

bigbasket-project-harsh-sharma-1

March 15, 2025

#

BIGBASKET PROJECT

```
[113]: from IPython.display import Image  
Image('download.png')
```

[113]:

SUCCESS STORY IN FOCUS!



1 STEPS TO FOLLOW

Step 1: Load DataSet.

Step 2: Use head function to look for first 12 rows.

Step 3: Get Description of the data in the DataFrame.

Step 4: Find Information about the DataFrame.

Step 5: Find out Top & least sold products.

Step 6: Measuring discount on a certain item.

Step 7: Find out the Missing Values from the Dataset.

Step 8: Find out the outliers from the dataset according to the columns and fill them with the mean.

Step 9: Create Plots or visualizations.

2 Importing liabraries

```
[55]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

3 Step 1: Load DataSet.

```
[56]: df=pd.read_csv('/content/BigBasket Products.csv')
```

```
[57]: df
```

```
[57]:
```

	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]

4 Step 2: Use head function to look for first 12 rows.

```
[58]: df["product"].value_counts().head(12)
```

```
[58]: product
Turmeric Powder/Arisina Pudi      26
Extra Virgin Olive Oil            15
Cow Ghee/Tuppa                   14
Soft Drink                       12
Colorsilk Hair Colour With Keratin 12
Olive Oil - Extra Virgin          12
Coriander Powder                 11
Powder - Coriander               11
Ghee/Tuppa                      11
Peanut Butter - Creamy           10
Casting Creme Gloss Hair Color   10
Hand Sanitizer                   10
Name: count, dtype: int64
```

5 Step 3: Get Description of the data in the DataFrame.

```
[59]: df.describe()
```

```
[59]:
```

	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

6 Step 4: Find Information about the DataFrame.

```
[60]: 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

```

7 Step 5: Find out Top & least sold products.

8 Top Five selling product

```
[61]: df["product"].value_counts().head(5)
```

```

[61]: product
Turmeric Powder/Arisina Pudi      26
Extra Virgin Olive Oil            15
Cow Ghee/Tuppa                   14
Soft Drink                       12
Colorsilk Hair Colour With Keratin 12
Name: count, dtype: int64

```

9 Least Five selling product

```
[62]: df["product"].value_counts().tail(5)
```

```

[62]: product
Pepper & Herb Salami Chicken      1
Nutmeg Powder                    1
Disney Mickey Mouse Plastic Kids Sipper Bottle - Coolest Dude 1
Chocolates-Roasted Peanut Chocolate 1
Green Tea - Pure Original         1
Name: count, dtype: int64

```

10 Step 6: Measuring discount on a certain item.

```
[63]: df["Discount_percentage"] = (((df["market_price"] - df["sale_price"]) /
↳ df["market_price"]) * 100).round(2)
```

```
[64]: df
```

```
[64]:
```

	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	Creame 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_percentage
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...	52.40
3	Multipurpose container with an attractive desi...	15.34
4	Nivea Creme Soft Soap gives your skin the best...	0.00
...
27550	Layerr brings you Wottagirl Classic fragrant b...	20.00
27551	Puramate rosemary is enough to transform a dis...	10.00
27552	We have taken the richness of Sweet Potatoes (...)	0.00
27553	Tetley Green Tea with its refreshing pure, ori...	20.00
27554	The new mens fragrance from the United Dreams ...	44.99

[27555 rows x 11 columns]

11 Step 7: Find out the Missing Values from the Dataset.

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

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

```
[66]: df[df["product"].isnull()]
```

```
[66]:      index product      category sub_category      brand sale_price \
14363  14364      NaN Beverages      Coffee Cothas Coffee      200.0

      market_price      type rating \
14363      240.0 Ground Coffee      4.2

      description Discount_percentage
14363  Cothas Specialty Blend Coffee and Chicory incl...      16.67
```

```
[67]: df[df["brand"].isnull()]
```

```
[67]:      index      product      category \
9765   9766 Food Package - Medium Cleaning & Household

      sub_category brand sale_price market_price \
9765  Disposables, Garbage Bag  NaN      50.0      50.0

      type rating description Discount_percentage
9765  Aluminium Foil, Clingwrap  NaN      NaN      0.0
```

```
[68]: df[df["sale_price"].isnull()]
```

```
[68]:      index      product \
1719   1720      Puja Flower Wicks - Puvvu Vathulu Batti
1720   1721      Powder - Sambar
1721   1722 Steel Fork - Medium, Premium Excel Series, BBCL08
1722   1723      Snack Mix - Dhokla
2401   2402      Battery AA 3UT Hi Top
2402   2403      Klassic Plain Cocktail Napkins (22 X 22 cm)

      category      sub_category      brand \
1719  Cleaning & Household      Pooja Needs      Jaji
1720  Foodgrains, Oil & Masala      Masalas & Spices  Annapoorna
1721  Kitchen, Garden & Pets      Crockery & Cutlery  BB Home
1722  Snacks & Branded Foods      Ready To Cook & Eat  MTR
2401  Kitchen, Garden & Pets  Appliances & Electricals  Nippo
2402  Cleaning & Household  Disposables, Garbage Bag  Origami

      sale_price market_price      type rating \
1719      NaN      20.0      Camphor & Wicks  NaN
1720      NaN      46.0      Blended Masalas  NaN
1721      NaN     299.0  Cutlery, Spoon & Fork  NaN
1722      NaN      65.0  Breakfast & Snack Mixes  4.1
2401      NaN     150.0  Battery & Electrical  4.0
2402      NaN      32.0      Toilet Paper  3.9

      description Discount_percentage
```


1719	Jaji Puja Flower Batti / Puvvu Vathulu is made...	NaN
1720	Annaporna Spices are the preserve for Connoiss...	NaN
1721	BB Home provides fine and classy cutlery that ...	NaN
1722	MTR Dhokla is extremely special in Gujarat. It...	NaN
2401	This Nippo battery incorporates state-of-the-a...	NaN
2402	Klassic Plain Cocktail Napkins (22 X 22 cm) 10...	NaN

```
[69]: df[df["rating"].isnull()]
```

```
[69]:
```

	index	product \	category	sub_category	brand \
55	56	Soothing Cucumber Facial Scrub With Apricot Seeds	Beauty & Hygiene	Skin Care	TJORI
59	60	Corporate Planner Diary With Premium PU Leathe...	Cleaning & Household	Stationery	Prozo Plus
65	66	Ayurvedic Anti-Tan Face Pack	Beauty & Hygiene	Skin Care	TJORI
68	69	Organic Carom Seeds/Ajwain/Om Kalu	Foodgrains, Oil & Masala	Masalas & Spices	Earthon
69	70	Padded Harness - 3/4 inch, Grey Colour	Kitchen, Garden & Pets	Pet Food & Accessories	Glenand
...
27509	27510	Deluxe Crackers - Veg	Gourmet & World Food	Chocolates & Biscuits	Kerk
27511	27512	Specialist Stain Remover Pen & Marker	Cleaning & Household	All Purpose Cleaners	365
27514	27515	Verge & Sheer Perfume For Pair	Beauty & Hygiene	Fragrances & Deos	Skinn by Titan
27530	27531	Tick'et to Fleadom Dry Shampoo For Dogs	Kitchen, Garden & Pets	Pet Food & Accessories	Captain Zack
27531	27532	Vanilla Ice Cream	Gourmet & World Food	Dairy & Cheese	Haagen-Dazs

	sale_price	market_price	type	rating \
55	299.4	499.0	Face Care	NaN
59	399.0	399.0	Notebooks, Files, Folders	NaN
65	269.4	449.0	Face Care	NaN
68	72.0	72.0	Whole Spices	NaN
69	840.0	840.0	Pet Collars & Leashes	NaN
...
27509	150.0	150.0	Cookies, Biscotti, Wafer	NaN
27511	449.0	449.0	Imported Cleaners	NaN
27514	1615.5	1795.0	Perfume	NaN
27530	99.0	99.0	Pet Cleaning & Grooming	NaN
27531	225.0	225.0	Gourmet Ice Cream	NaN

	description	Discount_percentage
55	The soothing feel of cucumber meets the gentle...	40.0
59	A5 Size (210x150mm) \n192 Pages Premium Natura...	0.0
65	A nourishing face pack that removes tan and br...	40.0
68	Earthon's Ajwain is Best quality, organically ...	0.0
69	These are soft padded harness for your active ...	0.0
...
27509	Kerk Biscuits-has been a household name synony...	0.0
27511	Mightier than the pen. The pen may be mightier...	0.0
27514	VERGE for men paints a picture of a classy out...	10.0
27530	1) No Rinse Defence Against Ticks and Fleas: C...	0.0
27531	Some things are so good in their simplest form...	0.0

[8636 rows x 11 columns]

