```
import pandas as pd
import seaborn as snas
import matplotlib.pyplot as plt
```

Step 1

loading dataset

```
df=pd.read_csv(r"C:\\Users\\ishus\\Downloads\\BigBasket Products.csv")
```

```
df.shape
```

```
(27555, 10)
```

```
df1=df.copy()
```

```
df1
```

```
       index                                                product  \
0          1             Garlic Oil - Vegetarian Capsule 500 mg
1          2                           Water Bottle - Orange
2          3                    Brass Angle Deep - Plain, No.2
3          4  Cereal Flip Lid Container/Storage Jar - Assort...
4          5                 Creme Soft Soap - For Hands & Body
...      ...                                                ...
27550  27551         Wottagirl! Perfume Spray - Heaven, Classic
27551  27552                                           Rosemary
27552  27553                      Peri-Peri Sweet Potato Chips
27553  27554                         Green Tea - Pure Original
27554  27555                     United Dreams Go Far Deodorant

                     category              sub_category  \
0           Beauty & Hygiene                 Hair Care
1      Kitchen, Garden & Pets     Storage & Accessories
2        Cleaning & Household               Pooja Needs
3        Cleaning & Household      Bins & Bathroom Ware
4           Beauty & Hygiene          Bath & Hand Wash
...                      ...                       ...
27550       Beauty & Hygiene         Fragrances & Deos
27551   Gourmet & World Food    Cooking & Baking Needs
27552   Gourmet & World Food  Snacks, Dry Fruits, Nuts
27553              Beverages                       Tea
27554       Beauty & Hygiene            Men's Grooming

                          brand  sale_price  market_price  \
0             Sri Sri Ayurveda       220.00         220.0
1                    Mastercook      180.00         180.0
2                           Trm      119.00         250.0
3                        Nakoda      149.00         176.0
```

```
4                            Nivea      162.00         162.0
...                            ...         ...           ...
27550                      Layerr       199.20         249.0
27551                    Puramate        67.50          75.0
27552                      FabBox       200.00         200.0
27553                      Tetley       396.00         495.0
27554  United Colors Of Benetton        214.53         390.0

                            type  rating  \
0              Hair Oil & Serum     4.1
1        Water & Fridge Bottles     2.3
2               Lamp & Lamp Oil     3.4
3      Laundry, Storage Baskets     3.7
4          Bathing Bars & Soaps     4.4
...                          ...     ...
27550                    Perfume     3.9
27551  Herbs, Seasonings & Rubs     4.0
27552            Nachos & Chips     3.8
27553                  Tea Bags     4.2
27554          Men's Deodorants     4.5

                                              description
0      This Product contains Garlic Oil that is known...
1      Each product is microwave safe (without lid), ...
2      A perfect gift for all occasions, be it your m...
3      Multipurpose container with an attractive desi...
4      Nivea Creme Soft Soap gives your skin the best...
...                                                  ...
27550  Layerr brings you Wottagirl Classic fragrant b...
27551  Puramate rosemary is enough to transform a dis...
27552  We have taken the richness of Sweet Potatoes (...
27553  Tetley Green Tea with its refreshing pure, ori...
27554  The new mens fragrance from the United Dreams ...

[27555 rows x 10 columns]
```

Step 2

getting first 12 rows of dataset

```
df1.head(12)
```

```
    index                                            product  \
0       1            Garlic Oil - Vegetarian Capsule 500 mg
1       2                           Water Bottle - Orange
2       3                    Brass Angle Deep - Plain, No.2
3       4  Cereal Flip Lid Container/Storage Jar - Assort...
4       5                Creme Soft Soap - For Hands & Body
```

```
5       6                  Germ - Removal Multipurpose Wipes
6       7                                      Multani Mati
7       8                  Hand Sanitizer - 70% Alcohol Base
8       9  Biotin & Collagen Volumizing Hair Shampoo + Bi...
9      10               Scrub Pad - Anti- Bacterial, Regular
10     11                           Wheat Grass Powder - Raw
11     12                     Butter Cookies Gold Collection

                   category            sub_category               brand
\
0         Beauty & Hygiene               Hair Care  Sri Sri Ayurveda

1   Kitchen, Garden & Pets   Storage & Accessories         Mastercook

2     Cleaning & Household             Pooja Needs                Trm

3     Cleaning & Household    Bins & Bathroom Ware              Nakoda

4         Beauty & Hygiene        Bath & Hand Wash               Nivea

5     Cleaning & Household    All Purpose Cleaners     Nature Protect

6         Beauty & Hygiene               Skin Care           Satinance

7         Beauty & Hygiene        Bath & Hand Wash             Bionova

8         Beauty & Hygiene               Hair Care         StBotanica

9     Cleaning & Household  Mops, Brushes & Scrubs       Scotch brite

10    Gourmet & World Food  Cooking & Baking Needs          NUTRASHIL

11    Gourmet & World Food   Chocolates & Biscuits           Sapphire


    sale_price  market_price                           type  rating  \
0        220.0         220.0              Hair Oil & Serum     4.1
1        180.0         180.0        Water & Fridge Bottles     2.3
2        119.0         250.0                Lamp & Lamp Oil     3.4
3        149.0         176.0      Laundry, Storage Baskets     3.7
4        162.0         162.0          Bathing Bars & Soaps     4.4
5        169.0         199.0  Disinfectant Spray & Cleaners     3.3
6         58.0          58.0                      Face Care     3.6
7        250.0         250.0        Hand Wash & Sanitizers     4.0
8       1098.0        1098.0          Shampoo & Conditioner     3.5
9         20.0          20.0       Utensil Scrub-Pad, Glove     4.3
10       261.0         290.0             Flours & Pre-Mixes     4.0
11       600.0         600.0       Luxury Chocolates, Gifts     2.2

                                          description
0   This Product contains Garlic Oil that is known...
```

```
1   Each product is microwave safe (without lid), ...
2   A perfect gift for all occasions, be it your m...
3   Multipurpose container with an attractive desi...
4   Nivea Creme Soft Soap gives your skin the best...
5   Stay protected from contamination with Multipu...
6   Satinance multani matti is an excellent skin t...
7   70%Alcohol based is gentle of hand leaves skin...
8   An exclusive blend with Vitamin B7 Biotin, Hyd...
9   Scotch Brite Anti- Bacterial Scrub Pad thoroug...
10  Wheatgrass is a superfood potent health food w...
11  Enjoy a tin full of delicious butter cookies m...
```

Step 3

Getting the description of data

```
df1.describe()
```

```
              index     sale_price  market_price        rating
count  27555.00000   27549.000000  27555.000000  18919.000000
mean   13778.00000     334.648391    382.056664      3.943295
std     7954.58767    1202.102113    581.730717      0.739217
min        1.00000       2.450000      3.000000      1.000000
25%     6889.50000      95.000000    100.000000      3.700000
50%    13778.00000     190.320000    220.000000      4.100000
75%    20666.50000     359.000000    425.000000      4.300000
max    27555.00000  112475.000000  12500.000000      5.000000
```

Step 4

Getting info about the dataset

```
df1.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 27555 entries, 0 to 27554
Data columns (total 10 columns):
 #   Column        Non-Null Count  Dtype  
---  ------        --------------  -----  
 0   index         27555 non-null  int64  
 1   product       27554 non-null  object 
 2   category      27555 non-null  object 
 3   sub_category  27555 non-null  object 
 4   brand         27554 non-null  object 
 5   sale_price    27549 non-null  float64
 6   market_price  27555 non-null  float64
```

```
 7   type          27555 non-null  object
 8   rating        18919 non-null  float64
 9   description   27440 non-null  object
dtypes: float64(3), int64(1), object(6)
memory usage: 2.1+ MB
```

