

In [1]:

import pandas as pd  
import seaborn as sns  
import matplotlib.pyplot as plt

In [2]:

fashion=pd.read\_csv('FashionDataset.csv')

In [3]:

fashion

Out[3]:

	Unnamed: 0	BrandName	Deatils	Sizes	MRP	SellPrice	Discou
0	0	life	solid cotton blend collar neck womens a-line d...	Size:Large,Medium,Small,X- Large,X-Small	Rs\n1699	849	50% c
1	1	only	polyester peter pan collar womens blouson dres...	Size:34,36,38,40	Rs\n3499	2449	30% c
2	2	fratini	solid polyester blend wide neck womens regular...	Size:Large,X-Large,XX- Large	Rs\n1199	599	50% c
3	3	zink london	stripes polyester sweetheart neck womens dress...	Size:Large,Medium,Small,X- Large	Rs\n2299	1379	40% c
4	4	life	regular fit regular length denim womens jeans ...	Size:26,28,30,32,34,36	Rs\n1699	849	50% c
...	...	...	...	...	...	...	...
30753	21	swarovski	crystal stylish womens rodhium earrings	Nan	Nan	8950	Na
30754	22	Nan	Nan	Nan	Nan	Nan	Na
30755	23	jewelz	ethnic gold plated jhumki earrings	Nan	Rs\n1839	643	65% c
30756	24	estelle	womens gold plated double line fancy white and...	Nan	Nan	2799	Na
30757	25	estelle	womens gold plated bridge designer mangalsutra...	Nan	Nan	1899	Na

In [5]:

30758 rows × 8 columns  
fashion.describe()

Out[5]:

In [4]:

fashion.info()

Out[4]:

<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 30758 entries, 0 to 30757  
Data columns (total 8 columns):  
# Column Non-Null Count Dtype  
--- --  
0 Unnamed: 0 30758 non-null int64  
1 BrandName 30758 non-null object  
2 Deatils 30758 non-null object  
3 Sizes 30758 non-null object  
4 MRP 30758 non-null object  
5 SellPrice 30758 non-null object  
6 Discount 30758 non-null object  
7 Category 30758 non-null object  
dtypes: object(7), int64(1)  
memory usage: 1.9+ MB

In [6]:

fashion.dropna(inplace = True)

In [7]:

fashion=fashion.rename(columns={'BrandName':'Brand'})  
fashion=fashion.rename(columns={'Deatils':'Details'})

In [8]:

fashion.isna().any()

Out[8]:

Brand False  
Details False

In [5]:

30758 rows x 8 columns  
fashion.describe()

Out[5]:

In [4]:

fashion.isna().sum()

count 30758

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 30758 entries, 0 to 30757

Data columns (total 8 columns):

# std min max Non-Null Count Dtype

0 Unnamed: 0 0.000000 30758 non-null int64

1 BrandName 0.000000 30758 non-null object

2 Details 0.000000 30758 non-null object

3 Sizes 0.000000 30758 non-null object

4 MRP 19.000000 30758 non-null object

5 SellPrice 25.000000 30758 non-null object

6 Discount 0.000000 30758 non-null object

7 Category 0.000000 30758 non-null object

In [6]:

fashion.dropna(inplace=True)

memory usage: 1.9+ MB

In [7]:

fashion=fashion.rename(columns={'BrandName':'Brand'})  
fashion=fashion.rename(columns={'Deatils':'Details'})

In [8]:

fashion.isna().any()

Out[8]:

Brand False  
Details False  
Sizes False  
MRP False  
SellPrice False  
Discount False  
Category False  
dtype: bool

In [9]:

fashion.replace('NaN',inplace = True)

In [10]:

fashion['MRP']=fashion['MRP'].str.strip('Rs\n').astype(int)  
fashion['MRP']

Out[10]:

0 1699  
1 3499  
2 1199  
3 2299  
4 1699  
...  
30753 449  
30754 449  
30755 1839  
30756 1839  
30757 1839  
Name: MRP, Length: 30758, dtype: int32

