



BigBasket Project

Submitted by: Rohan Negi

Submitted to: Neeraj Sharma

1. Loading Data

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

# Loading CSV file in Pandas
df = pd.read_csv("C:/Users/Rohan/Downloads/BigBasket Products.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]
```

2. Using Head function to looking for first 12 rows

```
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...

2. Using Tail function to looking for first 12 rows

```
df.tail(12)
```

	index	product \
27543	27544	Popcorn - French Butter & Pink Salt
27544	27545	Brass Dhoop Aarti With Wooden Handle,No.3
27545	27546	Toilet Cleaning Brush - Round With Holder (Big)
27546	27547	Organic Powder - Garam Masala
27547	27548	Powder - Baking
27548	27549	Apple Cider Vinegar Shampoo
27549	27550	Papad - Garlic Disco
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 \
27543	Gourmet & World Food	Snacks, Dry Fruits, Nuts
27544	Cleaning & Household	Pooja Needs
27545	Cleaning & Household	Mops, Brushes & Scrubs
27546	Foodgrains, Oil & Masala	Organic Staples
27547	Snacks & Branded Foods	Ready To Cook & Eat
27548	Beauty & Hygiene	Hair Care
27549	Snacks & Branded Foods	Ready To Cook & Eat
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 \
27543	4700BC	31.50	35.0
27544	Trm	239.00	495.0
27545	Liao	189.00	349.0
27546	Organic Tattva	152.00	160.0
27547	Kwality	38.00	38.0
27548	Morpheme Remedies	499.00	499.0
27549	Atish	61.00	61.0
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 \
27543	Gourmet Popcorn	4.1
27544	Other Pooja Needs	4.0
27545	Toilet & Other Brushes	3.8
27546	Organic Masalas & Spices	4.2
27547	Home Baking	3.7
27548	Shampoo & Conditioner	5.0
27549	Papads, Ready To Fry	4.0
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
27543	High-quality mushroom corn popped in olive oil...
27544	Dhoobakal is a vessel used for burning the Dho...
27545	This round toilet brush is made up of virgin q...
27546	Organic Tattva Garam masala is a famous spice ...
27547	Kwality Baking Powder Directions for use: Use ...
27548	Say no to dull, lifeless, dry and damaged hair...
27549	Papads are prepared from urad dal flour and sp...

```

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 ...

```

3. Getting Description of the data in the DataFrame

```
df.describe()
```

	index	sale_price	market_price	rating
count	27555.000000	27549.000000	27555.000000	18919.000000
mean	13778.000000	334.648391	382.056664	3.943295
std	7954.58767	1202.102113	581.730717	0.739217
min	1.000000	2.450000	3.000000	1.000000
25%	6889.500000	95.000000	100.000000	3.700000
50%	13778.000000	190.320000	220.000000	4.100000
75%	20666.500000	359.000000	425.000000	4.300000
max	27555.000000	112475.000000	12500.000000	5.000000

4. Getting info about the Data Frame

```
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

```

5. Finding out top and amp: least sold products

```

# Top 10 sold products
top_sold = df.sort_values(by='sale_price', ascending=False).head(10)
top_sold

```

	index	product \
1249	1250	Beard Kit

248	249	4mm Aluminium Induction Base Chapati Roti Tawa...
436	437	Balloon - Polka Dot, 12 Inch
288	289	Arrabbiata Tomato Pasta Sauce With Chilli
25301	25302	Bravura Clipper
21761	21762	Pet Food - N&D Team Breeder Puppy Top Farmina
12669	12670	Epilator SE9-9961 Legs-Body-Face
23082	23083	Gas Stove-4 Burner Royale Plus Schott Glass, B...
2781	2782	Extra Virgin Olive Oil
25797	25798	4 Burner Gas Stove - Marvel Plus Glass Tables,...

sale_price \	category	sub_category	brand
1249	Beauty & Hygiene	Men's Grooming	Uncle Tony
112475.00			
248	Kitchen, Garden & Pets	Cookware & Non Stick	HAZEL
111649.00			
436	Cleaning & Household	Party & Festive Needs	B Vishal
88899.00			
288	Gourmet & World Food	Sauces, Spreads & Dips	Montanini
22325.00			
25301	Kitchen, Garden & Pets	Pet Food & Accessories	Wahl
12500.00			
21761	Kitchen, Garden & Pets	Pet Food & Accessories	Farmina
10090.00			
12669	Beauty & Hygiene	Feminine Hygiene	Braun
8184.44			
23082	Kitchen, Garden & Pets	Cookware & Non Stick	Prestige
7999.00			
2781	Gourmet & World Food	Oils & Vinegar	Abbies
7299.00			
25797	Kitchen, Garden & Pets	Cookware & Non Stick	Prestige
7270.00			

	market_price	type	rating \
1249	3300.0	Combos & Gift Sets	NaN
248	1289.0	Tawa & Sauce Pan	NaN
436	129.0	Caps, Balloons & Candles	3.9
288	325.0	Mustard & Cheese Sauces	5.0
25301	12500.0	Pet Cleaning & Grooming	NaN
21761	10090.0	Pet Meals & Treats	NaN
12669	10769.0	Hair Removal	NaN
23082	12245.0	Gas Stove	NaN
2781	7299.0	Extra Virgin Olive Oil	NaN
25797	9695.0	Gas Stove	NaN

	description	
discount_percentage		
1249	The combination of a beard oil, a beard wash, ...	-
3308.33		
248	Hazel Aluminium Tawa has an ergonomic design f...	-

```

8561.68
436 Whether it is a party in the office, a Christm... -
68813.95
288 NaN -
6769.23
25301 The bravura clipper is a must-have clipper for...
0.00
21761 Dog Food Adult Health Nutritional Dog Food.
0.00
12669 This cordless epilator has a sonic exfoliation...
24.00
23082 Prestige Royale Plus Gas Stove. Add a Touch of...
34.68
2781 Suitable to cook Indian meals due to its neutr...
0.00
25797 Prestige Marvel Plus Gas Stove . A Smart and C...
25.01

```

lest 10 sold product

```

least_sold = df.sort_values(by='sale_price').head(10)
least_sold

```

	index	product \
26976	26977	Curry Leaves
21312	21313	Serum
11306	11307	Happy Happy Choco-Chip Cookies
2761	2762	Orbit Sugar-Free Chewing Gum - Lemon & Lime
3445	3446	Marie Light Biscuits - Active
9971	9972	Tomato - Local, Organically Grown
27490	27491	50-50 Timepass Salted Biscuits
22072	22073	Tiger Chocolate Cream Biscuits
18290	18291	Sugar Coated Chocolate
21228	21229	Dish Shine Bar

	category	sub_category	brand \
26976	Fruits & Vegetables	Herbs & Seasonings	Fresho
21312	Beauty & Hygiene	Hair Care	Livon
11306	Snacks & Branded Foods	Biscuits & Cookies	Parle
2761	Snacks & Branded Foods	Chocolates & Candies	Wrigleys
3445	Snacks & Branded Foods	Ready To Cook & Eat	Sunfeast
9971	Fruits & Vegetables	Fresh Vegetables	Fresho
27490	Snacks & Branded Foods	Biscuits & Cookies	Britannia
22072	Snacks & Branded Foods	Biscuits & Cookies	Britannia
18290	Snacks & Branded Foods	Chocolates & Candies	Cadbury Gems
21228	Cleaning & Household	Detergents & Dishwash	Exo

	sale_price	market_price	type	rating \
26976	2.45	15.00	Indian & Exotic Herbs	NaN
21312	3.00	3.00	Hair Oil & Serum	2.5
11306	5.00	5.00	Glucose & Milk Biscuits	4.2