Step 5

Find out Top & least sold products

```
top_sold=df1.sort_values(by='rating',ascending=False).head(10)

least_sold=df1.sort_values(by='rating',ascending=True).head(10)

print('10 top sold product:\n')
print(top_sold[['product', 'brand', 'category', 'rating',
'sale_price']])
```

```
10 top sold product:

                                                product
brand  \
27515                           EDT Spray - Musk For Men
Brut
12                         Face Wash - Oil Control, Active
Oxy
16                          Smooth Skin Oil - For Dry Skin    Aroma
Treasures
6133                            Peri Peri Stir-Fry Sauce      As Chefs
Cook
6178         Disney Frozen Sisters Metal Kids Pencil Box   Hm
International
6183                       Hand Made Organic Charcoal Soap      Barber's
Club
5935                      Amarant Cheela Mix - Gluten Free         Nutty
Yogi
5947   Hot Schezwan Dipping Sauce - Tasty & Nutrient-...            D-
Alive
5950                                      De Tan Face Pack
Nutrinorm
27445  Organic Virgin Coconut Oil - Pure Coldpressed ...  Morpheme
Remedies

                    category  rating  sale_price
27515      Beauty & Hygiene     5.0       595.0
12         Beauty & Hygiene     5.0       110.0
16         Beauty & Hygiene     5.0       324.0
6133   Gourmet & World Food     5.0       216.0
```

```
6178   Cleaning & Household     5.0       149.0
6183       Beauty & Hygiene     5.0       174.3
5935   Gourmet & World Food     5.0       149.0
5947   Gourmet & World Food     5.0       599.0
5950       Beauty & Hygiene     5.0       249.0
27445      Beauty & Hygiene     5.0       427.0
```

```
print('10 lest sold product:\n')
print(least_sold[['product', 'brand', 'category', 'rating',
'sale_price']])
```

```
10 lest sold product:

                                                product
brand  \
13324                                Tea - Black, Kadak CTC
Octavius
16579                                     Clear Décor Clips
Command
16552                      Glass Set - Long Drink, Allure  Sanjeev
Kapoor
27285  Dry Shampoo - Instant Hair Refresh, Coconut & ...
Batiste
3687                                  Caramel Chocolate Bar
Rage
16941                                          Coconut Oil     ST.
D'VENCE
16863                                   Piping Gel, Neutral
Foodecor
23399                           100% Natural Intimate Wipes
Plush
10453              Milk Chocolate - Salted Peanut Praline
Uns
10473                                     Joy Square Glass
Saura

                      category  rating  sale_price
13324                Beverages     1.0       390.0
16579  Kitchen, Garden & Pets     1.0       179.0
16552  Kitchen, Garden & Pets     1.0       445.0
27285       Beauty & Hygiene     1.0       649.0
3687     Gourmet & World Food     1.0       220.0
16941       Beauty & Hygiene     1.0       254.0
16863    Gourmet & World Food     1.0       104.5
23399       Beauty & Hygiene     1.0       120.0
10453  Snacks & Branded Foods     1.0       295.0
10473  Kitchen, Garden & Pets     1.0       109.0
```

Step 6

Measuring discount on a certain item.

```
df1["discount"]=((df1['market_price']-df1['sale_price'])/
df1['market_price'])*100

print(df1[['product','sale_price','market_price','discount']])
```

```
                                                product
sale_price  \
0                 Garlic Oil - Vegetarian Capsule 500 mg      220.00

1                                   Water Bottle - Orange      180.00

2                          Brass Angle Deep - Plain, No.2      119.00

3      Cereal Flip Lid Container/Storage Jar - Assort...      149.00

4                    Creme Soft Soap - For Hands & Body      162.00

...                                                  ...         ...

27550         Wottagirl! Perfume Spray - Heaven, Classic      199.20

27551                                           Rosemary       67.50

27552                       Peri-Peri Sweet Potato Chips      200.00

27553                          Green Tea - Pure Original      396.00

27554                       United Dreams Go Far Deodorant      214.53


       market_price   discount
0             220.0   0.000000
1             180.0   0.000000
2             250.0  52.400000
3             176.0  15.340909
4             162.0   0.000000
...             ...        ...
27550         249.0  20.000000
27551          75.0  10.000000
27552         200.0   0.000000
27553         495.0  20.000000
27554         390.0  44.992308

[27555 rows x 4 columns]
```

```
df1
```

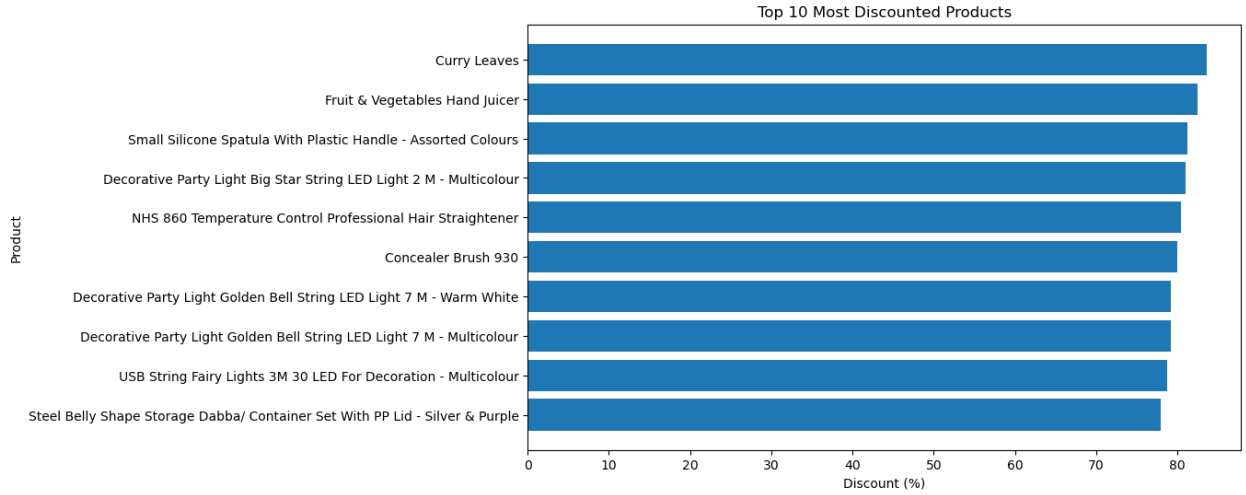
```
       index                                              product  \
0          1              Garlic Oil - Vegetarian Capsule 500 mg
1          2                                 Water Bottle - Orange
2          3                          Brass Angle Deep - Plain, No.2
3          4  Cereal Flip Lid Container/Storage Jar - Assort...
4          5                  Creme Soft Soap - For Hands & Body
...      ...                                                 ...
27550  27551         Wottagirl! Perfume Spray - Heaven, Classic
27551  27552                                           Rosemary
27552  27553                       Peri-Peri Sweet Potato Chips
27553  27554                          Green Tea - Pure Original
27554  27555                       United Dreams Go Far Deodorant

                     category              sub_category  \
0           Beauty & Hygiene                 Hair Care
1      Kitchen, Garden & Pets     Storage & Accessories
2         Cleaning & Household              Pooja Needs
3         Cleaning & Household      Bins & Bathroom Ware
4           Beauty & Hygiene          Bath & Hand Wash
...                       ...                       ...
27550       Beauty & Hygiene         Fragrances & Deos
27551    Gourmet & World Food    Cooking & Baking Needs
27552    Gourmet & World Food  Snacks, Dry Fruits, Nuts
27553              Beverages                       Tea
27554       Beauty & Hygiene            Men's Grooming

                          brand  sale_price  market_price  \
0              Sri Sri Ayurveda      220.00         220.0
1                     Mastercook     180.00         180.0
2                           Trm      119.00         250.0
3                        Nakoda      149.00         176.0
4                         Nivea      162.00         162.0
...                         ...         ...           ...
27550                    Layerr      199.20         249.0
27551                  Puramate       67.50          75.0
27552                    FabBox      200.00         200.0
27553                    Tetley      396.00         495.0
27554  United Colors Of Benetton      214.53         390.0

                          type  rating  \
0             Hair Oil & Serum     4.1
1       Water & Fridge Bottles     2.3
2               Lamp & Lamp Oil     3.4
3      Laundry, Storage Baskets     3.7
4          Bathing Bars & Soaps     4.4
...                         ...     ...
27550                  Perfume     3.9
27551  Herbs, Seasonings & Rubs     4.0
27552           Nachos & Chips     3.8
27553                Tea Bags     4.2
```

```
27554          Men's Deodorants     4.5
```

