age'].fillna(temp_survey['Age'].mean(), inplace=True)

temp_survey= temp_survey.drop('Age1', axis=1)
temp_survey

```

Out[69]:

	Timestamp	คุณมี ความ สนใจ ในสิ่ง เหล่านี้ มาก น้อย เพียง ใด [อาหาร ญี่ปุ่น]	คุณมี ความ สนใจ ในสิ่ง เหล่านี้ มาก น้อย เพียง ใด [อา หาร จีน]	คุณมี ความ สนใจ ในสิ่ง เหล่านี้ มาก น้อย เพียง ใด [อาหาร ไทย]	คุณมี ความ สนใจ ในสิ่ง เหล่านี้ มาก น้อย เพียง ใด [อาหาร อีสาน]	คุณมี ความ สนใจ ใน สิ่ง เหล่านี้ มาก น้อย เพียง ใด [อาหาร สุขภาพ]	คุณมี ความ สนใจ ใน สิ่ง เหล่านี้ มาก น้อย เพียง ใด [บุฟเฟต์]	คุณมี ความ สนใจ ในสิ่ง เหล่านี้ มาก น้อย เพียง ใด [ชา บูชา บู]	คุณมี ความ สนใจ ในสิ่ง เหล่านี้ มาก น้อย เพียง ใด [ปิ้ง ย่าง]	คุณมี ความ สนใจ ในสิ่ง เหล่านี้ มาก น้อย เพียง ใด [เหล้า เบียร์]	...	คุณ บริโภค สิ่งเหล่านี้ บ่อย ขนาด ไหน [Content เกี่ยวกับ ละคร]
0	2021-01-09 10:07:27.365	5	5	7	6	5	5	5	5	7	...	4
1	2021-01-09 10:07:34.010	6	4	7	6	5	6	7	6	3	...	2
2	2021-01-09 10:07:35.902	7	4	6	5	4	7	7	7	2	...	2

	Timestamp	คุณมีความสนใจในสิ่งเหล่านี้	คุณมีความสนใจในสิ่งเหล่านี้	คุณมีความสนใจในสิ่งเหล่านี้	คุณมีความสนใจในสิ่งเหล่านี้	คุณมีความสนใจในสิ่งเหล่านี้	คุณมีความสนใจในสิ่งเหล่านี้	คุณมีความสนใจในสิ่งเหล่านี้	คุณมีความสนใจในสิ่งเหล่านี้	คุณมีความสนใจในสิ่งเหล่านี้	...	คุณบริโภคสิ่งเหล่านี้บ่อยขนาดไหน [Content เกี่ยวกับละคร]
		มากน้อยเพียงใด [อาหารญี่ปุ่น]	มากน้อยเพียงใด [อาหารไทย]	มากน้อยเพียงใด [อาหารไทย]	มากน้อยเพียงใด [อาหารอีสาน]	น้อยนิด [อาหารสุขภาพ]	น้อยนิด [บุฟเฟต์]	มากน้อยเพียงใด [ชาบู]	มากน้อยเพียงใด [ปิ้งย่าง]	น้อยนิด [เหล้าเบียร์]		
3	2021-01-09 10:07:36.476	5	4	4	4	5	5	5	5	1	...	1
4	2021-01-09 10:07:41.965	6	5	5	5	7	4	4	4	4	...	1
...	...	...	...	...	...	...	...	...	...	...	...	...
58	2021-01-09 10:17:02.033	5	7	7	7	7	4	4	4	4	...	6
59	2021-01-09 10:17:52.578	6	5	6	5	7	6	7	6	4	...	2
60	2021-01-09 10:18:23.803	5	4	6	6	5	2	4	5	3	...	3
61	2021-01-09 10:20:53.832	7	5	6	6	6	7	7	7	4	...	2
62	2021-01-09 10:21:33.057	5	4	4	4	6	4	4	4	3	...	2

63 rows × 69 columns

## Rename of columns

```
In [70]: cm1 = temp_survey.iloc[:, :].columns.rename('col_name')
cm1_interest = temp_survey.loc[:, 'คุณมีความสนใจในสิ่งเหล่านี้มากน้อยเพียงใด [อาหารญี่ปุ่น]': 'คุณมีความสนใจในสิ่งเหล่านี้มากน้อยเพียงใด [อาหารไทย]'
cm1_consume = temp_survey.loc[:, 'คุณบริโภคสิ่งเหล่านี้บ่อยขนาดไหน [อาหารญี่ปุ่น]': 'คุณบริโภคสิ่งเหล่านี้บ่อยขนาดไหน [อาหารอีสาน]'
#cm1 = temp_survey.iloc[:, 1:-8].columns.rename('col_name')

doc_col = pd.DataFrame(cm1)
#u= 'sd'
def split_name_inter(x) :
    try :
        if 'สนใจ' in x :
```

```

word = x.split(' ')[1]
y = f'สนใจ[{word}]'
return y
elif 'บริโภค' in x :
word = x.split(' ')[1]
y = f'บริโภค[{word}]'
return y
else :
return x
except :
return x

#cm1[['']]
doc_col['col_name'] = doc_col['col_name'].apply( split_name_inter )
columns_new = doc_col['col_name'].to_list()
columns_old = temp_survey.columns.to_list()
columns_t = dict(zip(columns_old,columns_new))
temp_survey = temp_survey.rename(columns=columns_t )#temp_survey

```

## Display final data

In [156... temp\_survey.columns

Out[156... Index(['Timestamp', 'สนใจ[อาหารญี่ปุ่น]', 'สนใจ[อาหารจีน]', 'สนใจ[อาหารไทย]', 'สนใจ[อาหารอีสาน]', 'สนใจ[อาหารสุขภาพ]', 'สนใจ[บุฟเฟต์]', 'สนใจ[ชาบูชาบู]', 'สนใจ[ปิ้งย่าง]', 'สนใจ[เหล้าเบียร์]', 'สนใจ[เสื้อผ้าแฟชั่น]', 'สนใจ[แต่งหน้า]', 'สนใจ[skincare บำรุงผิว]', 'สนใจ[การออกกำลังกาย]', 'สนใจ[การเล่นเกมส์]', 'สนใจ[การท่องเที่ยว]', 'สนใจ[การอ่านหนังสือ]', 'สนใจ[การอ่านบทความ Online]', 'สนใจ[การเรียนรู้ e-learning]', 'สนใจ[ดู Series]', 'สนใจ[ดูหนัง]', 'สนใจ[ดู Youtube]', 'สนใจ[ดู Netflix]', 'สนใจ[Content ด้านบันเทิง]', 'สนใจ[Content เกี่ยวกับเกมส์]', 'สนใจ[Content เกี่ยวกับการ review ร้านอาหาร]', 'สนใจ[Content เกี่ยวกับการท่องเที่ยว]', 'สนใจ[Content เกี่ยวกับการแต่งบ้าน]', 'สนใจ[Content เกี่ยวกับการโทรศัพท์]', 'สนใจ[Content เกี่ยวกับละคร]', 'สนใจ[Content เกี่ยวกับการเมือง]', 'บริโภค[อาหารญี่ปุ่น]', 'บริโภค[อาหารจีน]', 'บริโภค[อาหารไทย]', 'บริโภค[อาหารอีสาน]', 'บริโภค[อาหารสุขภาพ]', 'บริโภค[บุฟเฟต์]', 'บริโภค[ชาบูชาบู]', 'บริโภค[ปิ้งย่าง]', 'บริโภค[เหล้าเบียร์]', 'บริโภค[เสื้อผ้าแฟชั่น]', 'บริโภค[แต่งหน้า]', 'บริโภค[skincare บำรุงผิว]', 'บริโภค[การออกกำลังกาย]', 'บริโภค[การเล่นเกมส์]', 'บริโภค[การท่องเที่ยว]', 'บริโภค[การอ่านหนังสือ]', 'บริโภค[การอ่านบทความ Online]', 'บริโภค[การเรียนรู้ e-learning]', 'บริโภค[ดู Series]', 'บริโภค[ดูหนัง]', 'บริโภค[ดู Youtube]', 'บริโภค[ดู Netflix]', 'บริโภค[Content ด้านบันเทิง]', 'บริโภค[Content เกี่ยวกับเกมส์]', 'บริโภค[Content เกี่ยวกับการ review ร้านอาหาร]', 'บริโภค[Content เกี่ยวกับการท่องเที่ยว]', 'บริโภค[Content เกี่ยวกับการแต่งบ้าน]', 'บริโภค[Content เกี่ยวกับการโทรศัพท์]', 'บริโภค[Content เกี่ยวกับละคร]', 'บริโภค[Content เกี่ยวกับการเมือง]', 'นอกเหนือจากรายการดังกล่าวแล้ว คุณมีความสนใจในเรื่องใดอีกบ้าง', 'คุณเกิดวันที่เท่าไร', 'เพศ', 'คุณเกิดวันที่เท่าไร\_s', 'Year', 'Month', 'Day', 'Age'], dtype='object')

## Graph Plot



```
In [72]: # rr = temp_survey.columns.to_list()
# import matplotlib.pyplot as plt
# import matplotlib.font_manager as fm
# jp = [ i for i in rr if (('คุณมีความสนใจ') in i) ]
# nrow = 3
# ncol = 10

# #plt.figure(figsize=(200,100))
# fig, axes = plt.subplots(nrow, ncol)

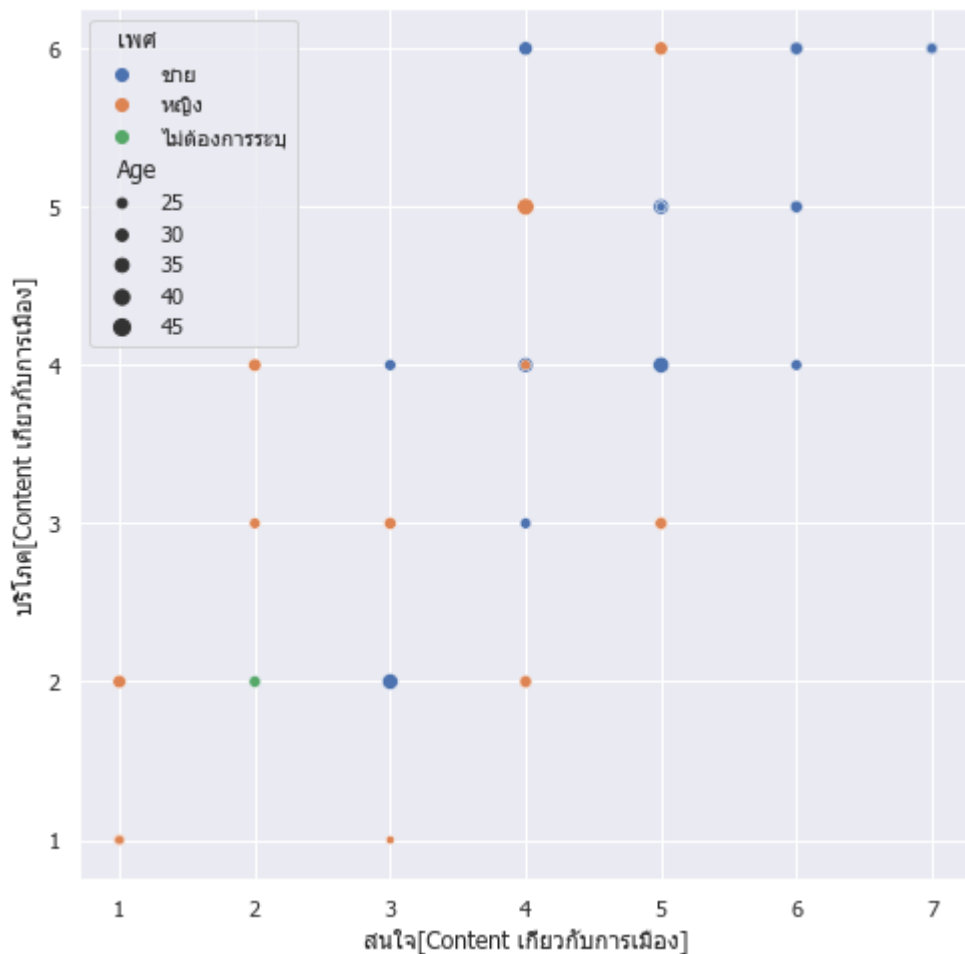
# #plt.figure(figsize=(30,4))
# c = 0
# #font_path = "/System/Library/Fonts/STHeiti Light.ttc"
# plt.rcParams["font.family"] = 'Tahoma'
# #prop = mfm.FontProperties(fname=font_path)
# count = 0
# for i in range(nrow) :
#     for j in range(ncol) :
#         #c = 0
#         #plt.rcParams['font.family'] = 'Tahoma'
#         temp_survey.groupby([jp[count], 'เพศ']).size().unstack().plot(kind='bar', stack
#         plt.rcParams["font.family"] = 'Tahoma'
#         #ax.set_xticklabels(rr.get_xticklabels(), fontproperties=prop)
#         count +=1
#         # print(count)
# #plt.rcParams['font.family'] = 'Tahoma'
# # plt.figure(figsize=(3,1000))
# # plt.show()
#         #temp_survey[temp_survey[[jp[i]],jp[2]]#.size().unstack().plot(kind='bar', sta
#
# temp_survey.groupby(['เพศ', 'Age']).size()#.plot(kind='line',xticks = 'Age')#.unstack()
```

## SAMPLE PLOT BETWEEN สนใจการเมือง และ การบริโภคข่าวการเมือง

หาความสัมพันธ์ของความสนใจในการเมืองและการบริโภคข่าวการเมือง พบว่าผู้ชายส่วนมากมีความสนใจการเมืองและบริโภคมากกว่าผู้หญิง

