

EDA

- EDA stands for Exploratory Data Analysis. It's an approach to analyzing datasets to summarize their main characteristics, often using statistical graphs and other data visualisation techniques.
- The primary goal of the EDA is to understand what the data can tell us beyond the formal modelling or hypothesis testing tasks. It involves looking at the data from different angles summarising it's main features and making sense of the data.

```
# import=ng majoe liberaries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import plotly.express as px

import warnings
warnings.filterwarnings('ignore')

df=pd.read_csv('BigBasket_Clean.csv')

df
```

	product \
0	Garlic Oil - Vegetarian Capsule 500 mg
1	Water Bottle - Orange
2	Brass Angle Deep - Plain, No.2
3	Cereal Flip Lid Container/Storage Jar - Assort...
4	Creme Soft Soap - For Hands & Body
...	...
27550	Wottagirl! Perfume Spray - Heaven, Classic
27551	Rosemary
27552	Peri-Peri Sweet Potato Chips
27553	Green Tea - Pure Original
27554	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 9 columns]

df.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 27555 entries, 0 to 27554
Data columns (total 9 columns):
#   Column          Non-Null Count  Dtype
---  -
```

```

0    product      27555 non-null object
1    category     27555 non-null object
2    sub_category 27555 non-null object
3    brand        27555 non-null object
4    sale_price   27555 non-null float64
5    market_price 27555 non-null float64
6    type         27555 non-null object
7    rating       27555 non-null float64
8    description  27555 non-null object
dtypes: float64(3), object(6)
memory usage: 1.9+ MB

```

Labeling columns as Numerical,Categorical or Mixed

- **Numeric:** sale_price, market_price, rating
- **Categorical:** Category, sub_category, brand, type
- **Mixed:** Description, Product

```
df[df['product'] == 'Soft Drink']
```

	product	category	sub_category	brand
sale_price \				
2208	Soft Drink	Beverages	Energy & Soft Drinks	Mountain Dew
145.80				
3287	Soft Drink	Beverages	Energy & Soft Drinks	Coca-Cola
210.00				
4583	Soft Drink	Beverages	Energy & Soft Drinks	Pepsi
90.00				
5446	Soft Drink	Beverages	Energy & Soft Drinks	Mirinda
64.03				
8335	Soft Drink	Beverages	Energy & Soft Drinks	Pepsi
35.20				
9429	Soft Drink	Beverages	Energy & Soft Drinks	Coca-Cola
160.00				
10005	Soft Drink	Beverages	Energy & Soft Drinks	Pepsi
68.29				
14916	Soft Drink	Beverages	Energy & Soft Drinks	Sprite
438.00				
18812	Soft Drink	Beverages	Energy & Soft Drinks	7 Up
140.00				
22997	Soft Drink	Beverages	Energy & Soft Drinks	Thums Up
59.40				
24608	Soft Drink	Beverages	Energy & Soft Drinks	Mirinda
128.06				
26884	Soft Drink	Beverages	Energy & Soft Drinks	Thums Up
38.00				
	market_price	type	rating	\
2208	180.0	Cold Drinks	4.3	
3287	240.0	Cold Drinks	4.2	

4583	90.0	Cold Drinks	4.2
5446	90.0	Cold Drinks	4.2
8335	40.0	Cold Drinks	4.3
9429	160.0	Cold Drinks	4.6
10005	80.0	Cold Drinks	4.2
14916	510.0	Cold Drinks	4.2
18812	180.0	Cold Drinks	4.3
22997	80.0	Cold Drinks	4.2
24608	180.0	Cold Drinks	4.2
26884	40.0	Cold Drinks	4.2

	description
2208	Mountain Dew exhilarates like no other because...
3287	It's delicious, uplifting and refreshing since...
4583	Pepsi is the pop that shakes things up! Pepsi ...
5446	Mirindas great orangey taste and sparkling bub...
8335	Pepsi is the pop that shakes things up. Pepsi ...
9429	It's delicious, uplifting and refreshing since...
10005	Pepsi is the pop that shakes things up! Pepsi ...
14916	Sprite is a clear lime drink truly meant to qu...
18812	7UP's sparkling and crisp lemon and lime flavo...
22997	Thums Up is one of the most popular soft drink...
24608	Mirindas great orangey taste and sparkling bub...
26884	Thums Up is well-known for its strong, fizzy f...

Univarient Analysis

- It focuses on analyzing each feature in the dataset independently
- **Benefits:**
 - **Distribution Analysis:** The distribution of each feature is examined to identify shape , central tendency and dispersion
 - **Potential Issues:** It helps in identifying potential problems with the data such as outliers, skewnwss and missing values

```
df.columns
Index(['product', 'category', 'sub_category', 'brand', 'sale_price',
      'market_price', 'type', 'rating', 'description'],
      dtype='object')

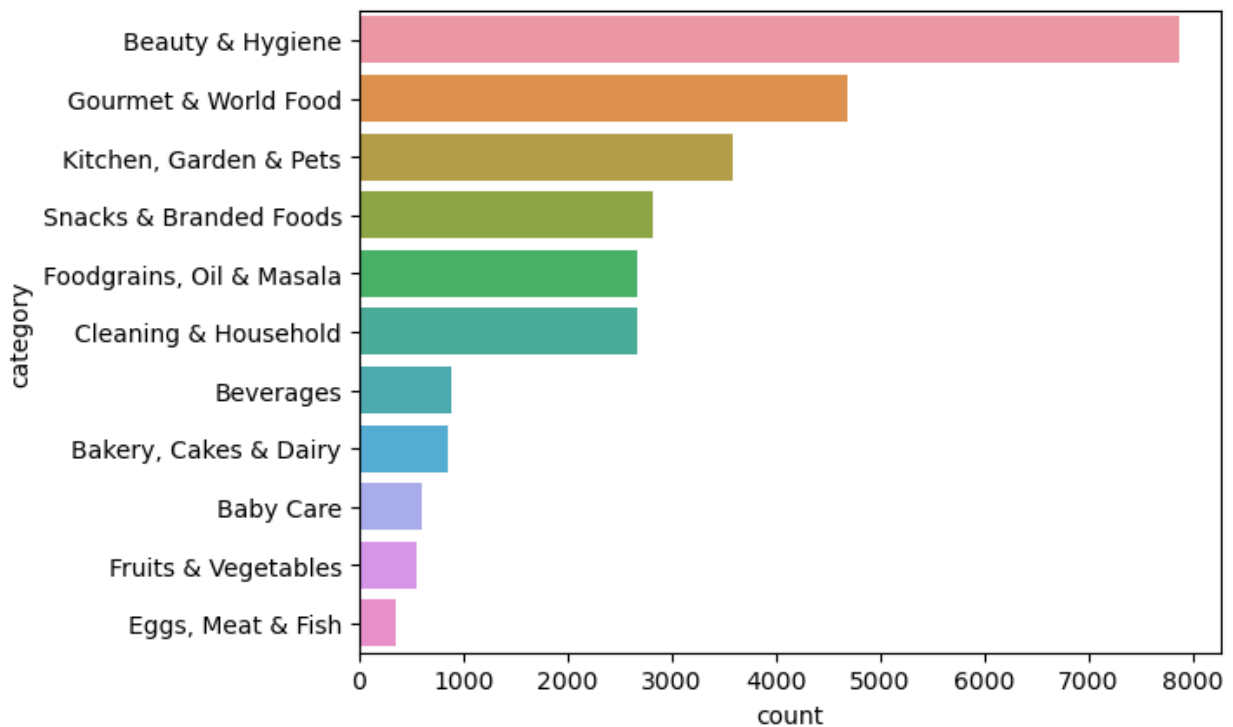
df.category.value_counts()

category
Beauty & Hygiene      7867
Gourmet & World Food   4690
Kitchen, Garden & Pets 3580
Snacks & Branded Foods 2814
Foodgrains, Oil & Masala 2676
Cleaning & Household   2675
```

```
Beverages 885
Bakery, Cakes & Dairy 851
Baby Care 610
Fruits & Vegetables 557
Eggs, Meat & Fish 350
Name: count, dtype: int64
```

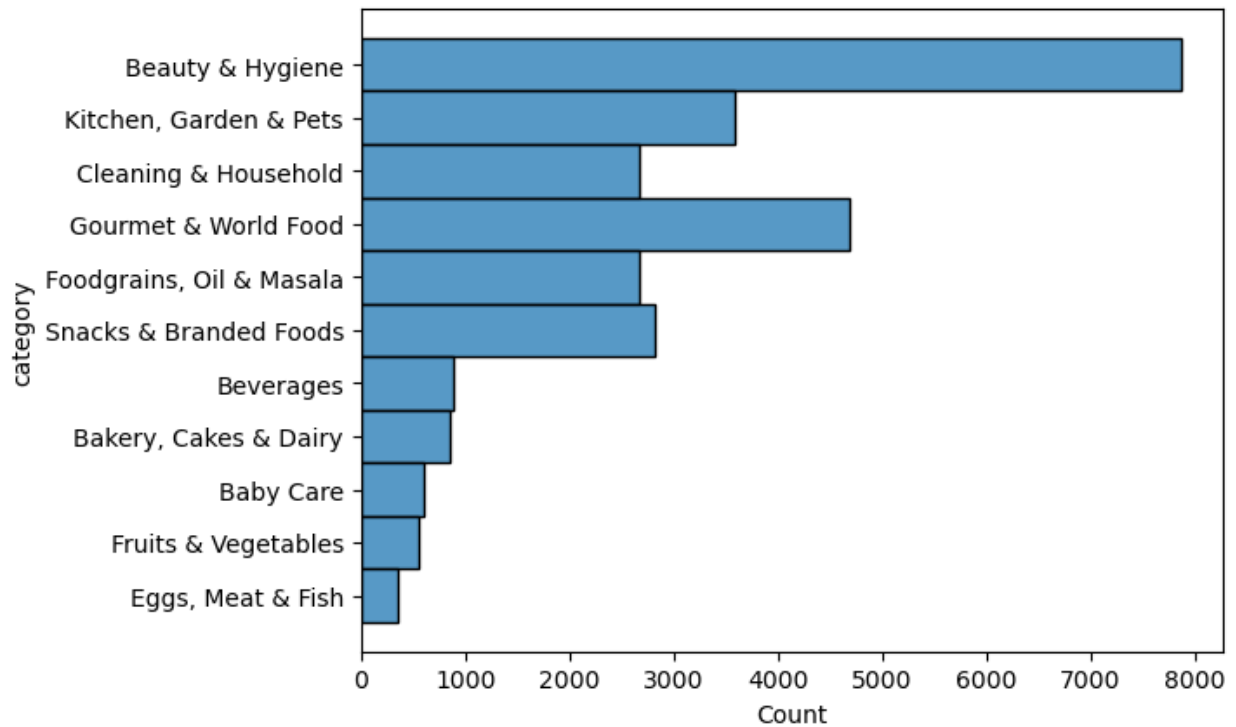
```
sns.barplot(y=df.category.value_counts().index,x=df.category.value_counts())
```

```
<Axes: xlabel='count', ylabel='category'>
```

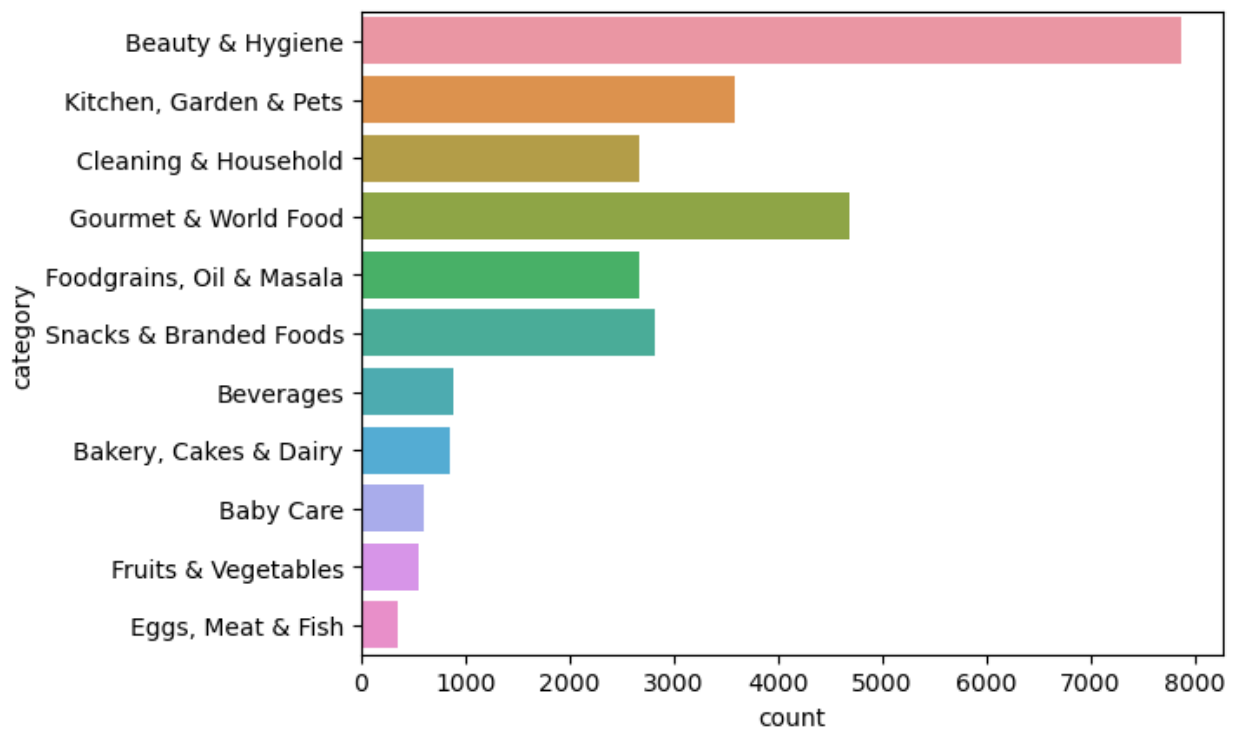