	description	discount
0	This Product contains Garlic Oil that is known...	0.000000
1	Each product is microwave safe (without lid), ...	0.000000
2	A perfect gift for all occasions, be it your m...	52.400000
3	Multipurpose container with an attractive desi...	15.340909
4	Nivea Creme Soft Soap gives your skin the best...	0.000000
...	...	...
27550	Layerr brings you Wottagirl Classic fragrant b...	20.000000
27551	Puramate rosemary is enough to transform a dis...	10.000000
27552	We have taken the richness of Sweet Potatoes (...	0.000000
27553	Tetley Green Tea with its refreshing pure, ori...	20.000000
27554	The new mens fragrance from the United Dreams ...	44.992308

```
[27555 rows x 11 columns]
```

```
top_10_discounted = (df1.sort_values(by='discount',
ascending=False).head(10))
plt.figure(figsize=(10,6))
plt.barh(top_10_discounted['product'],top_10_discounted['discount'])
plt.xlabel('Discount (%)')
plt.ylabel('Product')
plt.title('Top 10 Most Discounted Products')
plt.gca().invert_yaxis()   # highest discount on top
plt.show()
```



Step 7

Find out the Missing Values from the Dataset

```
df1.isnull().sum()
```

```
index              0
product            1
category           0
sub_category       0
brand              1
sale_price         6
market_price       0
type               0
rating          8636
description      115
discount           6
dtype: int64
```

```
df1['rating']=df1['rating'].fillna(df1['rating'].mean())
```

```
df1['rating'].isnull().sum()
```

```
np.int64(0)
```

```
df1['sale_price']=df1['sale_price'].fillna(df1['sale_price'].mean())
```

```
df1['sale_price'].isnull().sum()
```

```
np.int64(0)
```

```
df1['description']=df1['description'].fillna(df1['description'].mode()
[0])
```

```
df1['description'].isnull().sum()
```

```
np.int64(0)
```

```
df1['product']=df1['product'].fillna(df1['product'].mode()[0])
```

```
df1['brand']=df1['brand'].fillna(df1['brand'].mode()[0])
```

```
df1['brand'].isnull().sum()
```

```
np.int64(0)
```

```
df1.isnull().sum()
```

```
index           0
product         0
category        0
sub_category    0
brand           0
sale_price      0
market_price    0
type            0
rating          0
description     0
```

```
discount         6
dtype: int64
```

Step 8

Find out the outliers from the dataset according to the columnn

```
numeric_cols = ['sale_price', 'market_price', 'rating']
for col in numeric_cols:
    Q1 = df[col].quantile(0.25)
    Q3 = df[col].quantile(0.75)
    IQR = Q3 - Q1

lower_bound = Q1 - 1.5 * IQR
upper_bound = Q3 + 1.5 * IQR

outliers = df[(df[col] < lower_bound) | (df[col] > upper_bound)]

outliers['sale_price']
```

```
1          180.00
11         600.00
29         169.15
43         109.00
64         168.75
           ...
27415      700.00
27493      359.10
27498     1071.00
27503      169.00
27518      350.00
Name: sale_price, Length: 1312, dtype: float64
```

```
outliers['market_price']
```

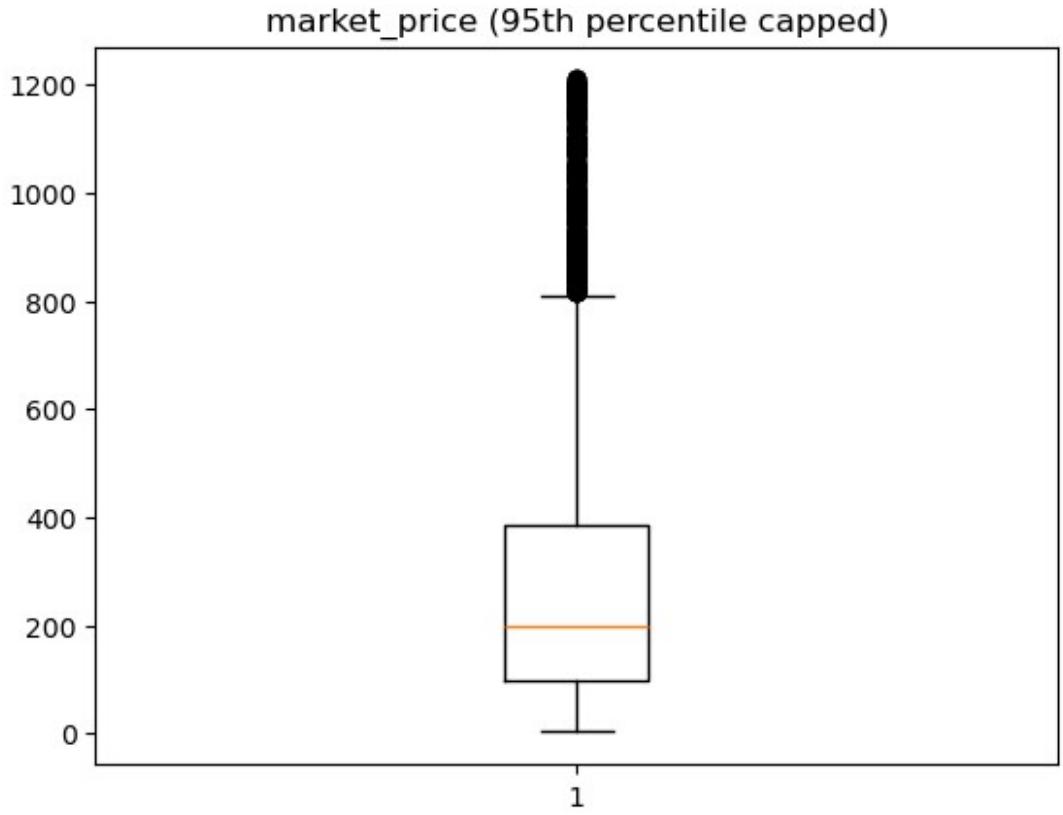
```
1          180.0
11         600.0
29         199.0
43         299.0
64         225.0
           ...
27415      700.0
27493      399.0
27498     1071.0
27503      246.0
27518      350.0
Name: market_price, Length: 1312, dtype: float64
```

```
outliers['rating']
```

```
1         2.3
11        2.2
29        1.5
43        2.6
64        1.0
         ...
27415     2.0
27493     2.8
27498     2.0
27503     2.7
27518     1.0
Name: rating, Length: 1312, dtype: float64
```

