BIG BASKET Mini Project

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
warnings.filterwarnings('ignore')
```

```
# Load DataSet.
df=pd.read_csv(r"C:\Users\TEMP.RYZENFIVE.008\OneDrive\Desktop\My Projects\Skill Cricle Projects\Big Basket Mini Project\BigBasket Products.csv")
df.head(12) #Use head function to look for first 12 rows.
```

	index	product	category	sub_category	brand	sale_price	market_price	type	rating	description
0	1	Garlic Oil - Vegetarian Capsule 500 mg	Beauty & Hygiene	Hair Care	Sri Sri Ayurveda	220.0	220.0	Hair Oil & Serum	4.1	This Product contains Garlic Oil that is known...
1	2	Water Bottle - Orange	Kitchen, Garden & Pets	Storage & Accessories	Mastercook	180.0	180.0	Water & Fridge Bottles	2.3	Each product is microwave safe (without lid), ...
2	3	Brass Angle Deep - Plain, No.2	Cleaning & Household	Pooja Needs	Trm	119.0	250.0	Lamp & Lamp Oil	3.4	A perfect gift for all occasions, be it your m...
3	4	Cereal Flip Lid Container/Storage Jar - Assort...	Cleaning & Household	Bins & Bathroom Ware	Nakoda	149.0	176.0	Laundry, Storage Baskets	3.7	Multipurpose container with an attractive desi...
4	5	Creme Soft Soap - For Hands & Body	Beauty & Hygiene	Bath & Hand Wash	Nivea	162.0	162.0	Bathing Bars & Soaps	4.4	Nivea Creme Soft Soap gives your skin the best...
5	6	Germ - Removal Multipurpose Wipes	Cleaning & Household	All Purpose Cleaners	Nature Protect	169.0	199.0	Disinfectant Spray & Cleaners	3.3	Stay protected from contamination with Multipu...
6	7	Multani Mati	Beauty & Hygiene	Skin Care	Satinance	58.0	58.0	Face Care	3.6	Satinance multani matti is an excellent skin t...
7	8	Hand Sanitizer - 70% Alcohol Base	Beauty & Hygiene	Bath & Hand Wash	Bionova	250.0	250.0	Hand Wash & Sanitizers	4.0	70%Alcohol based is gentle of hand leaves skin...
8	9	Biotin & Collagen Volumizing Hair Shampoo + Bi...	Beauty & Hygiene	Hair Care	StBotanica	1098.0	1098.0	Shampoo & Conditioner	3.5	An exclusive blend with Vitamin B7 Biotin, Hyd...
9	10	Scrub Pad - Anti- Bacterial, Regular	Cleaning & Household	Mops, Brushes & Scrubs	Scotch brite	20.0	20.0	Utensil Scrub-Pad, Glove	4.3	Scotch Brite Anti- Bacterial Scrub Pad thoroug...
10	11	Wheat Grass Powder - Raw	Gourmet & World Food	Cooking & Baking Needs	NUTRASHIL	261.0	290.0	Flours & Pre-Mixes	4.0	Wheatgrass is a superfood potent health food w...
11	12	Butter Cookies Gold Collection	Gourmet & World Food	Chocolates & Biscuits	Sapphire	600.0	600.0	Luxury Chocolates, Gifts	2.2	Enjoy a tin full of delicious butter cookies m...

```
# Get Description of the data in the DataFrame
df.describe(include='all')
```

	index	product	category	sub_category	brand	sale_price	market_price	type	rating	description
count	27555.00000	27554	27555	27555	27554	27549.000000	27555.000000	27555	18919.000000	27440
unique	NaN	23540	11	90	2313	NaN	NaN	426	NaN	21944
top	NaN	Turmeric Powder/Arisina Pudi	Beauty & Hygiene	Skin Care	Fresho	NaN	NaN	Face Care	NaN	A brand inspired by the Greek goddess of victo...
freq	NaN	26	7867	2294	638	NaN	NaN	1508	NaN	47
mean	13778.00000	NaN	NaN	NaN	NaN	334.648391	382.056664	NaN	3.943295	NaN
std	7954.58767	NaN	NaN	NaN	NaN	1202.102113	581.730717	NaN	0.739217	NaN
min	1.00000	NaN	NaN	NaN	NaN	2.450000	3.000000	NaN	1.000000	NaN
25%	6889.50000	NaN	NaN	NaN	NaN	95.000000	100.000000	NaN	3.700000	NaN
50%	13778.00000	NaN	NaN	NaN	NaN	190.320000	220.000000	NaN	4.100000	NaN
75%	20666.50000	NaN	NaN	NaN	NaN	359.000000	425.000000	NaN	4.300000	NaN
max	27555.00000	NaN	NaN	NaN	NaN	112475.000000	12500.000000	NaN	5.000000	NaN

```
# Find Information about the DataFrame
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
```

```
# Checking column names
df.columns
```

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

```
# Find out Top &amp;amp; least sold products
# Top sold product
top_market_price = df[df['market_price'] == df['market_price'].max()]
print("\nTop Market Price Product(s):")
print(top_market_price[['product','market_price']])

# Least sold product
least_market_price = df[df['market_price'] == df['market_price'].min()]
print("\nLeast Market Price Product(s):")
print(least_market_price[['product','market_price']])
```

```
Top Market Price Product(s):
                product  market_price
25301  Bravura Clipper       12500.0

Least Market Price Product(s):
       product  market_price
21312   Serum           3.0
```

```
# Measuring discount on a certain item
df['discount_percentage'] = ((df['market_price'] - df['sale_price']) / df['market_price']) * 100
print("\nSample Discount Calculations:")
print(df[['product','market_price','sale_price','discount_percentage']].head(10))
```

```
Sample Discount Calculations:
                                          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                Creme Soft Soap - For Hands & Body         162.0
5                 Germ - Removal Multipurpose Wipes         199.0
6                                     Multani Mati          58.0
```

```
7                 Hand Sanitizer - 70% Alcohol Base         250.0
8  Biotin & Collagen Volumizing Hair Shampoo + Bi...        1098.0
9               Scrub Pad - Anti- Bacterial, Regular          20.0

   sale_price  discount_percentage
0       220.0             0.000000
1       180.0             0.000000
2       119.0            52.400000
3       149.0            15.340909
4       162.0             0.000000
5       169.0            15.075377
6        58.0             0.000000
7       250.0             0.000000
8      1098.0             0.000000
9        20.0             0.000000
```

```
# Find out the Missing Values from the Dataset
print("\nMissing Values:")
print(df.isnull().sum())
```

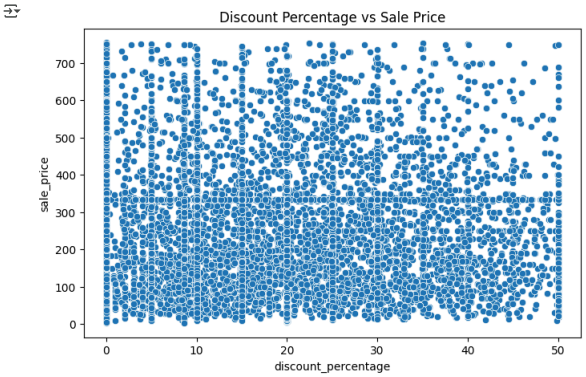
```
Missing Values:
index                    0
product                  1
category                 0
sub_category             0
brand                    1
sale_price               6
market_price             0
type                     0
rating                8636
description            115
discount_percentage      6
dtype: int64
```

```
# Find out the outliers and fill them with mean
numeric_cols = ['sale_price', 'market_price', 'rating', 'discount_percentage']

for col in numeric_cols:
    Q1 = df[col].quantile(0.25)
    Q3 = df[col].quantile(0.75)
    IQR = Q3 - Q1
    lower = Q1 - 1.5 * IQR
    upper = Q3 + 1.5 * IQR

    mean_value = df[col].mean()
    df[col] = np.where((df[col] < lower) | (df[col] > upper), mean_value, df[col])

print("\nOutliers handled successfully.")
```

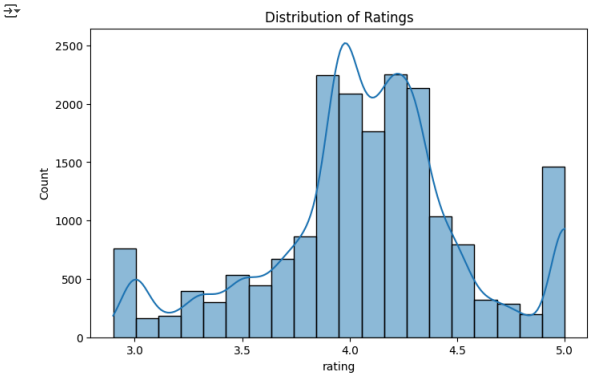
```
Outliers handled successfully.
```

```
# Create Plots or visualizations

# Distribution of Sale Price
plt.figure(figsize=(8, 5))
sns.histplot(df['sale_price'], bins=30, kde=True)
plt.title('Distribution of Sale Price')
plt.show()
```


```
# Discount vs Sale Price
plt.figure(figsize=(8, 5))
sns.scatterplot(x='discount_percentage', y='sale_price', data=df)
plt.title('Discount Percentage vs Sale Price')
plt.show()
```




```
# Ratings Distribution
plt.figure(figsize=(8, 5))
sns.histplot(df['rating'], bins=20, kde=True)
plt.title('Distribution of Ratings')
plt.show()
```


Distribution of Ratings
Count
rating
0
500
1000
1500
2000
2500
3.0
3.5
4.0
4.5
5.0