```
In [162... #temp = temp_survey.loc[['บริโภค[Content เกี่ยวกับการเมือง]', 'สนใจ[Content เกี่ยวกับการเมือง]']]
import matplotlib.font_manager as fm
sb.set_style("darkgrid",{"font.sans-serif":["simhei", 'Tahoma']})
tempee = temp_survey.loc[:, ['บริโภค[Content เกี่ยวกับการเมือง]', 'สนใจ[Content เกี่ยวกับการเมือง]']]
#plt.rcParams['font.family'] = 'Tahoma'
f, ax = plt.subplots(figsize =(8, 8))
#sb.rcParams['font.family'] = 'Tahoma'
sb.scatterplot(y='บริโภค[Content เกี่ยวกับการเมือง]',x='สนใจ[Content เกี่ยวกับการเมือง]', hue='เพศ')
#sb.scatterplot(x='บริโภค[Content เกี่ยวกับการเมือง]',y='สนใจ[Content เกี่ยวกับการเมือง]', hue='เพศ')
```

```
Out[162... <AxesSubplot:xlabel='สนใจ[Content เกี่ยวกับการเมือง]', ylabel='บริโภค[Content เกี่ยวกับการเมือง]'
```



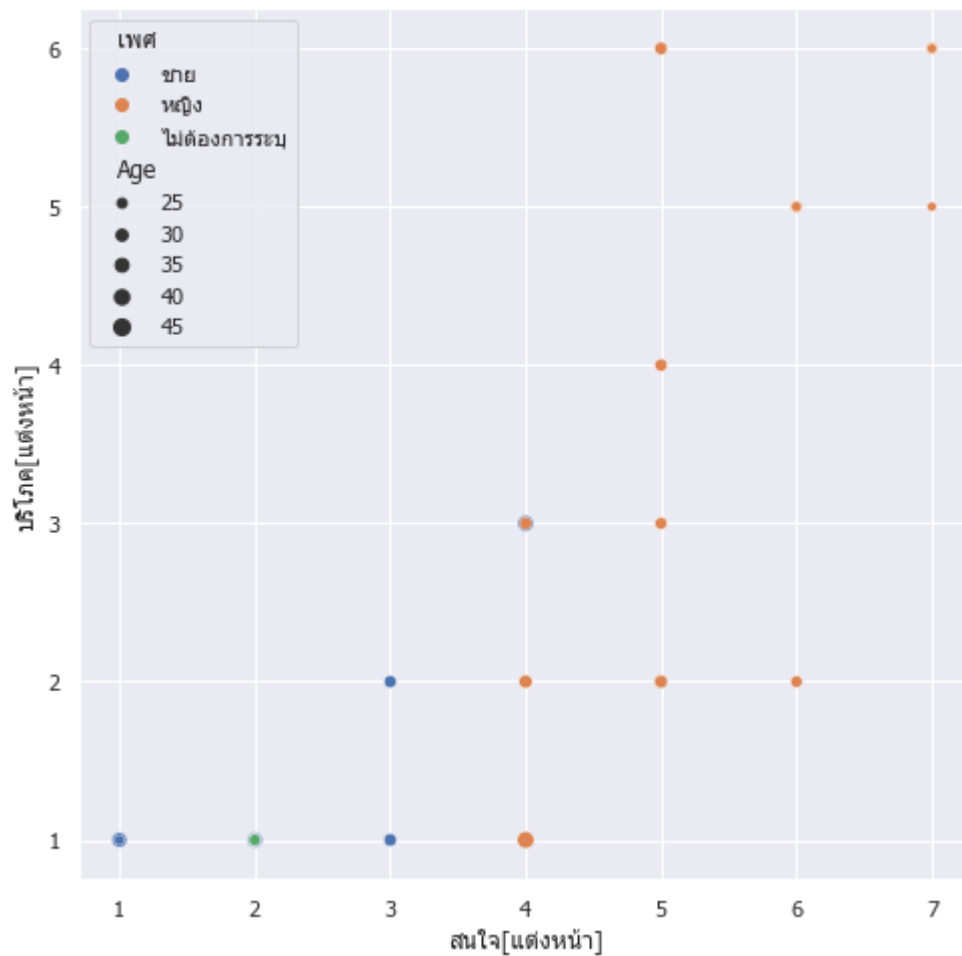
## SAMPLE PLOT BETWEEN สนใจ[แต่งงาน]และบริโภค[แต่งงาน]

จะเห็นว่าเพศหญิงจะมีความสนใจแต่งงานและใช้การแต่งงานที่คะแนนความสัมพันธ์ที่สูงโดยเฉพาะช่วงที่มีอายุน้อยๆ เมื่อเทียบกับผู้ชาย

```
In [160... sb.set_style("darkgrid",{"font.sans-serif":["simhei", 'Tahoma']})
tempee = temp_survey.loc[:, ['บริโภค[แต่งงาน]', 'สนใจ[แต่งงาน]', '', 'เพศ', 'Age']]
#plt.rcParams['font.family'] = 'Tahoma'
f, ax = plt.subplots(figsize=(8, 8))
#sb.rcParams['font.family'] = 'Tahoma'
sb.scatterplot(y='บริโภค[แต่งงาน]', x='สนใจ[แต่งงาน]', hue='เพศ', size='Age', data=tempee)
```

```
Out[160... <AxesSubplot:xlabel='สนใจ[แต่งงาน]', ylabel='บริโภค[แต่งงาน]'
```



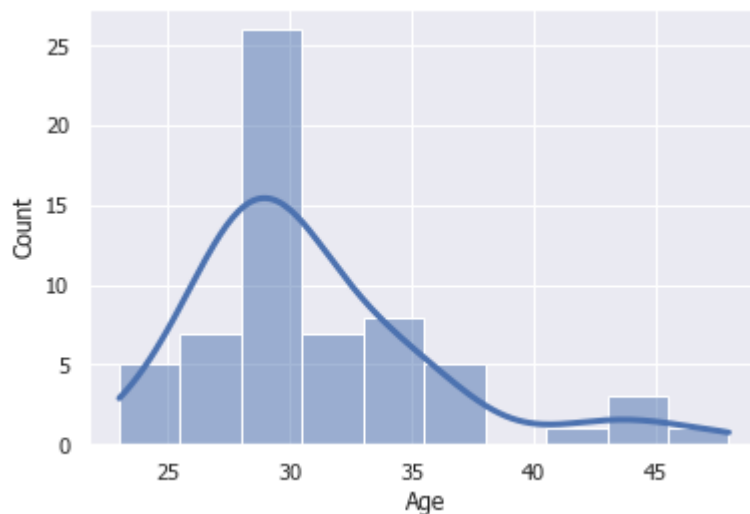


## Age Distributed

ดูการกระจายตัวของอายุของการทำสำรวจ พบว่าจะอยู่ที่ช่วงอายุ 30 ปี

```
In [74]: age = temp_survey.Age
sb.histplot(data=age,kde=True,line_kws={"linewidth":3})
```

```
Out[74]: <AxesSubplot:xlabel='Age', ylabel='Count'>
```

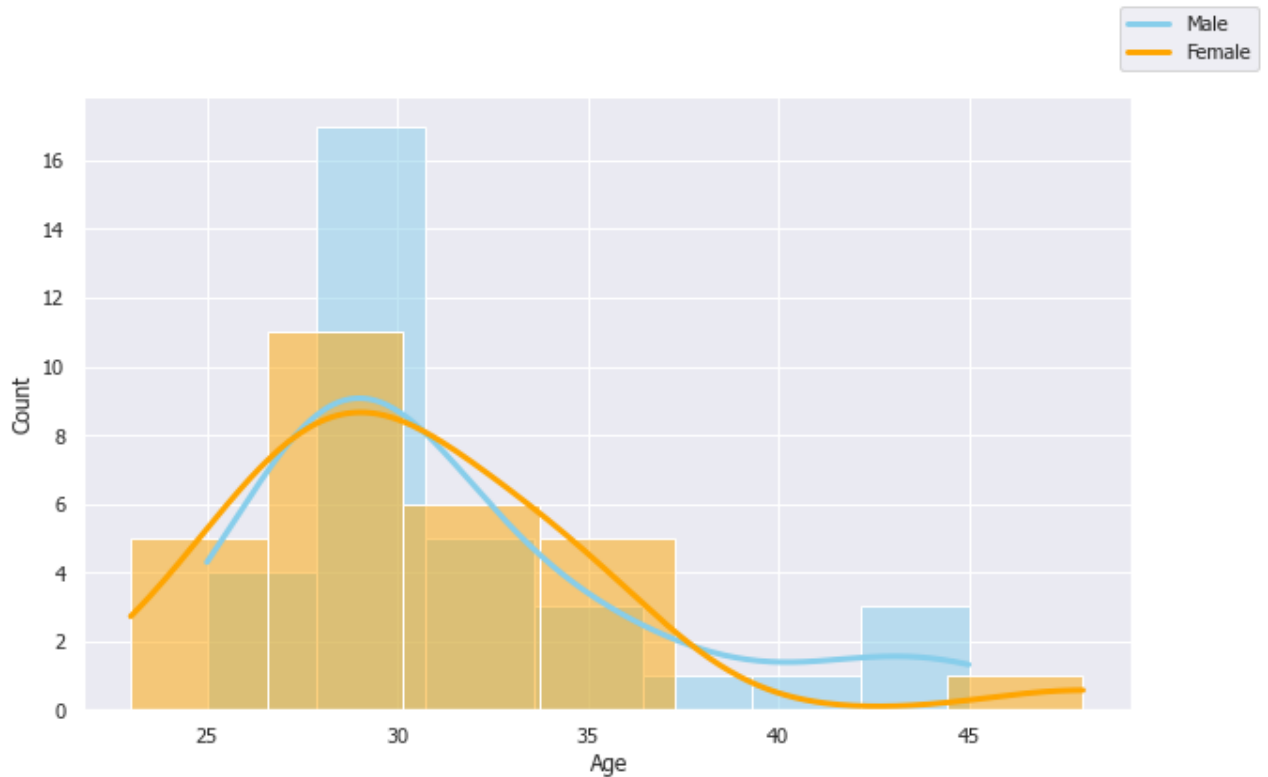


## Find Age Distributed Male & Female