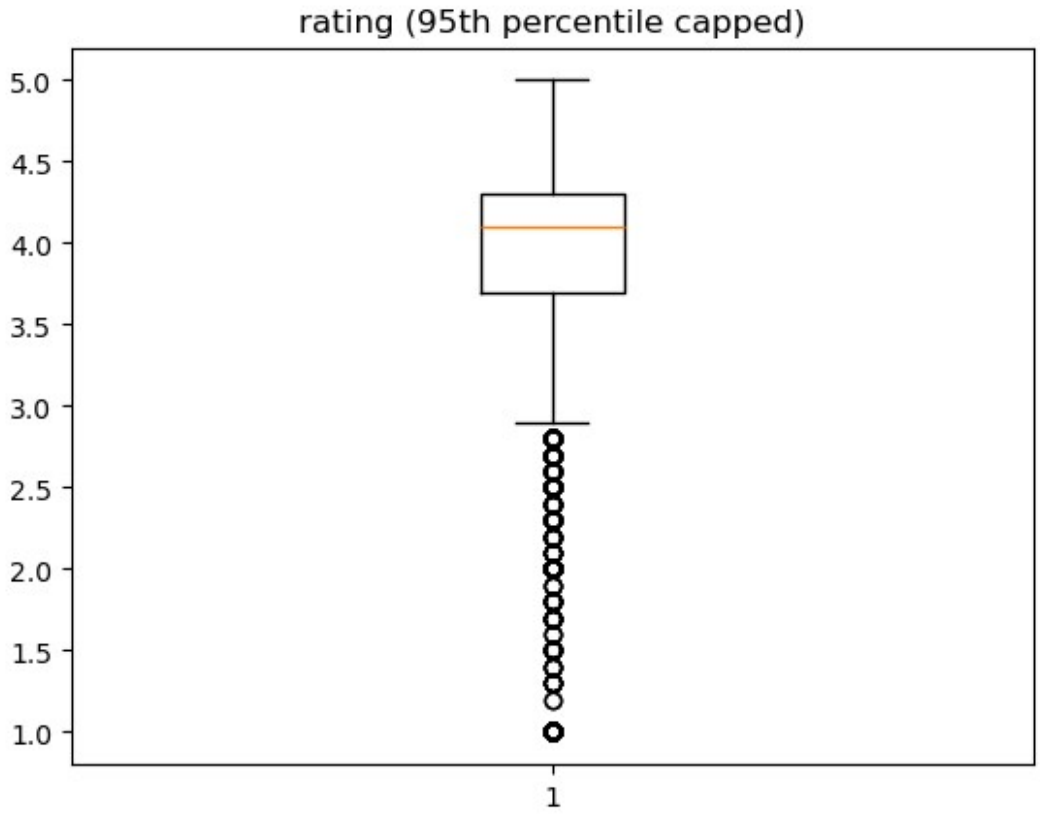
The boxplot visualizes the distribution of the market_price column and highlights the presence and approximate location of outliers beyond the whiskers. For better readability, values above the 95th percentile were capped, meaning extreme values beyond this threshold are not displayed in the plot. This improves visualization clarity without altering the underlying data

```
upper_cap = df['market_price'].quantile(0.95)
plt.boxplot(df[df['market_price'] <= upper_cap]['market_price'])
plt.title('market_price (95th percentile capped)')
plt.show()
```



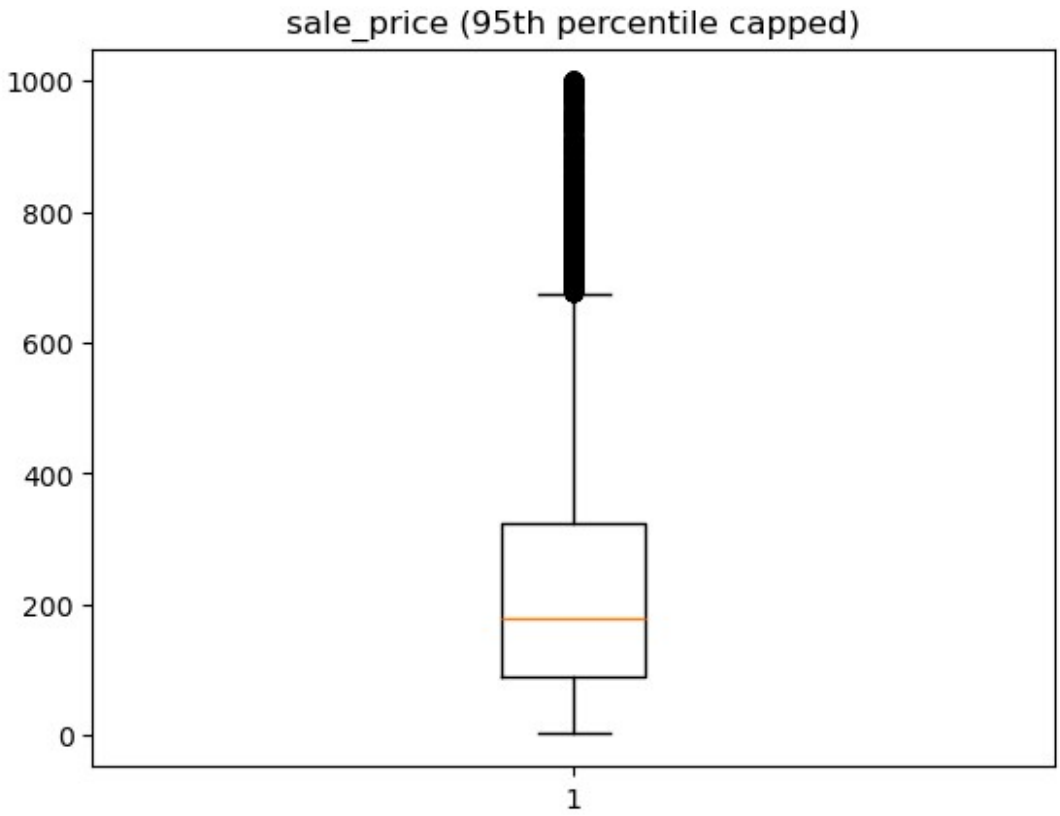
The boxplot visualizes the distribution of the rating column and highlights the presence and approximate location of outliers beyond the whiskers. For better readability, values above the 95th percentile were capped, meaning extreme values beyond this threshold are not displayed in the plot. This improves visualization clarity without altering the underlying data

```
upper_cap = df['rating'].quantile(0.95)
plt.boxplot(df[df['rating'] <= upper_cap]['rating'])
plt.title('rating (95th percentile capped)')
plt.show()
```



The boxplot visualizes the distribution of the sale_price column and highlights the presence and approximate location of outliers beyond the whiskers. For better readability, values above the 95th percentile were capped, meaning extreme values beyond this threshold are not displayed in the plot. This improves visualization clarity without altering the underlying data

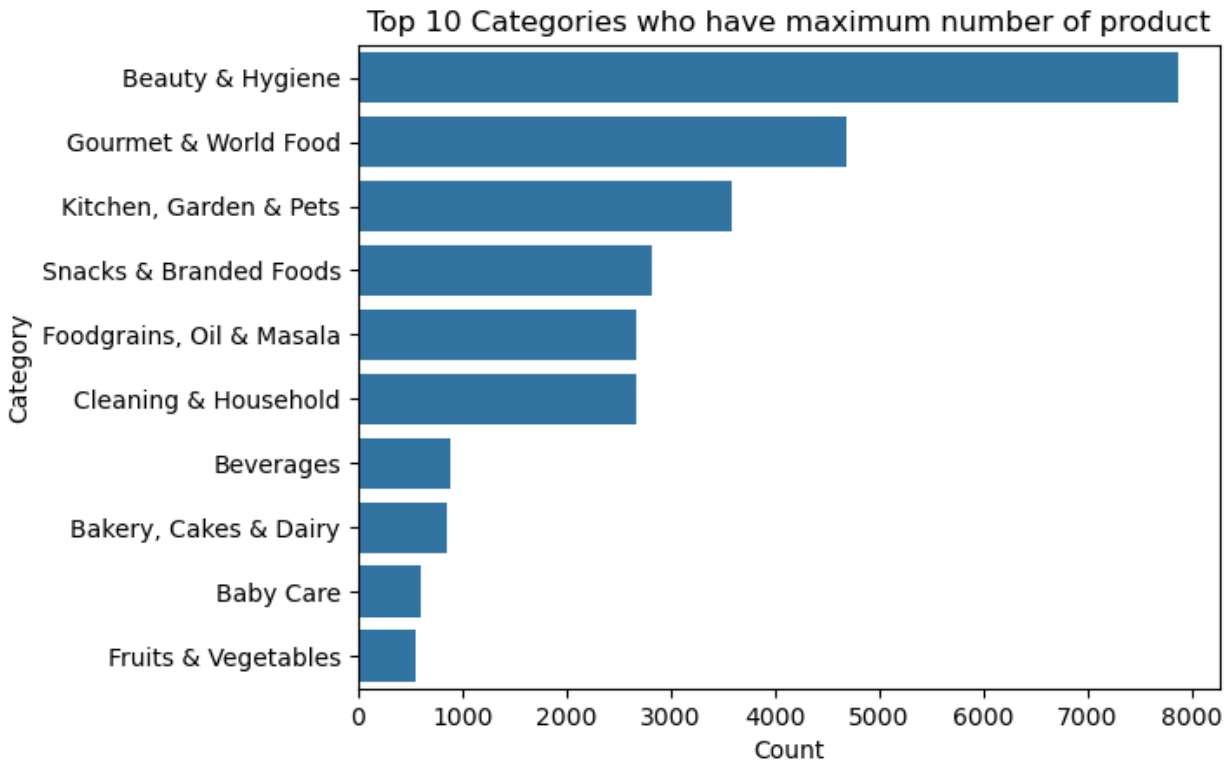
```
upper_cap = df['sale_price'].quantile(0.95)
plt.boxplot(df[df['sale_price'] <= upper_cap]['sale_price'])
plt.title('sale_price (95th percentile capped)')
plt.show()
```



Step 9

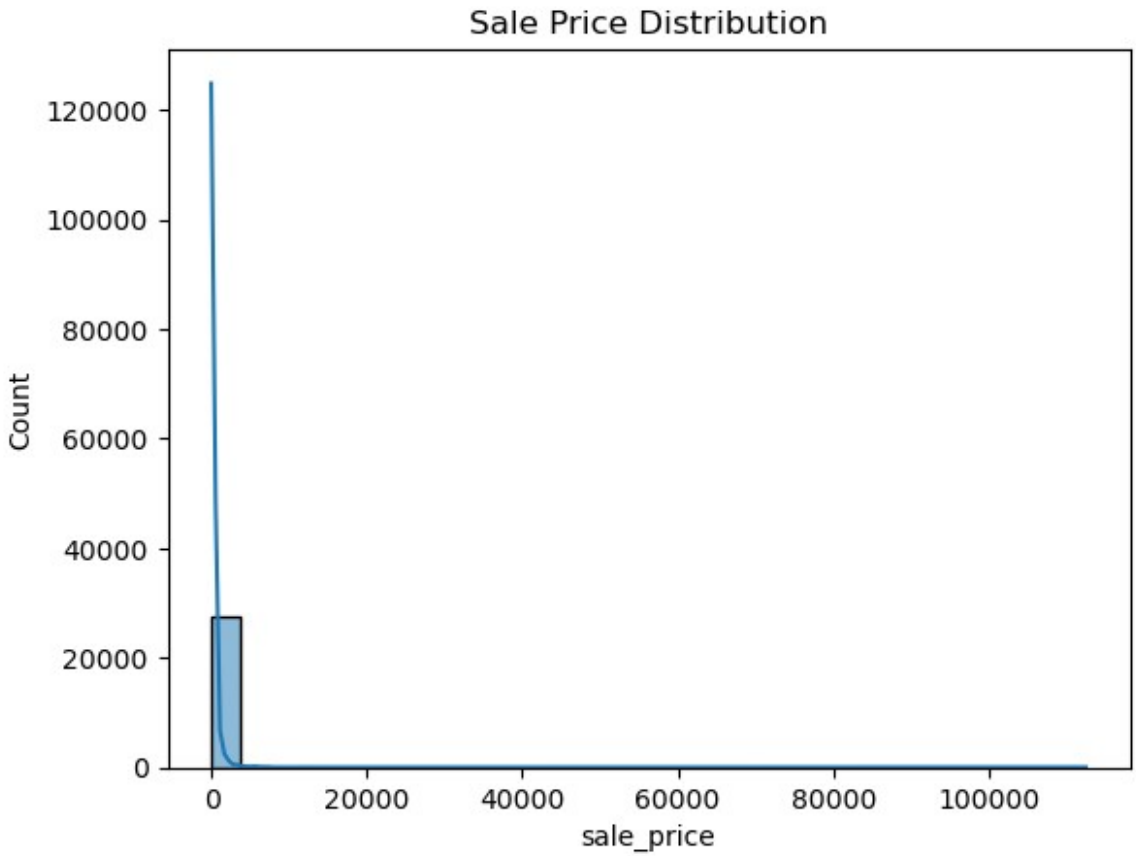
The bar plot shows the top 10 categories with the maximum number of products. Beauty & Hygiene has the highest product count, followed by Gourmet & World Food and Kitchen, Garden & Pets. This indicates that the dataset is heavily dominated by Beauty & Hygiene products, suggesting higher product variety and listing frequency in this category. Categories like Fruits & Vegetables and Baby Care have the lowest product count among the top 10

```
top10 = df1['category'].value_counts().head(10)
sns.barplot(x=top10.values, y=top10.index)
plt.title("Top 10 Categories who have maximum number of product")
plt.xlabel("Count")
plt.ylabel("Category")
plt.show()
```



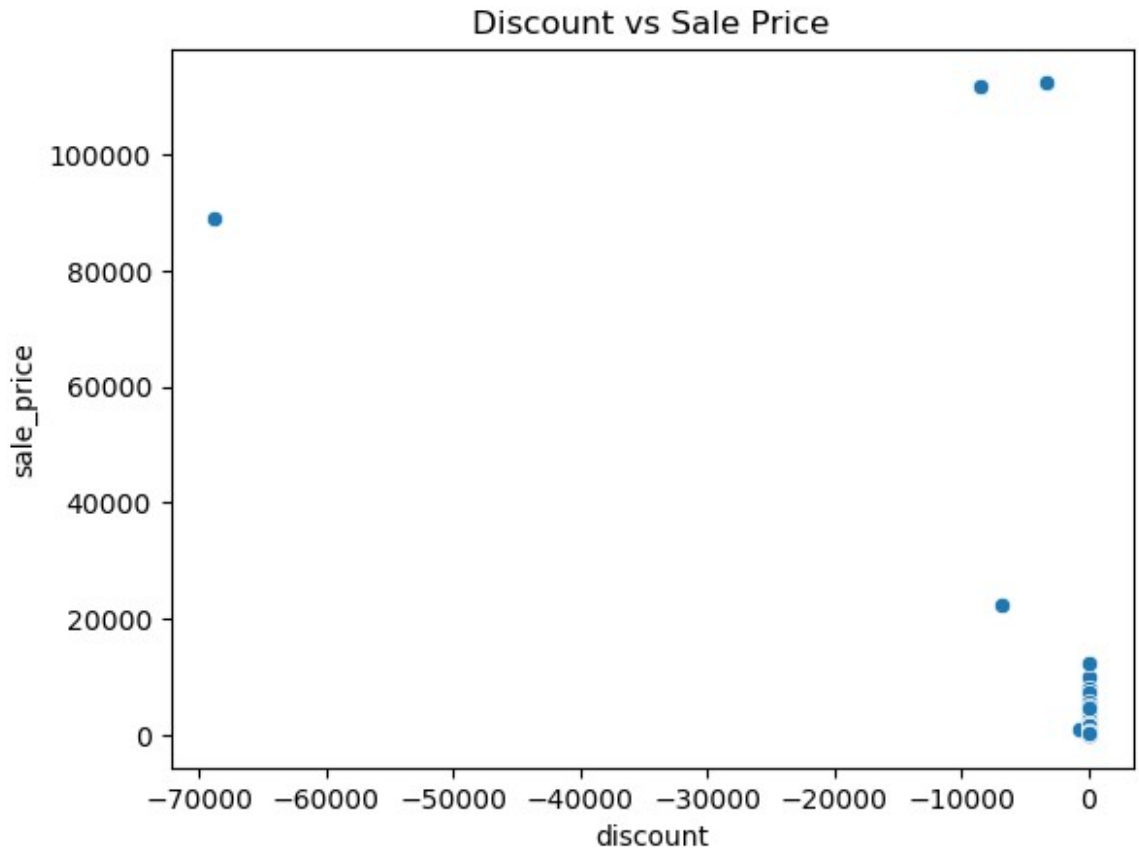
The histogram shows the distribution of sale_price. Most products have low sale prices, as seen from the high frequency near the left side. The distribution is highly right-skewed with a long tail, indicating the presence of a few very expensive products (outliers). Therefore, the sale_price is not normally distributed and median would be a better measure of central tendency than mean

```
sns.histplot(df1['sale_price'], bins=30, kde=True)
plt.title("Sale Price Distribution")
plt.show()
```



The scatter plot represents the relationship between discount and sale price. Most products are concentrated near zero discount, indicating that the majority of products have low discounts. A few products have very high sale prices (outliers), which increases the spread of the plot. Also, discount values appear negative and extremely large, which suggests possible data issues or incorrect discount calculation. Due to outliers and abnormal discount values, no clear trend can be concluded and the discount column needs cleaning for better analysis

```
sns.scatterplot(data=df1, x="discount", y="sale_price")
plt.title("Discount vs Sale Price")
plt.show()
```



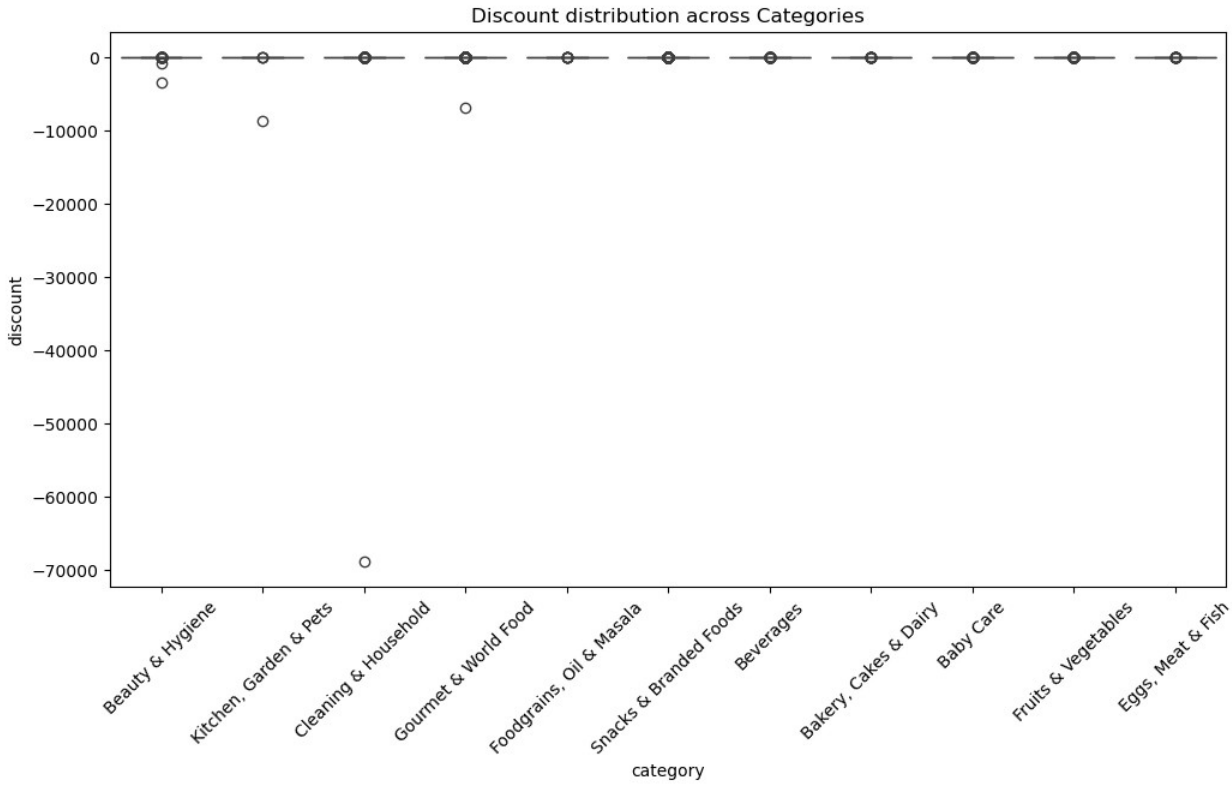

```
df1[df1['discount'] == df1['discount'].min()]
```

```
     index                         product              category  \
436    437  Balloon - Polka Dot, 12 Inch  Cleaning & Household

              sub_category     brand  sale_price  market_price  \
436  Party & Festive Needs  B Vishal     88899.0         129.0

                          type  rating  \
436  Caps, Balloons & Candles     3.9

                                              description      discount
436  Whether it is a party in the office, a Christm... -68813.953488
```

This plot shows the discount distribution across different product categories. Most discount values for all categories are clustered around 0, indicating that a majority of products have low discounts. However, some extreme negative discount values are present, which act as outliers and may indicate incorrect discount calculation or data issues. Due to these outliers, discount comparison between categories is not clearly visible and the discount column needs cleaning for accurate analysis.

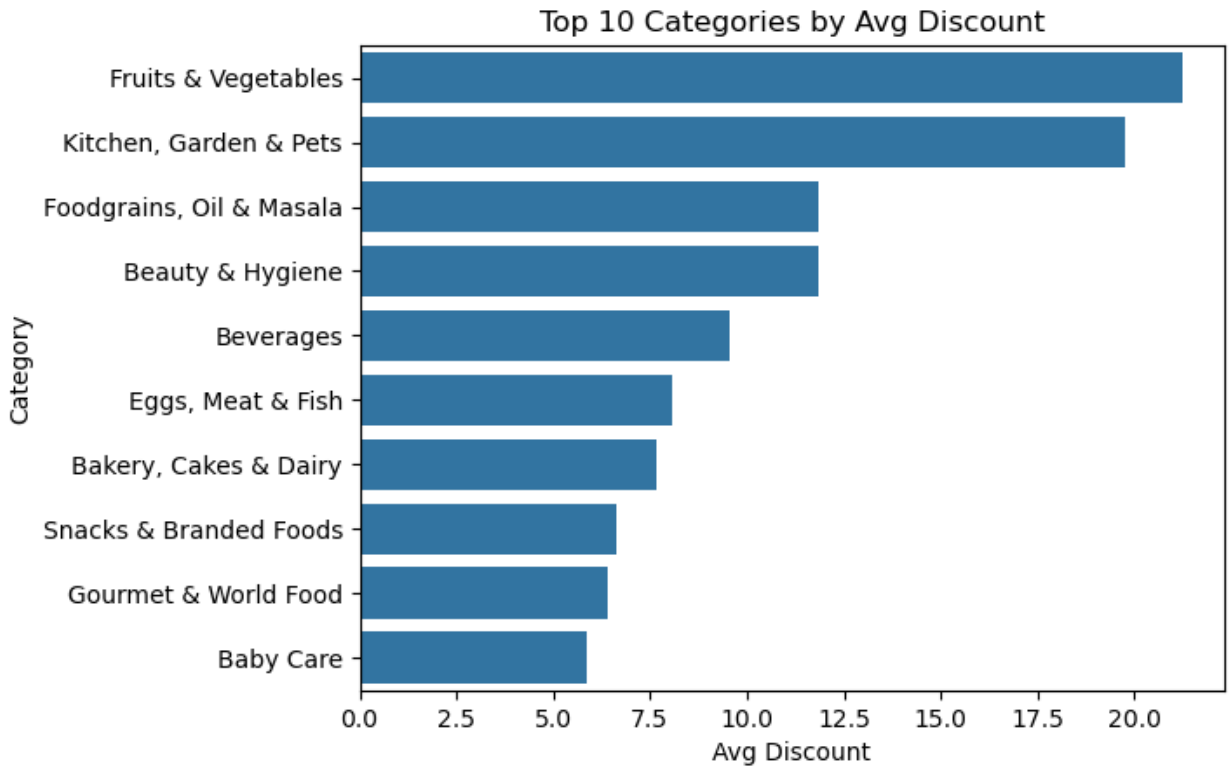
```
plt.figure(figsize=(12,6))
sns.boxplot(data=df1, x="category", y="discount")
```

```
plt.xticks(rotation=45)
plt.title("Discount distribution across Categories")
plt.show()
```



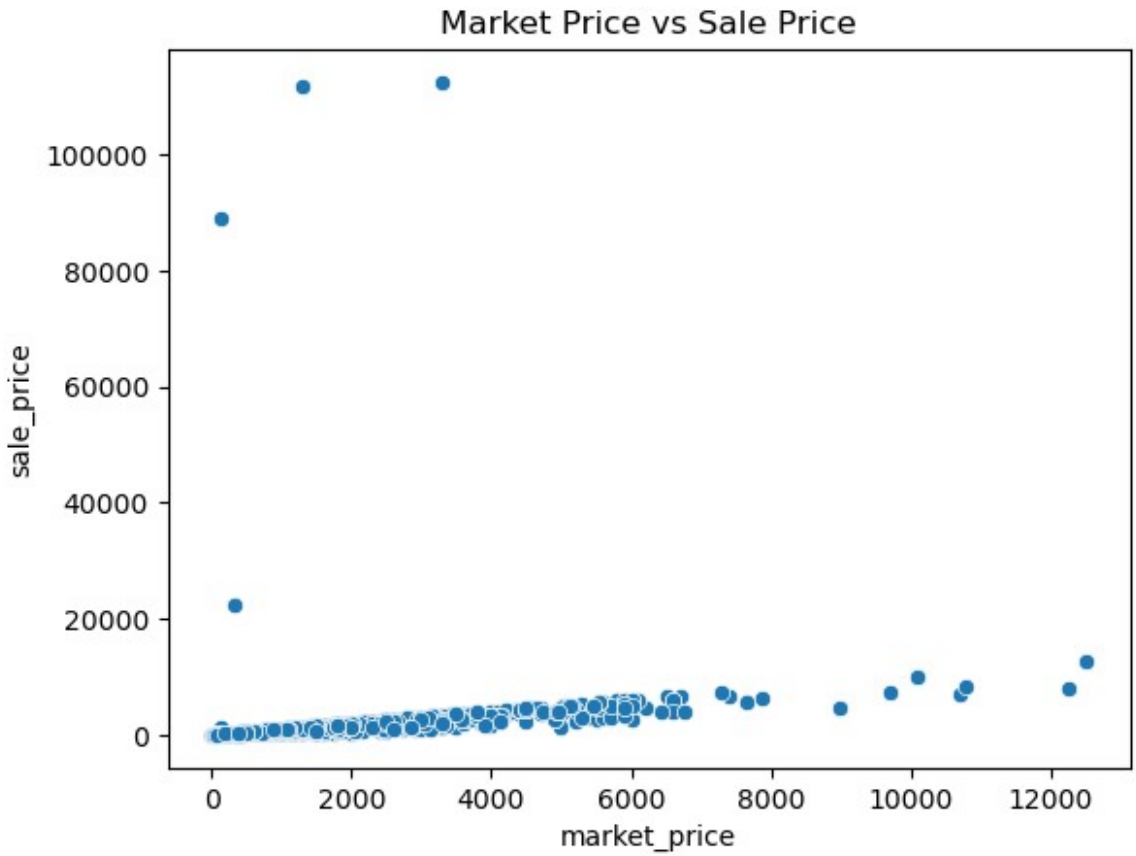
The bar plot shows the top 10 product categories based on their average discount. Fruits & Vegetables have the highest average discount, followed by Kitchen, Garden & Pets. This indicates that discounts are not uniformly distributed and some categories provide higher discounts than others. Categories with higher average discounts are likely used to attract customers or clear fast-moving/perishable inventory. Therefore, discount strategy varies across categories

```
avg_discount = df1.groupby("category")
["discount"].mean().sort_values(ascending=False).head(10)
sns.barplot(x=avg_discount.values, y=avg_discount.index)
plt.title("Top 10 Categories by Avg Discount")
plt.xlabel("Avg Discount")
plt.ylabel("Category")
plt.show()
```



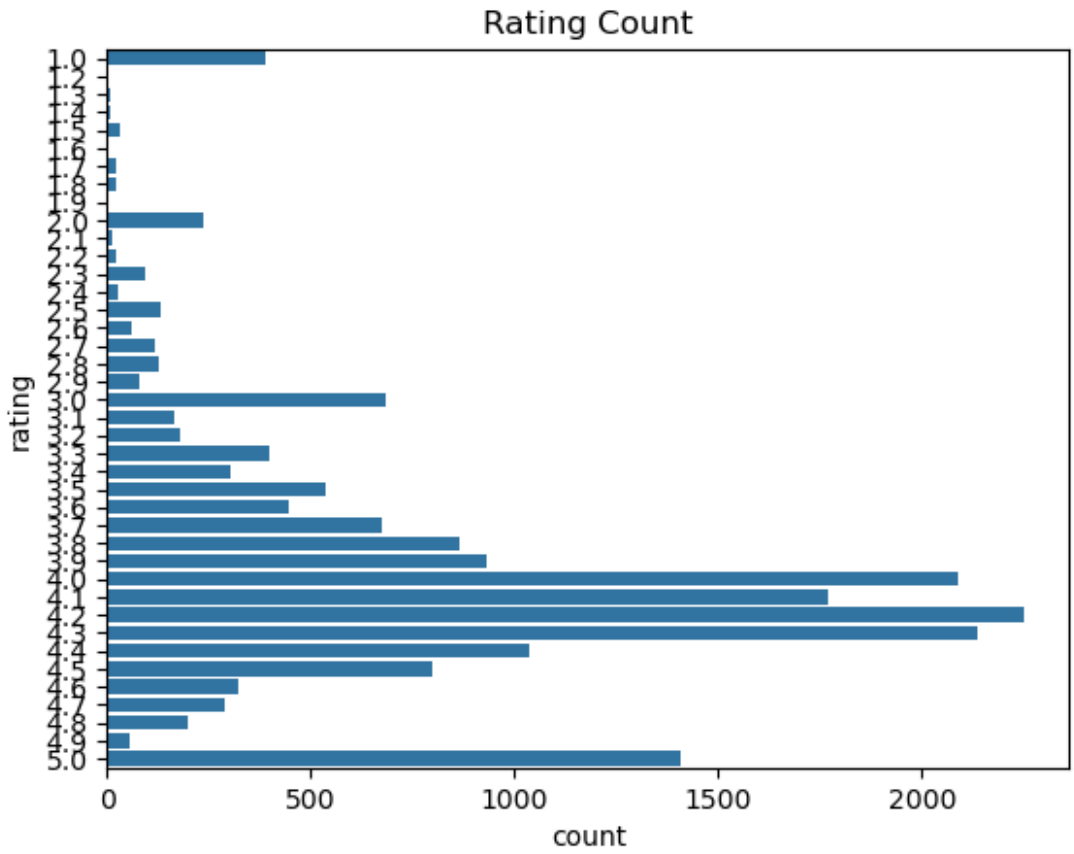
In the given plot we are seeing the realtionship betweeen market price and sale price most of the values were sell at a decent price which shows the relation is kind a linear but there exist some values which is extream high or too costly according to their market price we can say that these are the outliers

```
sns.scatterplot(data=df, x="market_price", y="sale_price")
plt.title("Market Price vs Sale Price")
plt.show()
```



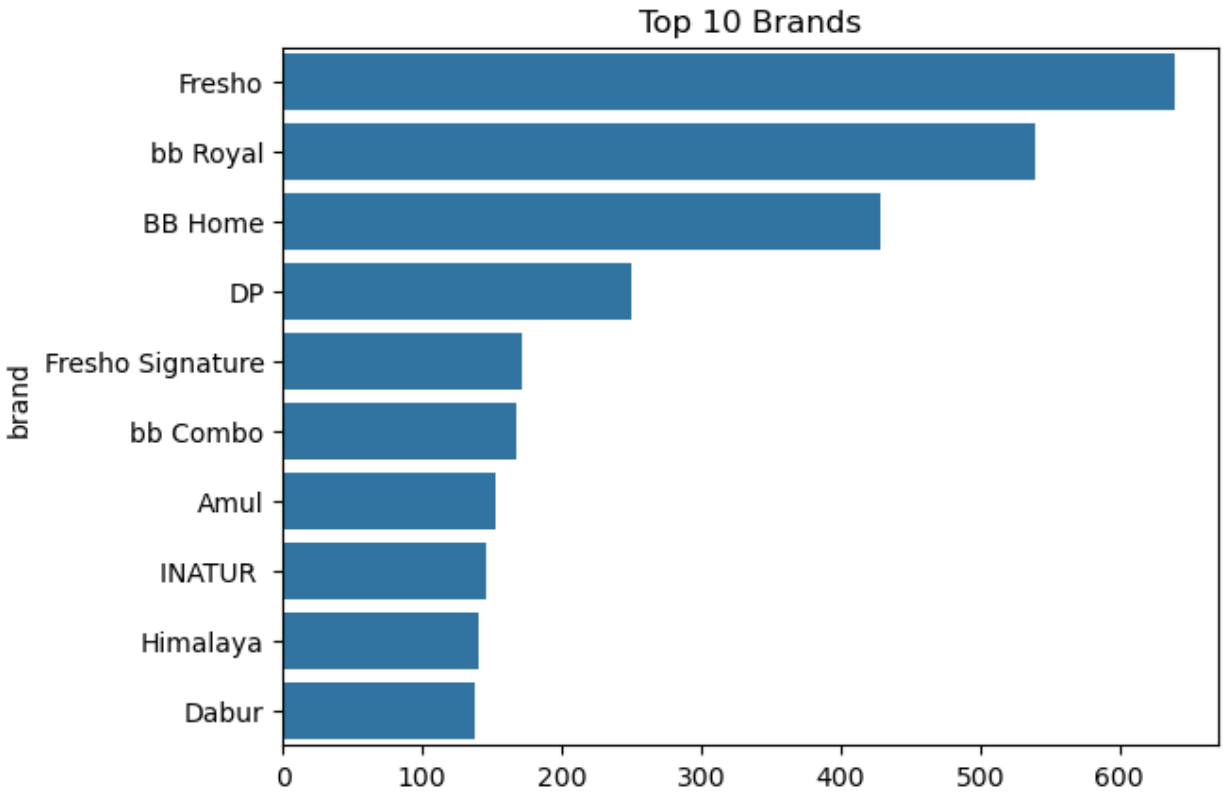
This countplot shows the frequency of product ratings in the dataset. Most products have ratings between 3.8 and 4.5, with the highest frequency around 4.0–4.2. This indicates that the majority of products receive good ratings, reflecting overall customer satisfaction. The distribution is skewed towards higher ratings, meaning low ratings are comparatively rare

```
sns.countplot(y="rating", data=df1)
plt.title("Rating Count")
plt.show()
```



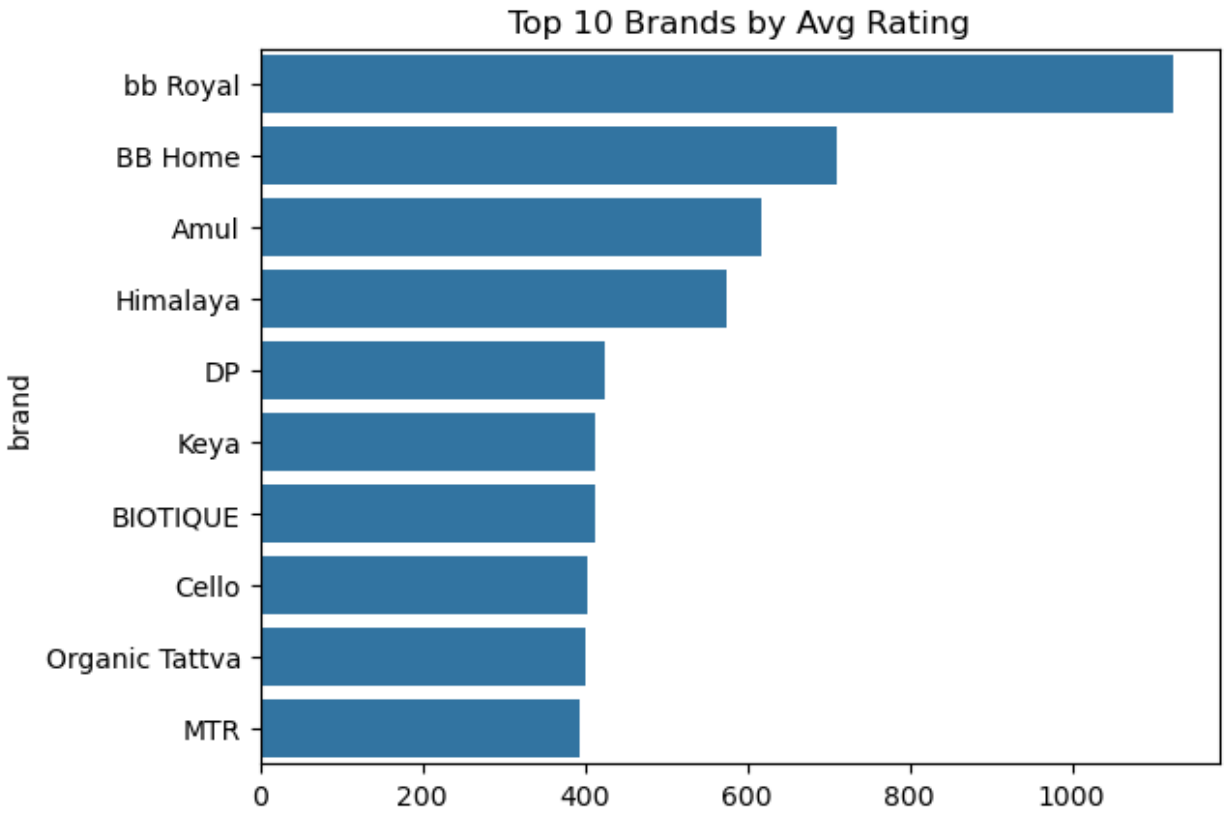
The given horizonatal bar plot shows the top brand who has higher number of product in the dataset and after analyzing the plot and the data we can say that 'Fresho','BB Royal'.'BB Home' are these top brand who have the most products in the dataset

```
top_brands = df1['brand'].value_counts().head(10)
sns.barplot(x=top_brands.values, y=top_brands.index)
plt.title("Top 10 Brands")
plt.show()
```



This plot represents the top 10 brands ranked by their average rating. bb Royal has the highest average rating, followed by BB Home and Amul. This suggests that customers are more satisfied with products from these brands. Overall, the plot helps identify brands with better customer feedback and product quality based on ratings

```
top10_brand_rating = df.groupby("brand")
["rating"].sum().sort_values(ascending=False).head(10)
sns.barplot(x=top10_brand_rating.values, y=top10_brand_rating.index)
plt.title("Top 10 Brands by Avg Rating")
plt.show()
```


Top 10 Brands by Avg Rating
brand
bb Royal
BB Home
Amul
Himalaya
DP
Keya
BIOTIQUE
Cello
Organic Tattva
MTR
0
200
400
600
800
1000

From the EDA analysis, we observed that the dataset is dominated by a few categories and brands. Beauty & Hygiene has the highest product count, and Fresho, bb Royal, and BB Home are the brands with maximum product listings. Sale price distribution is highly right-skewed, meaning most products are low-priced while a few expensive products exist as outliers. Market price and sale price show a positive linear relationship, but some extreme products act as anomalies. Most products are rated between 3.8 and 4.5, indicating overall customer satisfaction. However, discount values contain extreme and negative outliers, so discount data needs cleaning for accurate discount-related insights

Key Insights

Beauty & Hygiene has the highest number of products, meaning this category dominates the dataset and has maximum product variety.

Brands like Fresho, bb Royal, and BB Home have the highest product listings, showing they are the most frequently available brands in the dataset.

Market price and sale price show a positive linear relationship, which means as market price increases, sale price also increases — pricing is mostly consistent.

Sale price distribution is right-skewed, meaning:

~ most products are low-priced

~ Only a few products are very expensive

~ These costly products act as outliers.

Categories like Fruits & Vegetables and Kitchen, Garden & Pets show higher average discounts, which may be due to:

~ perishable items

~ competition

~ quick sales clearance

Negative discount values were observed due to inconsistencies in data where sale_price is greater than market_price. Such values were treated as anomalies and were either removed or clipped to 0 for better discount analysis