

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

```
amazon=pd.read_csv("/content/amazon.csv")
```

amazon

	product_id	product_name	category	discounted_price	av
0	B07JW9H4J1	Wayona Nylon Braided USB to Lightning Fast Cha...	Computers&Accessories Accessories&Peripherals ...	₹399	
1	B098NS6PVG	Ambrane Unbreakable 60W / 3A Fast Charging 1.5...	Computers&Accessories Accessories&Peripherals ...	₹199	
2	B096MSW6CT	Sounce Fast Phone Charging Cable & Data Sync U...	Computers&Accessories Accessories&Peripherals ...	₹199	
3	B08HDJ86NZ	boAt Deuce USB 300 2 in 1 Type-C & Micro USB S...	Computers&Accessories Accessories&Peripherals ...	₹329	
4	B08CF3B7N1	Portronics Konnect L 1.2M Fast Charging 3A 8 P...	Computers&Accessories Accessories&Peripherals ...	₹154	
...
1460	B08L7J3T31	Noir Aqua - 5pcs PP Spun Filter + 1 Spanner ...	Home&Kitchen Kitchen&HomeAppliances WaterPurif...	₹379	
1461	B01M6453MB	Prestige Delight PRWO Electric Rice Cooker (1 ...	Home&Kitchen Kitchen&HomeAppliances SmallKitch...	₹2,280	
1462	B009P2LIL4	Bajaj Majesty RX10 2000 Watts Heat Convector R...	Home&Kitchen Heating,Cooling&AirQuality RoomHe...	₹2,219	
1463	B00J5DYCCA	Havells Ventil Air DSP 230mm Exhaust Fan (Pist...	Home&Kitchen Heating,Cooling&AirQuality Fans E...	₹1,399	
1464	B01486F4G6	Borosil Jumbo 1000-Watt Grill Sandwich Maker (...)	Home&Kitchen Kitchen&HomeAppliances SmallKitch...	₹2,863	

1465 rows × 6 columns

Next steps: [New interactive sheet](#)

amazon.head()

	product_id	product_name	category	discounted_price	actual_p
0	B07JW9H4J1	Wayona Nylon Braided USB to Lightning Fast Cha...	Computers&Accessories Accessories&Peripherals ...	₹399	₹
1	B098NS6PVG	Ambrane Unbreakable 60W / 3A Fast Charging 1.5...	Computers&Accessories Accessories&Peripherals ...	₹199	
2	B096MSW6CT	Sounce Fast Phone Charging Cable & Data Sync U...	Computers&Accessories Accessories&Peripherals ...	₹199	₹
3	B08HDJ86NZ	boAt Deuce USB 300 2 in 1 Type-C & Micro USB S...	Computers&Accessories Accessories&Peripherals ...	₹329	
4	B08CF3B7N1	Portronics Konnect L 1.2M Fast Charging 3A 8 P...	Computers&Accessories Accessories&Peripherals ...	₹154	

Next steps: [New interactive sheet](#)

amazon.tail()

	product_id	product_name	category	discounted_price	actu
1460	B08L7J3T31	Noir Aqua - 5pcs PP Spun Filter + 1 Spanner ...	Home&Kitchen Kitchen&HomeAppliances WaterPurif...	₹379	
1461	B01M6453MB	Prestige Delight PRWO Electric Rice Cooker (1 ...	Home&Kitchen Kitchen&HomeAppliances SmallKitch...	₹2,280	
1462	B009P2LIL4	Bajaj Majesty RX10 2000 Watts Heat Convector R...	Home&Kitchen Heating,Cooling&AirQuality RoomHe...	₹2,219	
1463	B00J5DYCCA	Havells Ventil Air DSP 230mm Exhaust Fan (Pist...	Home&Kitchen Heating,Cooling&AirQuality Fans E...	₹1,399	
1464	B01486F4G6	Borosil Jumbo 1000-Watt Grill Sandwich Maker (...)	Home&Kitchen Kitchen&HomeAppliances SmallKitch...	₹2,863	

amazon.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1465 entries, 0 to 1464
Data columns (total 16 columns):
#   Column                Non-Null Count  Dtype
---  -
0   product_id            1465 non-null   object
1   product_name          1465 non-null   object
2   category              1465 non-null   object
3   discounted_price      1465 non-null   object
4   actual_price          1465 non-null   object
5   discount_percentage   1465 non-null   object
6   rating                1465 non-null   object
7   rating_count          1463 non-null   object
8   about_product         1465 non-null   object
9   user_id               1465 non-null   object
10  user_name             1465 non-null   object
11  review_id             1465 non-null   object
12  review_title          1465 non-null   object
13  review_content        1465 non-null   object
14  img_link              1465 non-null   object
15  product_link          1465 non-null   object
dtypes: object(16)
memory usage: 183.3+ KB
```

```
amazon.isnull().sum()
```

	0
product_id	0
product_name	0
category	0
discounted_price	0
actual_price	0
discount_percentage	0
rating	0
rating_count	2
about_product	0
user_id	0
user_name	0
review_id	0
review_title	0
review_content	0
img_link	0
product_link	0

```
dtype: int64
```

```
amazon['actual_price'] = pd.to_numeric(
    amazon['actual_price'].astype(str)
    .str.replace('₹,', '', regex=True),
    errors='coerce'
)

amazon['discount_percentage'] = pd.to_numeric(
    amazon['discount_percentage'].astype(str)
    .str.replace('%', '', regex=False),
    errors='coerce'
)

amazon['discounted_price'] = pd.to_numeric(
```

```

amazon['discounted_price'].astype(str)
.str.replace('₹', '', regex=False)
.str.replace(',', '', regex=False),
errors='coerce'
)

amazon['rating_count'] = pd.to_numeric(
amazon['rating_count'].astype(str)
.str.replace(',', '', regex=False),
errors='coerce'
)

amazon['rating'] = pd.to_numeric(
amazon['rating'].astype(str)
.str.replace(',', '', regex=False),
errors='coerce'
)

```

amazon.dtypes

	0
product_id	object
product_name	object
category	object
discounted_price	float64
actual_price	float64
discount_percentage	int64
rating	float64
rating_count	float64
about_product	object
user_id	object
user_name	object
review_id	object
review_title	object
review_content	object
img_link	object
product_link	object

dtype: object

```
amazon.dropna(subset=[ 'actual_price', 'discounted_price', ], inplace=True)
```

```
amazon.drop_duplicates(inplace=True)
```

Feature engineering

```

#discount amount
amazon['dis_amount'] = amazon['actual_price'] - amazon['discounted_price']
print(amazon['dis_amount'])

```

0	700.0
1	150.0

```

2      1700.0
3      370.0
4      245.0
...
1460    540.0
1461    765.0
1462    861.0
1463    491.0
1464    827.0
Name: dis_amount, Length: 1465, dtype: float64

```

```

#range category of price
amazon['price_range']=pd.cut(amazon['discounted_price'],bins=[0,500,2000,5000,20000,100000],labels=[

```

```

#estimated revenue
if 'rating_count' in amazon.columns:
    amazon['estimated_revenue']=amazon['discounted_price']*amazon['rating_count']

```

EDA

```
amazon.describe()
```

	discounted_price	actual_price	discount_percentage	rating	rating_count	dis_amount
count	1465.000000	1465.000000	1465.000000	1464.000000	1463.000000	1465.000000
mean	3125.310874	5444.990635	47.691468	4.096585	18295.541353	2319.679761
std	6944.304394	10874.826864	21.635905	0.291674	42753.864952	4604.473790
min	39.000000	39.000000	0.000000	2.000000	2.000000	0.000000
25%	325.000000	800.000000	32.000000	4.000000	1186.000000	371.000000
50%	799.000000	1650.000000	50.000000	4.100000	5179.000000	800.000000
75%	1999.000000	4295.000000	63.000000	4.300000	17336.500000	1953.000000
max	77990.000000	139900.000000	94.000000	5.000000	426973.000000	61910.000000

