BIG BASKET

Big Basket is an Indian online grocery delivery platform founded in 2011 that has grown to become one of the largest in the country. It offers a wide range of products, including fresh produce, dairy, bakery items, household essentials, and personal care products, with options for same-day or next-day delivery. Known for its efficient supply chain and customer-centric approach, Big Basket has expanded its services across multiple cities in India.

Table of Content

Importing Libraries:

```python
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import plotly.express as px
```

```python
import warnings
warnings.filterwarnings('ignore')
```

Loading the Big Basket Product Dataset:

In this step we will load our CSV file.

```
[3]:  df=pd.read_csv('BigBasket Products.csv')
```

```
[4]:  df.head(12)
```

[4]:	index	product	category	sub_category	brand	sale_price	market_price	type	rating	description
0	1	Garlic Oil - Vegetarian Capsule 500 mg	Beauty & Hygiene	Hair Care	Sri Sri Ayurveda	220.0	220.0	Hair Oil & Serum	4.1	This Product contains Garlic Oil that is known...
1	2	Water Bottle - Orange	Kitchen, Garden & Pets	Storage & Accessories	Mastercook	180.0	180.0	Water & Fridge Bottles	2.3	Each product is microwave safe (without lid), ...
2	3	Brass Angle Deep - Plain, No.2	Cleaning & Household	Pooja Needs	Trm	119.0	250.0	Lamp & Lamp Oil	3.4	A perfect gift for all occasions, be it your m...
3	4	Cereal Flip Lid Container/ Storage Jar - Assort...	Cleaning & Household	Bins & Bathroom Ware	Nakoda	149.0	176.0	Laundry, Storage Baskets	3.7	Multipurpose container with an attractive desi...
4	5	Creme Soft Soap - For Hands & Body	Beauty & Hygiene	Bath & Hand Wash	Nivea	162.0	162.0	Bathing Bars & Soaps	4.4	Nivea Creme Soft Soap gives your skin the best...
5	6	Germ - Removal Multipurpose Wipes	Cleaning & Household	All Purpose Cleaners	Nature Protect	169.0	199.0	Disinfectant Spray & Cleaners	3.3	Stay protected from contamination with Multipu...
6	7	Multani Mati	Beauty & Hygiene	Skin Care	Satinance	58.0	58.0	Face Care	3.6	Satinance multani matti is an excellent skin t...

We can find sample of this Dataset.

```
]:  df.sample(3)
```

]:	index	product	category	sub_category	brand	sale_price	market_price	type	rating	description
11423	11424	Body Wash - Lavender & Almond Oil, Nourishing ...	Beauty & Hygiene	Bath & Hand Wash	Vivel	216.0	240.0	Shower Gel & Body Wash	4.2	Pamper yourself with the care of Vivel Lavende...
18907	18908	Glass Bottle With Plastic Outer - Blue Lid, BB...	Kitchen, Garden & Pets	Storage & Accessories	DP	189.0	349.0	Water & Fridge Bottles	NaN	Keep yourself hydrated by carrying this smart-...
13991	13992	Fudgy Brownie Bar	Gourmet & World Food	Chocolates & Biscuits	Purple Hippie	300.0	300.0	Cookies, Biscotti, Wafer	1.0	Granola bar meets brownie! This hand cut bar, ...

Understanding Dataset:

To understand this Dataset, we have to find columns, shape, info, size, duplicated, describe.

```
:  df.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
```

```
:  df.columns
```

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

```
:  df.shape
```

```
:  (27555, 10)
```

```
df.size
```

275550

```
df.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

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

[11]:  index            0
       product          1
       category         0
       sub_category     0
       brand            1
       sale_price       6
       market_price     0
       type             0
       rating        8636
       description    115
       dtype: int64

[12]:  df.duplicated()

[12]:  0          False
       1          False
       2          False
       3          False
       4          False
                  ...
       27550      False
       27551      False
       27552      False
       27553      False
       27554      False
       Length: 27555, dtype: bool
```

Columns in a Big Basket Dataset:

1. **Index**: The index column is a unique identifier for each row in a dataset. This column might not have a specific name, as it can simply be an automatic numbering from 0 to n-1.
2. **Product:** The name or description of the product.
3. **Category:** The category to which the product belongs (e.g., fruits, vegetables, dairy).

4. **Sub-category:** A more specific classification within the category (e.g., citrus fruits, leafy greens).
5. **Brand:** The brand name of the product.
6. **Sale price:** column contains the final price at which the product is sold to customers on Big Basket. This price reflects any discounts, promotions, or offers applied to the original price.
7. **Market price:** often representing the Maximum Retail Price (MRP), lists the original price of the product before any discounts or offers. This is the price set by the manufacturer or the standard price at which the product is typically sold in the market.
8. **Type:** The type column typically categorizes products or items within the dataset. This can represent the type or category of a product, such as "Fruit," "Vegetable," "Dairy," or even subcategories like "Organic," "Packaged," etc.
9. **Rating:** Customer feedback rating for the product or service.
10. **Description:** The description column in a Big Basket dataset typically contains detailed information about each product.

