Import Necessary Liabraies

In [ ]:
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np
```

Load the data set

In [ ]:
```
df=pd.read_csv(r"C:\Users\DELL\Desktop\New folder (2)/FashionDataset.csv")
df.head()
```

Out[ ]:

	Unnamed: 0	BrandName	Deatils	Sizes	MRP	SellPrice	Disco
0	0	life	solid cotton blend collar neck womens a-line d...	Size:Large,Medium,Small,X-Large,X-Small	Rs\n1699	849	50%
1	1	only	polyester peter pan collar womens blouson dres...	Size:34,36,38,40	Rs\n3499	2449	30%
2	2	fratini	solid polyester blend wide neck womens regular...	Size:Large,X-Large,XX-Large	Rs\n1199	599	50%
3	3	zink london	stripes polyester sweetheart neck womens dress...	Size:Large,Medium,Small,X-Large	Rs\n2299	1379	40%
4	4	life	regular fit regular length denim womens jeans ...	Size:26,28,30,32,34,36	Rs\n1699	849	50%

Check data types of data

In [ ]:
```
df.info()
```

```
<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: int64(1), object(7)
memory usage: 1.9+ MB
```

In [ ]:
```
df.shape
```

Out[ ]:
```
(30758, 8)
```

In [ ]:
```
df.columns
```

Out[ ]:
```
Index(['Unnamed: 0', 'BrandName', 'Deatils', 'Sizes', 'MRP', 'SellPrice',
       'Discount', 'Category'],
      dtype='object')
```

rename columns names

In [ ]:
```
df.rename(columns={"Deatils":"Details"}, inplace=True)
df.columns
```

Out[ ]:
```
Index(['Unnamed: 0', 'BrandName', 'Details', 'Sizes', 'MRP', 'SellPrice',
       'Discount', 'Category'],
      dtype='object')
```

Data Cleaning

In [ ]:
```
# remove Unnames colums
df.drop(columns=['Unnamed: 0'],inplace=True)
