Big Basket Analysis

-Domain:- E-Commerce

-About:-

E-commerce (electronic commerce) is the activity of electronically buying or selling of products on online services or over the Internet. E- commerce draws on technologies such as mobile commerce, electronic funds transfer, supply chain management, Internet marketing, online transaction processing, electronic data interchange (EDI), inventory management systems, and automated data collection systems. E- commerce is in turn driven by the technological advances of the semiconductor industry, and is the largest sector of the electronics industry.

-EDA with Big Basket Data:-

Big basket is the largest online grocery supermarket in India. Was launched somewhere around in 2011 since then they have been expanding their business. Though some new competitors have been able to set their foot in the nation such as Blink it etc. but Big Basket has still not lost anything - thanks to ever expanding popular base and their shift to online buying.

-Data Dictionary:-

This dataset contains 10 attributes with simple meaning and which are described as follows:

1. index - Simply the Index!
2. product - Title of the product (as they're listed)
3. category - Category into which product has been classified
4. sub_category - Subcategory into which product has been kept
5. brand - Brand of the product
6. sale_price - Price at which product is being sold on the site
7. market_price - Market price of the product
8. type - Type into which product falls
9. rating - Rating the product has got from its consumers
10. description - Description of the dataset (in detail)

```python
##Importing Libraries
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt #For Visualization
import seaborn as sns #For Visualization
```

Step 1: Load DataSet

```
data = pd.read_csv("BigBasket Products.csv")
```

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

```
data.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...
```

Step 3: Get Description of the data in the DataFrame

```
x = pd.DataFrame(data["description"])
x
```

```
                                               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 1 columns]
```

Step 4: Find Information about the DataFrame

```
data.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
```

Step 5: Find out Top & least sold products.

```
top_sold_products = data.sort_values(by = "sale_price", ascending = 
False)
print("Top 10 sold products\n", top_sold_products[["product", 
"sale_price"]].head(10))

Top 10 sold products
                                                   product  sale_price
1249                                            Beard Kit   112475.00
248    4mm Aluminium Induction Base Chapati Roti Tawa...   111649.00
436                         Balloon - Polka Dot, 12 Inch    88899.00
288            Arrabbiata Tomato Pasta Sauce With Chilli    22325.00
25301                                     Bravura Clipper    12500.00
21761      Pet Food - N&D Team Breeder Puppy Top Farmina    10090.00
12669                Epilator SE9-9961 Legs-Body-Face        8184.44
23082  Gas Stove-4 Burner Royale Plus Schott Glass, B...     7999.00
2781                              Extra Virgin Olive Oil     7299.00
25797  4 Burner Gas Stove - Marvel Plus Glass Tables,...     7270.00

least_sold_product = data.sort_values( by = "sale_price", ascending = 
False)
print("Least 10 sold products\n", 
least_sold_product[["product","sale_price"]].tail(10))

Least 10 sold products
                                                   product  sale_price
18290                             Sugar Coated Chocolate         5.00
16551                        Biscuits - Magix Kreams Choc        5.00
21312                                               Serum        3.00
26976                                        Curry Leaves        2.45
1719                Puja Flower Wicks - Puvvu Vathulu Batti         NaN
```

```
1720                                     Powder - Sambar          NaN
1721   Steel Fork - Medium, Premium Excel Series, BBCL08          NaN
1722                                  Snack Mix - Dhokla          NaN
2401                               Battery AA 3UT Hi Top          NaN
2402         Klassic Plain Cocktail Napkins (22 X 22 cm)          NaN
```

Step 6: Measuring discount on a certain item

```
data["discount"] = data["market_price"] - data["sale_price"]
data.head(10)
```

```
   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

                 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


   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

                                         description  discount
0  This Product contains Garlic Oil that is known...       0.0
1  Each product is microwave safe (without lid), ...       0.0
2  A perfect gift for all occasions, be it your m...     131.0
3  Multipurpose container with an attractive desi...      27.0
4  Nivea Creme Soft Soap gives your skin the best...       0.0
5  Stay protected from contamination with Multipu...      30.0
6  Satinance multani matti is an excellent skin t...       0.0
7  70%Alcohol based is gentle of hand leaves skin...       0.0
8  An exclusive blend with Vitamin B7 Biotin, Hyd...       0.0
9  Scotch Brite Anti- Bacterial Scrub Pad thoroug...       0.0
```