```
[70]: df[df["description"].isnull()]
```

```
[70]:
```

	index	product \
288	289	Arrabbiata Tomato Pasta Sauce With Chilli
480	481	Pumice Stone & Emery Foot Filer - FFL3407, Col...
780	781	Peanut Butter - Chunky
845	846	Squash - Sarasaparilla/Nannari/Sugandhi Root
1047	1048	Natura Rice & Walnut Drink, Vegan
...
25510	25511	BBPopular Almond/Badam Californian 500gm + BBR...
25583	25584	Chai/Chai/Tea/ Coffee Bone China Microwave Ser...
26116	26117	Tea Strainer - Stainless Steel, Medium
26835	26836	Cold Pressed Virgin Groundnut Oil
27303	27304	Zipper Storage - Gallon

	category	sub_category	brand \
288	Gourmet & World Food	Sauces, Spreads & Dips	Montanini
480	Beauty & Hygiene	Bath & Hand Wash	Kaiv
780	Gourmet & World Food	Sauces, Spreads & Dips	American Garden
845	Beverages	Fruit Juices & Drinks	NaturoBell
1047	Gourmet & World Food	Drinks & Beverages	BORGES
...
25510	Foodgrains, Oil & Masala	Dry Fruits	bb Combo
25583	Kitchen, Garden & Pets	Crockery & Cutlery	BP Bharat
26116	Kitchen, Garden & Pets	Kitchen Accessories	Ritu
26835	Foodgrains, Oil & Masala	Edible Oils & Ghee	Chekkko
27303	Cleaning & Household	Disposables, Garbage Bag	Glad

	sale_price	market_price	type	rating \
288	22325.00	325.0	Mustard & Cheese Sauces	5.0
480	175.00	175.0	Bathing Accessories	3.8

780	299.00	299.0	Chocolate, Peanut Spread	4.1
845	140.00	140.0	Syrups & Concentrates	2.5
1047	275.01	349.0	Health Drinks	4.5
...
25510	443.00	725.0	Almonds	NaN
25583	269.00	373.0	Cups, Mugs & Tumblers	2.5
26116	99.00	135.0	Strainer, Ladle, Spatula	4.3
26835	325.00	325.0	Groundnut Oil	4.2
27303	350.00	350.0	Aluminium Foil, Clingwrap	4.7

	description	Discount_percentage
288	NaN	-6769.23
480	NaN	0.00
780	NaN	0.00
845	NaN	0.00
1047	NaN	21.20
...
25510	NaN	38.90
25583	NaN	27.88
26116	NaN	26.67
26835	NaN	0.00
27303	NaN	0.00

[115 rows x 11 columns]

```
[71]: df[df["Discount_percentage"].isnull()]
```

```
[71]:
```

	index	product \
1719	1720	Puja Flower Wicks - Puvvu Vathulu Batti
1720	1721	Powder - Sambar
1721	1722	Steel Fork - Medium, Premium Excel Series, BBCL08
1722	1723	Snack Mix - Dhokla
2401	2402	Battery AA 3UT Hi Top
2402	2403	Klassic Plain Cocktail Napkins (22 X 22 cm)

	category	sub_category	brand \
1719	Cleaning & Household	Pooja Needs	Jaji
1720	Foodgrains, Oil & Masala	Masalas & Spices	Annapoorna
1721	Kitchen, Garden & Pets	Crockery & Cutlery	BB Home
1722	Snacks & Branded Foods	Ready To Cook & Eat	MTR
2401	Kitchen, Garden & Pets	Appliances & Electricals	Nippo
2402	Cleaning & Household	Disposables, Garbage Bag	Origami

	sale_price	market_price	type	rating \
1719	NaN	20.0	Camphor & Wicks	NaN
1720	NaN	46.0	Blended Masalas	NaN
1721	NaN	299.0	Cutlery, Spoon & Fork	NaN

1722	NaN	65.0	Breakfast & Snack Mixes	4.1
2401	NaN	150.0	Battery & Electrical	4.0
2402	NaN	32.0	Toilet Paper	3.9

		description	Discount_percentage
1719	Jaji Puja Flower Batti / Puvvu Vathulu is made...		NaN
1720	Annaporna Spices are the preserve for Connoiss...		NaN
1721	BB Home provides fine and classy cutlery that ...		NaN
1722	MTR Dhokla is extremely special in Gujarat. It...		NaN
2401	This Nippo battery incorporates state-of-the-a...		NaN
2402	Klassic Plain Cocktail Napkins (22 X 22 cm) 10...		NaN

12 Cleaning the Missing Values

```
[72]: df["product"]=df["product"].fillna(df["product"].mode()[0])
df["product"]
```

```
[72]: 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
Name: product, Length: 27555, dtype: object
```

```
[73]: df["sale_price"]=df["sale_price"].fillna(df["sale_price"].median())
df["sale_price"]
```

```
[73]: 0          220.00
1          180.00
2          119.00
3          149.00
4          162.00
...
27550      199.20
27551        67.50
27552      200.00
27553      396.00
27554      214.53
Name: sale_price, Length: 27555, dtype: float64
```

```
[74]: df["brand"]=df["brand"].fillna(df["brand"].mode()[0])
df["brand"]
```

```
[74]: 0          Sri Sri Ayurveda
1          Mastercook
2          Trm
3          Nakoda
4          Nivea
...
27550         Layerr
27551        Puramate
27552        FabBox
27553        Tetley
27554  United Colors Of Benetton
Name: brand, Length: 27555, dtype: object
```

```
[75]: df["rating"]=df["rating"].fillna(df["rating"].median())
df["rating"]
```

```
[75]: 0          4.1
1          2.3
2          3.4
3          3.7
4          4.4
...
27550        3.9
27551        4.0
27552        3.8
27553        4.2
27554        4.5
Name: rating, Length: 27555, dtype: float64
```

```
[76]: df["description"]=df["description"].fillna(df["description"].mode()[0])
df["description"]
```

```
[76]: 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 ...
Name: description, Length: 27555, dtype: object
```

```
[77]: df["Discount_percentage"]=df["Discount_percentage"].
      ↪fillna(df["Discount_percentage"].median())
      df["Discount_percentage"]
```

```
[77]: 0          0.00
      1          0.00
      2         52.40
      3         15.34
      4          0.00
      ...
      27550       20.00
      27551       10.00
      27552        0.00
      27553       20.00
      27554       44.99
      Name: Discount_percentage, Length: 27555, dtype: float64
```

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

```
[78]: index          0
      product        0
      category        0
      sub_category    0
      brand           0
      sale_price      0
      market_price    0
      type            0
      rating          0
      description     0
      Discount_percentage  0
      dtype: int64
```

13 Step 8: Find out the outliers from the dataset according to the columns

```
[79]: numeric=df.select_dtypes(include=["int","float"])
      numeric
```

```
[79]:
```

	index	sale_price	market_price	rating	Discount_percentage
0	1	220.00	220.0	4.1	0.00
1	2	180.00	180.0	2.3	0.00
2	3	119.00	250.0	3.4	52.40
3	4	149.00	176.0	3.7	15.34
4	5	162.00	162.0	4.4	0.00
...
27550	27551	199.20	249.0	3.9	20.00

27551	27552	67.50	75.0	4.0	10.00
27552	27553	200.00	200.0	3.8	0.00
27553	27554	396.00	495.0	4.2	20.00
27554	27555	214.53	390.0	4.5	44.99

[27555 rows x 5 columns]

```
[80]: def select_outliers(data):
        q1=data.quantile(.25)
        q3=data.quantile(.75)
        iqr=q3-q1
        upper=q3+1.5*iqr
        lower=q1-1.5*iqr
        outliers=data[(data<lower)|(data>upper)]
        return outliers.count()
```

```
[81]: numeric_colm=[          "sale_price"          ,"market_price"          ,"rating"          ,"Discount_p
        outliers_count={col:select_outliers(df[col])for col in numeric_colm}
        outliers_count
```

```
[81]: {'sale_price': 2205,
        'market_price': 2147,
        'rating': 6372,
        'Discount_percentage': 438}
```

CLEANING DATA AND REPLACE THE OUTLIERS WITH HELP OF MEAN