2761	5.00	5.00	Mints & Chewing Gum	4.2
3445	5.00	5.00	Breakfast & Snack Mixes	4.5
9971	5.00	6.25	Organic Vegetables	NaN
27490	5.00	5.00	Salted Biscuits	4.2
22072	5.00	5.00	Cream Biscuits & Wafers	4.2
18290	5.00	5.00	Chocolates	4.2
21228	5.00	5.00	Dishwash Bars & Powders	4.2

	description
discount_percentage	
26976	With dark green and glossy appearance, curry l...
83.67	
21312	Instantly Softens and Smoothens Dry, Rough, Ta...
0.00	
11306	Baked to perfection and topped with delicious ...
0.00	
2761	With Orbit Sugarfree Chewing Gums, there's no ...
0.00	
3445	EAN Code: 8901725114848 Country of origin: In...
0.00	
9971	Fresho brings to you an exquisite range of loc...
20.00	
27490	Britannia 50-50 Timepass Salted Biscuits are a...
0.00	
22072	Britannia Tiger Chocolate Cream Biscuits will ...
0.00	
18290	Cadbury Gems is chocolate flavor, rich and cho...
0.00	
21228	Exo Dish Shine Bar makes your vessels aromatic...
0.00	

6. Measuring discount on a certain item.

```
df['discount'] = ((df['market_price'] - df['sale_price']) /
df['market_price']) * 100
print(df[['product', 'market_price', 'sale_price', 'discount']])
```

	product	market_price
0	Garlic Oil - Vegetarian Capsule 500 mg	220.0
1	Water Bottle - Orange	180.0
2	Brass Angle Deep - Plain, No.2	250.0
3	Cereal Flip Lid Container/Storage Jar - Assort...	176.0
4	Creame Soft Soap - For Hands & Body	162.0
...

27550	Wottagirl! Perfume Spray - Heaven, Classic	249.0
27551	Rosemary	75.0
27552	Peri-Peri Sweet Potato Chips	200.0
27553	Green Tea - Pure Original	495.0
27554	United Dreams Go Far Deodorant	390.0

	sale_price	discount
0	220.00	0.000000
1	180.00	0.000000
2	119.00	52.400000
3	149.00	15.340909
4	162.00	0.000000
...
27550	199.20	20.000000
27551	67.50	10.000000
27552	200.00	0.000000
27553	396.00	20.000000
27554	214.53	44.992308

[27555 rows x 4 columns]

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

```
data.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
discount       6
dtype: int64

# Finding Blank values
blank_values = (df==" ").sum()
print('Blank Values count:\n',blank_values)

Blank Values count:
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
discount          0
dtype: int64
```

```
df[df.isna().any(axis=1)]
```

	index	product \
55	56	Soothing Cucumber Facial Scrub With Apricot Seeds
59	60	Corporate Planner Diary With Premium PU Leathe...
65	66	Ayurvedic Anti-Tan Face Pack
68	69	Organic Carom Seeds/Ajwain/0m Kalu
69	70	Padded Harness - 3/4 inch, Grey Colour
...
27509	27510	Deluxe Crackers - Veg
27511	27512	Specialist Stain Remover Pen & Marker
27514	27515	Verge & Sheer Perfume For Pair
27530	27531	Tick'et to Fleadom Dry Shampoo For Dogs
27531	27532	Vanilla Ice Cream

	category	sub_category
brand \		
55	Beauty & Hygiene	Skin Care
TJORI		
59	Cleaning & Household	Stationery
Plus		
65	Beauty & Hygiene	Skin Care
TJORI		
68	Foodgrains, Oil & Masala	Masalas & Spices
Earthon		
69	Kitchen, Garden & Pets	Pet Food & Accessories
Glenand		
...
.		
27509	Gourmet & World Food	Chocolates & Biscuits
Kerk		
27511	Cleaning & Household	All Purpose Cleaners
365		
27514	Beauty & Hygiene	Fragrances & Deos
Skinn by		
Titan		
27530	Kitchen, Garden & Pets	Pet Food & Accessories
Captain		
Zack		
27531	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	40.0	The soothing feel of cucumber meets the gentle...
59	0.0	A5 Size (210x150mm) \n192 Pages Premium Natura...
65	40.0	A nourishing face pack that removes tan and br...
68	0.0	Earthon's Ajwain is Best quality, organically ...
69	0.0	These are soft padded harness for your active ...
...		...
...		...
27509	0.0	Kerk Biscuits-has been a household name synony...
27511	0.0	Mightier than the pen. The pen may be mightier...
27514	10.0	VERGE for men paints a picture of a classy out...
27530	0.0	1) No Rinse Defence Against Ticks and Fleas: C...
27531	0.0	Some things are so good in their simplest form...

	discount
55	40.0
59	0.0
65	40.0
68	0.0
69	0.0
...	...
27509	0.0
27511	0.0
27514	10.0

```
27530      0.0
27531      0.0
```

```
[8728 rows x 12 columns]
```

```
# Checking NaN value in Product column
```

```
df[df['product'].isna()]
```

```
      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
```

```
      discount
14363  16.666667
```

```
# Handling NaN value with "Unknown"
```

```
df.loc[df['product'].isna(), 'product'] = 'Unknown'
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
Name: product, Length: 27555, dtype: object
```

```
# Checking NaN value in Brand column
```

```
df[df['brand'].isna()]
```

```
      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			

discount
9765 0.0

```
# Handling NaN value with "Unknown"
df.loc[df['brand'].isna(), 'brand'] = 'Unknown'
df['brand']
```

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

```
# Checking NaN value in Sale_Price column
df[df['sale_price'].isna()]
```

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

```

```

discount
1719 NaN
1720 NaN
1721 NaN
1722 NaN
2401 NaN
2402 NaN

```

Handling NaN value with "Median"

```
df['sale_price'] = df['sale_price'].fillna(df['sale_price'].median())
```

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

```

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

```

```
# Checking NaN value in rating column
```

```
df[df['rating'].isna()]
```

	index	product \
55	56	Soothing Cucumber Facial Scrub With Apricot Seeds
59	60	Corporate Planner Diary With Premium PU Leathe...
65	66	Ayurvedic Anti-Tan Face Pack
68	69	Organic Carom Seeds/Ajwain/0m Kalu
69	70	Padded Harness - 3/4 inch, Grey Colour
...
27509	27510	Deluxe Crackers - Veg
27511	27512	Specialist Stain Remover Pen & Marker
27514	27515	Verge & Sheer Perfume For Pair
27530	27531	Tick'et to Fleadom Dry Shampoo For Dogs
27531	27532	Vanilla Ice Cream

	brand \	category	sub_category	
55		Beauty & Hygiene	Skin Care	
TJORI				
59		Cleaning & Household	Stationery	Prozo
Plus				
65		Beauty & Hygiene	Skin Care	
TJORI				
68		Foodgrains, Oil & Masala	Masalas & Spices	
Earthon				
69		Kitchen, Garden & Pets	Pet Food & Accessories	
Glenand				
...	
.				
27509		Gourmet & World Food	Chocolates & Biscuits	
Kerk				
27511		Cleaning & Household	All Purpose Cleaners	
365				
27514		Beauty & Hygiene	Fragrances & Deos	Skin by
Titan				
27530		Kitchen, Garden & Pets	Pet Food & Accessories	Captain
Zack				
27531		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

```

```

discount
55 40.0
59 0.0
65 40.0
68 0.0
69 0.0
...
27509 0.0
27511 0.0
27514 10.0
27530 0.0
27531 0.0

```

```
[8636 rows x 12 columns]
```

```
# Handling NaN value with "Median"
```

```
df['rating'] = df['rating'].fillna(df['rating'].median())
```

```
# Checking NaN value in description column
```

```
df[df['description'].isna()]
```

	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	
Chekko			
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	4.1
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	discount
288	NaN	-6769.23	-6769.230769
480	NaN	0.00	0.000000
780	NaN	0.00	0.000000
845	NaN	0.00	0.000000
1047	NaN	21.20	21.200573
...
25510	NaN	38.90	38.896552
25583	NaN	27.88	27.882038
26116	NaN	26.67	26.666667
26835	NaN	0.00	0.000000
27303	NaN	0.00	0.000000

[115 rows x 12 columns]

Handling NaN value with "Unknown"

```
df.loc[df['description'].isna(), 'description'] = 'Unknown'
df['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 ...
```

Name: description, Length: 27555, dtype: object

Checking NaN value in discount_percentage column

```
df[df['discount_percentage'].isna()]
```

	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)

	brand \	category	sub_category	
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	190.32	20.0	Camphor & Wicks	4.1	
1720	190.32	46.0	Blended Masalas	4.1	
1721	190.32	299.0	Cutlery, Spoon & Fork	4.1	
1722	190.32	65.0	Breakfast & Snack Mixes	4.1	
2401	190.32	150.0	Battery & Electrical	4.0	
2402	190.32	32.0	Toilet Paper	3.9	

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

	discount
1719	NaN
1720	NaN
1721	NaN
1722	NaN
2401	NaN
2402	NaN

Handling NaN value with "Median"

```
df['discount_percentage'] =
df['discount_percentage'].fillna(df['discount_percentage'].median())
```

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

index	0
product	0
category	0
sub_category	0
brand	0
sale_price	0

```
market_price      0
type              0
rating            0
description        0
discount_percentage 0
discount          6
dtype: int64
```

```
# Checking NaN value in discount column
df[df['discount'].isna()]
```

```
      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  Klasic 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      190.32      20.0      Camphor & Wicks      4.1
1720      190.32      46.0      Blended Masalas      4.1
1721      190.32      299.0  Cutlery, Spoon & Fork      4.1
1722      190.32      65.0  Breakfast & Snack Mixes      4.1
2401      190.32      150.0  Battery & Electrical      4.0
2402      190.32      32.0      Toilet Paper      3.9
```

```
      description
discount_percentage \
1719  Jaji Puja Flower Batti / Puvvu Vathulu is made...
5.0
1720  Annapoorna Spices are the preserve for Connoiss...
5.0
1721  BB Home provides fine and classy cutlery that ...
5.0
1722  MTR Dhokla is extremely special in Gujarat. It...
```

```

5.0
2401 This Nippo battery incorporates state-of-the-a...
5.0
2402 Klassic Plain Cocktail Napkins (22 X 22 cm) 10...
5.0

      discount
1719      NaN
1720      NaN
1721      NaN
1722      NaN
2401      NaN
2402      NaN

# Handling NaN value with "Median"
df['discount'] = df['discount'].fillna(df['discount'].median())

df.isna().sum()

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

```

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

.

```

# calculate the IQR
Q1=df['sale_price'].quantile(0.25)
Q3=df['sale_price'].quantile(0.75)
IQR=Q3-Q1

# define outlier range
lower_bound = Q1 - 1.5*IQR
upper_bound = Q3 + 1.5*IQR

# Identify outliers
outliers = df[(df['sale_price'] < lower_bound) | (df['sale_price'] >

```

```

upper_bound)]

# Display the calculated values
print(f"Q1: {Q1}")
print(f"Q3: {Q3}")
print(f"IQR: {IQR}")
print(f"lower: {lower_bound}")
print(f"upper: {upper_bound}")

Q1: 95.0
Q3: 359.0
IQR: 264.0
lower: -301.0
upper: 755.0

# Create a histogram to visualize the distribution of the data
plt.figure(figsize=(8, 6))
sns.histplot(df['rating'], bins=10, kde=True, color='skyblue',
edgecolor='black')

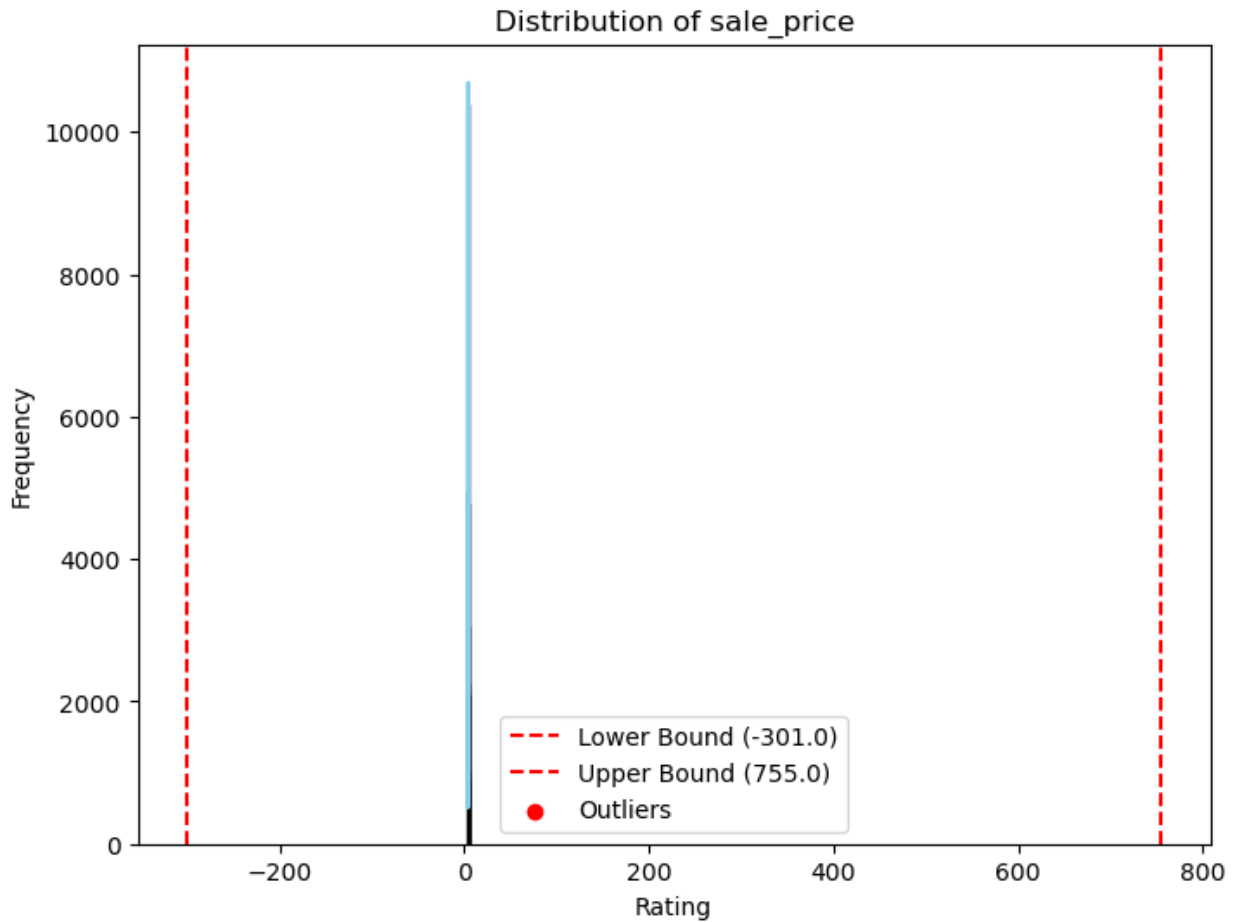
# Add lines for the lower and upper bounds
plt.axvline(x=lower_bound, color='red', linestyle='--', label=f'Lower
Bound ({lower_bound})')
plt.axvline(x=upper_bound, color='red', linestyle='--', label=f'Upper
Bound ({upper_bound})')

# Highlight the outliers
outlier_values = df[(df['rating'] < lower_bound) | (df['rating'] >
upper_bound)]['rating']
plt.scatter(outlier_values, np.zeros_like(outlier_values),
color='red', label='Outliers', zorder=5)

# Add title and labels
plt.title('Distribution of sale_price')
plt.xlabel('Rating')
plt.ylabel('Frequency')
plt.legend()

# Show the plot
plt.show()

```

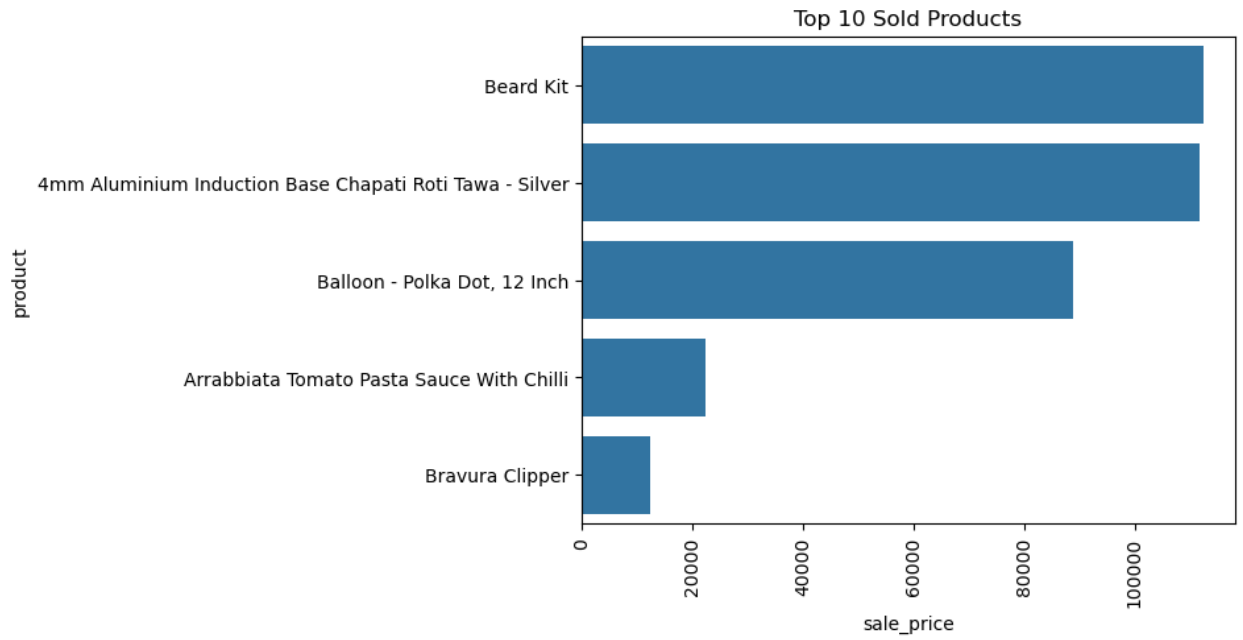


Create Plots or visualizations.

9. Create Plots or visualizations.

```
import matplotlib.pyplot as plt
import seaborn as sns

# Bar plot of top 10 sold products
sns.barplot(x='sale_price', y='product', data=top_sold)
plt.xticks(rotation=90)
plt.title("Top 10 Sold Products")
plt.show()
```

Observations:¶

Beard Kit: This is the clear top-selling product, with the highest sales volume among the 10 listed.

4mm Aluminium Induction Base Chapati Roti Tawa - Silver: This product ranks second in sales, indicating strong demand for kitchenware items.

Balloon - Polka Dot; 12 Inch: This product occupies the third position, suggesting popularity for party supplies or decorative items.

Arrabbiata Tomato Pasta Sauce With Chilli: This product has lower sales compared to the top three, suggesting a potential niche market or less demand for this specific flavor.

Bravura Clipper: This product has the lowest sales among the top 10, indicating it might be a less popular or niche product.

Recommendations:

Focus on Top Performers: Invest in marketing and promotions for the top-selling products like the Beard Kit and 4mm Aluminium Induction Base Chapati Roti Tawa to further boost their sales.

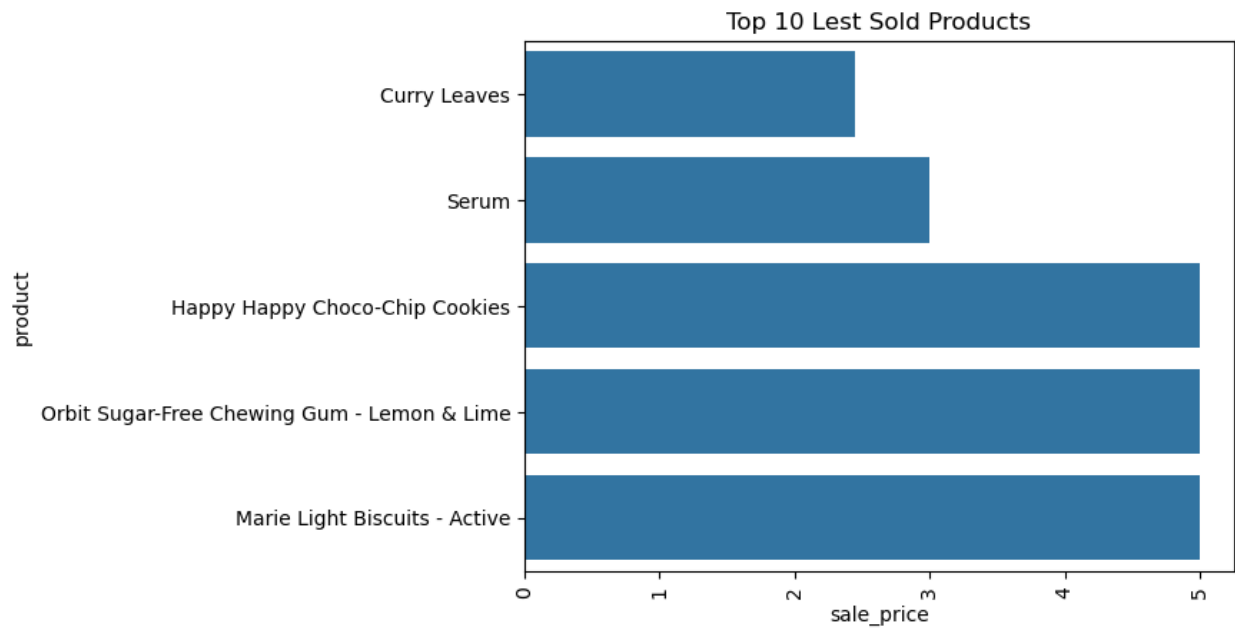
Analyze Underperformers: Investigate the reasons behind the low sales of products like Arrabbiata Tomato Pasta Sauce With Chilli and Bravura Clipper. Consider price adjustments, targeted marketing, or product improvements to increase their appeal.

Product Bundling: Create attractive product bundles combining top-selling items with underperforming ones to incentivize purchases and increase overall sales.

Customer Feedback: Gather customer feedback on the underperforming products to identify specific areas for improvement and tailor your strategies accordingly.

```
sns.barplot(x='sale_price', y='product', data=least_sold)
plt.xticks(rotation=90)
```

```
plt.title("Top 10 Lest Sold Products")  
plt.show()
```



Observations:

Product Performance:

Marie Light Biscuits - Active and Orbit Sugar-Free Chewing Gum - Lemon & Lime are the least sold products, with sales significantly lower than the others.

Happy Happy Choco-Chip Cookies and Serum have slightly higher sales compared to the top two least sold products.

Curry Leaves is the most sold product among the top 10 least sold items.

Potential Insights:

Product Category: The chart includes a mix of food items (biscuits, cookies, gum, curry leaves) and a non-food item (serum). This suggests that the underperformance might not be solely due to product category.

Price Point: It's important to consider the pricing of these products compared to similar items. Lower-priced items might have lower sales due to lower profit margins.

Marketing and Promotion: The lack of effective marketing and promotion for these products could be contributing to their low sales.

Customer Preferences: It's possible that these products don't align with current customer preferences or trends.

Recommendations:

Targeted Marketing: Implement targeted marketing campaigns to promote these products to specific customer segments.

Price Optimization: Analyze pricing strategies and consider adjustments to make these products more competitive.

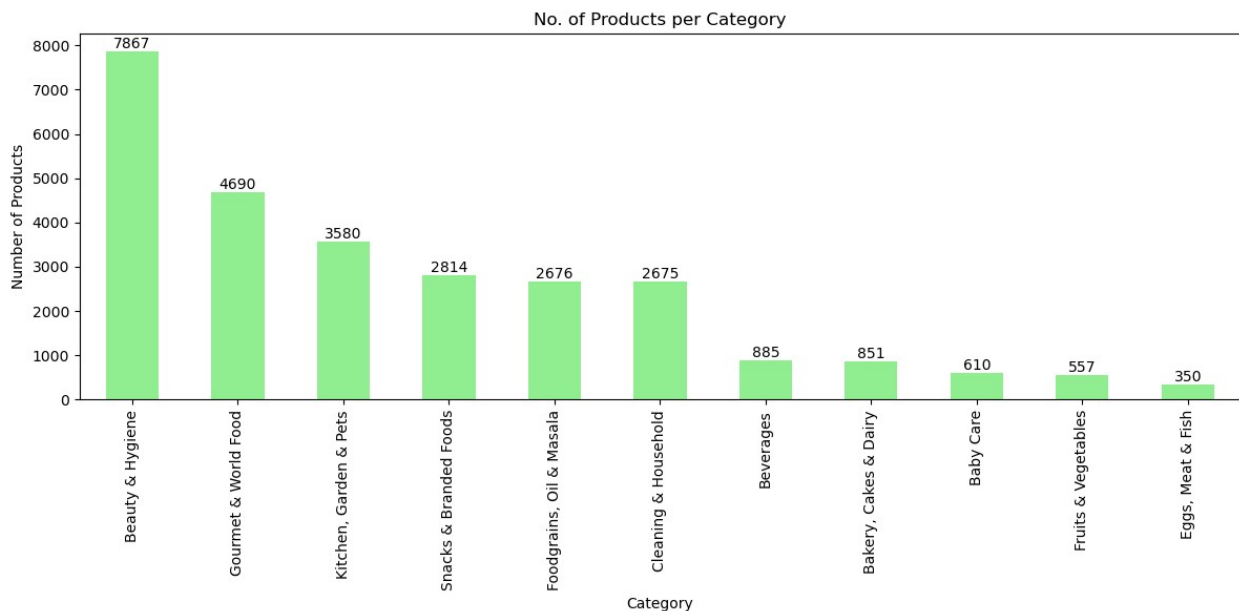
Product Bundling: Create attractive product bundles that include these items to encourage sales.

Customer Feedback: Gather customer feedback to understand why these products are not performing well and make necessary improvements.

```
# Count the number of products per category
category_counts = df['category'].value_counts()

# Plotting the number of products per category
plt.figure(figsize=(12,6))
ax = category_counts.plot(kind='bar', color='lightgreen')
plt.title('No. of Products per Category')
plt.xlabel('Category')
plt.ylabel('Number of Products')
# Adding labels on top of the bars
for i in ax.patches: ax.text(i.get_x() + i.get_width() / 2,
                             i.get_height() + 0.2,
                             f'{int(i.get_height())}', ha='center',
                             va='bottom')

plt.tight_layout()
plt.show()
```



Observations:

The distribution of products across categories is uneven, with some categories having a significantly larger number of products than others.

This uneven distribution might indicate different levels of demand or market size for different product categories.

It's important to consider the sales performance and profitability of each category to understand the overall business impact.

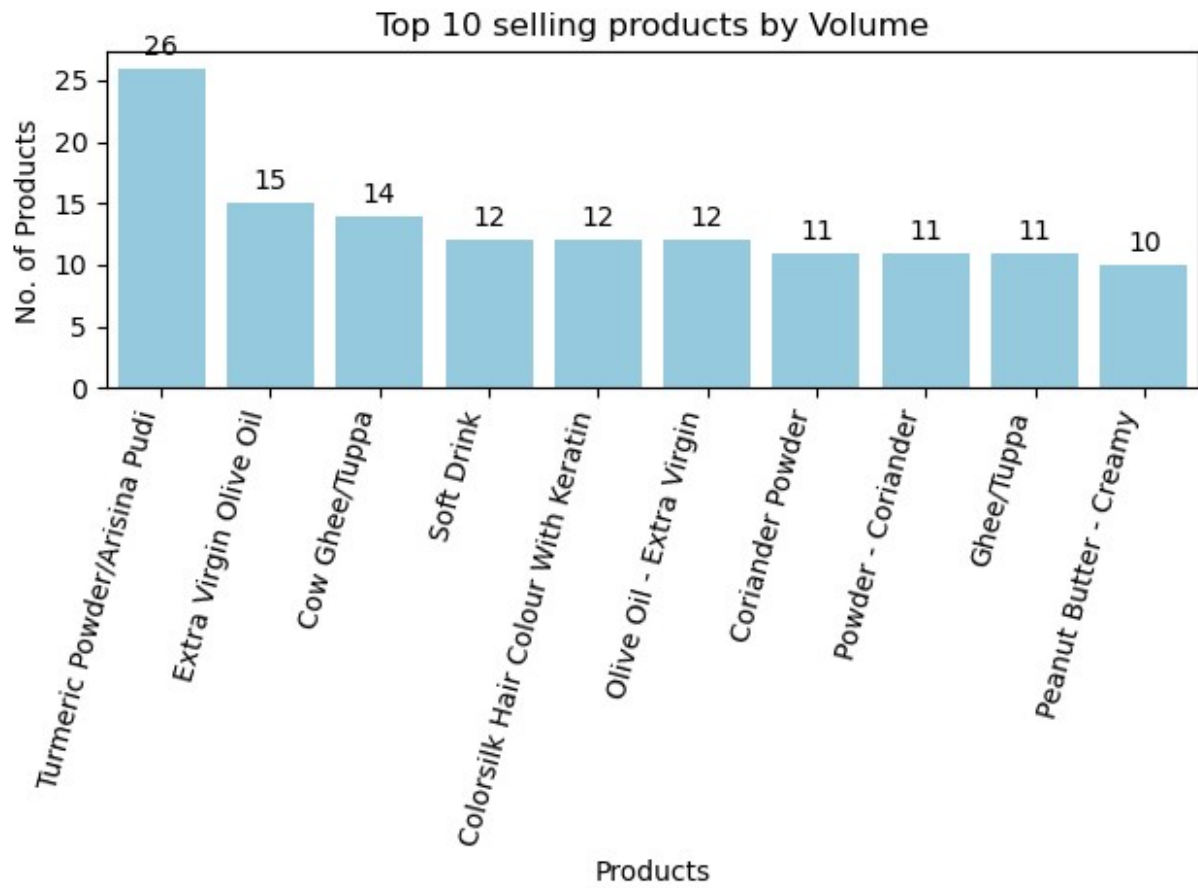
Recommendations:

Category Analysis: Analyze the sales and profit margins of each category to identify opportunities for improvement.

Product Portfolio Management: Review the product portfolio within each category to ensure a balanced mix of high-performing and low-performing items.

Targeted Marketing: Implement targeted marketing strategies for categories with high potential but lower product numbers.

```
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='skyblue')
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=(0, 5), textcoords='offset points')
plt.tight_layout()
plt.show()
```



Observations:

The chart highlights a wide range of product categories, including food items, personal care products, and beverages.

The top-selling products are a mix of essential household items and popular consumer goods.

It's important to consider the sales volume in relation to the overall product portfolio and market demand.

Recommendations:

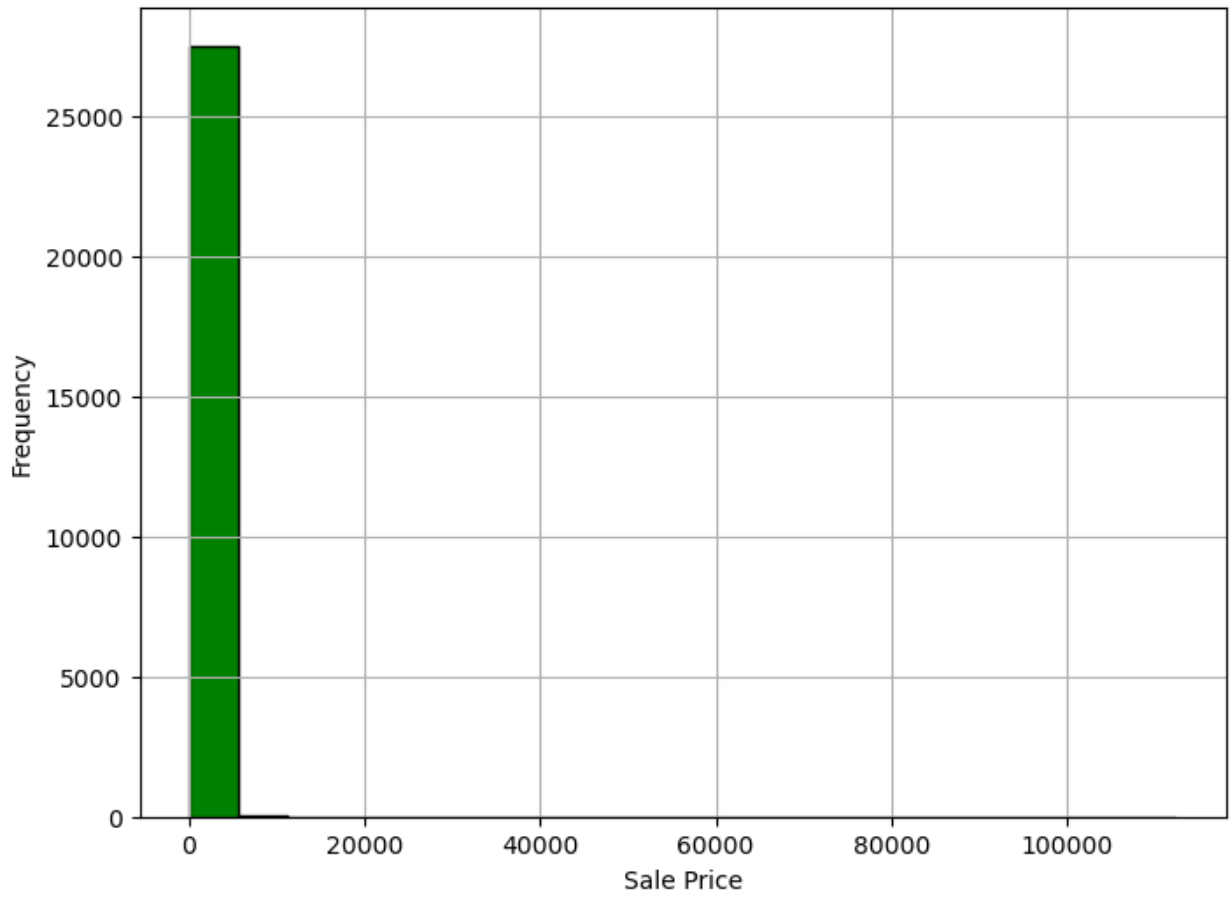
Inventory Management: Optimize inventory levels for the top-selling products to avoid stockouts and meet customer demand.

Marketing and Promotion: Continue to promote the top-selling products through effective marketing campaigns.

Customer Insights: Analyze customer purchasing behavior to identify trends and opportunities for cross-selling and upselling.

```
# Plot histogram of sale prices
plt.figure(figsize=(8,6))
plt.hist(df['sale_price'], bins=20, color='green', edgecolor='black')
plt.title('Distribution of Sale Prices')
plt.xlabel('Sale Price')
plt.ylabel('Frequency')
plt.grid(True)
plt.show()
```


Distribution of Sale Prices



Observations:

The distribution is right-skewed, suggesting that most sales are for lower-priced items.

This pattern is common in retail settings where a large portion of sales comes from lower-priced products.

The distribution might be influenced by factors such as product mix, target market, and pricing strategy.

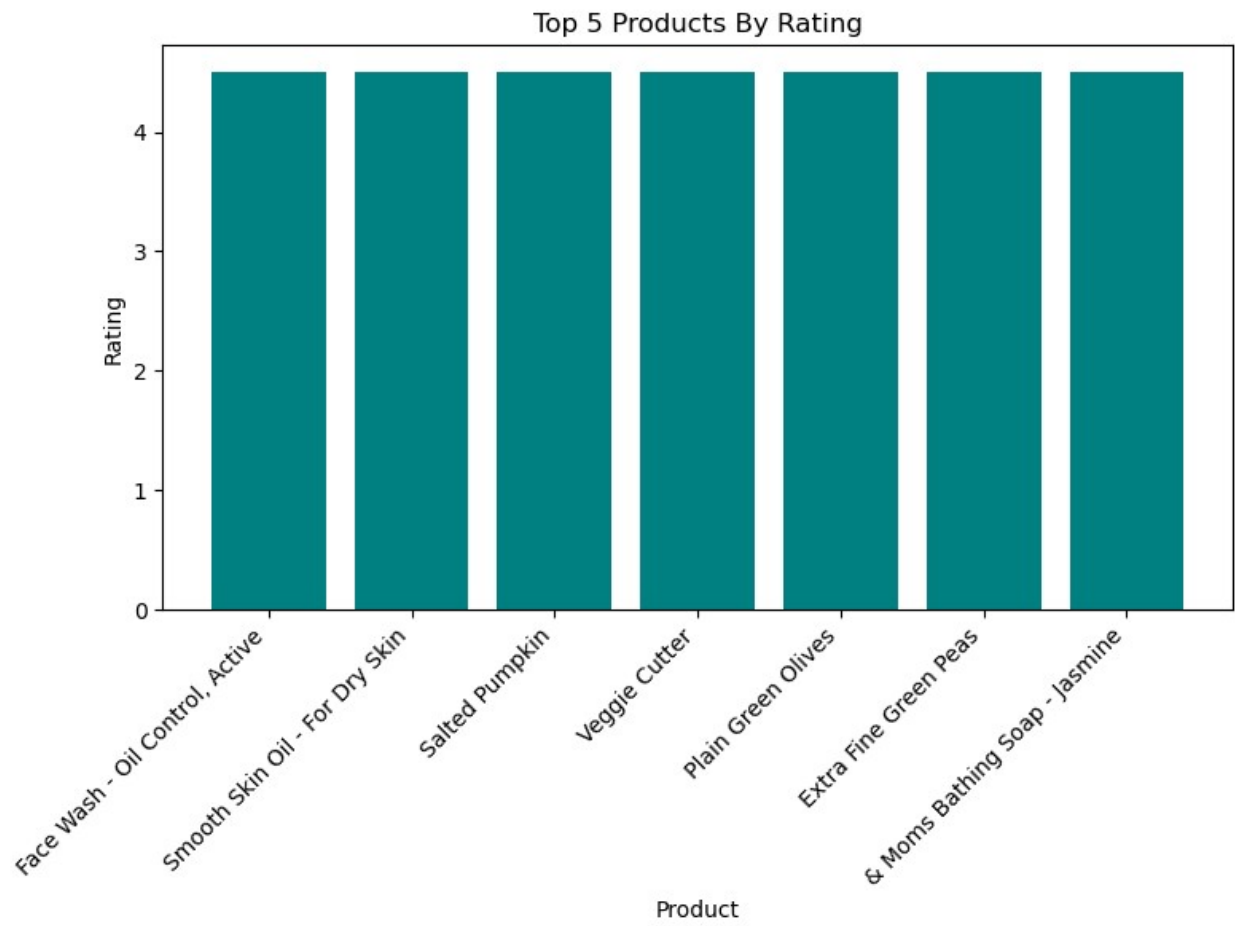
Recommendations:

Pricing Strategy: Analyze the pricing strategy for high-priced items to identify potential adjustments that could increase sales.

Product Mix: Consider expanding the product mix to include more mid-priced items to attract a wider range of customers.

Marketing and Promotion: Implement targeted marketing campaigns to promote high-priced items and highlight their unique value proposition.

```
top_5_products = df.nlargest(7, 'rating')[['product', 'rating']]
# Plot histogram of ratings
plt.figure(figsize=(8,6))
plt.bar(top_5_products['product'],
top_5_products['rating'],color="#008080")
plt.title('Top 5 Products By Rating')
plt.xlabel('Product')
plt.ylabel('Rating')
plt.xticks(rotation=45,ha='right')
plt.tight_layout()
plt.show()
```



Observations:

The chart highlights the importance of customer reviews and ratings in assessing product performance.

Positive ratings can significantly impact sales and brand reputation.

It's important to maintain consistent quality and customer service to sustain high ratings.

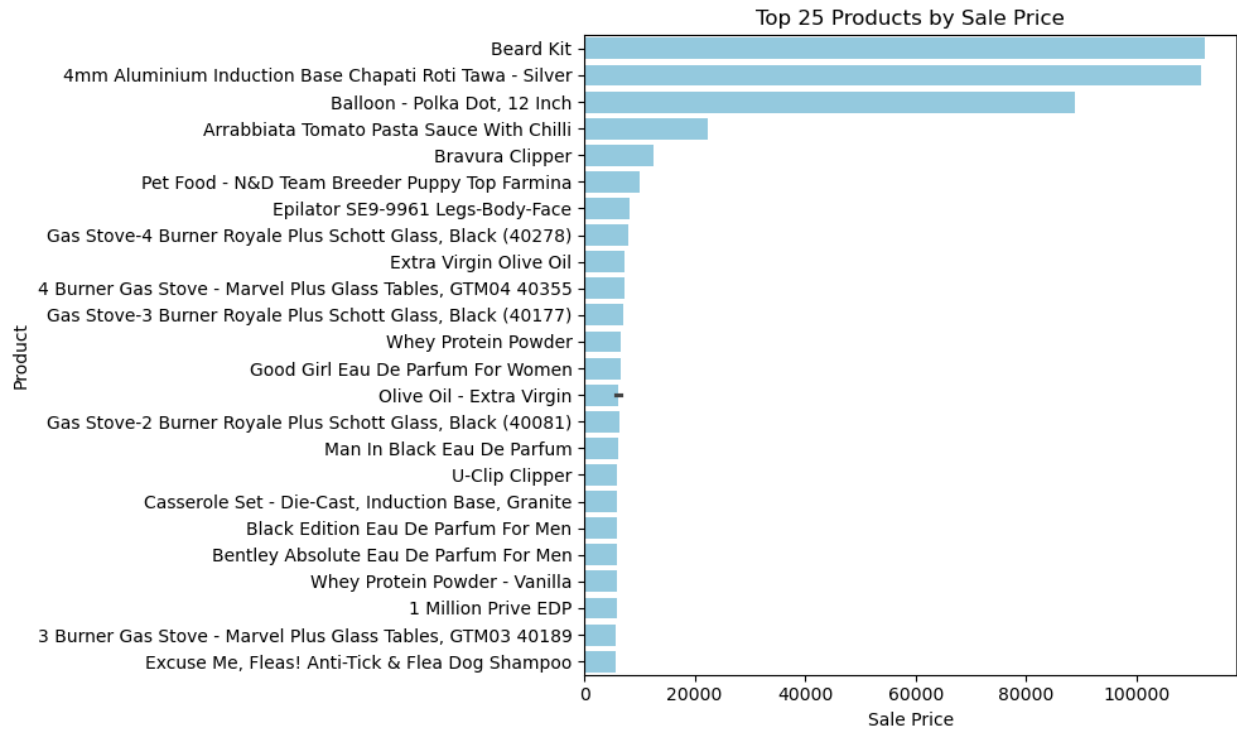
Recommendations:

Customer Feedback: Continuously monitor customer feedback and reviews to identify areas for improvement.

Quality Control: Implement rigorous quality control measures to ensure product consistency.

Marketing and Promotion: Leverage positive ratings in marketing and promotional materials to attract new customers.

```
top_25_products = df[['product',  
'sale_price']].sort_values(by='sale_price', ascending=False).head(25)  
plt.figure(figsize=(10, 6))  
ax = sns.barplot(x='sale_price', y='product', data=top_25_products,  
color='skyblue')  
ax.set_xlabel('Sale Price')  
ax.set_ylabel('Product')  
ax.set_title('Top 25 Products by Sale Price')  
plt.tight_layout()  
plt.show()
```



Observations:

The chart highlights the diversity of products offered, ranging from household appliances to personal care items and food products.

The distribution of prices suggests that the platform caters to a wide range of customer budgets and preferences.

It's important to consider the sales volume and profit margins of each product to understand the overall revenue impact.

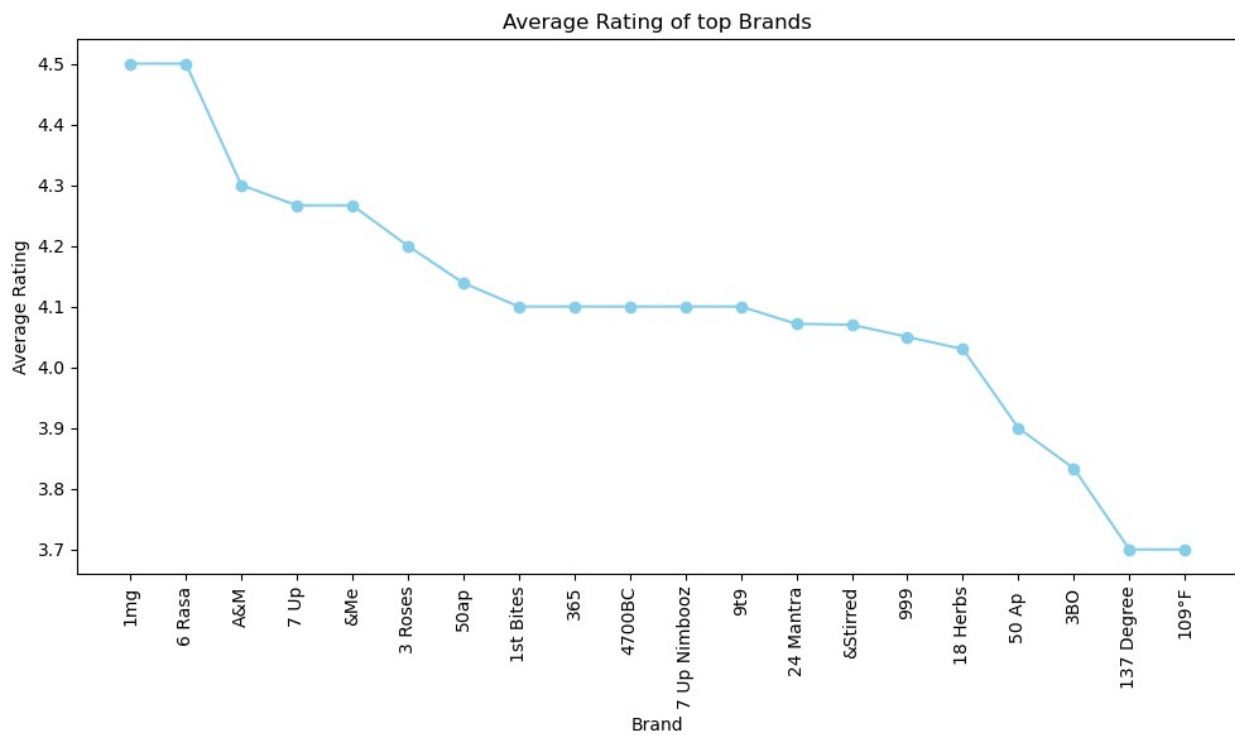
Recommendations:

Pricing Strategy: Analyze the pricing strategy for high-priced items to ensure they are competitively priced and aligned with customer expectations.

Product Mix: Consider expanding the product mix to include more mid-priced items to attract a wider range of customers.

Marketing and Promotion: Implement targeted marketing campaigns to promote high-priced items and highlight their unique value proposition.

```
brand_ratings = df.groupby('brand')
['rating'].mean().reset_index().head(20)
brand_ratings_sorted = brand_ratings.sort_values(by='rating',
ascending=False)
plt.figure(figsize=(10,6))
plt.plot(brand_ratings_sorted['brand'],
brand_ratings_sorted['rating'], marker='o', color='skyblue')
plt.xticks(rotation=90)
plt.xlabel('Brand')
plt.ylabel('Average Rating')
plt.title('Average Rating of top Brands')
plt.tight_layout()
plt.show()
```



Observations:

The chart highlights the importance of customer reviews and ratings in assessing brand performance.

Positive ratings can significantly impact brand reputation and customer trust.

It's important for brands to actively monitor and respond to customer feedback to maintain high ratings.

Recommendations:

Customer Feedback: Continuously monitor customer feedback and reviews to identify areas for improvement.

Quality Control: Implement rigorous quality control measures to ensure product consistency.

Marketing and Promotion: Leverage positive ratings in marketing and promotional materials to attract new customers.

Overall Conclusion

The analysis highlights a diverse range of products and categories, revealing disparities in sales performance and demand across product types. Top-selling products like Beard Kit and Curry Leaves suggest strong demand in specific niches, while underperforming items like Bravura Clipper and Arrabbiata Sauce point to potential issues in pricing, marketing, or market alignment. The skewed sales distribution toward lower-priced items emphasizes the need for a balanced pricing strategy and product mix. To drive growth, the focus should be on optimizing inventory, targeted marketing for high-potential categories, leveraging customer feedback, and promoting positive product ratings to enhance brand trust and visibility.