Step 7: Find out the Missing Values from the Dataset

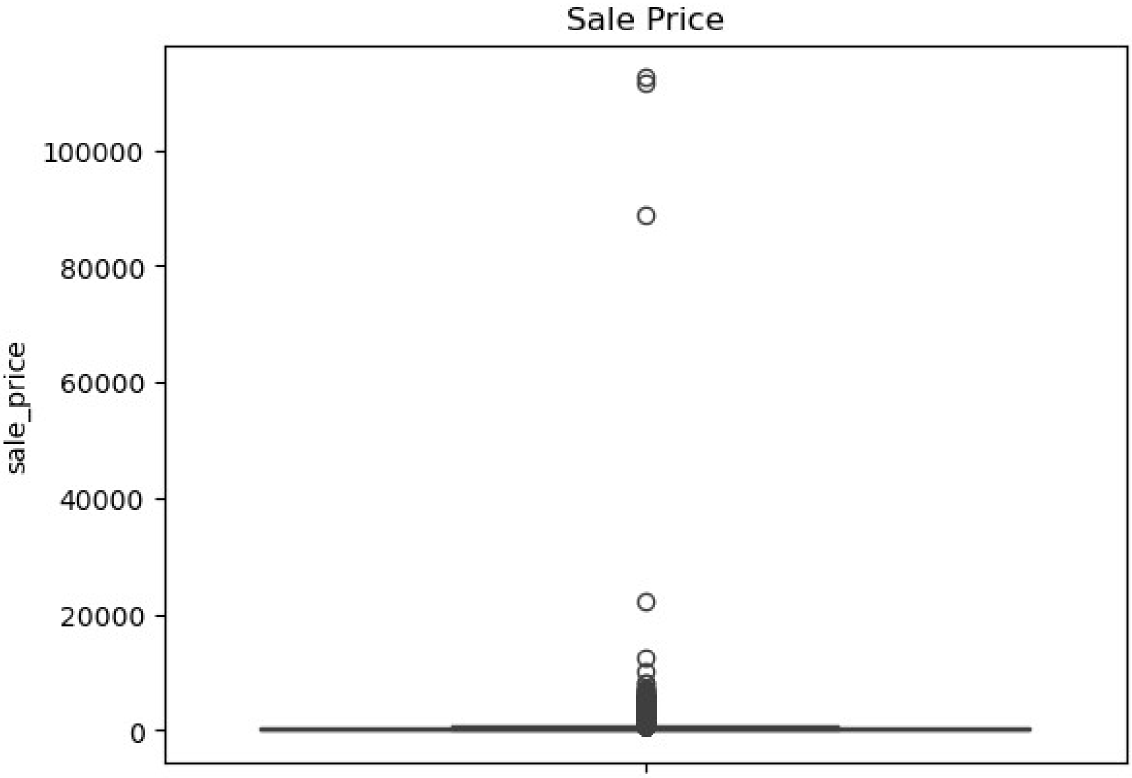
```
data.isnull().sum()
```

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

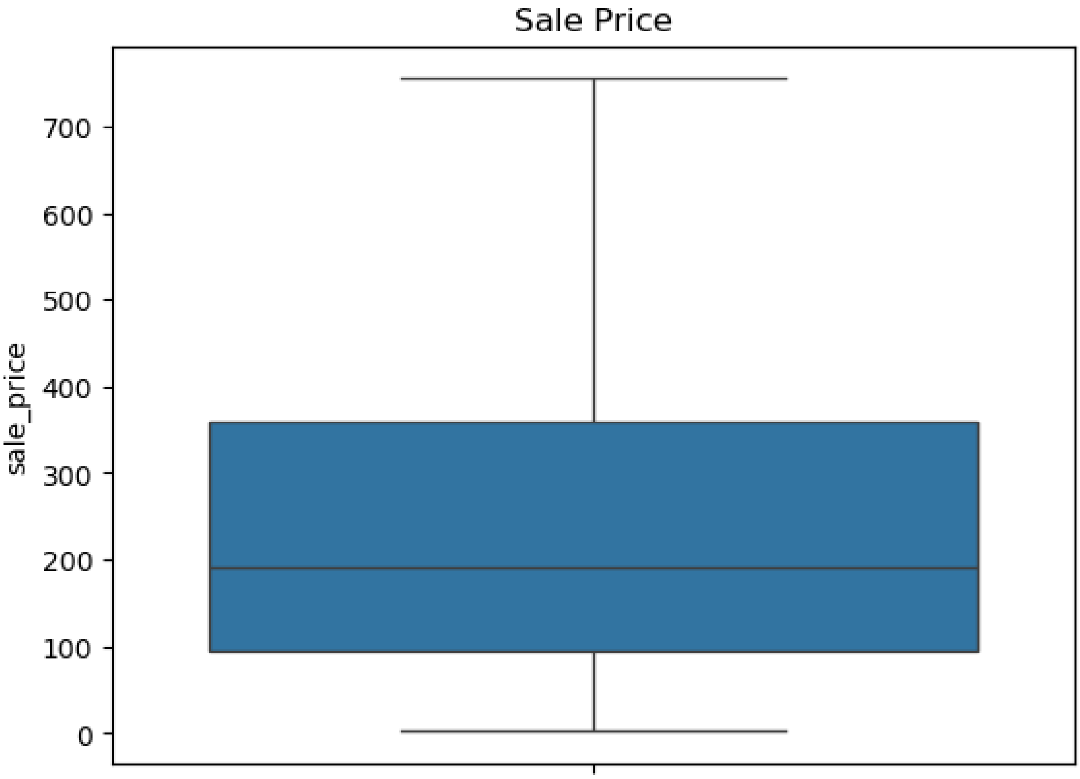
Step 8: Find out the outliers from the dataset according to the columns and fill Null with the mean.

```
##To find outliers we use IQR method

##IQR
q1 = data["sale_price"].quantile(0.25)
q3 = data["sale_price"].quantile(0.75)

iqr = q3-q1

ul = q3+1.5*iqr
```