```
[82]: def replace_outliers(df,colm):
        q1=df[colm].quantile(.25)
        q3=df[colm].quantile(.75)
        iqr=q3-q1
        upper=q3+1.5*iqr
        lower=q1-1.5*iqr
        df.loc[(df[colm]<lower)|(df[colm]>upper),colm]=df[colm].mean()
        return df
```

```
[83]: df1 = replace_outliers(df, "sale_price")
        df2 = replace_outliers(df, "market_price")
        df3 = replace_outliers(df, "rating")
        df4 = replace_outliers(df, "Discount_percentage")
        df1
        df2
        df3
        df4
```

```
[83]:          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

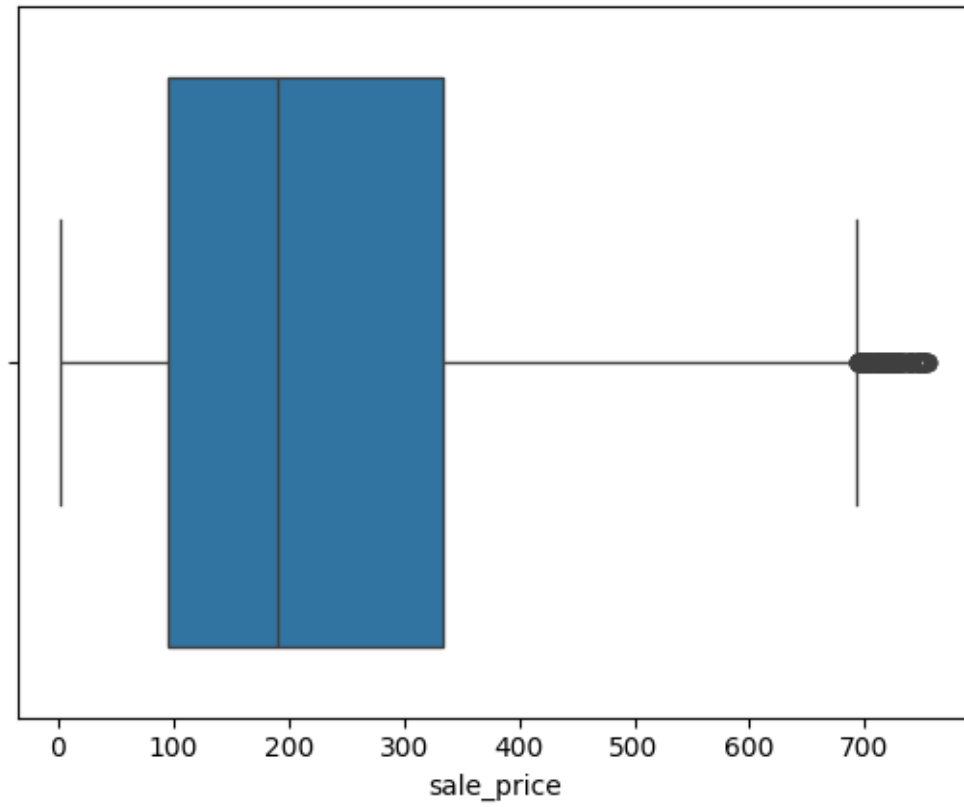
	type	rating \
0	Hair Oil & Serum	4.100000
1	Water & Fridge Bottles	3.992408
2	Lamp & Lamp Oil	3.992408
3	Laundry, Storage Baskets	3.700000
4	Bathing Bars & Soaps	4.400000
...
27550	Perfume	3.900000
27551	Herbs, Seasonings & Rubs	4.000000
27552	Nachos & Chips	3.800000

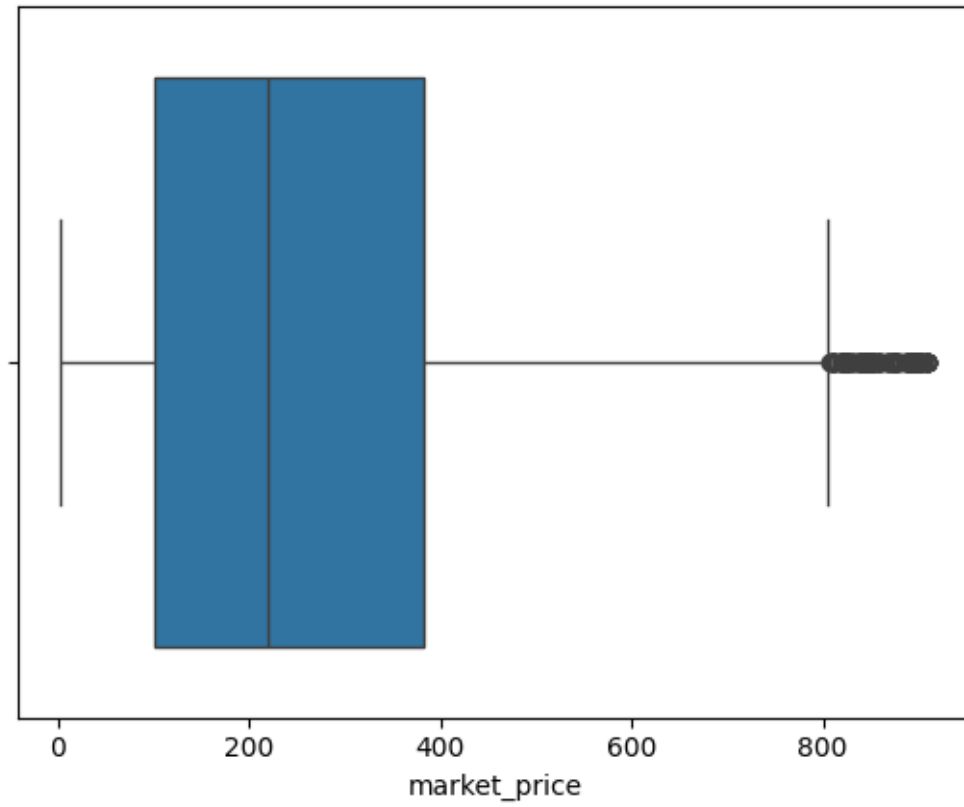
27553	Tea Bags	4.200000
27554	Men's Deodorants	4.500000

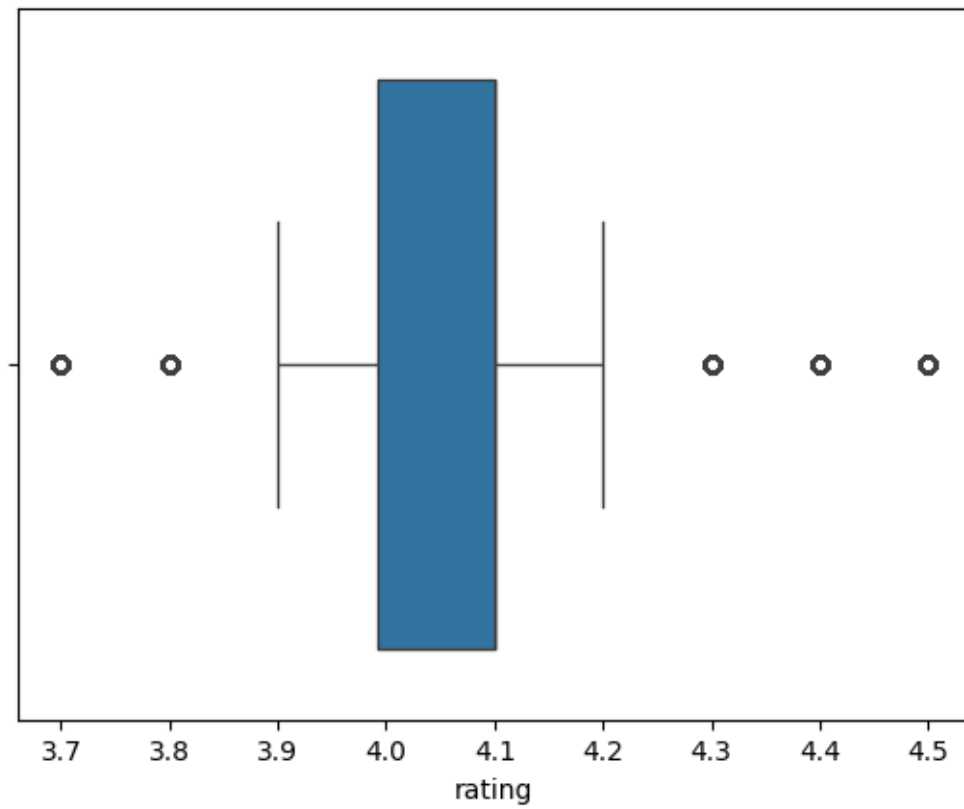
	description	Discount_percentage
0	This Product contains Garlic Oil that is known...	0.000000
1	Each product is microwave safe (without lid), ...	0.000000
2	A perfect gift for all occasions, be it your m...	8.620643
3	Multipurpose container with an attractive desi...	15.340000
4	Nivea Creme Soft Soap 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 ...	44.990000

