EstasyBD Scrape: Scrape All Man Product Items

In [ ]:
```python
# Trial Extraction to validate function

import requests
from bs4 import BeautifulSoup
import pandas as pd
import time
from  urllib.parse import urlparse, urljoin
url = "https://ecstasybd.com/?page=product-list&sid=42"

HEADERS = {
    "Accept": "text/html",
    "Accept-Encoding": "UTF-8",
    "User-Agent": "Mozilla/5.0"


}
response = requests.get(url, headers=HEADERS, timeout=5)
print(response.status_code)
soup = BeautifulSoup(response.text, 'html.parser')
```

200

Extract Page Links

In [7]:
```python
def extract_page_links(html):
    soup = BeautifulSoup(html, "html.parser")

    pagination = soup.find("div", class_="paginatoin-area text-center")
    if not pagination:
        return []

    a_tags = pagination.find_all("a")

    base_url = "https://ecstasybd.com/"
    page_urls = [
        urljoin(base_url, a['href'])
        for a in a_tags
        if a.get("href")
        ]

    return list(dict.fromkeys(page_urls))
```

Extract showcased Product Links

In [9]:
```python
def extract_showcased_product_links(html):
    soup = BeautifulSoup(html, 'html.parser')
    main_wrapper_section = soup.find("div",class_= "shop-main-wrapper section-space
```

```python
    if not main_wrapper_section:
        return []
    all_product_section = main_wrapper_section.find_all("div", class_="col-lg-3 col

    base_url = "https://ecstasybd.com/"
    product_urls = []

    for product in all_product_section:
        product_a_tag = product.find("div", class_="product-item").find("a")
        product_url = urljoin(base_url, product_a_tag['href'])
        product_urls.append(product_url)
    return product_urls
```

In [59]:
```python
extract_showcased_product_links(response.text)
```

Out[59]:
```
['https://ecstasybd.com/?page=product-details&pid=4970',
 'https://ecstasybd.com/?page=product-details&pid=4969',
 'https://ecstasybd.com/?page=product-details&pid=4967',
 'https://ecstasybd.com/?page=product-details&pid=4968',
 'https://ecstasybd.com/?page=product-details&pid=3928',
 'https://ecstasybd.com/?page=product-details&pid=3927',
 'https://ecstasybd.com/?page=product-details&pid=3926',
 'https://ecstasybd.com/?page=product-details&pid=1822',
 'https://ecstasybd.com/?page=product-details&pid=1821',
 'https://ecstasybd.com/?page=product-details&pid=1818']
```

Extract each product Details

In [10]:
```python
def extract_each_product_details(product_details_html_content):
    soup = BeautifulSoup(product_details_html_content, 'html.parser')

    product_detail_page = soup.body.main.find("div", class_="shop-main-wrapper sect
    product_detail_page

    product_quick_details = product_detail_page.find("div", class_ = "product-detai

    related_products = soup.body.main.find("section", class_ = "related-products").

    ## Product Images extraction
    base_url = "https://ecstasybd.com/"

    product_images_links = []

    for link in product_detail_page.find_all("div", class_="pro-large-img img-zoom"
        img_tag = link.find("img")

        if not img_tag or not img_tag.get("src"):
            continue

        img = img_tag["src"]

        if not img.startswith("http"):
            img = base_url + img
```

```python
        product_images_links.append(img)

    product_details = {
        "product-name": product_quick_details.find("h3", class_ = "product-name").g

        "price-regular" : product_quick_details.find("span", class_="price-regular

        "sizes" : product_quick_details.table.find_all("td")[1].get_text(strip=True

        "availability" : product_quick_details.find("span", class_ = "product-att a

        "SKU" : product_quick_details.find("div", class_="sku-wrap half-wrap").find

        "PID" : product_quick_details.find("div", class_="pid-container").find("spa

        "related-PID" : list(set([pid['href'].split("pid=")[1] for pid in related_p

        "info_guide" : product_quick_details.find("div", class_="product-review-inf

        "product-images-links": product_images_links

    }
    return product_details
```

Scrool to botton

```python
In [11]:  # Scrool_to_bottom function
          def scroll_to_bottom(driver):
              last_height = driver.execute_script("return document.body.scrollHeight")

              while True:
                  driver.execute_script("window.scrollTo(0, document.body.scrollHeight);")
                  time.sleep(2)  # allow content to load

                  new_height = driver.execute_script("return document.body.scrollHeight")

                  if new_height == last_height:
                      break

                  last_height = new_height
```

