Lab Assignment- PDV

Scraped the website - https://shekouwoman.com/collections/home

To run Scrapy, install the libraries and create a folder to store the files needed for the scrapy project. The command 'scrapy startproject' creates the required .py files and spider to scrape the website.

```
Anaconda Powershell Prompt (anaconda3)
(base) PS C:\Users\aditi> cd C:\Users\aditi\OneDrive\Desktop\ScrapyAssignment
(base) PS C:\Users\aditi\OneDrive\Desktop\ScrapyAssignment> scrapy startproject clothshopping
New Scrapy project 'clothshopping', using template directory 'C:\Users\aditi\anaconda3\lib\site-packages\scrapy\templates\project', created in:
    C:\Users\aditi\OneDrive\Desktop\ScrapyAssignment\clothshopping

You can start your first spider with:
    cd clothshopping
    scrapy genspider example example.com
(base) PS C:\Users\aditi\OneDrive\Desktop\ScrapyAssignment> cd clothshopping
(base) PS C:\Users\aditi\OneDrive\Desktop\ScrapyAssignment\clothshopping> cd clothshopping
(base) PS C:\Users\aditi\OneDrive\Desktop\ScrapyAssignment\clothshopping\clothshopping> ls


    Directory: C:\Users\aditi\OneDrive\Desktop\ScrapyAssignment\clothshopping\clothshopping


Mode                LastWriteTime         Length Name
----                -------------         ------ ----
d-----        23-12-2021     16:19                spiders
-a----        28-01-2022     23:49            269 items.py
-a----        28-01-2022     23:49           3662 middlewares.py
-a----        28-01-2022     23:49            367 pipelines.py
-a----        28-01-2022     23:49           3134 settings.py
-a----        12-10-2020     17:32              0 __init__.py


(base) PS C:\Users\aditi\OneDrive\Desktop\ScrapyAssignment\clothshopping\clothshopping>
```

Scraping the website directly in the shell:

The 'scrapy shell' command is used to scrape the website on the shell.

```
Anaconda Powershell Prompt (anaconda3)
(base) PS C:\Users\aditi\OneDrive\Desktop\ScrapyAssignment\clothshopping\clothshopping> scrapy shell "https://shekouwoman.com/collections/home"
2022-01-28 23:51:00 [scrapy.utils.log] INFO: Scrapy 2.4.0 started (bot: clothshopping)
2022-01-28 23:51:00 [scrapy.utils.log] INFO: Versions: lxml 4.6.3.0, libxml2 2.9.10, cssselect 1.1.0, parsel 1.5.2, w3lib 1.21.0, Twisted 20.3.0, Python 3.8.8 (default, Apr
 13 2021, 15:08:03) [MSC v.1916 64 bit (AMD64)], pyOpenSSL 20.0.1 (OpenSSL 1.1.1k  25 Mar 2021), cryptography 3.4.7, Platform Windows-10-10.0.22000-SP0
2022-01-28 23:51:00 [scrapy.utils.log] DEBUG: Using reactor: twisted.internet.selectreactor.SelectReactor
2022-01-28 23:51:00 [scrapy.crawler] INFO: Overridden settings:
{'BOT_NAME': 'clothshopping',
 'DUPEFILTER_CLASS': 'scrapy.dupefilters.BaseDupeFilter',
 'LOGSTATS_INTERVAL': 0,
 'NEWSPIDER_MODULE': 'clothshopping.spiders',
 'ROBOTSTXT_OBEY': True,
 'SPIDER_MODULES': ['clothshopping.spiders']}
2022-01-28 23:51:00 [scrapy.extensions.telnet] INFO: Telnet Password: 296616c9baa65c3e
2022-01-28 23:51:00 [scrapy.middleware] INFO: Enabled extensions:
['scrapy.extensions.corestats.CoreStats',
 'scrapy.extensions.telnet.TelnetConsole']
2022-01-28 23:51:00 [scrapy.middleware] INFO: Enabled downloader middlewares:
['scrapy.downloadermiddlewares.robotstxt.RobotsTxtMiddleware',
 'scrapy.downloadermiddlewares.httpauth.HttpAuthMiddleware',
 'scrapy.downloadermiddlewares.downloadtimeout.DownloadTimeoutMiddleware',
 'scrapy.downloadermiddlewares.defaultheaders.DefaultHeadersMiddleware',
 'scrapy.downloadermiddlewares.useragent.UserAgentMiddleware',
 'scrapy.downloadermiddlewares.retry.RetryMiddleware',
 'scrapy.downloadermiddlewares.redirect.MetaRefreshMiddleware',
 'scrapy.downloadermiddlewares.httpcompression.HttpCompressionMiddleware',
 'scrapy.downloadermiddlewares.redirect.RedirectMiddleware',
 'scrapy.downloadermiddlewares.cookies.CookiesMiddleware',
 'scrapy.downloadermiddlewares.httpproxy.HttpProxyMiddleware',
 'scrapy.downloadermiddlewares.stats.DownloaderStats']
2022-01-28 23:51:00 [scrapy.middleware] INFO: Enabled spider middlewares:
['scrapy.spidermiddlewares.httperror.HttpErrorMiddleware',
 'scrapy.spidermiddlewares.offsite.OffsiteMiddleware',
 'scrapy.spidermiddlewares.referer.RefererMiddleware',
 'scrapy.spidermiddlewares.urllength.UrlLengthMiddleware',
 'scrapy.spidermiddlewares.depth.DepthMiddleware']
2022-01-28 23:51:00 [scrapy.middleware] INFO: Enabled item pipelines:
[]
2022-01-28 23:51:00 [scrapy.extensions.telnet] INFO: Telnet console listening on 127.0.0.1:6023
2022-01-28 23:51:00 [scrapy.core.engine] INFO: Spider opened
2022-01-28 23:51:00 [scrapy.core.engine] DEBUG: Crawled (200) <GET https://shekouwoman.com/robots.txt> (referer: None)
Show desktop
```

The 'response.css' command is used to along with the extract() to extract the tag we want and the relevant information.

```
Anaconda Powershell Prompt (anaconda3)
In [1]: response.css('title::text').extract()
Out[1]:
['New Arrivals | Shekou Woman',
 'American Express',
 'Apple Pay',
 'Google Pay',
 'JCB',
 'Mastercard',
 'Shop Pay',
 'Visa']
```

When there are multiple similar tags can use the dot operator along with the class name of tag to refer a particular tag for example response.css('div.product-details-container').get().

```
Anaconda Powershell Prompt (anaconda3)
In [6]: response.css('div.product--details-container').get()
Out[6]: '<div class="product--details-container">\n      <div class="product--details">\n        <div class="product--details-wrapper">\n          \n\n          <p class="p
roduct--title" data-item="paragraph">\n            PRE-ORDER: On Your Mark Sweatpants- Mens\n          </p>\n\n          \n\n          \n        </div>\n\n        \n  <div
class="product--price-container">\n    <div class="product--price-wrapper"><span class="product--compare-price money" data-item="light-accent-text">\n          Rs. 3,100.00
\n        </span><span class="product--price money" data-item="accent-text">\n        Rs. 1,600.00\n      </span>\n    </div>\n\n    <div class="product--unit-price-contain
er" data-item="small-text"></div>\n  </div>\n\n      </div>\n\n      \n      \n    </div>'
```

All the product details that is the name actual price and the price after the discount is stored in a variable called product for easy access.

```
In [11]: product=response.css('div.product--details-container')

In [12]: len(product)
Out[12]: 48
```

We use the variable for further accessing the details stored in it.

```
In [15]: product.css('p::text').get()
Out[15]: '\n            PRE-ORDER: On Your Mark Sweatpants- Mens\n

In [16]:
```

The get () method is used to return the first instance of the data we want to scrape.

```
Anaconda Powershell Prompt (anaconda3)

In [16]: product.css('span::text').get()
Out[16]: '\n          Rs. 3,100.00\n        '

In [17]: product.css('span.product--price.money::text').get()
Out[17]: '\n        Rs. 1,600.00\n      '

In [18]:
```

Similarly, we can use the getall () method to return all the data of a particular tag.

Anaconda Powershell Prompt (anaconda3)

```
In [22]: product.css('p.product--title::text').getall()
Out[22]:
['\n            PRE-ORDER: On Your Mark Sweatpants- Mens\n          ',
 '\n            Jeans Bundle (4 ITEMS)\n          ',
 '\n            My Baby Blue Sweatpants\n          ',
 '\n            Bikini Bundle (8 Separates)\n          ',
 '\n            Game Player Jeans\n          ',
 '\n            Kyoto Jeans\n          ',
 '\n            My Baby Blue Hoodie\n          ',
 '\n            Tropicana Motel Sweatpants\n          ',
 '\n            Tropicana Motel Sweatshirt\n          ',
 '\n            PRE-ORDER: In the Woodlands Sweatpants\n          ',
 '\n            In the Woodlands Hoodie\n          ',
 '\n            Jeans Bundle (4 ITEMS)\n          ',
 '\n            PRE-ORDER: Fleece Lined Hoodie\n          ',
 '\n            Sweatpants + Hoodie Bundle - Grey (2 ITEMS)\n          ',
 "\n            I'm Not In The Mood Bundle (4 ITEMS)\n          ",
 '\n            Hoodie & Sweatpants Bundle (2 ITEMS)\n          ',
 '\n            Sweatpants Mega Bundle (9 ITEMS)\n          ',
 '\n            Soft Girl Bundle (4 ITEMS)\n          ',
 '\n            Beachy Bundle (5 Items)\n          ',
 '\n            Indie Bundle (5 ITEMS)\n          ',
 '\n            Astrology Bundle (4 ITEMS)\n          ',
 '\n            Unlock Your Dreams Phone Chain\n          ',
 '\n            Running On Time Necklace\n          ',
 "\n            90's Bundle (4 ITEMS)\n          ",
 '\n            Centre Of Attention Necklace\n          ',
 '\n            Key West Kitten Bundle (5 ITEMS)\n          ',
 '\n            Danish Pastel Bundle (5 ITEMS)\n          ',
 '\n            Sweatpants Bundle (3 ITEMS)\n          ',
 '\n            PRE-ORDER: Sweatpants Bundle (3 ITEMS)\n          ',
 '\n            PRE-ORDER: Sweatpants Bundle (3 ITEMS)\n          ',
 '\n            Sweatpants Bundle (3 ITEMS)\n          ',
 '\n            PRE-ORDER: Angel Aura Halter Top\n          ',
 "\n            PRE-ORDER: Midsummer Night's Dream Dress\n          ",
 "\n            PRE-ORDER: Don't @ Me Hoodie\n          ",
 "\n            PRE-ORDER: There's No Better Time Hoodie\n          ",
 '\n            PRE-ORDER: Collect Moments Hoodie\n          ',
 '\n            PRE-ORDER: Take What You Need Hoodie\n          ',
In [23]:
```

Anaconda Powershell Prompt (anaconda3)

```
In [21]: product.css('span.product--price.money::text').getall()
Out[21]:
['\n        Rs. 1,600.00\n      ',
 '\n        Rs. 5,500.00\n      ',
 '\n        Rs. 1,600.00\n      ',
 '\n        Rs. 5,000.00\n      ',
 '\n        Rs. 1,400.00\n      ',
 '\n        Rs. 1,400.00\n      ',
 '\n        Rs. 2,700.00\n      ',
 '\n        Rs. 1,600.00\n      ',
 '\n        Rs. 2,700.00\n      ',
 '\n        Rs. 1,600.00\n      ',
 '\n        Rs. 2,700.00\n      ',
 '\n        Rs. 5,400.00\n      ',
 '\n        Rs. 2,700.00\n      ',
 '\n        Rs. 3,900.00\n      ',
 '\n        Rs. 4,900.00\n      ',
 '\n        Rs. 3,900.00\n      ',
 '\n        Rs. 13,100.00\n      ',
 '\n        Rs. 3,100.00\n      ',
 '\n        Rs. 3,500.00\n      ',
 '\n        Rs. 7,300.00\n      ',
 '\n        Rs. 5,400.00\n      ',
 '\n        Rs. 200.00\n      ',
 '\n        Rs. 300.00\n      ',
 '\n        Rs. 3,500.00\n      ',
 '\n        Rs. 300.00\n      ',
 '\n        Rs. 3,100.00\n      ',
 '\n        Rs. 5,400.00\n      ',
 '\n        Rs. 4,300.00\n      ',
 '\n        Rs. 4,300.00\n      ',
 '\n        Rs. 4,300.00\n      ',
 '\n        Rs. 4,300.00\n      ',
 '\n        Rs. 1,400.00\n      ',
 '\n        Rs. 1,400.00\n      ',
 '\n        Rs. 2,700.00\n      ',
 '\n        Rs. 2,700.00\n      ',
 '\n        Rs. 2,700.00\n      ',
 '\n        Rs. 2,700.00\n      ',
 '\n        Rs. 1,600.00\n      ',
```

We can extract the links using the following command.

Anaconda Powershell Prompt (anaconda3)

```
In [26]: products.css('a').attrib['href']
Out[26]: '/collections/home/products/pre-order-on-your-mark-sweatpants-mens'

In [27]:
```

To scrape the website using spider:

First, we need to create a .py file in the spider's folder and then this folder can be used to scrape the website.

Shopcloth.py

```python
# -*- coding: utf-8 -*-
"""
Created on Sat Jan 29 01:19:35 2022

@author: aditi
"""

import scrapy

class ShopclothSpider(scrapy.Spider):
    name='Clothes'
    start_urls=['https://shekouwoman.com/collections/home']

    def parse(self, response):
        title=response.css('title::text').extract()
        yield{'titleset':title}
        for products in response.css('div.product--details-container'):
            yield {
                'name': products.css('p::text').get(),
                'real_price':products.css('span.product--compare-price.money::text').get(),
                'discount_price':products.css('span.product--price.money::text').get()
                    }
```

The command 'scrapy crawl spidername' is run on the command prompt to execute the spider. Once the data is scraped the command scrapy crawl spidername -o json/csv is used to save the scraped data in the json/csv format.

Shopcloth.py items.json


```
[
{"titleset": ["New Arrivals | Shekou Woman", "American Express", "Apple Pay", "Google Pay", "JCB", "Mastercard", "Sho
{"name": "\n            PRE-ORDER: On Your Mark Sweatpants- Mens\n          ", "real_price": "\n          Rs. 3,100.0
{"name": "\n            Jeans Bundle (4 ITEMS)\n          ", "real_price": "\n          Rs. 10,700.00\n          ", "di
{"name": "\n            My Baby Blue Sweatpants\n          ", "real_price": null, "discount_price": "\n          Rs. 1,
{"name": "\n            Bikini Bundle (8 Separates)\n          ", "real_price": "\n          Rs. 10,100.00\n          "
{"name": "\n            Game Player Jeans\n          ", "real_price": "\n          Rs. 2,700.00\n          ", "discount
{"name": "\n            Kyoto Jeans\n          ", "real_price": "\n          Rs. 2,700.00\n          ", "discount_price
{"name": "\n            My Baby Blue Hoodie\n          ", "real_price": "\n          Rs. 5,400.00\n          ", "discou
{"name": "\n            Tropicana Motel Sweatpants\n          ", "real_price": "\n          Rs. 3,100.00\n          ",
{"name": "\n            Tropicana Motel Sweatshirt\n          ", "real_price": "\n          Rs. 5,400.00\n          ",
{"name": "\n            PRE-ORDER: In the Woodlands Sweatpants\n          ", "real_price": "\n          Rs. 3,100.00\
{"name": "\n            In the Woodlands Hoodie\n          ", "real_price": "\n          Rs. 5,400.00\n          ", "di
{"name": "\n            Jeans Bundle (4 ITEMS)\n          ", "real_price": "\n          Rs. 10,700.00\n          ", "di
{"name": "\n            PRE-ORDER: Fleece Lined Hoodie\n          ", "real_price": "\n          Rs. 5,400.00\n
{"name": "\n            Sweatpants + Hoodie Bundle - Grey (2 ITEMS)\n          ", "real_price": "\n          Rs. 8,40
{"name": "\n            I'm Not In The Mood Bundle (4 ITEMS)\n          ", "real_price": "\n          Rs. 10,100.00\n
{"name": "\n            Hoodie & Sweatpants Bundle (2 ITEMS)\n          ", "real_price": "\n          Rs. 8,400.00\n
{"name": "\n            Sweatpants Mega Bundle (9 ITEMS)\n          ", "real_price": "\n          Rs. 24,100.00\n
{"name": "\n            Soft Girl Bundle (4 ITEMS)\n          ", "real_price": "\n          Rs. 7,400.00\n          ",
{"name": "\n            Beachy Bundle (5 Items)\n          ", "real_price": "\n          Rs. 7,400.00\n          ", "di
{"name": "\n            Indie Bundle (5 ITEMS)\n          ", "real_price": "\n          Rs. 15,200.00\n          ", "di
{"name": "\n            Astrology Bundle (4 ITEMS)\n          ", "real_price": "\n          Rs. 11,600.00\n          ",
{"name": "\n            Unlock Your Dreams Phone Chain\n          ", "real_price": "\n          Rs. 300.00\n          "
{"name": "\n            Running On Time Necklace\n          ", "real_price": null, "discount_price": "\n          Rs. 3
{"name": "\n            90's Bundle (4 ITEMS)\n          ", "real_price": "\n          Rs. 7,600.00\n          ", "disc
{"name": "\n            Centre Of Attention Necklace\n          ", "real_price": null, "discount_price": "\n          R
{"name": "\n            Key West Kitten Bundle (5 ITEMS)\n          ", "real_price": "\n          Rs. 6,400.00\n
{"name": "\n            Danish Pastel Bundle (5 ITEMS)\n          ", "real_price": "\n          Rs. 11,300.00\n
{"name": "\n            Sweatpants Bundle (3 ITEMS)\n          ", "real_price": "\n          Rs. 9,200.00\n          ",
{"name": "\n            PRE-ORDER: Sweatpants Bundle (3 ITEMS)\n          ", "real_price": "\n          Rs. 9,200.00\
{"name": "\n            PRE-ORDER: Sweatpants Bundle (3 ITEMS)\n          ", "real_price": "\n          Rs. 9,200.00\
{"name": "\n            Sweatpants Bundle (3 ITEMS)\n          ", "real_price": "\n          Rs. 9,200.00\n          ",
{"name": "\n            PRE-ORDER: Angel Aura Halter Top\n          ", "real_price": "\n          Rs. 2,700.00\n
```

items - Notepad

File Edit Format View Help


```
[
{"titleset": ["New Arrivals | Shekou Woman", "American Express", "Apple Pay", "Google Pay", "JCB", "Mastercard", "Shop Pay", "Visa"]},
{"name": "\n            PRE-ORDER: On Your Mark Sweatpants- Mens\n          ", "real_price": "\n          Rs. 3,100.00\n        ", "discount_price": "\n
{"name": "\n            Jeans Bundle (4 ITEMS)\n          ", "real_price": "\n          Rs. 10,700.00\n        ", "discount_price": "\n        Rs. 5,500.00\n
{"name": "\n            My Baby Blue Sweatpants\n          ", "real_price": null, "discount_price": "\n        Rs. 1,600.00\n      "},
{"name": "\n            Bikini Bundle (8 Separates)\n          ", "real_price": "\n          Rs. 10,100.00\n        ", "discount_price": "\n        Rs. 5,000
{"name": "\n            Game Player Jeans\n          ", "real_price": "\n          Rs. 2,700.00\n        ", "discount_price": "\n        Rs. 1,400.00\n
{"name": "\n            Kyoto Jeans\n          ", "real_price": "\n          Rs. 2,700.00\n        ", "discount_price": "\n        Rs. 1,400.00\n      "},
{"name": "\n            My Baby Blue Hoodie\n          ", "real_price": "\n          Rs. 5,400.00\n        ", "discount_price": "\n        Rs. 2,700.00\n
{"name": "\n            Tropicana Motel Sweatpants\n          ", "real_price": "\n          Rs. 3,100.00\n        ", "discount_price": "\n        Rs. 1,600.0
{"name": "\n            Tropicana Motel Sweatshirt\n          ", "real_price": "\n          Rs. 5,400.00\n        ", "discount_price": "\n        Rs. 2,700.0
{"name": "\n            PRE-ORDER: In the Woodlands Sweatpants\n          ", "real_price": "\n          Rs. 3,100.00\n        ", "discount_price": "\n
{"name": "\n            In the Woodlands Hoodie\n          ", "real_price": "\n          Rs. 5,400.00\n        ", "discount_price": "\n        Rs. 2,700.00\n
{"name": "\n            Jeans Bundle (4 ITEMS)\n          ", "real_price": "\n          Rs. 10,700.00\n        ", "discount_price": "\n        Rs. 5,400.00\n
{"name": "\n            PRE-ORDER: Fleece Lined Hoodie\n          ", "real_price": "\n          Rs. 5,400.00\n        ", "discount_price": "\n        Rs. 2,7
{"name": "\n            Sweatpants + Hoodie Bundle - Grey (2 ITEMS)\n          ", "real_price": "\n          Rs. 8,400.00\n        ", "discount_price": "\n
{"name": "\n            I'm Not In The Mood Bundle (4 ITEMS)\n          ", "real_price": "\n          Rs. 10,100.00\n        ", "discount_price": "\n
{"name": "\n            Hoodie & Sweatpants Bundle (2 ITEMS)\n          ", "real_price": "\n          Rs. 8,400.00\n        ", "discount_price": "\n        R
{"name": "\n            Sweatpants Mega Bundle (9 ITEMS)\n          ", "real_price": "\n          Rs. 24,100.00\n        ", "discount_price": "\n        Rs.
{"name": "\n            Soft Girl Bundle (4 ITEMS)\n          ", "real_price": "\n          Rs. 7,400.00\n        ", "discount_price": "\n        Rs. 3,100.0
{"name": "\n            Beachy Bundle (5 Items)\n          ", "real_price": "\n          Rs. 7,400.00\n        ", "discount_price": "\n        Rs. 3,500.00\n
{"name": "\n            Indie Bundle (5 ITEMS)\n          ", "real_price": "\n          Rs. 15,200.00\n        ", "discount_price": "\n        Rs. 7,300.00\n
{"name": "\n            Astrology Bundle (4 ITEMS)\n          ", "real_price": "\n          Rs. 11,600.00\n        ", "discount_price": "\n        Rs. 5,400.
{"name": "\n            Unlock Your Dreams Phone Chain\n          ", "real_price": "\n          Rs. 300.00\n        ", "discount_price": "\n        Rs. 200.0
{"name": "\n            Running On Time Necklace\n          ", "real_price": null, "discount_price": "\n        Rs. 300.00\n      "},
{"name": "\n            90's Bundle (4 ITEMS)\n          ", "real_price": "\n          Rs. 7,600.00\n        ", "discount_price": "\n        Rs. 3,500.00\n
{"name": "\n            Centre Of Attention Necklace\n          ", "real_price": null, "discount_price": "\n        Rs. 300.00\n      "},
{"name": "\n            Key West Kitten Bundle (5 ITEMS)\n          ", "real_price": "\n          Rs. 6,400.00\n        ", "discount_price": "\n        Rs. 3
{"name": "\n            Danish Pastel Bundle (5 ITEMS)\n          ", "real_price": "\n          Rs. 11,300.00\n        ", "discount_price": "\n        Rs. 5,
{"name": "\n            Sweatpants Bundle (3 ITEMS)\n          ", "real_price": "\n          Rs. 9,200.00\n        ", "discount_price": "\n        Rs. 4,300.
{"name": "\n            PRE-ORDER: Sweatpants Bundle (3 ITEMS)\n          ", "real_price": "\n          Rs. 9,200.00\n        ", "discount_price": "\n
{"name": "\n            PRE-ORDER: Sweatpants Bundle (3 ITEMS)\n          ", "real_price": "\n          Rs. 9,200.00\n        ", "discount_price": "\n
{"name": "\n            Sweatpants Bundle (3 ITEMS)\n          ", "real_price": "\n          Rs. 9,200.00\n        ", "discount_price": "\n        Rs. 4,300.
{"name": "\n            PRE-ORDER: Angel Aura Halter Top\n          ", "real_price": "\n          Rs. 2,700.00\n        ", "discount_price": "\n        Rs. 1
```


Ln 1, Col 1 | 100% | Unix (LF) | UTF-8

To iterate through the next pages of the website –

Used to scrape the subsequent pages.

```
next_page=response.css('a.pagination--link')[2].attrib['href']
if next_page is not None:
    yield response.follow(next_page, callback = self.parse)
```

Using the Items.py folder as a container –

The entire website can be scraped and placed into the items.py file which acts like a container. From this file we can then display the results or use this file as an intermediary and send the information to the pipeline and store it in database or display directly.

Shopcloth.py | items.py | pipelines.py | settings.py

```
# Define here the models for your scraped items
#
# See documentation in:
# https://docs.scrapy.org/en/latest/topics/items.html

import scrapy


class ClothshoppingItem(scrapy.Item):
    # define the fields for your item here like:
    name = scrapy.Field()
    real_price= scrapy.Field()
    discount_price= scrapy.Field()
   # pass
```

```python
"""

import scrapy

from ..items import ClothshoppingItem

class ShopclothSpider(scrapy.Spider):

    name='Clothes'
    start_urls=['https://shekouwoman.com/collections/home']

    def parse(self, response):
        #title=response.css('title::text').extract()
        #yield{'titleset':title}
        item=ClothshoppingItem()

        for products in response.css('div.product--details-container'):

            item['name']=products.css('p::text').get(),
            item['real_price']=products.css('span.product--compare-price.money::text').get(),
            item['discount_price']=products.css('span.product--price.money::text').get()

            yield item
```

Using Pipelines-

We use pipeline to push the scraped item into the output file/database or to perform any pre-processing steps by making use of middleware.

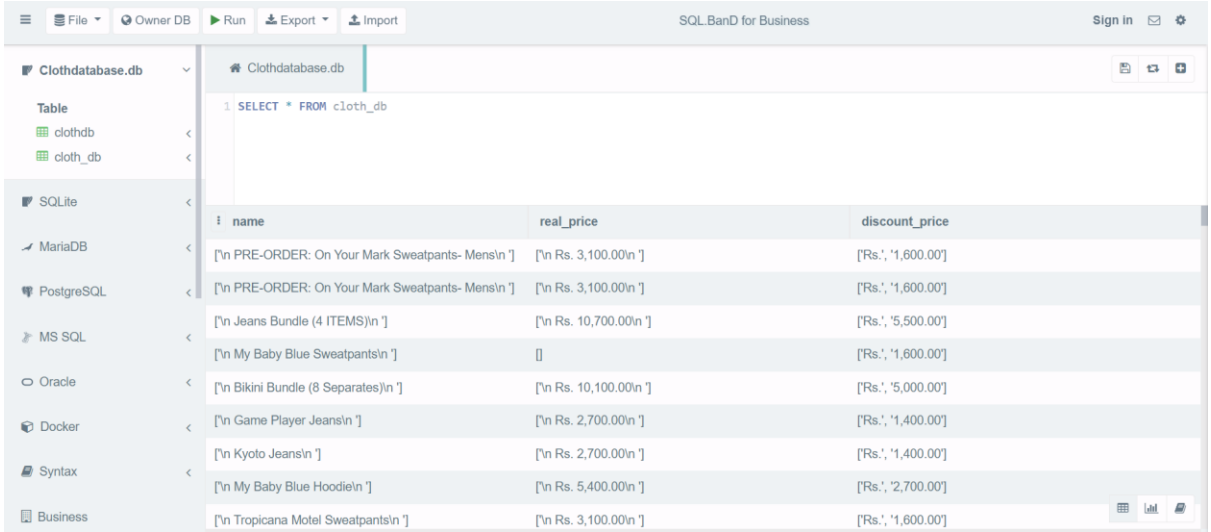
In the settings.py file, uncomment the code for “Item_Pipeline”-

```python

# Enable or disable extensions
# See https://docs.scrapy.org/en/latest/topics/extensions.html
#EXTENSIONS = {
#    'scrapy.extensions.telnet.TelnetConsole': None,
#}

# Configure item pipelines
# See https://docs.scrapy.org/en/latest/topics/item-pipeline.html
ITEM_PIPELINES = {
    'clothshopping.pipelines.ClothshoppingPipeline': 300,
}
```

In the pipelines.py file, we define classes to create the database –

```python
# See: https://docs.scrapy.org/en/latest/topics/item-pipeline.html


# useful for handling different item types with a single interface
import sqlite3
from itemadapter import ItemAdapter


class ClothshoppingPipeline:
    def __init__(self):
        self.create_connection()
        self.create_table()

    def create_connection(self):
        self.con = sqlite3.connect('Clothdatabase.db')
        self.cur = self.con.cursor()

    def create_table(self):
        self.cur.execute("""  DROP TABLE IF EXISTS Clothdb""")
        self.cur.execute("""  CREATE TABLE Clothdb(
            name TEXT
            real_price REAL
            discount_price REAL)""")

    def process_item(self, item, spider):
        self.strore_db(item)
        return item

    def store_db(self,item):
        self.cur.execute("""INSERT INTO Clothdb VALUES(?,?,?)""",
                         (item['name'],item['real_price'],item['discount_price']))
        self.con.commit()
        return item
```

Using sqliteonline.com to run and display the database



The database cloth_db is created with the columns -name, real_price and discount_price. when we use pipelines to scrape the data, Sqliteonline.com can then be used to view the scraped data.

To create an offline database sqlitebrowser.com is used and the database can be downloaded directly from the website.