

In [1]:

```
# This project is based on web scrapping using BeautifulSoup.  
# Most of the brands are promoting Black Friday and Christmas Discounts on their products.  
# For this project, I have selected a website which sells dj controllers and other related equipment with different discount offers  
# Through this project, we can calculate the % Discount on each Dj Controller and choose the right product with maximum savings.
```

In [2]:

```
import bs4
```

In [3]:

```
#Importing urlopen from urllib.request to fetch internet resources  
#Importing BeautifulSoup  
from urllib.request import urlopen as uReq  
from bs4 import BeautifulSoup as soup
```

In [4]:

```
my_url = 'https://www.storedj.com.au/dj-equipment/dj-controllers?SortProduct=SaveValue%20DESC&PageProduct=1'
```

In [5]:

```
#Opening the connection and grabbing the page  
uClient = uReq(my_url)  
page_html = uClient.read()  
  
#Closing the client  
uClient.close()
```

In [6]:

```
#Data Derived from BeautifulSoup Package | HTML Parsing  
page_soup = soup(page_html, "html.parser")
```

In [7]:

```
# Checking h1 tags  
page_soup.h1
```

Out[7]:

```
<h1 class="page-title ico-heading">DJ Controllers</h1>
```

In [8]:

```
# Checking P tags  
page_soup.p
```

Out[8]:

```
<p class="promo-title">5 Store Locations</p>
```

In [9]:

```
# Checking Span Class  
page_soup.body.span
```

Out[9]:

```
<span class="topbar-copy">Australia's Largest Range Of Pro Audio Equipment</span>
```

In [10]:

```
#Fetching all div with class "Product"
containers = page_soup.findAll("div", {"class": "product"})
```

In [11]:

```
#Checking Container Length
len(containers)
```

Out[11]:

15

In [12]:

```
#Checking container at index 0
containers[0]
```

Out[12]:

```
<div class="product " data-current-pickup-store-name="" data-enable-store-availability-click-and-c
ollect="true" data-hide-footer-zone="false" data-is-favourite="false" data-is-user-favourites-
page="false" data-items-on-order="0" data-product-code="PIO-ZDDJRZX" data-product-data="" data-
qty-on-back-order="0" data-rc-quick-view-close-button-html='&lt;div class="close-overlay cv-ico-ge
neral-cross3"&gt;Close&lt;/div&gt;' data-rc-quick-view-slide-toggle-selector=".cv-zone-product-3 .c
lose-overlay" data-rc-quick-view-zone-selector=".cv-zone-product-3" data-role="product" data-show-c
elivery-available="true" data-show-delivery-available-back-order="false" data-show-delivery-
available-five-to-seven-days="false" data-show-delivery-available-three-to-six-days="false" data-s
how-delivery-available-two-to-five-days="false" data-show-delivery-indent-product-call="false" dat
a-show-delivery-select-product-attributes="false" data-show-delivery-special-order="false" data-sh
ow-delivery-store-not-set="false" data-show-delivery-unavailable="false" data-show-pickup-
available="false" data-show-pickup-available-back-order="false" data-show-pickup-indent-product-ca
ll="false" data-show-pickup-low-stock-call="false" data-show-pickup-select-product-
attributes="false" data-show-pickup-store-not-set="true" data-show-pickup-unavailable="false"
data-store-avail-click-and-collect-delivery-display-mode="Standard" data-trigger-favourites-
messages="false" data-trigger-messages="true"><div data-view="true">
<div class="cv-zone-product-1">
<div class="widget-productlist-image ">
<a data-product-link="PIO-ZDDJRZX" href="/open-box-pioneer-ddjrzx-pro-rekordbox-dj-vj-controller">

</a>
<!-- Product Ribbon Group -->
<aside class="product-ribbons">

</aside>
</div>
</div>
<div class="cv-zone-product-2">
<div class="cv-zone-product-4">
</div>
<div class="cv-zone-product-5 ">
<span class="widget-productlist-title "><a data-product-link="PIO-ZDDJRZX" href="/open-box-
pioneer-ddjrzx-pro-rekordbox-dj-vj-controller">OPEN BOX Pioneer DDJRZX Pro Rekordbox DJ/VJ
Controller</a></span>
<div class="widget-html-snippet ">
<div class="our-price">Our Price</div>
</div>
<span class="widget-productlist-price ">
<span class="item-price" data-bind="invisible: showAltPrice">$3499</span>
</span>
<div class="product-sale-info cv-ico-general-tag3">
<span class="price-sale-save-item">Save $500</span>
</div>
<span class="widget-productlist-availability cart-product-availability in-stock"> In Stock </span>
</div>
<div class="cv-zone-product-6">
<div class="widget-productlist-add-button">Add to cart</div>
</div>
```

```

<button class="widget-productlist-atc-button btn cv-add-to-cart" data-bind="click: addToCart, addi
sProcessingClass: isAddingToCart, disabled: isAddingToCart" type="button">Add to cart</button>
</div>
</div>
<div class="cv-zone-product-3">
</div>
</div></div>

```

In [13]:

```

#Assigning the containers to variable container
container = containers[0]

```

In [14]:

```

#Checking the details inside container 'a' class
container.a

```

Out[14]:

```

<a data-product-link="PIO-ZDDJRZX" href="/open-box-pioneer-ddjrzx-pro-rekordbox-dj-vj-controller">

</a>

```

In [15]:

```

#Checking Container 'div'
container.div

```

Out[15]:

```

<div data-view="true">
<div class="cv-zone-product-1">
<div class="widget-productlist-image ">
<a data-product-link="PIO-ZDDJRZX" href="/open-box-pioneer-ddjrzx-pro-rekordbox-dj-vj-controller">

</a>
<!-- Product Ribbon Group -->
<aside class="product-ribbons">

</aside>
</div>
</div>
<div class="cv-zone-product-2">
<div class="cv-zone-product-4">
</div>
<div class="cv-zone-product-5 ">
<span class="widget-productlist-title "><a data-product-link="PIO-ZDDJRZX" href="/open-box-
pioneer-ddjrzx-pro-rekordbox-dj-vj-controller">OPEN BOX Pioneer DDJRZX Pro Rekordbox DJ/VJ
Controller</a></span>
<div class="widget-html-snippet ">
<div class="our-price">Our Price</div>
</div>
<span class="widget-productlist-price ">
<span class="item-price" data-bind="invisible: showAltPrice">$3499</span>
</span>
<div class="product-sale-info cv-ico-general-tag3">
<span class="price-sale-save-item">Save $500</span>
</div>
<span class="widget-productlist-availability cart-product-availability in-stock"> In Stock </span>
</div>
<div class="cv-zone-product-6">

```

```

<button class="widget-productlist-atc-button btn cv-add-to-cart" data-bind="click: addToCart, addI
sProcessingClass: isAddingToCart, disabled: isAddingToCart" type="button">Add to cart</button>
</div>
</div>
<div class="cv-zone-product-3">
</div>
</div>

```

In [16]:

```

#Checking inside the right 'div' and 'a' class to fetch required data
container.div.div.a

```

Out[16]:

```

<a data-product-link="PIO-ZDDJRZX" href="/open-box-pioneer-ddjrzx-pro-rekordbox-dj-vj-controller">

</a>

```

In [17]:

```

# Creating file with the required fetched data
# Using 'For loop' to iterate over each line item - brand name, product price & discount price
filename = "products1.csv"
f = open(filename,"w")
headers = "brand , price , discount \n"
f.write(headers)

for container in containers:
    brand = container.div.div.a.img["data-image-alt-tag"]
    product_price = container.findAll("span", {"class": "item-price"})
    price = product_price[0].text
    discount_price = container.findAll("span", {"class": "price-sale-save-item"})
    discount = discount_price[0].text
    print("brand: " + brand)
    print("price: " + price)
    print("discount:" + discount)

    f.write(brand + "," + price + "," + discount + "\n")

f.close()

```

```

brand: OPEN BOX Pioneer DDJRZX Pro Rekordbox DJ/VJ Controller
price: $3499
discount:Save $500
brand: OPEN BOX Native Instruments Traktor Kontrol S8
price: $1299
discount:Save $400
brand: Native Instruments Traktor Kontrol S8 Flagship All-In-One 4-Channel DJ System
price: $1388
discount:Save $311
brand: Korg Kaoss DJ Controller w/ Built-In Kaoss Pad
price: $295
discount:Save $304
brand: Native Instruments Traktor Kontrol S5 Pro 4-Channel DJ System
price: $648
discount:Save $251
brand: OPEN BOX Pioneer DDJSX3 4-Channel Serato DJ Pro Controller
price: $1349
discount:Save $250
brand: Pioneer DDJSX3 DJ Controller with Free HDJX5 Headphones
price: $1499
discount:Save $249
brand: Roland DJ808 Serato DJ Controller w/ Drum Machine, Vocoder & Aira Link
price: $1650
discount:Save $229
brand: Native Instruments Traktor Kontrol D2
price: $345
discount:Save $224

```

```

discount:Save $204
brand: Numark Mixtrack Quad 4-Ch DJ Controller
price: $379
discount:Save $170
brand: Numark NS6 II 4-Ch Premium Serato DJ Controller
price: $979
discount:Save $170
brand: Native Instruments Traktor Kontrol S8 4-ch DJ System w/ Bonus Decksaver Cover
price: $1290
discount:Save $163
brand: Native Instruments Traktor Kontrol S4 MK2 4-Deck DJ Controller
price: $599
discount:Save $150
brand: Numark NS7II Display 3-Screen Display to Convert NS7II to a NS7mk3
price: $149
discount:Save $150
brand: OPEN BOX Native Instruments Traktor Kontrol S2 Mk2
price: $369
discount:Save $130

```

In [18]:

```

#Importing Pandas
import pandas as pd

```

In [19]:

```

#Reading the fetched CSV File
df = pd.read_csv("C:/Users/rmokashi/Documents/Individual Assignment/products1.csv")

df.to_csv('C:/Users/rmokashi/Documents/Individual Assignment/products1.csv')

```

In [20]:

```

#Showing the column names
list(df.columns.values)

```

Out[20]:

```
['brand ', ' price ', ' discount ']
```

In [21]:

```

#Showing the head of the dataframe df
df.head()

```

Out[21]:

	brand	price	discount
0	Pioneer DDJRZX Pro 4-Channel Rekordbox DJ Cont...	\$3999	Save \$0
1	Pioneer DDJRZ Pro 4-Channel Rekordbox DJ Contr...	\$2999	Save \$0
2	Pioneer DDJSZ2 Four Channel Premium Serato DJ ...	\$2999	Save \$0
3	OPEN BOX Pioneer DDJRZX Pro Rekordbox DJ/VJ Co...	\$3499	Save \$500
4	Pioneer XDJR2 All-in-One DJ System for Rekordbox	\$2349	Save \$0

In [22]:

```

#Calculating Percent Discount
df2 = pd.DataFrame({"Price": [1184, 295, 345, 648, 1388, 1299, 3499, 1349, 1699, 2999, 3999, 1599, 2999, 1699, 2349],
                    "Discount": [215, 304, 204, 251, 311, 400, 500, 250, 0, 0, 0, 0, 0, 0, 0]})

df['Percent_Discount'] = ((df2["Discount"] / (df2["Price"]+df2["Discount"])))*100

```

In [23]:

```
#Overview of df dataframe updated with percent discount
df
```

Out[23]:

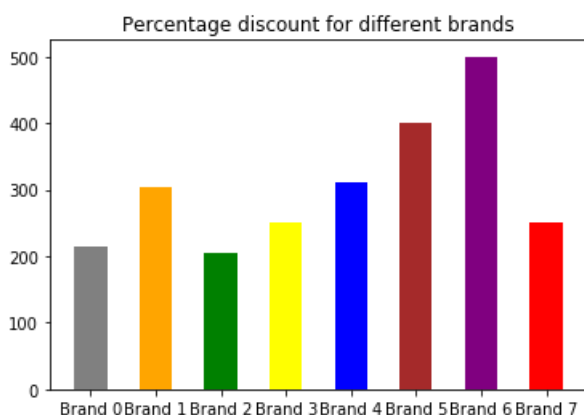
	brand	price	discount	Percent_Discount
0	Pioneer DDJRZX Pro 4-Channel Rekordbox DJ Cont...	\$3999	Save \$0	15.368120
1	Pioneer DDJRZ Pro 4-Channel Rekordbox DJ Contr...	\$2999	Save \$0	50.751252
2	Pioneer DDJSZ2 Four Channel Premium Serato DJ ...	\$2999	Save \$0	37.158470
3	OPEN BOX Pioneer DDJRZX Pro Rekordbox DJ/VJ Co...	\$3499	Save \$500	27.919911
4	Pioneer XDJRX2 All-in-One DJ System for Rekordbox	\$2349	Save \$0	18.304885
5	OPEN BOX Native Instruments Traktor Kontrol S8	\$1299	Save \$400	23.543261
6	Pioneer XDJRR All-in-One DJ System for Rekordbox	\$1699	Save \$0	12.503126
7	Pioneer DDJ1000 4 Channel Rekordbox DJ Control...	\$1699	Save \$0	15.634772
8	Pioneer DDJSX3 4-Channel Serato DJ Pro Controller	\$1599	Save \$0	0.000000
9	Native Instruments Traktor Kontrol S8 Flagship...	\$1388	Save \$311	0.000000
10	Korg Kaoss DJ Controller w/ Built-In Kaoss Pad	\$295	Save \$304	0.000000
11	Native Instruments Traktor Kontrol S5 Pro 4-Ch...	\$648	Save \$251	0.000000
12	OPEN BOX Pioneer DDJSX3 4-Channel Serato DJ Pr...	\$1349	Save \$250	0.000000
13	Denon MC7000 Professional DJ Controller w/ Dua...	\$1184	Save \$215	0.000000
14	Native Instruments Traktor Kontrol D2	\$345	Save \$204	0.000000

In [24]:

```
#Importing matplotlib
import matplotlib.pyplot as plt
```

In [26]:

```
#Plotting the bar diagram of products with percent discount
#Numbers 0 to 7 on 'x axis' are the brands according to index position in dataframe
ax = (1,2,3,4,5,6,7)
ay = df.iloc[:7,3]
a1= plt.bar("Brand 0", df.iloc[0,3], color= 'gray',width=.5)
a2 =plt.bar('Brand 1',df.iloc[1,3], color= 'orange',width=.5 )
a3 =plt.bar('Brand 2',df.iloc[2,3], color= 'green',width=.5)
a4 =plt.bar('Brand 3',df.iloc[3,3], color= 'yellow',width=.5)
a4 =plt.bar('Brand 4',df.iloc[4,3], color= 'blue',width=.5)
a5 =plt.bar('Brand 5',df.iloc[5,3], color= 'brown',width=.5)
a6 =plt.bar('Brand 6',df.iloc[6,3], color= 'purple',width=.5)
a7 =plt.bar('Brand 7',df.iloc[7,3], color= 'red',width=.5)
plt.title("Percentage discount for different brands")
plt.show()
df
```



Out[26]:

	Unnamed: 0	Brand	Price	Discount	Percent_Discount
0	0	Denon MC7000 Professional	1184	215	15.368120
1	1	Korg Kaoss DJ Controller	295	304	50.751252
2	2	Native Instruments Traktor Kontrol D2	345	204	37.158470
3	3	Native Instruments Traktor Kontrol S5	648	251	27.919911
4	4	Native Instruments Traktor Kontrol S8	1388	311	18.304885
5	5	OPEN BOX Native Instruments Traktor Kontrol S8	1299	400	23.543261
6	6	OPEN BOX Pioneer DDJRZX Pro Rekordbox	3499	500	12.503126
7	7	OPEN BOX Pioneer DDJSX3	1349	250	15.634772
8	8	Pioneer DDJ1000 4	1699	0	0.000000
9	9	Pioneer DDJRZ Pro 4	2999	0	0.000000
10	10	Pioneer DDJRZX Pro 4	3999	0	0.000000
11	11	Pioneer DDJSX3 4	1599	0	0.000000
12	12	Pioneer DDJSZ2 4	2999	0	0.000000
13	13	Pioneer XDJRR	1699	0	0.000000
14	14	Pioneer XDJRX2	2349	0	0.000000