

นาย กิตติภณ สรุ่งเรืองสกุล 61070278

Assignment 1

```
In [100]: #!pip install urllib requests  
#ລັງ urllib ແລະ requests ກ່ອນດ້າວື່ນໃນນີ້  
import json, urllib  
import pandas as pd  
import numpy as np
```

ดัชนีราคาผู้บริโภคทั่วไป (Consumer Price Index)

ວິເຄາະທີ່ດັชนີ່ราคาຜູ້ບໍ່ຮົມທຳກຳທີ່ມີສະແດງໃນກຸງເຖິງແຫຼ່ງທີ່ມີສະແດງໃນກຸງເຖິງພິຈາລະນາ ຕັ້ງແຕ່ປີ.ສ. 2017-2020 ແລະ ໃນຍຸດໂຄວິດ-19

ແທລັງຂ່ອງນຸ້ລຸ: <https://data.moc.go.th/OpenData/CPIIndexes>

Query Parameters

Name	Data type	Description	Default
region_id	Integer	ລັດສຸກມີກັດ (0 = ກຽມເທິງແລະປ່ຽນແຫຼ່ງ, 1 = ກາດກລາງ, 2 = ກາດຕະວັນອອກເຊີຍເຫັນ, 3 = ກາດເກີນອືອ, 4 = ກາດໄດ້, 5 = ປະເທດ)	-
index_id	String	ຮັດສົມວັດ / ກລຸມຕົ້ນ	000000000000000000
from_year	Integer	ປີທີ່ຈົດທ່າດ້ານນີ້ເລີ່ມຕົ້ນ (ປ.ສ.)	-
to_year	Integer	ປີທີ່ຈົດທ່າດ້ານນີ້ສິ້ນສຸດ (ປ.ສ.)	-

Link API: <https://dataapi.moc.go.th/cpig-indexes>

Example: https://dataapi.moc.go.th/cpig-indexes?region_id=1&index_id=10000000000000000000&from_year=2017&to_year=2020

- គັນທາງຂ່ອງນຸ້ລຸໃນ ກາດກລາງ
- ຮັດສົມວັດທີ່ 10000000000000000000
- ມີທີ່ຈົດທ່າດ້ານນີ້ເລີ່ມຕົ້ນທີ່ປີ 2017
- ມີທີ່ຈົດທ່າດ້ານນີ້ສິ້ນສຸດທີ່ປີ 2020

Responses

Type: Array

- ຜລລັບຮັບຈະມີກາຣເຮັງລໍາດັບໂດຍປີທີ່ຈົດທ່າດ້ານນີ້ ດາມຕ່າຍເຕືອນທີ່ຈົດທ່າດ້ານນີ້ ດາມຕ່າຍຮ້າສຸກມີກັດ

Name	Data type	Description
index_id	String	ຮັດສົມວັດ / ກລຸມຕົ້ນ
index_description	String	ຊື່ອໝາດ / ກລຸມຕົ້ນ
region_id	Integer	ລັດສຸກມີກັດ
region_name	String	ຊື່ອໝາດ
base_year	Integer	ປີສ້າງຕົ້ນ
year	Integer	ປີທີ່ຈົດທ່າດ້ານນີ້
month	Integer	ເດືອນທີ່ຈົດທ່າດ້ານນີ້
price_index	Float	ຕົ້ນຕົ້ນ
mon	Float	ອັດຮາການປ່ສົນແປລັງເທິນກົນຕົວກອນທຳນາ
yoy	Float	ອັດຮາການປ່ສົນແປລັງເທິນກົນເຕືອນເຕີຍກົນນີ້ກອນທຳນາ
aoa	Float	ອັດຮາການປ່ສົນແປລັງເທິນກົນນ່ວງເຕີຍກົນນີ້ກອນທຳນາ

```
[  
  {  
    "index_id": "10000000000000000000",  
    "index_description": "ໜ້າວດອາຫານແລະເຄື່ອງດືມໄມ້ມີແອລກອອສອລ്",  
    "region_id": 1,  
    "region_name": "ກາດກລາງ",  
    "base_year": 2015,  
    "year": 2017,  
    "month": 1,  
    "price_index": 101.53,  
    "mon": 0.4,  
    "yoy": 1.61,  
    "aoa": 1.61  
  },  
]
```

รหัสหมวด/กลุ่มตัวชี้ (Index_id)

แสดงรหัสหมวด/กลุ่มตัวชี้ (Index_id) ที่ใช้ในลิงก์ API ค่าไม่ใช่เป็นค่าเริ่มต้นคือ 0000000000000000

```
In [101]: index_id_df = pd.read_csv('data/รหัสหมวดตัวชี้ CPI_INDEX 31-8-2020.csv', dtype={'รหัสตัวชี้': np.int64})  
index_id_df
```

Out[101]:

รหัสตัวชี้	ชื่อหมวด/กลุ่มตัวชี้
0	รวมทุกรายการ
1	หมวดอาหารและเครื่องดื่มน้ำมันและก่ออํส
2	อาหารปรุงสำปัน
3	ข้าว แป้งและผลิตภัณฑ์จากแป้ง
4	ข้าว
...	...
185	น้ำมัน 91 95 ตีเซล
186	CPI ในรวมน้ำมัน 91 95 ตีเซล
187	CPI ในรวมน้ำมัน 91 95 ตีเซล และรายการของ คน.
188	CPI ในรวมอาหารสด
189	CPI ในรวมพลังงาน

190 rows × 2 columns

Data collection and storage

ดึงข้อมูลจากลิงก์ API

อ่านข้อมูลรหัสหมวด/กลุ่มตัวชี้ (Index_id) จากไฟล์ รหัสหมวดตัวชี้ CPI_INDEX 31-8-2020.csv ที่เก็บไว้ในโฟลเดอร์ data

ข้อมูลที่อ่านมาจะเก็บไว้ในตัวแปร **CPI_Index** แต่ค่าที่ต้องการใช้มีแค่รหัสหมวด/กลุ่มตัวชี้ (Index_id) จึงทำการเลือกเฉพาะคอลัมน์รหัสตัวชี้แล้วเก็บไว้ในตัวแปร **CPI_index_id**

ข้อมูลที่เก็บไว้ในตัวแปร **CPI_index_id** อยู่ในรูปแบบของ pandas.Series ซึ่งไม่สะดวกต่อการใช้งานจึงทำการเปลี่ยนให้อยู่ในรูปแบบของ list โดยเก็บข้อมูลไว้ในตัวแปร **CPI_index_id_list**

```
In [3]: CPI_Index = pd.read_csv("data/รหัสหมวดตัวชี้ CPI_INDEX 31-8-2020.csv", dtype={'รหัสตัวชี้': np.int64})  
CPI_index_id = CPI_Index['รหัสตัวชี้']  
CPI_index_id_list = CPI_index_id.tolist()  
print('Sample index_id:', CPI_index_id_list[:3])
```

Sample index_id: [0, 1000000000000000, 1100000000000000]

สร้างตัวแปร data เพื่อรอเก็บข้อมูลที่อ่านเข้ามา

```
In [4]: data = []  
  
def addData(lists): #รับข้อมูลที่อยู่ในรูปแบบของ List  
    for i in range(len(lists)): #อ่านข้อมูลตามความยาวของ List นั้น  
        data.append(lists[i]) #เอาข้อมูลใน List แต่ละตัวซึ่งอยู่ในรูปของ dict ไปเก็บไว้ในตัวแปร data
```

ในงานนี้จะรีเคราะห์เฉพาะตัวชี้นิรภัยริก็ท์ว่าไปในกรุงเทพและปริมณฑล ตั้งแต่ปีค.ศ. 2017-2020 เลยทำการสร้างและกำหนดค่าตัวแปร

- **region_id** เท่ากับ 0 (กรุงเทพและปริมณฑล)
- **from_year** เท่ากับ 2017 (เริ่มอ่านข้อมูลตั้งแต่ปีค.ศ. 2017)
- **to_year** เท่ากับ 2020 (อ่านข้อมูลถึงปีค.ศ. 2020)

ตัวต้องการดึงข้อมูลในกฎีก้าว มีเรื่องคันและปีสุดท้ายอ่อน ๆ สามารถแก้ไขค่าที่ตัวแปรได้เลย

เพื่อจัดข้อมูลที่ต้องการดึงมีรายละเอียดต้องทำการสร้างสิ่ง API แล้วค่อย ๆ ตั้งที่ล่ะรอบ

โดยเริ่มจากลูปในตัวแปร **CPI_index_id_list** ที่เก็บรหัสหมวด/กลุ่มตัวชี้ (Index_id) โดย i ในแต่ละรอบคือรหัสหมวดที่ต้องใส่ไปในสิ่งที่จัดทำ การสร้างตัวแปร url เพื่อใช้เก็บลิงก์ที่จะขอข้อมูลในแต่ละรอบ จากนั้นก็เอา parameters ต่าง ๆ ใส่เข้าไป

โดยที่ต้องใส่โคดสามนี้ str(i) + ('0' * (16 - len(str(i)))) เพราะว่า index_id ที่ใช้ในสิ่ง API ต้องมีทั้งหมด 16 หลักและเลข 0 ต่อท้ายถ้าหากไม่ครบ 16 หลัก (0000000000000000) แต่ข้อมูลใน **CPI_index_id_list** บางตัวมีไม่ถึง 16 หลัก เช่น 0

- โคดในส่วนนี้ str(i) จึงทำการแปลงข้อมูลเป็น string เพื่อร่วมกับส่วนด้านหลัง
- โคดในส่วนนี้ ('0' * (16 - len(str(i)))) จะทำการพับความยาวของ i ซึ่งต่อ index_id ที่วน回来ในแต่ละรอบ จึงทำการแปลงข้อมูล string แล้วค่อยทำการนับความยาวของ i และป่าว 16 - ความยาวของ i เพื่อนำไปปะกับ '0' เพื่อสร้าง 0 ที่เหลือมาต่อท้ายกับโคดในส่วนด้านหน้าเพื่อให้ index_id ครบ 16 หลัก

Example รอบที่ 1-3:

1. https://dataapi.moc.go.th/cpig-indexes?region_id=0&index_id=0000000000000000&from_year=2017&to_year=2020
2. https://dataapi.moc.go.th/cpig-indexes?region_id=0&index_id=1000000000000000&from_year=2017&to_year=2020
3. https://dataapi.moc.go.th/cpig-indexes?region_id=0&index_id=1100000000000000&from_year=2017&to_year=2020

```
In [7]: region_id = "0"  
from_year="2017"
```

```
to_year="2020"

# url = 'https://dataapi.moc.go.th/cpig-indexes?region_id=0&index_id=1000000000000000&from_year=2017&t
o_year=2020'

for i in CPI_index_id_list:
    url = "https://dataapi.moc.go.th/cpig-indexes?region_id=" + region_id + "&index_id=" + str(i) +
('0' * (16 - len(str(i)))) + "&from_year=" + from_year + "&to_year=" + to_year

    response = urllib.request.urlopen(url) #ส่ง request ขอข้อมูล
    raw_json = response.read().decode("utf-8") #อ่านข้อมูลที่รับกลับมา ข้อมูลมีส่วนที่เป็นภาษาไทยและต้อง decode ด้วย
    utf_8
    addData(json.loads(raw_json)) #Loads json ที่อ่านค่าได้ซึ่งเมื่ออ่านแล้วจะอยู่ในรูป [{key:value}] สงค่าที่อ่านได้ที่อยู่
    ในรูปของ List ไปที่ function addData() เพื่อเอาข้อมูลมาเก็บไว้

print(len(data))

7748
```

7748

```
In [8]: df = pd.json_normalize(data) #นำตัวไปรับ data ที่เก็บข้อมูลทั้งหมดไว้มาแปลงให้อยู่ในรูปของ dataframe  
df
```