```
In [210]: import matplotlib.pyplot as plt
fig = plt.figure(figsize=(10,6))

temp_man = temp_survey[temp_survey['เพศ'] == 'ชาย']
temp_woman = temp_survey[temp_survey['เพศ'] == 'หญิง']
dd = sb.histplot( temp_man["Age"] , color="skyblue",kde=True,line_kws={"linewidth":3}, 1
ee = sb.histplot( temp_woman["Age"] , color="orange",kde=True,line_kws={"linewidth":3},
fig.legend(labels=['Male', 'Female'])
#plt.legend(handles = [dd,ee])
#sb.show()plt.legend(handles=[blue_line])
```

Out[210]: <matplotlib.legend.Legend at 0x15bb723ac08>



## Find mean

```
In [76]: display(temp_survey.groupby('เพศ').mean())
```

	สนใจ[อาหารญี่ปุ่น]	สนใจ[อาหารจีน]	สนใจ[อาหารไทย]	สนใจ[อาหารอีสาน]	สนใจ[อาหารสุขภาพ]	สนใจ[บุฟเฟต์]	สนใจ[ชาบู]	สนใจ[กาแฟ]
เพศ								
ชาย	6.147059	4.647059	5.647059	5.117647	4.941176	5.470588	5.764706	5.794118
หญิง	6.035714	4.821429	5.678571	5.250000	5.000000	5.392857	5.642857	5.428571
ไม่ต้องการระบุ	5.000000	4.000000	4.000000	4.000000	6.000000	4.000000	4.000000	4.000000

3 rows × 61 columns

## Find Frequency of word

ใช้ pythai nlp ในการตัดsentence และ ตัด word จาก free type columns  
'นอกเหนือจากรายการดังกล่าวแล้ว คุณมีความสนใจในเรื่องใดอีกบ้าง'

```
In [79]: #temp_survey.to_excel('lll.xlsx')
import matplotlib.pyplot as plt
from pythainlp.tokenize import sent_tokenize
from pythainlp.tokenize import word_tokenize
# from pythainlp.tokenize import sent_tokenize
location = 'C:\\Users\\rangsarid.p\\Desktop\\python_code\\crm\\'
#temp_survey3 = pd.read_excel(f'{location}lll.xlsx')
temp_survey3 = temp_survey.copy()
lis_word = temp_survey3['นอกเหนือจากรายการดังกล่าวแล้ว คุณมีความสนใจในเรื่องใดอีกบ้าง'].to_list()
oo = ' '.join(lis_word)
ss = sent_tokenize(oo, engine="whitespace")

stop_word = [' ', ',', '-', '/', '(', ')', '\n', 'และ', 'เกี่ยวกับ', 'ด้าน']
oo2 = [i for i in oo if i not in stop_word]
ss
hh = []
dict_word = {}
for g in ss :
    dd= word_tokenize(g, engine="newmm")
    count=0
    for ddd in dd:
        if ddd not in stop_word :
            if ddd.lower() not in dict_word:
                dict_word[ddd.lower()] = 1
            else :
                dict_word[ddd.lower()] += 1
sorted_values = dict(sorted(dict_word.items(),key= lambda item : item[1], reverse=True))
wordw = list(sorted_values.keys())
count = list(sorted_values.values())
dcit_w = {'word': wordw , 'count':count}
w_c = pd.DataFrame(dcit_w)
w_c
#dict(sorted(x.items(), key=lambda item: item[1]))
#print(sorted_values)
#tt_word = pd.DataFrame(sorted_values, columns = ['name', 'count'])
#tt_word #= tt_word.T
#sorted_values.plt('bar')
#for i , j in sorted_values.items():
#    plt.barh(i,j)

#plt.tight_layout(pad=0.7)
#plt.show()
```

Out[79]:

	word	count
0	สัตว์เลี้ยง	10
1	กีฬา	5
2	รถยนต์	5
3	เทคโนโลยี	4

	word	count
4	ฟุตบอล	4
...	...	...
140	podcast	1
141	พื้ม	1
142	นา	1
143	ตนเอง	1
144	ซีรีส์	1

145 rows × 2 columns

## Find top 10 the frequency of word

พบว่าคนส่วนใหญ่สนใจในเรื่อง สัตว์เลี้ยงมาเป็นอันดับแรกที่ 10 คน และ กีฬาและรถยนต์มาเป็นอันดับ 2 ตามลำดับ

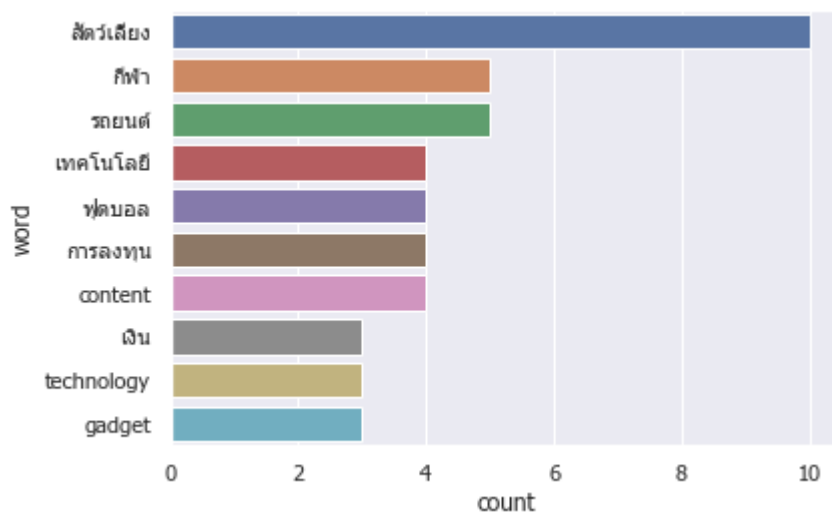
```
In [80]: w_c_10 = w_c.head(10)
display(w_c_10)
```

	word	count
0	สัตว์เลี้ยง	10
1	กีฬา	5
2	รถยนต์	5
3	เทคโนโลยี	4
4	ฟุตบอล	4
5	การลงทุน	4
6	content	4
7	เงิน	3
8	technology	3
9	gadget	3

## Plot bar Graph 10 Tops list

```
In [81]: #sb.set_theme(style="whitegrid")
sb.barplot(y= 'word', x = 'count' ,data = w_c_10)
```

```
Out[81]: <AxesSubplot:xlabel='count', ylabel='word'>
```



```
In [82]: ## word cloud
# import matplotlib.pyplot as plt
# from wordcloud import WordCloud
# text = "แมว กิน ปลา แมว มัน ชอบ นอน นอน กลางวัน นอน แล้ว นอน อีก เป็น สัตว์ ที่ ชี้ เกี้ยว จริงๆ เล
# wordcloud = WordCloud(font_path='TH Sarabun New.ttf',
#                         relative_scaling = 1.0,
#                         min_font_size=4,
#                         background_color="white",
#                         width=1024,
#                         height=768,
#                         scale=3,
#                         font_step=1,
#                         collocations=False,
#                         regexp=r"[\u0E00-\u0E7Fa-zA-Z']+",
#                         margin=2
#                         ).generate(text)
# plt.imshow(wordcloud, cmap=plt.cm.gray, interpolation='bilinear')
# plt.axis("off")
# plt.show()
```

## Find similarity by euclidean distance

เป็นการหาความเป็น similarity ของ person โดยใช้คะแนนจากผลการประเมินในแต่ละคนมาหา ความสัมพันธ์ by euclidean distance

```
In [24]: distancew = pd.read_excel('./l11.xlsx')
```

```
In [25]: distancew.shape
```

```
Out[25]: (63, 70)
```

```
In [60]: ##rename
from scipy.spatial import distance
import numpy as np
import pandas as pd
location = 'C:\\Users\\rangsarid.p\\Desktop\\python_code\\crm\\'
distancew = pd.read_excel(f'{location}l11.xlsx')
distancew = distancew.rename(columns = {'Unnamed: 0': 'ID'}) ##2,61
distancew.iloc[1,2:62]
i_index = []
```

```

j_index = []
sim_list = []
p1_sex = []
p2_sex=[]
age1_sex=[]
age2_sex=[]
month1 = []
month2 = []
for i in range(63):
    for j in range(63) :
        i_index.append(i)
        j_index.append(j)
        sim = distance.euclidean(distancew.iloc[i,2:62],distancew.iloc[j,2:62])
        sim_list.append(round(sim,4))
        p1_sex.append(distancew.iloc[i,64])
        p2_sex.append(distancew.iloc[j,64])
        age1_sex.append(distancew.iloc[i,69])
        age2_sex.append(distancew.iloc[j,69])
        month1.append(distancew.iloc[i,67])
        month2.append(distancew.iloc[j,67])
data = {'similarity-distance' : sim_list,'first' : i_index,'seconde' : j_index , "Sexu
ddd = pd.DataFrame(data)
ddd =ddd.sort_values('similarity-distance')
ddd = ddd[ddd['similarity-distance'] > 0]
#ddd['similarity-distance']= ddd['similarity-distance'].apply(lambda x : str(x))
display(ddd.head(40))
#ddd.drop_duplicates('similarity-distance')

```

	similarity-distance	first	seconde	Sexual_1	Sexual_2	Age1	Age2	Month1	Month2
567	7.6811	9	0	ชาย	ชาย	36	33	11	9
9	7.6811	0	9	ชาย	ชาย	33	36	9	11
1168	8.0623	18	34	หญิง	หญิง	27	33	11	10
2160	8.0623	34	18	หญิง	หญิง	33	27	10	11
1351	8.1854	21	28	ชาย	ชาย	43	25	9	3
1785	8.1854	28	21	ชาย	ชาย	25	43	3	9
1991	8.2462	31	38	หญิง	หญิง	26	29	4	12
2425	8.2462	38	31	หญิง	หญิง	29	26	12	4
2929	8.3066	46	31	หญิง	หญิง	36	26	11	4
1999	8.3066	31	46	หญิง	หญิง	26	36	4	11
1761	8.5440	27	60	หญิง	หญิง	30	27	7	5
3807	8.5440	60	27	หญิง	หญิง	27	30	5	7
2141	8.6023	33	62	ชาย	ไม่ต้องการระบุ	30	30	3	3
3939	8.6023	62	33	ไม่ต้องการระบุ	ชาย	30	30	3	3
1751	8.7178	27	50	หญิง	หญิง	30	36	7	5
3177	8.7178	50	27	หญิง	หญิง	36	30	5	7
1710	8.8318	27	9	หญิง	ชาย	30	36	7	11
594	8.8318	9	27	ชาย	หญิง	36	30	11	7

	similarity-distance	first	seconde	Sexual_1	Sexual_2	Age1	Age2	Month1	Month2
3247	8.9443	51	34	หญิง	หญิง	33	33	2	10
2193	8.9443	34	51	หญิง	หญิง	33	33	10	2
539	9.0554	8	35	ชาย	ชาย	28	29	12	11
1194	9.0554	18	60	หญิง	หญิง	27	27	11	5
490	9.0554	7	49	ชาย	ชาย	28	30	4	10
3798	9.0554	60	18	หญิง	หญิง	27	27	5	11
2003	9.0554	31	50	หญิง	หญิง	26	36	4	5
3094	9.0554	49	7	ชาย	ชาย	30	28	10	4
2213	9.0554	35	8	ชาย	ชาย	29	28	11	12
3181	9.0554	50	31	หญิง	หญิง	36	26	5	4
1385	9.1104	21	62	ชาย	ไม่ต้องการระบุ	43	30	9	3
3927	9.1104	62	21	ไม่ต้องการระบุ	ชาย	30	43	3	9
1954	9.2195	31	1	หญิง	หญิง	26	28	4	8
94	9.2195	1	31	หญิง	หญิง	28	26	8	4
1563	9.2736	24	51	หญิง	หญิง	27	33	11	2
3237	9.2736	51	24	หญิง	หญิง	33	27	2	11
994	9.3274	15	49	หญิง	ชาย	29	30	3	10
3102	9.3274	49	15	ชาย	หญิง	30	29	10	3
598	9.3808	9	31	ชาย	หญิง	36	26	11	4
2664	9.3808	42	18	ชาย	หญิง	44	27	5	11
1962	9.3808	31	9	หญิง	ชาย	26	36	4	11
1176	9.3808	18	42	หญิง	ชาย	27	44	11	5

จาก result พบว่า คนที่มีเพศเดียวกันและอายุใกล้เคียงกันจะมีความเหมือนกัน ที่มีค่า euclidean distance ที่ต่ำๆ

## Pearson Correlation

### Overall Interesting

Correlation ของสิ่งที่มีความสัมพันธ์กันสูงจะสังเกตได้ว่าทั้งความสนใจและการบริโภคจะเป็นไปทางเดียวกันกับ กลุ่มๆเดียวกัน เช่น ในกลุ่ม สนใจ/บริโภคบุฟเฟต์ บ้าง่าง อาหารญี่ปุ่นก็จะมีความสัมพันธ์กันสูง

```
In [88]: ### interesting
temp_survey.columns
```

```
Out[88]: Index(['Timestamp', 'สนใจ[อาหารญี่ปุ่น]', 'สนใจ[อาหารจีน]', 'สนใจ[อาหารไทย]',
              'สนใจ[อาหารอีสาน]', 'สนใจ[อาหารสุขภาพ]', 'สนใจ[บุฟเฟต์]',
              'สนใจ[ชาบูชาวน]', 'สนใจ[บึงย่าง]', 'สนใจ[เหล้าเบียร์]',
```



```
'สนใจ[เสื้อผ้าแฟชั่น]', 'สนใจ[แต่งหน้า]', 'สนใจ[skincare บำรุงผิว]',
'สนใจ[การออกกำลังกาย]', 'สนใจ[การเล่นเกมส์]', 'สนใจ[การท่องเที่ยว]',
'สนใจ[การอ่านหนังสือ]', 'สนใจ[การอ่านบทความ Online]',
'สนใจ[การเรียนรู้ e-learning]', 'สนใจ[ดู Series]', 'สนใจ[ดูหนัง]',
'สนใจ[ดู Youtube]', 'สนใจ[ดู Netflix]', 'สนใจ[Content ด้านบันเทิง]',
'สนใจ[Content เกี่ยวกับเกมส์]',
'สนใจ[Content เกี่ยวกับการ review ร้านอาหาร]',
'สนใจ[Content เกี่ยวกับการท่องเที่ยว]',
'สนใจ[Content เกี่ยวกับการแต่งบ้าน]',
'สนใจ[Content เกี่ยวกับการโทรทัศน์]', 'สนใจ[Content เกี่ยวกับละคร]',
'สนใจ[Content เกี่ยวกับการเมือง]', 'บริโภค[อาหารญี่ปุ่น]',
'บริโภค[อาหารจีน]', 'บริโภค[อาหารไทย]', 'บริโภค[อาหารอีสาน]',
'บริโภค[อาหารสุขภาพ]', 'บริโภค[บุฟเฟต์]', 'บริโภค[ชาบูชาบู]',
'บริโภค[ปิ้งย่าง]', 'บริโภค[เหล้าเบียร์]', 'บริโภค[เสื้อผ้าแฟชั่น]',
'บริโภค[แต่งหน้า]', 'บริโภค[skincare บำรุงผิว]',
'บริโภค[การออกกำลังกาย]', 'บริโภค[การเล่นเกมส์]',
'บริโภค[การท่องเที่ยว]', 'บริโภค[การอ่านหนังสือ]',
'บริโภค[การอ่านบทความ Online]', 'บริโภค[การเรียนรู้ e-learning]',
'บริโภค[ดู Series]', 'บริโภค[ดูหนัง]', 'บริโภค[ดู Youtube]',
'บริโภค[ดู Netflix]', 'บริโภค[Content ด้านบันเทิง]',
'บริโภค[Content เกี่ยวกับเกมส์]',
'บริโภค[Content เกี่ยวกับการ review ร้านอาหาร]',
'บริโภค[Content เกี่ยวกับการท่องเที่ยว]',
'บริโภค[Content เกี่ยวกับการแต่งบ้าน]',
'บริโภค[Content เกี่ยวกับการโทรทัศน์]', 'บริโภค[Content เกี่ยวกับละคร]',
'บริโภค[Content เกี่ยวกับการเมือง]',
'นอกเหนือจากรายการดังกล่าวแล้ว คุณมีความสนใจในเรื่องใดอีกบ้าง',
'คุณเกิดวันที่เท่าไร', 'เพศ', 'คุณเกิดวันที่เท่าไร_s', 'Year',
'Month', 'Day', 'Age'],
dtype='object')
```

```
In [106... interested_cols = [col for col in temp_survey.columns if ('สนใจ' in col) or col == 'Age']
consume_cols = [col for col in temp_survey.columns if ('บริโภค' in col) or col == 'Age']
#print(list(df.columns))
temp_survey_inter = temp_survey[interested_cols]
temp_survey_consume = temp_survey[consume_cols]
```

```
In [107... correlation_pear_inter = temp_survey_inter.corr(method='pearson')
```

```
In [108... correlation_pear_inter
```

```
Out[108...
```

	สนใจ[อาหารญี่ปุ่น]	สนใจ[อาหารจีน]	สนใจ[อาหารไทย]	สนใจ[อาหารอีสาน]	สนใจ[อาหารสุขภาพ]	สนใจ[บุฟเฟต์]	สนใจ[ชาบูชาบู]
สนใจ[อาหารญี่ปุ่น]	1.000000	0.183296	0.145602	0.147661	-0.065076	0.419898	0.565748
สนใจ[อาหารจีน]	0.183296	1.000000	-0.019348	-0.001542	0.184708	-0.046306	0.053874
สนใจ[อาหารไทย]	0.145602	-0.019348	1.000000	0.587620	0.139504	0.028239	0.122320
สนใจ[อาหารอีสาน]	0.147661	-0.001542	0.587620	1.000000	0.272640	-0.021135	0.002153
สนใจ[อาหารสุขภาพ]	-0.065076	0.184708	0.139504	0.272640	1.000000	-0.114337	-0.071215

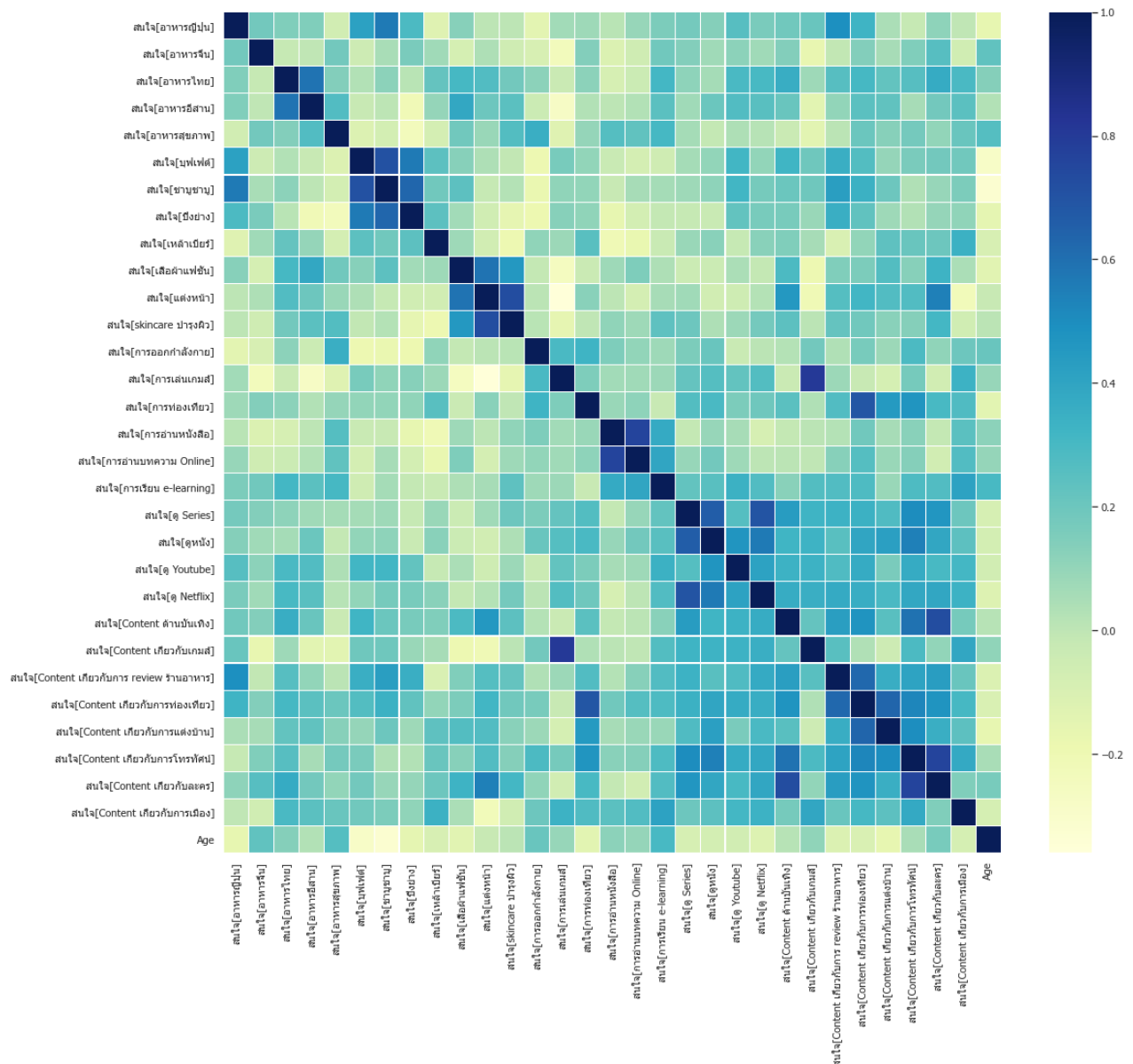
	สนใจ[อาหารญี่ปุ่น]	สนใจ[อาหารจีน]	สนใจ[อาหารไทย]	สนใจ[อาหารอีสาน]	สนใจ[อาหารสุขภาพ]	สนใจ[นุฟเฟด]	สนใจ[ชาบูชาบู]
สนใจ[นุฟเฟด]	0.419898	-0.046306	0.028239	-0.021135	-0.114337	1.000000	0.708651
สนใจ[ชาบูชาบู]	0.565748	0.053874	0.122320	0.002153	-0.071215	0.708651	1.000000
สนใจ[บึงย่าง]	0.291514	0.164691	0.012510	-0.214018	-0.234293	0.569601	0.630016
สนใจ[เหล้าเบียร์]	-0.122948	0.074457	0.223797	0.098865	-0.078235	0.243643	0.187260
สนใจ[เสื้อผ้าแฟชั่น]	0.132005	-0.080016	0.305590	0.388254	0.191345	0.132237	0.241247
สนใจ[แต่งหน้า]	0.001303	0.042797	0.275282	0.202281	0.088967	0.038492	-0.019996
สนใจ[skincare บำรุงผิว]	0.005521	-0.052147	0.182062	0.246808	0.266950	-0.008650	0.027500
สนใจ[การออกกำลังกาย]	-0.143173	-0.089687	0.123245	-0.038560	0.356954	-0.194633	-0.175782
สนใจ[การเล่นเกมส์]	0.069919	-0.235995	-0.034493	-0.272227	-0.125488	0.170261	0.112420
สนใจ[การท่องเที่ยว]	0.079148	0.141182	0.119156	0.032107	0.098817	0.103019	0.106393
สนใจ[การอ่านหนังสือ]	0.016923	-0.106191	-0.086073	0.004239	0.264369	-0.030130	-0.021827
สนใจ[การอ่านบทความ Online]	0.093847	-0.055293	-0.041236	0.028929	0.231391	-0.078027	0.058942
สนใจ[การเรียนรู้ e-learning]	0.160462	0.190193	0.312424	0.247754	0.308598	-0.053961	0.064127
สนใจ[ดู Series]	0.181694	0.141822	0.116650	0.073666	0.061409	0.064635	0.080434
สนใจ[ดูหนัง]	0.138813	0.071990	0.060891	0.203251	-0.014171	0.106652	0.116314
สนใจ[ดู Youtube]	0.259322	0.119347	0.296258	0.272805	0.039122	0.315541	0.318935
สนใจ[ดู Netflix]	0.162962	0.071265	0.301399	0.251159	0.019482	0.104080	0.174847
สนใจ[Content ด้านบันเทิง]	0.189610	0.142805	0.358933	0.213325	-0.031257	0.321266	0.205141
สนใจ[Content เกี่ยวกับเกมส์]	0.211447	-0.164434	0.080494	-0.138965	-0.138591	0.193600	0.197524
สนใจ[Content เกี่ยวกับการ review ร้านอาหาร]	0.492472	-0.008829	0.258793	0.104730	0.009823	0.354044	0.426675

	สนใจ[อาหารญี่ปุ่น]	สนใจ[อาหารจีน]	สนใจ[อาหารไทย]	สนใจ[อาหารอีสาน]	สนใจ[อาหารสุขภาพ]	สนใจ[นุฟเฟด]	สนใจ[ชาบู]
สนใจ[Content เกี่ยวกับการท่องเที่ยว]	0.326259	0.142515	0.307045	0.246205	0.206435	0.266865	0.345573
สนใจ[Content เกี่ยวกับการแต่งงาน]	0.047509	0.055277	0.228765	0.233475	0.132436	0.100076	0.199605
สนใจ[Content เกี่ยวกับการโทรทัศน์]	-0.020065	0.146242	0.258939	0.059229	0.177285	0.164792	0.034024
สนใจ[Content เกี่ยวกับการละคร]	0.113752	0.256343	0.368052	0.183730	0.105656	0.185538	0.092039
สนใจ[Content เกี่ยวกับการเมือง]	-0.002893	-0.061932	0.294486	0.245932	0.210777	0.188379	0.155494
Age	-0.159958	0.230792	0.136509	0.031228	0.266435	-0.283205	-0.318132

31 rows × 31 columns