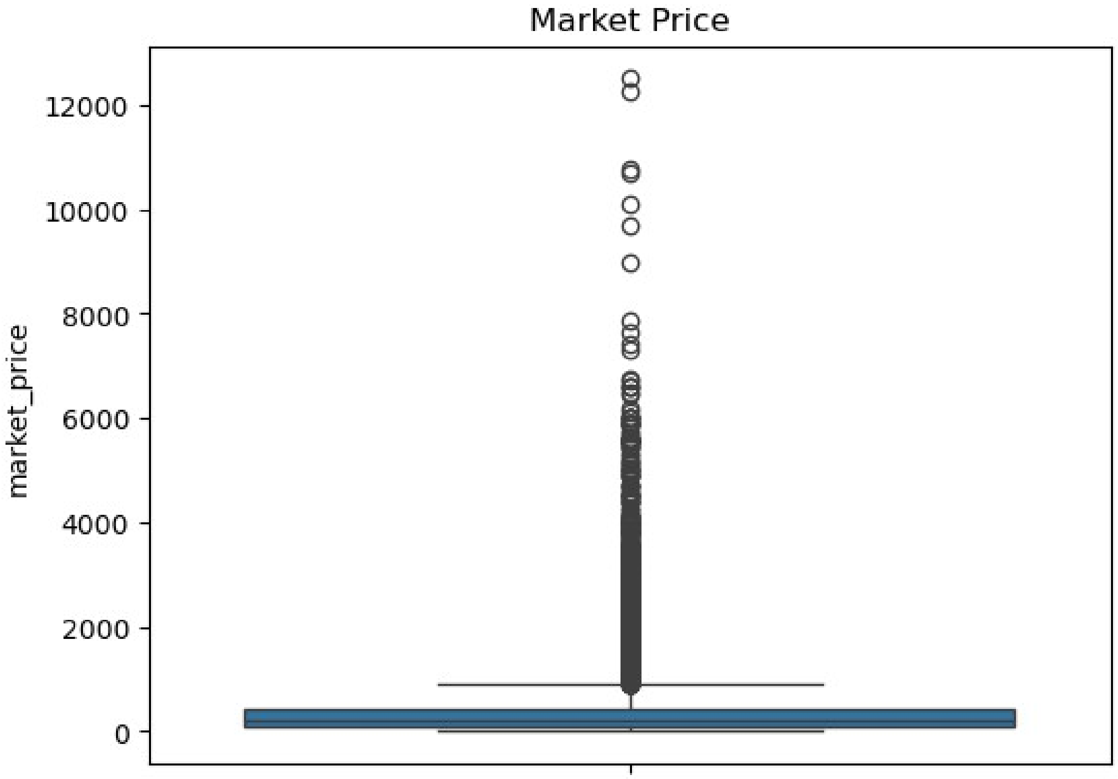
```
ll = q1-1.5*iqr



sns.boxplot(data["sale_price"])
plt.title("Sale Price")
plt.show()
```



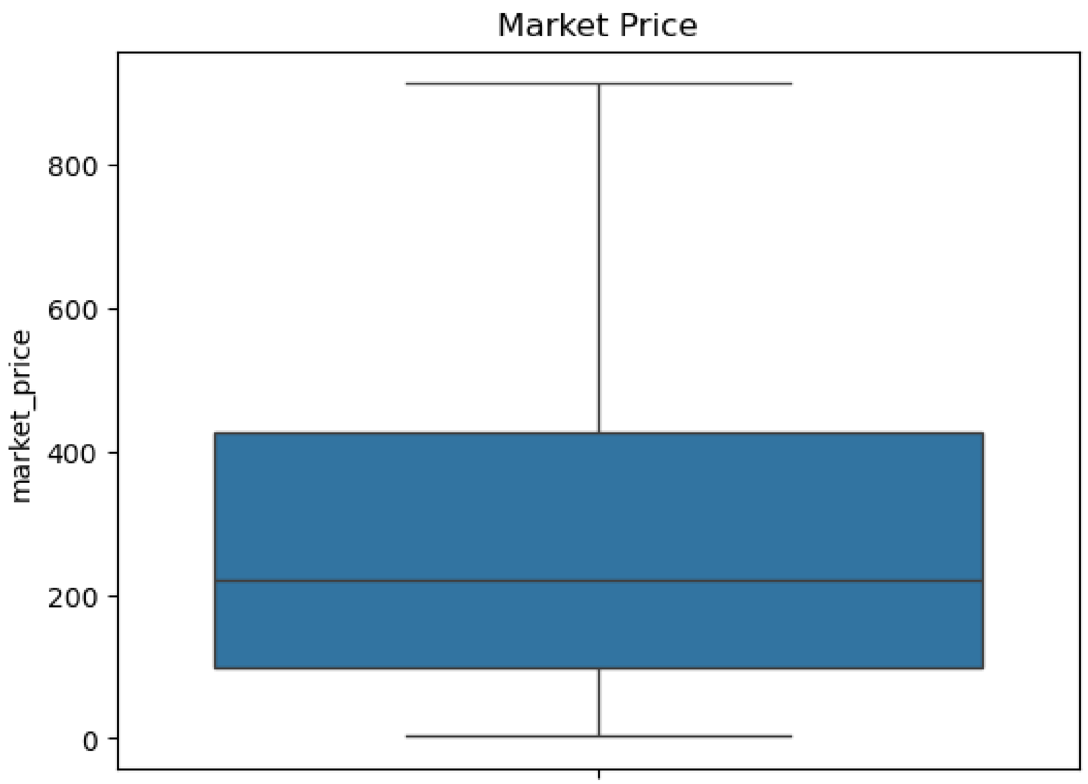
```
##setting upper limit and lower limit values in dataset
data["sale_price"] = np.where(data["sale_price"]>ul,ul,
                              np.where(data["sale_price"]<ll,ll,
                              data["sale_price"]))

##Outliers setteled in upper limit and lower limit
sns.boxplot(data["sale_price"])
plt.title("Sale Price")
plt.show()
```



