capstone-project

July 16, 2024

- **Name - Kailash Konkal**
- **Project - Capstone Clothing Brand Project**

[1]:
```
# import nessesary library
import selenium
from selenium.webdriver.common.by import By
import time
import pandas as pd
from selenium import webdriver
import warnings
warnings.filterwarnings('ignore')
from bs4 import BeautifulSoup
from selenium.common.exceptions import StaleElementReferenceException,␣
 ↪NoSuchElementException
import requests
import re
```

Brand - Old Navy

[2]:
```
#open the wikipedia page on automated chrome browser
driver = webdriver.Chrome()
```

[3]:
```
driver.get("https://oldnavy.gap.com/")
```

[4]:
```
women = driver.find_element(By.XPATH,'//div[@class="sitewide-1ivvj99"]/a[1]')
women.click()
```

[5]:
```
# 500 time we scroll down by 10000 in order to generate more product on the␣
 ↪website.
print("scroll page for generate more products")
for _ in range(500):
    driver.execute_script("window.scrollBy(0,10000)")
```

```
scroll page for generate more products
```

[6]:
```
Description=[]
Price = []
```

[11]:
```
# scraping the description of product
des = driver.find_elements(By.XPATH,'//div[@class="category-page-1r1wcud"]')
for a in des:
    Description.append(a.text)
# scraping the price of product
prices = driver.find_elements(By.XPATH,'//
 ↪div[@class="product-price__highlight"]')
for b in prices:
    Price.append(b.text)
```

[12]:
```
#print len
print(len(Description),len(Price))
```

```
300 300
```

[13]:
```
brand = ['Old Navy'] * len(Description)
category = ['Womens Wear'] * len(Description)
```

[14]:
```
Old_Navy = pd.DataFrame({})
Old_Navy['Brand1']= brand
Old_Navy['Price1'] = Price
Old_Navy['Description1']= Description
Old_Navy['Category1']= category
Old_Navy
```

[14]:

	Brand1	Price1	Description1	Category1
0	Old Navy	$10.47	Loose Satin Button-Down Shirt	Womens Wear
1	Old Navy	$15.47	High-Waisted OG Loose Cotton-Hemp Blend Jeans	Womens Wear
2	Old Navy	$6.47	Luxe Tunic T-Shirt	Womens Wear
3	Old Navy	$8.97	Loose Crepe Button-Down Shirt	Womens Wear
4	Old Navy	$14.97	Tie-Shoulder Underwire One-Piece Swimsuit	Womens Wear
..	…	…	…	…
295	Old Navy	$22.97	High-Waisted OG Loose Faux-Leather Pants	Womens Wear
296	Old Navy	$4.96	Ribbed Halter Top	Womens Wear
297	Old Navy	$27.47	High-Waisted OG Loose Faux-Leather Pants	
298	Old Navy	$13.47	High-Waisted Pixie Mini Skirt	
299	Old Navy	$4.96	Rib-Knit Crop Tank Top	

```
297  Womens Wear
298  Womens Wear
299  Womens Wear

[300 rows x 4 columns]
```

[15]:
```
driver.close()
```

Brand - Allen Solley

[16]:
```
#open the allen solley page on automated chrome browser
driver = webdriver.Chrome()
```

[17]:
```
driver.get("https://allensolly.abfrl.in/")
```

[18]:
```
search = driver.find_element(By.XPATH,'//*[@id="headerMenu"]/div/div/div[3]/
 ↪div[1]/div/input')
search.send_keys("Kids")
```

[20]:
```
click = driver.find_element(By.XPATH,'//p[@class="MuiTypography-root␣
 ↪MuiTypography-body1 Header_searchresult__Tvqzn css-12a9y8i"]')
click.click()
```

[21]:
```
load_more = driver.find_element(By.XPATH,'//*[@id="__next"]/main/div/div[4]/div/
 ↪div[2]/button')
load_more.click()
```

[22]:
```
Brand = []
Description = []
Price = []
Color = []
Category = []
```

[23]:
```
des = driver.find_elements(By.XPATH,'//
 ↪div[@class="ProductCard_description__BQzle"]')
for a in des:
    Description.append(a.text)
# scraping the price of product
prices = driver.find_elements(By.XPATH,'//span[@class="price"]')
for b in prices:
    Price.append(b.text)
```

[24]:
```
Brand = ['Allen Solley'] * len(Description)
Category = ['Kids Wear'] * len(Description)
```

[25]:
```
As = pd.DataFrame({})
As['Brand2']= Brand
```

```
As['Price2'] = Price
As['Description2']= Description
As['Category2']= Category
As
```

[25]:

```
            Brand2   Price2                                        Description2  \
0    Allen Solley    1,399                          Boys Blue Regular Fit Jeans
1    Allen Solley    1,999           Boys Black Regular Fit Solid Casual Shirt
2    Allen Solley    1,399                          Boys Navy Regular Fit Jeans
3    Allen Solley    2,099      Girls Pink Embellished Regular Fit Party Dress
4    Allen Solley    2,299  Girls Purple Embellished Regular Fit Casual Dress
..            …       …                                                    …
539  Allen Solley    1,699                Boys Navy Slim Fit Check Casual Shirt
540  Allen Solley    2,999  Boys Black Solid Regular Fit Sweatshirt and Tr…
541  Allen Solley    1,299      Boys Black Regular Fit Print Crew Neck T-shirt
542  Allen Solley    1,499                   Boys Navy Regular Fit Solid Shorts
543  Allen Solley      899              Girls Purple Embellished Casual T-shirt

     Category2
0    Kids Wear
1    Kids Wear
2    Kids Wear
3    Kids Wear
4    Kids Wear
..         …
539  Kids Wear
540  Kids Wear
541  Kids Wear
542  Kids Wear
543  Kids Wear

[544 rows x 4 columns]
```

[26]:

```
driver.close()
```

Brand-Biba

[27]:

```
#open the biba page on automated chrome browser
driver = webdriver.Chrome()
```

[28]:

```
driver.get("https://www.biba.in/")
```

[29]:

```
product = driver.find_element(By.XPATH,'//*[@id="new-arrival"]')
product.click()
```

[30]:

```
Brand = []
Description = []
```

```
Price = []
Category = []
```

[38]:
```
des = driver.find_elements(By.XPATH,'//h3[@class="pdp-link"]')
for a in des:
    Description.append(a.text)
# scraping the price of product
prices = driver.find_elements(By.XPATH,'//span[@class="sales"]')
for b in prices:
    Price.append(b.text)
```

[39]:
```
# print len
print(len(Description),len(Price))
```

```
1164 1164
```

[40]:
```
Brand = ['Biba'] * len(Description)
Category = ['ladies Wear'] * len(Description)
```

[41]:
```
biba = pd.DataFrame({})
biba['Brand3']= Brand
biba['Price3'] = Price
biba['Description3']= Description
biba['Category3']= Category
biba
```

[41]:

	Brand3	Price3	Description3	Category3
0	Biba	4,999	Blue Polyester Straight Suit Set	ladies Wear
1	Biba	2,999	Beige Rayon Straight Suit Set	ladies Wear
2	Biba	2,999	Ice Blue Rayon Straight Suit Set	ladies Wear
3	Biba	3,999	White Pink Rayon Gathered Suit Set	ladies Wear
4	Biba	2,999	Turquoise Rayon Straight Suit Set	ladies Wear
…	…	…	…	…
1159	Biba	1,299	Denim Blue Polyester Straight Printed Kurti	ladies Wear
1160	Biba	1,599	Black Rayon Straight Printed Kurta	
1161	Biba	1,599	Mustard Rayon Straight Printed Kurta	
1162	Biba	1,999	Burnt Orange Poly Viscose Straight Printed Top	
1163	Biba	2,599	Black Polyester Straight Printed Top	

```
1160  ladies Wear
1161  ladies Wear
1162  ladies Wear
1163  ladies Wear

[1164 rows x 4 columns]
```

[35]:
```
driver.close()
```

Brand - FabIndia

[49]:
```
#open the fab india page on automated chrome browser
driver = webdriver.Chrome()
```

[50]:
```
driver.get("https://www.fabindia.com/")
```

[51]:
```
search = driver.find_element(By.XPATH,'//div[@class="search-icon"]')
search.click()
```

[52]:
```
inputt = driver.find_element(By.XPATH,'//*[@id="searchBox"]')
inputt.send_keys("Mens")
```

[54]:
```
select = driver.find_element(By.XPATH,'//span[@class="highlight fw-bold"]')
select.click()
```

[55]:
```
Brand = []
Description = []
Price = []
Category = []
```

[56]:
```
des = driver.find_elements(By.XPATH,'//a[@class="text-decoration-none␣
 ↪text-muted text-secondary"]')
for a in des:
    Description.append(a.text)
# scraping the price of product
prices = driver.find_elements(By.XPATH,'//span[@class="offer-amount"]')
for b in prices:
    Price.append(b.text)
```

[57]:
```
# print len
print(len(Description),len(Price))
```

```
1325 1325
```

[58]:
```
Brand = ['Fab India'] * len(Description)
Category = ['Mens Wear'] * len(Description)
```

[59]:
```
fab = pd.DataFrame({})
fab['Brand4']= Brand
fab['Price4'] = Price
fab['Description4']= Description
fab['Category4']= Category
fab
```

[59]:

	Brand4	Price4	Description4	Category4
0	Fab India	899	White Cotton Short Kurta	Mens Wear
1	Fab India	2,999	Beige Silk Checks Regular Shirt	Mens Wear
2	Fab India	999	Natural Cotton Woven Short Kurta	Mens Wear
3	Fab India	999	Pink Cotton Woven Short Kurta	Mens Wear
4	Fab India	1,599	White Cotton Dobby Slim Fit Long Kurta	Mens Wear
…	…	…	…	…
1320	Fab India	1,899	Pink Cotton Slim Fit Geometric Long Kurta	Mens Wear
1321	Fab India	1,899	Blue Cotton Slim Fit Geometric Long Kurta	Mens Wear
1322	Fab India	2,299	Beige Cotton Shirt	Mens Wear
1323	Fab India	1,699	Black Cotton Printed Slim Fit Shirt	Mens Wear
1324	Fab India	2,499	Red Cotton Printed Slim Fit Long Kurta	Mens Wear