```
sns.histplot(y=df.category)
```

```
<Axes: xlabel='Count', ylabel='category'>
```



```
sns.countplot(y=df.category)  
<Axes: xlabel='count', ylabel='category'>
```



```
#Sub_Category
```

```
temp=df[df.category=='Beauty & Hygiene']
```

```
temp
```

	category \	product	
0	Hygiene	Garlic Oil - Vegetarian Capsule 500 mg	Beauty &
4	Hygiene	Creme Soft Soap - For Hands & Body	Beauty &
6	Hygiene	Multani Mati	Beauty &
7	Hygiene	Hand Sanitizer - 70% Alcohol Base	Beauty &
8	Hygiene	Biotin & Collagen Volumizing Hair Shampoo + Bi...	Beauty &
...		...	
27541	Hygiene	Herbal Rose & Apricot Walnut Facial Scrub	Beauty &
27542	Hygiene	Embellish Skin Lightening Serum	Beauty &
27548	Hygiene	Apple Cider Vinegar Shampoo	Beauty &
27550	Hygiene	Wottagirl! Perfume Spray - Heaven, Classic	Beauty &
27554	Hygiene	United Dreams Go Far Deodorant	Beauty &

	sub_category	brand	sale_price \
0	Hair Care	Sri Sri Ayurveda	220.00
4	Bath & Hand Wash	Nivea	162.00
6	Skin Care	Satinance	58.00
7	Bath & Hand Wash	Bionova	250.00
8	Hair Care	StBotanica	1098.00
...
27541	Skin Care	Khadi Natural	250.00
27542	Skin Care	Organic Harvest	1525.75
27548	Hair Care	Morpheme Remedies	499.00
27550	Fragrances & Deos	Layerr	199.20
27554	Men's Grooming	United Colors Of Benetton	214.53

	market_price	type	rating \
0	220.0	Hair Oil & Serum	4.1
4	162.0	Bathing Bars & Soaps	4.4
6	58.0	Face Care	3.6
7	250.0	Hand Wash & Sanitizers	4.0
8	1098.0	Shampoo & Conditioner	3.5

...
27541	250.0	Face Care	4.4
27542	1795.0	Face Care	4.2
27548	499.0	Shampoo & Conditioner	5.0
27550	249.0	Perfume	3.9
27554	390.0	Men's Deodorants	4.5

	description
0	This Product contains Garlic Oil that is known...
4	Nivea Creme Soft Soap gives your skin the best...
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...
...	...
27541	Rose and apricot walnut face scrub helps in re...
27542	Achieve an everlasting illuminated skin by inc...
27548	Say no to dull, lifeless, dry and damaged hair...
27550	Layerr brings you Wottagirl Classic fragrant b...
27554	The new mens fragrance from the United Dreams ...

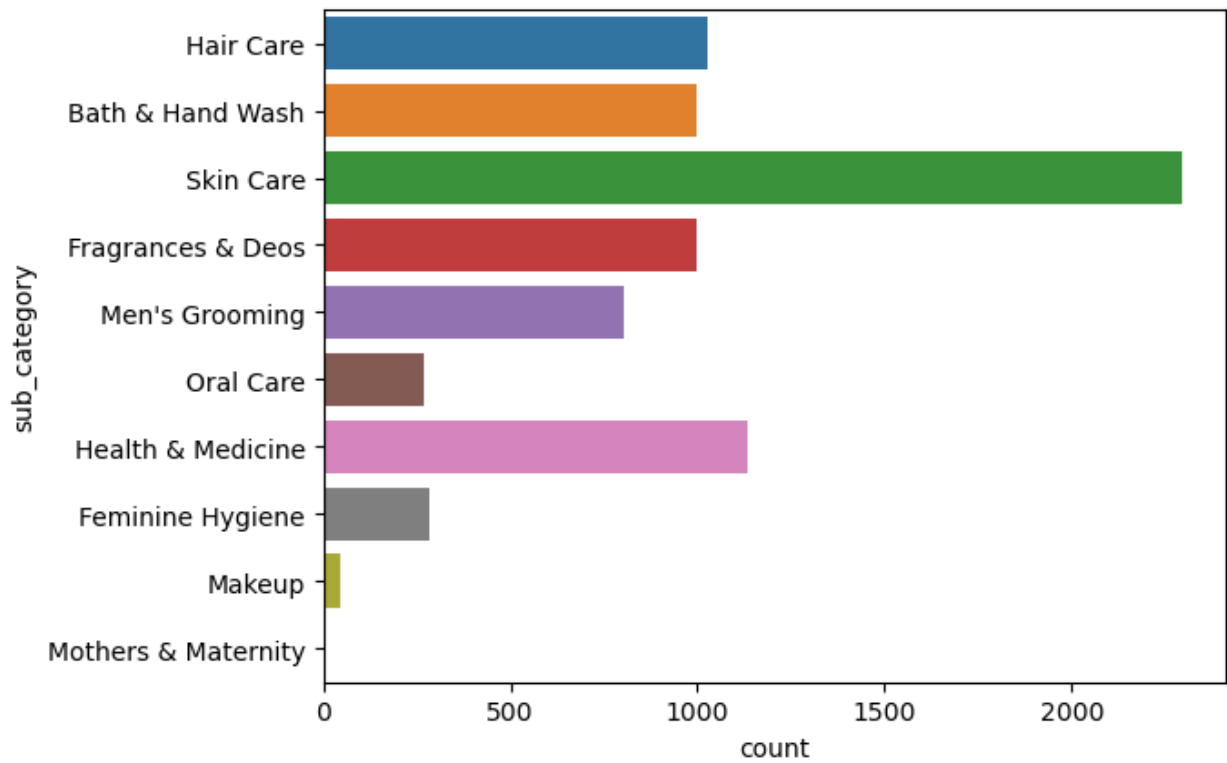
[7867 rows x 9 columns]

```
temp.sub_category.value_counts()
```

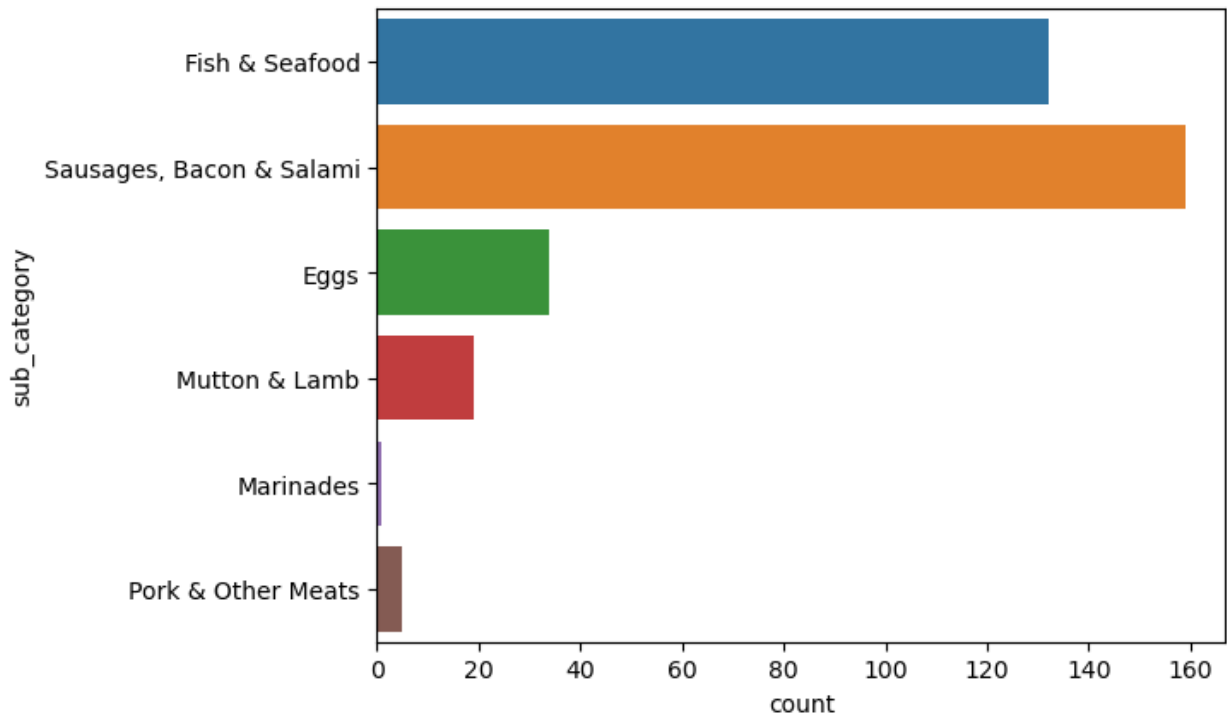
```
sub_category
Skin Care      2294
Health & Medicine  1133
Hair Care      1028
Fragrances & Deos  1000
Bath & Hand Wash   996
Men's Grooming    805
Feminine Hygiene  285
Oral Care        271
Makeup           48
Mothers & Maternity 7
Name: count, dtype: int64
```

```
sns.countplot(y=temp.sub_category)
```

```
<Axes: xlabel='count', ylabel='sub_category'>
```

```
temp_1=df[df.category=='Eggs, Meat & Fish']  
sns.countplot(y=temp_1.sub_category)  
<Axes: xlabel='count', ylabel='sub_category'>
```



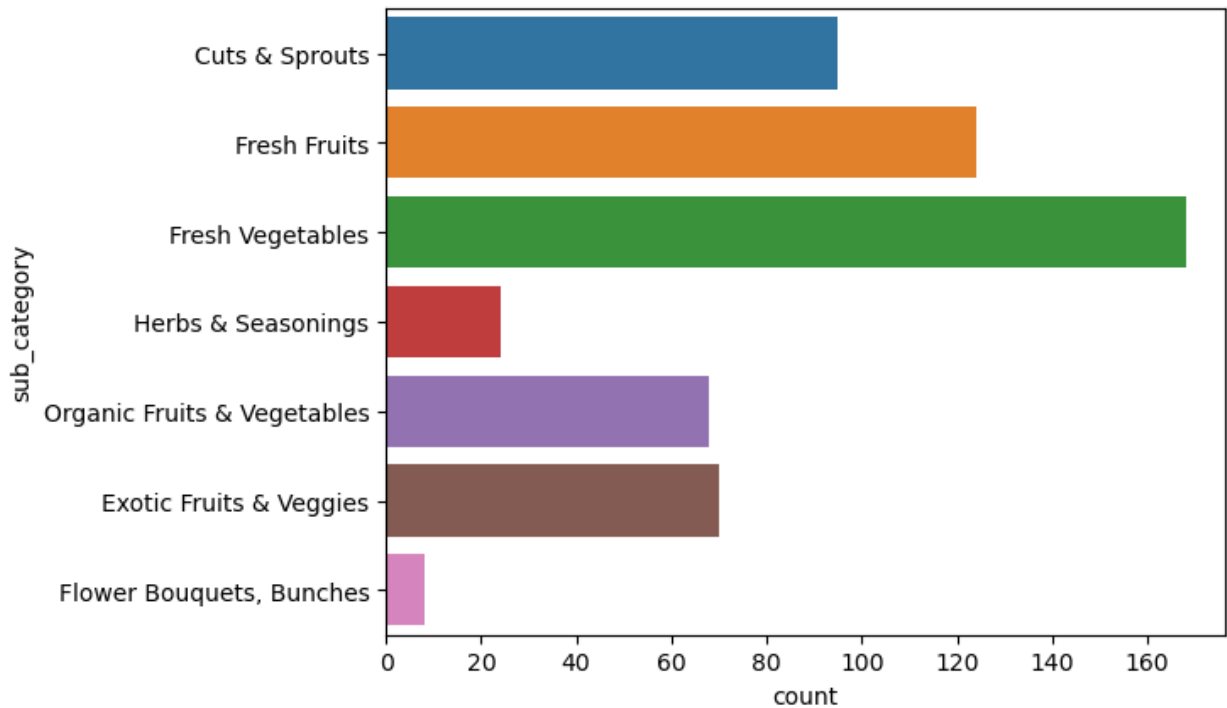
```
temp_2=df[df.category=='Fruits & Vegetables']
```