Preparing Dataset for Analysis:

To do EDA we have to clean the data first, to cleaned up I look up all necessary steps to ensure that the data is accurate and correct.

```
df.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
dtype: int64
```

```
df.brand.fillna(value='Unknown',inplace=True)
```

```
df['product'].fillna(value='Unknown',inplace=True)
```

```
df.description.fillna(value='Unknown',inplace=True)
```

I change string columns with 'Unknown' using 'fillna' and Numeric column with 'mean'.

With this method we fill all the null values in the Big Basket Dataset.

```
df.isnull().sum()
```

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

```
df.rating.fillna(value=df.rating.mean(),inplace=True)
```

```
df.sale_price.fillna(value=df.sale_price.mean(),inplace=True)
```
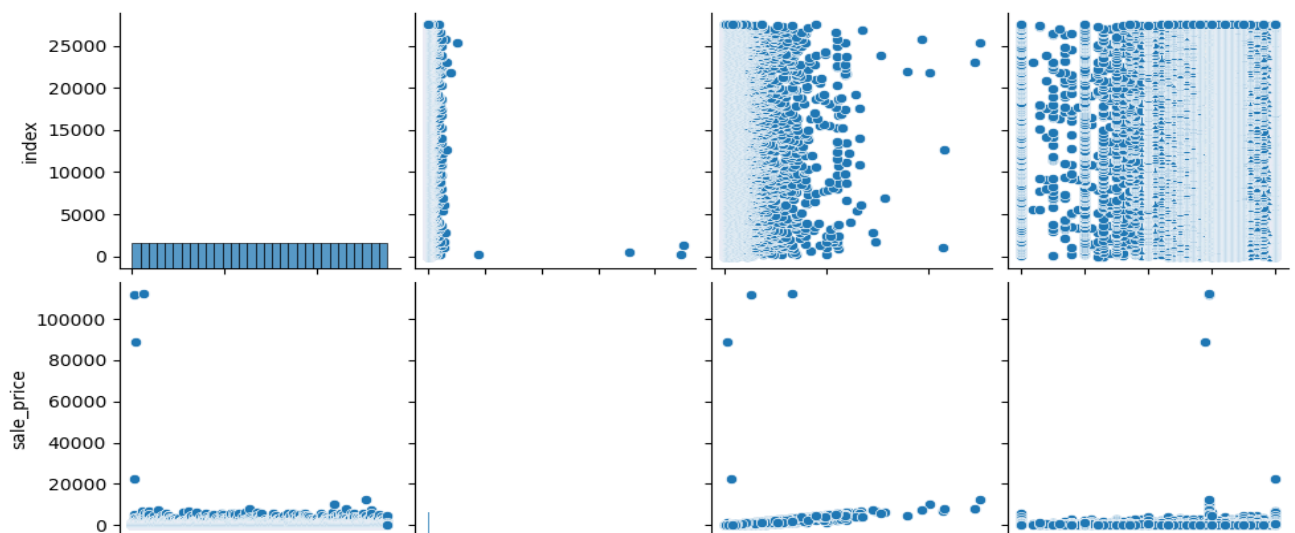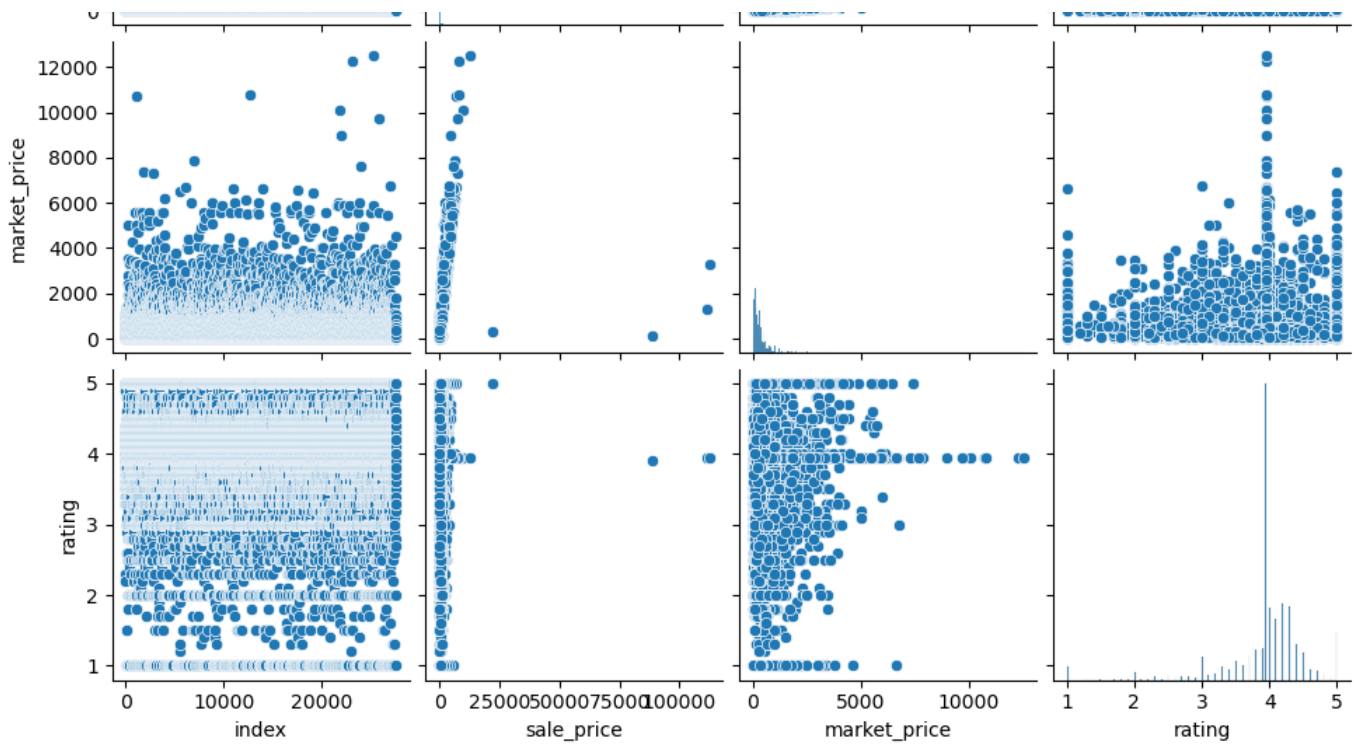
```
df.isnull().sum()
```

```
index           0
product         0
category        0
sub_category    0
brand           0
sale_price      0
market_price    0
type            0
rating          0
description     0
dtype: int64
```