[1325 rows x 4 columns]

[60]:
```
driver.close()
```

Brand - Levis

[61]:
```
#open the levis page on automated chrome browser
driver = webdriver.Chrome()
```

[62]:
```
driver.get("https://levi.in/")
```

[63]:
```
search = driver.find_element(By.XPATH,'//*[@id="Details-HeaderMenu-3"]/summary')
search.click()
```

[64]:
```
Brand = []
Description = []
Price = []
Category = []
```

[65]:
```
des = driver.find_elements(By.XPATH,'//a[@class="product-card__title␣
 ↪tw-text-[12px] tw-min-h-[3rem] md:tw-text-[14px] lg:tw-mb-[8px]␣
 ↪tw-line-clamp-2 tw-tracking-normal tw-leading-normal tw-text-[#333]␣
 ↪tw-font-primary tw-font-normal"]')
for a in des:
    Description.append(a.text)
# scraping the price of product
```

```
prices = driver.find_elements(By.XPATH,'//div[@class="tw-text-[12px] lg:
 ↪tw-text-[14px] tw-flex tw-flex-wrap tw-items-center tw-min-h-[2rem]"]//
 ↪span[1]')
for b in prices:
    Price.append(b.text)
```

[66]:
```
# print len
print(len(Description),len(Price))
```

```
785 785
```

[67]:
```
Brand = ['Levis'] * len(Description)
Category = ['Womens Wear'] * len(Description)
```

[68]:
```
levis = pd.DataFrame({})
levis['Brand5']= Brand
levis['Price5'] = Price
levis['Description5']= Description
levis['Category5']= Category
levis
```

[68]:

	Brand5	Price5	Description5	Category5
0	Levis	949	Women's Striped Slim Fit T-shirt Purple	Womens Wear
1	Levis	1,749	Women's Striped Round Neck Top Coral	Womens Wear
2	Levis	1,699	Women's Solid Spread Collar Shirt Green	Womens Wear
3	Levis	1,549	Women's Solid Spread Collar Shirt Yellow	Womens Wear
4	Levis	1,149	Women's Solid Round Neck Top Green	Womens Wear
..	…	…	…	…
780	Levis	1,649	Women's 711 Skinny Fit Jeans	Womens Wear
781	Levis	1,249	Women's 711 Skinny Fit Jeans	Womens Wear
782	Levis	1,649	Women's 711 Skinny Fit Jeans	Womens Wear
783	Levis	2,999	Women's 711 Skinny Fit Jeans	Womens Wear
784	Levis	2,999	Women's High RiseWide Leg Trousers	Womens Wear

[785 rows x 4 columns]

[69]:
```
driver.close()
```

Brand - Banetton

[70]:
```
#open the Monte Carlo page on automated chrome browser
driver = webdriver.Chrome()
```

[71]:
```
driver.get("https://in.benetton.com/")
```

[72]:
```
kids = driver.find_element(By.XPATH,'//*[@id="kids-menu-trigger"]')
kids.click()
```

[73]:
```
boys = driver.find_element(By.XPATH,'//*[@id="maincontent"]/div[2]/div/div[1]/
  ↪div/div/div/div/div/div/div[2]/div/div[2]/a[1]')
boys.click()
```

[74]:
```
Brand = []
Description = []
Price = []
Category = []
```

[75]:
```
# scraping the description of product
des = driver.find_elements(By.XPATH,'//div[@class="pdp-link"]')
for b in des:
    Description.append(b.text)
# scraping the price of product
prices = driver.find_elements(By.XPATH,'//span[@class="value"]')
for a in prices:
    Price.append(a.text)
```

[76]:
```
# print len
print(len(Description),len(Price))
```

```
106 106
```

[77]:
```
Brand = ['Banetton'] * len(Description)
Category = ['Kids Wear'] * len(Description)
```

[78]:
```
bane = pd.DataFrame({})
bane['Brand6']= Brand
bane['Price6'] = Price
bane['Description6']= Description
bane['Category6']= Category
bane
```

[78]:

	Brand6	Price6	Description6	Category6
0	Banetton	799.00	REGULAR FIT ROUND NECK PRINTED T-SHIRT	Kids Wear
1	Banetton	1,299.00	REGULAR FIT ROUND NECK STRIPED T-SHIRT	Kids Wear
2	Banetton	1,299.00	REGULAR FIT ROUND NECK PRINTED T-SHIRT	Kids Wear
3	Banetton	1,499.00	REGULAR FIT SPREAD COLLAR CHECKERED SHIRT	Kids Wear
4	Banetton	1,999.00	REGULAR FIT SPREAD COLLAR CHECKERED SHIRT	Kids Wear
..	…	…	…	…
101	Banetton	999.00	REGULAR FIT POLO NECK SOLID T-SHIRT	Kids Wear
102	Banetton	1,299.00	REGULAR FIT ROUND NECK PRINTED T-SHIRT	Kids Wear
103	Banetton	999.00	REGULAR FIT POLO NECK SOLID T-SHIRT	Kids Wear
104	Banetton	1,999.00	REGULAR FIT MANDARIN CHECKERED SHIRT	Kids Wear
105	Banetton	1,499.00	REGULAR FIT SPREAD COLLAR TIE & DYE SHIRT	Kids Wear

[106 rows x 4 columns]

[79]:
```
girls = driver.find_element(By.XPATH,'//*[@id="maincontent"]/div[2]/div/div[1]/
  ↪div/div/div/div/div/div/div[2]/div/div[2]/a[2]')
girls.click()
```

[80]:
```
Brand = []
Description = []
Price = []
Category = []
```

[82]:
```
# scraping the description of product
des = driver.find_elements(By.XPATH,'//h2[@class="link"]')
for b in des:
    Description.append(b.text)
# scraping the price of product
prices = driver.find_elements(By.XPATH,'//span[@class="value"]')
for a in prices:
    Price.append(a.text)
```

[83]:
```
# print len
print(len(Description),len(Price))
```

```
106 106
```

[84]:
```
Brand = ['Banetton'] * len(Description)
Category = ['Kids Wear'] * len(Description)
```

[85]:
```
ban = pd.DataFrame({})
ban['Brand7']= Brand
ban['Price7'] = Price
ban['Description7']= Description
ban['Category7']= Category
ban
```

[85]:

	Brand7	Price7	Description7	Category7
0	Banetton	2,499.00	REGULAR FIT V-NECK PRINTED A-LINE DRESS	Kids Wear
1	Banetton	1,299.00	REGULAR FIT POLO NECK PRINTED TOP	Kids Wear
2	Banetton	1,999.00	REGULAR FIT ROUND NECK SOLID A-LINE DRESS	Kids Wear
3	Banetton	1,999.00	SOLID MID RISE CULLOTTES	Kids Wear
4	Banetton	1,799.00	REGULAR FIT POLO NECK PRINTED DRESS	Kids Wear
..	…	…	…	…
101	Banetton	1,799.00	REGULAR FIT SPREAD COLLAR PRINTED TOP	Kids Wear
102	Banetton	1,299.00	REGULAR FIT ROUND NECK SEQUINED T-SHIRT	Kids Wear
103	Banetton	2,799.00	SOLID STRAIGHT FIT FIT MID RISE CARGO	Kids Wear
104	Banetton	999.00	REGULAR FIT ROUND NECK SOLID TOP	Kids Wear
105	Banetton	999.00	REGULAR FIT ROUND NECK GRAPHIC PRINT TOP	Kids Wear

[106 rows x 4 columns]

[86]:
```
driver.close()
```

Brand - Peter England

[87]:
```
#open the fab india page on automated chrome browser
driver = webdriver.Chrome()
```

[88]:
```
driver.get("https://peterengland.abfrl.in/")
```

[91]:
```
search = driver.find_element(By.XPATH,'//*[@id="headerMenu"]/div/div/div[3]/
 ↪div[1]/div/input')
search.send_keys("Formals")
```

[92]:
```
click = driver.find_element(By.XPATH,'//p[@class="MuiTypography-root␣
 ↪MuiTypography-body1 Header_searchresult__Tvqzn css-1bcywis"]')
click.click()
```

[93]:
```
Brand = []
Description = []
Price = []
Category = []
```

[94]:
```
# scraping the description of product
des = driver.find_elements(By.XPATH,'//
 ↪div[@class="ProductCard_description__BQzle"]')
for b in des:
    Description.append(b.text)
# scraping the price of product
prices = driver.find_elements(By.XPATH,'//
 ↪div[@class="ProductCard_productPrice__lHsjc productPrice"]')
for a in prices:
    Price.append(a.text)
```

[95]:
```
# print len
print(len(Description),len(Price))
```

```
288 288
```

[96]:
```
Brand = ['Peter England'] * len(Description)
Category = ['Mens Wear'] * len(Description)
```

[97]:
```
Pe = pd.DataFrame({})
Pe['Brand8']= Brand
Pe['Price8'] = Price
Pe['Description8']= Description
Pe['Category8']= Category
Pe
```

[97]:
```
             Brand8   Price8  \
0    Peter England    1,799
1    Peter England    2,499
2    Peter England    1,799
3    Peter England    1,599
4    Peter England    1,799
..             …      …
283  Peter England    1,599
284  Peter England    2,499
285  Peter England    1,399
286  Peter England    2,499
287  Peter England    2,499

                                          Description8  Category8
0    Men Blue Slim Fit Formal Full Sleeves Formal S…  Mens Wear
1    Men Red Slim Fit Formal Full Sleeves Formal Shirt  Mens Wear
2              Men Navy Check Slim Fit Formal Trousers  Mens Wear
3    Men Blue Regular Fit Formal Half Sleeves Forma…  Mens Wear
4    Men Blue Slim Fit Formal Full Sleeves Formal S…  Mens Wear
..                                                 …        …
283  Men White Regular Fit Formal Half Sleeves Form…  Mens Wear
284            Men Grey Check Slim Fit Formal Trousers  Mens Wear
285  Men Grey Regular Fit Formal Full Sleeves Forma…  Mens Wear
286  Men White Slim Fit Formal Full Sleeves Formal …  Mens Wear
287  Men Blue Slim Fit Formal Full Sleeves Formal S…  Mens Wear

[288 rows x 4 columns]
```

[98]:
```
driver.close()
```

Now we will combined all dataset and make a one dataframe.

[99]:
```
import pandas as pd

# Assuming you have DataFrames Old_Navy,As,biba,fab,levis,bane,ban,Pe

# Concatenate DataFrames vertically (stack them on top of each other)
combined_df = pd.concat([Old_Navy,As,biba,fab,levis,bane,ban,Pe],␣
 ↪ignore_index=True)

# Create new DataFrame with concatenated columns
data = pd.DataFrame(columns=['Brand', 'Price', 'Description', 'Category'])

# Assign values from original DataFrames to new DataFrame
```

```
data['Brand'] = combined_df['Brand1'].fillna('') + ' ' + combined_df['Brand2'].
 ↪fillna('') + ' ' + combined_df['Brand3'].fillna('')+ ' ' +␣
 ↪combined_df['Brand4'].fillna('') + ' ' + combined_df['Brand5'].fillna('') +␣
 ↪' ' + combined_df['Brand6'].fillna('')+ ' ' + combined_df['Brand7'].
 ↪fillna('') + ' ' + combined_df['Brand8'].fillna('')
data['Price'] = combined_df['Price1'].fillna('') + ' ' + combined_df['Price2'].
 ↪fillna('') + ' ' + combined_df['Price3'].fillna('')+ ' ' +␣
 ↪combined_df['Price4'].fillna('') + ' ' + combined_df['Price5'].fillna('') +␣
 ↪' ' + combined_df['Price6'].fillna('')+ ' ' + combined_df['Price7'].
 ↪fillna('') + ' ' + combined_df['Price8'].fillna('')
data['Description'] = combined_df['Description1'].fillna('') + ' ' +␣
 ↪combined_df['Description2'].fillna('') + ' ' + combined_df['Description3'].
 ↪fillna('') + ' ' + combined_df['Description4'].fillna('') + ' ' +␣
 ↪combined_df['Description5'].fillna('') + ' ' + combined_df['Description6'].
 ↪fillna('') + ' ' + combined_df['Description7'].fillna('') + ' ' +␣
 ↪combined_df['Description8'].fillna('')
data['Category'] = combined_df['Category1'].fillna('') + ' ' +␣
 ↪combined_df['Category2'].fillna('') + ' ' + combined_df['Category3'].
 ↪fillna('') + ' ' + combined_df['Category4'].fillna('') + ' ' +␣
 ↪combined_df['Category5'].fillna('') + ' ' + combined_df['Category6'].
 ↪fillna('') + ' ' + combined_df['Category7'].fillna('') + ' ' +␣
 ↪combined_df['Category8'].fillna('')
```

[100]:
```
import numpy as np
# Now we will add new columns those are difficult to extract so will add column␣
 ↪with random choose.
data['Color'] = np.random.choice(['red', 'yellow', 'blue', 'black', 'white',␣
 ↪'green', 'brown', 'pink', 'olive'], size=len(data))
data['Product Code'] = np.random.randint(1000, 9999, size=len(data))
data['Origin'] = np.random.choice(['India', 'US','UK'], size=len(data))
data['Size'] = np.where(data['Category'] == 'Kids Wear',
                                np.random.choice(['1-2y', '2-3y', '4-5y',␣
 ↪'5-6y', '6-7y'], size=len(data)),
                                np.random.choice(['M', 'L', 'XL', 'XXL'],␣
 ↪size=len(data)))
data['Product Type'] = np.random.choice(['Sportswear','Formal Shirt','Formal␣
 ↪Trouser','Tshirt','Jeans'], size=len(data))
data['Material Type'] = np.random.choice(['Cotton','Denim','Polyster'],␣
 ↪size=len(data))
data['Wear Type'] = np.random.choice(['Bottom Wear','Top Wear'], size=len(data))
data['Return Time'] = np.random.choice(['14 Days Return & Excahnge','7 Days␣
 ↪Return & Excahnge'], size=len(data))
data['Occasion'] = np.random.choice(['Regular','Casual','Celebration'],␣
 ↪size=len(data))
```

[101]:
```
data.head()
```

[101]:
```
             Brand          Price  \
0  Old Navy          $10.47
1  Old Navy          $15.47
2  Old Navy           $6.47
3  Old Navy           $8.97
4  Old Navy          $14.97

                                          Description            Category  \
0               Loose Satin Button-Down Shirt         Womens Wear
1  High-Waisted OG Loose Cotton-Hemp Blend Jeans …  Womens Wear
2                          Luxe Tunic T-Shirt         Womens Wear
3               Loose Crepe Button-Down Shirt         Womens Wear
4   Tie-Shoulder Underwire One-Piece Swimsuit         Womens Wear

    Color  Product Code Origin Size    Product Type Material Type  \
0    pink          4636  India   XL      Sportswear         Denim
1  yellow          1082     US    M  Formal Trouser      Polyster
2     red          8375     UK    M  Formal Trouser         Denim
3   olive          3106  India    M    Formal Shirt        Cotton
4     red          9260     US  XXL           Jeans         Denim

     Wear Type                Return Time     Occasion
0     Top Wear   7 Days Return & Excahnge       Casual
1  Bottom Wear  14 Days Return & Excahnge       Casual
2  Bottom Wear  14 Days Return & Excahnge      Regular
3  Bottom Wear  14 Days Return & Excahnge  Celebration
4  Bottom Wear   7 Days Return & Excahnge  Celebration
```

[102]:
```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4618 entries, 0 to 4617
Data columns (total 13 columns):
 #   Column         Non-Null Count  Dtype
---  ------         --------------  -----
 0   Brand          4618 non-null   object
 1   Price          4618 non-null   object
 2   Description    4618 non-null   object
 3   Category       4618 non-null   object
 4   Color          4618 non-null   object
 5   Product Code   4618 non-null   int32
 6   Origin         4618 non-null   object
 7   Size           4618 non-null   object
 8   Product Type   4618 non-null   object
 9   Material Type  4618 non-null   object
 10  Wear Type      4618 non-null   object
 11  Return Time    4618 non-null   object
```

```
 12  Occasion       4618 non-null   object
dtypes: int32(1), object(12)
memory usage: 451.1+ KB
```

[103]:
```
def convert_and_remove_currency(price):
    if price.startswith('$'):
        price_inr = float(price[1:]) * 83.20  # Assuming 1 USD = 83.20 INR (you␣
 ↪can adjust this conversion rate)
        return ' ' + str(round(price_inr, 2))
    else:
        return ' ' + price[1:]

# Apply the function to the price column
data['Price'] = data['Price'].apply(convert_and_remove_currency)
```

[104]:
```
data
```

[104]:

	Brand	Price	Description	Category	Color	Product Code	Origin	Size	Product Type	Material Type
0	Old Navy	871.1	Loose Satin Button-Down Shirt	Womens Wear	pink	4636	India	XL	Sportswear	Denim
1	Old Navy	1287.1	High-Waisted OG Loose Cotton-Hemp Blend Jeans …	Womens Wear	yellow	1082	US	M	Formal Trouser	Polyster
2	Old Navy	538.3	Luxe Tunic T-Shirt	Womens Wear	red	8375	UK	M	Formal Trouser	Denim
3	Old Navy	746.3	Loose Crepe Button-Down Shirt	Womens Wear						
4	Old Navy	1245.5	Tie-Shoulder Underwire One-Piece Swimsuit	Womens Wear						
…	…	…	…	…						
4613	Peter England	1,599	Men White Regular Fit Formal Half Sleev…	Mens Wear						
4614	Peter England	2,499	Men Grey Check Slim Fit Formal Trousers	Mens Wear						
4615	Peter England	1,399	Men Grey Regular Fit Formal Full Sleeve…	Mens Wear						
4616	Peter England	2,499	Men White Slim Fit Formal Full Sleeves …	Mens Wear						
4617	Peter England	2,499	Men Blue Slim Fit Formal Full Sleeves F…	Mens Wear						

```
3      olive          3106  India    M    Formal Shirt        Cotton
4        red          9260     US  XXL           Jeans         Denim
…        …             …     …   …              …             …
4613    blue          4067     UK   XL    Formal Shirt        Cotton
4614   green          6384     US  XXL    Formal Shirt         Denim
4615    pink          9989     US    L          Tshirt        Cotton
4616    pink          1380     US   XL          Tshirt        Cotton
4617   brown          6115     UK    M          Tshirt         Denim

        Wear Type                Return Time     Occasion
0        Top Wear   7 Days Return & Excahnge       Casual
1     Bottom Wear  14 Days Return & Excahnge       Casual
2     Bottom Wear  14 Days Return & Excahnge      Regular
3     Bottom Wear  14 Days Return & Excahnge  Celebration
4     Bottom Wear   7 Days Return & Excahnge  Celebration
…           …                          …          …
4613     Top Wear   7 Days Return & Excahnge  Celebration
4614     Top Wear  14 Days Return & Excahnge       Casual
4615  Bottom Wear  14 Days Return & Excahnge      Regular
4616     Top Wear  14 Days Return & Excahnge  Celebration
4617     Top Wear  14 Days Return & Excahnge       Casual

[4618 rows x 13 columns]
```