```
temp_2.sub_category.value_counts()
```

```
sub_category
Fresh Vegetables    168
Fresh Fruits        124
Cuts & Sprouts       95
Exotic Fruits & Veggies  70
Organic Fruits & Vegetables  68
Herbs & Seasonings   24
Flower Bouquets, Bunches  8
Name: count, dtype: int64
```

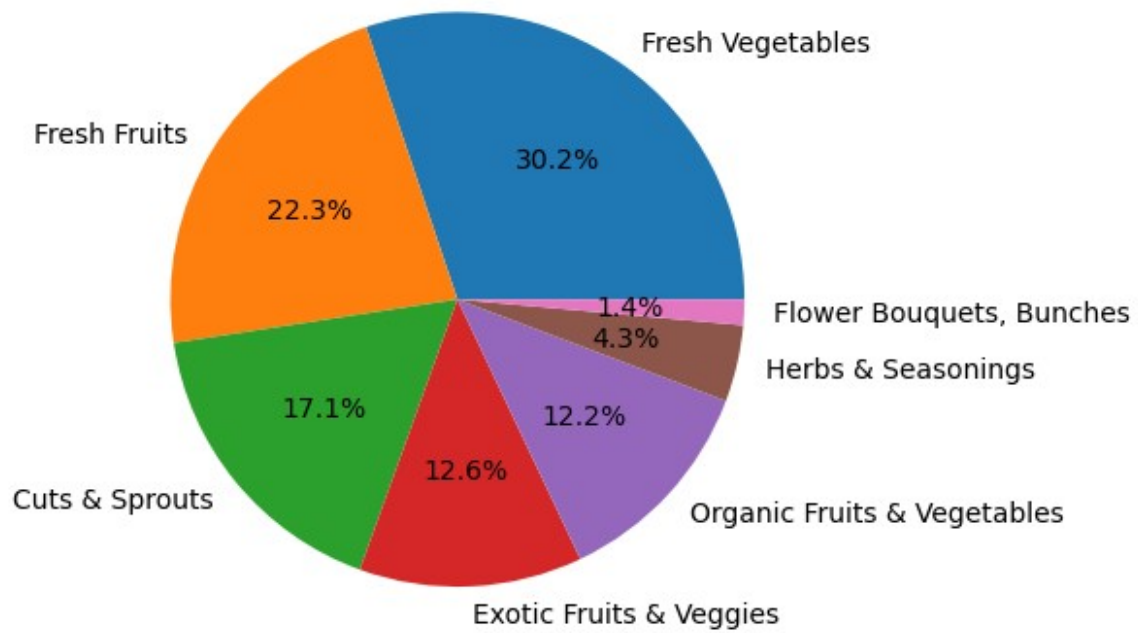
```
sns.countplot(y=temp_2.sub_category)
```

```
<Axes: xlabel='count', ylabel='sub_category'>
```



```
f=temp_2.sub_category.value_counts()
f
sub_category
Fresh Vegetables      168
Fresh Fruits          124
Cuts & Sprouts         95
Exotic Fruits & Veggies 70
Organic Fruits & Vegetables 68
Herbs & Seasonings     24
Flower Bouquets, Bunches 8
Name: count, dtype: int64

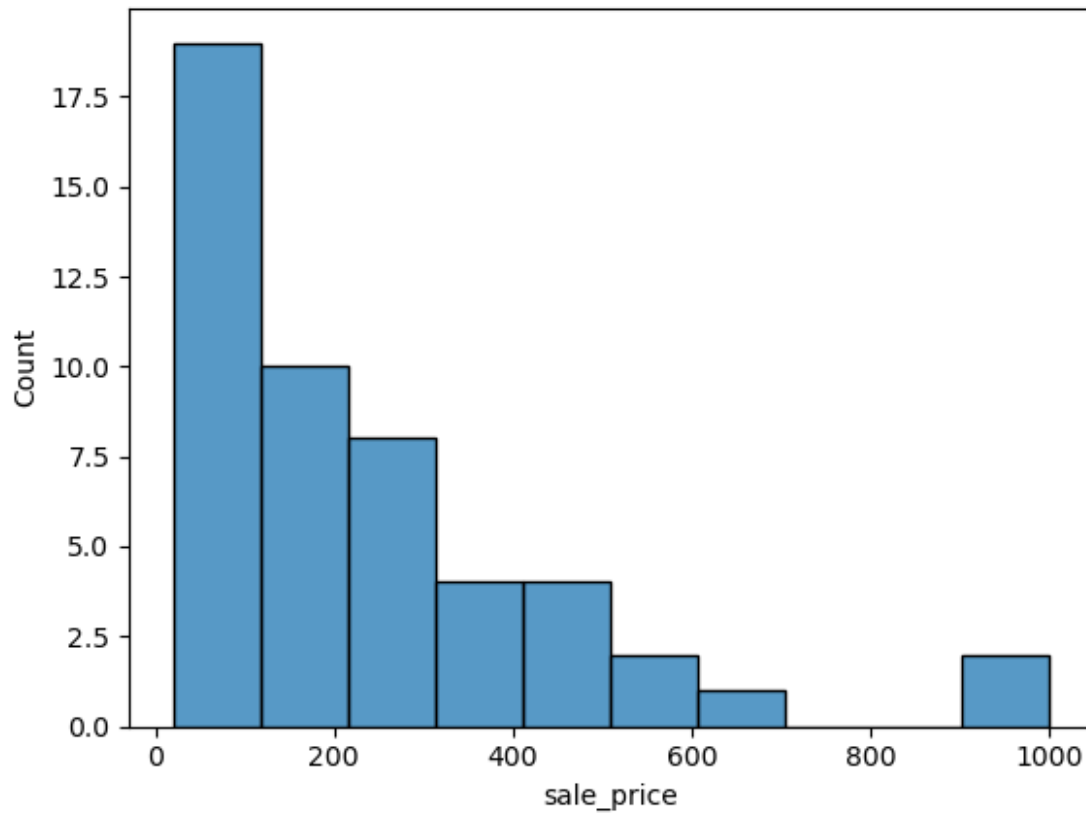
plt.pie(f,labels=f.index,autopct='%1.1f%%')
plt.show()
```



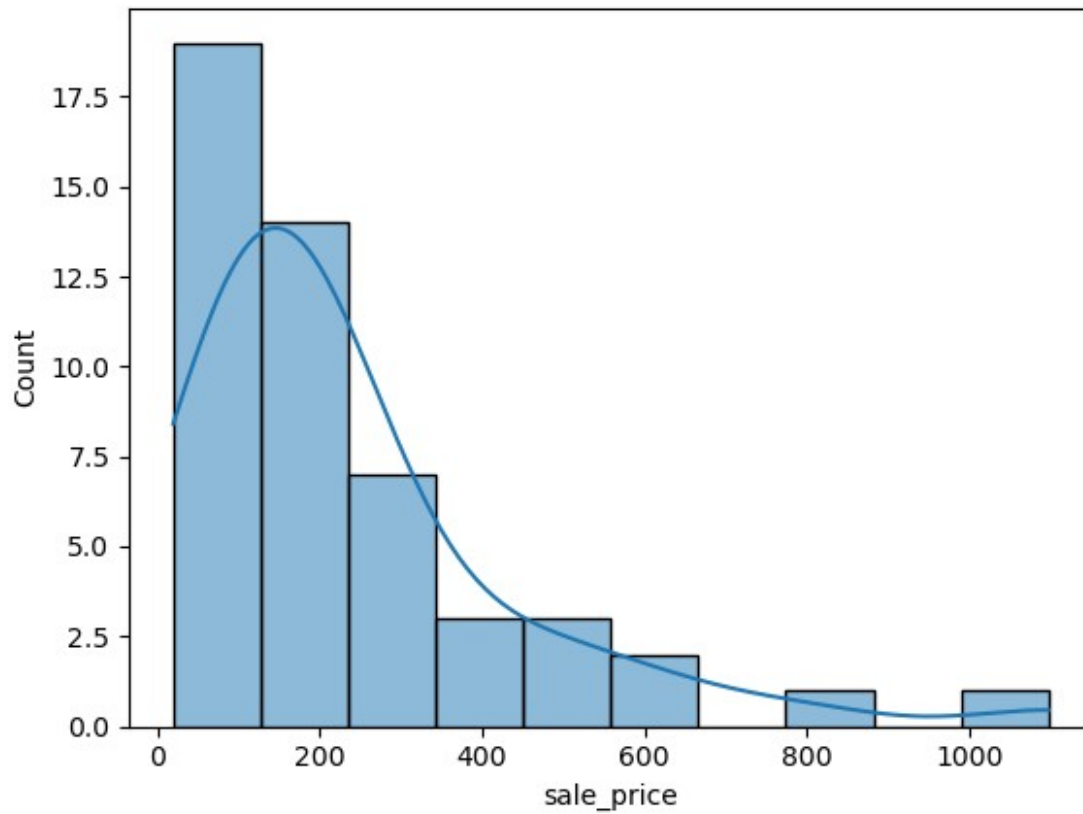
#Histogram

```
sns.histplot(df.sale_price.sample(50),bins=10)
```

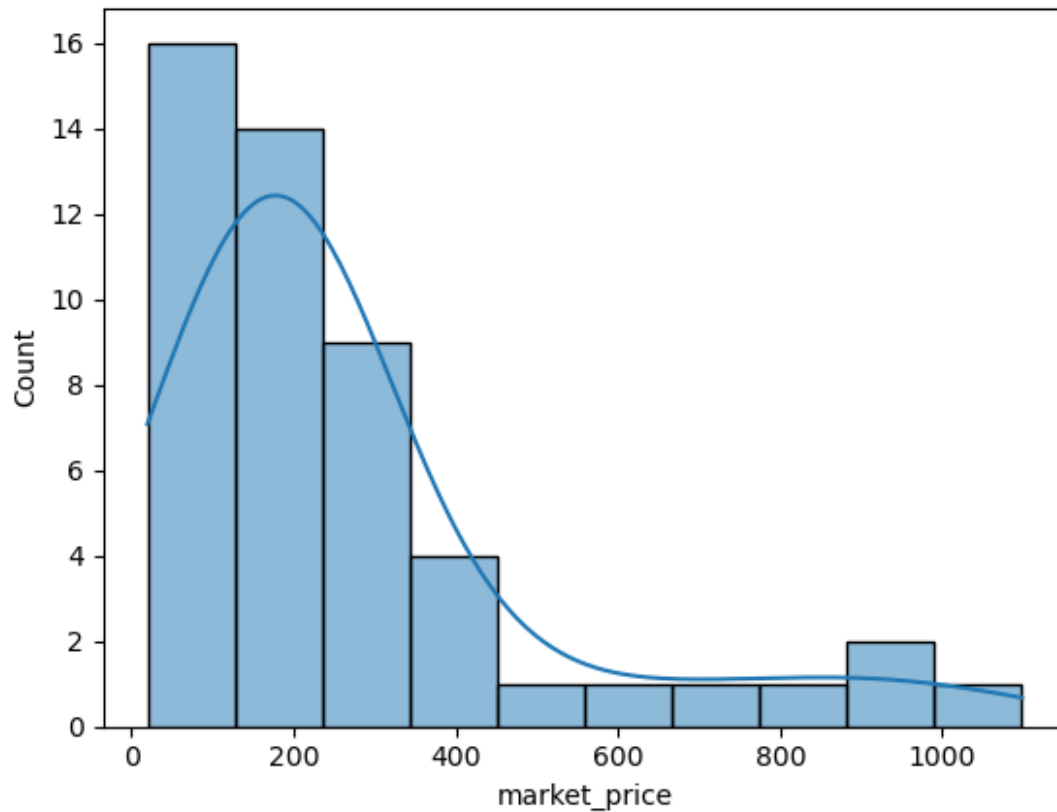
```
<Axes: xlabel='sale_price', ylabel='Count'>
```