df
```

Out[ ]:

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

30758 rows × 7 columns

Check null values in to data

In [ ]:
```
df.isnull().any()
```

Out[ ]:
```
BrandName    False
Details      False
Sizes        False
MRP          False
SellPrice    False
Discount     False
Category     False
dtype: bool
```

Missing Values Treatment:

Check null values and remove it

In [ ]:
```
nan_value=df.loc[df["BrandName"]=="Nan"]
nan_value
```

Out[ ]:

	BrandName	Details	Sizes	MRP	SellPrice	Discount	Category
22	Nan	Nan	Nan	Nan	Nan	Nan	Westernwear-Women
48	Nan	Nan	Nan	Nan	Nan	Nan	Westernwear-Women
74	Nan	Nan	Nan	Nan	Nan	Nan	Westernwear-Women
100	Nan	Nan	Nan	Nan	Nan	Nan	Westernwear-Women
126	Nan	Nan	Nan	Nan	Nan	Nan	Westernwear-Women
...	...	...	...	...	...	...	...
30650	Nan	Nan	Nan	Nan	Nan	Nan	Jewellery-Women
30676	Nan	Nan	Nan	Nan	Nan	Nan	Jewellery-Women
30702	Nan	Nan	Nan	Nan	Nan	Nan	Jewellery-Women
30728	Nan	Nan	Nan	Nan	Nan	Nan	Jewellery-Women
30754	Nan	Nan	Nan	Nan	Nan	Nan	Jewellery-Women

1183 rows × 7 columns

Replace Nan in to blank and remove columns where values are not fill

In [ ]:
```
df.replace('Nan',np.nan,inplace=True)
df.dropna(axis=0,inplace=True)
```

In [ ]:
```
df.isnull().sum()
```

Out[ ]:
```
BrandName    0
Details      0
Sizes        0
MRP          0
SellPrice    0
Discount     0
Category     0
dtype: int64
```

Change data types of columns ---SellPrice

In [ ]:
```
df["SellPrice"]=df["SellPrice"].astype(int)
df['MRP'] = df['MRP'].str.replace('Rs\n','')