Out[8]:

	index_id	index_description	region_id	region_name	base_year	year	month	price_index	mon	yoy	aoa
0	0000000000000000	รวมทุกรายการ	0	กรุงเทพมหานคร และปริมณฑล	2015	2017	1	100.53	0.07	1.18	1.18
1	0000000000000000	รวมทุกรายการ	0	กรุงเทพมหานคร และปริมณฑล	2015	2017	2	100.53	0.00	1.00	1.09
2	0000000000000000	รวมทุกรายการ	0	กรุงเทพมหานคร และปริมณฑล	2015	2017	3	100.18	-0.35	0.49	0.88
3	0000000000000000	รวมทุกรายการ	0	กรุงเทพมหานคร และปริมณฑล	2015	2017	4	100.34	0.16	0.18	0.71
4	0000000000000000	รวมทุกรายการ	0	กรุงเทพมหานคร และปริมณฑล	2015	2017	5	100.48	0.14	-0.04	0.56
...
7743	9520000000000000	CPI ไม่รวมบ้านมีน 91, 95, เดชล	0	กรุงเทพมหานคร และปริมณฑล	2015	2020	8	102.67	0.23	-0.21	-0.69
7744	9700000000000000	CPI ไม่รวมอาหารสด	0	กรุงเทพมหานคร และปริมณฑล	2015	2019	11	102.45	0.04	0.19	0.39
7745	9700000000000000	CPI ไม่รวมอาหารสด	0	กรุงเทพมหานคร และปริมณฑล	2015	2019	12	102.48	0.03	0.67	0.41
7746	9800000000000000	CPI ไม่รวมพลังงาน	0	กรุงเทพมหานคร และปริมณฑล	2015	2019	11	103.05	-0.13	0.84	1.27
7747	9800000000000000	CPI ไม่รวมพลังงาน	0	กรุงเทพมหานคร และปริมณฑล	2015	2019	12	103.04	-0.01	0.99	1.25

7748 rows x 11 columns

แปลงข้อมูลจาก DataFrame ไปเป็น .csv

ເລີ້ມໄຕໃນໜີ່ອ data.csv

```
In [ ]: #เก็บข้อมูลในรูปแบบไฟล์ csv โดยเก็บไว้ในชื่อว่า data.csv และต้องตั้งค่า encoding เป็น utf-8-sig เพราะมีข้อมูลที่เป็นภาษาไทย
df.to_csv('data/data.csv', header=True, index=False, encoding='utf-8-sig')
```

Data pre-processing and analysis

```
In [102]: #อ่านข้อมูล data.csv และเก็บไว้ในตัวแปร df เพื่อใช้ในการทำ pre-processing  
df = pd.read_csv('data/data.csv')  
df
```

Out[102]:

	index_id	index_description	region_id	region_name	base_year	year	month	price_index	mon	yoy	aoa
0	0	รวมทุกรายการ	0	กรุงเทพมหานคร และปริมณฑล	2015	2017	1	100.53	0.07	1.18	1.18
1	0	รวมทุกรายการ	0	กรุงเทพมหานคร และปริมณฑล	2015	2017	2	100.53	0.00	1.00	1.09
2	0	รวมทุกรายการ	0	กรุงเทพมหานคร และปริมณฑล	2015	2017	3	100.18	-0.35	0.49	0.88
3	0	รวมทุกรายการ	0	กรุงเทพมหานคร และปริมณฑล	2015	2017	4	100.34	0.16	0.18	0.71
4	0	รวมทุกรายการ	0	กรุงเทพมหานคร และปริมณฑล	2015	2017	5	100.48	0.14	-0.04	0.56
...
7743	9520000000000000	CPI 'ไม่รวมบ้านปั้น 91, 95, ตีเข็ล'	0	กรุงเทพมหานคร และปริมณฑล	2015	2020	8	102.67	0.23	-0.21	-0.69
7744	9700000000000000	CPI 'ไม่รวมอาหารสด'	0	กรุงเทพมหานคร และปริมณฑล	2015	2019	11	102.45	0.04	0.19	0.39
7745	9700000000000000	CPI 'ไม่รวมอาหารสด'	0	กรุงเทพมหานคร และปริมณฑล	2015	2019	12	102.48	0.03	0.67	0.41
7746	9800000000000000	CPI 'ไม่รวมพสจงงาน'	0	กรุงเทพมหานคร และปริมณฑล	2015	2019	11	103.05	-0.13	0.84	1.27

7748 rows × 11 columns

หากเรา missing value โดยใช้คำสั่ง `isna()` ในการตรวจสอบว่าเป็นค่า missing value หรือไม่ ถ้าเป็นจะชี้เป็น True แต่ถ้าไม่เป็นจะชี้เป็น False และใช้คำสั่ง `sum()` ในการนับจำนวน missing value ทั้งหมด

In [103]: `df.isna().sum()`

```
Out[103]: index_id      0
index_description  0
region_id        0
region_name       0
base_year         0
year              0
month             0
price_index       0
mon               0
yoy                0
aoa                0
dtype: int64
```

หลังจากตรวจสอบแล้วข้อมูลที่ดึงมาไม่มีค่า missing value

ทำการ `groupby` และที่มี `index_description` เพื่อนับกันแล้วทำการนับจำนวนแคล ข้อมูลที่ได้จะอยู่ในรูปของ `pandas.Series` และเก็บไว้ในตัวแปร `countGroup`

เนื่องจากตัวแปร `countGroup` ยังค่าจำนวนมาก เลยแปลงให้อยู่ในรูปของ dict เพื่อที่จะได้เห็นชัดเจนว่า `index_description` แต่ละตัวมีจำนวนแคลทั้งหมดเท่าไรและเก็บไว้ในตัวแปร `dictCountGroup`

เอาตัวแปร `dictCountGroup` มาเรียงลำดับจากน้อยไปมากโดยเรียงจาก `value` ซึ่งคือจำนวนแคลทั้งหมด เพื่อให้ง่ายต่อการวิเคราะห์

```
In [104]: countGroup = df['index_description'].groupby(df['index_description']).count()
dictCountGroup = countGroup.to_dict()
{k: v for k, v in sorted(dictCountGroup.items(), key=lambda item: item[1])}
```

```
Out[104]: {'CPI ไม่รวมพลังงาน': 2,
'CPI ไม่รวมอาหารสด': 2,
'CPI ไม่รวมบ้าน 91 , 95 , ตีเซล': 44,
'กลุ่มอาหารสดและผลิตภัณฑ์': 44,
'กางเกงบุรุษ': 44,
'กางเกงสตรีและกระโปรงสตรี': 44,
'กางเกงเด็ก': 44,
'การบันทึกและการอ่าน': 44,
'การศึกษา': 44,
'การศึกษาและกีฬา': 44,
'การศึกษาระดับมัธยมศึกษา ดำเนินการนิดและอาชีวะ': 44,
'การศึกษาระดับมัธยมศึกษาสายสามัญ': 44,
'การศึกษาระดับอุดมศึกษา': 44,
'การลือสาร': 44,
'ของเล่น': 44,
'ข้าว': 44,
'ข้าว แป้งและผลิตภัณฑ์จากแป้ง': 44,
'คนรู้ใช้ / คณาน': 44,
'คนไข้': 44,
'คนไข้ใน': 44,
'ค้าการบันทึก': 44,
'ค้าของใช้ส่วนบุคคล': 44,
'ค้าจัง': 44,
'ค้าจ้างตัดเย็บและอุปกรณ์': 44,
'ค้าเชื้อยานพาหนะ': 44,
'ค้าข้อมูล': 44,
'ค้าข้อมูลและดูแลรักษา': 44,
'ค้าครัวรักษา ค้าบริการ และค้ายาของคนไข้ใน': 44,
'ค้าครัวรักษาและคาย': 44,
'ค้าครัวรักษาและอุปกรณ์ตัดกาว': 44,
'ค้าครัวรักษาสายตา': 44,
'ค้าครัวรักษา ค้าบริการ และค้ายาของคนไข้อก': 44,
'ค้าทำฟัน': 44,
'ค้าที่พักอาศัย': 44,
'ค้าธรรมเนียมและค่าสมាជິກ': 44,
'ค้าบริการการลือสาร': 44,
'ค้าบริการค่าปรับ': 44,
'ค้าบริการน้ำดื่ม': 44,
'ค้าบริการน้ำดื่มน้ำดื่ม': 44,
'ค้าบริการสำนักงานบุคคล': 44,
'ค้าบริการสำนักงานบุคคลเพื่อความฝันคลาย': 44,
'ค้าบริการสำนักงานบุคคลเพื่อความสวยงาม': 44,
'ค้าบริการสำนักงานบุคคลเพื่อความสวยงาม': 44,
'ค้าประภัยน้ำดื่ม': 44,
'ค้าภาษี': 44,
'ค้าภาษีและประภัยน้ำดื่ม': 44,
'ค้ายา': 44,
'ค้ายาและเบนซิน': 44,
'ค้ารักษากายภาพ': 44,
'ค้าห้องพักคนไข้': 44,
'ค้าอาหาร ของหวานและอัน ฯ': 44,
'ค้าอุปกรณ์การบันทึก': 44,
'ค้าอุปกรณ์กีฬาและเครื่องเส่น': 44,
'ค้าชา': 44,
'ค้าเรียนพิเศษ': 44,
'ค่าเล่าเรียนและค่าธรรมเนียมการศึกษา': 44,
'ค่าแรง': 44,
'ค่าวัสดุและอุปกรณ์': 44}
```



```
'เบื้องล้วงสุด': 44,
'เบื้องล้วงมีประป.': 44,
'เปิด ไก': 44,
'เปิด ไก ส่าเร็จรูปและแปรรูป': 44,
'เบชเก็ตช์': 44,
'เลื่อนราก': 44,
'เสิร์ฟอาหารรุ่งสืบฯ': 44,
'เสือผ่าสดรือเนา': 44,
'เสือผ่าสดรือเนา': 44,
'เสือสดร': 44,
'เสือลึก': 44,
'แฟ้มและผลิตภัณฑ์จากแฟ้ม': 44,
'แสงสว่าง': 44,
'ไข': 44,
'ไขและผลิตภัณฑ์เม': 44,
'ไฟฟ้า': 44,
'ไฟฟ้า เขื่อเหลือง น้ำประปาและแสงสว่าง': 44,
'ไม่รวมอาหาร เครื่องดื่ม และพลังงาน': 44,
'ค่าบริการอื่น ๆ': 88}
```

หลังจากทำการรันแล้วจะเห็นได้ว่าหมวดส่วนใหญ่จะมีแค่ว่าทั้งหมด 44 แต่ โดย 44 ตามมาจากปีค.ศ. 2017-2020 โดยปีค.ศ. 2017-2019 แต่ละปีจะมีทั้งหมด 12 แคว แต่ปีค.ศ. 2020 ยังไม่จบปีเลยมีข้อมูลสิ่งแวดล้อมสิ่งที่มาทั้งหมด 8 แคว เมื่อรวมรวมกันแล้ว ($12+12+12+8$) จึงมีทั้งหมด 44 แคว ($1 \text{ แคว} = 1 \text{ เม็ดอ่อน}$)

แต่จะมี 3 หมวดที่แยกต่างหากหมวดอื่น ๆ ได้แก่ CPI ไม่รวมพลังงาน, CPI ไม่รวมอาหารสด มีทั้งหมด 2 แควและ ค่าบริการอื่น ๆ มีทั้งหมด 88 แคว

ลองแควที่มี index_description เท่ากับ CPI ไม่รวมพลังงานและ CPI ไม่รวมอาหารสด เพราะว่ามีข้อมูลน้อยไป เป็นอย่างไรก็ตามนี้จะวิเคราะห์ดูว่ามีราคายุ่นต่ำกว่าทั่วไปในกรุงเทพและปริมณฑล ตั้งแต่ปีค.ศ. 2017-2020 ซึ่งต้องมีข้อมูลทั้งหมดหมวดละ 44 แคว แต่ 2 หมวดดังกล่าวมีข้อมูลไม่พอจึงต้องลบหัก

```
In [105]: df = df[df.index_description != 'CPI ไม่รวมพลังงาน']
df = df[df.index_description != 'CPI ไม่รวมอาหารสด']
```