In [11]:

fashion.dtypes

Out[11]:

Brand object  
Details object  
Sizes object  
MRP int32  
SellPrice object  
Discount object  
Category object  
dtype: object

In [15]:

fashion

Out[15]:

	Brand	Details	Sizes	MRP	SellPrice	Discount	Category
0	solid wide	regular fit	Size:Large,Medium,Small,X-Large,X-Small	1299	649	50	Westernwear-Women
1	viscose	regular top -					
2	womens	regular top -					
3	life	regular fit					
4	jeans ...	regular					
...	...	...	...	...	...	...	...
30753	solid wide	regular fit	Size:Large,Medium,Small,X-Large,X-Small	1299	649	50	Westernwear-Women
30754	viscose	regular top -					
30755	womens	regular top -					
30756	life	regular fit					
30757	jeans ...	regular					

In [12]:

fashion['SellPrice']=fashion['SellPrice'].astype(int)

In [13]:

fashion['Discount']=fashion['Discount'].str.strip('%').astype(int)

In [14]:

fashion=fashion.loc[fashion.duplicated()]

48

life

viscose wide neck

Size:Large,Medium,Small,X-Large,X-Small

1299

649

50

Westernwear-Women

74

stop

printed viscose regular neck womens top - white

Size:Large,Medium,Small,X-Large,X-Small,XX-Large

999

499

50

Westernwear-Women

100

stop

solid polyester regular neck

Size:Large,Medium,Small,X-Large,X-Small

1299

649

50

Westernwear-Women

```
Out[11]:
In [15]: fashion
Out[15]: MRP      int32
          SellPrice object
          Discount object
          Category object
          dtype: object
```

```
In [12]: fashion['SellPrice']=fashion['SellPrice'].astype(int)
```

```
In [13]: fashion['Discount']=fashion['Discount'].astype(int)
```

```
In [14]: fashion=fashion.loc[fashion.duplicated()]
```

	Brand	Details	Sizes	MRP	SellPrice	Discount	Category
22	life	viscose wide neck womens regular top - blue	Size:Large,Medium,Small,X-Large,X-Small	1299	649	50	Westernwear-Women
74	stop	printed viscose regular neck womens top - white	Size:Large,Medium,Small,X-Large,X-Small,XX-Large	999	499	50	Westernwear-Women
100	stop	solid polyester regular neck womens top - sage	Size:Large,Medium,Small,X-Large,XX-Large	1299	649	50	Westernwear-Women
...	...	...	...	...	...	...	...
30702	skagen	elin rose gold bracelet - skj0851791	Size:Error Size	599	5795	50	Jewellery-Women
30716	ayesha	womens stone studded drop earrings - multi	Size:Error Size	698	349	50	Jewellery-Women
30728	fossil	vintage gl two tone bracelet - jf02311040	Size:Error Size	2399	4795	5	Jewellery-Women
30735	jewelz	cotton multicolour girls hair bands	Size:Error Size	999	349	65	Jewellery-Women
30754	swarovski	crystal stylish womens rodhium earrings	Size:Error Size	449	8950	15	Jewellery-Women

1958 rows × 7 columns

```
In [16]: split=fashion['Category'].str.split('-',n=1,expand=True)
```

```
In [19]: fashion
```

```
Out[19]: fashion.insert(fashion.columns.get_loc('Discount')+1,'Sub Category',split[0])
```

```
In [18]: fashion.insert(fashion.columns.get_loc('Sub Category')+1,'Main Category',split[1])
```

	Brand	Details	Sizes	MRP	SellPrice	Discount	Sub Category	Main Category
22	life	viscose wide neck womens regular top - blue	Size:Large,Medium,Small,X-Large,X-Small	1299	649	50	Westernwear	W
24	life	regular fit regular length denim womens jeans ...	Size:26,28,30,32,34,36	1999	999	50	Westernwear	W
48	life	printed viscose wide neck womens regular dress...	Size:Large,Medium,Small,X-Large,X-Small	1299	649	50	Westernwear	W
74	stop	printed viscose regular neck womens top - white	Size:Large,Medium,Small,X-Large,X-Small,XX-Large	999	499	50	Westernwear	W

```
In [19]: fashion
Out[19]: fashion.insert(fashion.columns.get_loc('Discount')+1,'Sub Category',split[0])

In [18]: fashion.insert(fashion.columns.get_loc('Sub Category')+1,'Main Category',split[1])
```