Exploratory Data Analysis:

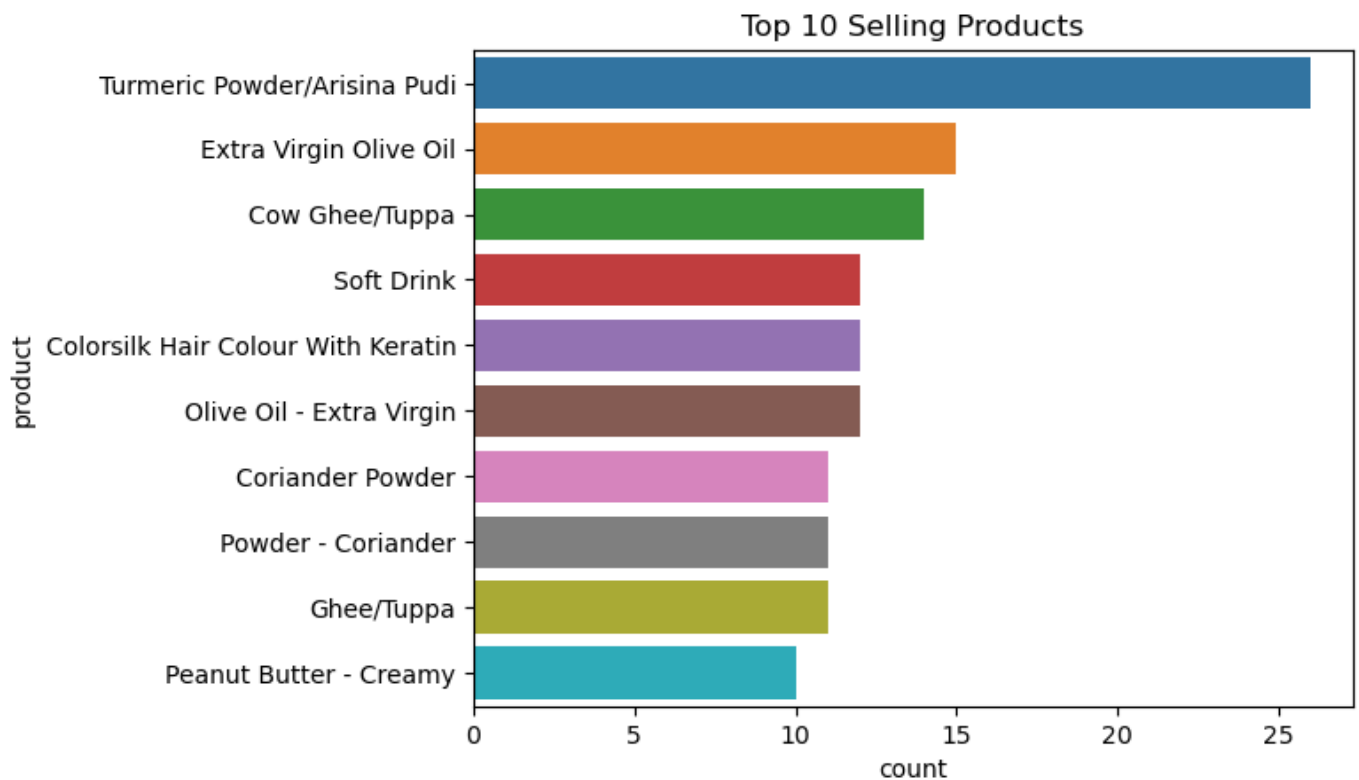Relation between Numerical Data in Dataset: We use pairplot.

```
[23]:  sns.pairplot(df)
       plt.show()
```

Top and Least selling Product

```python
sns.barplot(data=top,y='product',x='count')
plt.title('Top 10 Selling Products')
plt.show()
```
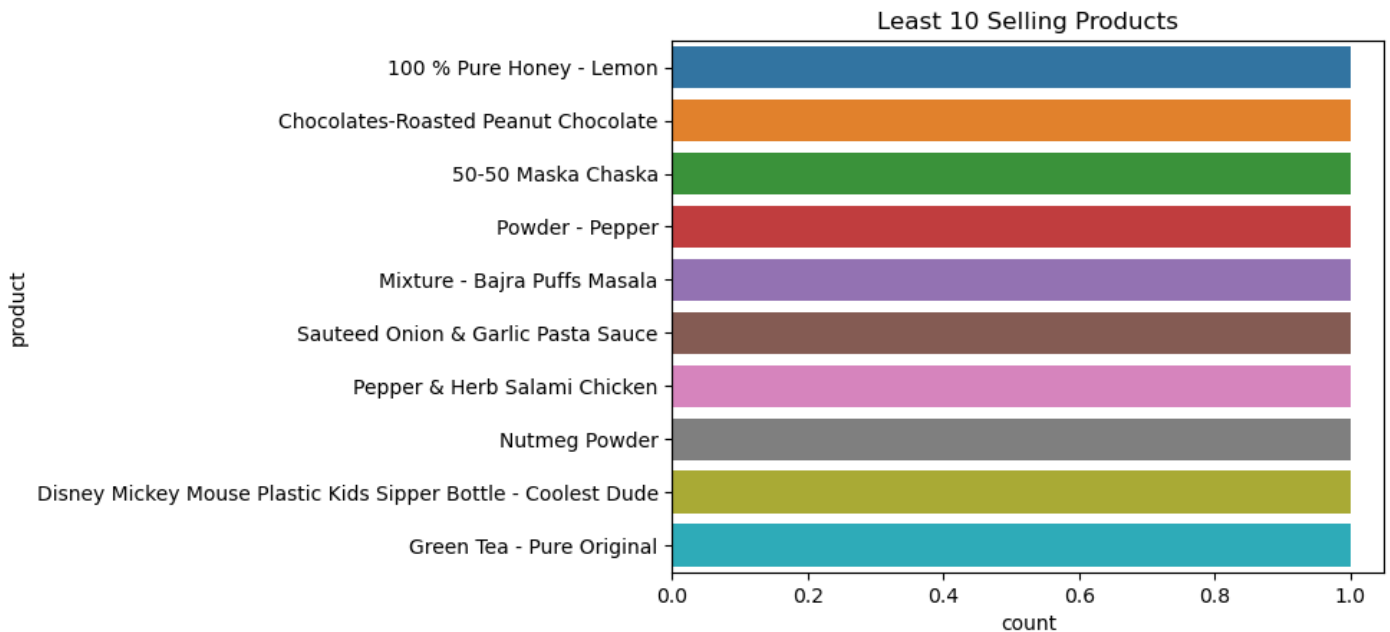

Top 10 Selling Products

- Analyzing the top 10 selling products on Big Basket can provide valuable insights into customer preferences, popular product categories, pricing strategies, and market trends. Assess the types

of products that are most frequently purchased. Are they daily essentials, luxury items, or specialty goods? This helps in understanding consumer behavior.

- **Example**: If most top-selling products are daily essentials like milk, bread, and rice, it suggests that customers primarily use Big Basket for their regular grocery needs.

```
sns.barplot(data=least,y='product',x='count')
plt.title('Least 10 Selling Products')
plt.show()
```


Least 10 Selling Products

Analyzing the top 10 least-selling products on Big Basket can provide crucial insights into various aspects of the business, helping identify areas for improvement, inventory management, and potential strategies to boost sales. Understanding why these products are not selling well can reveal gaps in customer demand. It could be due to low awareness, lack of interest, or poor product-market fit.

Distribution of rating:

```
]: sns.histplot(data=df,x= df.rating,bins=20)
   plt.title('Distribution of Rating')
   plt.show()
```


Distribution of Rating
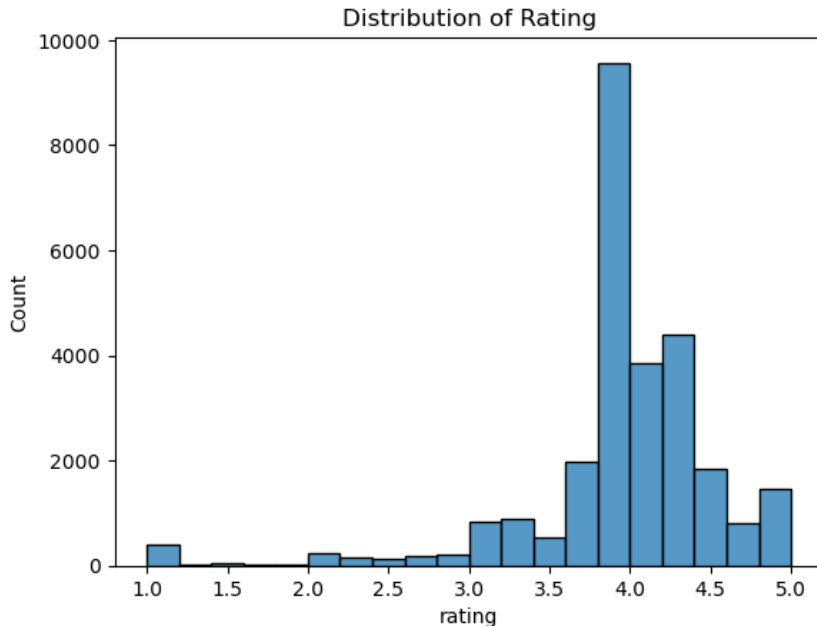
The average rating across products can give a sense of overall customer satisfaction. A higher average rating indicates that customers are generally happy with the products, while a lower average may signal issues that need addressing.

Top Five Brands

```
:  px.bar(TopB,x='brand',y='count',color='brand')
```
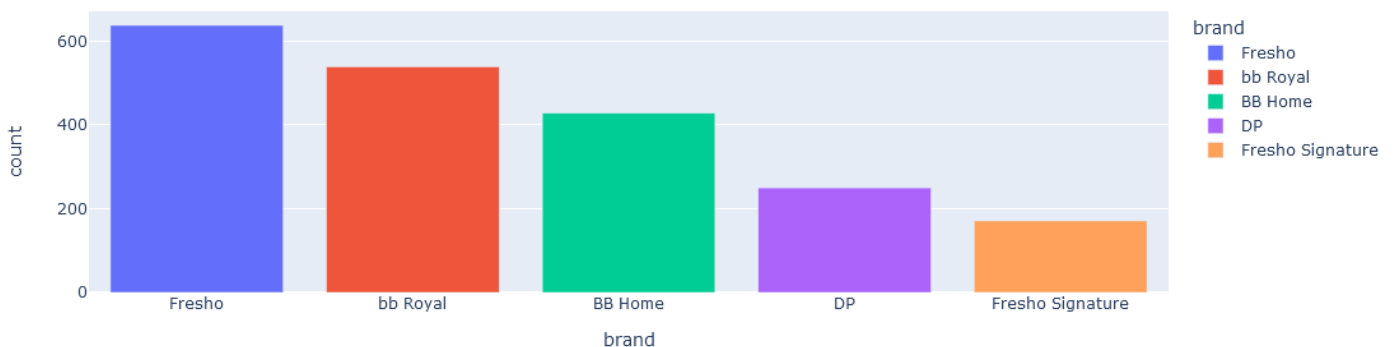

Understanding which brands are top sellers can reveal market dominance and share. Brands that consistently rank high in sales are likely leaders in their category and have strong market presence. Also, analyzing the top brands helps identify customer preferences and trends. Brands that frequently appear at the top are likely preferred due to factors like quality, pricing, and brand reputation.

The categories with the highest average sale price:

The categories with the highest average sale_price
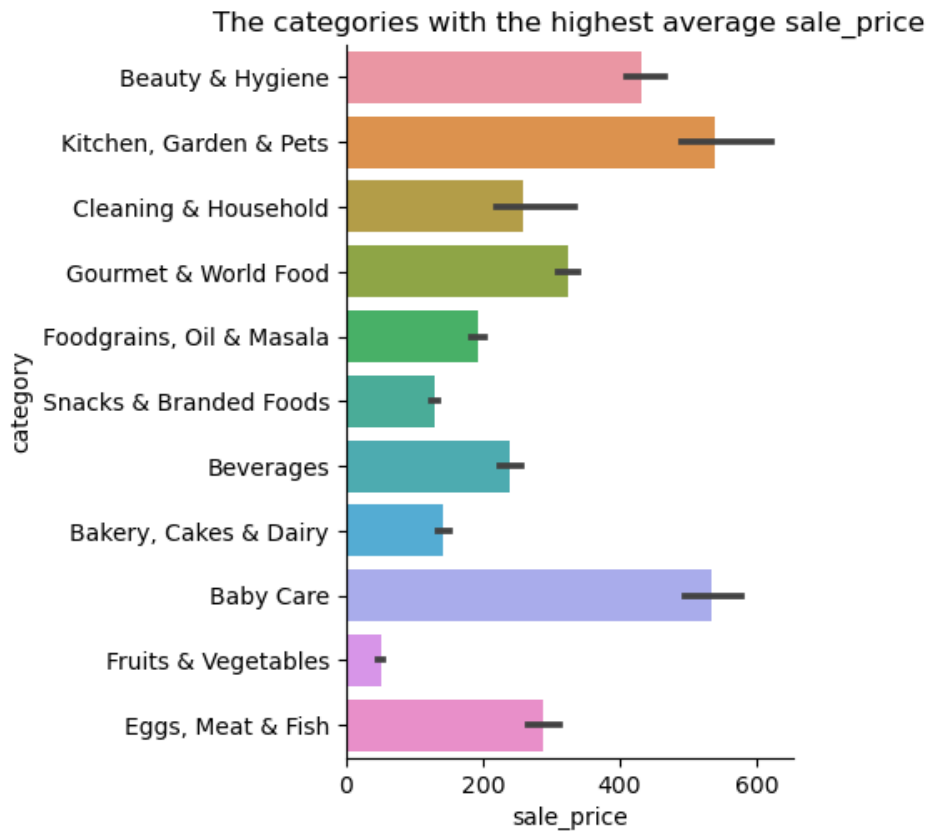
Baby Care category has the highest average sale price, approaching around 600 units. **Kitchen, Garden & Pets** is the second highest, with an average sale price close to 500 units. **Fruits & Vegetables** category has the lowest average sale price, close to zero. This analysis can be useful for understanding which categories tend to have higher value products and may influence pricing strategies, promotions, or inventory management

Finding Outliers of market price:

The majority of the data points for `market price` are concentrated towards the lower end, with the bulk of the values being below 2000 units. There are significant outliers in the `market price` variable. These outliers extend up to around 12,000 units. These outliers indicate that there are products in the dataset with much higher market prices compared to the majority of products.

```python
sns.boxplot(df['market_price'],orient= 'h')
plt.ylabel('market_price')
plt.title('Outliers of market_price')
plt.show()
```

Outliers of market_price
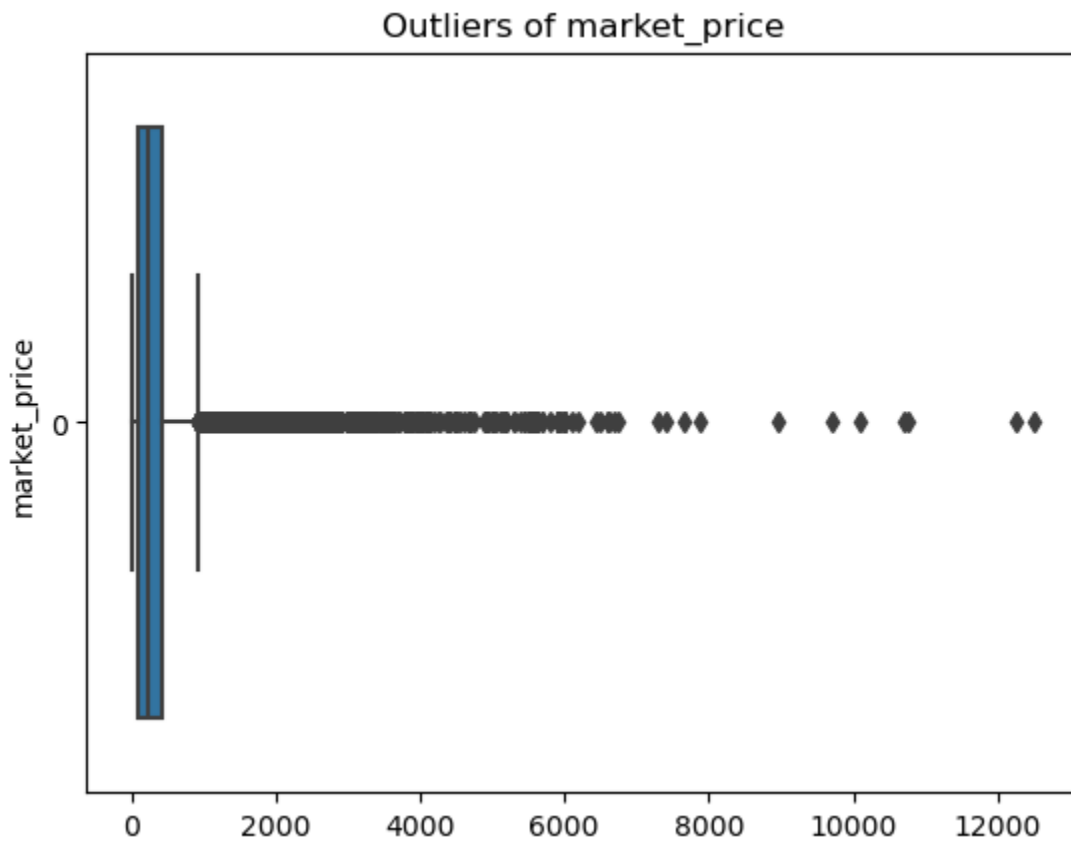
Contribution of each product
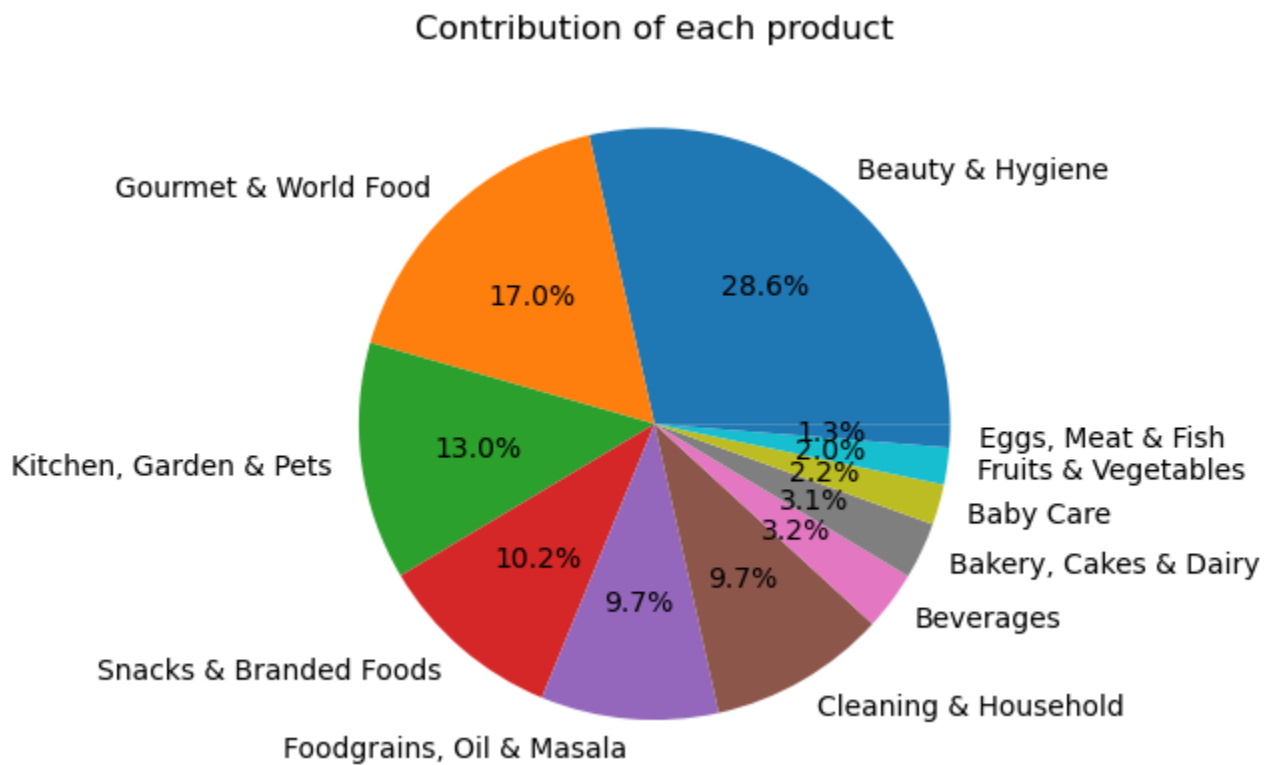
This pie chart is useful for understanding the distribution of product categories and where the majority of sales or products lie. While Beauty & Hygiene is dominant, there is still an important contribution from other categories like Gourmet & World Food, and Kitchen, Garden & Pets, showing a diverse product portfolio.

Top 10 sub-category:
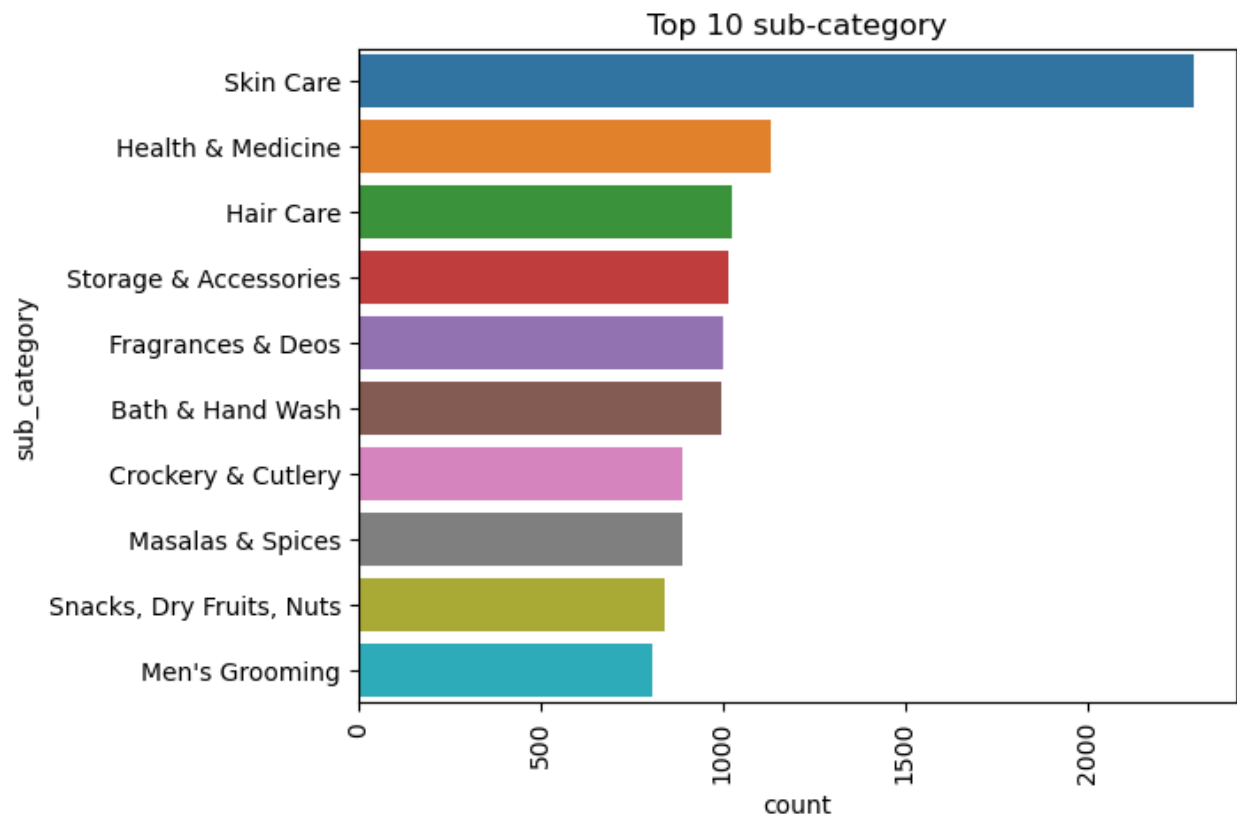
understanding the top-performing sub-categories can help in identifying trends, seasonal demands, and potential areas for growth or further investment. By focusing on these high-performing segments, a company can optimize its strategies to enhance profitability and customer satisfaction.
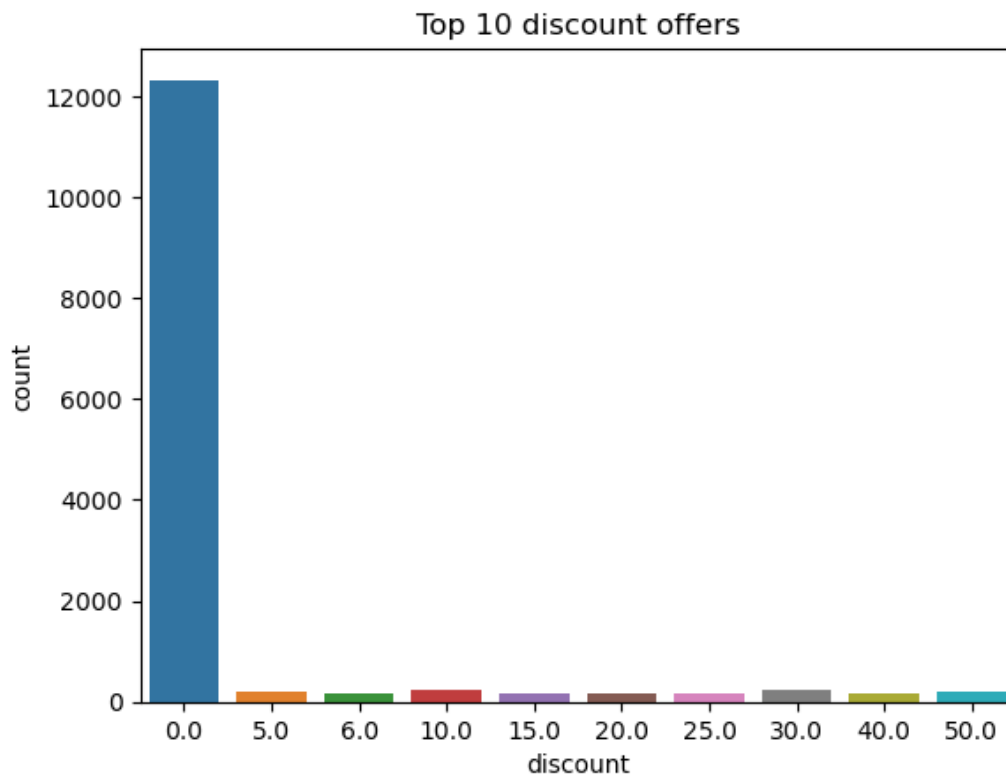
Top 10 discount offers:

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

```
df.head()
```

	index	product	category	sub_category	brand	sale_price	market_price	type	rating	description	discount
0	1	Garlic Oil - Vegetarian Capsule 500 mg	Beauty & Hygiene	Hair Care	Sri Sri Ayurveda	220.0	220.0	Hair Oil & Serum	4.1	This Product contains Garlic Oil that is known...	0.0
1	2	Water Bottle - Orange	Kitchen, Garden & Pets	Storage & Accessories	Mastercook	180.0	180.0	Water & Fridge Bottles	2.3	Each product is microwave safe (without lid), ...	0.0
2	3	Brass Angle Deep - Plain, No.2	Cleaning & Household	Pooja Needs	Trm	119.0	250.0	Lamp & Lamp Oil	3.4	A perfect gift for all occasions, be it your m...	131.0
3	4	Cereal Flip Lid Container/ Storage Jar - Assort...	Cleaning & Household	Bins & Bathroom Ware	Nakoda	149.0	176.0	Laundry, Storage Baskets	3.7	Multipurpose container with an attractive desi...	27.0
4	5	Creme Soft Soap - For Hands & Body	Beauty & Hygiene	Bath & Hand Wash	Nivea	162.0	162.0	Bathing Bars & Soaps	4.4	Nivea Creme Soft Soap gives your skin the best...	0.0

In this step we make a new column by subtracting market price and sales price and name it as discount column.

Analyzing the top 10 discounts offered can reveal important insights into the effectiveness of promotional strategies and their impact on sales. By identifying which discounts are most attractive to customers, you can gauge the price sensitivity and purchasing triggers of your target audience.

Conclusion:

As a student working on an exploratory data analysis (EDA) of Big Basket's dataset, I've learned a lot about the company's products, pricing, and what customers like to buy. The analysis showed that categories like Beauty & Hygiene and Gourmet & World Food are very popular, with high sales and contributions, meaning customers really like these products. I also noticed some very high prices in the data, which suggests there are premium products that may affect the overall pricing strategy. Additionally, looking at the discounts offered, it became clear that customers are quite responsive to promotions, which can be used strategically to boost sales.

This project helped me better understand how to use data to make informed business decisions and highlighted the importance of analyzing different parts of a business to find trends and improve performance.