```
In [109... f, ax = plt.subplots(figsize =(18, 16))
sb.heatmap(correlation_pear_inter, ax = ax, cmap ="YlGnBu", linewidths = 0.1)
```

```
Out[109... <AxesSubplot:>
```



## Correlation of interest split by Male/Female

### Male Interest

In [112]...

# specify by man

```
cor_man = temp_survey_inter[temp_survey_inter['เพศ']=='ชาย'].corr(method='pearson')
cor_man
```

Out[112]...

	สนใจ[อาหารญี่ปุ่น]	สนใจ[อาหารจีน]	สนใจ[อาหารไทย]	สนใจ[อาหารอีสาน]	สนใจ[อาหารสุขภาพ]	สนใจ[บุฟเฟต์]	สนใจ[ชาบู]
สนใจ[อาหารญี่ปุ่น]	1.000000	0.380732	0.157459	1.896269e-01	0.158708	0.494592	0.622619
สนใจ[อาหารจีน]	0.380732	1.000000	-0.067920	-1.200667e-01	0.042025	0.114980	0.192861

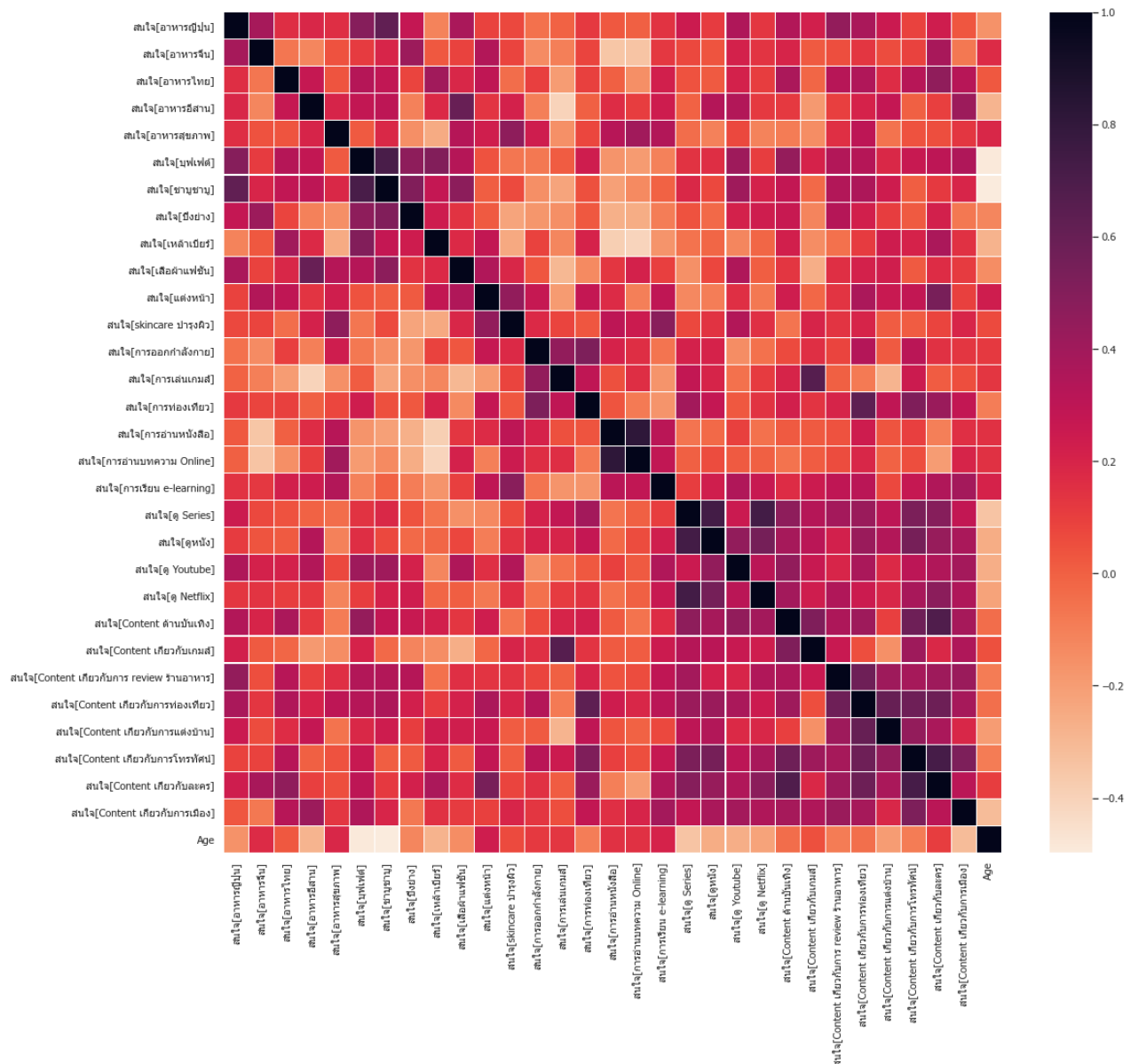
	สนใจ[อาหารญี่ปุ่น]	สนใจ[อาหารจีน]	สนใจ[อาหารไทย]	สนใจ[อาหารอีสาน]	สนใจ[อาหารสุขภาพ]	สนใจ[บุฟเฟต์]	สนใจ[ชาบู]
สนใจ[อาหารไทย]	0.157459	-0.067920	1.000000	2.732872e-01	0.035847	0.327804	0.291628
สนใจ[อาหารอีสาน]	0.189627	-0.120067	0.273287	1.000000e+00	0.200475	0.285578	0.301159
สนใจ[อาหารสุขภาพ]	0.158708	0.042025	0.035847	2.004747e-01	1.000000	0.017470	0.183404
สนใจ[บุฟเฟต์]	0.494592	0.114980	0.327804	2.855784e-01	0.017470	1.000000	0.711382
สนใจ[ชาบู]	0.622619	0.192861	0.291628	3.011591e-01	0.183404	0.711382	1.000000
สนใจ[ปิ้งย่าง]	0.280125	0.417618	0.086203	-1.006460e-01	-0.152283	0.463457	0.511995
สนใจ[เหล้าเบียร์]	-0.109057	0.023500	0.397987	1.787587e-01	-0.247893	0.504125	0.277423
สนใจ[เสื้อผ้าแฟชั่น]	0.371818	0.091441	0.192306	5.978660e-01	0.331520	0.329081	0.470055
สนใจ[แต่งหน้า]	0.089614	0.341789	0.293687	1.345784e-01	0.228492	0.042878	0.007642
สนใจ[skincare บำรุงผิว]	0.072834	0.085038	-0.041434	2.101375e-01	0.466854	-0.066060	0.066546
สนใจ[การออกกำลังกาย]	-0.049823	-0.138751	0.100342	-9.675366e-02	0.237576	-0.075797	-0.148706
สนใจ[การเล่นเกมส์]	-0.002990	-0.094493	-0.198724	-4.006539e-01	-0.157431	0.015407	-0.225660
สนใจ[การท่องเที่ยว]	0.117665	0.085482	0.096943	1.268418e-16	0.077926	0.238192	0.050170
สนใจ[การอ่านหนังสือ]	0.028406	-0.348647	0.003285	1.637455e-01	0.325753	-0.165080	-0.207811
สนใจ[การอ่านบทความ Online]	0.002260	-0.342618	-0.152278	1.088475e-01	0.403143	-0.190773	-0.129774
สนใจ[การเรียนรู้ e-learning]	0.142535	0.125832	0.223967	2.410562e-01	0.344131	-0.102266	-0.006154
สนใจ[ดู Series]	0.250097	0.072932	0.044500	-2.736185e-03	-0.038492	0.143910	0.184736
สนใจ[ดูหนัง]	0.114372	0.035207	0.017036	3.372943e-01	-0.100138	0.162674	0.074388
สนใจ[ดู Youtube]	0.353228	0.210694	0.214438	3.422512e-01	0.073874	0.413291	0.407438
สนใจ[ดู Netflix]	0.128664	0.138233	0.107498	1.182419e-01	-0.111614	0.108051	0.213257

	สนใจ[อาหารญี่ปุ่น]	สนใจ[อาหารจีน]	สนใจ[อาหารไทย]	สนใจ[อาหารอีสาน]	สนใจ[อาหารสุขภาพ]	สนใจ[นุฟเฟด]	สนใจ[ชาบู]
สนใจ[Content ด้านบันเทิง]	0.343853	0.200341	0.367046	1.199931e-01	-0.083360	0.440531	0.283954
สนใจ[Content เกี่ยวกับเกมส์]	0.224738	0.020525	-0.016929	-1.825515e-01	-0.146287	0.207974	-0.027378
สนใจ[Content เกี่ยวกับการ review ร้านอาหาร]	0.449090	0.056339	0.322179	1.040500e-01	0.154079	0.341481	0.343887
สนใจ[Content เกี่ยวกับการท่องเที่ยว]	0.369451	0.129532	0.345856	2.083727e-01	0.299996	0.327534	0.357517
สนใจ[Content เกี่ยวกับการแต่งงาน]	0.253527	0.060911	0.166665	2.770509e-01	-0.058172	0.184177	0.240598
สนใจ[Content เกี่ยวกับการโทรทัศน์]	0.090865	0.092110	0.323441	4.377169e-03	0.045716	0.263402	0.010812
สนใจ[Content เกี่ยวกับละคร]	0.235961	0.379027	0.456864	1.005909e-01	0.055251	0.303393	0.122111
สนใจ[Content เกี่ยวกับการเมือง]	0.032278	-0.071465	0.322499	4.178766e-01	0.133350	0.353010	0.196609
Age	-0.158955	0.174396	0.028509	-2.822230e-01	0.189060	-0.491919	-0.497942

31 rows × 31 columns