df["Discount"]=df["Discount"].str.replace("% off","")
df['MRP'] = df['MRP'].astype(int)
df['Sizes']=df['Sizes'].str.replace("Size:","")
df["Category"]=df["Category"].str.replace("-Women","")
df=df.astype({"Discount":"float","MRP": "float","SellPrice":"float"})
```

In [ ]:
```
df.head(5)
```

Out[ ]:

	BrandName	Details	Sizes	MRP	SellPrice	Discount	Category
0	life	solid cotton blend collar neck womens a-line d...	Large,Medium,Small,X-Large,X-Small	1699.0	849.0	50.0	Westernwear
1	only	polyester peter pan collar womens blouson dres...	34,36,38,40	3499.0	2449.0	30.0	Westernwear
2	fratini	solid polyester blend wide neck womens regular...	Large,X-Large,XX-Large	1199.0	599.0	50.0	Westernwear
3	zink london	stripes polyester sweetheart neck womens dress...	Large,Medium,Small,X-Large	2299.0	1379.0	40.0	Westernwear
4	life	regular fit regular length denim womens jeans ...	26,28,30,32,34,36	1699.0	849.0	50.0	Westernwear

done cleaning data is ready to next step of get insights

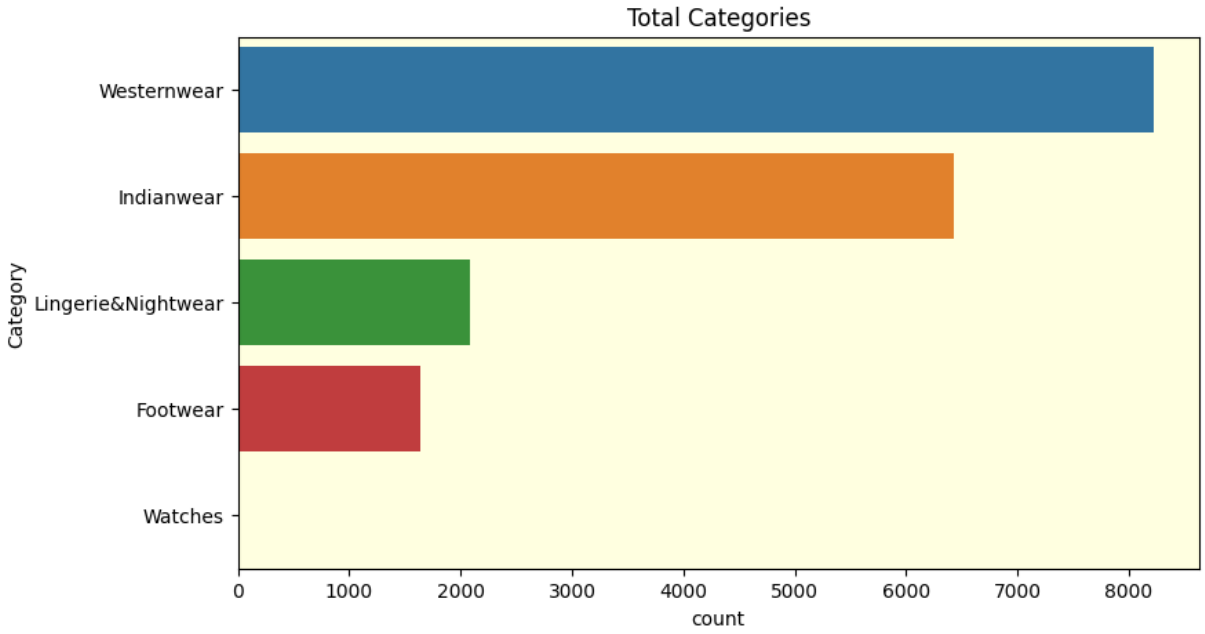
1.Categories

In [ ]:
```
category=df["Category"].value_counts()
category
```

Out[ ]:
```
Westernwear           8221
Indianwear            6428
Lingerie&Nightwear    2084
Footwear              1638
Watches                  3
Name: Category, dtype: int64
```

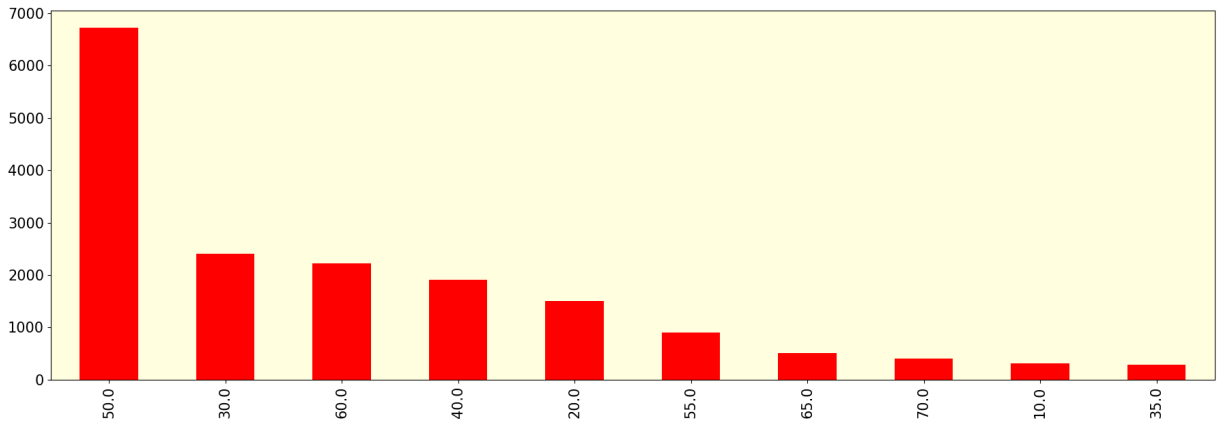
In [ ]:
```
plt.figure(figsize=(9,5))
sns.countplot(data=df,y="Category").set_facecolor("lightyellow")
```

```
plt.title("Total Categories")
plt.show()
```



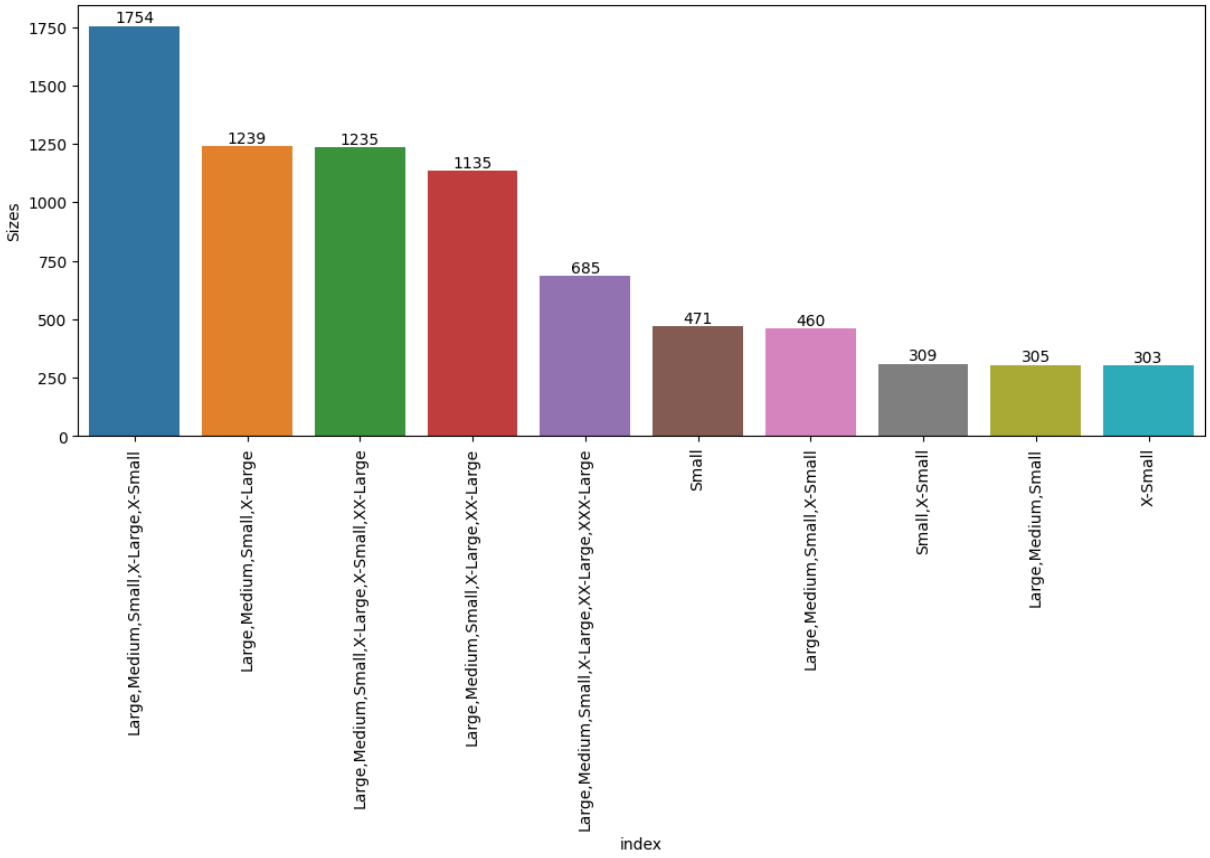
Western-Women Product -----Discount Graph