ตรวจสอบแควที่มี index_description เท่ากับ ค่าบริการอื่น ๆ เพาะมีจำนวนแควที่แยกต่างจากข้อมูลหักหักอื่น ๆ

```
In [106]: df[df.index_description == 'ค่าบริการอื่น ๆ']
```

Out[106]:

index_id	index_description	region_id	region_name	base_year	year	month	price_index	mon	yoy	aoa
4576	4121500000000000	ค่าบริการอื่น ๆ	0 กรุงเทพมหานคร และปริมณฑล	2015	2017	1	102.08	0.0	1.46	1.46
4577	4121500000000000	ค่าบริการอื่น ๆ	0 กรุงเทพมหานคร และปริมณฑล	2015	2017	2	102.08	0.0	1.46	1.46
4578	4121500000000000	ค่าบริการอื่น ๆ	0 กรุงเทพมหานคร และปริมณฑล	2015	2017	3	102.08	0.0	1.46	1.46
4579	4121500000000000	ค่าบริการอื่น ๆ	0 กรุงเทพมหานคร และปริมณฑล	2015	2017	4	102.08	0.0	1.46	1.46
4580	4121500000000000	ค่าบริการอื่น ๆ	0 กรุงเทพมหานคร และปริมณฑล	2015	2017	5	102.08	0.0	1.46	1.46
...
4835	4122400000000000	ค่าบริการอื่น ๆ	0 กรุงเทพมหานคร และปริมณฑล	2015	2020	4	103.82	0.0	0.00	0.00
4836	4122400000000000	ค่าบริการอื่น ๆ	0 กรุงเทพมหานคร และปริมณฑล	2015	2020	5	103.82	0.0	0.00	0.00
4837	4122400000000000	ค่าบริการอื่น ๆ	0 กรุงเทพมหานคร และปริมณฑล	2015	2020	6	103.82	0.0	0.00	0.00
4838	4122400000000000	ค่าบริการอื่น ๆ	0 กรุงเทพมหานคร และปริมณฑล	2015	2020	7	103.82	0.0	0.00	0.00
4839	4122400000000000	ค่าบริการอื่น ๆ	0 กรุงเทพมหานคร และปริมณฑล	2015	2020	8	103.82	0.0	0.00	0.00

88 rows × 11 columns

หลังจากตรวจสอบจะเห็นว่าทั้ง 2 ตัวมีข้อมูล index_id, price_index ที่แยกต่างกัน แต่มี index_description ที่เหมือนกันเลยทำให้มีจำนวนมากกว่าหมวดอื่น ๆ

ซึ่งท่าการแก้ไขโดยนำ price_index ของทั้ง 2 ตัวมาเฉลี่ยกัน และนำค่าเฉลี่ยที่ได้แทนลงไว้ใน price_index ของ index_id ที่เท่ากับ 4121500000000000 และทำการลบแควที่ index_id เท่ากับ 4122400000000000

```
In [107]: print('index_id = 4121500000000000\n', 'price_index =', df[df.index_id == 4121500000000000]['price_index'].drop_duplicates().iloc[0],
'\n\n',
'index_id = 4122400000000000\n', 'price_index =', df[df.index_id == 4122400000000000]['price_index'].drop_duplicates().iloc[0])
```

```
index_id = 4121500000000000
price_index = 102.08

index_id = 4122400000000000
price_index = 103.82
```

```
In [108]: #กำหนดให้ priceIndexFirst เก็บค่า price_index ของ index_id ที่เท่ากับ 4121500000000000
priceIndexFirst = df[df.index_id == 4121500000000000]['price_index'].iloc[0]
```

#เพิ่มต่อไปนี้เพื่อให้คำแนะนำว่า ค่า price_index ของ index_id ที่เท่ากับ 4121500000000000 คือ

```
#ກ່າວເພີມຕົວທີ່ສອງດ້ວຍລວມຂອງຕົວທີ່ໄວ້ໃນຕົວນີ້
priceIndexSecond = df[df.index_id == 41224000000000000000]['price_index'].iloc[0]
```

```
In [109]: #ນໍາ price_index ຂອງທີ່ສອງດ້ວຍລວມຂອງຕົວທີ່ໄວ້ໃນຕົວນີ້ໃຫຍ່
newPriceIndex = ((df[df.index_id == 41215000000000000000]['price_index'].add(priceIndexSecond, fill_value=0))/2).iloc[0]
print(newPriceIndex)

#ນໍາ newPriceIndex ທີ່ເປັນຄ່າເອົ້າຂອງທີ່ສອງດ້ວຍມາແທນຕາລູໄປໃນ index_id ທີ່ເທົ່າກັນ 4121500000000000
df[df.index_id == 4121500000000000] = df[df.index_id == 4121500000000000].replace(priceIndexFirst, newPriceIndex)
```

```
102.94999999999999
```

```
In [110]: #ກ່າວຈັກແທນຄໍາໃໝ່ມລົງໄປແລ້ວ ຈຶ່ງທ່າງຮອບ index_id ທີ່ເທົ່າກັນ 4122400000000000 ອອກຈາກ df
df = df[df.index_id != 4122400000000000]
df[df.index_description == 'ຄານຮັກຮືນ ຖ']
```

Out[110]:

	index_id	index_description	region_id	region_name	base_year	year	month	price_index	mon	yoy	aoa
4576	4121500000000000	ຄານຮັກຮືນ ພ	0	ກຽງເທິພນານຄຣ ແລະປ່ວນິມຄເຫລ	2015	2017	1	102.95	0.0	1.46	1.46
4577	4121500000000000	ຄານຮັກຮືນ ພ	0	ກຽງເທິພນານຄຣ ແລະປ່ວນິມຄເຫລ	2015	2017	2	102.95	0.0	1.46	1.46
4578	4121500000000000	ຄານຮັກຮືນ ພ	0	ກຽງເທິພນານຄຣ ແລະປ່ວນິມຄເຫລ	2015	2017	3	102.95	0.0	1.46	1.46
4579	4121500000000000	ຄານຮັກຮືນ ພ	0	ກຽງເທິພນານຄຣ ແລະປ່ວນິມຄເຫລ	2015	2017	4	102.95	0.0	1.46	1.46
4580	4121500000000000	ຄານຮັກຮືນ ພ	0	ກຽງເທິພນານຄຣ ແລະປ່ວນິມຄເຫລ	2015	2017	5	102.95	0.0	1.46	1.46
4581	4121500000000000	ຄານຮັກຮືນ ພ	0	ກຽງເທິພນານຄຣ ແລະປ່ວນິມຄເຫລ	2015	2017	6	102.95	0.0	1.46	1.46
4582	4121500000000000	ຄານຮັກຮືນ ພ	0	ກຽງເທິພນານຄຣ ແລະປ່ວນິມຄເຫລ	2015	2017	7	102.95	0.0	1.46	1.46
4583	4121500000000000	ຄານຮັກຮືນ ພ	0	ກຽງເທິພນານຄຣ ແລະປ່ວນິມຄເຫລ	2015	2017	8	102.95	0.0	1.46	1.46
4584	4121500000000000	ຄານຮັກຮືນ ພ	0	ກຽງເທິພນານຄຣ ແລະປ່ວນິມຄເຫລ	2015	2017	9	102.95	0.0	1.46	1.46
4585	4121500000000000	ຄານຮັກຮືນ ພ	0	ກຽງເທິພນານຄຣ ແລະປ່ວນິມຄເຫລ	2015	2017	10	102.95	0.0	1.46	1.46
4586	4121500000000000	ຄານຮັກຮືນ ພ	0	ກຽງເທິພນານຄຣ ແລະປ່ວນິມຄເຫລ	2015	2017	11	102.95	0.0	0.00	1.33
4587	4121500000000000	ຄານຮັກຮືນ ພ	0	ກຽງເທິພນານຄຣ ແລະປ່ວນິມຄເຫລ	2015	2017	12	102.95	0.0	0.00	1.21
4588	4121500000000000	ຄານຮັກຮືນ ພ	0	ກຽງເທິພນານຄຣ ແລະປ່ວນິມຄເຫລ	2015	2018	1	102.95	0.0	0.00	0.00
4589	4121500000000000	ຄານຮັກຮືນ ພ	0	ກຽງເທິພນານຄຣ ແລະປ່ວນິມຄເຫລ	2015	2018	2	102.95	0.0	0.00	0.00
4590	4121500000000000	ຄານຮັກຮືນ ພ	0	ກຽງເທິພນານຄຣ ແລະປ່ວນິມຄເຫລ	2015	2018	3	102.95	0.0	0.00	0.00
4591	4121500000000000	ຄານຮັກຮືນ ພ	0	ກຽງເທິພນານຄຣ ແລະປ່ວນິມຄເຫລ	2015	2018	4	102.95	0.0	0.00	0.00
4592	4121500000000000	ຄານຮັກຮືນ ພ	0	ກຽງເທິພນານຄຣ ແລະປ່ວນິມຄເຫລ	2015	2018	5	102.95	0.0	0.00	0.00
4593	4121500000000000	ຄານຮັກຮືນ ພ	0	ກຽງເທິພນານຄຣ ແລະປ່ວນິມຄເຫລ	2015	2018	6	102.95	0.0	0.00	0.00
4594	4121500000000000	ຄານຮັກຮືນ ພ	0	ກຽງເທິພນານຄຣ ແລະປ່ວນິມຄເຫລ	2015	2018	7	102.95	0.0	0.00	0.00
4595	4121500000000000	ຄານຮັກຮືນ ພ	0	ກຽງເທິພນານຄຣ ແລະປ່ວນິມຄເຫລ	2015	2018	8	102.95	0.0	0.00	0.00
4596	4121500000000000	ຄານຮັກຮືນ ພ	0	ກຽງເທິພນານຄຣ ແລະປ່ວນິມຄເຫລ	2015	2018	9	102.95	0.0	0.00	0.00
4597	4121500000000000	ຄານຮັກຮືນ ພ	0	ກຽງເທິພນານຄຣ ແລະປ່ວນິມຄເຫລ	2015	2018	10	102.95	0.0	0.00	0.00
4598	4121500000000000	ຄານຮັກຮືນ ພ	0	ກຽງເທິພນານຄຣ ແລະປ່ວນິມຄເຫລ	2015	2018	11	102.95	0.0	0.00	0.00
4599	4121500000000000	ຄານຮັກຮືນ ພ	0	ກຽງເທິພນານຄຣ ແລະປ່ວນິມຄເຫລ	2015	2018	12	102.95	0.0	0.00	0.00
4600	4121500000000000	ຄານຮັກຮືນ ພ	0	ກຽງເທິພນານຄຣ ແລະປ່ວນິມຄເຫລ	2015	2019	1	102.95	0.0	0.00	0.00
4601	4121500000000000	ຄານຮັກຮືນ ພ	0	ກຽງເທິພນານຄຣ ແລະປ່ວນິມຄເຫລ	2015	2019	2	102.95	0.0	0.00	0.00
4602	4121500000000000	ຄານຮັກຮືນ ພ	0	ກຽງເທິພນານຄຣ ແລະປ່ວນິມຄເຫລ	2015	2019	3	102.95	0.0	0.00	0.00
4603	4121500000000000	ຄານຮັກຮືນ ພ	0	ກຽງເທິພນານຄຣ ແລະປ່ວນິມຄເຫລ	2015	2019	4	102.95	0.0	0.00	0.00
4604	4121500000000000	ຄານຮັກຮືນ ພ	0	ກຽງເທິພນານຄຣ ແລະປ່ວນິມຄເຫລ	2015	2019	5	102.95	0.0	0.00	0.00
4605	4121500000000000	ຄານຮັກຮືນ ພ	0	ກຽງເທິພນານຄຣ ແລະປ່ວນິມຄເຫລ	2015	2019	6	102.95	0.0	0.00	0.00
4606	4121500000000000	ຄານຮັກຮືນ ພ	0	ກຽງເທິພນານຄຣ ແລະປ່ວນິມຄເຫລ	2015	2019	7	102.95	0.0	0.00	0.00

```
In [111]: #ทำการนับค่าของ column หลังจากแก้ไขหมวดที่มีจำนวนแผลกดค้างจากหมวดเดิมแล้วจะเรียบร้อย  
#จะเห็นได้ว่าทุกหมวดมีจำนวนแผลกดค้างเท่ากันทั้งหมด  
countGroup = df['index_description'].groupby(df['index_description']).count()  
dictCountGroup = countGroup.to_dict()  
{k: v for k, v in sorted(dictCountGroup.items(), key=lambda item: item[1])}
```

```
Out[111]: {'CPI' ไม่รวมหน้ามัน 91 , 95 , ตีเซล': 44,
'กลุ่มอาหารสดและพืชผัก': 44,
'กาแฟรักษา': 44,
'กาแฟสดรีดและกระป๋องสดรี': 44,
'กาแฟเก็บเด็ก': 44,
'การบันทึกและการอ่าน': 44,
'การคำสาน': 44,
'การศึกษา': 44,
'การศึกษาระดับประถมศึกษา': 44,
'การศึกษาระดับมัธยมศึกษา ดำเนินเทคโนโลยีและภาษาไทย': 44,
'การศึกษาระดับมัธยมศึกษาสามัญ': 44,
'การศึกษาระดับอนุมัติศึกษา': 44,
'การเสื่อสาร': 44,
'ของเล่น': 44,
'ข่าว': 44,
'ข่าว แบ่งและผลิตภัณฑ์จากแฟ้ม': 44,
'คนรับใช้ / คนงาน': 44,
'คนไข้เมือง': 44,
'คนไข้ใน': 44,
'ค่าการบันทึก': 44,
'ค่าของใช้ส่วนบุคคล': 44,
'ค่าจ้าง': 44,
'ค่าจ้างเด็กรักษาและอุปกรณ์': 44,
'ค่าเชื้อ yan พาหนะ': 44,
'ค่าเชื้อภัยชั่ว': 44,
'ค่าเชื้อภัยและดูแลรักษา': 44,
'ค่าตรวจรักษา ค่าบริการ และค่ายาของคนไข้ใน': 44,
'ค่าตรวจรักษาและค่าตัว': 44,
'ค่าตรวจรับนิรภัยทางกองปฏิบัติการ': 44,
'ค่าตรวจสภาพด้': 44,
'ค่าตรวจโรครักษา ค่าบริการ และค่ายาของคนไข้ข้างนอก': 44,
'ค่าไฟฟ้า': 44,
'ค่าที่พักอาศัย': 44,
'ค่าธรรมเนียมและค่าสามารถ': 44,
'ค่าบริการค่าเสื่อสาร': 44,
'ค่าบริการค่าห้องสี': 44,
'ค่าบริการรักษาพยาบาล': 44,
'ค่าบริการล่วงบุคคลเพื่อความผ่านคลาย': 44,
'ค่าบริการล่วงบุคคลเพื่อความส่ายงวน': 44,
'ค่าบริการอื่น ๆ': 44,
'ค่าวัสดุปักรักษาพยาบาลและประกันชีริต': 44,
'ค่าประกันภัยรถ': 44,
'ค่าภาระ': 44,
'ค่าภาระและประกันภัยรถยนต์': 44,
'ค่าย': 44,
'ค่ายและแขวงภัยชาติ': 44,
'ค่ารักษาพยาบาล': 44,
'ค่าห้องพักคนไข้': 44,
'ค่าอาหาร ของหวานพร้อม ไหว้เจ้าและอื่น ๆ': 44,
'ค่าอุปกรณ์การบันทึก': 44,
'ค่าอุปกรณ์กีฬาและเครื่องเล่น': 44,
'ค่าเช่า': 44,
'ค่าเรียนพิเศษ': 44,
'ค่าเลาเรียนและค่าธรรมเนียมการศึกษา': 44,
'ค่าแรง': 44,
'ค่าโฆษณาออกอากาศท่องเที่ย': 44,
'ค่าโฆษณาออกอากาศท่องเที่ยว': 44,
```