	Brand	Details	Sizes	MRP	SellPrice	Discount	Sub Category	Cat
22	life	solid wide neck viscose womens regular top - blue	Size:Large,Medium,Small,X-Large,X-Small	1299	649	50	Westernwear	V
24	life	regular fit regular length denim womens jeans ...	Size:26,28,30,32,34,36	1999	999	50	Westernwear	V
48	life	printed viscose wide neck womens regular dress...	Size:Large,Medium,Small,X-Large,X-Small	1299	649	50	Westernwear	V
74	stop	printed viscose regular neck womens top - white	Size:Large,Medium,Small,X-Large,X-Small,XX-Large	999	499	50	Westernwear	V
100	stop	solid polyester regular neck womens top - sage	Size:Large,Medium,Small,X-Large,XX-Large	1299	649	50	Westernwear	V
...	...	...	...	...	...	...	...	...
30702	skagen	elin rose gold bracelet - skj0851791	Size:Error Size	599	5795	50	Jewellery	V
30716	ayesha	womens stone studded drop earrings - multi	Size:Error Size	698	349	50	Jewellery	V
30728	fossil	vintage gl two tone bracelet - jf02311040	Size:Error Size	2399	4795	5	Jewellery	V
30735	jewelz	cotton multicolour girls hair bands	Size:Error Size	999	349	65	Jewellery	V
30754	swarovski	crystal stylish womens rodhium earrings	Size:Error Size	449	8950	15	Jewellery	V

```
In [21]: fashion
Out[21]: 1958 rows × 9 columns

In [20]: fashion.drop('Category',axis=1,inplace = True)
```

	Brand	Details	Sizes	MRP	SellPrice	Discount	Sub Category	Cat
22	life	solid wide neck viscose womens regular top - blue	Size:Large,Medium,Small,X-Large,X-Small	1299	649	50	Westernwear	V
24	life	regular fit regular length denim womens jeans ...	Size:26,28,30,32,34,36	1999	999	50	Westernwear	V
48	life	printed viscose wide neck womens regular dress...	Size:Large,Medium,Small,X-Large,X-Small	1299	649	50	Westernwear	V
74	stop	printed viscose regular neck womens top - white	Size:Large,Medium,Small,X-Large,X-Small,XX-Large	999	499	50	Westernwear	V

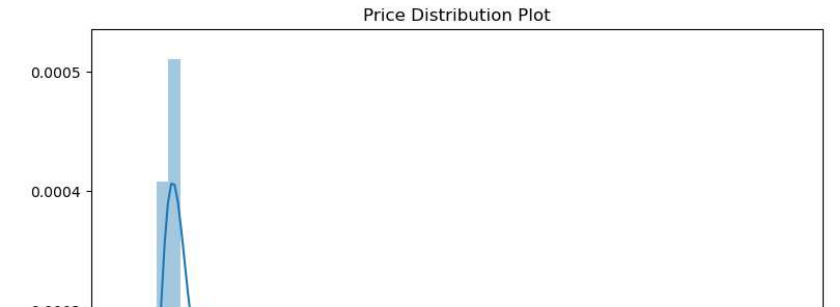
```
In [21]: fashion
Out[21]:
In [20]: fashion.drop('Category',axis=1,inplace = True)
```

	Brand	Details	Sizes	MRP	SellPrice	Discount	Sub Category	Cat
22	life	solid wide neck viscose womens regular top - blue	Size:Large,Medium,Small,X-Large,X-Small	1299	649	50	Westernwear	V
24	life	regular fit regular length denim womens jeans ...	Size:26,28,30,32,34,36	1999	999	50	Westernwear	V
48	life	printed viscose wide neck womens regular dress...	Size:Large,Medium,Small,X-Large,X-Small	1299	649	50	Westernwear	V
74	stop	printed viscose regular neck womens top - white	Size:Large,Medium,Small,X-Large,X-Small,XX-Large	999	499	50	Westernwear	V
100	stop	solid polyester regular neck womens top - sage	Size:Large,Medium,Small,X-Large,XX-Large	1299	649	50	Westernwear	V
...	...	...	...	...	...	...	...	...
30702	skagen	elin rose gold bracelet - skj0851791	Size:Error Size	599	5795	50	Jewellery	V
30716	ayesha	womens stone studded drop earrings - multi	Size:Error Size	698	349	50	Jewellery	V
30728	fossil	vintage gl two tone bracelet - jf02311040	Size:Error Size	2399	4795	5	Jewellery	V
30735	jewelz	cotton multicolour girls hair bands	Size:Error Size	999	349	65	Jewellery	V
30754	swarovski	crystal stylish womens rodhium earrings	Size:Error Size	449	8950	15	Jewellery	V

```
In [22]: plt.figure(figsize=(20,8))
plt.subplot(1,2,1)
plt.title('Price Distribution Plot')
sns.distplot(fashion.SellPrice)
plt.show()
```

C:\Users\kashan\Anaconda3\lib\site-packages\seaborn\distributions.py:2619: FutureWarning: `distplot` is a deprecated function and will be removed in a future version. Please adapt your code to use either `displot` (a figure-level function with similar flexibility) or `histplot` (an axes-level function for histograms).

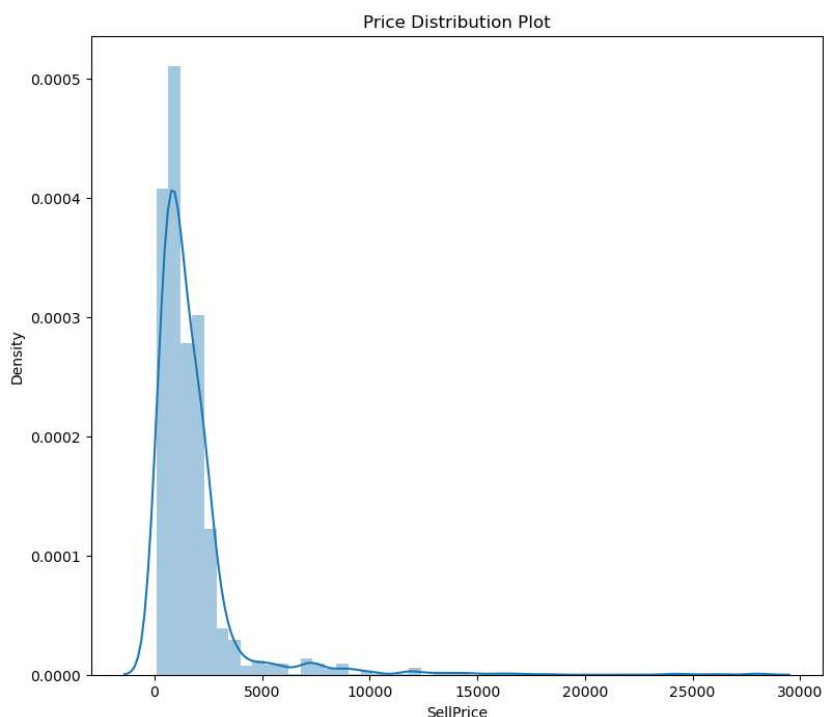
warnings.warn(msg, FutureWarning)



```
In [22]: plt.figure(figsize=(20,8))
plt.subplots(1,2,1)
plt.title('Price Distribution Plot')
sns.distplot(fashion.SellPrice)
plt.show()
```

C:\Users\kashan\Anaconda3\lib\site-packages\seaborn\distributions.py:2619: FutureWarning: `distplot` is a deprecated function and will be removed in a future version. Please adapt your code to use either `displot` (a figure-level function with similar flexibility) or `histplot` (an axes-level function for histograms).

warnings.warn(msg, FutureWarning)



## Which Brand is Most Expensive

```
In [23]: max_price=fashion['SellPrice'].max()
max_price
```

Out[23]: 28000

```
In [24]: brand=fashion.loc[ max_price == fashion['SellPrice'], 'Brand'].iloc[0]
brand
```

```
In [27]: max_price
```

Out[24]: tissot

Out[27]:

```
In [25]: category=fashion.loc[ max_price == fashion['SellPrice'], 'Main Category'].iloc[0]
category
```

Out[25]: 'Watches'

**Tissot is the most expensive brand in women category with price 28000 Rs**

```
In [26]: max_price=fashion.groupby(['Brand', 'Sub Category'])['SellPrice'].max().reset_index()
```

Brand	Sub Category	SellPrice
tissot	Watches	28000
porio armani	Watches	27995
michael kors	Watches	25995
fossil	Watches	23995
titan	Watches	16495
swarovski	Jewellery	15500
tommy hilfiger	Watches	13495
coach	Watches	11995
anne klein	Watches	11995

```
In [28]: sns.barplot(data=max_price,x='Brand',y='SellPrice')
plt.xticks(rotation=90)
plt.show()
```

```

brand
In [27]: max_price
Out[24]: tissot
Out[27]:
In [25]: category=fashion.loc[max_price == fashion['SellPrice'], 'Main Category'].iloc[0]
category    tissot    Watches    28000

```

```

Out[25]: 'women'
113    michael kors    Watches    25995
61     fossil        Watches    23995
160    titan         Watches    16495
154    swarovski     Jewellery   15500
161    tommy hilfiger Watches    13495

```

**Tissot is the most expensive brand in women category with price 28000 Rs**

```

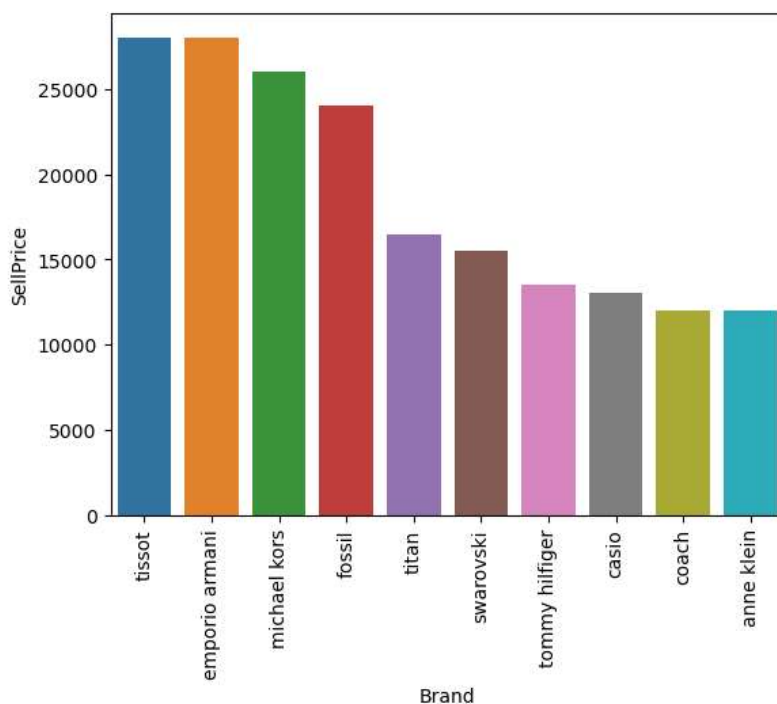
In [26]: max_price=fashion.groupby(['Brand', 'Sub Category'])['SellPrice'].max().reset_index()
33     coach    Watches    11995
15     anne klein    Watches    11995

```

```

In [28]: sns.barplot(data=max_price, x='Brand', y='SellPrice')
plt.xticks(rotation=90)
plt.show()

```



```

In [32]: min_price=fashion.groupby(['Brand', 'Sub Category'])['SellPrice'].min().reset_index()
Most Expensive Brand and Category is Tissot and watches
min_price

```

```

Out[29]: min_price=fashion['SellPrice'].min()
min_price

```

```

20    ayesha    Jewellery    89
89    insense  Lingerie&Nightwear    119
82

```

```

In [30]: min_brand=fashion.loc[min_price == fashion['SellPrice'], 'Brand'].iloc[0]
min_brand
145    sole    Lingerie&Nightwear    192

```

```

Out[30]: 'ayesha'
81    infuse    Westernwear    199

```

```

In [31]: min_category=fashion.loc[ min_price == fashion['SellPrice'], 'Main Category'].iloc[0]
min_category
150    stop    Indianwear    199

```

```

Out[31]: 'women'
83    insense    Westernwear    199
48    enamor    Lingerie&Nightwear    199
11    alllife    Westernwear    199

```

```

In [33]: sns.barplot(data=min_price, x='Brand', y='SellPrice')
plt.xticks(rotation=90)
plt.show()

```

```
In [32]: min_price=fashion.groupby(['Brand','Sub_Category'])['SellPrice'].min().reset_index()
Out[32]: min_price
```

```
In [29]: min_price=fashion['SellPrice'].min()
Out[29]: min_price
```

```
In [30]: min_brand=fashion['Brand'].min()
Out[30]: min_brand
```

```
In [31]: min_category=fashion['Main Category'].min()
Out[31]: min_category
```

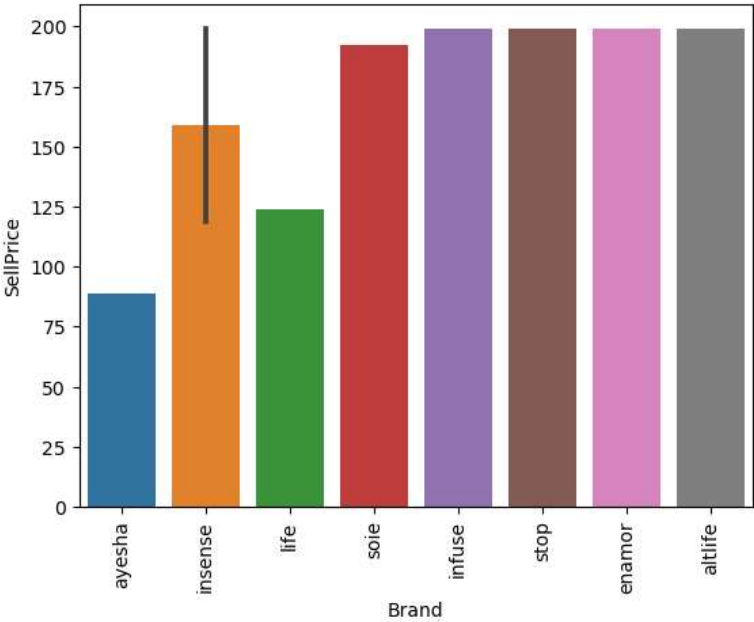
```
In [32]: min_price=fashion.groupby(['Brand','Sub_Category'])['SellPrice'].min().reset_index()
Out[32]: min_price
```

```
In [33]: min_price=fashion['SellPrice'].min()
Out[33]: min_price
```

```
In [34]: min_brand=fashion['Brand'].min()
Out[34]: min_brand
```

```
In [35]: min_category=fashion['Main Category'].min()
Out[35]: min_category
```