```
amazon['category'].value_counts().head(5)
```

	count
Computers&Accessories Accessories&Peripherals Cables&Accessories Cables USBCables	233
Electronics WearableTechnology SmartWatches	76
Electronics Mobiles&Accessories Smartphones&BasicMobiles Smartphones	68
Electronics HomeTheater,TV&Video Televisions SmartTelevisions	63
Electronics Headphones,Earbuds&Accessories Headphones In-Ear	52

```
dtype: int64
```

```
#avg rating by categories
```

```
amazon.groupby('category')['rating'].mean().sort_values(ascending=False).head()
```

	rating
category	
Computers&Accessories Tablets	4.6
Computers&Accessories NetworkingDevices NetworkAdapters PowerLANAdapters	4.5
Electronics Cameras&Photography Accessories Film	4.5
Computers&Accessories Components Memory	4.5
Electronics HomeAudio MediaStreamingDevices StreamingClients	4.5

dtype: float64

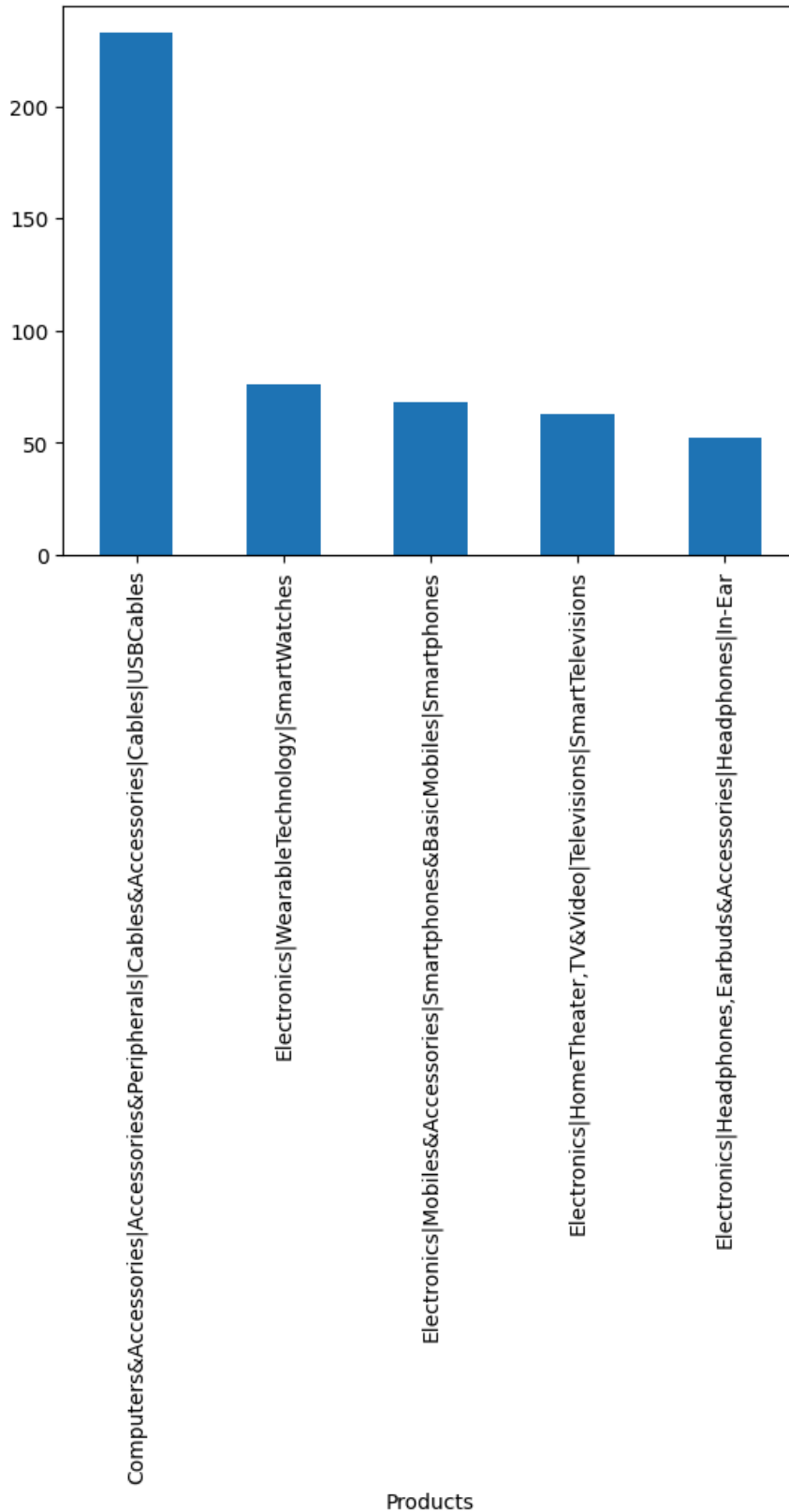
```
amazon.sort_values(by='discounted_price',ascending=False).head()
```

	product_id	product_name	category	discounted_price	actual
249	B09WN3SRC7	Sony Bravia 164 cm (65 inches) 4K Ultra HD Sma...	Electronics HomeTheater,TV&Video Televisions S...	77990.0	13
325	B095JPKPH3	OnePlus 163.8 cm (65 inches) U Series 4K LED S...	Electronics HomeTheater,TV&Video Televisions S...	61999.0	6
255	B0BC8BQ432	VU 164 cm (65 inches) The GloLED Series 4K Sma...	Electronics HomeTheater,TV&Video Televisions S...	54990.0	8
192	B0B15GSPQW	Samsung 138 cm (55 inches) Crystal 4K Neo Seri...	Electronics HomeTheater,TV&Video Televisions S...	47990.0	7
283	B0B3XXSB1K	LG 139 cm (55 inches) 4K Ultra HD Smart LED TV...	Electronics HomeTheater,TV&Video Televisions S...	47990.0	7

Dashboard Visualizations