Product Navigation

```python
In [54]:  import requests
          from bs4 import BeautifulSoup
          import pandas as pd
          from  urllib.parse import urlparse, urljoin
          from selenium import webdriver
          from selenium.webdriver.common.by import By
          from selenium.webdriver.support.ui import WebDriverWait
          from selenium.webdriver.support import expected_conditions as EC
          from bs4 import BeautifulSoup
```

```python
import time
import random


# -----------------------------
# Start Driver
# -----------------------------
driver = webdriver.Chrome()
wait = WebDriverWait(driver, 10)

try:
    driver.get("https://ecstasybd.com/")

    # Click "Men"
    catg_btn = wait.until(
        EC.element_to_be_clickable((By.XPATH, '//*[@id="tris-bottom-link-1"]'))
    )
    catg_btn.click()

    # Click subcategory
    subcat_btn = wait.until(
        EC.element_to_be_clickable((By.CSS_SELECTOR, '//*[@id="tris-bottom-link-4"]
    )
    subcat_btn.click()
    ## Use this below for polo category
    # driver.get("https://ecstasybd.com/?page=product-list&sid=3")

    time.sleep(2)

    # Scroll main  page
    scroll_to_bottom(driver)

    # Extract  shirt page links
    all_page_links = extract_page_links(driver.page_source)

    all_product_details = []

    # Loop through pagination pages
    for page_link in all_page_links:
        try:
            driver.get(page_link)
            time.sleep(random.uniform(2, 4))

            scroll_to_bottom(driver)

            all_showcase_product_links = extract_showcased_product_links(
                driver.page_source
            )

            # Loop through product links
            for product_link in all_showcase_product_links:
                try:
                    driver.get(product_link)
                    time.sleep(random.uniform(2, 5))

                    product_details_info = extract_each_product_details(
```

```python
                            driver.page_source
                        )

                        all_product_details.append(product_details_info)

                except Exception as e:
                    print(f"XXXX Error in product: {product_link}")
                    print(e)
                    continue

            except Exception as e:
                print(f"XXXX Error in page: {page_link}")
                print(e)
                continue

        print("Scraping completed successfully.")
        print(f"Total products collected: {len(all_product_details)}")

finally:
    driver.quit()
```

```
Scraping completed successfully.
Total products collected: 35
```