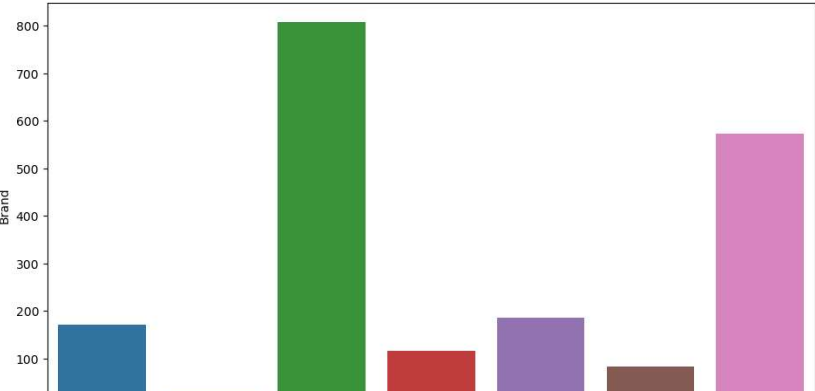
```
In [36]: sns.barplot(data=min_price,x='Brand',y='SellPrice')
plt.xticks(rotation=90)
plt.show()
```



The most cheapest brand and category are Ayesha and Jewellery

```
In [35]: plt.figure(figsize=(25,6))
In [34]: optional_clothing=fashion.groupby(['Sub_Category']).count().reset_index()
Out[34]: optional_clothing
```

```
In [36]: sns.barplot(x='Sub_Category',y='Brand',data=optional_clothing)
plt.xticks(rotation=90)
plt.show()
```

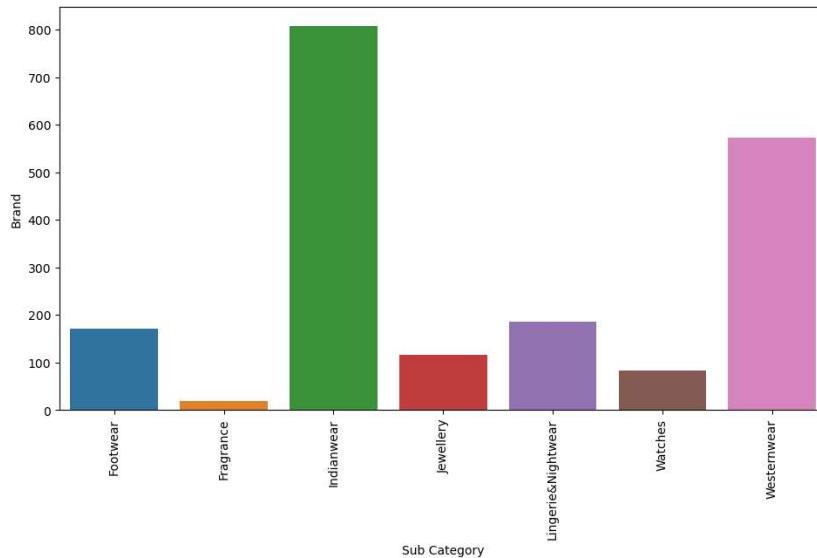




```
In [35]: plt.figure(figsize=(25,6))
In [34]: optional_clothing=fashion.groupby(['Sub Category']).count().reset_index()
optional_clothing
plt.show()
```

```
Out[34]:
sns.barplot(x='Sub Category', y='Brand', data=optional_clothing)
plt.xticks(rotation=90)
plt.show()
```

Sub Category	Brand	Details	Sizes	MRP	Sell Price	Discount	Main Category
0	Footwear	171	171	171	171	171	171
1	Fragrance	20	20	20	20	20	20
2	Indianwear	807	807	807	807	807	807

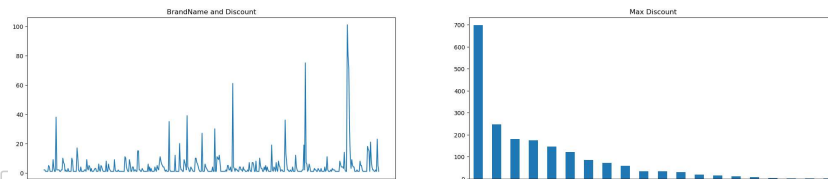


IndianWear category have max number of brands

```
In [66]: plt.figure(figsize=(25,5))

plt.subplot(1,2,1)
plt.title('BrandName and Discount')
fashion.groupby(['Brand'])['Discount'].value_counts().plot()

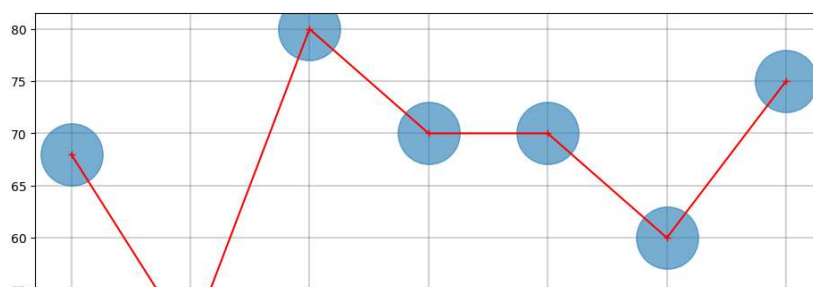
plt.subplot(1,2,2)
plt.title('Max Discount')
fashion['Discount'].value_counts().head(20).plot(kind='bar')
plt.show()
```

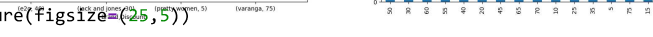


```
In [75]: plt.figure(figsize=(25,5))

plt.subplot(1,2,1)
max_dis=fashion.groupby(['Sub Category'])['Discount'].max()
max_dis
plt.grid(color='black',linewidth = 0.25)
plt.plot(max_dis,'.', alpha=0.6, markersize=50,marker='o')
plt.plot(max_dis,color='red',marker='+')
plt.show()
```

C:\Users\kashan\AppData\Local\Temp\ipykernel\_8332\776001359.py:7: UserWarning: marker is redundantly defined by the 'marker' keyword argument and the fmt string "." (-> marker='.'). The keyword argument will take precedence.  
 plt.plot(max\_dis,'.', alpha=0.6, markersize=50,marker='o')





```
plt.figure(figsize=(25,5))

plt.subplot(1,2,1)
max_dis=fashion.groupby(['Sub Category'])['Discount'].max()
max_dis
plt.grid(color='black',linewidth = 0.25)
plt.plot(max_dis,'.', alpha=0.6, markersize=50,marker='o')
plt.plot(max_dis,color='red',marker='+')
plt.show()
```

Category	Percentage
Footwear	68
Fragrance	50
Indianwear	80
Jewellery	70
Lingerie&Nightwear	70
Watches	60
Westernwear	75

```
plt.figure(figsize=(25,5))

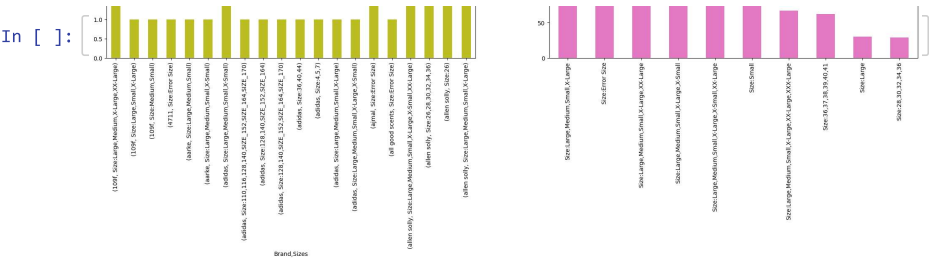
plt.subplot(1,2,1)
plt.title('Brand And Size')
size=fashion.groupby(['Brand'])['Sizes'].value_counts()
size.head(20).plot(kind='bar',color='#bcbd22')

plt.subplot(1,2,2)
plt.title('Size')
size2=fashion['Sizes'].value_counts().head(10)
size2.plot(kind='bar',color='#e377c2')

plt.show()
```



localhost:8888/notebooks/Data Visual Practice/Fashion Ecommerce EDA.ipynb



As fashion industry is progressing, the brands are trying their best to fit into every customer size from small to XL sizes as well. Thus, from the above analysis it is clear that though the gap is shortneing but their is still high demand for sizes in Small,X-Small, Medium, Large and X-Large