In [ ]:
```
plt.figure(figsize=(22,7))
a=df['Discount'].value_counts().head(10).plot(kind='bar' , color = 'red',fontsize=1
```



Analyze for Size columns

In [ ]:
```
plt.figure(figsize=(13,5))
a=df.Sizes.value_counts().head(10).reset_index()
chart=sns.barplot(x='index', y='Sizes', data=a,errwidth=0)
plt.xticks(rotation=90)
chart.bar_label(chart.containers[0])
plt.show()
```



Visuliaze High sellPrice

In [ ]:
```
hp=df[['BrandName','SellPrice' , 'Category']].sort_values(by='SellPrice', ascending
hp.style.background_gradient(cmap="coolwarm")
```

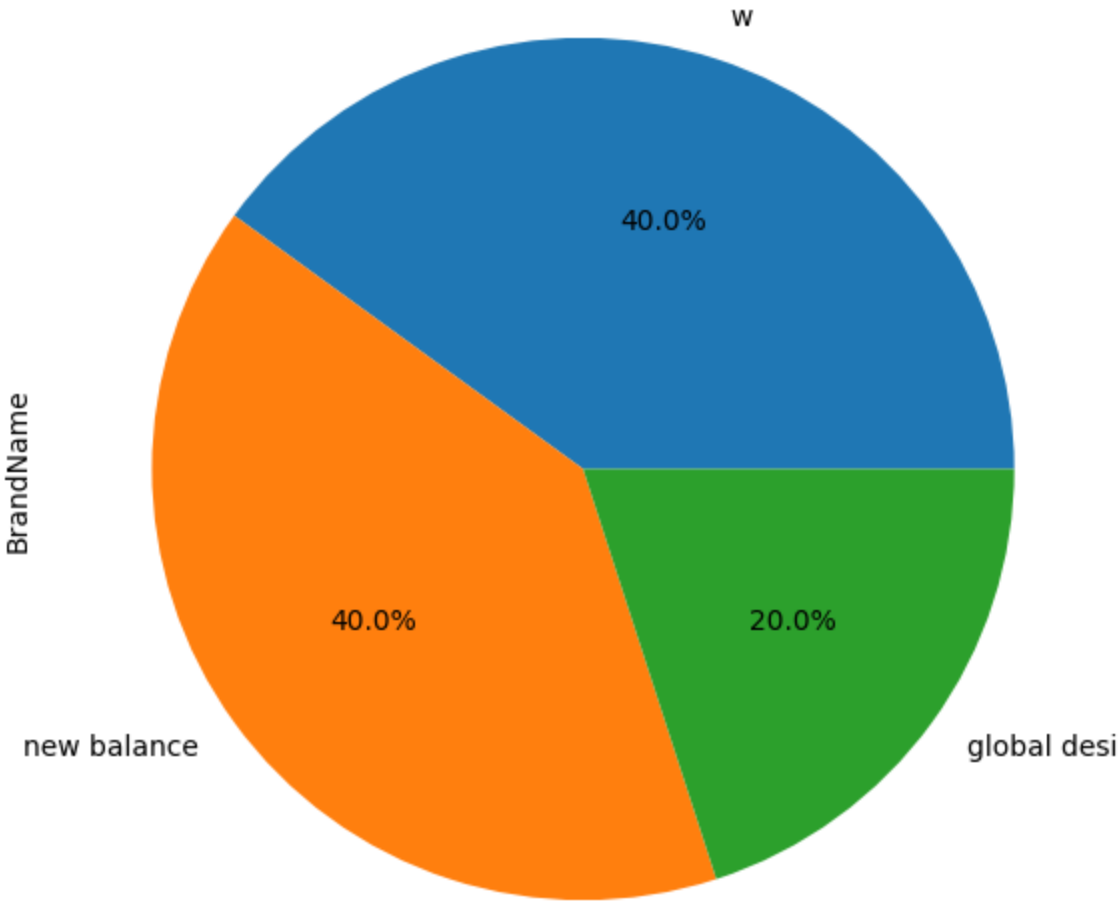
Out[ ]:

	BrandName	SellPrice	Category
15744	global desi	13599.000000	Indianwear
14154	global desi	12799.000000	Indianwear
18629	w	7499.000000	Indianwear
24281	new balance	6999.000000	Footwear
24641	new balance	6999.000000	Footwear
25635	new balance	6999.000000	Footwear
24742	new balance	6999.000000	Footwear
14053	w	6749.000000	Indianwear
16718	w	6749.000000	Indianwear
15053	w	6749.000000	Indianwear

Top costly BrandName

In [ ]:
```
hp["BrandName"].value_counts().plot(kind="pie",figsize=(12,7),autopct='%1.1f%%')
```

Out[ ]: <Axes: ylabel='BrandName'>



Highest Brand Sale-

In [ ]:
```
cdf=df.groupby("BrandName").agg(high_sale=("SellPrice", "sum"))
cdf=cdf.sort_values(by="high_sale", ascending=False).reset_index()
cdf
```

Out[ ]:

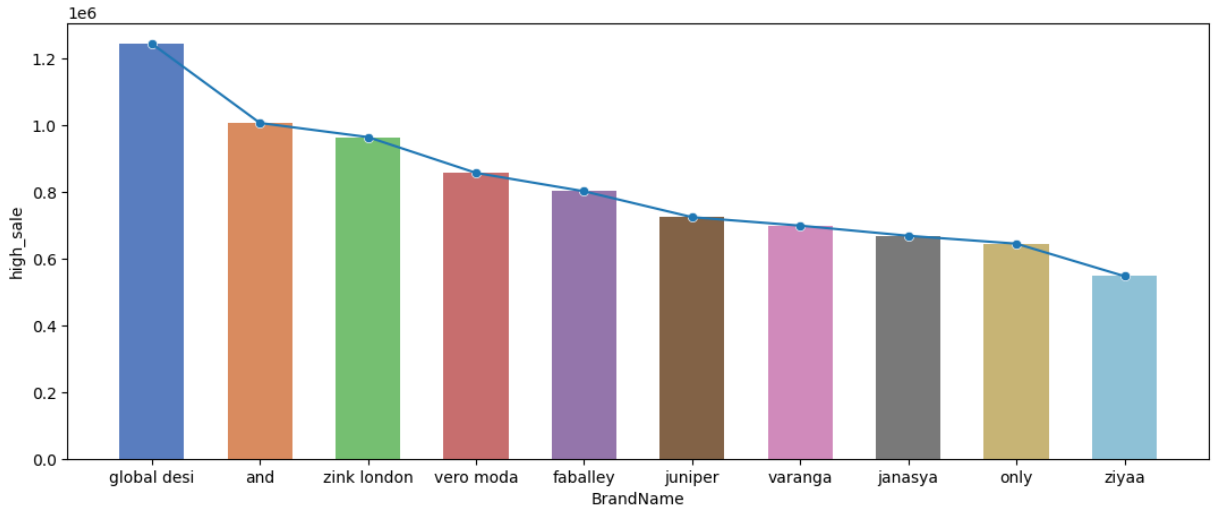
	BrandName	high_sale
0	global desi	1243094.0
1	and	1005731.0
2	zink london	963353.0
3	vero moda	856105.0
4	faballey	801103.0
...	...	...