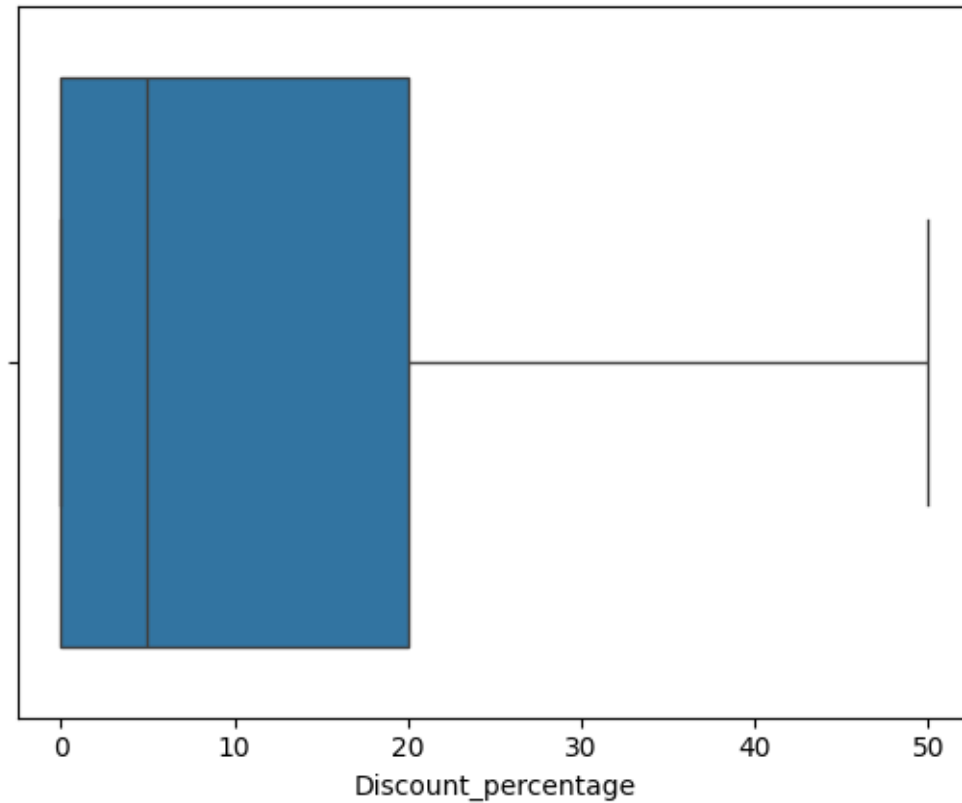
[27555 rows x 11 columns]

```
[84]: sns.boxplot(data=df1,x="sale_price")
plt.show()
sns.boxplot(data=df2,x="market_price")
plt.show()
sns.boxplot(data=df3,x="rating")
plt.show()
sns.boxplot(data=df4,x="Discount_percentage")
plt.show()
```









```
[85]: numeric_colm=[          "sale_price"          ,"market_price"          ,"rating"          ,"Discount_percentage"]
      outliers_count={col:select_outliers(df[col])for col in numeric_colm}
      outliers_count
```

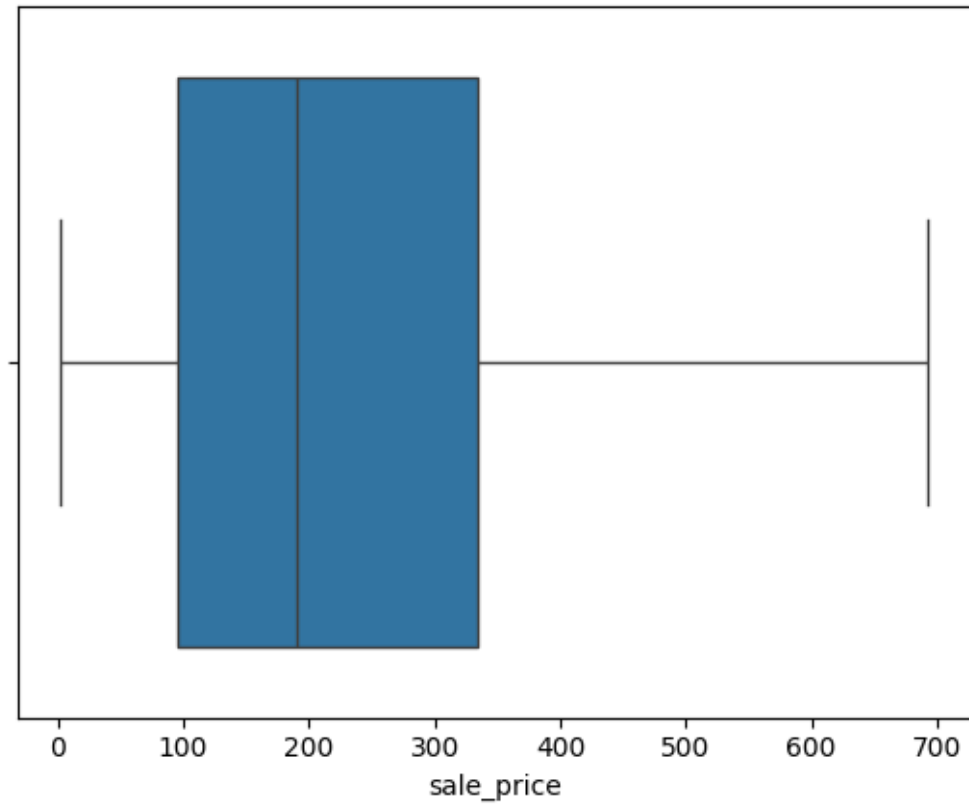
```
[85]: {'sale_price': 365,
      'market_price': 374,
      'rating': 5510,
      'Discount_percentage': 0}
```

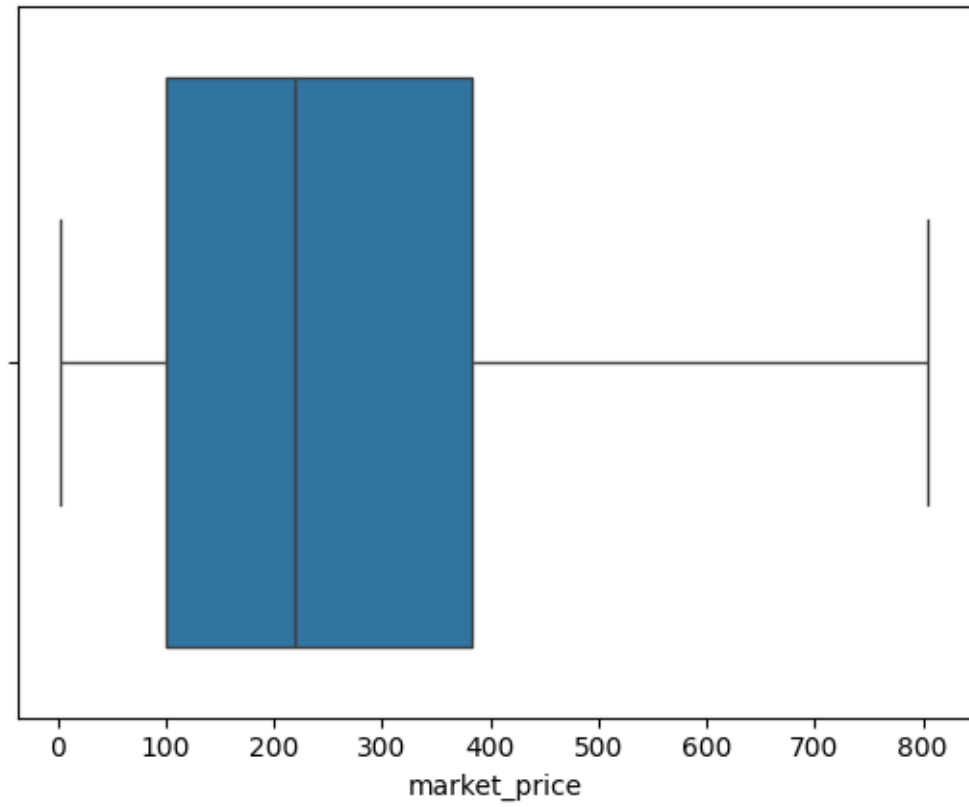
```
[86]: df1 = replace_outliers(df, "sale_price")
      df2 = replace_outliers(df, "market_price")
      df3 = replace_outliers(df, "rating")
      df4 = replace_outliers(df, "Discount_percentage")
```

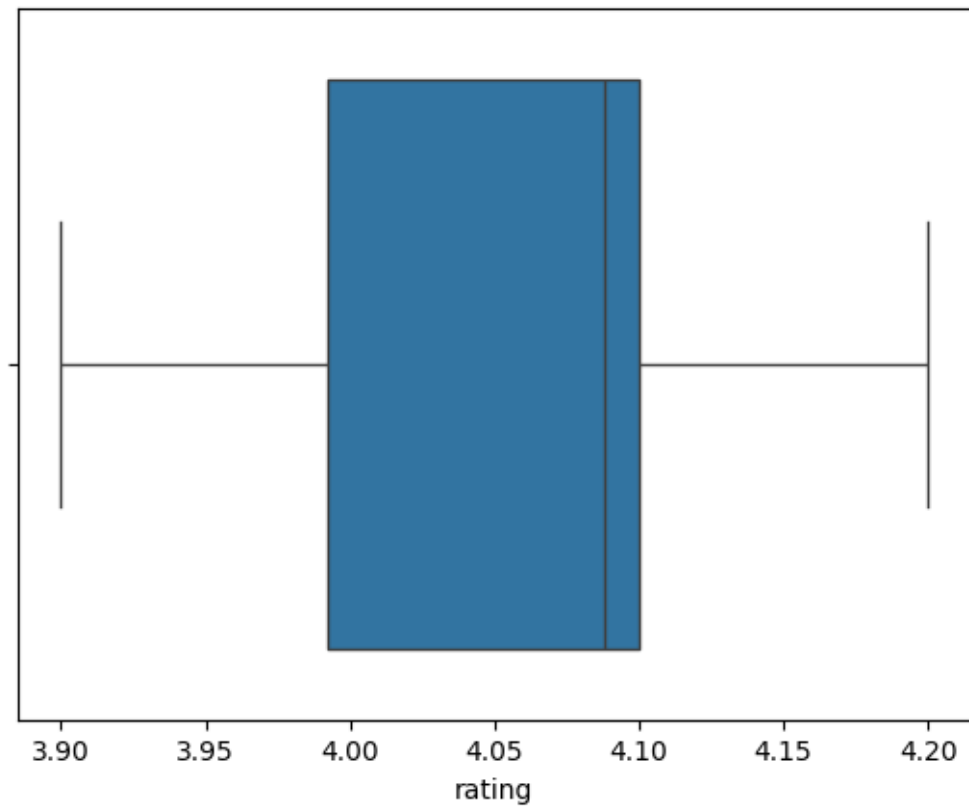
```
[87]: numeric_colm=[          "sale_price"          ,"market_price"          ,"rating"          ,"Discount_percentage"]
      outliers_count={col:select_outliers(df[col])for col in numeric_colm}
      outliers_count
```

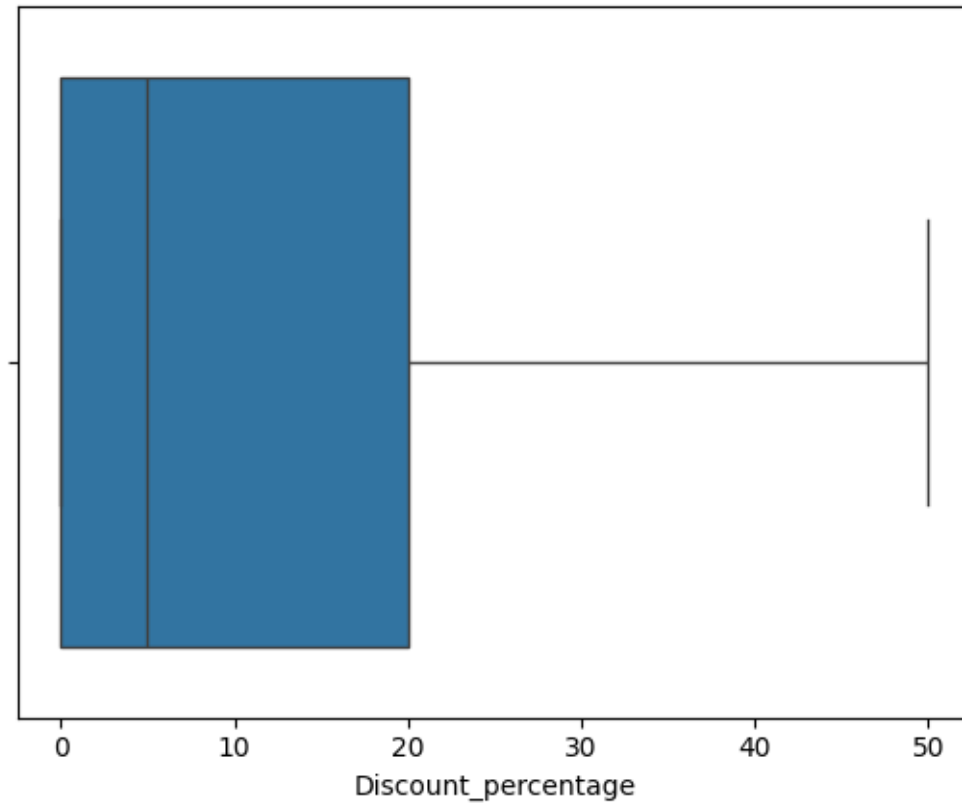
```
[87]: {'sale_price': 0, 'market_price': 0, 'rating': 0, 'Discount_percentage': 0}
```

```
[88]: sns.boxplot(data=df1,x="sale_price")  
plt.show()  
sns.boxplot(data=df2,x="market_price")  
plt.show()  
sns.boxplot(data=df3,x="rating")  
plt.show()  
sns.boxplot(data=df4,x="Discount_percentage")  
plt.show()
```









14 Step 9: Create Plots or visualizations.

```
[89]: df["category"].value_counts()
```

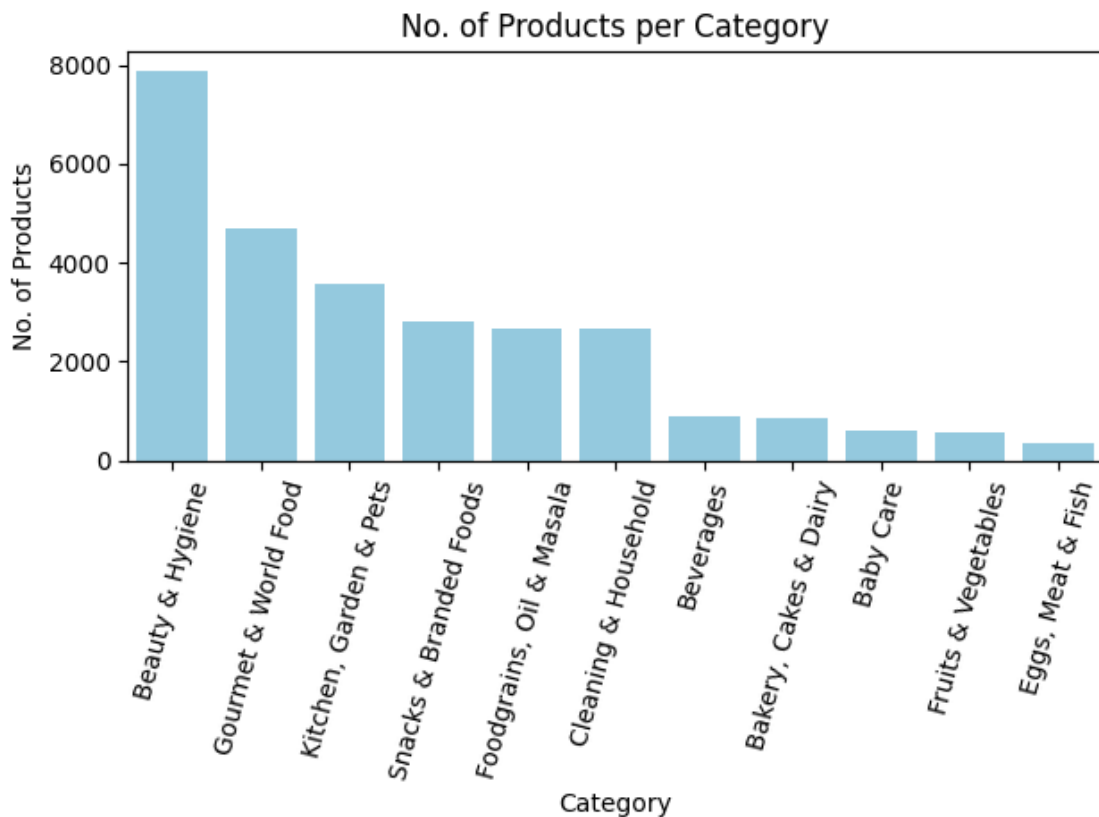
```
[89]: category
Beauty & Hygiene          7867
Gourmet & World Food      4690
Kitchen, Garden & Pets    3580
Snacks & Branded Foods     2814
Foodgrains, Oil & Masala   2676
Cleaning & Household       2675
Beverages                  885
Bakery, Cakes & Dairy       851
Baby Care                  610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

14.0.1 A bar plot showing the count of products within each category

```
[90]: # Data
data = {'category': ["Beauty & Hygiene", "Gourmet & World Food", "Kitchen, Garden & Pets",
                    "Snacks & Branded Foods", "Foodgrains, Oil & Masala",
                    "Cleaning & Household",
                    "Beverages", "Bakery, Cakes & Dairy", "Baby Care", "Fruits & Vegetables",
                    "Eggs, Meat & Fish"]}

df2 = pd.DataFrame(data)
counts = df['category'].value_counts()

sns.barplot(x=counts.index, y=counts.values, color='skyblue')
plt.xticks(rotation=75)
plt.xlabel('Category')
plt.ylabel('No. of Products')
plt.title('No. of Products per Category')
plt.tight_layout()
plt.show()
```

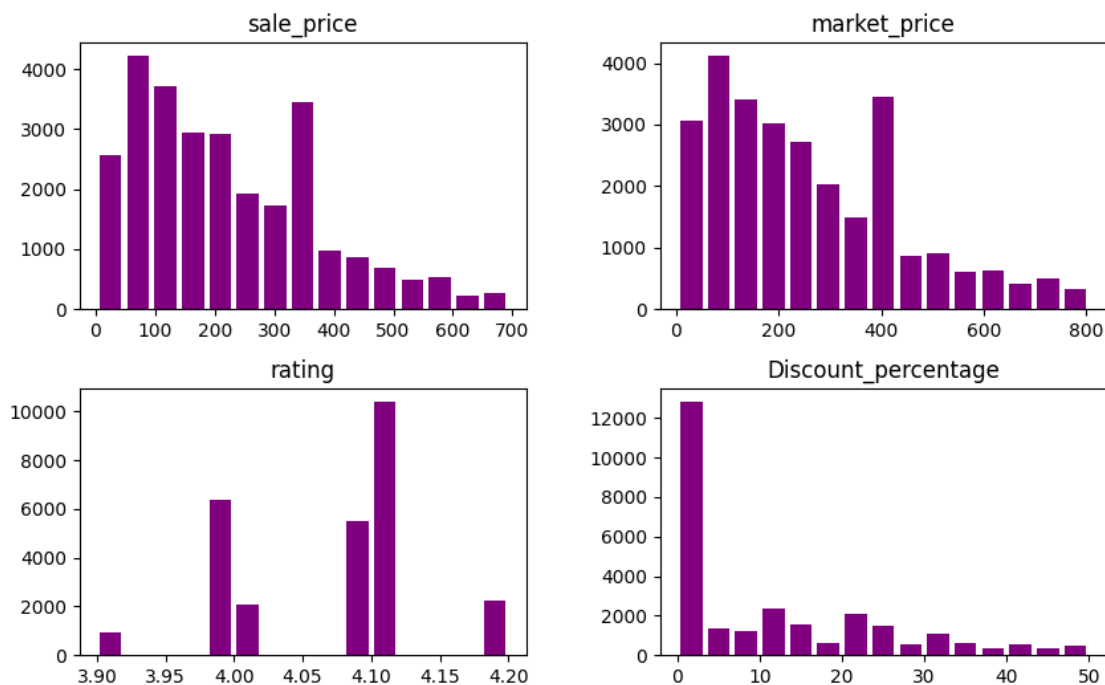


Key Takeaways: ##### • Beauty & Hygiene dominates with the highest number of products (5458). ##### • Kitchen, Garden & Pets follows closely with 2493 products. ##### • Snacks & Branded Foods and Gourmet & World Food have a similar product count, around 2400. ##### • Foodgrains, Oil & Masala and Cleaning & Household have a moderate number of products, just over 2000. ##### • Bakery, Cakes & Dairy has significantly fewer products (665). ##### • Beverages and Baby Care have the lowest product counts, around 600 and 500, respectively.

Recommendations: ##### • Focus on High-Performing Categories: Invest in marketing and promotions for Beauty & Hygiene, Kitchen, Garden & Pets, and Snacks & Branded Foods to drive sales. ##### • Optimize Low-Performing Categories: Analyze the product mix and customer demand for Bakery, Cakes & Dairy, Beverages, and Baby Care to identify opportunities for improvement. ##### • Consider Product Expansion: Explore opportunities to expand the product range in categories with lower product counts to attract a wider customer base.

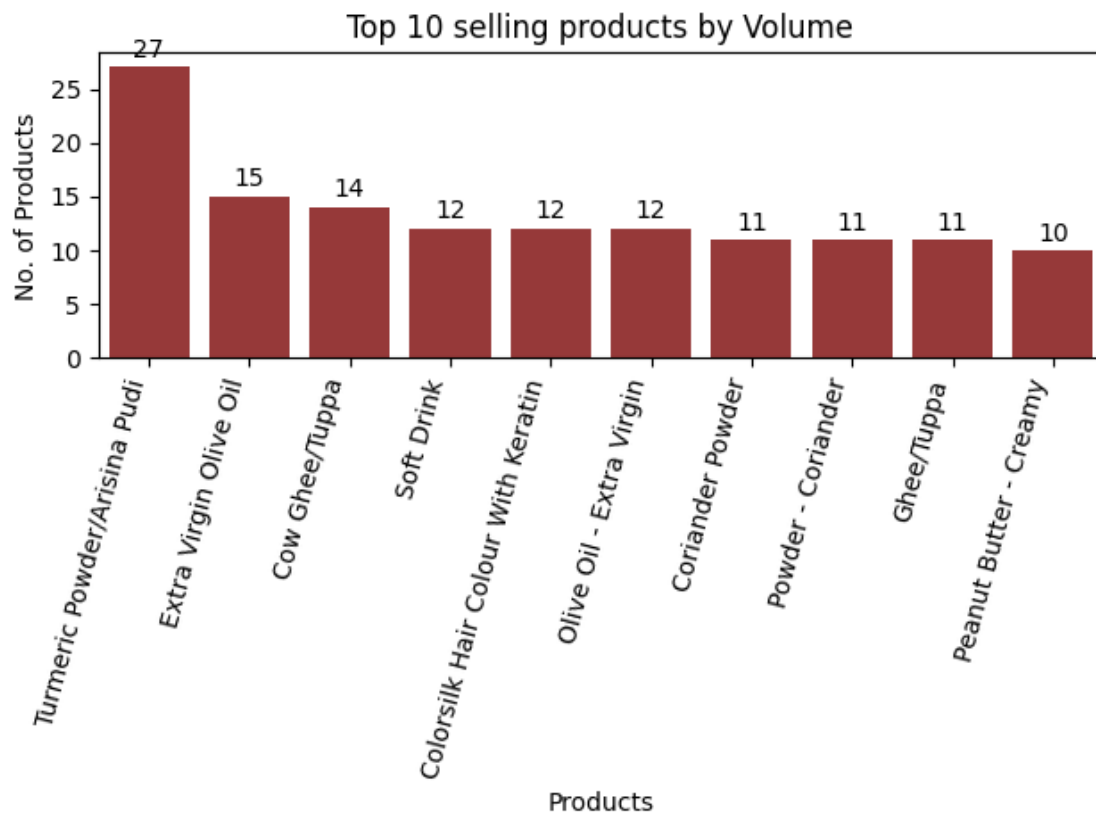
14.0.2 outliers removed from data

```
[91]: df[["sale_price", "market_price", "rating", "Discount_percentage"]].
      ↪ hist(bins=15, figsize=(10,6), grid=False, rwidth=0.8, color="purple")
      plt.show()
```



15 Top 10 Selling Products by Volume

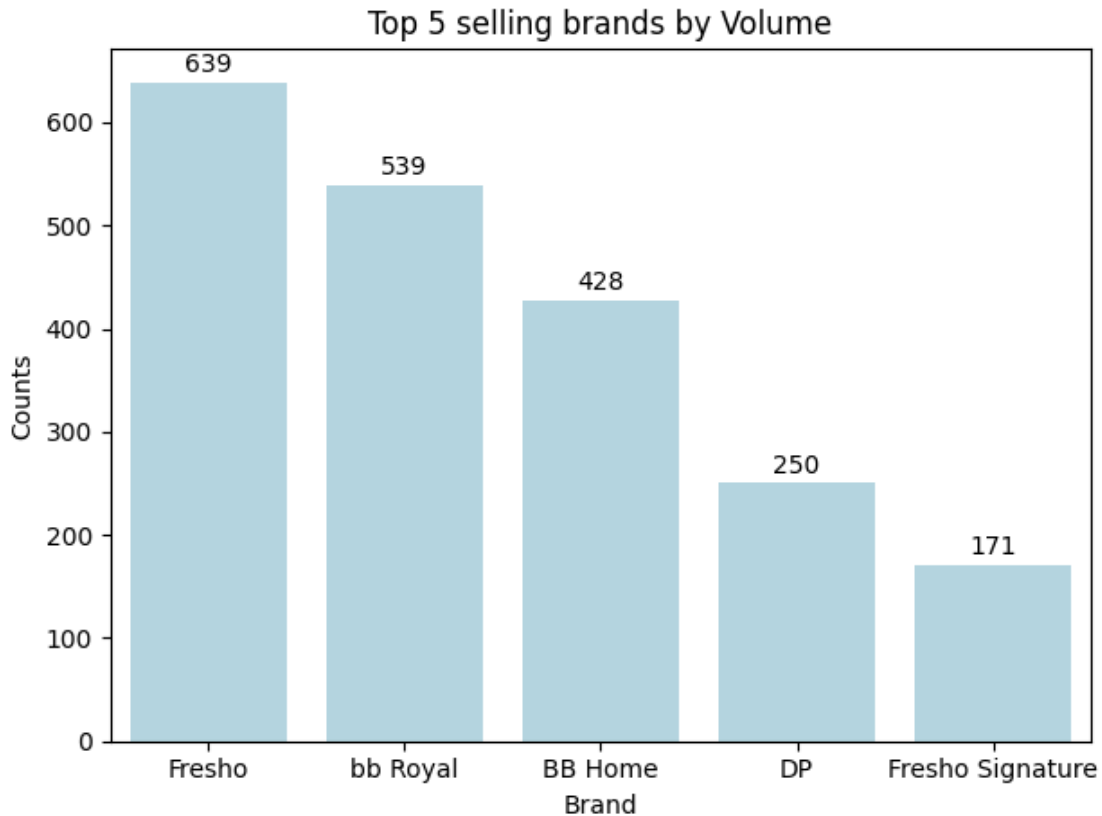
```
[92]: #This bar plot displays the top 10 products
      ↪
      counts = df['product'].value_counts()
      counts_p = pd.DataFrame({'product': counts.
      ↪index, 'Counts': counts.values})[:10]
      bx = sns.
      ↪barplot(x='product', y='Counts', data=counts_p, color='brown')
      plt.xticks(rotation=75, ha='right')
      bx.set_xlabel('Products')
      bx.set_ylabel('No. of Products')
      bx.set_title('Top 10 selling products by Volume')
      for p in bx.patches:
          bx.annotate(f'{int(p.get_height())}', (p.get_x()+p.get_width()/2., p.
      ↪get_height()),
      ↪ha='center', va='baseline', fontsize=10, color='black', xytext=
      textcoords='offset points')
      plt.tight_layout()
```



Key Takeaways: ##### .Turmeric Powder/Arisina Pudi is the top-selling product with a significantly higher volume (26 units) compared to others. ##### .Extra Virgin Olive Oil is the second-highest selling product with 15 units, followed closely by Cow Ghee/Tuppa with 14 units. ##### .Soft Drink, Olive Oil - Extra Virgin, and Colorsilk Hair Colour With Keratin share the same volume (12 units), showcasing moderate sales performance. ##### .Coriander Powder, Powder - Coriander, and Ghee/Tuppa each have 11 units sold, indicating steady demand. ##### .Casting Creme Gloss Hair Color rounds out the top 10 with 10 units sold, slightly lower than the rest. ## Recommendations: ##### .Leverage Top-Selling Products: ##### .Increase promotions and visibility for Turmeric Powder/Arisina Pudi and Extra Virgin Olive Oil to further capitalize on their popularity. ##### .Bundle these products with related items to boost cross-selling opportunities. ##### .Optimize Moderate-Performing Products: ##### .Analyze factors driving the performance of products like Soft Drink and Colorsilk Hair Colour With Keratin to replicate success in similar categories. ##### .Explore opportunities to enhance their visibility through targeted campaigns. ##### .Enhance Sales for Lower-Performing Products: ##### .Investigate the lower sales volume of Casting Creme Gloss Hair Color and develop strategies to increase its appeal, such as offering discounts or showcasing customer testimonials. ##### .Evaluate the product placement and pricing strategies to drive higher sales

16 Top 5 Selling Brands by Volume

```
[93]: #
↳ This bar plot illustrates the five brands
↳
counts=df['brand'].value_counts()
counts_b=pd.DataFrame({'brand': counts.index,'Counts':counts.values})[:5]
cx = sns.
↳ barplot(x='brand',y='Counts',data=counts_b, color='lightblue')
cx.set_xlabel('Brand')
cx.set_title('Top 5 selling brands by Volume')
for p in cx.patches:
    cx.annotate(f'{int(p.get_height())}',(p.get_x()+p.get_width()/2.,p.
↳ get_height()),
↳
↳ ha='center',va='baseline',fontsize=10,color='black',xytext=(0,4),textcoords='offset_
↳ points')
plt.tight_layout()
plt.show()
```

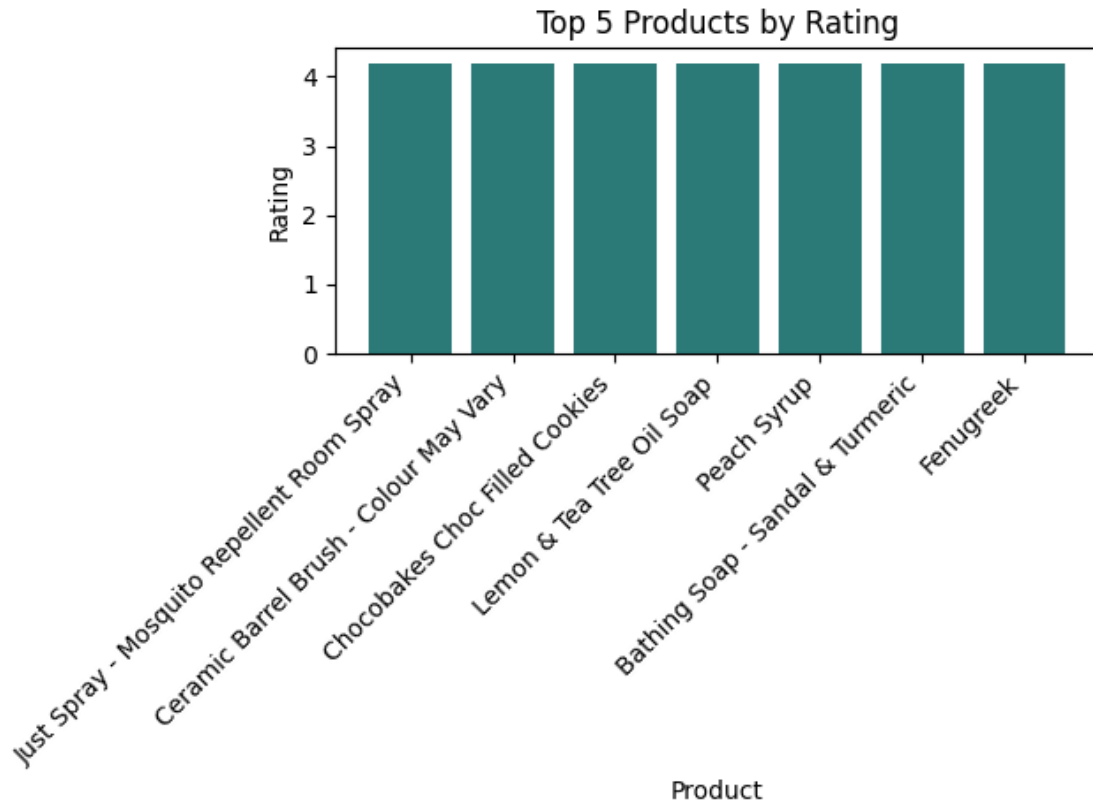


Key Insights: ##### Fresho is the top-selling brand with a volume of 638 units, significantly outperforming others. ##### bb Royal follows with 539 units, showing strong market presence. ##### BB Home, DP, and Fresho Signature occupy the remaining top 5 positions, with sales volumes of 428, 250, and 171, respectively. ##### The volume difference between the top-performing brands (Fresho and bb Royal) and the lower performers (DP and Fresho Signature) is notable, indicating possible brand preference disparities. ## Recommendations: ##### Focus on Fresho and bb Royal: Strengthen marketing strategies and maintain inventory for these high-demand brands. ##### Improve Mid-Tier Brands: Investigate the factors limiting the sales of DP and Fresho Signature, such as pricing, placement, or promotions. ##### Optimize Brand Performance: Perform a detailed analysis on BB Home to identify opportunities to increase its market share further. ##### Customer Insights: Conduct surveys or feedback sessions to understand customer preferences for the top-selling brands. ##### Expand Offerings: Consider introducing new products under the Fresho and bb Royal brands to capture additional market share.

17 Top 5 Products by Rating:

```
[94]: top_5_products = df.nlargest(7, 'rating')[['product', 'rating']]
plt.bar(top_5_products['product'], top_5_products['rating'], color='#2B7A78')
plt.xlabel('Product')
plt.ylabel('Rating')
```

```
plt.title('Top 5 Products by Rating')
plt.xticks(rotation=45,ha='right')
plt.tight_layout()
plt.show()
```

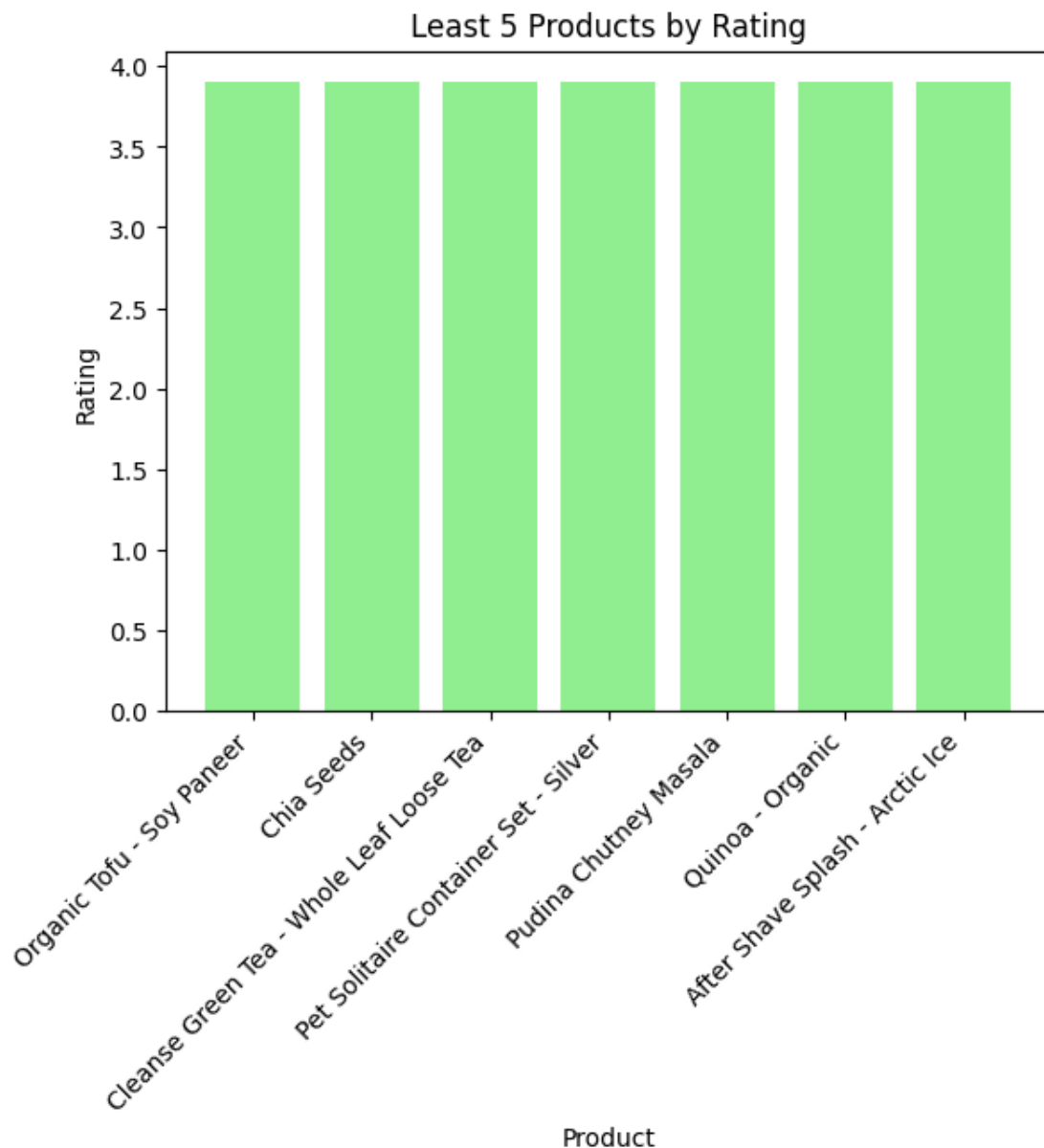


17.1 Key Insights:

Top 5 Products by Rating: #### All five products have achieved a high rating of 4+, indicating consistent customer satisfaction: ##### 1. Face Wash - Oil Control, Active ##### 2. Smooth Skin Oil - For Dry Skin ##### 3. Salted Pumpkin ##### 4. Veggie Cutter ##### 5. Plain Green Olives ### Recommendations: #### Maintain High Standards: ##### Continue to focus on quality and performance for these products to sustain high customer ratings. ##### Gather and showcase customer reviews for these top-rated products to build trust and attract new buyers. ### Expand Related Offerings: ##### Introduce variations or complementary products in the same categories (e.g., more types of skincare products or kitchen tools) to capture broader interest. ##### Use these high-rated products as anchors to promote bundles or combos. ### Leverage Marketing: ##### Highlight these products in promotional campaigns or feature them in “Top Rated” sections on websites or stores. ##### Offer limited-time discounts or loyalty rewards for these products to boost sales further

18 Least 5 Products by Rating

```
[95]: #A bar plot of the least-rated products, showing the top 5
      ↪products with the least ratings.
top_5_products=df.nsmallest(7,'rating')[['product','rating']]
plt.bar(top_5_products['product'],top_5_products['rating'],color='lightgreen')
plt.xlabel('Product')
plt.ylabel('Rating')
plt.title('Least 5 Products by Rating')
plt.xticks(rotation=45,ha='right')
plt.show()
```



18.1 Key Takeaways:

Water Bottle - Orange ##### Brass Angle Deep - Plain, No.2 ##### Germ - Removal Multipurpose Wipes ##### Multani Mati ##### Biotin & Collagen Volumizing Hair Shampoo + Biotin & Collagen Hair Conditioner ##### Butter Cookies Gold Collection ##### Insulated Hot Fresh Casserole For Roti/Chapati - White ##### All of these products are the lowest-rated, each with a rating of 1. ### Recommendations: ##### Review and Improve Low-Rated Products: ##### Analyze the feedback for these products to identify areas for improvement, such as quality, packaging, or product description. ### Consider Product Removal: ##### If the issues with low-rated products cannot be resolved, consider removing them from the inventory. ### Gather Customer Feedback: ##### Collect feedback from customers to identify areas for improvement and maintain high ratings. ## Overall Conclusion ##### Big Basket's strategy effectively balances high-frequency essentials, strategic discounting, and in-house brand development to build customer loyalty. ##### Improvements could include: ##### Addressing quality issues for lower-rated products. ##### Expanding the product mix in underperforming categories

[]: