Project Task-1

Write a python program to displayed products count number by products main categories and sub categories. Enter 1 put value link = Yoshops.com Enter 2 put categories link= Output = create excel file with Name Category, Products count,Total products count. After completing task please convert your PY file to EXE file and uploaded both py file and Exe file in mega.io or Google Drive

TASK

In [1]:

```
pip install requests beautifulsoup4 pandas openpyxl
```

```
Requirement already satisfied: requests in d:\users\hp\anaconda3\lib\site-
packages (2.27.1)
Requirement already satisfied: beautifulsoup4 in d:\users\hp\anaconda3\lib
\site-packages (4.11.1)
Requirement already satisfied: pandas in d:\users\hp\anaconda3\lib\site-pa
ckages (1.4.2)
Requirement already satisfied: openpyxl in d:\users\hp\anaconda3\lib\site-
packages (3.0.9)
Requirement already satisfied: certifi>=2017.4.17 in d:\users\hp\anaconda3
\lib\site-packages (from requests) (2023.5.7)
Requirement already satisfied: idna<4,>=2.5 in d:\users\hp\anaconda3\lib\s
ite-packages (from requests) (3.3)
Requirement already satisfied: urllib3<1.27,>=1.21.1 in d:\users\hp\anacon
da3\lib\site-packages (from requests) (1.26.9)
Requirement already satisfied: charset-normalizer~=2.0.0 in d:\users\hp\an
aconda3\lib\site-packages (from requests) (2.0.4)
Requirement already satisfied: soupsieve>1.2 in d:\users\hp\anaconda3\lib
\site-packages (from beautifulsoup4) (2.3.1)
Requirement already satisfied: pytz>=2020.1 in d:\users\hp\anaconda3\lib\s
ite-packages (from pandas) (2021.3)
Requirement already satisfied: numpy>=1.18.5 in d:\users\hp\anaconda3\lib
\site-packages (from pandas) (1.22.4)
Requirement already satisfied: python-dateutil>=2.8.1 in d:\users\hp\anaco
nda3\lib\site-packages (from pandas) (2.8.2)
Requirement already satisfied: et-xmlfile in d:\users\hp\anaconda3\lib\sit
e-packages (from openpyxl) (1.1.0)
Requirement already satisfied: six>=1.5 in d:\users\hp\anaconda3\lib\site-
packages (from python-dateutil>=2.8.1->pandas) (1.16.0)
Note: you may need to restart the kernel to use updated packages.
```

In [6]:

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

# Function to scrape product counts by category
def scrape_product_counts(url):
    # Send an HTTP request to the provided URL
    response = requests.get(url)
    if response.status_code != 200:
        print("Failed to retrieve data from the website.")
        return None

    # Parse the HTML content of the page
    soup = BeautifulSoup(response.content, 'html.parser')

    # Initialize dictionaries to store category and product count data
    category_counts = {}

    # Find and process main categories
    main_categories = soup.find_all('div', class_='main-category')
    for main_category in main_categories:
        main_category_name = main_category.find('h2').text.strip()
        sub_categories = main_category.find_all('div', class_='sub-category')
        total_count = 0
        for sub_category in sub_categories:
            sub_category_name = sub_category.find('h3').text.strip()
            product_count = len(sub_category.find_all('div', class_='product'))
            total_count += product_count
            category_counts[f"{main_category_name} - {sub_category_name}"] = product_cour

        # Add the main category with the total product count
        category_counts[f"{main_category_name} - Total"] = total_count

    return category_counts

# Input URLs
main_url = input("Enter the main website URL: ")
categories_url = input("Enter the categories URL: ")

# Scrape product counts
category_counts = scrape_product_counts(categories_url)

if category_counts:
    # Create a DataFrame from the scraped data
    df = pd.DataFrame(list(category_counts.items()), columns=['Category', 'Products Count

    # Create a new DataFrame with the total product count
    total_count = df[df['Category'].str.contains('Total')]['Products Count'].sum()
    total_df = pd.DataFrame({'Category': ['Total Products'], 'Products Count': [total_cou

    # Concatenate both DataFrames
    df = pd.concat([df, total_df], ignore_index=True)

    # Create an Excel file
    excel_file = 'product_counts.xlsx'
    df.to_excel(excel_file, index=False)

    print(f"Product counts have been saved to {excel_file}.")
else:
```

```
    print("No data to save.")
```

```
Enter the main website URL: https://yoshops.com/ (https://yoshops.com/)
Enter the categories URL: https://yoshops.com/ (https://yoshops.com/)
Failed to retrieve data from the website.
No data to save.
```

In [3]:

```
pip install pyinstaller
```

```
Collecting pyinstallerNote: you may need to restart the kernel to use upda
ted packages.

  Downloading pyinstaller-5.13.1-py3-none-win_amd64.whl (1.3 MB)
Collecting pefile>=2022.5.30
  Downloading pefile-2023.2.7-py3-none-any.whl (71 kB)
Collecting pywin32-ctypes>=0.2.1
  Downloading pywin32_ctypes-0.2.2-py3-none-any.whl (30 kB)
Collecting pyinstaller-hooks-contrib>=2021.4
  Downloading pyinstaller_hooks_contrib-2023.7-py2.py3-none-any.whl (276 k
B)
Collecting altgraph
  Downloading altgraph-0.17.3-py2.py3-none-any.whl (21 kB)
Requirement already satisfied: setuptools>=42.0.0 in d:\users\hp\anaconda3
\lib\site-packages (from pyinstaller) (61.2.0)
Installing collected packages: pywin32-ctypes, pyinstaller-hooks-contrib,
pefile, altgraph, pyinstaller
  Attempting uninstall: pywin32-ctypes
    Found existing installation: pywin32-ctypes 0.2.0
    Uninstalling pywin32-ctypes-0.2.0:
      Successfully uninstalled pywin32-ctypes-0.2.0
Successfully installed altgraph-0.17.3 pefile-2023.2.7 pyinstaller-5.13.1
pyinstaller-hooks-contrib-2023.7 pywin32-ctypes-0.2.2
```

In [ ]:

```

```