[105]:
```python
# Remove currency symbol, commas, newline characters, and percentages
data['Price'] = data['Price'].str.replace(' ', '')
data['Price'] = data['Price'].str.replace(',', '')
data['Price'] = data['Price'].str.replace('\n', '')
data['Price'] = data['Price'].str.replace('% OFF', '')

# Remove any non-numeric characters
data['Price'] = data['Price'].str.replace(r'\D', '')

# Convert to float
data['Price'] = data['Price'].astype(float)

# Convert float to int
data['Price'] = data['Price'].astype(int)
```

[106]:
```python
data
```

[106]:
```
                  Brand  Price  \
0          Old Navy          8711
1          Old Navy         12871
2          Old Navy          5383
3          Old Navy          7463
4          Old Navy         12455
```

```
…                    …    …
4613         Peter England   1599
4614         Peter England   2499
4615         Peter England   1399
4616         Peter England   2499
4617         Peter England   2499

                                                Description            Category  \
0                  Loose Satin Button-Down Shirt         Womens Wear
1     High-Waisted OG Loose Cotton-Hemp Blend Jeans …  Womens Wear
2                              Luxe Tunic T-Shirt         Womens Wear
3                  Loose Crepe Button-Down Shirt         Womens Wear
4      Tie-Shoulder Underwire One-Piece Swimsuit         Womens Wear
…                                                    …                 …
4613         Men White Regular Fit Formal Half Sleev…           Mens Wear
4614            Men Grey Check Slim Fit Formal Trousers           Mens Wear
4615         Men Grey Regular Fit Formal Full Sleeve…           Mens Wear
4616         Men White Slim Fit Formal Full Sleeves …           Mens Wear
4617         Men Blue Slim Fit Formal Full Sleeves F…           Mens Wear

       Color  Product Code Origin Size    Product Type Material Type  \
0       pink          4636  India   XL      Sportswear         Denim
1     yellow          1082     US    M  Formal Trouser      Polyster
2        red          8375     UK    M  Formal Trouser         Denim
3      olive          3106  India    M    Formal Shirt        Cotton
4        red          9260     US  XXL           Jeans         Denim
…       …           …    …  …             …           …
4613    blue          4067     UK   XL    Formal Shirt        Cotton
4614   green          6384     US  XXL    Formal Shirt         Denim
4615    pink          9989     US    L          Tshirt        Cotton
4616    pink          1380     US   XL          Tshirt        Cotton
4617   brown          6115     UK    M          Tshirt         Denim

        Wear Type                Return Time     Occasion
0        Top Wear   7 Days Return & Excahnge       Casual
1     Bottom Wear  14 Days Return & Excahnge       Casual
2     Bottom Wear  14 Days Return & Excahnge      Regular
3     Bottom Wear  14 Days Return & Excahnge  Celebration
4     Bottom Wear   7 Days Return & Excahnge  Celebration
…           …                          …          …
4613     Top Wear   7 Days Return & Excahnge  Celebration
4614     Top Wear  14 Days Return & Excahnge       Casual
4615  Bottom Wear  14 Days Return & Excahnge      Regular
4616     Top Wear  14 Days Return & Excahnge  Celebration
4617     Top Wear  14 Days Return & Excahnge       Casual

[4618 rows x 13 columns]
```

[107]:
```
data.to_csv("CapstoneClothingBrand", index=False)
```

[1]:
```
import pandas as pd
data = pd.read_csv("CapstoneClothingBrand")
```

[2]:
```
data
```

[2]:

```
                    Brand  Price  \
0              Old Navy          8711
1              Old Navy         12871
2              Old Navy          5383
3              Old Navy          7463
4              Old Navy         12455
…                     …    …
4613         Peter England   1599
4614         Peter England   2499
4615         Peter England   1399
4616         Peter England   2499
4617         Peter England   2499

                                                Description            Category  \
0                      Loose Satin Button-Down Shirt          Womens Wear
1     High-Waisted OG Loose Cotton-Hemp Blend Jeans …  Womens Wear
2                                Luxe Tunic T-Shirt          Womens Wear
3                      Loose Crepe Button-Down Shirt          Womens Wear
4       Tie-Shoulder Underwire One-Piece Swimsuit          Womens Wear
…                                                                 …                 …
4613         Men White Regular Fit Formal Half Sleev…           Mens Wear
4614            Men Grey Check Slim Fit Formal Trousers           Mens Wear
4615         Men Grey Regular Fit Formal Full Sleeve…           Mens Wear
4616         Men White Slim Fit Formal Full Sleeves …           Mens Wear
4617         Men Blue Slim Fit Formal Full Sleeves F…           Mens Wear

       Color  Product Code Origin Size    Product Type Material Type  \
0       pink          4636  India   XL      Sportswear         Denim
1     yellow          1082     US    M  Formal Trouser      Polyster
2        red          8375     UK    M  Formal Trouser         Denim
3      olive          3106  India    M    Formal Shirt        Cotton
4        red          9260     US  XXL           Jeans         Denim
…      …           …    …  …             …           …
4613    blue          4067     UK   XL    Formal Shirt        Cotton
4614   green          6384     US  XXL    Formal Shirt         Denim
4615    pink          9989     US    L          Tshirt        Cotton
4616    pink          1380     US   XL          Tshirt        Cotton
4617   brown          6115     UK    M          Tshirt         Denim

        Wear Type                Return Time     Occasion
```

```
0        Top Wear   7 Days Return & Excahnge       Casual
1     Bottom Wear  14 Days Return & Excahnge       Casual
2     Bottom Wear  14 Days Return & Excahnge      Regular
3     Bottom Wear  14 Days Return & Excahnge  Celebration
4     Bottom Wear   7 Days Return & Excahnge  Celebration
…           …                          …          …
4613     Top Wear   7 Days Return & Excahnge  Celebration
4614     Top Wear  14 Days Return & Excahnge       Casual
4615  Bottom Wear  14 Days Return & Excahnge      Regular
4616     Top Wear  14 Days Return & Excahnge  Celebration
4617     Top Wear  14 Days Return & Excahnge       Casual

[4618 rows x 13 columns]
```