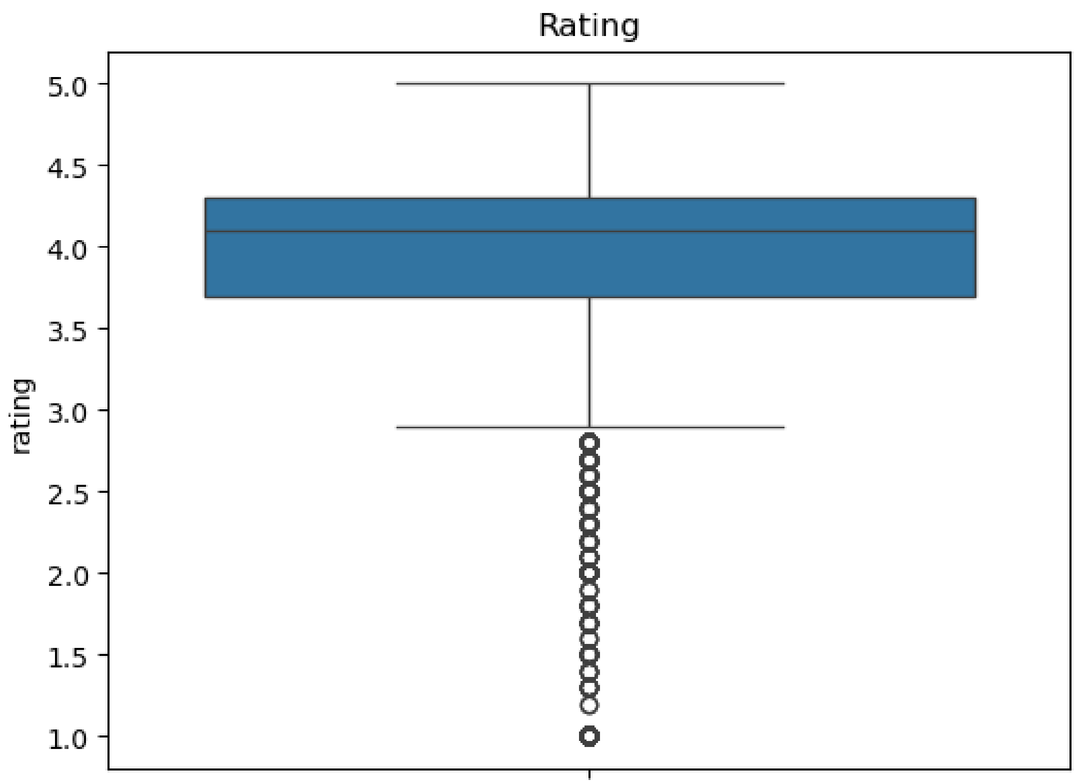
```
##IQR
q1 = data["market_price"].quantile(0.25)
q3 = data["market_price"].quantile(0.75)

iqr = q3-q1

ul = q3+1.5*iqr
ll = q1-1.5*iqr



sns.boxplot(data["market_price"])
plt.title("Market Price")
plt.show()
```



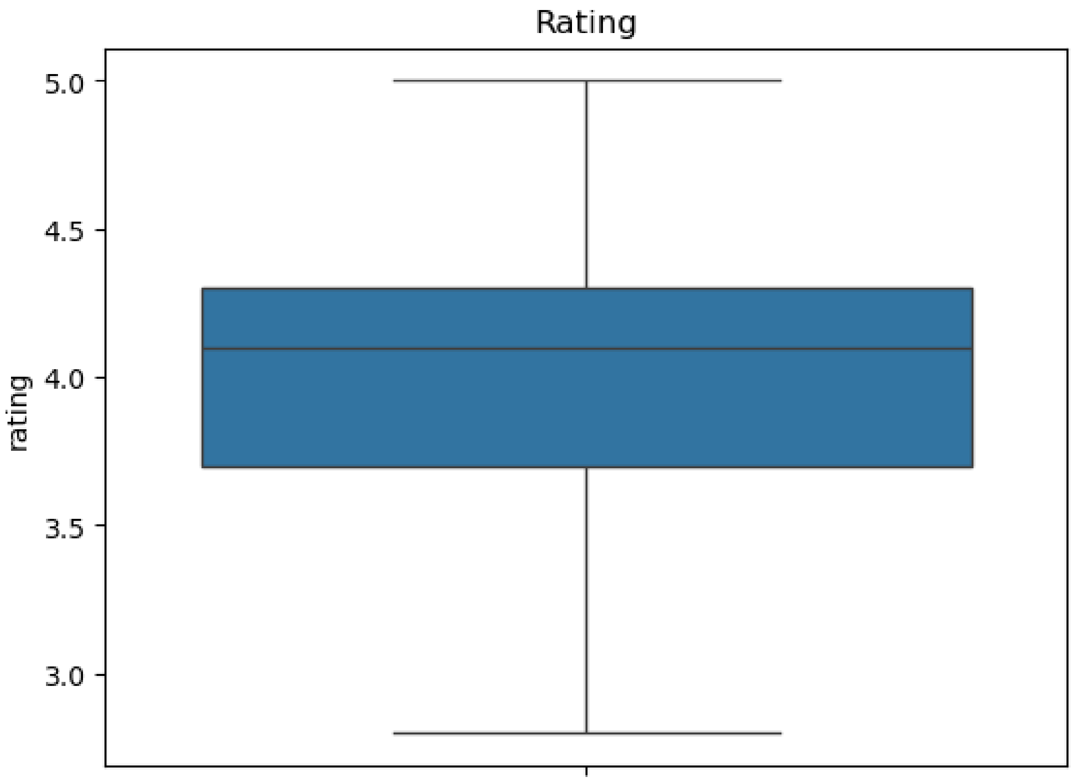
```
##setting upper limit and lower limit values in dataset
data["market_price"] = np.where(data["market_price"]>ul,ul,
                                np.where(data["market_price"]<ll,ll,
                                data["market_price"]))

##Outliers setteled in upper limit and lower limit
sns.boxplot(data["market_price"])
plt.title("Market Price")
plt.show()
```



```python
q1 = data["rating"].quantile(0.25)
q3 = data["rating"].quantile(0.75)

iqr = q3-q1

ul = q3+1.5*iqr
ll = q1-1.5*iqr


sns.boxplot(data["rating"])
plt.title("Rating")
plt.show()
```



```
##setting upper limit and lower limit values in dataset
data["rating"] = np.where(data["rating"]>ul,ul,
                             np.where(data["rating"]<ll,ll,
                             data["rating"]))

##Outliers setteled in upper limit and lower limit
sns.boxplot(data["rating"])
plt.title("Rating")
plt.show()
```



```
##Check Missing values in Dataset
data.isnull().sum()

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

##Handling Missing Values with mean,median,mode

data["sale_price"] =
data["sale_price"].fillna(data["sale_price"].mean())

data["market_price"] =
data["market_price"].fillna(data["market_price"].mean())

data["rating"] = data["rating"].fillna(data["rating"].median())
```

```
data["product"] = data["product"].fillna(data["product"].mode()[0]) ##

data["brand"] = data["brand"].fillna(data["brand"].mode()[0])

data.isnull().sum()
```