```
#category distribution
amazon['category'].value_counts().head(5).plot(kind='bar')
plt.xlabel("Products")
plt.ylabel("Counts")
plt.title("TOP 5 CATEGORIES")
plt.show()
```

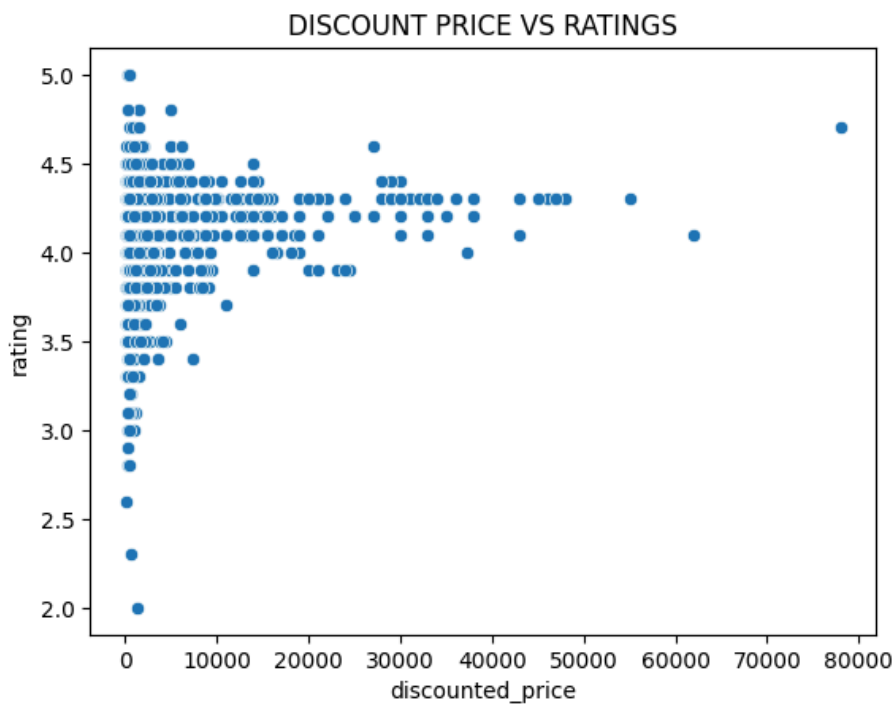
```
-----  
AttributeError                                Traceback (most recent call last)  
/tmp/ipynthon-input-574604856.py in <cell line: 0>()  
    2 amazon['category'].value_counts().head(5).plot(kind='bar')  
    3 plt.xlabel("Products")  
----> 4 plt.ylabel("Counts")  
    5 plt.title("TOP 5 CATEGORIES")  
    6 plt.show()  
  