```
sns.histplot(df.sale_price.head(50),bins=10,kde=True)  
<Axes: xlabel='sale_price', ylabel='Count'>
```



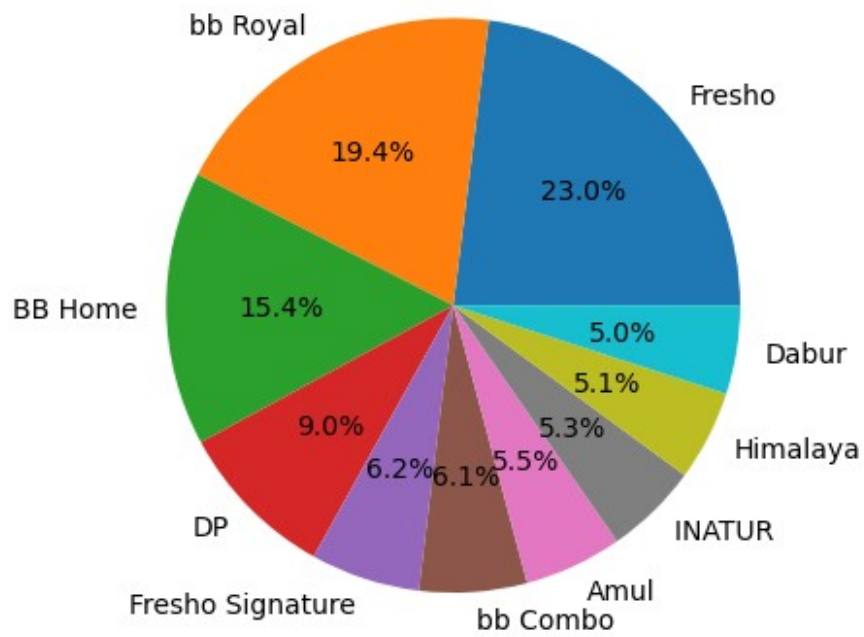
```
sns.histplot(df.market_price.head(50),bins=10,kde=True)  
<Axes: xlabel='market_price', ylabel='Count'>
```



```
a=df.brand.value_counts().head(10)
a

brand
Fresho          638
bb Royal        539
BB Home         428
DP              250
Fresho Signature 171
bb Combo        168
Amul            153
INATUR          146
Himalaya        141
Dabur           138
Name: count, dtype: int64

plt.pie(a,labels=a.index,autopct='%1.1f%%')
plt.show()
```

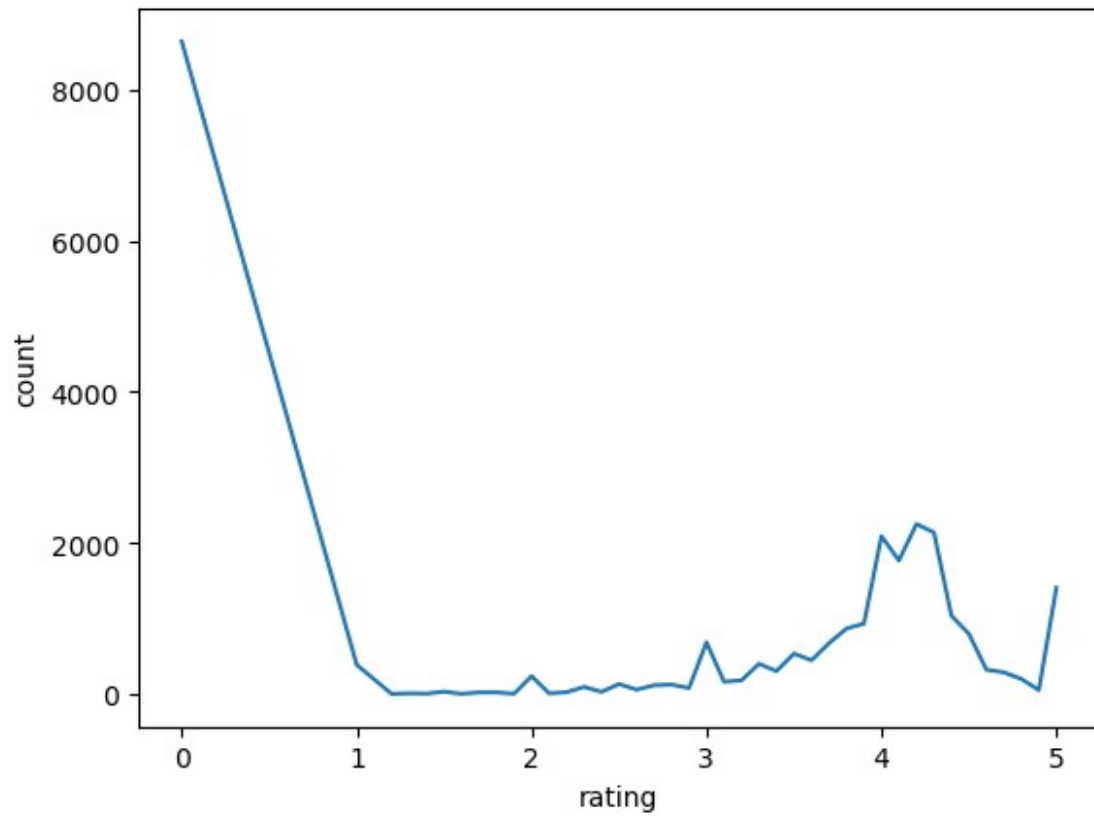


```
#ratings
```

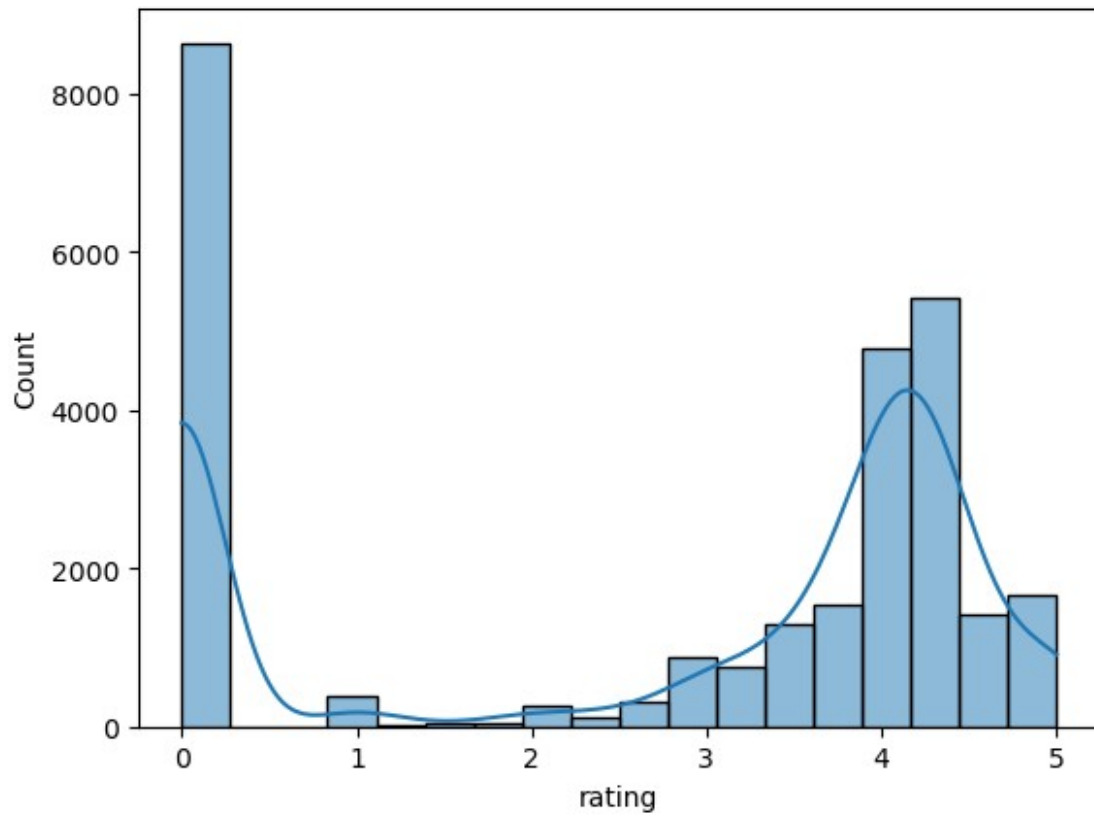
```
c=df.rating.value_counts()
```

```
sns.lineplot(c)
```

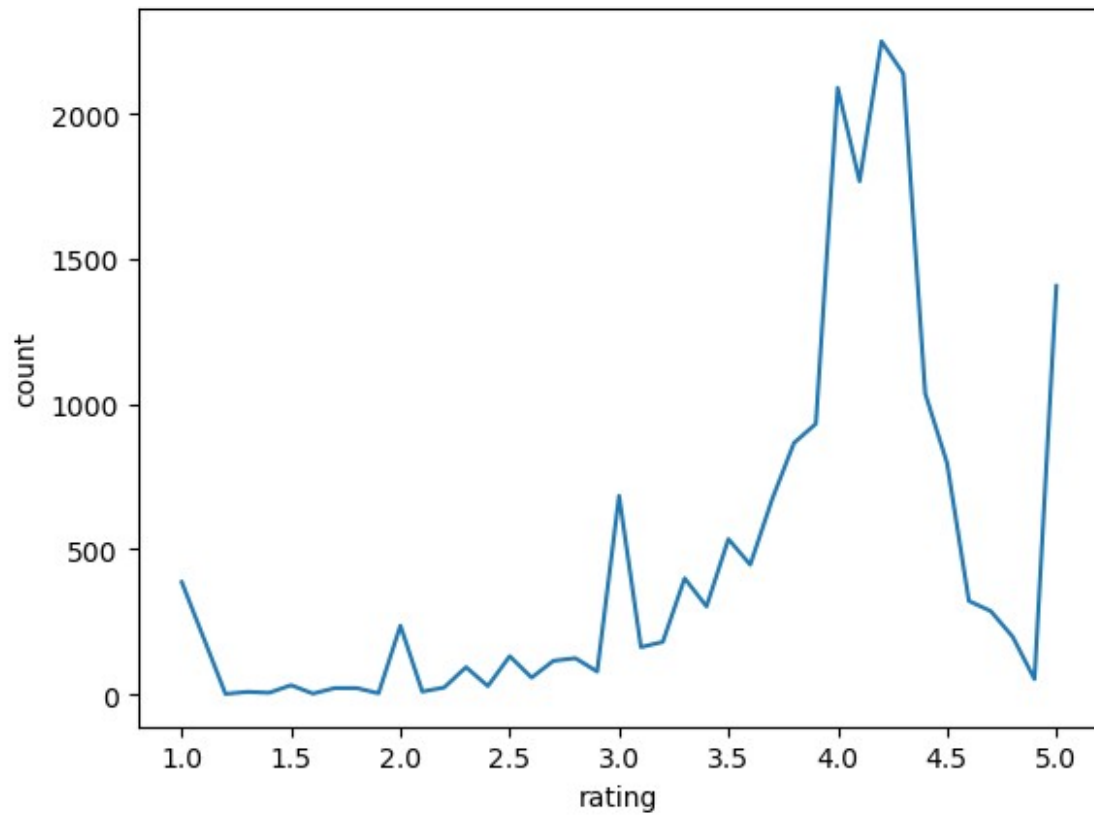
```
<Axes: xlabel='rating', ylabel='count'>
```

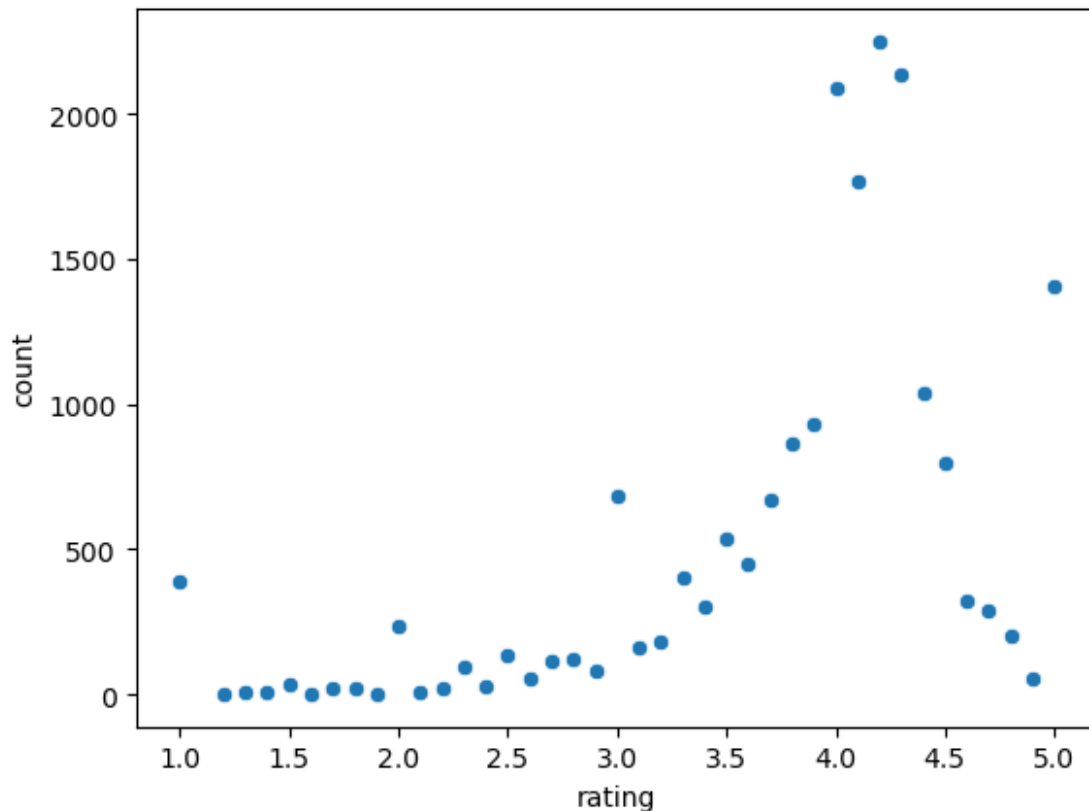
```
sns.histplot(df.rating,kde=10)  
<Axes: xlabel='rating', ylabel='Count'>
```



```
g=df[df.rating>=1].rating.value_counts()  
sns.lineplot(g)  
<Axes: xlabel='rating', ylabel='count'>
```



```
sns.scatterplot(g)  
<Axes: xlabel='rating', ylabel='count'>
```



Bivariate Analysis

- Bivariate Analysis is a statistical method that involves the analysis of two variables simultaneously to determine the relationship between them.
- It focuses on understanding how changes in one variable affect changes in another variable

Graph Preferences

- Scatterplot
- lineplot
- bubbleplots
- heatmap
- barplot
- pie

Steps:

- Select two columns
- Understanding relationship between the columns
 - **Numerical-Numerical:** Check coorelations. You can plot graphs like lineplot, heatmaps

- **Categorical-Numerical:** Create visualisation that compare the distribution of the numerical data across different categories of the categorical data
- **Categorical-Categorical:** You can create a cross-tabulation of contingency tables that show the distribution of values in one categorical column grouped by the values in the other categorical column

```
df.columns
```

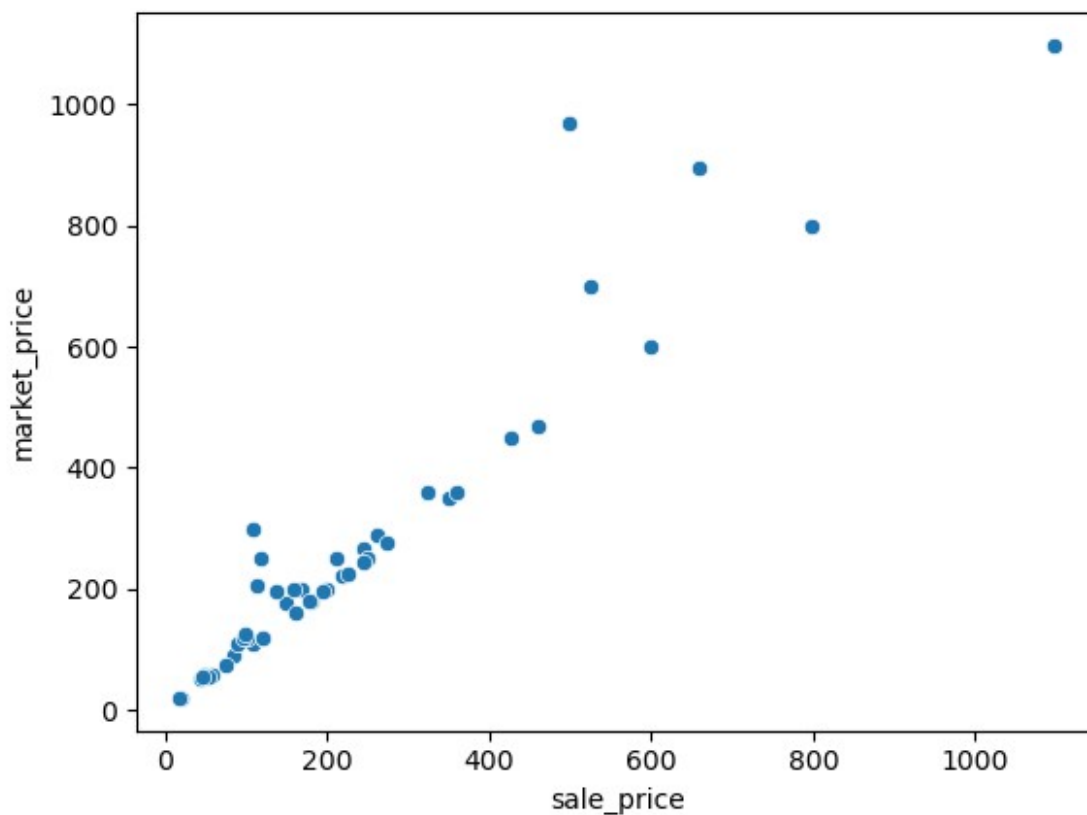
```
Index(['product', 'category', 'sub_category', 'brand', 'sale_price',  
      'market_price', 'type', 'rating', 'description'],  
      dtype='object')
```

Numerical-Numerical

```
tempdf=df.head(50)
```

```
sns.scatterplot(data=tempdf,y=tempdf.market_price,x=tempdf.sale_price)
```

```
<Axes: xlabel='sale_price', ylabel='market_price'>
```



- Discounts are given

```
df['discount']=df.market_price-df.sale_price
```

```
df
```

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

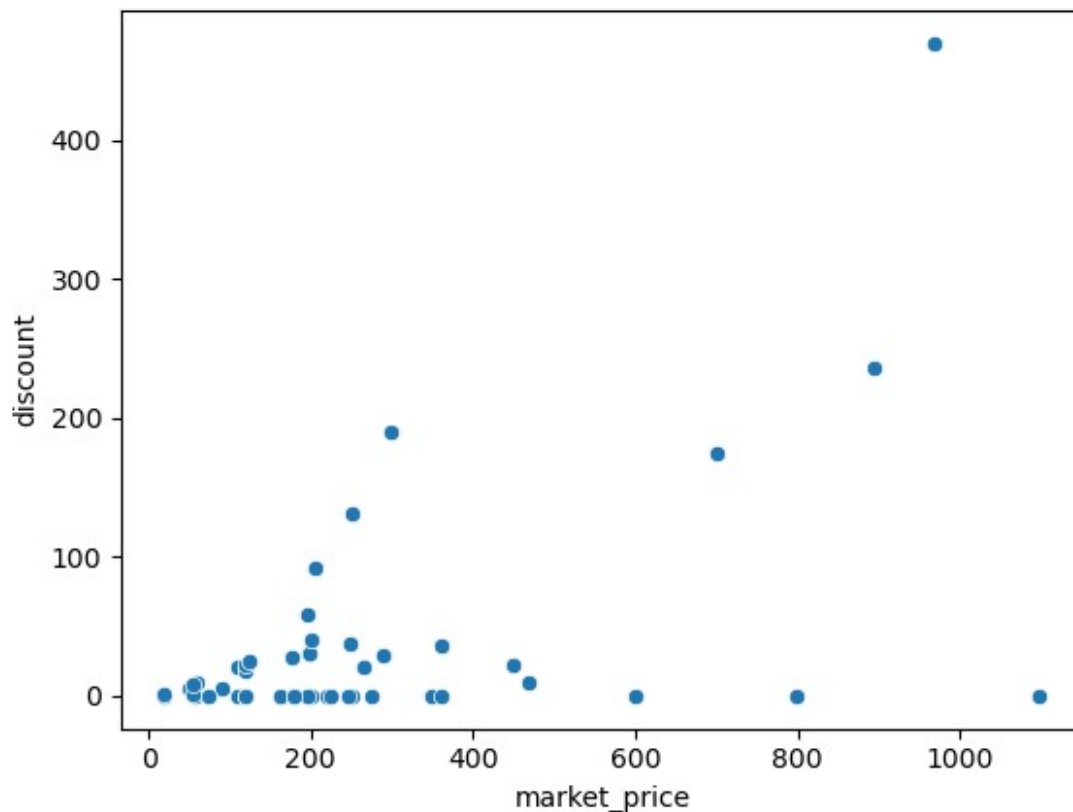
		description	discount
0	This Product contains Garlic Oil that is known...		0.00
1	Each product is microwave safe (without lid), ...		0.00
2	A perfect gift for all occasions, be it your m...		131.00
3	Multipurpose container with an attractive desi...		27.00
4	Nivea Creme Soft Soap gives your skin the best...		0.00
...			...
27550	Layerr brings you Wottagirl Classic fragrant b...		49.80
27551	Puramate rosemary is enough to transform a dis...		7.50
27552	We have taken the richness of Sweet Potatoes (...)		0.00
27553	Tetley Green Tea with its refreshing pure, ori...		99.00
27554	The new mens fragrance from the United Dreams ...		175.47

[27555 rows x 10 columns]

```
tempdf=df.head(50)
```

```
sns.scatterplot(data=tempdf,x=tempdf.market_price,y=tempdf.discount)
```

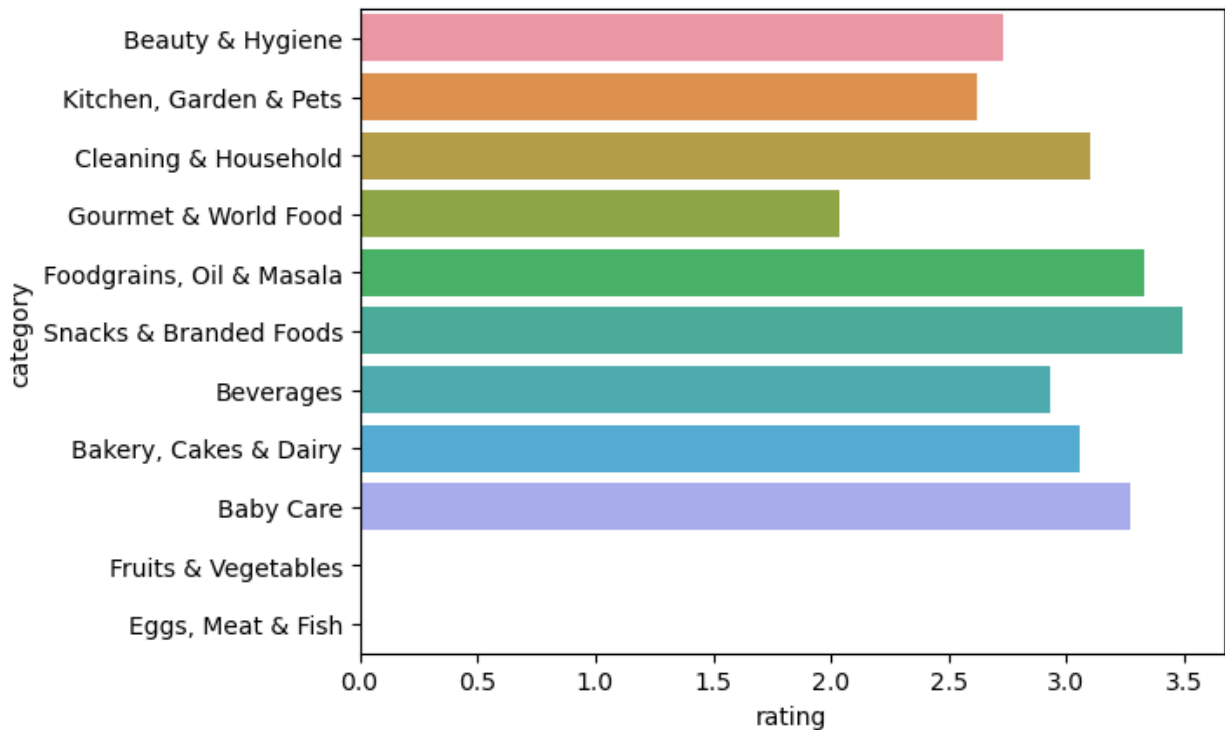
<Axes: xlabel='market_price', ylabel='discount'>



Categorical-Numerical

```
# Category-Rating
sns.barplot(data=df, y=df.category, x=df.rating, ci=False)

<Axes: xlabel='rating', ylabel='category'>
```



- Fruits and Vegetables have null ratings
- Eggs, Meats and Fish have null ratings

```
df[df.category=='Snacks & Branded Foods']
```

	product \
26	Granola - Happy Berries
28	Paratha Puff
39	Choco Deck - Mini Delights
49	Super Hot and Sweet Mango Chutney
53	Choco Deck - French Dessert Inspired Layered Bar
...	...
27506	Ready To Eat Tamarind Poha
27513	Sandwich Biscuits Fab - Jam In Fruit Flavour
27525	FunFoods Veg Mayonnaise Original
27547	Powder - Baking
27549	Papad - Garlic Disco

	category	sub_category
brand \		
26	Snacks & Branded Foods	Breakfast Cereals Fit & Flex

28	Snacks & Branded Foods	Frozen Veggies & Snacks	Switz
39	Snacks & Branded Foods	Chocolates & Candies	Fabelle
49	Snacks & Branded Foods	Pickles & Chutney	Spice Story
53	Snacks & Branded Foods	Chocolates & Candies	Fabelle
...
27506	Snacks & Branded Foods	Ready To Cook & Eat	Instafeast
27513	Snacks & Branded Foods	Biscuits & Cookies	Parle
27525	Snacks & Branded Foods	Spreads, Sauces, Ketchup	Dr. Oetker
27547	Snacks & Branded Foods	Ready To Cook & Eat	Kwality
27549	Snacks & Branded Foods	Ready To Cook & Eat	Atish

	sale_price	market_price	type	rating \
26	245.0	265.0	Granola & Cereal Bars	3.5
28	90.0	110.0	Frozen Veg Snacks	4.3
39	160.0	200.0	Chocolates	4.4
49	100.0	125.0	Chutney Powder	3.0
53	56.0	70.0	Chocolates	4.0
...
27506	55.0	55.0	Breakfast & Snack Mixes	3.4
27513	21.0	35.0	Cream Biscuits & Wafers	4.1
27525	39.0	39.0	Mayonnaise	4.2
27547	38.0	38.0	Home Baking	3.7
27549	61.0	61.0	Papads, Ready To Fry	4.0

	description	discount
26	Fit & Flex granola is not just crunchy and tas...	20.0
28	Switz Puff Parathas are made with the highest ...	20.0
39	Fabelle brings you Fabelle Choco Deck – Mini D...	40.0
49	From the Bylanes & Streets of Eastern UP, come...	25.0
53	Fabelle Choco Deck is a unique multi-layered c...	14.0
...
27506	Express Feast is the brainchild of two dynamic...	0.0
27513	Parle Sandwich Biscuits Jam in Fruit Flavour r...	14.0
27525	Veg Mayonnaise is a FunFoods innovation that w...	0.0
27547	Kwality Baking Powder Directions for use: Use ...	0.0
27549	Papads are prepared from urad dal flour and sp...	0.0

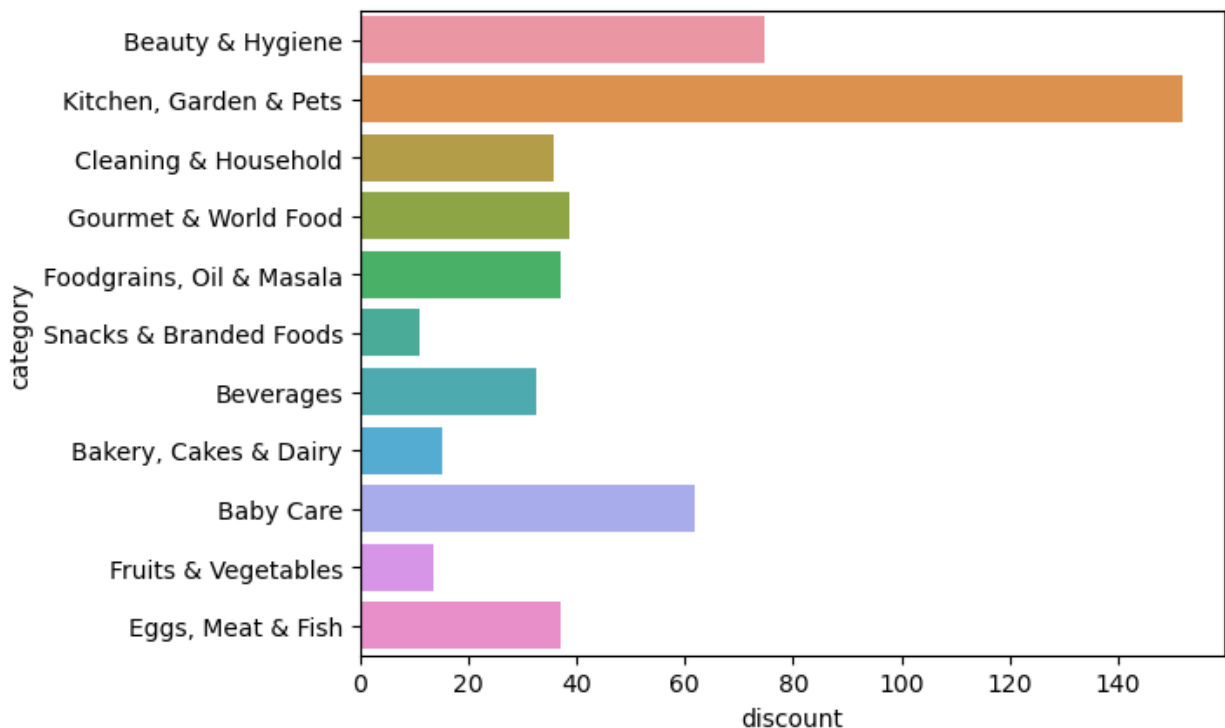
[2814 rows x 10 columns]

df.category.value_counts()

```
category
Beauty & Hygiene      7867
Gourmet & World Food  4690
Kitchen, Garden & Pets 3580
Snacks & Branded Foods 2814
Foodgrains, Oil & Masala 2676
Cleaning & Household  2675
Beverages             885
Bakery, Cakes & Dairy  851
Baby Care             610
Fruits & Vegetables    557
Eggs, Meat & Fish     350
Name: count, dtype: int64
```

```
sns.barplot(data=df, y=df.category, x=df.discount, ci=False)
```

```
<Axes: xlabel='discount', ylabel='category'>
```



```
df[df.category=='Fruits & Vegetables']
```

```

      product      category \
89  Sweet Corn Kernels - Chipotle  Fruits & Vegetables
115  Sapota - Organically Grown  Fruits & Vegetables
122  Cabbage - Red  Fruits & Vegetables
131  Banana - Red  Fruits & Vegetables
132  Chilli - Bajji, Mild  Fruits & Vegetables
...      ...      ...
```

27310		Leeks	Fruits & Vegetables
27319	Tapioca - Organically Grown		Fruits & Vegetables
27377	Fried Rice Mix		Fruits & Vegetables
27487	Broad Beans - Diced		Fruits & Vegetables
27491	Green Cabbage - Grated		Fruits & Vegetables

	sub_category	brand	sale_price
--	--------------	-------	------------

market_price \

89	Cuts & Sprouts	Supa Corn	55.0
55.00			
115	Fresh Fruits	Fresho	35.0
43.75			
122	Fresh Vegetables	Fresho	23.0
28.75			
131	Fresh Fruits	Fresho	39.0
48.75			
132	Herbs & Seasonings	Fresho	14.0
17.50			

...
-----	-----	-----	-----	----

27310	Exotic Fruits & Veggies	Fresho	16.0
20.00			
27319	Organic Fruits & Vegetables	Fresho	18.0
22.50			
27377	Cuts & Sprouts	Fresho	45.0
56.25			
27487	Fresh Vegetables	Fresho	33.0
41.25			
27491	Fresh Vegetables	Fresho	28.0
35.00			

	type	rating \
--	------	----------

89	Cut & Peeled Veggies	0.0
115	Organic Fruits	0.0
122	Cabbage & Cauliflower	0.0
131	Banana, Sapota & Papaya	0.0
132	Lemon, Ginger & Garlic	0.0

...
-----	-----	-----

27310	Exotic Vegetables	0.0
27319	Organic Vegetables	0.0
27377	Recipe Packs	0.0
27487	Cut & Peeled Veggies	0.0
27491	Cut & Peeled Veggies	0.0

	description	discount
--	-------------	----------

89	Ready To Eat Sweet Corn kernels, packaged usin...	0.00
115	Brown skinned sapotas are smooth to grainy tex...	8.75
122	It is round, wrapped in purplish red leaf laye...	5.75
131	Standing apart from the common varieties of ye...	9.75
132	Fresho is our brand of fresh fruits and vegeta...	3.50

```

...
27310 Leek is a long bundle of leaf sheaths with a m... 4.00
27319 It Is Organically Grown And Handpicked From Fa... 4.50
27377 This pack of cut Beans, Carrots, Sweetcorn, Ba... 11.25
27487 Fresh broad beans have a delicious earthy flav... 8.25
27491 Sweet and grassy, grated green cabbage is a wo... 7.00

```

```
[557 rows x 10 columns]
```

```
mrp=df.groupby('category')['market_price'].sum()
```

```
mrp
```

```

category
Baby Care          364020.00
Bakery, Cakes & Dairy  134357.00
Beauty & Hygiene    3882642.22
Beverages          240927.00
Cleaning & Household  701150.35
Eggs, Meat & Fish    114042.42
Foodgrains, Oil & Masala  615833.00
Fruits & Vegetables   35889.55
Gourmet & World Food  1680993.95
Kitchen, Garden & Pets  2361574.40
Snacks & Branded Foods  396141.50
Name: market_price, dtype: float64

```

```
srp=df.groupby('category')['sale_price'].sum()
```

```
srp
```

```

category
Baby Care          326317.17
Bakery, Cakes & Dairy  121525.14
Beauty & Hygiene    3294599.44
Beverages          212186.67
Cleaning & Household  605043.09
Eggs, Meat & Fish    101114.12
Foodgrains, Oil & Masala  516916.23
Fruits & Vegetables   28345.36
Gourmet & World Food  1500115.31
Kitchen, Garden & Pets  1817723.12
Snacks & Branded Foods  364675.08
Name: sale_price, dtype: float64

```

```
dis=(mrp-srp)/mrp*100
```

```
dis=dis.round(2)
```

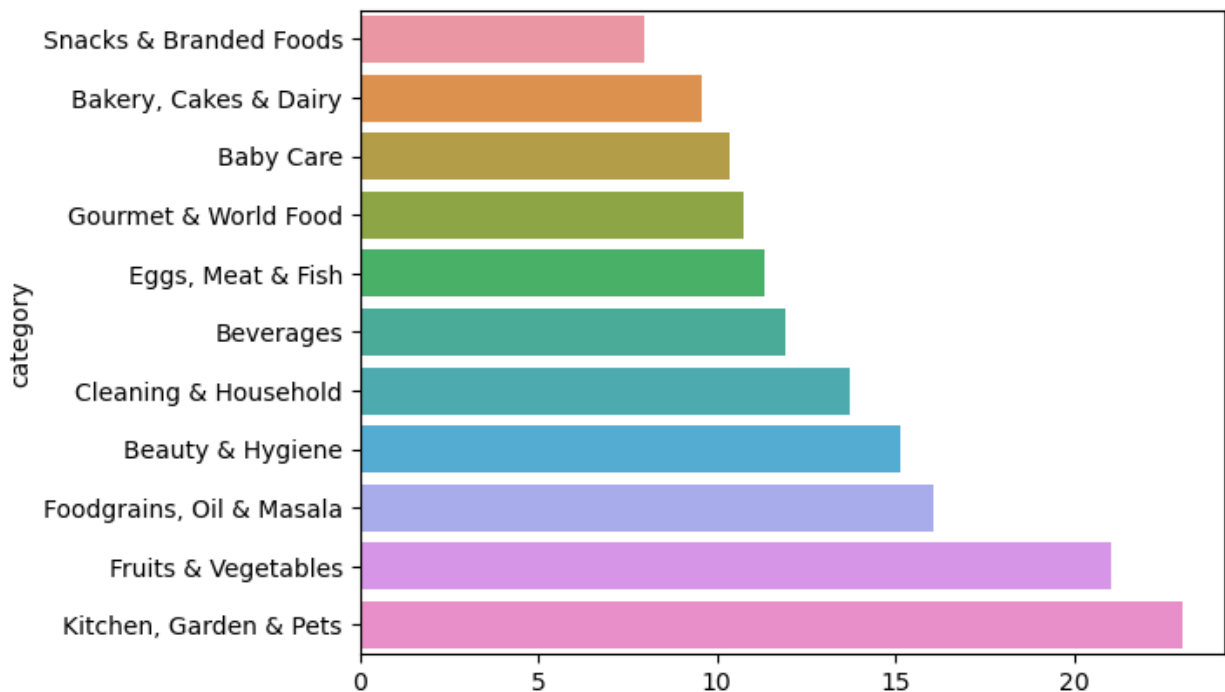
```
dis=dis.sort_values()
```