100	nayomi	847.0
101	charchit	841.0
102	u.s. polo assn.	750.0
103	mothercare	699.0
104	gas	624.0

105 rows × 2 columns

In [ ]:
```
s=pd.DataFrame(cdf)
s=s.head(10)
```

In [ ]:
```
plt.figure(figsize=(13,5))
sns.barplot(data=s,x="BrandName",y="high_sale", palette="muted",width=0.6)
sns.lineplot(data=s,x="BrandName",y="high_sale")
sns.scatterplot(data=s,x="BrandName",y="high_sale")
```

Out[ ]:
```
<Axes: xlabel='BrandName', ylabel='high_sale'>
```



In [ ]:
```
df
```

Out[ ]:

	BrandName	Details	Sizes	MRP	SellPrice	Discount	Cate
0	life	solid cotton blend collar neck womens a-line d...	Large,Medium,Small,X-Large,X-Small	1699.0	849.0	50.0	Western
1	only	polyester peter pan collar womens blouson dres...	34,36,38,40	3499.0	2449.0	30.0	Western
2	fratini	solid polyester blend wide neck womens regular...	Large,X-Large,XX-Large	1199.0	599.0	50.0	Western
3	zink london	stripes polyester sweetheart neck womens dress...	Large,Medium,Small,X-Large	2299.0	1379.0	40.0	Western
4	life	regular fit regular length denim womens jeans ...	26,28,30,32,34,36	1699.0	849.0	50.0	Western
...	...	...	...	...	...	...	
26673	lemon & pepper	womens casual wear buckle closure flats - navy	36,37,38,39,40	2999.0	1499.0	50.0	Foot
26674	haute curry	womens casual wear slip on heels - black	36,37,38,39,40	2199.0	1099.0	50.0	Foot
26885	swiss eagle	womens analogue metallic watch	Error Size	13990.0	4197.0	70.0	Wa

	BrandName	Details	Sizes	MRP	SellPrice	Discount	Cate
27290	lawman watches	womens rose gold dial stainless steel analogue...	Error Size	7499.0	4999.0	33.0	Wa
28418	lawman watches	womens silver dial stainless steel analogue wa...	Error Size	5999.0	3999.0	33.0	Wa

Which discount range attracts the user?

In [ ]:
```
def dis_range(Discount):
    if(Discount<=10):
        return "0-10"
    elif(Discount<=20):
        return "11-20"
    elif(Discount<=30):
        return "21-30"
    elif(Discount<=40):
        return "31-40"
    elif(Discount<=50):
        return "41-50"
    elif(Discount<=60):
        return "51-60"
    elif(Discount<=70):
        return "61-70"
    elif(Discount<=80):
        return "71-80"
    else:
        return "Above 80 "
```

In [ ]:
```
df["discount_range"]=df['Discount'].apply(dis_range)
df
```

Out[ ]:

	BrandName	Details	Sizes	MRP	SellPrice	Discount	Cate
0	life	solid cotton blend collar neck womens a-line d...	Large,Medium,Small,X-Large,X-Small	1699.0	849.0	50.0	Western
1	only	polyester peter pan collar womens blouson dres...	34,36,38,40	3499.0	2449.0	30.0	Western
2	fratini	solid polyester blend wide neck womens regular...	Large,X-Large,XX-Large	1199.0	599.0	50.0	Western
3	zink london	stripes polyester sweetheart neck womens dress...	Large,Medium,Small,X-Large	2299.0	1379.0	40.0	Western
4	life	regular fit regular length denim womens jeans ...	26,28,30,32,34,36	1699.0	849.0	50.0	Western
...	...	...	...	...	...	...	
26673	lemon & pepper	womens casual wear buckle closure flats - navy	36,37,38,39,40	2999.0	1499.0	50.0	Foot
26674	haute curry	womens casual wear slip on heels - black	36,37,38,39,40	2199.0	1099.0	50.0	Foot
26885	swiss eagle	womens analogue metallic watch	Error Size	13990.0	4197.0	70.0	Wa

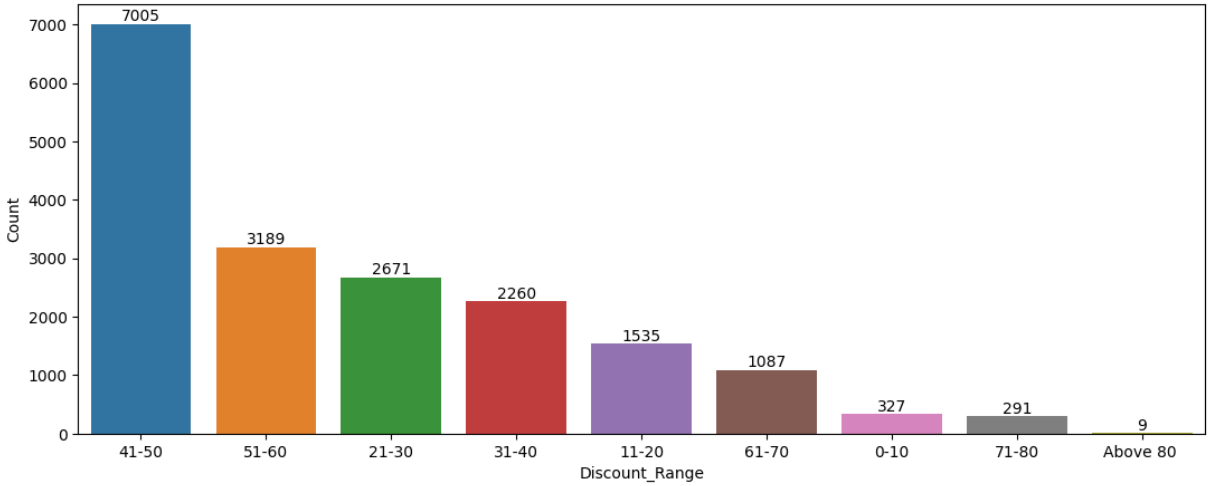
	BrandName	Details	Sizes	MRP	SellPrice	Discount	Cate
27290	lawman watches	womens rose gold dial stainless steel analogue...	Error Size	7499.0	4999.0	33.0	Wa
28418	lawman watches	womens silver dial stainless steel analogue wa...	Error Size	5999.0	3999.0	33.0	Wa

In [ ]:
```
c=df.discount_range.value_counts().reset_index()
c=c.rename(columns={"index":"Discount_Range","discount_range":"Count"})
c
```

Out[ ]:

	Discount_Range	Count
0	41-50	7005
1	51-60	3189
2	21-30	2671
3	31-40	2260
4	11-20	1535
5	61-70	1087
6	0-10	327
7	71-80	291
8	Above 80	9

In [ ]:
```
plt.figure(figsize=(13,5))
d=sns.barplot(data=c,x="Discount_Range",y="Count")
# d.bar_label(d.containers[0])
for count in d.containers:
    d.bar_label(count)
```



Conclusion : 41-50% discount is more attract to the user

Add new column "Discount Range" in the dataframe using pd.cut

In [ ]:
```
m=df.Discount.max()
a=list(range(1,90,5))
print(len(a))
print(f" Max Dis:{m},List Range: {a}")
```

```
18
 Max Dis:85.0,List Range: [1, 6, 11, 16, 21, 26, 31, 36, 41, 46, 51, 56, 61, 66, 71,
76, 81, 86]
```

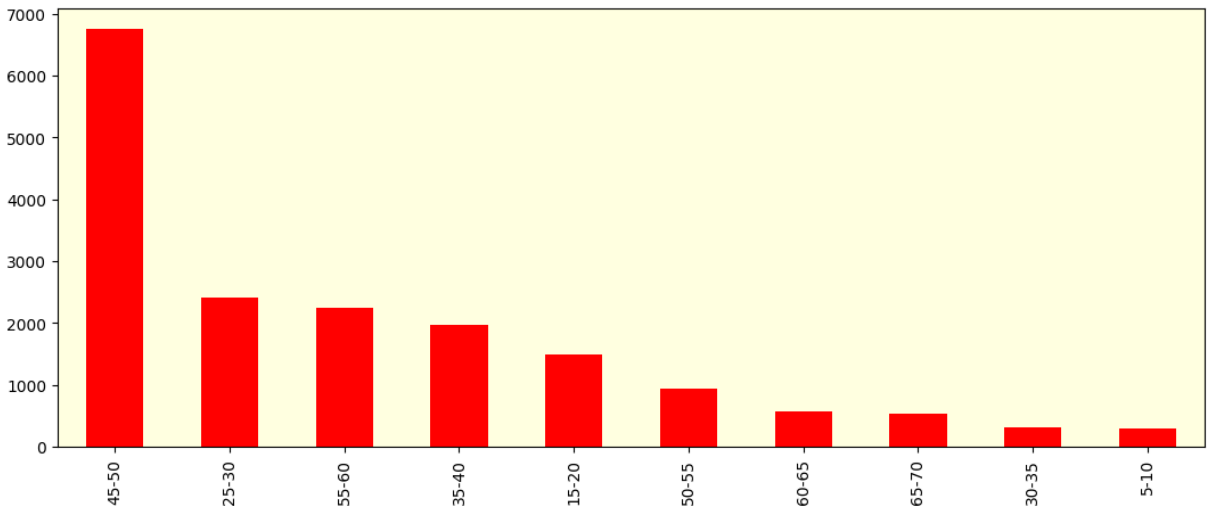
In [ ]:
```
l=["0-5","5-10","10-15","15-20","20-25","25-30","30-35","35-40","40-45","45-50","50
print(len(l))
df["Discountrange"]=pd.cut(x=df["Discount"],bins=a,labels=l)
df.head()
```

```
17
```

Out[ ]:

	BrandName	Details	Sizes	MRP	SellPrice	Discount	Category
0	life	solid cotton blend collar neck womens a-line d...	Large,Medium,Small,X-Large,X-Small	1699.0	849.0	50.0	Westernwear
1	only	polyester peter pan collar womens blouson dres...	34,36,38,40	3499.0	2449.0	30.0	Westernwear
2	fratini	solid polyester blend wide neck womens regular...	Large,X-Large,XX-Large	1199.0	599.0	50.0	Westernwear
3	zink london	stripes polyester sweetheart neck womens dress...	Large,Medium,Small,X-Large	2299.0	1379.0	40.0	Westernwear
4	life	regular fit regular length denim womens jeans ...	26,28,30,32,34,36	1699.0	849.0	50.0	Westernwear

In [ ]:
```
n=df.Discountrange.value_counts().head(10).plot(kind="bar", color="red", figsize=(1
```



In [ ]:
```