In [55]:
```python
all_product_details
```

```
Out[55]:  [{'product-name': 'TANJIM PREMIUM POLO',
           'price-regular': 'TK. 3498',
           'sizes': 'M, L, XL, 2XL,',
           'availability': 'In Stock',
           'SKU': '10004149',
           'PID': 'PI6822',
           'related-PID': ['6821', '6796'],
           'info_guide': 'Model typically wears\xa0Size L\xa0(tops) and\xa033" waist\xa0(bo
          ttoms)\nModel is wearing\xa0Size M\nTrue to size\xa0fit\nModel Measurements\nHeigh
          t:\xa06'\nChest:\xa039"\nShoulder:\xa020"',
           'product-images-links': ['https://ecstasybd.com/all-images/product/Product-Image
          -1769679834.jpg',
            'https://ecstasybd.com/all-images/product/Product-Image-11769679834.jpg',
            'https://ecstasybd.com/all-images/product/Product-Image-21769679834.jpg',
            'https://ecstasybd.com/all-images/product/Product-Image-31769679834.PNG']},
          {'product-name': 'T SQUAD POLO T-SHIRT - BLACK',
           'price-regular': 'TK. 3741',
           'sizes': 'S, L,',
           'availability': 'In Stock',
           'SKU': '10003723',
           'PID': 'PI6820',
           'related-PID': ['6816', '6799'],
           'info_guide': 'Polo\nSleeve: Short\nSquare Bottom\nColor: Black\nRegular Fit\nCo
          mposition: 100% Cotton\nThe model Sergio usually wears a size L top.\nModel Heigh
          t: 180 / Waist: 83 / Chest: 103 / Size: L',
           'product-images-links': ['https://ecstasybd.com/all-images/product/Product-Image
          -1769584935.jpg',
            'https://ecstasybd.com/all-images/product/Product-Image-11769584935.jpg',
            'https://ecstasybd.com/all-images/product/Product-Image-21769584935.jpg',
            'https://ecstasybd.com/all-images/product/Product-Image-31769595347.PNG']},
          {'product-name': 'TANJIM PREMIUM POLO',
           'price-regular': 'TK. 3498',
           'sizes': 'M, L, XL, 2XL,',
           'availability': 'In Stock',
           'SKU': '10004148',
           'PID': 'PI6817',
           'related-PID': ['6770'],
           'info_guide': 'Model typically wears\xa0Size L\xa0(tops) and\xa033" waist\xa0(bo
          ttoms)\nModel is wearing\xa0Size M\nTrue to size\xa0fit\nModel Measurements\nHeigh
          t:\xa06'\nChest:\xa039"\nShoulder:\xa020"',
           'product-images-links': ['https://ecstasybd.com/all-images/product/Product-Image
          -1769501152.jpg',
            'https://ecstasybd.com/all-images/product/Product-Image-11769501152.jpg',
            'https://ecstasybd.com/all-images/product/Product-Image-21769501152.jpg',
            'https://ecstasybd.com/all-images/product/Product-Image-31769501152.PNG']},
          {'product-name': 'TANJIM PREMIUM POLO',
           'price-regular': 'TK. 3608',
           'sizes': 'M, XL, 2XL, L,',
           'availability': 'In Stock',
           'SKU': '10004178',
           'PID': 'PI6816',
           'related-PID': ['6773', '6770'],
           'info_guide': 'Model typically wears\xa0Size L\xa0(tops) and\xa033" waist\xa0(bo
          ttoms)\nModel is wearing\xa0Size M\nTrue to size\xa0fit\nModel Measurements\nHeigh
          t:\xa06'\nChest:\xa039"\nShoulder:\xa020"',
           'product-images-links': ['https://ecstasybd.com/all-images/product/Product-Image
```

```
-1769500809.jpg',
   'https://ecstasybd.com/all-images/product/Product-Image-11769500809.jpg',
   'https://ecstasybd.com/all-images/product/Product-Image-21769500809.jpg',
   'https://ecstasybd.com/all-images/product/Product-Image-31769500809.PNG']},
{'product-name': 'TANJIM PREMIUM POLO',
 'price-regular': 'TK. 3608',
 'sizes': 'XL,',
 'availability': 'In Stock',
 'SKU': '10004178',
 'PID': 'PI6815',
 'related-PID': ['6773', '6796'],
 'info_guide': 'Model typically wears\xa0Size L\xa0(tops) and\xa033" waist\xa0(bo
ttoms)\nModel is wearing\xa0Size M\nTrue to size\xa0fit\nModel Measurements\nHeigh
t:\xa06'\nChest:\xa039"\nShoulder:\xa020"',
 'product-images-links': ['https://ecstasybd.com/all-images/product/Product-Image
-1769500423.jpg',
   'https://ecstasybd.com/all-images/product/Product-Image-11769500423.jpg',
   'https://ecstasybd.com/all-images/product/Product-Image-21769500423.jpg',
   'https://ecstasybd.com/all-images/product/Product-Image-31769500423.PNG']},
{'product-name': 'TANJIM PREMIUM POLO',
 'price-regular': 'TK. 3498',
 'sizes': 'M, L, XL, 2XL,',
 'availability': 'In Stock',
 'SKU': '10004148',
 'PID': 'PI6814',
 'related-PID': ['6800', '6813'],
 'info_guide': 'Model typically wears\xa0Size L\xa0(tops) and\xa033" waist\xa0(bo
ttoms)\nModel is wearing\xa0Size M\nTrue to size\xa0fit\nModel Measurements\nHeigh
t:\xa06'\nChest:\xa039"\nShoulder:\xa020"',
 'product-images-links': ['https://ecstasybd.com/all-images/product/Product-Image
-1769500045.jpg',
   'https://ecstasybd.com/all-images/product/Product-Image-11769500045.jpg',
   'https://ecstasybd.com/all-images/product/Product-Image-21769500045.jpg',
   'https://ecstasybd.com/all-images/product/Product-Image-31769501031.PNG']},
{'product-name': 'TANJIM PREMIUM POLO',
 'price-regular': 'TK. 3498',
 'sizes': 'M, L, XL, 2XL,',
 'availability': 'In Stock',
 'SKU': '10004148',
 'PID': 'PI6806',
 'related-PID': ['6801', '6796'],
 'info_guide': 'Model typically wears\xa0Size L\xa0(tops) and\xa033" waist\xa0(bo
ttoms)\nModel is wearing\xa0Size M\nTrue to size\xa0fit\nModel Measurements\nHeigh
t:\xa06'\nChest:\xa039"\nShoulder:\xa020"',
 'product-images-links': ['https://ecstasybd.com/all-images/product/Product-Image
-1769420881.jpg',
   'https://ecstasybd.com/all-images/product/Product-Image-11769420881.jpg',
   'https://ecstasybd.com/all-images/product/Product-Image-21769420881.jpg',
   'https://ecstasybd.com/all-images/product/Product-Image-31769420881.PNG']},
{'product-name': 'TANJIM PREMIUM POLO',
 'price-regular': 'TK. 3608',
 'sizes': 'M, XL, 2XL,',
 'availability': 'In Stock',
 'SKU': '10004178',
 'PID': 'PI6804',
 'related-PID': ['6797', '6800'],
```

```
    {'product-name': 'TANJIM PREMIUM POLO - PRIVATE LABEL',
     'price-regular': 'TK. 3487',
     'sizes': 'M, L,',
     'availability': 'In Stock',
     'SKU': '10003917',
     'PID': 'PI6110',
     'related-PID': [],
     'info_guide': '•\tColor: Deep Forest Green\n•\tStyle: Classic Fit Polo\n•\tDesig
n: Solid color design with minimal red accent detailing along the button placket f
or a refined touch.\n•\tMaterial: Premium cotton pique fabric – soft, breathable,
and durable.\n•\tCollar: Ribbed polo collar with a two-button placket.\n•\tSleeve
s: Short sleeves with a standard hem finish.\n•\tFit: Regular fit for all-day comf
ort and smart appearance.\n•\tCare Instructions: Machine washable – gentle cycle r
ecommended for long-lasting color and fabric quality.',
     'product-images-links': ['https://ecstasybd.com/all-images/product/Product-Image
-1745912053.jpg',
      'https://ecstasybd.com/all-images/product/Product-Image-11745912053.jpg',
      'https://ecstasybd.com/all-images/product/Product-Image-21745912053.jpg',
      'https://ecstasybd.com/all-images/product/Product-Image-31745912053.jpeg']},
    {'product-name': 'TANJIM POLO',
     'price-regular': 'TK. 1298',
     'sizes': 'XL, XXL,',
     'availability': 'In Stock',
     'SKU': '10002924',
     'PID': 'PI3546',
     'related-PID': [],
     'info_guide': 'Polo\nTanjim Logo in front\nSleeve: Short\nColor: Red\nSlim Fit\n
Cotton Fabric',
     'product-images-links': ['https://ecstasybd.com/all-images/product/Product-Image
-1684929142.jpg',
      'https://ecstasybd.com/all-images/product/Product-Image-11684929142.jpg',
      'https://ecstasybd.com/all-images/product/Product-Image-21684929142.png']}]
```

```python
In [56]:  df = pd.DataFrame(all_product_details)
          df.head()
```