```
In [113...  ## Drawn graph
f, ax = plt.subplots(figsize =(18, 16))
sb.heatmap(cor_man, ax = ax, cmap ="rocket_r", linewidths = 0.1)
```

```
Out[113... <AxesSubplot:>
```



## conclusion

พบว่า ความสัมพันธ์โดยส่วนใหญ่เป็นไปตามเทรนของภาพรวม เช่น ความสนใจกลุ่มอาหารญี่ปุ่น ความสนใจอาหารปิ้งย่าง จะมีความสัมพันธ์กันในทางบวก แต่มี บางที่ค่อนข้างมีความสัมพันธ์เชิงลบเช่นสนใจอาหารอีสานกับการสนใจเกมส์ เป็นต้น

## Female

```
In [114...] cor_woman = temp_survey_inter[temp_survey_inter['เพศ']=='หญิง'].corr(method='pearson')
cor_woman
```

Out[114...]

	สนใจ[อาหารญี่ปุ่น]	สนใจ[อาหารจีน]	สนใจ[อาหารไทย]	สนใจ[อาหารอีสาน]	สนใจ[อาหารสุขภาพ]	สนใจ[นุฟเฟ็ต]	สนใจ[ชาบู]
สนใจ[อาหารญี่ปุ่น]	1.000000	-0.060017	0.099929	0.081772	-2.893822e-01	0.321728	0.467467



	สนใจ[อาหารญี่ปุ่น]	สนใจ[อาหารจีน]	สนใจ[อาหารไทย]	สนใจ[อาหารอีสาน]	สนใจ[อาหารสุขภาพ]	สนใจ[นุฟเฟด]	สนใจ[ชาบูชาบู]
สนใจ[อาหารจีน]	-0.060017	1.000000	-0.013822	0.095626	3.722514e-01	-0.231462	-0.136992
สนใจ[อาหารไทย]	0.099929	-0.013822	1.000000	0.785821	2.599183e-01	-0.200399	-0.057875
สนใจ[อาหารอีสาน]	0.081772	0.095626	0.785821	1.000000	3.754471e-01	-0.317031	-0.348730
สนใจ[อาหารสุขภาพ]	-0.289382	0.372251	0.259918	0.375447	1.000000e+00	-0.221040	-0.326814
สนใจ[นุฟเฟด]	0.321728	-0.231462	-0.200399	-0.317031	-2.210396e-01	1.000000	0.696589
สนใจ[ชาบูชาบู]	0.467467	-0.136992	-0.057875	-0.348730	-3.268142e-01	0.696589	1.000000
สนใจ[บึงย่าง]	0.243181	-0.193448	-0.124935	-0.417267	-3.190656e-01	0.702514	0.769199
สนใจ[เหล้าเบียร์]	-0.240700	0.206600	0.089356	0.017172	2.172537e-01	-0.116624	-0.006242
สนใจ[เสื้อผ้าแฟชั่น]	-0.121125	-0.389916	0.395738	0.166821	7.640014e-02	-0.061418	0.003567
สนใจ[แต่งหน้า]	-0.005373	-0.480523	0.449146	0.323573	-1.820194e-16	0.096438	-0.005601
สนใจ[skincare บำรุงผิว]	-0.020126	-0.397647	0.427972	0.309980	8.321989e-02	0.078288	0.038853
สนใจ[การออกกำลังกาย]	-0.261443	0.009991	0.195230	0.063115	5.159578e-01	-0.324427	-0.201888
สนใจ[การเล่นเกม]	0.123245	-0.367380	0.100634	-0.160665	-1.216124e-01	0.347078	0.473692
สนใจ[การท่องเที่ยว]	0.035462	0.212605	0.148892	0.065195	1.268949e-01	-0.043409	0.189582
สนใจ[การอ่านหนังสือ]	-0.004632	0.213782	-0.172799	-0.177107	1.939058e-01	0.112028	0.209225
สนใจ[การอ่านบทความ Online]	0.147870	0.362000	-0.050081	-0.152448	5.395000e-02	-0.023992	0.231730
สนใจ[การเรียนรู้ e-learning]	0.162291	0.302914	0.426087	0.273355	2.875987e-01	-0.020660	0.131588
สนใจ[ดู Series]	0.088286	0.212749	0.160472	0.136339	2.021871e-01	-0.035447	-0.073157
สนใจ[ดูหนัง]	0.145311	0.097019	0.064963	0.064205	9.088562e-02	0.038086	0.133662
สนใจ[ดู Youtube]	0.126232	0.030566	0.362814	0.217217	2.348835e-02	0.208910	0.193005

	สนใจ[อาหารญี่ปุ่น]	สนใจ[อาหารจีน]	สนใจ[อาหารไทย]	สนใจ[อาหารอีสาน]	สนใจ[อาหารสุขภาพ]	สนใจ[บุฟเฟต์]	สนใจ[ชาบู]
สนใจ[ดู Netflix]	0.176864	-0.024133	0.470759	0.401909	2.158000e-01	0.075668	0.085639
สนใจ[Content ด้านบันเทิง]	-0.012854	-0.025642	0.419007	0.318368	6.522840e-02	0.222393	0.111663
สนใจ[Content เกี่ยวกับเกมส์]	0.174405	-0.350402	0.162049	-0.086239	-1.278896e-01	0.179918	0.428410
สนใจ[Content เกี่ยวกับการ review ร้านอาหาร]	0.609870	-0.172312	0.207970	0.062557	-1.948159e-01	0.398196	0.591208
สนใจ[Content เกี่ยวกับการท่องเที่ยว]	0.274366	0.114543	0.261465	0.248723	1.320311e-01	0.192708	0.320540
สนใจ[Content เกี่ยวกับการแต่งบ้าน]	-0.187331	0.010685	0.272516	0.166487	3.678393e-01	0.013782	0.161922
สนใจ[Content เกี่ยวกับการโทรศัพท์]	-0.172840	0.186439	0.221048	0.087889	3.877538e-01	0.056959	0.070284
สนใจ[Content เกี่ยวกับละคร]	-0.038148	0.013967	0.356242	0.264693	2.289771e-01	0.061880	0.062004
สนใจ[Content เกี่ยวกับการเมือง]	-0.191057	-0.032210	0.307096	0.165613	4.420792e-01	0.003151	-0.001720
Age	-0.193125	0.325240	0.232693	0.388388	3.834562e-01	-0.082453	-0.134015

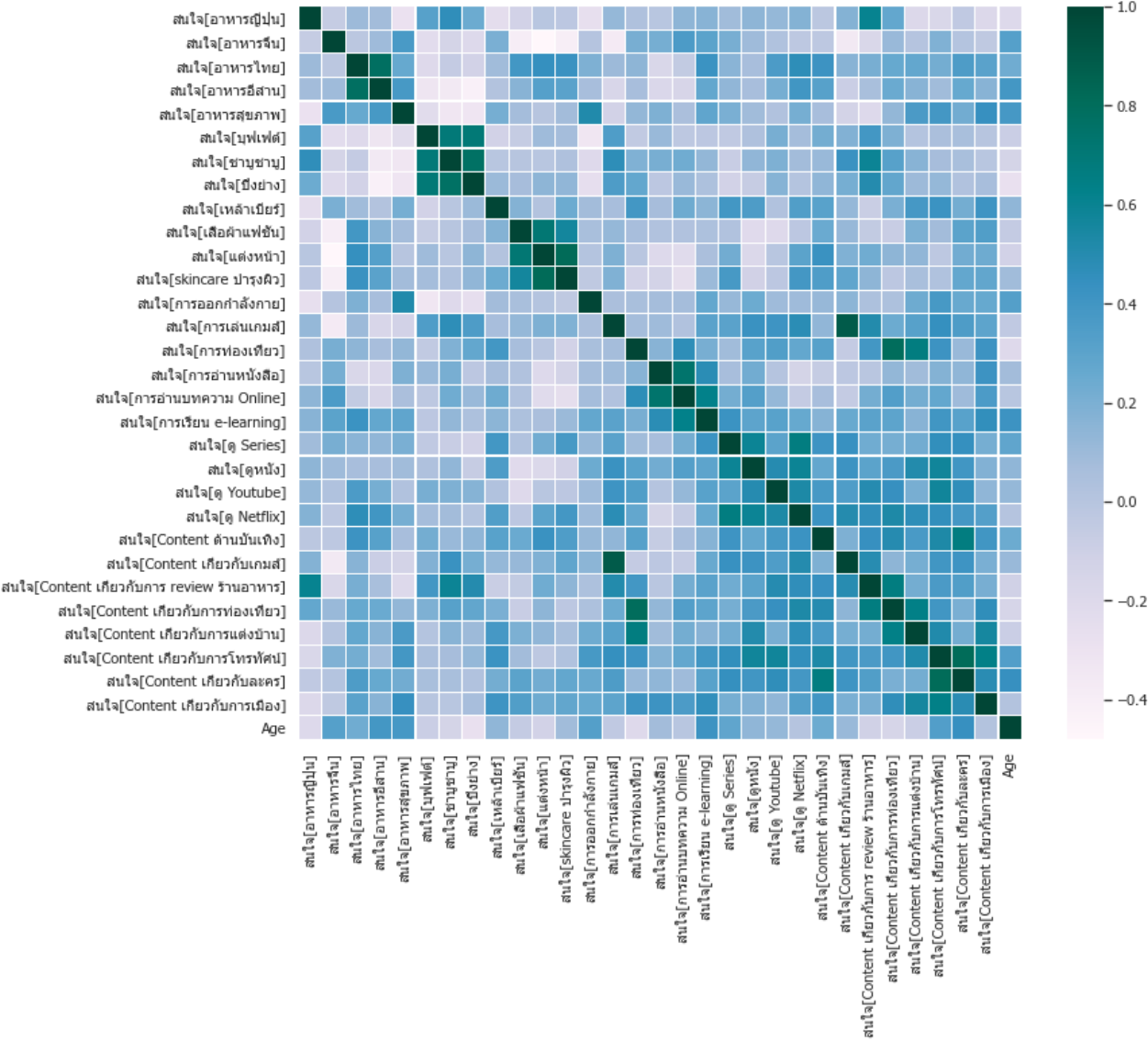
31 rows × 31 columns

In [234...