df.columns
```

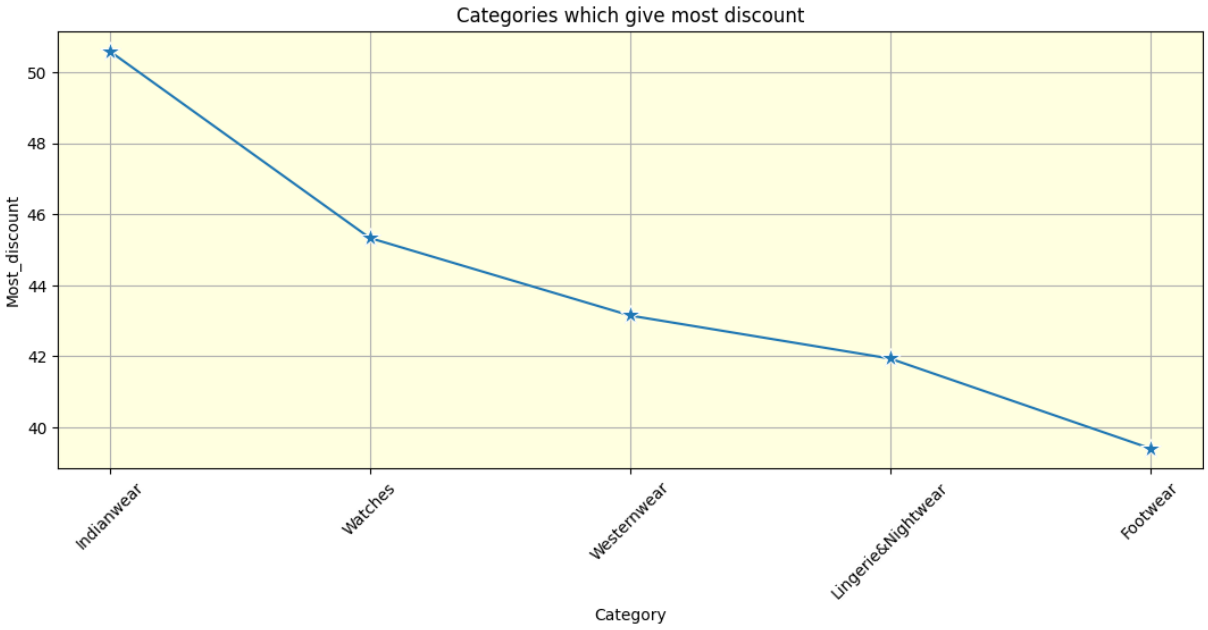
Out[ ]:
```
Index(['BrandName', 'Details', 'Sizes', 'MRP', 'SellPrice', 'Discount',
       'Category', 'discount_range', 'Discountrange'],
      dtype='object')
```

In [ ]:
```
s=df.groupby( "Category").agg(
    Most_discount=("Discount", "mean")
).reset_index().sort_values(by="Most_discount",ascending=False)
s.head()
```

Out[ ]:

	Category	Most_discount
1	Indianwear	50.591008
3	Watches	45.333333
4	Westernwear	43.151198
2	Lingerie&Nightwear	41.939060
0	Footwear	39.410256

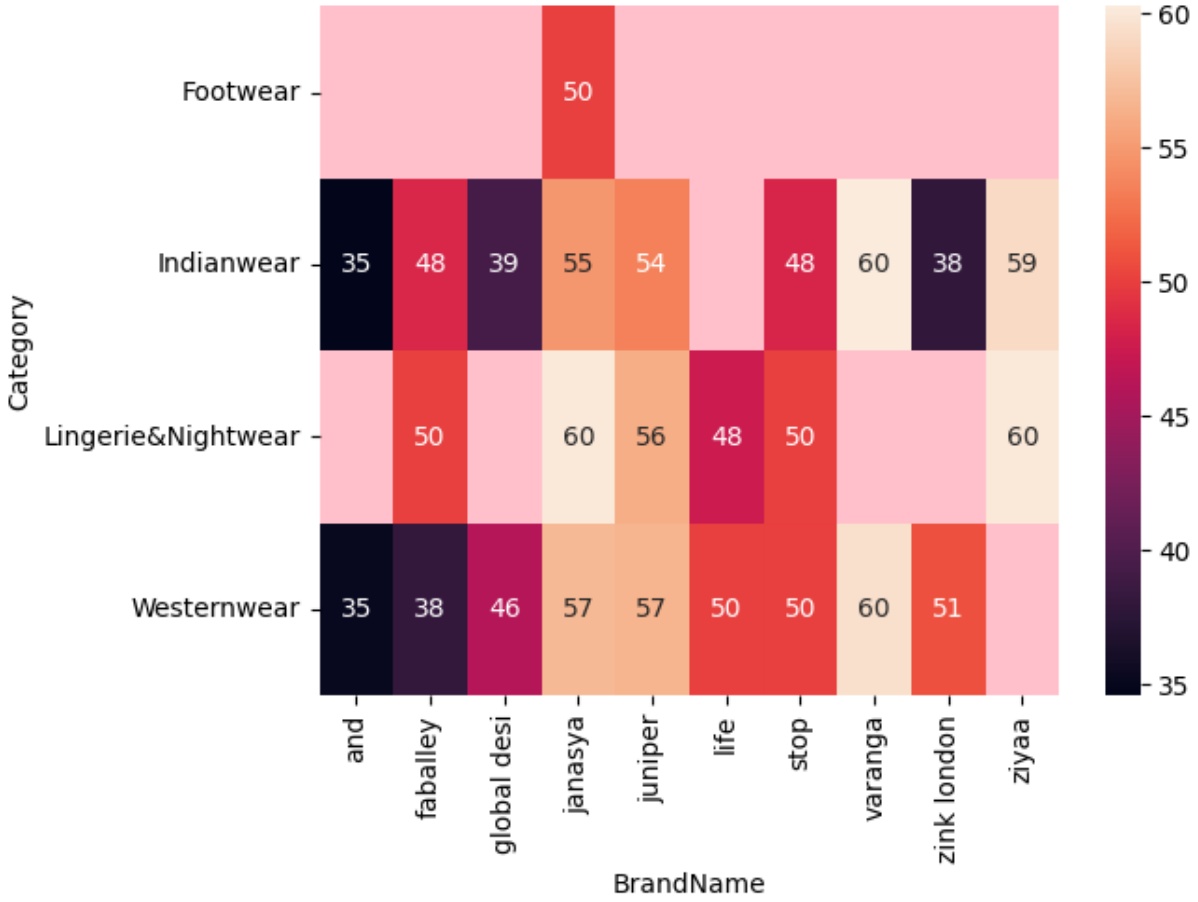
In [ ]:
```
plt.figure(figsize=(13,5))
sns.lineplot(data=s,x="Category",y="Most_discount",marker= "*", markersize=13).set_
plt.xticks(rotation=45)
plt.title("Categories which give most discount")
plt.grid()
plt.show()
```



Conclusion : Indianwear is most discounted category of the product

Correlation Analysis

In [ ]:
```
top_brand=df["BrandName"].value_counts().head(10).index.to_list()
heat_data=df[df["BrandName"].isin(top_brand)].pivot_table(
    index="Category",
    columns="BrandName",
    values="Discount")
sns.heatmap(data=heat_data, annot=True, color='blue').set_facecolor("pink")
```



In [ ]: