BigBasket EDA

```
# importing mojor libraries
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
import plotly.express as px

# addition libraries
import warnings
warnings.filterwarnings('ignore')

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

df
```

```
       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]
```

```
df.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...
```

```
df.describe()
```

```
              index     sale_price  market_price        rating
count  27548.000000   27548.000000  27548.000000  27548.000000
mean   13780.555213     334.653279    382.122818      2.707529
std     7953.679836    1202.123658    581.787524      1.929041
min        1.000000       2.450000      3.000000      0.000000
25%     6893.750000      95.000000    100.000000      0.000000
50%    13780.500000     190.200000    220.000000      3.800000
75%    20668.250000     359.000000    425.000000      4.200000
max    27555.000000  112475.000000  12500.000000      5.000000
```

```
df.isnull().mean()*100
```

```
index            0.000000
product          0.003629
category         0.000000
sub_category     0.000000
brand            0.003629
sale_price       0.021775
market_price     0.000000
type             0.000000
rating          31.340954
description      0.417347
dtype: float64
```

```
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.sample()
```

```
      index      product                         category      sub_category
brand  \
4189   4190  Meat Masala  Foodgrains, Oil & Masala  Masalas & Spices
Orika

      sale_price  market_price             type  rating  \
4189       57.75          77.0  Blended Masalas     5.0

                                            description
4189  Crafted especially for all the mutton aficiona...
```

```
df.isna().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
```

```
# droping null value from rows
df = df.dropna(subset=['product'])
df = df.dropna(subset=['sale_price'])

# filling null value with its mean
df['brand'].fillna('Unknown',inplace=True)
df['rating'].fillna(0,inplace=True)
df['description'].fillna('No Description',inplace=True)

missing_values = df.isnull().sum()

print("Missing values in each column:")
print(missing_values)
```

```
Missing values in each column:
index               0
product             0
category            0
sub_category        0
brand               0
sale_price          0
market_price        0
type                0
rating              0
```

```
description         0
discount_percent    0
dtype: int64
```

```python
# choose the item
item_name = "Baby Care"

item_df = df[df["product"].str.contains(item_name, case=False)]

item_df["discount_amount"] = item_df["market_price"] -
item_df["sale_price"]
item_df["discount_percent"] = (item_df["discount_amount"] /
item_df["market_price"]) * 100

print(item_df[["product","market_price","sale_price","discount_amount"
,"discount_percent"]])
```

```
                                                  product  market_price  
\
2317   Absorbent Soft Cotton Wool/Roll - For Makeup R...          60.0   

3920                                 Baby Care Travel Kit         499.0   

11322  Baby Care Collection - with Organic Cotton Bib...         200.0   

15785        Mom & Baby Care Essentials Suitcase Gift Box        2399.0   

17220  Baby Care Collection Baby Gift Set - with Orga...         570.0   

20513  Baby Care Collection Baby Gift Set - with Orga...         350.0   

24031     Baby Care Collection - with Organic Cotton Bib         125.0   


       sale_price  discount_amount  discount_percent  
2317        60.00             0.00               0.0  
3920       424.15            74.85              15.0  
11322      200.00             0.00               0.0  
15785     2399.00             0.00               0.0  
17220      570.00             0.00               0.0  
20513      350.00             0.00               0.0  
24031      125.00             0.00               0.0  
```

Data Card

Dataset Overview

- **Dataset Name:** BigBasket Product Dataset

- **Domain:** E-commerce / Online Grocery Retail
- **Data Type:** Structured, Product-level data
- **Granularity:** Individual product records

Dataset Description

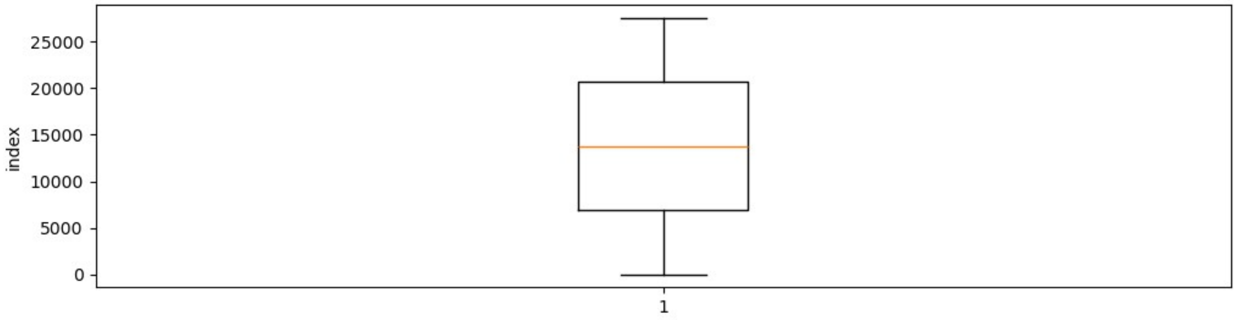
This dataset contains detailed information about products listed on **BigBasket**, including **categories**, **brands**, **pricing**, **discounts**, and **customer ratings**.
It is useful for analyzing **pricing strategies**, **category performance**, **brand positioning**, and **customer engagement** in online retail.

Key Features (Columns)

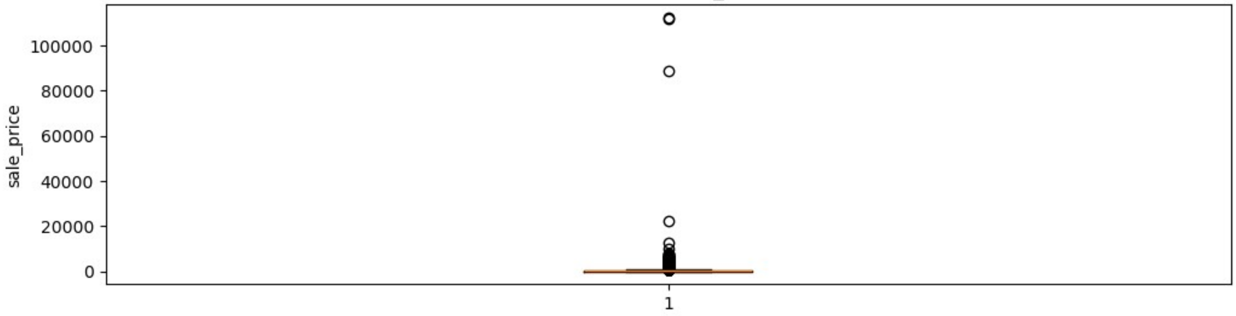
Column Name	Description
`product_name`	Name of the product
`category`	Product category
`sub_category`	Sub-category of the product
`brand`	Brand name
`market_price`	Original price (MRP)
`sale_price`	Discounted selling price
`discount`	Discount amount or percentage
`rating`	Customer rating (0–5 scale)
`description`	Product description

```python
numeric_cols = df.select_dtypes(include=np.number).columns

for col in numeric_cols:
    plt.figure(figsize=(12,3))
    plt.boxplot(df[col].dropna())
    plt.title(f"Outliers in {col}")
    plt.ylabel(col)
    plt.show()
```

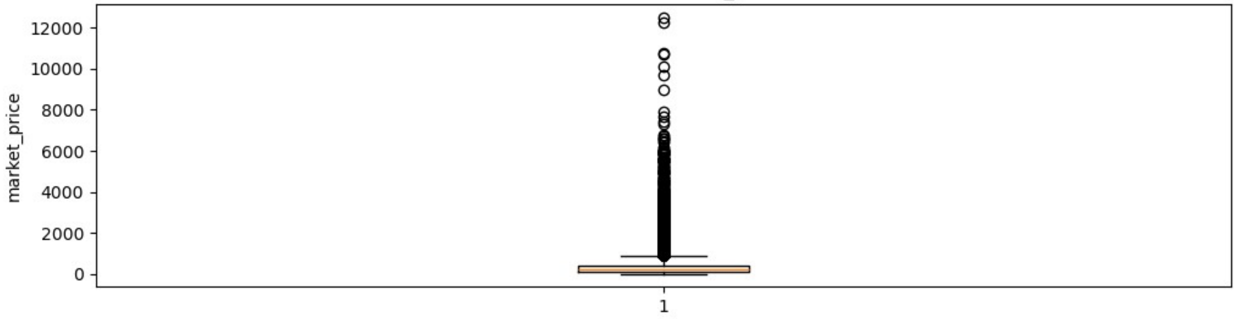
Outliers in index



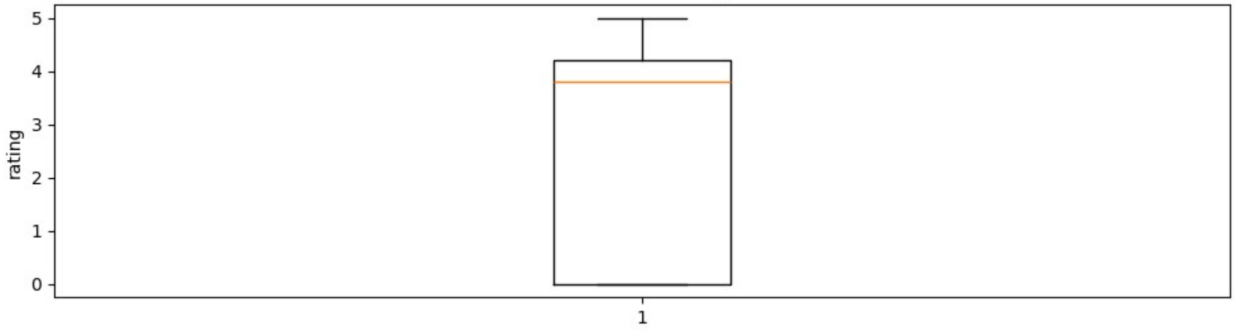
Outliers in sale_price

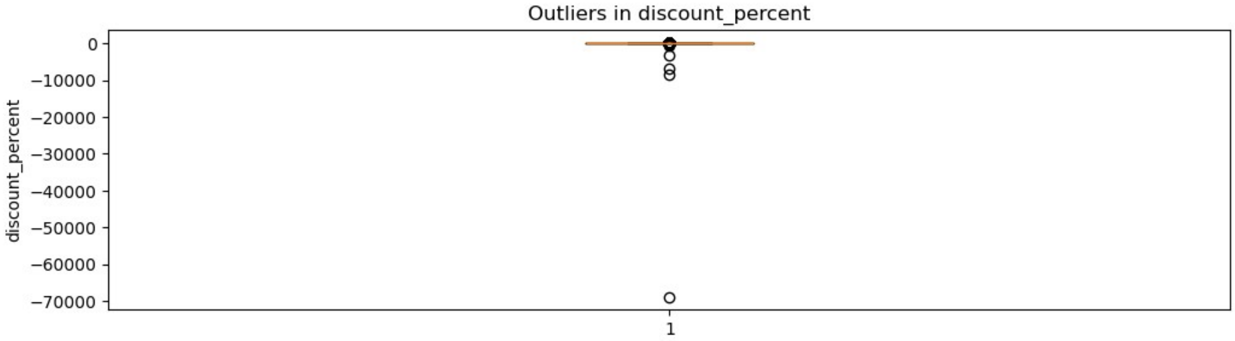


Outliers in market_price



Outliers in rating

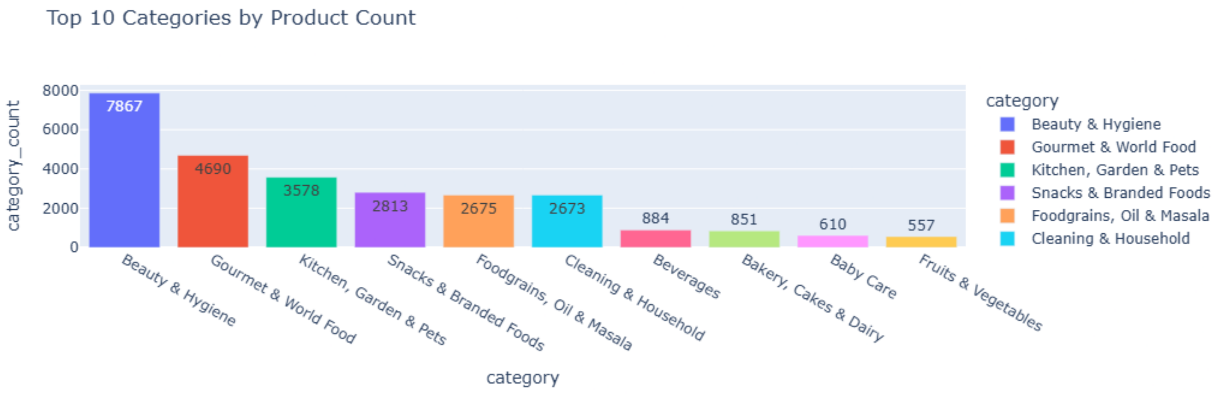




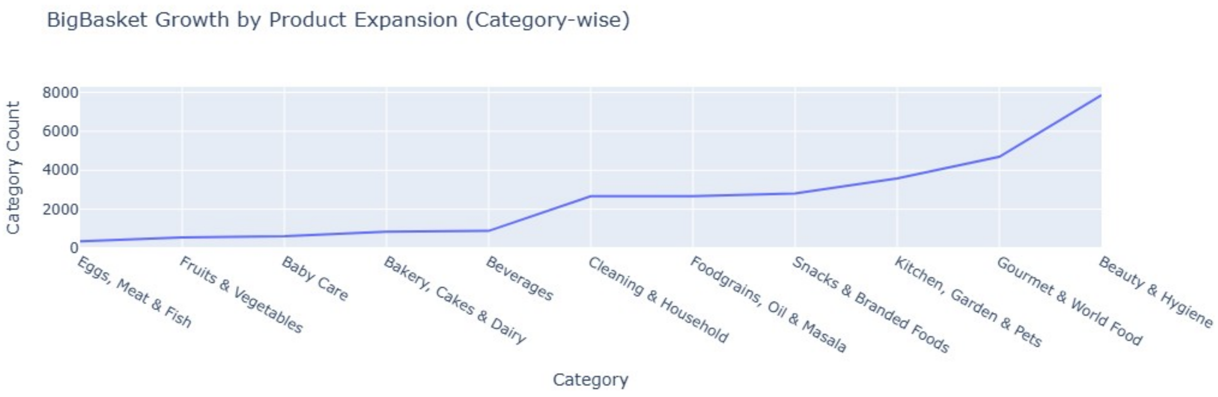
```
# categories and its product counts
temp = df['category'].value_counts().reset_index()
temp.columns = ['category', 'category_count']
temp
```

```
                     category  category_count
0            Beauty & Hygiene            7867
1        Gourmet & World Food            4690
2      Kitchen, Garden & Pets            3578
3      Snacks & Branded Foods            2813
4    Foodgrains, Oil & Masala            2675
5        Cleaning & Household            2673
6                   Beverages             884
7       Bakery, Cakes & Dairy             851
8                   Baby Care             610
9          Fruits & Vegetables            557
10          Eggs, Meat & Fish             350
```

```
# Product Distribution by Category
px.bar(temp.head(10),x='category',y='category_count',color='category',
text='category_count',title='Top 10 Categories by Product Count')
```

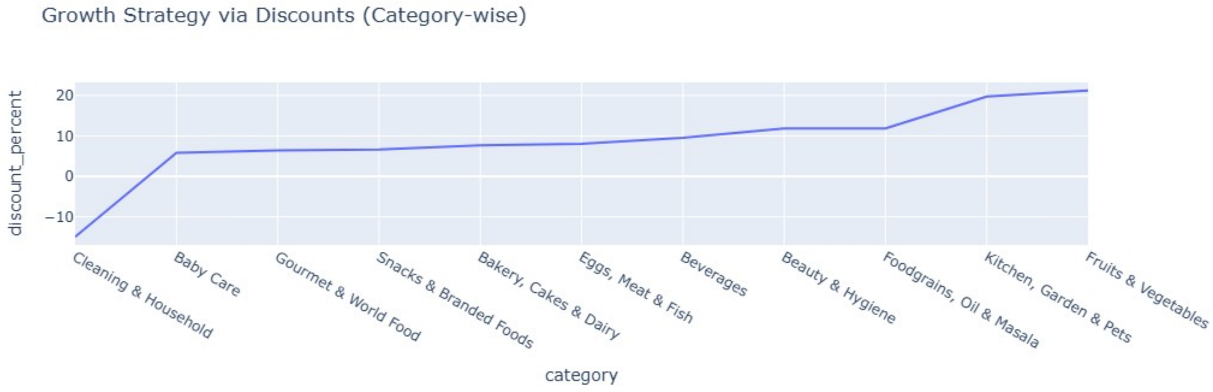


```
temp = temp.sort_values('category_count')
px.line(temp,x='category',y='category_count',labels={'category':'Categ
ory','category_count':'Category Count'},
        title='BigBasket Growth by Product Expansion (Category-wise)')
```



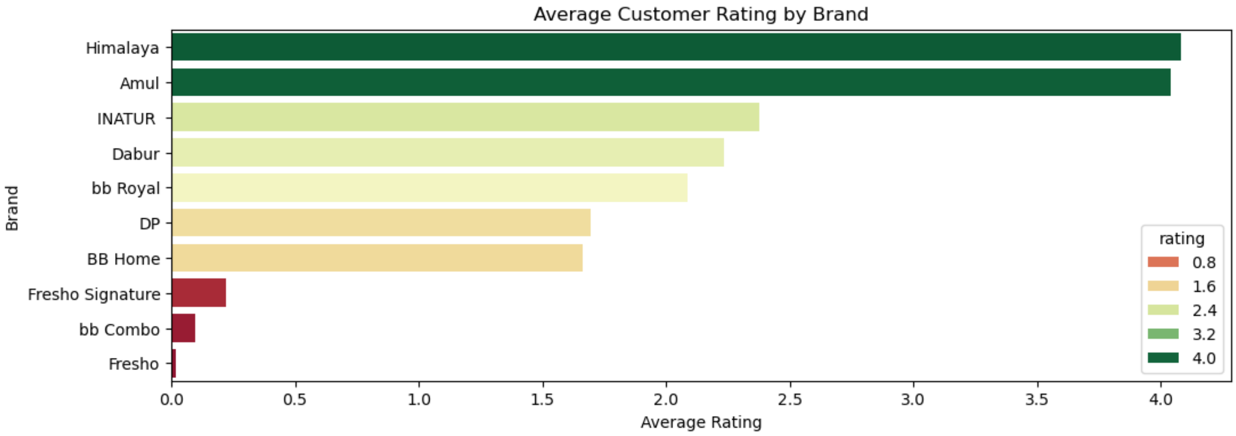
BigBasket Growth by Product Expansion (Category-wise)

```
df['discount_percent'] = ((df['market_price'] - df['sale_price']) /
df['market_price']) * 100
discount_growth = (df.groupby('category')
['discount_percent'].mean().reset_index().sort_values('discount_percen
t'))
px.line(discount_growth,x='category',y='discount_percent',
        title='Growth Strategy via Discounts (Category-wise)')
```

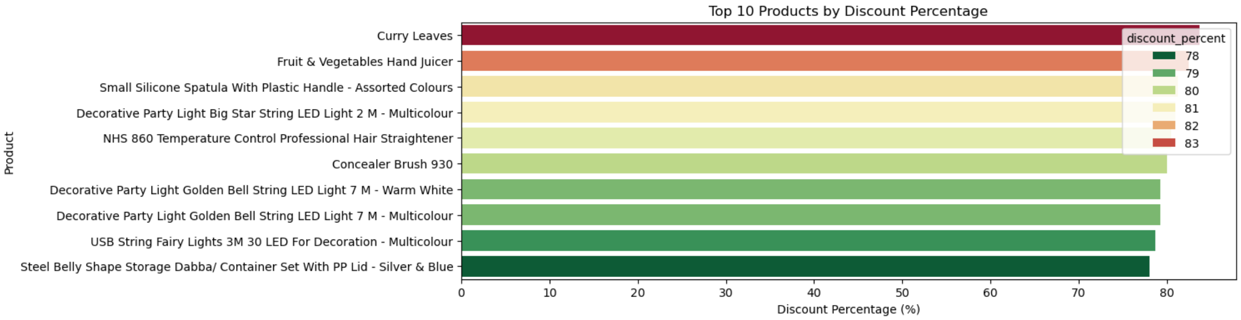


Growth Strategy via Discounts (Category-wise)

```
# checking brand by it rating
top_brands = df['brand'].value_counts().head(10).index
brand_rating = (df[df['brand'].isin(top_brands)].groupby('brand')
['rating'].mean().reset_index().sort_values('rating',ascending=False))
plt.figure(figsize=(12,4))
sns.barplot(data=brand_rating,x='rating',y='brand',hue='rating',palett
e = "RdYlGn")
plt.title('Average Customer Rating by Brand')
plt.xlabel('Average Rating')
plt.ylabel('Brand')
plt.show()
```

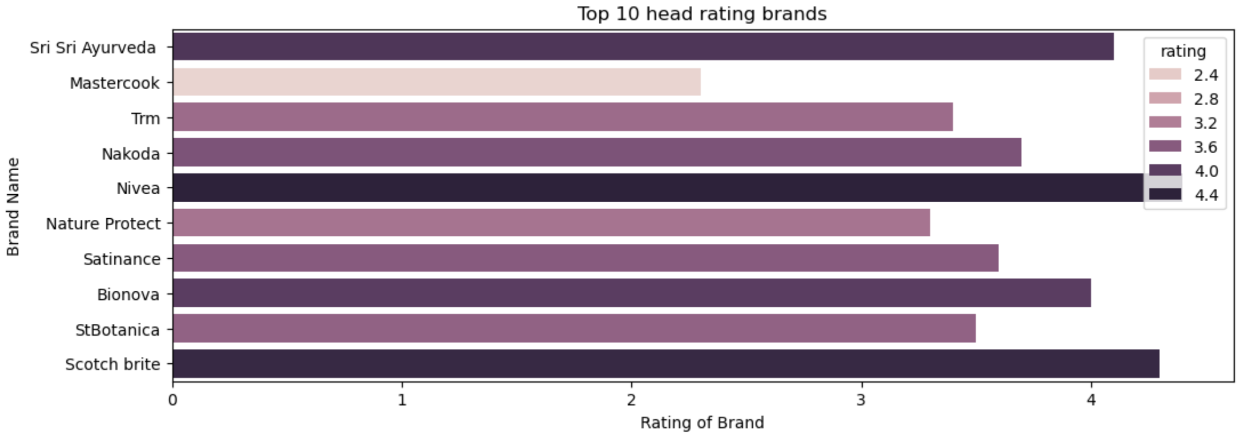


```
# Top 10 products with highest discount
temp3 = (df[['product',
'discount_percent']].sort_values('discount_percent',

ascending=False).head(10)).reset_index()
plt.figure(figsize=(12,4))
sns.barplot(data=temp3,x='discount_percent',y='product',palette='RdYlG
n_r',hue='discount_percent',estimator='sum')
plt.title('Top 10 Products by Discount Percentage')
plt.xlabel('Discount Percentage (%)')
plt.ylabel('Product')
plt.show()
temp3
```

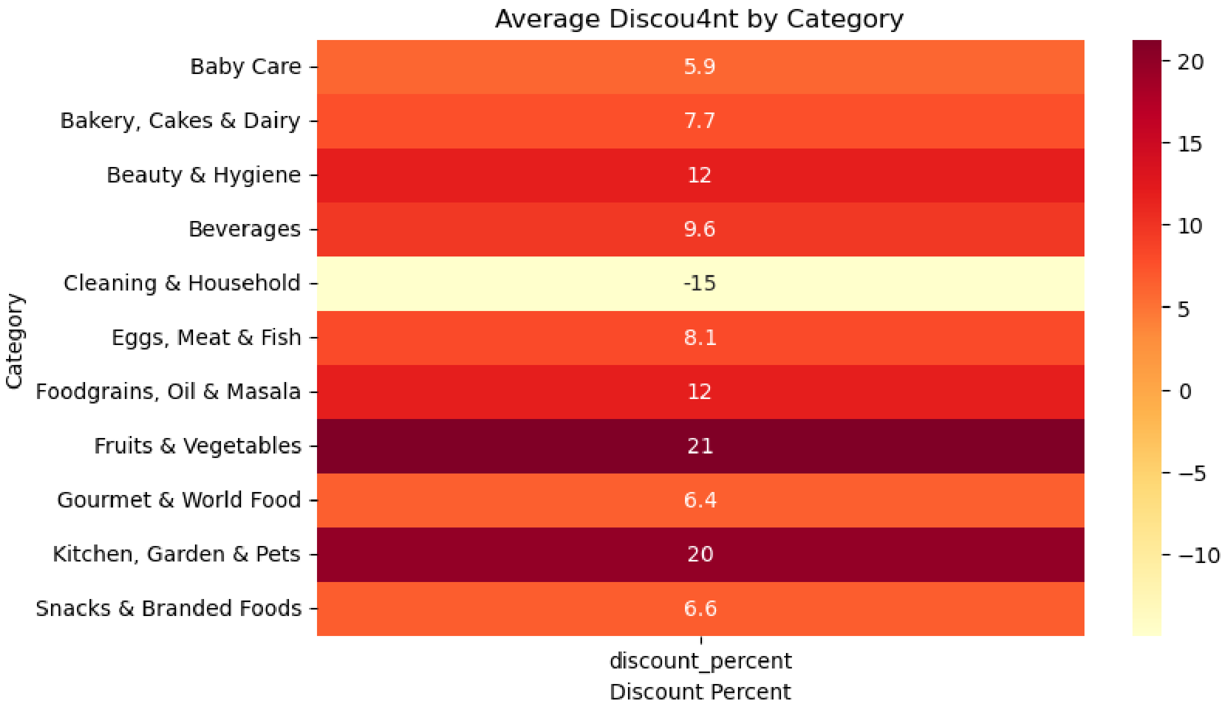


```
   index                                              product  
discount_percent
0  26976                                         Curry Leaves  
83.666667
1  17713                       Fruit & Vegetables Hand Juicer  
82.506266
2  13318  Small Silicone Spatula With Plastic Handle - A...  
```

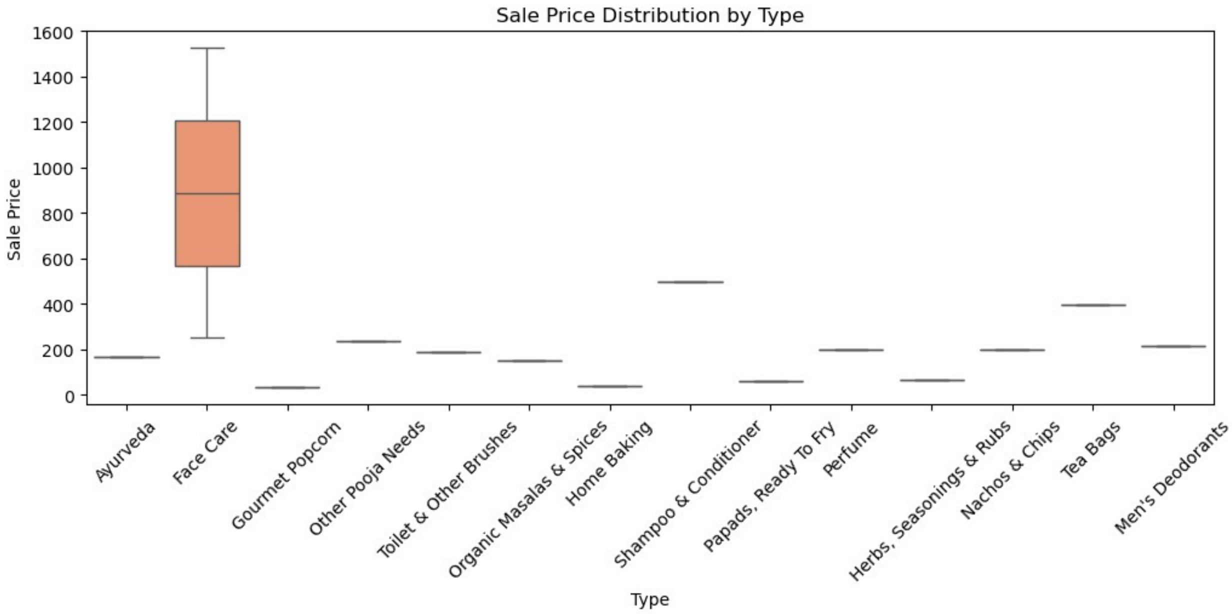
```
81.203008
3  13740  Decorative Party Light Big Star String LED Lig...
80.982712
4  10438  NHS 860 Temperature Control Professional Hair ...
80.499791
5   4562                              Concealer Brush 930
80.000000
6  11473  Decorative Party Light Golden Bell String LED ...
79.239620
7  13265  Decorative Party Light Golden Bell String LED ...
79.239620
8  10092  USB String Fairy Lights 3M 30 LED For Decorati...
78.696742
9    398  Steel Belly Shape Storage Dabba/ Container Set...
77.989950

# top 10 rating brand of head 10
plt.figure(figsize=(12,4))
sns.barplot(data=df.head(10),x='rating',y='brand',hue='rating')
plt.xlabel('Rating of Brand')
plt.ylabel('Brand Name')
plt.title('Top 10 head rating brands')
plt.show()
```



```
# Average Discount by Category representing on heatmap
temp4 = df.groupby('category')['discount_percent'].mean().to_frame()
plt.figure(figsize=(8,5))
sns.heatmap(temp4,annot=True,cmap='YlOrRd')
plt.xlabel('Discount Percent')
plt.ylabel('Category')
plt.title('Average Discou4nt by Category')
plt.show()
```



```
# Type vs Sale Price
plt.figure(figsize=(12,4))
sns.boxplot(data=df.tail(15),x='type',y='sale_price',palette='Set2')
plt.xticks(rotation=45)
plt.title("Sale Price Distribution by Type")
plt.xlabel("Type")
plt.ylabel("Sale Price")
plt.show()
```



EDA Insights

- The dataset includes **product**, **category**, **brand**, **pricing**, **discount**, and **rating** information, with **missing values handled** using logical defaults to ensure **data quality**.
- **Product distribution** is highly **skewed**, with a few **dominant categories** contributing most of the products, indicating a **demand-driven inventory strategy**.
- **Category-wise analysis** shows **uneven growth**, suggesting **selective expansion** based on **customer demand**.
- **Discount percentages** vary significantly across **categories**, reflecting **targeted** and **competitive pricing strategies**.
- A **small number of products** receive **very high discounts**, likely for **promotions**, **clearance**, or **traffic generation**.
- **Brand-wise analysis** reveals that **high product availability** does not always correspond to **higher customer ratings**.
- A **large number of products** have **zero ratings**, indicating **newly launched items** or **low customer engagement**.
- **Outliers** observed in **pricing** and **discounts** represent **premium** or **bulk products** and are **meaningful business cases**.
- **Heatmap analysis** highlights clear differences in **discount strategies** across **categories**.
- Overall, **BigBasket** focuses on **category dominance** and **strategic discounting**, with opportunities to improve **customer engagement** and **brand trust**.

Business Conclusion

BigBasket follows a **demand-driven product strategy**, concentrating on **high-performing categories** while maintaining selective expansion in niche segments.

The platform uses **strategic discounting** to stay competitive, applying higher discounts only where price sensitivity is high.

While **brand availability** is strong, **customer ratings** show that visibility alone does not ensure satisfaction.

The presence of many **unrated products** highlights an opportunity to improve **customer engagement and trust** through reviews.

Overall, BigBasket demonstrates a **data-backed retail strategy** with clear scope for enhancing **customer experience and loyalty**.