```
## Drawn graph
f, ax = plt.subplots(figsize =(12, 10))
sb.heatmap(cor_woman, ax = ax, cmap ="PuBuGn", linewidths = 0.1)
```

Out[234...

<AxesSubplot:>



conclusion

พบว่า ความสัมพันธ์โดยส่วนใหญ่เป็นไปตามเทรนของภาพรวม เช่น ความสนใจกลุ่มอาหารญี่ปุ่น ความสนใจอาหารปิ้งย่าง หรือ สนในดู Netflix กับ สนใจดูหนัง จะมีความสัมพันธ์กันในทางบวก แต่มี บางที่ค่อนข้างมีความสัมพันธ์เชิงลบเช่นสนใจการแต่งหน้าแต่ก็ไม่สนใจอาหารจีน

Correlation with consume of overall

```
In [120...] cor_consume = temp_survey_consume.corr(method='pearson')
cor_consume
```

Out[120...]

	บริโภค[อาหารญี่ปุ่น]	บริโภค[อาหารจีน]	บริโภค[อาหารไทย]	บริโภค[อาหารอีสาน]	บริโภค[อาหารสุขภาพ]	บริโภค[บุฟเฟ่ต์]	บ
บริโภค[อาหารญี่ปุ่น]	1.000000	0.201019	-0.008656	0.255596	0.211347	0.452719	
บริโภค[อาหารจีน]	0.201019	1.000000	-0.212573	-0.071733	0.090394	-0.025623	

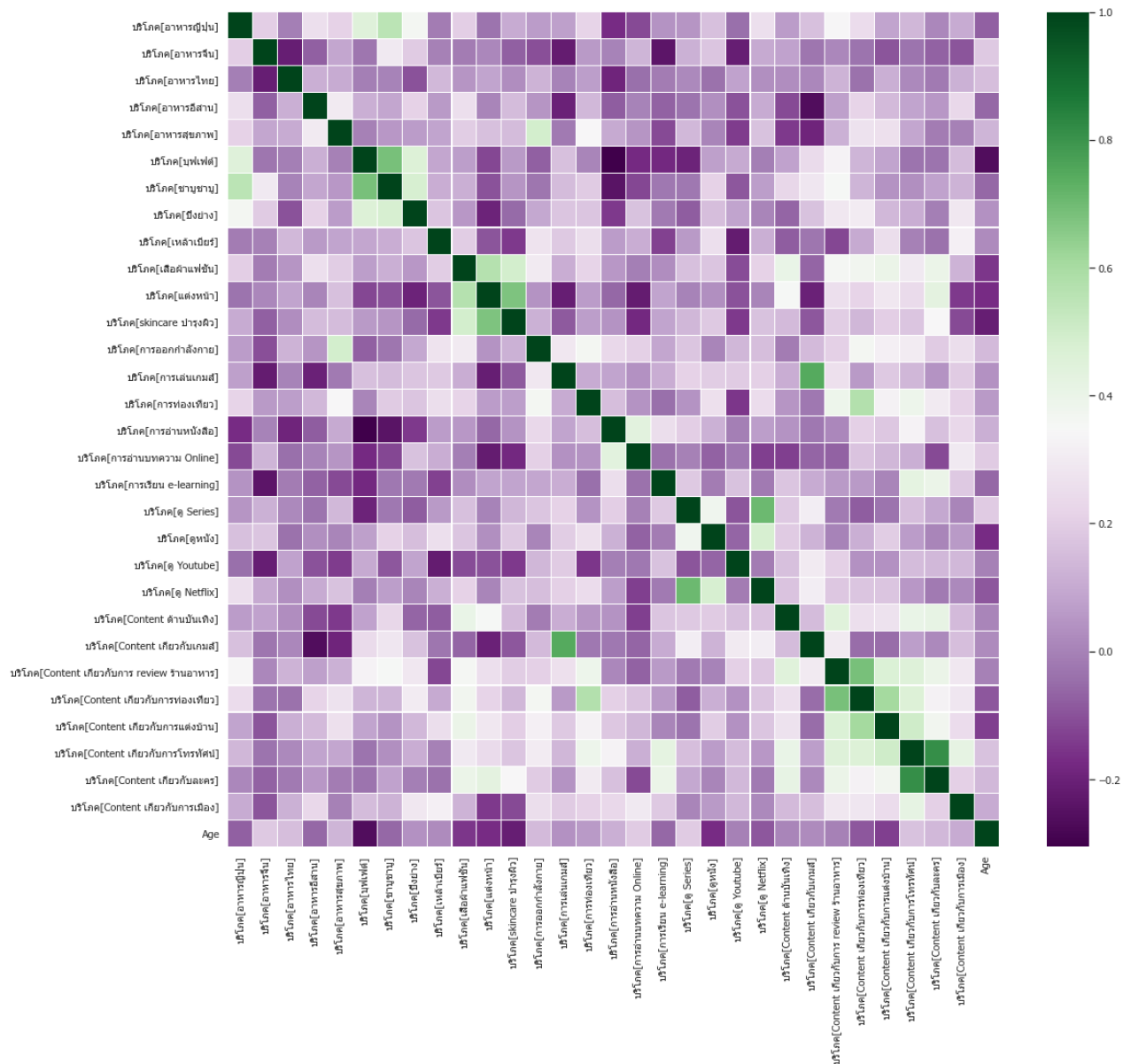
	บริโภค[อาหารญี่ปุ่น]	บริโภค[อาหารจีน]	บริโภค[อาหารไทย]	บริโภค[อาหารอีสาน]	บริโภค[อาหารสุขภาพ]	บริโภค[นพเฟด]	บ
บริโภค[อาหารไทย]	-0.008656	-0.212573	1.000000	0.115638	0.104190	-0.006981	
บริโภค[อาหารอีสาน]	0.255596	-0.071733	0.115638	1.000000	0.301185	0.117072	
บริโภค[อาหารสุขภาพ]	0.211347	0.090394	0.104190	0.301185	1.000000	-0.004224	
บริโภค[นพเฟด]	0.452719	-0.025623	-0.006981	0.117072	-0.004224	1.000000	
บริโภค[ขานูชา]	0.553200	0.295660	0.003685	0.088747	0.051553	0.684628	
บริโภค[ปิ้งย่าง]	0.363131	0.190619	-0.101217	0.213695	0.064078	0.459750	
บริโภค[เหล้าเบียร์]	-0.012661	-0.000307	0.142392	0.058324	0.086168	0.088721	
บริโภค[เสื้อผ้าแฟชั่น]	0.202090	-0.014136	0.047480	0.256849	0.196783	0.069644	
บริโภค[แต่งหน้า]	-0.031336	0.011682	0.083184	0.005878	0.161871	-0.119363	.
บริโภค[skincare บำรุงผิว]	0.102666	-0.073968	0.029978	0.152694	0.149077	0.058530	
บริโภค[การออกกำลังกาย]	0.066206	-0.104944	0.124647	0.046908	0.490575	-0.075917	.
บริโภค[การเล่นเกม]	0.074254	-0.215720	0.003172	-0.200812	-0.017427	0.160547	
บริโภค[การท่องเที่ยว]	0.216578	0.060478	0.071637	0.141576	0.346806	0.008786	
บริโภค[การอ่านหนังสือ]	-0.171102	0.004862	-0.189560	-0.077279	0.100062	-0.305777	.
บริโภค[การอ่านบทความ Online]	-0.113055	0.132516	-0.044027	0.000383	0.047519	-0.175798	.
บริโภค[การเรียนรู้ e-learning]	0.045123	-0.238541	-0.013361	-0.070282	-0.112677	-0.178289	.
บริโภค[ดู Series]	0.047166	0.120069	0.016668	-0.031710	0.142121	-0.196997	.
บริโภค[ดูหนัง]	0.160551	0.174797	-0.037654	0.031255	0.006863	0.073693	
บริโภค[ดู Youtube]	-0.045509	-0.210392	0.077145	-0.096891	-0.137625	0.097369	.
บริโภค[ดู Netflix]	0.258592	0.101328	0.032100	0.107015	0.163558	-0.007337	

	บริโภค[อาหารญี่ปุ่น]	บริโภค[อาหารจีน]	บริโภค[อาหารไทย]	บริโภค[อาหารอีสาน]	บริโภค[อาหารสุขภาพ]	บริโภค[นพเฟด]	บ
บริโภค[Content ด่านบันเทิง]	0.061995	0.079898	0.025961	-0.113162	-0.143664	0.077842	
บริโภค[Content เกี่ยวกับเกมส์]	0.160320	-0.028166	0.014210	-0.262201	-0.188840	0.242414	
บริโภค[Content เกี่ยวกับการ review ร้านอาหาร]	0.339288	0.011285	0.124639	0.073360	0.119070	0.327226	
บริโภค[Content เกี่ยวกับการท่องเที่ยว]	0.248484	-0.034153	-0.039817	0.216912	0.266012	0.128419	
บริโภค[Content เกี่ยวกับการแต่งบ้าน]	0.084140	-0.096467	0.113282	0.143191	0.263639	0.082857	
บริโภค[Content เกี่ยวกับการโทรศัพท์]	0.136615	-0.034750	0.018119	0.082662	0.089254	-0.033132	
บริโภค[Content เกี่ยวกับละคร]	0.011370	-0.070747	-0.019274	0.054111	-0.014929	-0.030946	
บริโภค[Content เกี่ยวกับการเมือง]	0.102948	-0.093133	0.119378	0.230956	-0.001469	0.149979	
Age	-0.073189	0.186531	0.154396	-0.058663	0.132620	-0.262057	

31 rows × 31 columns

```
In [128...  ## Drawn graph
f, ax = plt.subplots(figsize =(18, 16))
sb.heatmap(cor_consume, ax = ax, cmap ="PRGn", linewidths = 0.1)
```

```
Out[128... <AxesSubplot:>
```

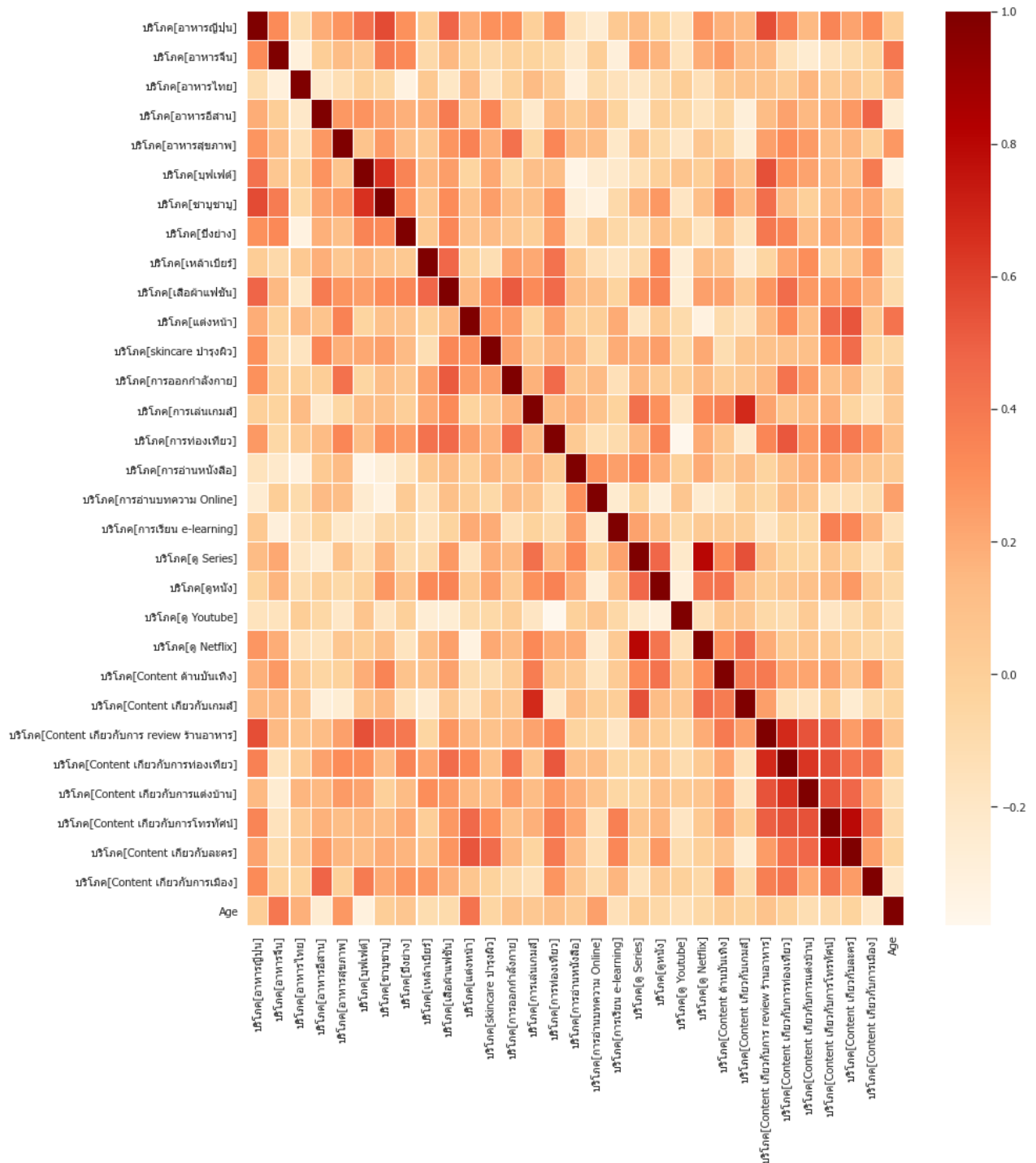


## Correlation with consume of male/female

### Male