พ.ร.บ. พัฒนาสังคม ปี ๒๕๖๔ : ๔๔,
'ค่าโดยสารในเมืองที่น้ำ': 44,
'ค่าใช้จ่ายที่เกี่ยวกับรถ': 44,
'ค่าใช้จ่ายส่วนบุคคล': 44,
'ค่าใช้จ่ายอื่น ๆ เกี่ยวกับการศึกษา': 44,
'ค่าใช้จ่ายเกี่ยวกับการทำอาชีวประดิษฐ์': 44,
'ค่าใช้จ่ายเกี่ยวกับการอ่าน': 44,
'ค่าใช้จ่ายเกี่ยวกับสัตว์เลี้ยง': 44,
'ค่าใช้จ่ายในการเดินทางไปสถานที่เชิงและท่องเที่ยว': 44,
'เด็กขาด้าผู้มีโรคพื้นฐาน (ไม่รวมค่าเข้าบ้าน)': 44,
'น้ำแร่และน้ำหวาน': 44,
'น้ำประปา': 44,
'น้ำฝน ๙๑ , ๙๕ , ตัวเลข': 44,
'น้ำอันเชื่อเพลิง': 44,
'น้ำมันและไข่น้ำ': 44,
'บริสุทธิ์อื่น ๆ': 44,
'บุรุษ': 44,
'ปลาน้ำจืด': 44,
'ปลาทะเลสด': 44,
'ปลาบรรจุ': 44,
'ปลาและสัตว์น้ำ': 44,
'ผลิตภัณฑ์จากเมือง': 44,
'ผลิตภัณฑ์จากตลาด': 44,
'ผลิตภัณฑ์ยาสูบ': 44,
'ผลิตภัณฑ์ของท่า': 44,
'ผลิตภัณฑ์สร้าง': 44,
'ผลิตภัณฑ์มีรู': 44,
'ผลิตภัณฑ์ไวไฟ': 44,
'ผลไม้สด': 44,
'ผลไม้สดบรรจุและอื่นๆ': 44,
'ผลไม้บรรจุและอื่น ๆ': 44,
'ผักสด': 44,
'ผักสดบรรจุและอื่นๆ': 44,
'ผักและผลไม้': 44,
'ผ้าและเสื้อผ้า': 44,
'พลงาน': 44,
'ยานพาหนะ': 44,
'ยานพาหนะและน้ำมันเชื้อเพลิง': 44,
'รวมทุกรายการ': 44,
'รองเท้ากีฬาและผ้าใบบุรุษ': 44,
'รองเท้ากีฬาพื้นและหัวไม่สอดรี': 44,
'รองเท้ากีฬาไม้เด็': 44,
'รองเท้าบุรุษ': 44,
'รองเท้าสตีฟ': 44,
'รองเท้าหุ้มส้นบุรุษ': 44,
'รองเท้าหุ้มสีสอดรี': 44,
'รองเท้าเด็ก': 44,
'รองเท้าแตะบุรุษ': 44,
'รองเท้าแตะสตีฟ': 44,
'รองเท้าแตะเด็ก': 44,
'วัสดุก่อสร้าง': 44,
'สตอรี': 44,
'สีครีม': 44,
'สีต้นน้ำแบบบรรจุ': 44,
'สีสันและน้ำ': 44,
'สิ่งที่เกี่ยวกับการทำความสะอาด': 44,
'หนังสือ': 44,
'หนังสือแบบเรียนเอกสาร': 44,
'หนังสือและอุปกรณ์การศึกษา': 44,
'หมวดการตรวจสอบรายการและรายการส่วนบุคคล': 44,
'หมวดการบันทึกการอ่าน การศึกษา และการศาสนา': 44,
'หมวดพาราเนส การขันส่ง และการสื่อสาร': 44,
'หมวดอาหารและเครื่องดื่มน้ำไม่มีแอลกอฮอล์': 44,
'หมวดอื่น ๆ ในใช้ชีวหาย และเครื่องดื่ม': 44,
'หมวดเครื่องบันทึกและรองเท้า': 44,
'หมวดเด็กสถาน': 44,
'อาหารรีไซเคิล-օบกัน': 44,
'อาหารบริโภค-ในบ้าน': 44,
'อาหารรุ่งที่บ้าน': 44,
'อาหารสด': 44,
'อาหารสำเร็จรูป': 44,
'อาหารเสริมที่ดื่ม (ไม่รวมอาหารสด)': 44,
'อุปกรณ์กีฬา': 44,
'อุปกรณ์เชื่อมเข็นบ้าน': 44,
'อุปกรณ์ยานพาหนะ': 44,
'อุปกรณ์เกี่ยวกับการศึกษา': 44,
'เครื่องดื่มน้ำไม่มีแอลกอฮอล์': 44,
'เครื่องดื่มน้ำไม่มีแอลกอฮอล์': 44,
'เครื่องดื่มชา': 44,
'เครื่องดื่มเครื่องดื่ม': 44,
'เครื่องดื่มน้ำ': 44,
'เครื่องดื่มน้ำและน้ำอื่น ๆ': 44,
'เครื่องแบบชาย': 44,
'เครื่องแบบหญิง': 44,
'เครื่องแบบเด็ก': 44,
'เครื่องใช้มีเดลล์ในบ้าน': 44,
'เชือกเหลืองในบ้าน': 44,
'เด็ก': 44,
'เตือสัตว์ เปิดໄກ และสัตว์น้ำ': 44,
'เต้อ้วย': 44,

```
'เมืองลั่นแม่พรป': 44,
'เม็ด ไก': 44,
'เม็ด ไก สด': 44,
'เม็ด ไก ส่าเร็วปะแม่พรป': 44,
'เวชภัณฑ์': 44,
'เส็บบูชา': 44,
'เสือผ่านรชอัน': 44,
'เสือผาสึรีอิน': 44,
'เสือผ้าตีกอี้ๆ': 44,
'เสือสดรี': 44,
'เสือเต็ก': 44,
'แปงและผลิตภัณฑ์จากแปง': 44,
'แสงว่าง': 44,
'ไข': 44,
'ไข่และผลิตภัณฑ์นม': 44,
'ไฟฟ': 44,
'ไฟฟ้า เซื้อเพลิง น้ำประปาและแสงสว่าง': 44,
'ในรวมอาหาร เครื่องดื่ม และผลิตภัณฑ์': 44}
```

Save ข้อมูลที่ pre-processing เรียบร้อยแล้ว จาก DataFrame ไปเป็น .csv

เก็บไว้ในชื่อ clean_data.csv

```
In [112]: #เก็บข้อมูลในรูปแบบไฟล์ csv โดยเก็บไว้ในชื่อว่า clean_data.csv และต้องตั้งค่า encoding เป็น utf-8-sig เพราะมีข้อมูลที่เป็นภาษาไทย
df.to_csv('data/clean_data.csv', header=True, index=False, encoding='utf-8-sig')
```

Visualization

```
In [137]: # -*- coding: utf-8 -*-
import matplotlib.pyplot as plt
%matplotlib inline
```

```
In [138]: #อ่านข้อมูล clean_data.csv และเก็บไว้ในตัวแปร df เพื่อใช้ในการทำ Visualization
df = pd.read_csv('data/clean_data.csv')
df
```