Out[56]:

	product-name	price-regular	sizes	availability	SKU	PID	related-PID	info_guide	pr
0	TANJIM PREMIUM POLO	TK. 3498	M, L, XL, 2XL,	In Stock	10004149	PI6822	[6821, 6796]	Model typically wears Size L (tops) and 33" wa...	[https:/ ima
1	T SQUAD POLO T-SHIRT - BLACK	TK. 3741	S, L,	In Stock	10003723	PI6820	[6816, 6799]	Polo\nSleeve: Short\nSquare Bottom\nColor: Bla...	[https:/ ima
2	TANJIM PREMIUM POLO	TK. 3498	M, L, XL, 2XL,	In Stock	10004148	PI6817	[6770]	Model typically wears Size L (tops) and 33" wa...	[https:/ ima
3	TANJIM PREMIUM POLO	TK. 3608	M, XL, 2XL, L,	In Stock	10004178	PI6816	[6773, 6770]	Model typically wears Size L (tops) and 33" wa...	[https:/ ima
4	TANJIM PREMIUM POLO	TK. 3608	XL,	In Stock	10004178	PI6815	[6773, 6796]	Model typically wears Size L (tops) and 33" wa...	[https:/ ima

◄ ▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬ ►

In [57]:
```python
df.to_excel("Polo.xlsx", index=False)
```

For each Subcategory I changed X-path then i changed the file name to save