```
In [233... #temp_survey_consume
cor_man_consume = temp_survey_consume[temp_survey_consume['เพศ']=='ชาย'].corr(method='p
f, ax = plt.subplots(figsize=(14, 16))
sb.heatmap(cor_man_consume, ax = ax, cmap="OrRd", linewidths = 0.1)
```

```
Out[233... <AxesSubplot:~>
```



## conclusion

พบว่า ความสัมพันธ์โดยส่วนใหญ่ของผู้ชายเป็นไปตามเทรนของภาพรวม เช่น การบริโภคกลุ่มชาบู การบริโภคอาหารปิ้งย่าง หรือ การบริโภคดู Netflix กับ การบริโภคหนัง หรือจะเป็นพวกในกลุ่ม บริโภค content แต่งบ้าน กับ conten ทางโทรทัศน์จะมีความสัมพันธ์กันในทางบวก แต่จะมีบาง factor ที่มีความสัมพันธ์เชิงลบ เช่นการดู youtube กับ การท่องเที่ยว

```
In [139... cor_woman_consume = temp_survey_consume[temp_survey_consume['เพศ']=='หญิง'].corr(method='spearmanr', axis=1)
cor_woman_consume
```

Out[139...]



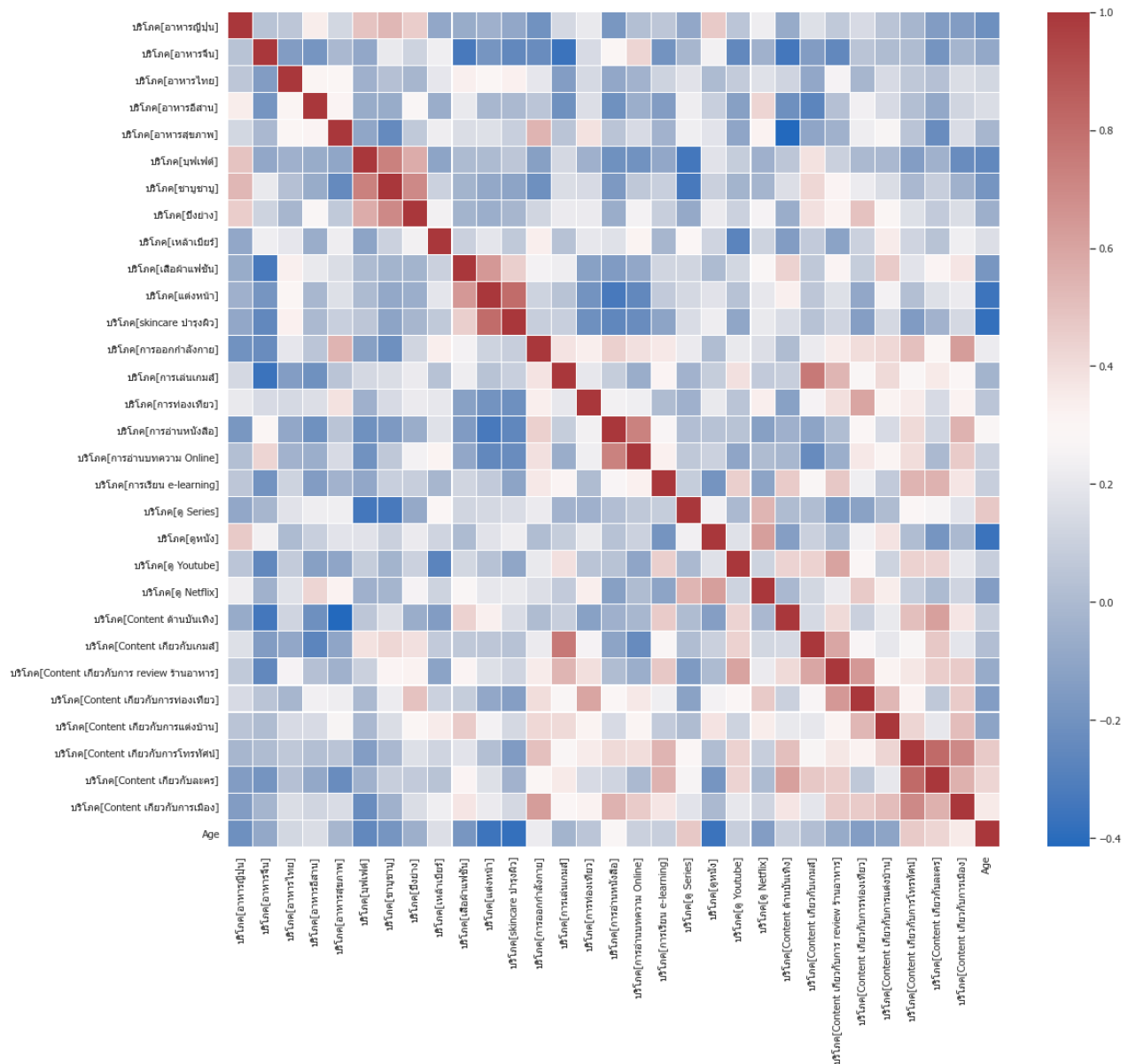
	บริโภค[อาหารญี่ปุ่น]	บริโภค[อาหารจีน]	บริโภค[อาหารไทย]	บริโภค[อาหารอีสาน]	บริโภค[อาหารสุขภาพ]	บริโภค[นพเฟด]	บ
บริโภค[อาหารญี่ปุ่น]	1.000000	0.044120	0.055298	0.342548	0.129312	0.488899	
บริโภค[อาหารจีน]	0.044120	1.000000	-0.145191	-0.187549	-0.009555	-0.093744	
บริโภค[อาหารไทย]	0.055298	-0.145191	1.000000	0.307937	0.298203	-0.047585	
บริโภค[อาหารอีสาน]	0.342548	-0.187549	0.307937	1.000000	0.313728	-0.071029	
บริโภค[อาหารสุขภาพ]	0.129312	-0.009555	0.298203	0.313728	1.000000	-0.096500	
บริโภค[นพเฟด]	0.488899	-0.093744	-0.047585	-0.071029	-0.096500	1.000000	
บริโภค[ชาบูชาบู]	0.530674	0.206613	0.030954	-0.067947	-0.240936	0.731549	
บริโภค[ปิ้งย่าง]	0.457332	0.110581	-0.018344	0.279290	0.066960	0.570713	
บริโภค[เหล้าเบียร์]	-0.089896	0.230736	0.196274	-0.063690	0.223893	-0.109829	
บริโภค[เสื้อผ้าแฟชั่น]	-0.067043	-0.329491	0.333489	0.205196	0.156813	0.002849	
บริโภค[แต่งหน้า]	-0.048155	-0.185935	0.295248	-0.014449	0.169317	0.005417	
บริโภค[skincareบำรุงผิว]	-0.089558	-0.249491	0.328924	-0.005251	0.094304	0.057087	
บริโภค[การออกกำลังกาย]	-0.202842	-0.230806	0.196080	0.054650	0.542008	-0.132028	
บริโภค[การเล่นเกมส์]	0.134946	-0.364184	-0.138482	-0.206708	0.048373	0.137184	
บริโภค[การท่องเที่ยว]	0.205965	0.148567	0.142450	0.162600	0.382340	-0.046888	
บริโภค[การอ่านหนังสือ]	-0.174070	0.301892	-0.092679	-0.205901	0.049172	-0.209853	
บริโภค[การอ่านบทความ Online]	0.025072	0.430430	-0.038081	-0.056843	0.146609	-0.202436	
บริโภค[การเรียนรู้ e-learning]	0.057592	-0.198712	0.113443	-0.144851	-0.034300	-0.090247	
บริโภค[ดู Series]	-0.090707	-0.018320	0.178757	0.223640	0.228663	-0.338819	
บริโภค[ดูหนัง]	0.469266	0.242819	0.001104	0.093970	0.186461	0.176754	

	บริโภค[อาหารญี่ปุ่น]	บริโภค[อาหารจีน]	บริโภค[อาหารไทย]	บริโภค[อาหารอีสาน]	บริโภค[อาหารสุขภาพ]	บริโภค[นพเฟด]	บ
บริโภค[ดู Youtube]	0.053281	-0.252886	0.081937	-0.131364	-0.100390	0.084906	.
บริโภค[ดู Netflix]	0.225795	-0.039191	0.172814	0.425925	0.325496	-0.049475	.
บริโภค[Content ด่านบันเทิง]	-0.076853	-0.341487	0.119283	-0.210719	-0.414443	0.057561	.
บริโภค[Content เกี่ยวกับเกมส์]	0.170487	-0.144398	-0.090839	-0.266511	-0.103622	0.382107	.
บริโภค[Content เกี่ยวกับการ review ร้านอาหาร]	0.068821	-0.252390	0.251694	0.030795	-0.042361	0.101589	.
บริโภค[Content เกี่ยวกับการท่องเที่ยว]	0.144268	0.044827	-0.020016	0.234300	0.221114	-0.001179	.
บริโภค[Content เกี่ยวกับการแต่งบ้าน]	0.046891	0.009471	0.146892	0.144846	0.259975	-0.008967	.
บริโภค[Content เกี่ยวกับการโทรทัศน์]	-0.018789	0.009487	0.065995	0.023962	0.046797	-0.141727	.
บริโภค[Content เกี่ยวกับละคร]	-0.144089	-0.192957	0.029438	-0.086550	-0.229972	-0.040931	.
บริโภค[Content เกี่ยวกับการเมือง]	-0.147036	-0.025274	0.156803	0.122347	0.150411	-0.227332	.
Age	-0.213251	-0.087191	0.130349	0.157423	-0.016897	-0.249318	.

31 rows × 31 columns

```
In [227... cor_woman_consume
f, ax = plt.subplots(figsize =(18, 16))
sb.heatmap(cor_woman_consume, ax = ax, cmap ="vlag", linewidths = 0.1)
```

Out[227... <AxesSubplot:>



## conclusion

พบว่า ความสัมพันธ์การบริโภคโดยส่วนใหญ่ของผู้หญิงเป็นไปตามเทรนของภาพรวม เช่น การแต่งหน้า การบริโภคครีมบำรุงผิว การเลือกแฟชั่น หรือ จะมีความสัมพันธ์กันในทางบวก แต่จะมีบาง factor ที่มีความสัมพันธ์กันในเชิงลบ เช่น อายุกับการบริโภค skin care or ผลิตภัณฑ์แต่งหน้า