Out[138]:

	index_id	index_description	region_id	region_name	base_year	year	month	price_index	mon	yoy	aoa
0	0	รวมทุกรายการ	0	กรุงเทพมหานคร และปริมณฑล	2015	2017	1	100.53	0.07	1.18	1.18
1	0	รวมทุกรายการ	0	กรุงเทพมหานคร และปริมณฑล	2015	2017	2	100.53	0.00	1.00	1.09
2	0	รวมทุกรายการ	0	กรุงเทพมหานคร และปริมณฑล	2015	2017	3	100.18	-0.35	0.49	0.88
3	0	รวมทุกรายการ	0	กรุงเทพมหานคร และปริมณฑล	2015	2017	4	100.34	0.16	0.18	0.71
4	0	รวมทุกรายการ	0	กรุงเทพมหานคร และปริมณฑล	2015	2017	5	100.48	0.14	-0.04	0.56
...
7695	9520000000000000	CPI "ในรวมน้ำมัน 91 , 95 , ตีเซลล"	0	กรุงเทพมหานคร และปริมณฑล	2015	2020	4	100.43	-1.72	-2.14	-0.14
7696	9520000000000000	CPI "ในรวมน้ำมัน 91 , 95 , ตีเซลล"	0	กรุงเทพมหานคร และปริมณฑล	2015	2020	5	100.29	-0.14	-2.78	-0.66
7697	9520000000000000	CPI "ในรวมน้ำมัน 91 , 95 , ตีเซลล"	0	กรุงเทพมหานคร และปริมณฑล	2015	2020	6	101.70	1.41	-1.31	-0.78
7698	9520000000000000	CPI "ในรวมน้ำมัน 91 , 95 , ตีเซลล"	0	กรุงเทพมหานคร และปริมณฑล	2015	2020	7	102.43	0.72	-0.64	-0.75
7699	9520000000000000	CPI "ในรวมน้ำมัน 91 , 95 , ตีเซลล"	0	กรุงเทพมหานคร และปริมณฑล	2015	2020	8	102.67	0.23	-0.21	-0.69

7700 rows × 11 columns

แสดงดัชนีราคาผู้บริโภคทั่วไปในกรุงเทพและปริมณฑล ตั้งแต่ปีค.ศ. 2017-2020

Line plot

อ่านข้อมูลรหัสหมวด/กลุ่มดัชนี (Index_id) จากไฟล์ รหัสหมวดดัชนี CPI_INDEX 31-8-2020.csv ที่เก็บไว้ในโฟลเดอร์ data

ข้อมูลที่อ่านมาจะเก็บไว้ในตัวแปร CPI_Index แต่ค่าที่ต้องการใช้มีแค่รหัสหมวด/กลุ่มดัชนี (Index_id) จึงทำการเลือกเฉพาะคอลัมน์รหัสดัชนีแล้วเก็บไว้ในตัวแปร CPI_index_id

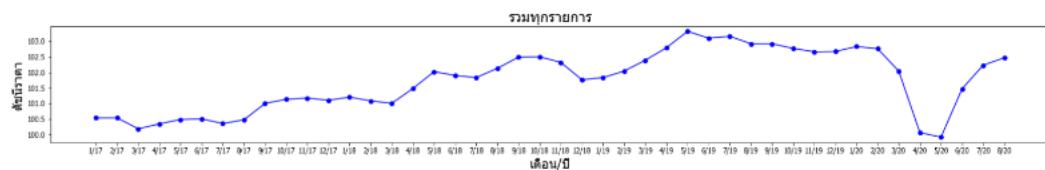
ข้อมูลที่เก็บไว้ในตัวแปร CPI_index_id อù ในรูปแบบของ pandas.Series ซึ่งไม่สะดวกต่อการใช้งานจึงทำการเปลี่ยนให้อยู่ในรูปแบบของ list โดยเก็บข้อมูลไว้ในตัวแปร CPI_index_id_list

```
In [139]: #แปลงข้อมูลในคอลัมน์ 'รหัสดัชนี' ให้เป็นประเภท int64 เพื่อกำหนดจุดทดสอบ
CPI_Index = pd.read_csv("data/รหัสหมวดดัชนี CPI_INDEX 31-8-2020.csv", dtype={'รหัสดัชนี': np.int64})
CPI_index_id = CPI_Index['รหัสดัชนี']
CPI_index_id_list = CPI_index_id.tolist()
# CPI_index_id_list
```

ทำการพล็อต line chart ของหมวดรวมทุกรายการเพื่อดูภาพรวม

```
In [140]: index_ = df[df.index_id == 0] #เลือก index_id ที่เท่ากับ 0 เพื่อเป็นหมวดรวมทุกรายการ แล้วเก็บไว้ในตัวแปร index_
month = index_[‘month’].tolist() #เลือกคอลัมน์ month จากตัวแปร index_ และแปลงให้เป็น List จากนั้นเก็บไว้ในตัวแปร month
year = index_[‘year’].tolist() #เลือกคอลัมน์ year จากตัวแปร index_ และแปลงให้เป็น List จากนั้นเก็บไว้ในตัวแปร year
price_index = index_[‘price_index’].tolist() #เลือกคอลัมน์ price_index จากตัวแปร index_ และแปลงให้เป็น List จากนั้นเก็บไว้ในตัวแปร price_index
month_year = [] #สร้างตัวแปร month_year เพื่อรักษาลักษณะของ month กับ year มาต่อกันเพื่อนำไปใช้ในการผลลัพธ์กราฟ
for i in range(len(month)): #วนลูปตามความยาวของตัวแปร month
    #นำตัวแปร month index ที่ i มาต่อกับตัวแปร year 2 ตัวท้าย index ที่ i โดยขึ้นลงกลางด้วย /
    #หลังจากนั้นมาต่อที่ได้เพิ่มเข้าไปใน month_year
    month_year.append(str(month[i]) + ‘/’ + str(year[i])[2:] )
#example i = 0
# 1/17

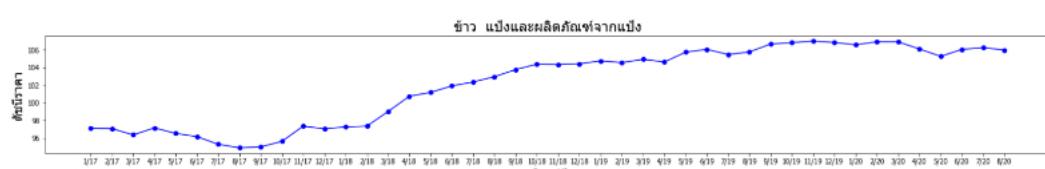
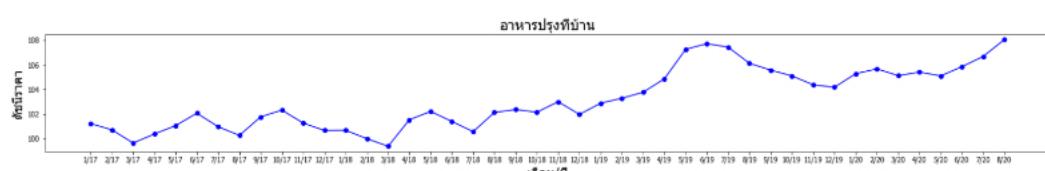
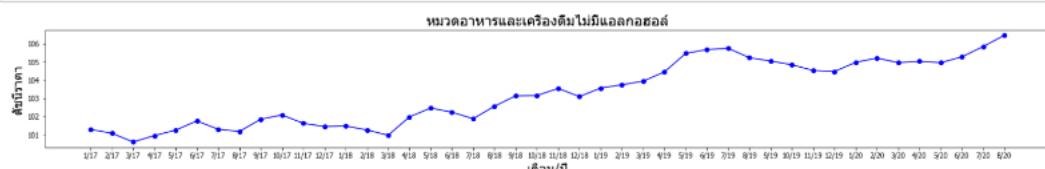
plt.figure(figsize=(25, 3))
#ทำการเปลี่ยนฟอนต์ที่รองรับภาษาไทย
plt.title(‘รวมทุกรายการ’, fontname=‘Tahoma’, fontsize=‘20’)
plt.xlabel(‘เดือน/ปี’, fontname=‘Tahoma’, fontsize=‘18’)
plt.ylabel(‘ตัวชี้มีรaca’, fontname=‘Tahoma’, fontsize=‘18’)
#ผลลัพธ์กราฟโดยแยก x คือเดือน/ปี (จากตัวแปร month_year) และ y คือตัวชี้มีรaca (จากตัวแปร price_index) และตั้งค่าให้สีเป็นสีเขียวและมีจุดวงกลม
plt.plot(month_year, price_index, ‘-bo’);
```

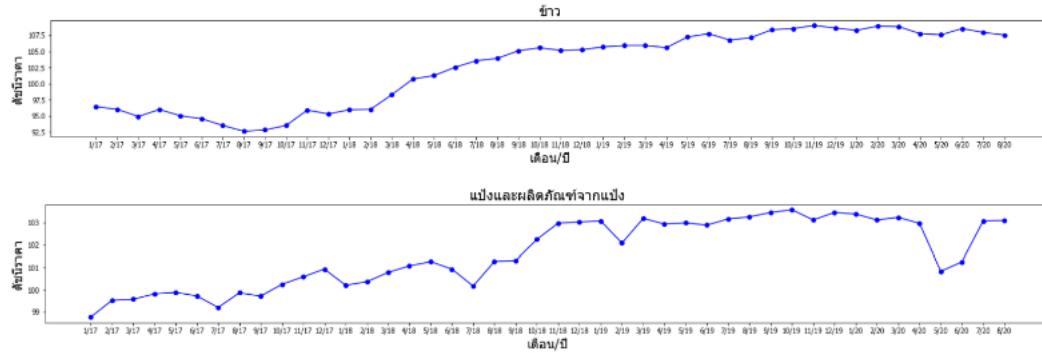


จากราฟเส้น รวมทุกรายการ จะเห็นว่าเราเพาะขายไว้รัสโคริด-19 ในช่วงปีนี้2020 สงผลให้คุณเริ่มไม่กล้าใช้เงิน ทำให้ต้องปรับราคาสินค้าลง (ลดราคางามเพื่อให้ขายออก) และเนื่องจากมีประกาศพรก.ฉกเงินจำกัด NORMAL เมื่อวันที่ 25/03/2020 (<https://www.bangkokbiznews.com/news/detail/872653>) จะเห็นได้ว่าราคาสินค้าต่างๆลดลงอย่างชัดเจน ซึ่งเมื่อขายไว้รัสโคริด-19 เริ่มต้นราคาสินค้าต่างๆก็เริ่มกลับขึ้นมา ซึ่งในช่วงโควิดก็มีสินค้าที่ราคาเพิ่มขึ้นและลดลง

```
In [141]: #มีทั้งหมด 189 หมวด (1-189) ไม่บันหมวดรวมทุกรายการ ถ้าใส่เลขເຍອະຈະໃຫ້ລາຍການ
show = 5 #ตัวแปร show จะแสดงจำนวนกราฟตามที่กำหนด
for i in CPI_index_id_list[1:show+1]:
    #นับถูกนำไปตัวแปร CPI_index_id_list ซึ่งเก็บ index_id ทั้งหมด แต่จะเริ่มพื้นด้วยตัวแปร index_ ที่ 1 เพราะได้ทำการเพลิดเพลินหัวหน้า
    #ก่อนแล้ว
    #index_ = df[df.index_id == i] #เลือก index_id ที่เท่ากับ i และเก็บไว้ในตัวแปร index_
    month = index_[‘month’].tolist() #เลือกคอลัมน์ month จากตัวแปร index_ และแปลงให้เป็น List จากนั้นเก็บไว้ในตัวแปร month
    year = index_[‘year’].tolist() #เลือกคอลัมน์ year จากตัวแปร index_ และแปลงให้เป็น List จากนั้นเก็บไว้ในตัวแปร year
    price_index = index_[‘price_index’].tolist() #เลือกคอลัมน์ price_index จากตัวแปร index_ และแปลงให้เป็น List จากนั้นเก็บไว้ในตัวแปร price_index
    month_year = [] #สร้างตัวแปร month_year เพื่อรักษาลักษณะของ month กับ year มาต่อกันเพื่อนำไปใช้ในการผลลัพธ์กราฟ
    for i in range(len(month)): #วนลูปตามความยาวของตัวแปร month
        #นำตัวแปร month index ที่ i มาต่อกับตัวแปร year 2 ตัวท้าย index ที่ i โดยขึ้นลงกลางด้วย /
        #หลังจากนั้นมาต่อที่ได้เพิ่มเข้าไปใน month_year
        month_year.append(str(month[i]) + ‘/’ + str(year[i])[2:] )

    #เช็คความยาวของ index_ ต้องมากกว่า 0 ถึงจะทำการเพลิดเพลินว่า index_id มากกว่านี้ไม่มีข้อมูล
    if len(index_[‘index_description’].tolist()) > 0:
        plt.figure(figsize=(25, 3))
        #กำหนดให้ title และค่าใน index_description ของแต่ละรอบและเปลี่ยนฟอนต์ที่รองรับภาษาไทย
        plt.title(‘+index_[‘index_description’].tolist()[0], fontname=‘Tahoma’, fontsize=‘20’)
        plt.xlabel(‘เดือน/ปี’, fontname=‘Tahoma’, fontsize=‘18’)
        plt.ylabel(‘ตัวชี้มีรaca’, fontname=‘Tahoma’, fontsize=‘18’)
        #ผลลัพธ์กราฟโดยแยก x คือเดือน/ปี (จากตัวแปร month_year) และ y คือตัวชี้มีรaca (จากตัวแปร price_index) และตั้งค่าให้สีเป็นสีเขียวและมีจุดวงกลม
        plt.plot(month_year, price_index, ‘-bo’);
    else:
        continue
```