AttributeError: module 'matplotlib.pyplot' has no attribute 'ylabel'
```



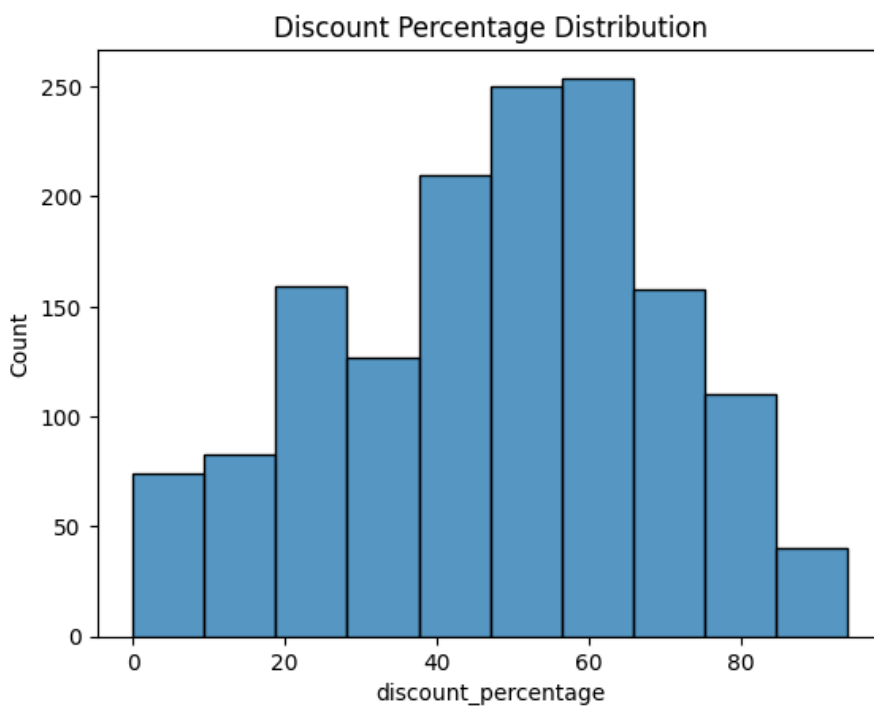
```
#rating distribution  
sns.histplot(amazon['rating'],bins=10)
```

```
plt.title("Rating Distribution")  
plt.show()
```

```
#discount price vs rating  
sns.scatterplot(x='discounted_price',y='rating',data=amazon)  
plt.title("DISCOUNT PRICE VS RATINGS")  
plt.show()
```



```
#Eiscount% distribution  
sns.histplot(amazon['discount_percentage'], bins=10)  
plt.title("Discount Percentage Distribution")  
plt.show()
```



Double-click (or enter) to edit

correlation analysis


```
sns.heatmap(amazon.corr(numeric_only=True),annot=True)
plt.title("Correlation matrix")
plt.show()
```

