In [1]:

```
import pandas as pd
from bs4 import BeautifulSoup
import requests, re

cat_list = ['best-selling','vegetables-fruits','breakfast-dairy','beverages',
            'biscuits-snacks-packed-foods','personal-care','household-items']
url_base  = "https://towness-blr.storese.in/collections/"

# Form URLs and iterate through all pages
url_arr=[]
for i in range(len(cat_list)):
    page_url = url_base + cat_list[i]
    url_arr.append(page_url)
    html = requests.get(page_url)
    pages = BeautifulSoup(html.content,'html.parser').find('ul',{'class':'list--inline
 pagination'})
    total = 0
    if pages is not None:
        pages = pages.find('li',{'class':'pagination__text'}).text
        arr = re.split('^[a-zA-Z_\s]+', pages)[1].split()
        for j in range(len(arr)):
            if arr[j].isdigit():
                total+=int(arr[j])
        for k in range(total-2):
            url_arr.append(page_url + '?page=' + str(k+2))
    else:
        continue
for i in range(len(url_arr)):
    print(url_arr[i])
# Now use the url_arr to get all the items in each page
```

https://towness-blr.storese.in/collections/best-selling
https://towness-blr.storese.in/collections/best-selling?page=2
https://towness-blr.storese.in/collections/best-selling?page=3
https://towness-blr.storese.in/collections/best-selling?page=4
https://towness-blr.storese.in/collections/best-selling?page=5
https://towness-blr.storese.in/collections/best-selling?page=6
https://towness-blr.storese.in/collections/best-selling?page=7
https://towness-blr.storese.in/collections/best-selling?page=8
https://towness-blr.storese.in/collections/best-selling?page=9
https://towness-blr.storese.in/collections/best-selling?page=10
https://towness-blr.storese.in/collections/vegetables-fruits
https://towness-blr.storese.in/collections/vegetables-fruits?page=2
https://towness-blr.storese.in/collections/breakfast-dairy
https://towness-blr.storese.in/collections/breakfast-dairy?page=2
https://towness-blr.storese.in/collections/breakfast-dairy?page=3
https://towness-blr.storese.in/collections/beverages
https://towness-blr.storese.in/collections/beverages?page=2
https://towness-blr.storese.in/collections/biscuits-snacks-packed-foods
https://towness-blr.storese.in/collections/biscuits-snacks-packed-foods?pa
ge=2
https://towness-blr.storese.in/collections/biscuits-snacks-packed-foods?pa
ge=3
https://towness-blr.storese.in/collections/personal-care
https://towness-blr.storese.in/collections/personal-care?page=2
https://towness-blr.storese.in/collections/personal-care?page=3
https://towness-blr.storese.in/collections/household-items
https://towness-blr.storese.in/collections/household-items?page=2

In [2]:

```
# Sample scraping from a test page
sale_price=[]
regular_price=[]
names = []
url = "https://towness-blr.storese.in/collections/vegetables-fruits?page=2"
html_soup = BeautifulSoup(requests.get(url).content,'html.parser')
items = html_soup.find('ul',{'class':'grid grid--uniform grid--view-items'})
itemDiv = items.findAll('div',{'class':'grid-view-item product-card'})
for itr in itemDiv:
    # Get the name here and store in names
    name = itr.find('div',{'class':'h4 grid-view-item__title product-card__title'}).tex
t
    print(name)
    names.append(name)
    prices = itr.find('dl',{'class',re.compile('([price price\-\-listing(\.+)])')})
    # regular
    r_price = prices.find('span',{'class':'price-item price-item--regular'}).text
    print(r_price)
    regular_price.append(r_price)
    # sale
    s_price= prices.find('span',{'class':'price-item price-item--sale'}).text
    print(s_price)
    sale_price.append(s_price)

# Create a sample DataFrame
pd.DataFrame({'Product_Name':names, 'Regular_Price':regular_price, 'Sale_Price':sale_pr
ice,})
```

```
Orinko Spinach (Palaksoppu) 200Gms
Rs. 35.00

Rs. 35.00

Bottle Gourd (Sorekayi) Fresh Rush 1pc ( 800 Gms to 1kg )
Rs. 7.50

Rs. 7.50

Coconut Medium (Tenginakayi) Town 1 No
Rs. 21.00

Rs. 21.00
```

Out[2]:

	Product_Name	Regular_Price	Sale_Price
0	Orinko Spinach (Palaksoppu) 200Gms	Rs. 35.00\n	Rs. 35.00\n
1	Bottle Gourd (Sorekayi) Fresh Rush 1pc (800 G...	Rs. 7.50\n	Rs. 7.50\n
2	Coconut Medium (Tenginakayi) Town 1 No	Rs. 21.00\n	Rs. 21.00\n

In [3]:

```
# Search for direct product links, can use this approach to get to all products individ
ually
page = requests.get("https://towness-blr.storese.in/")
html_soup = BeautifulSoup(page.content,'html.parser')
html_soup.find_all('script',{'id':'ProductJson-product-template'})
[k.get('href') for k in items.find_all('a')]
```

Out[3]:

```
['/collections/vegetables-fruits/products/99666',
 '/collections/vegetables-fruits/products/95102',
 '/collections/vegetables-fruits/products/92912']
```