จากที่ก้าวข้างต้นเราในช่วงโควิดก็มีทั้งสินค้าที่ราคาเพิ่มขึ้นและลดลง เมื่อต่อจากกราฟเส้นที่แสดงจะเห็นได้ว่าหมวดอาหารและเครื่องดื่มไม่มีผลออกห้อส่วนอาหารปูจงที่มีบ้าน เป็นเด่น มีแนวโน้มดับเบิลร่าคาที่เพิ่มสูงขึ้นในช่วงโควิด-19 ส่วนหมวดที่แนวโน้มดับเบิลร่าคาลดลงในช่วงโควิด-19 เช่น หมวดแป้งและผลิตภัณฑ์จากแป้ง

หมวดที่ราคาลดลง/เพิ่มขึ้น จากปีก่อนมากที่สุด 10 อันดับแรก

Bar plot

In [142]: #เดือนที่จะเทียบคือเดือน 8 และปี 2020 กับ 2019

```
#หมวดที่ราคาลดลงจากปีก่อนมากที่สุด 10 อันดับแรก เก็บไว้ในตัวแปร decreaseFiveYOY
decreaseTenYOY = df[(df.year == 2020) & (df.month == 8)].sort_values(by=['yoy']).head(10)[::-1]

#หมวดที่ราคาน้ำเพิ่มขึ้นจากปีก่อนมากที่สุด 10 อันดับแรก เก็บไว้ในตัวแปร increaseFiveYOY
increaseTenYOY = df[(df.year == 2020) & (df.month == 8)].sort_values(by=['yoy']).tail(10)

print(decreaseTenYOY[['index_description', 'yoy']], '\n\n', increaseTenYOY[['index_description', 'yoy']])
```

	index_description	yoy
6555	การศึกษาระดับมัธยมศึกษา ด้านเทคโนโลยีอาชีวะ	-1.83
7435	กลุ่มอาหารสดและพืชผัก	-2.86
5103	หมวดพาหนะ การขนส่ง และการสื่อสาร	-3.89
5279	ยานพาหนะและนำมี่เชือเพลิง	-6.58
1231	ผลไม้สดแบบรูปและอันๆ	-6.79
1275	ผลไม้สด	-6.86
7523	พัลจังงาน	-8.77
3695	เชือเพลิงในบ้าน	-10.59
7655	นำมี่ 91 , 95 , ตีเซล	-12.49
5455	นำมี่เชือเพลิง	-14.46
	index_description	yoy
1187	ผักผลไม้สดและอันๆ	3.43
351	เนื้อสัตว์ เป็ดไก่ และตัวน้ำ	3.75
1583	เครื่องปรุงรส	3.94
1495	เครื่องปรุงอาหาร	4.78
1539	นำมี่และไข่ม่าน	6.28
4311	beans กับตัว	6.43
395	เนื้อสัตว์สด	9.13
439	เนื้อสัตว์สด	10.95
1099	ผักผลไม้สดและอันๆ	11.35
1143	ผักสด	12.66

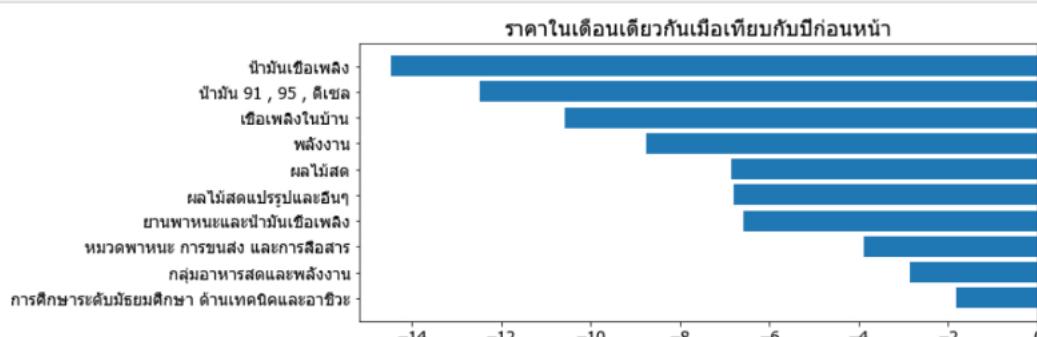
In [143]: #เปลี่ยนเป็นฟอนต์ที่รองรับภาษาไทย ที่ด้องกากัดครองนี้เพราจะได้เปลี่ยนในส่วนของ index_description ที่แสดงในแกน y
plt.rc('font', **{'family': 'Tahoma',
'size': '14'})

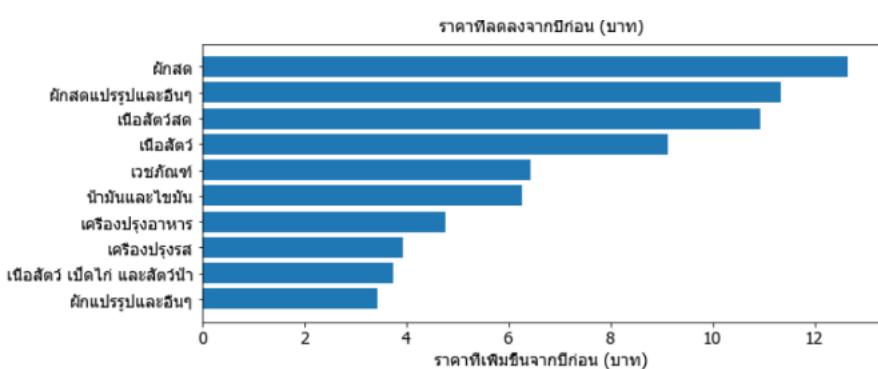
```
plt.figure(figsize=(10, 7))
plt.subplots_adjust(top = 10, bottom=9)

plt.subplot(2,1,1)
plt.title(u'ราคainเดือนเดียวกันเมื่อเทียบกับปีก่อนหน้า', fontname='Tahoma', fontsize='18')

#แกน y ต้องคู่ลิมที่ index_description และแกน x ต้องคู่ลิมที่ yoy ของ decreaseFiveYOY
plt.barh(decreaseTenYOY['index_description'], decreaseTenYOY['yoy'])
plt.xlabel(u'ราคainเดือนเดียวกัน (บาท)')

plt.subplot(2,1,2)
#แกน y ต้องคู่ลิมที่ index_description และแกน x ต้องคู่ลิมที่ yoy ของ increaseFiveYOY
plt.barh(increaseTenYOY['index_description'], increaseTenYOY['yoy'])
plt.xlabel(u'ราคainเดือนเดียวกันเมื่อเทียบกับปีก่อน (บาท)');
```





จาก Bar plot ข้างบนจะเห็นได้ว่าหมวดที่ราคาลดลงจากปีก่อนหน้ามากที่สุดคือ น้ำมันเชื้อเพลิง โดยราคาลดลงไป 14.46 บาทและหมวดที่ราคาเพิ่มขึ้นจากปีก่อนหน้ามากที่สุดคือ ผ้ากันสาด โดยราคาเพิ่มขึ้น 12.66 บาท

กราฟเส้นหมวดที่ราคาลดลงจากปีก่อนมากที่สุด 10 อันดับแรก

Line chart

```
In [144]: #ดึงมาปรับ relatedCategoriesDecrease เก็บค่า index_description ของหมวดที่ราคาลดลงจากปีก่อนมากที่สุด 10 อันดับแรก
#คัดกรองที่ไม่ใช่ใน relatedCategoriesDecrease จะถูกตัดออกที่นี่
relatedCategoriesDecrease = decreaseTenYOY['index_description'].tolist()
for i in relatedCategoriesDecrease:
    print(i)
```

การศึกษาระดับมัธยมศึกษา ดำเนินเทคโนโลยีและอาชีวะ
กลุ่มอาหารสดและพัฒนา
หมวดพาหนะ การขนส่ง และการสื่อสาร
ยานพาหนะและน้ำมันเชื้อเพลิง
ผลไม้สดปรับรูปและอื่นๆ
ผลไม้สด
พลังงาน
เชื้อเพลิงในบ้าน
น้ำมัน 91 , 95 , ดีเซล
น้ำมันเชื้อเพลิง

```
In [145]: #function relatedCategoriesPriceIndexDecrease จะรับค่าตัวเลขเข้ามา
#ตัวเลขตั้งแต่ค่าตัว index ที่เอามาไปใช้ตั้งข้อมูลจากตัวปรับ relatedCategoriesDecrease
#ค่าที่ function นี้จะส่งกลับไปก็คือ price_index
def relatedCategoriesPriceIndexDecrease(index):
    return df[df.index_description == relatedCategoriesDecrease[index]]['price_index'].tolist()
```

```
In [146]: #เปลี่ยนเป็นฟอนต์ที่รองรับภาษาไทย ที่ด้องกับหนังสือที่เราจะใช้เปลี่ยนในส่วนของ index_description ที่แสดงใน Legend
plt.rc('font', **{'family': 'Tahoma',
                  'size' : '14'})

#สร้างตัวแปร month_year เพื่อใช้เก็บค่าเดือน/ปี เพื่อนำไปกำหนดค่าในแกน x
index_ = df[df.index_id == 0]
month = index_[‘month’].tolist()
year = index_[‘year’].tolist()
month_year = []
for i in range(len(month)):
    month_year.append(str(month[i]) + ‘/’ + str(year[i])[2:])

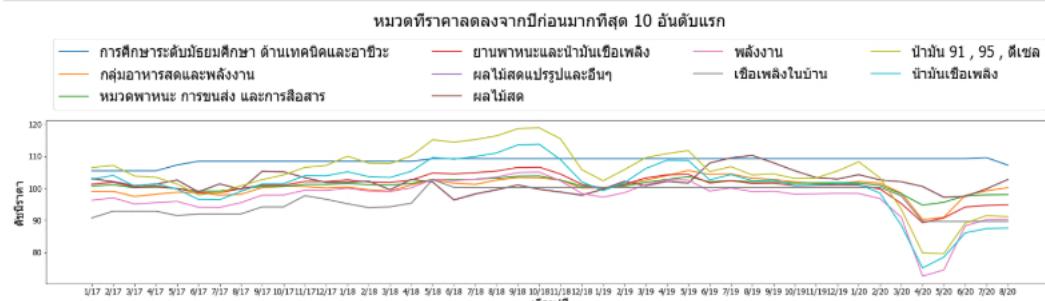
plt.figure(figsize=(30, 5))

#วนลูปตามความยาวของ relatedCategoriesDecrease
for i in range(len(relatedCategoriesDecrease)):
    #สร้างตัวปรับ relatedCategories_i เพื่อกำหนดค่าให้กับ function relatedCategoriesPriceIndexDecrease ลงมา
    relatedCategories_i = relatedCategoriesPriceIndexDecrease(i)
    #พล็อตกราฟเส้นข้อมูลที่ได้รับจากตัวปรับ relatedCategories_i
    plt.plot(month_year, relatedCategories_i, label=relatedCategoriesDecrease[i], linewidth=2)

plt.title(‘หมวดที่ราคาลดลงจากปีก่อนมากที่สุด 10 อันดับแรก’, y=1.55, fontsize=’28’)
plt.xlabel(‘เดือน/ปี’, fontsize=’20’)
plt.ylabel(‘ตัวชี้มีรยา’, fontsize=’20’)

plt.legend(bbox_to_anchor=(0., 1.02, 1., .102), loc=’lower left’,
           ncol=4, mode=’expand’, fontsize=’24’)

plt.show();
```



จากการเเส่นข้างบนจะเห็นว่า ผ่านพ้นนี้เเม้ก็เพิ่ม เสื่อเพลิง, พลังงาน, น้ำมัน 91 95 ตีเซล, กوليอาหารสดและพลังงาน, เสื่อเพลิงในบ้าน, น้ำมันเชื้อเพลิง และหมวดพาหนะ การขนส่ง และการสื่อสาร แนวโน้มของราฝจะไปในทศทางเดียวกัน โดยสินค้าดังกล่าวจะจัดอยู่ในหมวดของเชื้อเพลิง พลังงานและพาหนะ เมื่อถูกกราฟในช่วงโควิดจะเห็นว่าตัวเป็นราคานิ่งในช่วงเวลาดังกล่าวมีค่าลดลงเพราะประชาชนส่วนใหญ่ต้องกักตัวอยู่บ้าน เดินทางน้อยลง ทำงานอยู่บ้าน เหตุผลเหล่านี้ส่งผลให้สินค้านิ่งหมวดดังต้นเป็นที่ต้องการมีอยู่ลงเรื่งท่าให้ราคานิ่งค่าลดลงตามไปด้วย (<https://www.posttoday.com/economy/news/621491>)

กราฟเเส่นหมวดที่ราคามีเพิ่มขึ้นจากปีก่อนมากที่สุด 10 อันดับแรก

Line chart

```
In [147]: #ดูแบบ relatedCategoriesIncrease เป็นค่า index_description ของหมวดที่ราคามีเพิ่มขึ้นจากปีก่อนมากที่สุด 10 อันดับแรก
#ค่าตัวในตัวแบบ relatedCategoriesIncrease จะมีตัวที่เกี่ยวข้องกัน
relatedCategoriesIncrease = increaseTenYOY['index_description'].tolist()
for i in relatedCategoriesIncrease:
    print(i)
```

ผักแปรรูปและอื่นๆ
เมือสัตว์ เปิดไก่ และสัตว์ป่า
เครื่องปรุงรส
เครื่องปัจจุบันอาหาร
น้ำมันและไขมัน
เวชภัณฑ์
เมือสัตว์สลด
ผักสดแปรรูปและอื่นๆ
ผักสด

```
In [148]: #function relatedCategoriesPriceIndexIncrease จะรับค่าตัวเลขเข้ามา
#ดูเลขตั้งกล่าวด้วย index ที่อาไปใช้ตั้งชื่อมาจากการ relatedCategoriesIncrease
#ค่าที่ function นี้จะส่งกลับไปก็คือ price_index
def relatedCategoriesPriceIndexIncrease(index):
    return df[df.index_description == relatedCategoriesIncrease[index]]['price_index'].tolist()
```

```
In [149]: #เปลี่ยนเป็นฟอนต์ที่รองรับภาษาไทย ที่ต้องกำหนดตรงนี้เพื่อเปลี่ยนในส่วนของ index_description ที่แสดงใน legend
plt.rcParams['font', **{'family': 'Tahoma',
                      'size': '14'});

#สร้างตัวแบบ month_year เพื่อใช้เก็บค่าเดือน/ปี เพื่อนำไปกำหนดค่าในแกน x
index_ = df[df.index_id == 0]
month = index_[['month']].tolist()
year = index_[['year']].tolist()
month_year = []
for i in range(len(month)):
    month_year.append(str(month[i]) + '/' + str(year[i])[2:])

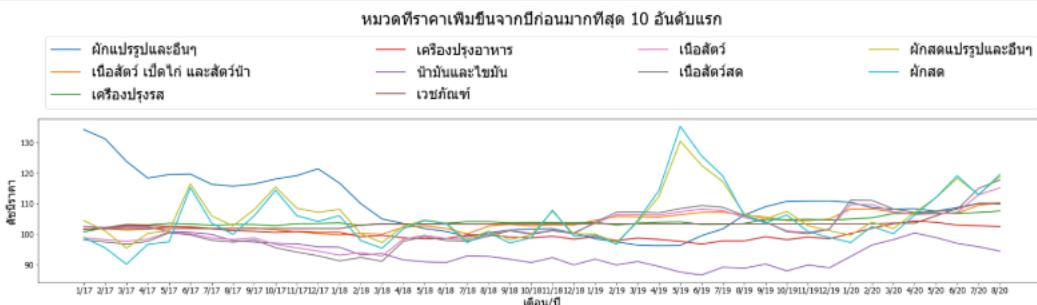
plt.figure(figsize=(30, 5))

#วนลูปตามความหมายของ relatedCategoriesIncrease
for i in range(len(relatedCategoriesIncrease)):
    #สร้างตัวแบบ relatedCategories_i เพื่อกัน price_index ที่ function relatedCategoriesPriceIndexIncrease สำหรับมา
    relatedCategories_j = relatedCategoriesPriceIndexIncrease(i)
    #เพล้อกราฟเพื่อข้อมูลนี้ไม่เรียง ๆ จนกว่าจะจบสูบ
    #แทน x คือเดือน/ปีที่เก็บไว้ในตัวแบบ month_year
    #แทน y คือตัวซึ่งราคานี้เก็บไว้ในตัวแบบ relatedCategoriesDecrease
    #label คือ index_description และกำหนดให้รวมหน้าของสินค้าที่กัน 2
    plt.plot(month_year, relatedCategories_j, label=relatedCategoriesIncrease[i], linewidth=2)

plt.title('หมวดที่ราคามีเพิ่มขึ้นจากปีก่อนมากที่สุด 10 อันดับแรก', y=1.05, fontsize='28')
plt.xlabel('เดือน/ปี', fontsize='20')
plt.ylabel('ตัวเป็นราคานี้', fontsize='20')

plt.legend(bbox_to_anchor=(0., 1.02, 1., .102), loc='lower left',
           ncol=4, mode="expand", fontsize='24')

plt.show();
```



จากการเเส่นข้างบนจะเห็นว่า ผักแปรรูปและอื่นๆ, เครื่องปัจจุบันอาหาร, เมือสัตว์, ผักสดแปรรูปและอื่นๆ, เมือสัตว์ เปิดไก่ และสัตว์ป่า, น้ำมันและไขมัน, เมือสัตว์สลด, ผักสด, เครื่องปรุงรส แนวโน้มของราฝจะไปในทศทางเดียวกัน โดยสินค้าดังกล่าวจะจัดอยู่ในหมวดของอาหาร เมื่อถูกกราฟในช่วงโควิดจะเห็นว่าตัวเป็นราคานิ่งในช่วงเวลาดังกล่าวมีค่าที่เพิ่มสูงขึ้นเพราะประชาชนส่วนใหญ่ต้องกักตัวอยู่บ้าน เดินทางน้อยลง ทำงานอยู่บ้าน เหตุผลเหล่านี้ส่งผลให้สินค้านิ่งหมวดดังต้นเป็นที่ต้องการมีอยู่ลงเรื่งท่าให้ราคานิ่งค่าลดลงตามไปด้วย

แต่จะมีหมวดหนึ่งที่แตกต่างจากพวกต่อ เวชภัณฑ์ (อุปกรณ์ทางการแพทย์) เหตุที่ทำให้หมวดเวชภัณฑ์มีราคากลางๆเพราในช่วงโควิดมีคนต้องการอุปกรณ์ทางการแพทย์เป็นจำนวนมาก ราคานิ่งสินค้าในหมวดนี้จึงสูงขึ้นเช่นกัน

หาสัดส่วนราคานิคัลลดลง/เพิ่มขึ้น/เท่าเดิม

Pie chart

```
In [150]: #ดูว่าปร LenDfNow เก็บค่าขนาดของ Dataframe ในปีจุบัน (เดือน 8 ปี 2020) เพื่อใช้ในการหาสัดส่วน
lenDfNow = len(df[(df.year == 2020) & (df.month == 8)])

#ดูว่าปร percentDecrease เก็บเปอร์เซ็นต์หมวดสินค้าที่มีตัวชี้ว่าราคาลดลง
percentDecrease = round(len(df[(df.yoy < 0) & (df.year == 2020) & (df.month == 8)]) / lenDfNow *100, 3)

#ดูว่าปร percentSame เก็บเปอร์เซ็นต์หมวดสินค้าที่มีตัวชี้ว่าราคากลับเดิม
percentSame = round(len(df[(df.yoy == 0) & (df.year == 2020) & (df.month == 8)]) / lenDfNow*100, 3)

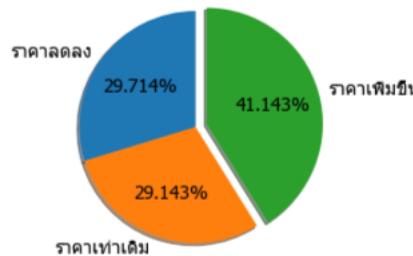
#ดูว่าปร percentIncrease เก็บเปอร์เซ็นต์หมวดสินค้าที่มีตัวชี้ว่าราคายังคงเพิ่มขึ้น
percentIncrease = round(len(df[(df.yoy > 0) & (df.year == 2020) & (df.month == 8)]) / lenDfNow*100, 3)

#คูณ 100 เพื่อแปลงเป็นเปอร์เซ็นต์ และใช้ค่าสิ่ง round เพื่อบรดุกดศนิยมให้เหลือ 3 หลัก
print('ราคานิคัลลดลง: \t', percentDecrease, '%\nราคากลับเดิม: \t', percentSame, '%\nราคายังคงเพิ่มขึ้น: \t', percentIncrease, '%')
print('รวมทั้งหมด: \t', percentDecrease + percentSame + percentIncrease, '%')

ราคานิคัลลดลง: 29.714 %
ราคากลับเดิม: 29.143 %
ราคายังคงเพิ่มขึ้น: 41.143 %
รวมทั้งหมด: 100.0 %
```

```
In [151]: values = [percentDecrease, percentSame, percentIncrease]
labels = ['ราคานิคัลลดลง', 'ราคากลับเดิม', 'ราคายังคงเพิ่มขึ้น']
explode = (0, 0, 0.1) #ทำให้ pie หักออกมานิดหน่อย
plt.title("สัดส่วนอัตราราคาเปลี่ยนแปลงที่ยังกับเดือนเดียวกันปีก่อนหน้า (YOY)", fontname='Tahoma', fontsize='18')
plt.pie(values, explode=explode, labels=labels, shadow=True, autopct='%1.3f%%', startangle=90);
```

สัดส่วนอัตราราคาเปลี่ยนแปลงที่ยังกับเดือนเดียวกันปีก่อนหน้า (YOY)



จาก Pie chart ข้างบนจะเห็นว่าตัวชี้ว่าราคานิคัลลดลงใหญ่กว่าเพิ่มขึ้น ก็จะเป็น 41.143% จากทั้งหมด