```
dis
```

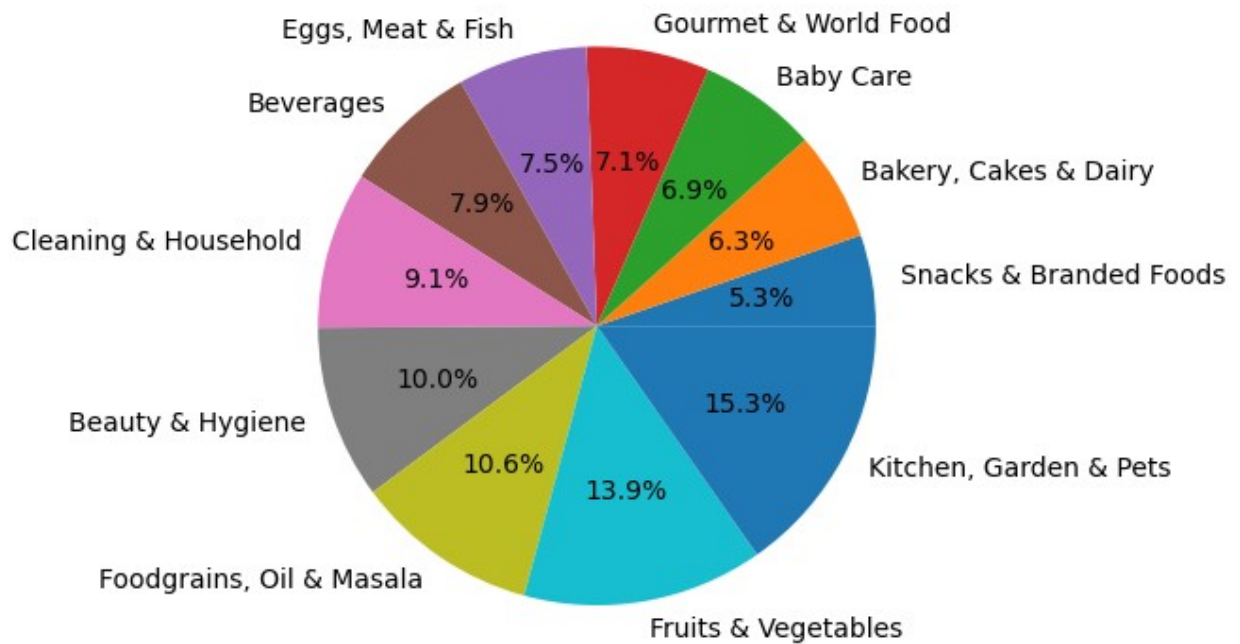
```
category
Snacks & Branded Foods      7.94
Bakery, Cakes & Dairy       9.55
Baby Care                   10.36
Gourmet & World Food        10.76
Eggs, Meat & Fish           11.34
Beverages                   11.93
Cleaning & Household        13.71
Beauty & Hygiene            15.15
Foodgrains, Oil & Masala    16.06
Fruits & Vegetables         21.02
Kitchen, Garden & Pets     23.03
dtype: float64
```

```
sns.barplot(y=dis.index,x=dis.values)
```

```
<Axes: ylabel='category'>
```



```
plt.pie(dis,labels=dis.index,autopct='%1.1f%%')
plt.show()
```

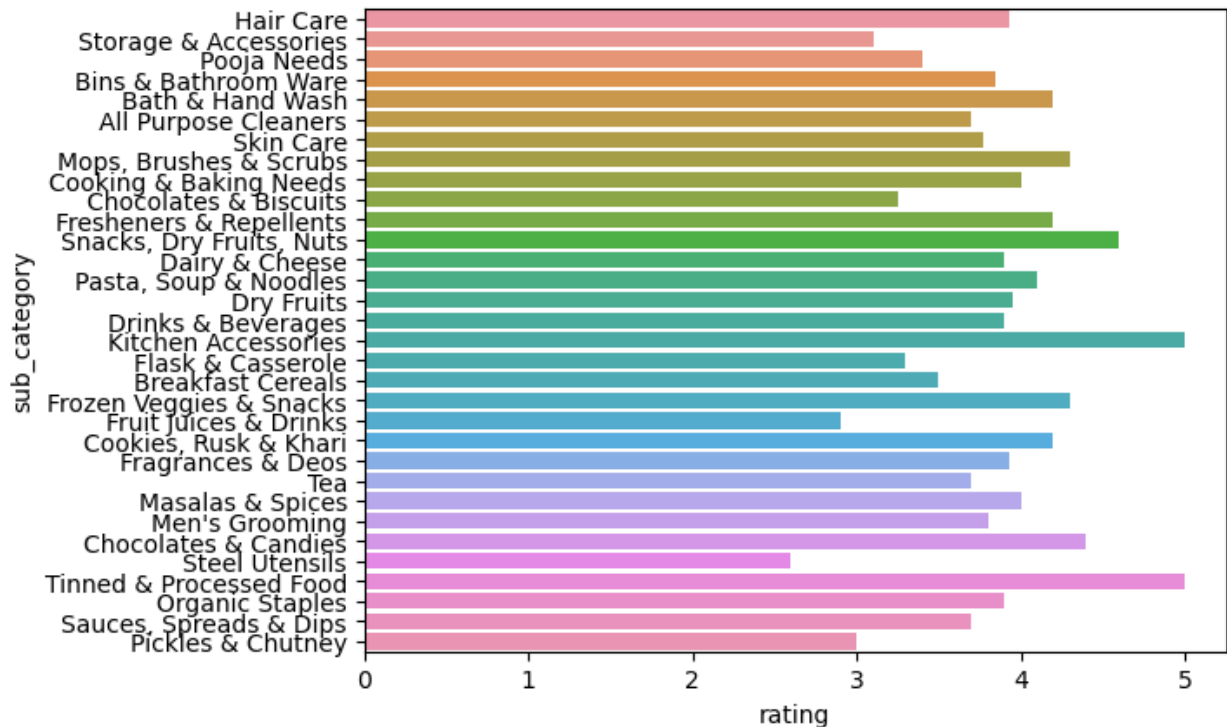


```
# Sub_Category and Rating
```

```
tempdf=df.head(50)
```

```
sns.barplot(data=tempdf,y=tempdf.sub_category,x=tempdf.rating,ci=False  
)
```

```
<Axes: xlabel='rating', ylabel='sub_category'>
```



Brand: Top 10 Brands

```
df.brand.value_counts()
```

brand

Fresho 638

bb Royal 539

BB Home 428

DP 250

Fresho Signature 171

...

GO CHOCO 1

JLT 1

MTR FOODS 1

Swiss Beauty 1

Baron De Bercy 1

Name: count, Length: 2314, dtype: int64

```
df.brand.value_counts().head(10)
```

brand

Fresho 638

bb Royal 539

BB Home 428

DP 250

Fresho Signature 171

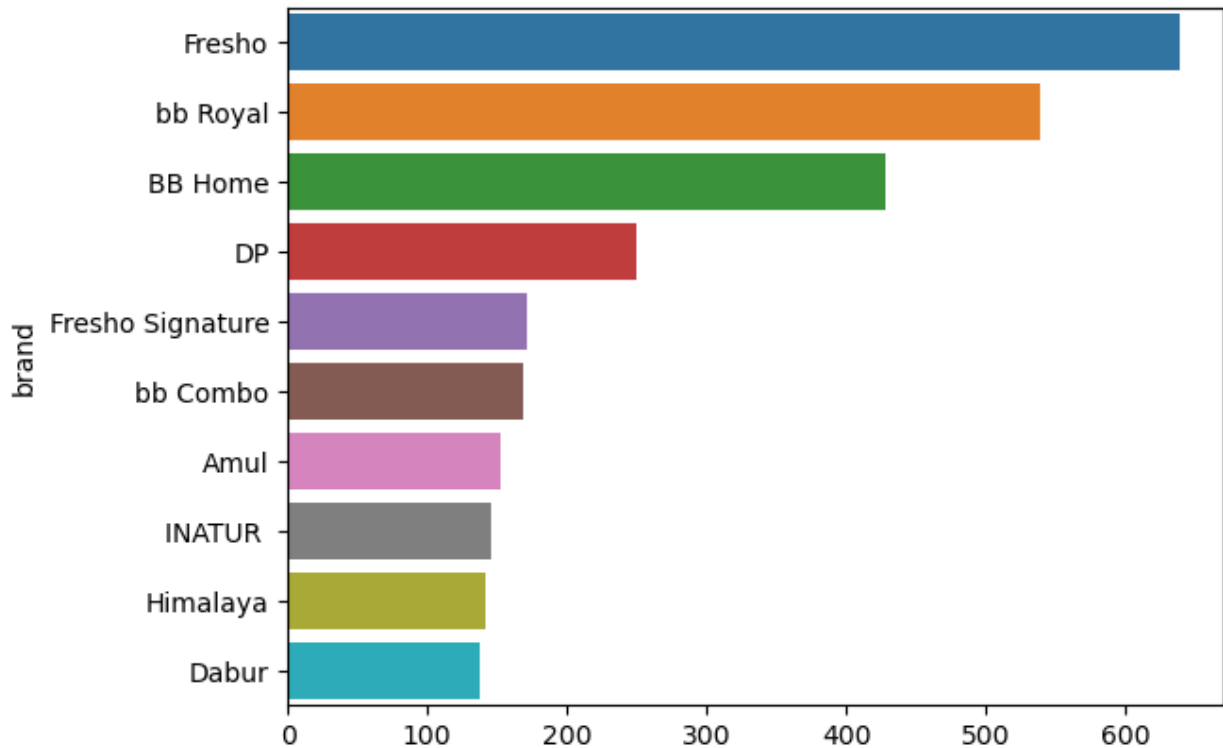
bb Combo 168

Amul 153

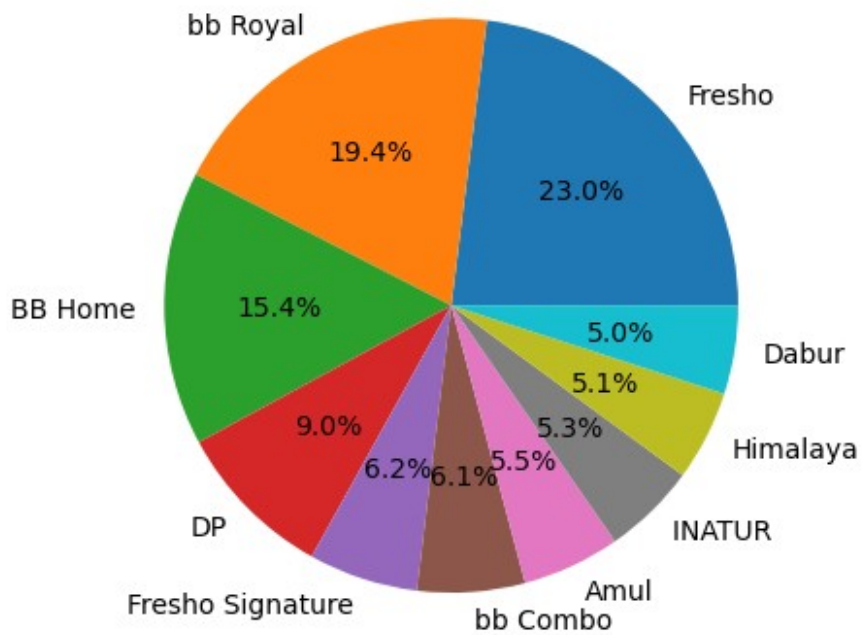
```
INATUR      146
Himalaya    141
Dabur       138
Name: count, dtype: int64
```

```
sns.barplot(df,y=df.brand.value_counts().head(10).index,x=df.brand.value_counts().head(10).values)
```

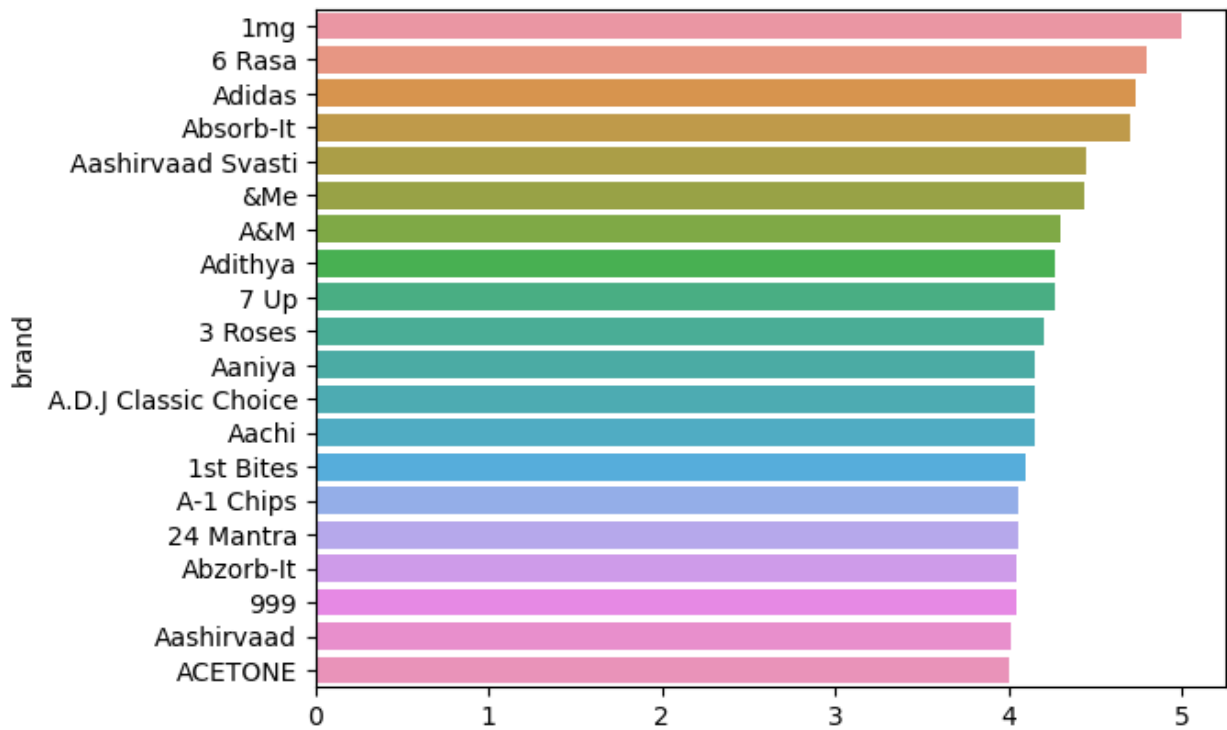
```
<Axes: ylabel='brand'>
```



```
plt.pie(df.brand.value_counts().head(10),labels=df.brand.value_counts(
).head(10).index,autopct='%1.1f%%')
plt.show()
```

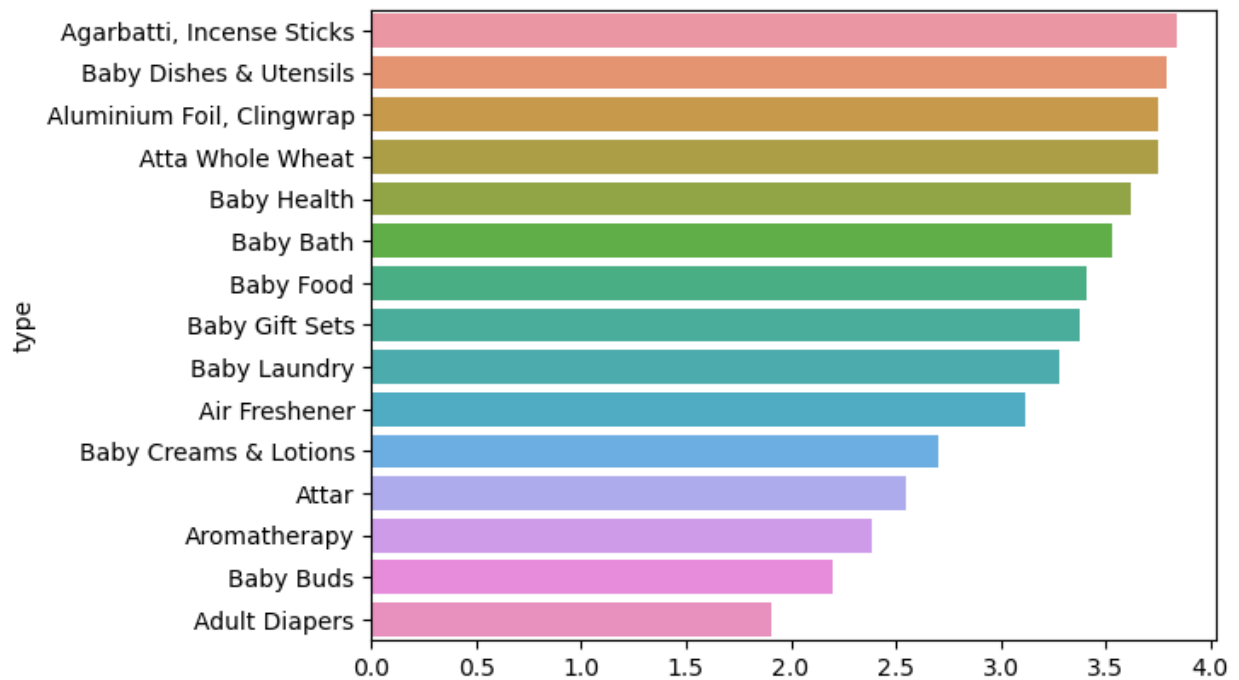



```
temp=df.groupby('brand')['rating'].mean()  
tempx=temp.head(50).sort_values(ascending=False)  
sns.barplot(y=tempx.head(20).index,x=tempx.head(20).values)  
<Axes: ylabel='brand'>
```

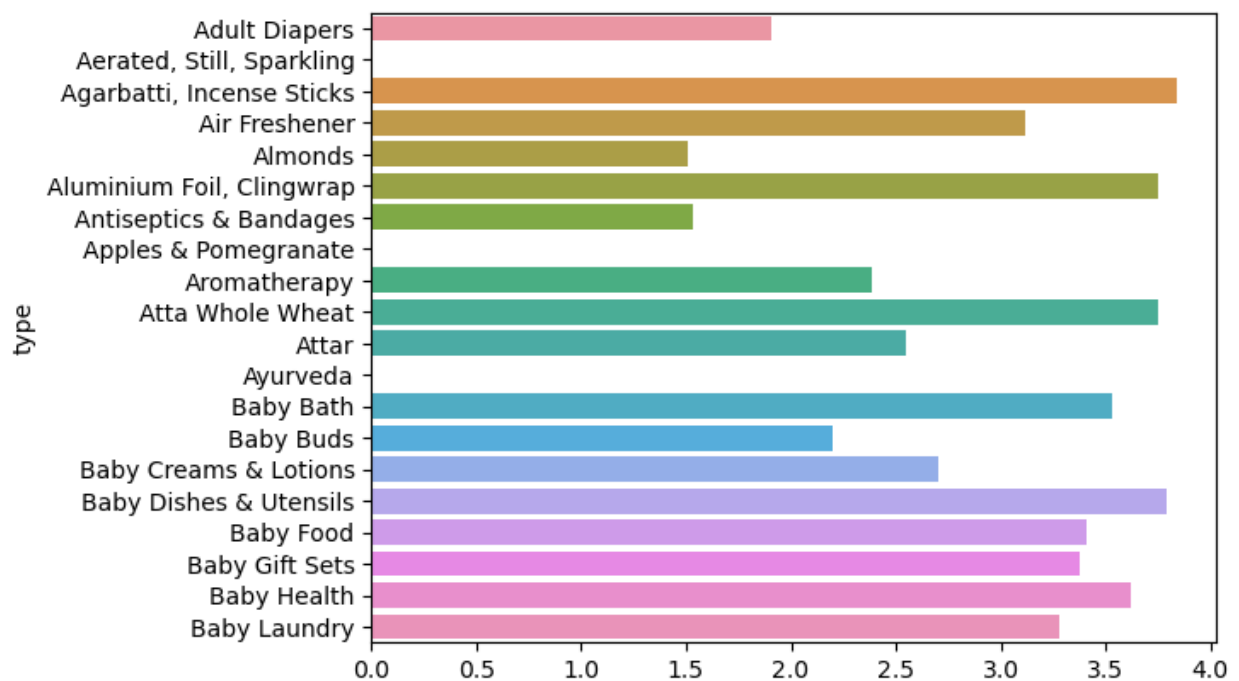


#Type and Rating

```
temp=df.groupby('type')['rating'].mean()
tempx=temp.head(20).sort_values(ascending=False)
sns.barplot(y=tempx.head(15).index,x=tempx.head(15).values)
<Axes: ylabel='type'>
```



```
tempc=temp.head(20)
sns.barplot(y=tempc.index,x=tempc.head(15).values)
<Axes: ylabel='type'>
```



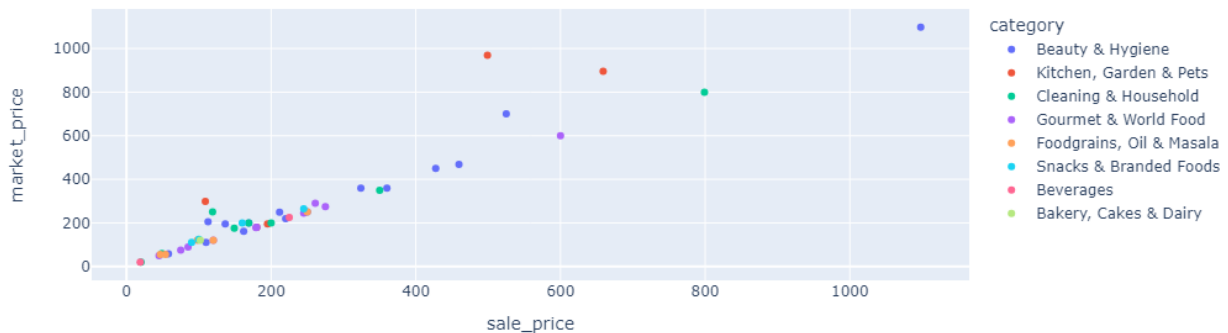
MultiVariant Analysis

- Multivariant Analysis is a statistical method used to analyze and understand relationship, patterns, and interactions among multiple variables within a data set simultaneously
- It involves examining the collective influence of several variables on one another and on the outcomes of interest

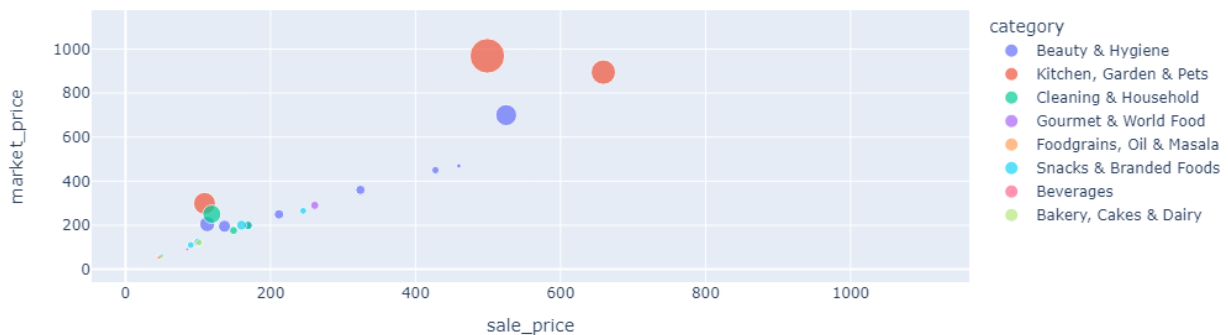
#Types of Graphs

- Bubble plot
- Cluster plot
- barplot

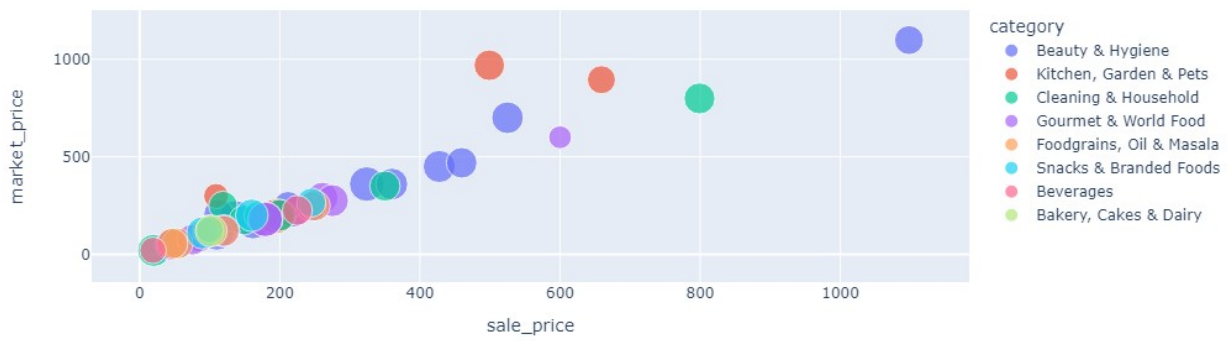
```
px.scatter(df.head(50),y='market_price',x='sale_price',color='category')
```



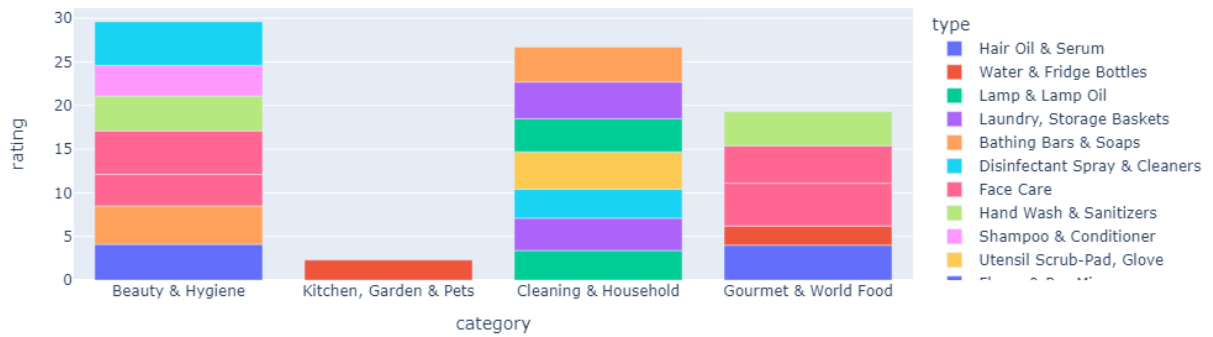
```
px.scatter(df.head(50),y='market_price',x='sale_price',color='category',size='discount')
```



```
px.scatter(df.head(50),y='market_price',x='sale_price',color='category',size='rating')
```



```
px.bar(df.head(20), x='category', y='rating', color='type')
```



```
sns.catplot(data=df.head(10), y='category', x='rating', hue='type', kind='bar')
```

```
<seaborn.axisgrid.FacetGrid at 0x259b4900f90>
```