[3]:
```
# I want rename some columns
data.rename(columns={'Product Type': 'Product_Type', 'Wear Type': 'Wear_Type',␣
 ↪'Material Type':'Material_Type', 'Return Time':'Return_Time'}, inplace=True)
```

[4]:
```
# Adding a 'S.no' column with sequence serial numbers
data.insert(0, 'Sno', range(1, 1 + len(data)))
```

[5]:
```
data
```

[5]:
```
       Sno                 Brand  Price  \
0        1       Old Navy          8711
1        2       Old Navy         12871
2        3       Old Navy          5383
3        4       Old Navy          7463
4        5       Old Navy         12455
…      …                  …     …
4613  4614         Peter England   1599
4614  4615         Peter England   2499
4615  4616         Peter England   1399
4616  4617         Peter England   2499
4617  4618         Peter England   2499

                                            Description            Category  \
0                 Loose Satin Button-Down Shirt         Womens Wear
1     High-Waisted OG Loose Cotton-Hemp Blend Jeans …  Womens Wear
2                             Luxe Tunic T-Shirt         Womens Wear
3                 Loose Crepe Button-Down Shirt         Womens Wear
4      Tie-Shoulder Underwire One-Piece Swimsuit         Womens Wear
…                                                 …                …
4613         Men White Regular Fit Formal Half Sleev…           Mens Wear
4614            Men Grey Check Slim Fit Formal Trousers             Mens Wear
4615         Men Grey Regular Fit Formal Full Sleeve…           Mens Wear
4616         Men White Slim Fit Formal Full Sleeves …           Mens Wear
```

```
4617         Men Blue Slim Fit Formal Full Sleeves F…           Mens Wear

       Color  Product Code Origin Size    Product_Type Material_Type  \
0       pink          4636  India   XL      Sportswear         Denim
1     yellow          1082     US    M  Formal Trouser      Polyster
2        red          8375     UK    M  Formal Trouser         Denim
3      olive          3106  India    M    Formal Shirt        Cotton
4        red          9260     US  XXL           Jeans         Denim
…        …            …    …   …              …           …
4613    blue          4067     UK   XL    Formal Shirt        Cotton
4614   green          6384     US  XXL    Formal Shirt         Denim
4615    pink          9989     US    L          Tshirt        Cotton
4616    pink          1380     US   XL          Tshirt        Cotton
4617   brown          6115     UK    M          Tshirt         Denim

        Wear_Type                Return_Time     Occasion
0        Top Wear   7 Days Return & Excahnge       Casual
1     Bottom Wear  14 Days Return & Excahnge       Casual
2     Bottom Wear  14 Days Return & Excahnge      Regular
3     Bottom Wear  14 Days Return & Excahnge  Celebration
4     Bottom Wear   7 Days Return & Excahnge  Celebration
…           …                          …          …
4613     Top Wear   7 Days Return & Excahnge  Celebration
4614     Top Wear  14 Days Return & Excahnge       Casual
4615  Bottom Wear  14 Days Return & Excahnge      Regular
4616     Top Wear  14 Days Return & Excahnge  Celebration
4617     Top Wear  14 Days Return & Excahnge       Casual

[4618 rows x 14 columns]
```

Now,according to project we are extract 3 tables from this dataset.

[8]:
```
# Extracting Table 1
cb1 = data[["Sno","Brand", "Category", "Product_Type", "Description", "Product␣
 ↪Code", "Price"]]
# Extracting Table 2
cb2 = data[["Sno","Wear_Type", "Material_Type","Color"]]
# Extracting Table 3
cb3 = data[["Sno","Size", "Origin", "Return_Time", "Occasion"]]
```

[11]:
```
cb1.to_csv("cb1.csv", index=False)
cb2.to_csv("cb2.csv", index=False)
cb3.to_csv("cb3.csv", index=False)
```

[ ]:
```

```