อีกสองส่วนที่เหลือคือ ราคานิคัลลดลงและราคากลับเดิม มีสัดส่วนเป็น 29.714% และ 29.143% ตามลำดับ ถือว่ามีสัดส่วนพอ ๆ กัน

สรุปได้ว่าในช่วง 1 ปีที่ผ่านมาตัวชี้ว่าราคานิคัลลดลงหรือคงที่ไปมีแนวโน้มที่เพิ่มสูงขึ้น

วิเคราะห์ดัชนีราคานิคัลลดลง/เพิ่มขึ้น/คงที่

Histogram

```
In [152]: #ดูว่าปร priceIndexYOY เก็บค่าอัตราราคาเปลี่ยนแปลงที่ยังกับเดือนเดียวกันปีก่อนหน้า (เดือน 8 ปี 2020 ที่เทียบกับเดือนเดียวกันในปี 2019)
priceIndexYOY = df[(df.year == 2020) & (df.month == 8)][['yoy']]

#ดูว่าปร priceIndexMON เก็บค่าอัตราราคาเปลี่ยนแปลงที่ยังกับเดือนก่อนหน้า (เดือน 8 ปี 2020 ที่เทียบกับปีเดียวกันในเดือน 7)
priceIndexMON = df[(df.year == 2020) & (df.month == 8)][['mon']]

#ดูว่าปร priceIndexAOA เก็บค่าอัตราราคาเปลี่ยนแปลงที่ยังกับเดือนก่อนหน้า (เดือน 8 ปี 2020 ที่เทียบกับปีเดียวกันในเดือน 8)
priceIndexAOA = df[(df.year == 2020) & (df.month == 8)][['aoa']]

priceIndexList = [priceIndexYOY, priceIndexMON, priceIndexAOA]
priceIndexNameList = ['อัตราราคาเปลี่ยนแปลงราคานิคัลลดลงเดียวกันปีก่อนหน้า (YOY)', 'อัตราราคาเปลี่ยนแปลงราคานิคัลลดลงเดือนก่อนหน้า (MON)', 'อัตราราคาเปลี่ยนแปลงราคานิคัลลดลงเดียวกันปีก่อนหน้า (AOA)']
```

ใช้ Rice Rule ในการหาว่าควรจะกำหนดจำนวน bin เท่าไร

ก คือจำนวนข้อมูล

$$binSize = 2\sqrt{n}^{(1/3)}$$

$$binSize = 2\sqrt{175}^{(1/3)}$$

$$binSize \approx 11$$

```
In [153]: binSize = int(2 * (len(priceIndexYOY)**(1/3)))
binSize
```

Out[153]: 11

```
In [154]: #แสดงความถี่ของราคานิคัลลดลง/เพิ่มขึ้น/คงที่
fig, axes = plt.subplots(3,1, figsize = (6,12))
colorlist = ['#31699b', '#008080', '#108700']
```

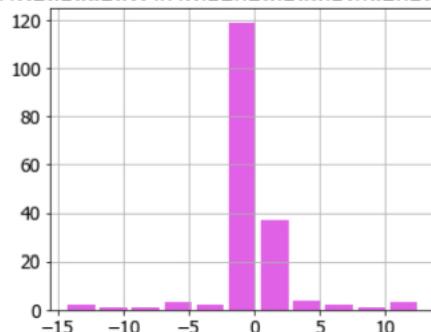
```

for i in range(len(priceIndexList)):
    axes[i].hist(priceIndexList[i], binSize, color=colorList[i], alpha=0.7, rwidth=0.85)
    axes[i].grid(True)
    axes[i].set_title(priceIndexNameList[i])

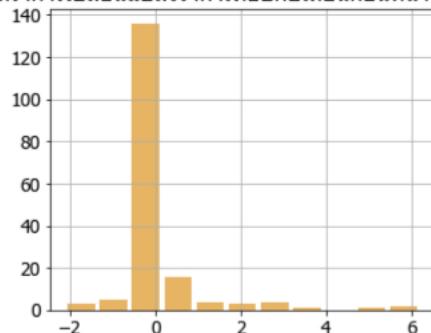
fig.tight_layout()
plt.show();

```

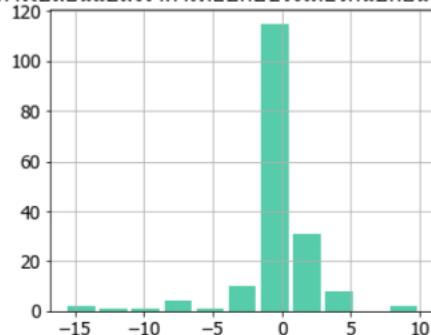
ອັດຕາການປັບປຸງແປ່ງຮາຄາເທິຍນກັນເຕືອນເດືອນເດືອນກັນປີກ່ອນໜ້າ (YOY)



ອັດຕາການປັບປຸງແປ່ງຮາຄາເທິຍນກັນເຕືອນກ່ອນໜ້າ (MON)



ອັດຕາການປັບປຸງແປ່ງຮາຄາເທິຍນກັນປ່າງເດືອນກັນປີກ່ອນໜ້າ (AOA)



ຈາກ Histogram ຂອງອັດຕາການປັບປຸງແປ່ງຮາຄາເທິຍນກັນເຕືອນເດືອນເດືອນກັນປີກ່ອນໜ້າ (YOY) ຈະເຫັນວ່າສັງແມ່ຮາຄາຈະມີການປັບປຸງແປ່ງໃນວ່າງ
ຈະເພີ່ມທີ່ອຸດ ຮາຄາຈະ \pm ໄຟເກີນ 5 ນາທ

ຈາກ Histogram ຂອງອັດຕາການປັບປຸງແປ່ງຮາຄາເທິຍນກັນເຕືອນກ່ອນໜ້າ (MON) ຈະເຫັນວ່າສັງແມ່ຮາຄາຈະມີການປັບປຸງແປ່ງໃນວ່າງເພີ່ມ
ທີ່ອຸດ ຮາຄາຈະ \pm ໄຟເກີນ 2 ນາທ

ຈາກ Histogram ຂອງອັດຕາການປັບປຸງແປ່ງຮາຄາເທິຍນກັນປ່າງເດືອນກັນປີກ່ອນໜ້າ (AOA) ຈະເຫັນວ່າສັງແມ່ຮາຄາຈະມີການປັບປຸງແປ່ງໃນວ່າ
ຈະເພີ່ມທີ່ອຸດ ຮາຄາຈະ \pm ໄຟເກີນ 5 ນາທ

ເນື້ອງກວດວິເຄາະທີ່ໄດ້ຈາກ Pie chart ຂ້າງບັນນາດູຕໍ່ກັນ ຈະເຫັນວ່າໃນປ່າງ 1 ປີທີ່ຜ່ານມາດູຕໍ່ນີ້ຮາຄາຊູນຮົກຄວາມທີ່ໄປສຸວນໃຫຍ່ມີຮາຄາທີ່ເພີ່ມສູງຂຶ້ນໃນ
ປ່າງ 0.1 - 5 ນາທ

ສຽງຜົນ

ຈາກການຝ່າຍທຸກຮາຍການ ຈະເຫັນໄດ້ວ່າເພຣະໜ້າໄວ້ວິດີ-19 ໃນປ່າງຕົ້ນປີ2020 ສົງລົບໄທດີນີ້ຮາຄາຊູນຮົກທີ່ໄປມີແນວໄນ້ມີລົດລົງໃນປ່າງກົງ
ປີປັກແລ້ວຄວາມຝຶ່ນພູ້ຂຶ້ນມາເຮົອຍໆໃນປ່າງເຮົ່ານັ້ນຂອງຄຽງປິ່ງສຳລັບ ຈາກທີ່ກາໄທແລດງຈະເຫັນວ່າເຄຮັດກົງໃນຄຽງປິ່ງທີ່ລົງເຊີ້ນເຊື່ອຍໆ ຖື່ນ
ທ່າກຮາສັດລົນຂອງຮາຄາສິນຄາເພີ່ມ/ອຸດ/ທ່າດິນ ຈະພົບວ່າເນື້ອທີ່ໃນເຕືອນເດືອນກັນປີກ່ອນໜ້າ ສັດສົນຂອງຮາຄາສິນຄາເພີ່ມຂຶ້ນມີສັດສົນທີ່
ນາກທີ່ສຸດ ທີ່ສັດສົນດັ່ງກ່າວແລດງໃຫ້ເຫັນວ່າເຄຮັດກົງຈະກໍາລັງສິ້ນເຊື່ອຍໆ ຈະ

ໃນປ່າງໂຄວັດກົງທີ່ມີທີ່ໜ່ວຍກົດທີ່ຮາຄາເພີ່ມຂຶ້ນແລະລົດລົງ ແໜວດທີ່ຮາຄາສິນຄາເພີ່ມຂຶ້ນຄື່ອມໝວດອາຫານ ໄດ້ແກ່ ຜັກແປຣູບແລະອື່ນາ, ເຄຣິອງປຸງຈາກກາ, ເນື້ອ
ສັດ, ຜັກສົດແປຣູບແລະອື່ນາ, ເນື້ອສັດ ເປີດໄກ ແລະສັດວັນ, ນ້ານັນແລະໃນນັນ, ເນື້ອສັດວັດ, ຜັກສົດ, ເຄຣິອງປຸງຈາກສົດ
ອື່ອມໝວດຈັກກົດຈົບ (ອຸປະກອນທ່າງການແພ່ຍໍ) ແຕ່ກົມົງວັດທີ່ຮາຄາສິນຄາລົດລົງເທິງກົມົງວັດທີ່ອົງເຫຼິນ ພົບລົງຈາກ
ໄດ້ແກ່ ຍານພັນແລະນ້ານັນເຊື້ອເພັນ, ພົບລົງຈາກ, ນ້ານັນ 91 95 ຕີ່ເຊົ່າລົງ, ກຸ່ມອາຫານສົດແລະພົບລົງຈາກ, ເຊື້ອເພັນໃນນັນ, ນ້ານັນເຊື້ອເພັນ ແລະໜ່ວຍ
ກາທັນ ກາຮັນລົງ ແລະກາລົງສາຮາ

ຖື່ນຈະມີທີ່ໜ່ວຍກົດທີ່ຮາຄາເພີ່ມຂຶ້ນແລະລົດລົງແຕ່ເນື້ອເທິຍນດີນີ້ຮາຄາເດືອນເດືອນກັນປີກ່ອນໜ້າ ສິນຄາສົວນໃຫຍ່ມີຮາຄາຈະ \pm ໄຟເກີນ 5 ນາທ

ແນວ່າສິນຄາສົວນໃຫຍ່ຮາຄາ \pm ໄຟເກີນ 5 ນາທແຕ່ກົງມີສິນຄາທີ່ຮາຄາເພີ່ມຂຶ້ນແລະລົດລົງນັກງວ່ານັນເຊັ່ນກັນ ແຊ່ນ

- ໜ່ວຍກົດຮາຄາເພີ່ມຂຶ້ນ 12.66 ນາທ
- ໜ່ວຍກົດນ້ານັນເຊື້ອເພັນ ຮາຄາລົດລົງໄປ 14.46 ນາທ

ปัญหาที่พบ

1. ตัวอักษรภาษาไทยยังมีปัญหาเรื่องสะกด
2. พจนบัญชาаницื่องของการพัฒนาระบบที่ไม่ดี
3. โคดในบางส่วนยังมีการซ้ำกัน อาจจะส่งผลให้รันนานขึ้น
4. ใช้ข้อมูลจากแหล่งเดียว ทำให้ไม่สามารถอ้างอิงแหล่งที่มาได้
5. ไม่เข้าใจคอลัมน์ `aoa` เพราทางผู้ให้บริการให้คำอธิบายที่ไม่ชัดเจน สงสัยว่าทำให้ไปเสียชื่อชื่อ