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

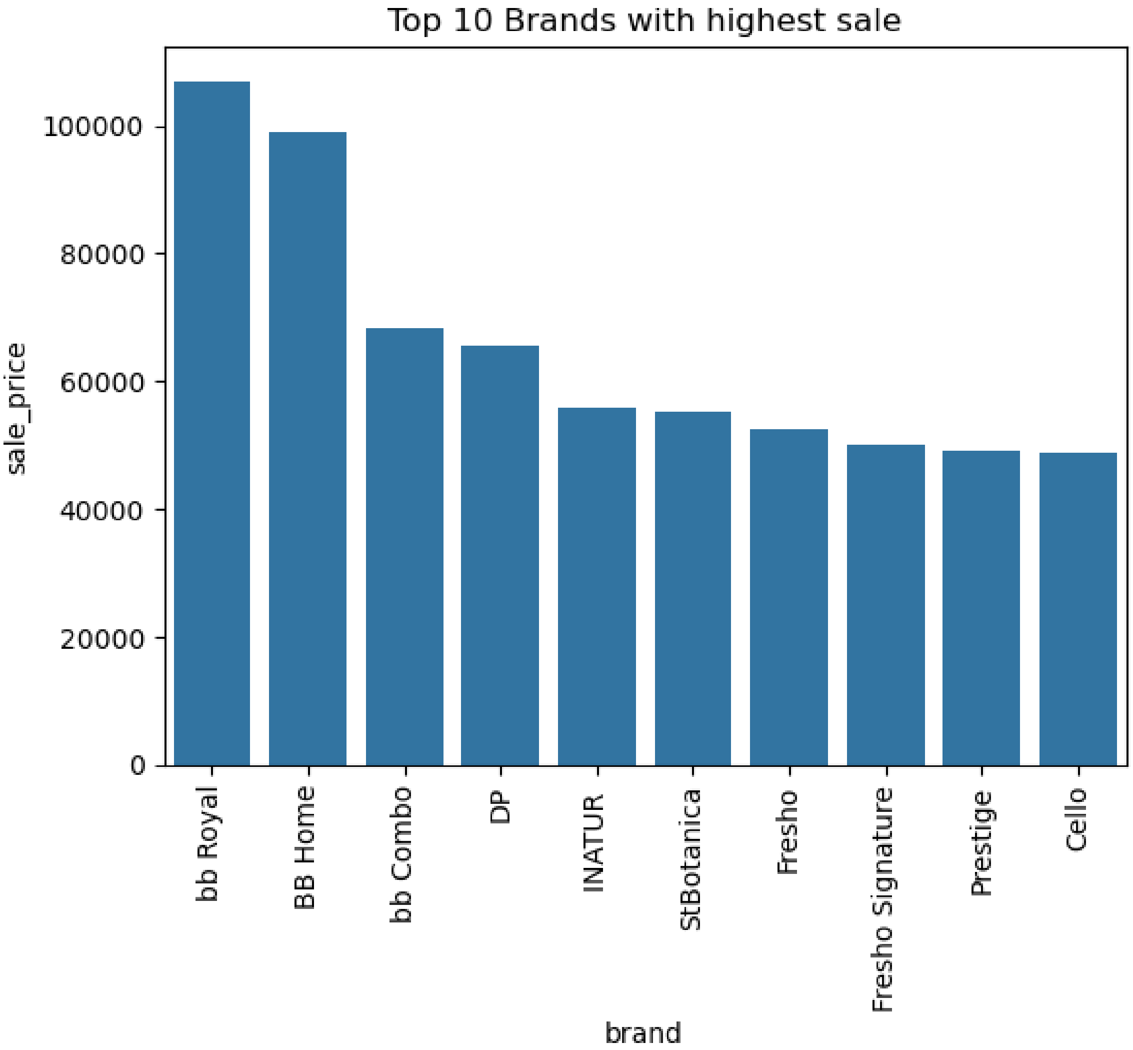
Step 9: Create Plots or visualizations.

1. 10 Highest-selling brands by sale price.

```
x = data.groupby(["brand"]).sum()

top_brands = x.sort_values(by = "sale_price", ascending = False)
print("Top 10 sold brands \n", top_brands[["sale_price"]].head(10))
```

```
Top 10 sold brands
                       sale_price
brand
bb Royal            106881.030000
BB Home              98884.333321
bb Combo             68201.950000
DP                   65560.410000
INATUR               55709.000000
StBotanica           55295.000000
Fresho               52571.810000
Fresho Signature     49910.520000
Prestige             49099.000000
Cello                48932.100000
```

```
sns.barplot(top_brands.head(10), x = "brand", y = "sale_price")
plt.title("Top 10 Brands with highest sale")
plt.xticks(rotation = 90)
plt.show()
```

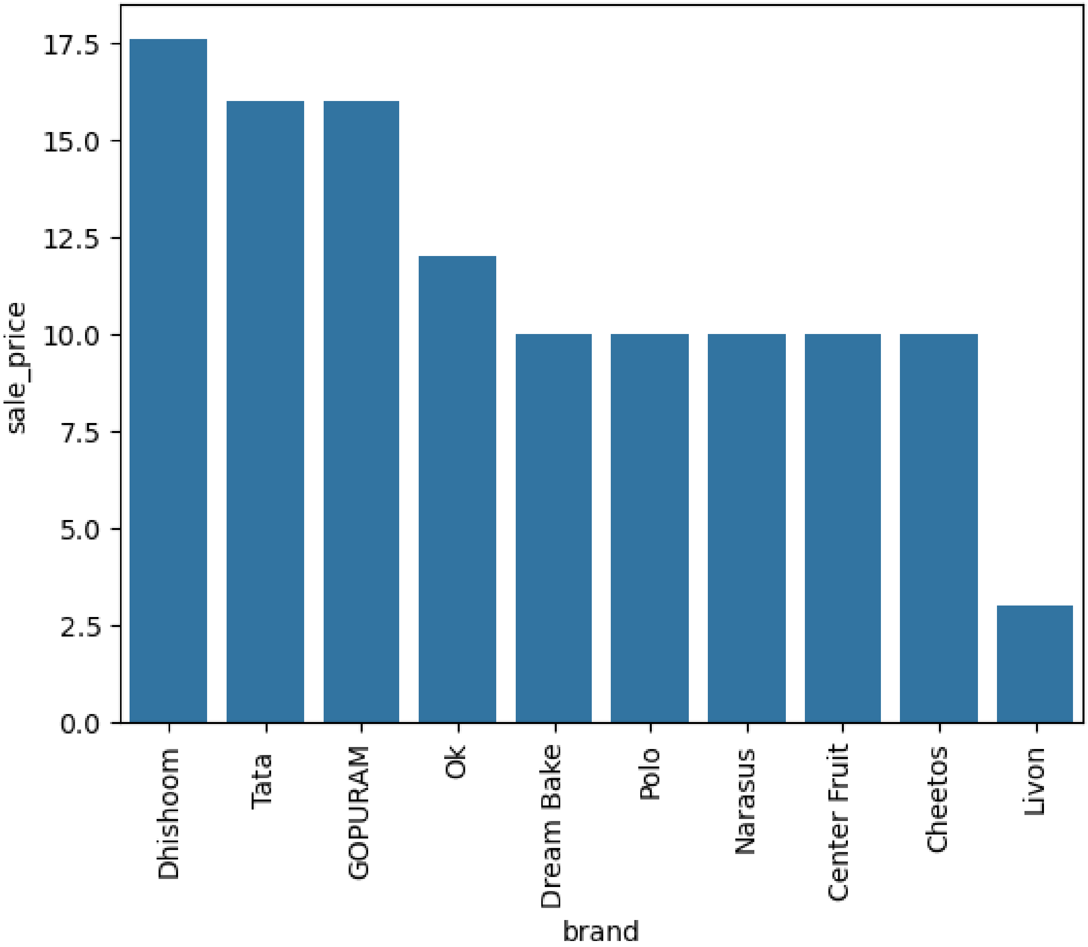


2. 10 lowest-selling brands by sale price

```
bottom_brands = x.sort_values(by = "sale_price", ascending = False)
print("Bottom 10 sold brands \
n",bottom_brands[["sale_price"]].tail(10))
```

```
Bottom 10 sold brands
               sale_price
brand
Dhishoom             17.6
Tata                 16.0
GOPURAM              16.0
Ok                   12.0
Dream Bake           10.0
Polo                 10.0
Narasus              10.0
Center Fruit         10.0
```

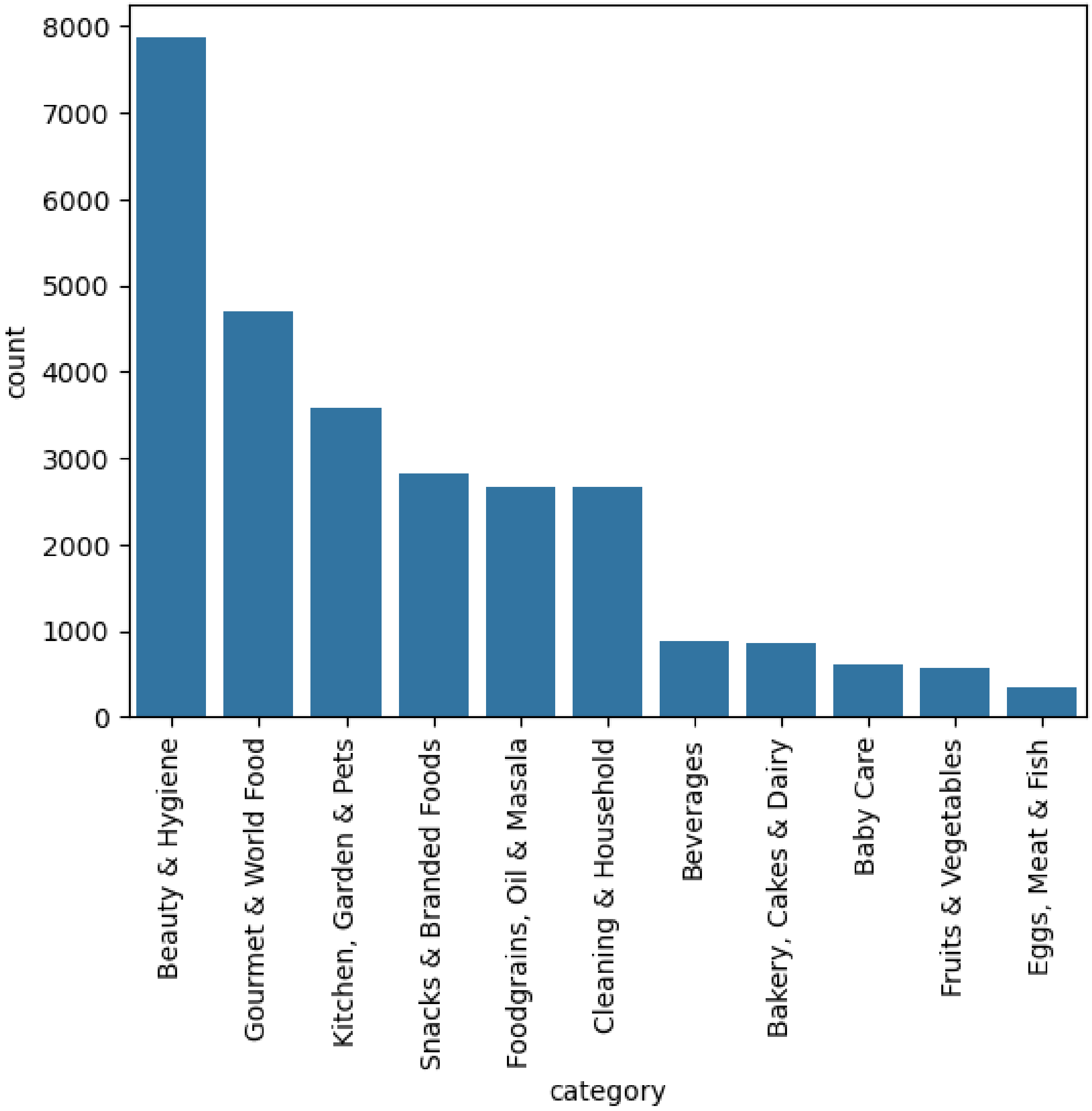
```
Cheetos                   10.0
Livon                      3.0

sns.barplot(bottom_brands.tail(10), x = "brand", y = "sale_price")
plt.xticks(rotation = 90)
plt.show()
```



3. How many products were sold in each category?

```
a = pd.DataFrame(data["category"].value_counts())
a

                             count
category
Beauty & Hygiene              7867
Gourmet & World Food          4690
Kitchen, Garden & Pets        3580
Snacks & Branded Foods        2814
Foodgrains, Oil & Masala      2676
Cleaning & Household          2675
```

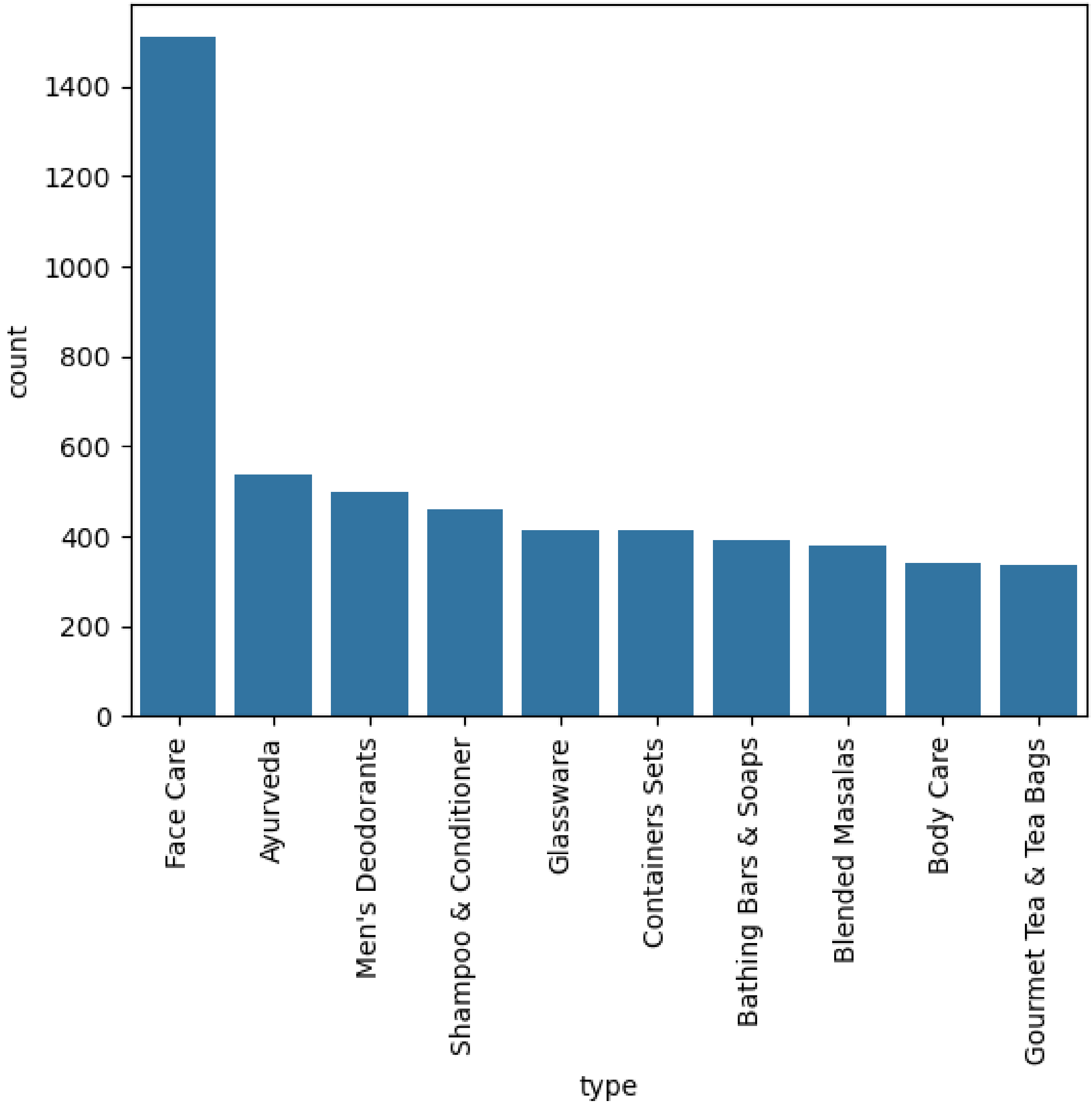
```
Beverages                     885
Bakery, Cakes & Dairy         851
Baby Care                     610
Fruits & Vegetables           557
Eggs, Meat & Fish             350

sns.barplot(a, x = a.index, y = "count")
plt.xticks(rotation = 90)
plt.show()
```



4. Which type of product was sold the most?

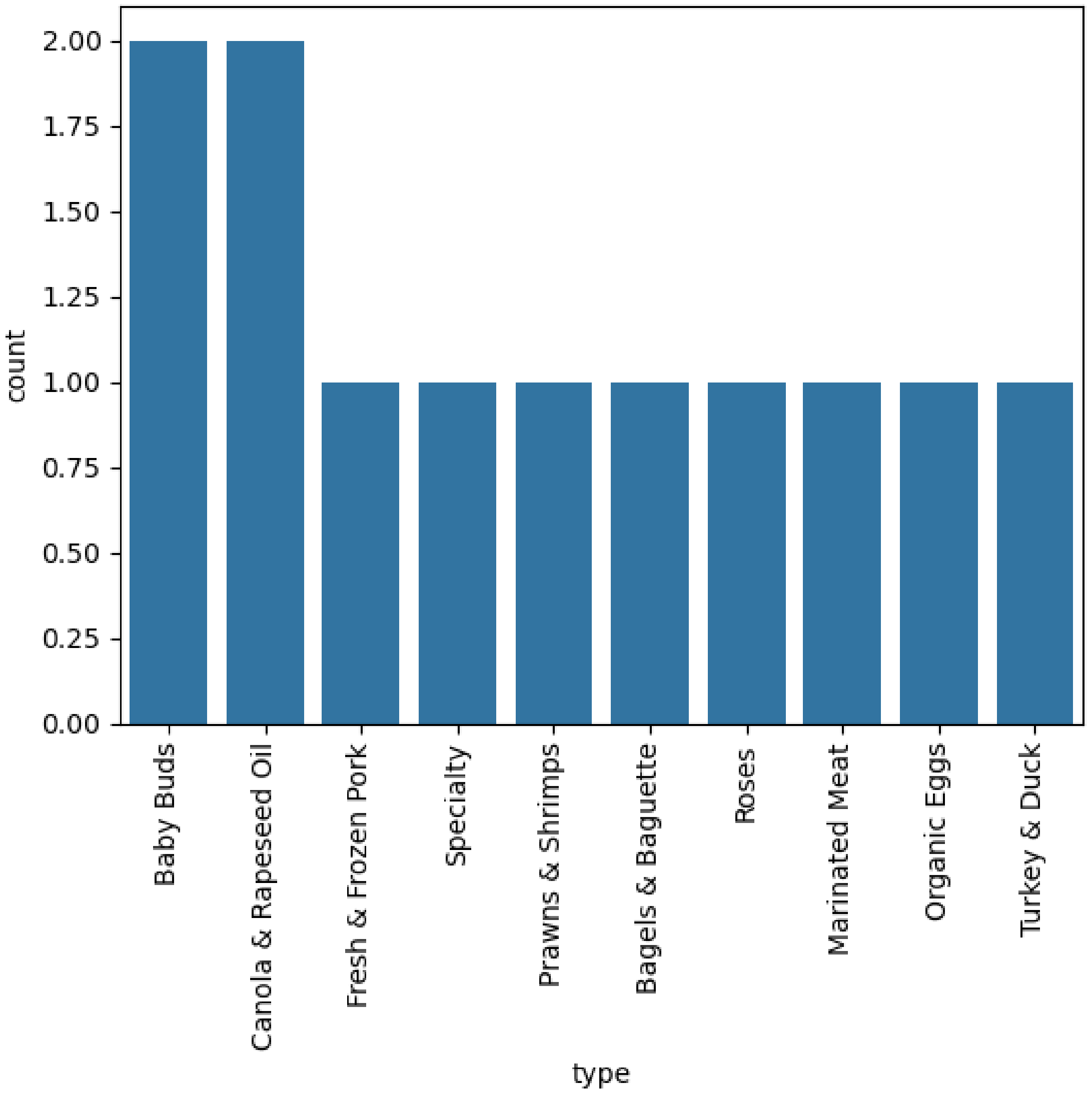
```
b = pd.DataFrame(data["type"].value_counts()).head(10)
```

```
sns.barplot(b, x = b.index, y = "count")
plt.xticks(rotation = 90)
plt.show()
```



5. Which type of product was sold the least?

```
c = pd.DataFrame(data["type"].value_counts()).tail(10)

sns.barplot(c, x = c.index, y = "count")
plt.xticks(rotation = 90)
plt.show()
```


2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
count
Baby Buds
Canola & Rapeseed Oil
Fresh & Frozen Pork
Specialty
Prawns & Shrimps
Bagels & Baguette
Roses
Marinated Meat
Organic Eggs
Turkey & Duck
type