

Big Basket Products Data Analysis

An in-depth look into product sales, pricing, and distribution within the Big Basket dataset.



Project Overview

- Analyze Big Basket's product dataset to gain insights into pricing, discounts, and product popularity.
- Clean and process the data for better visualization and understanding.
- Identify top and least sold products.
- Understand discount distribution and rating patterns.

Dataset Overview

- Total rows: 27,555
- Key columns: Product, Category, Sub-category, Brand, Sale Price, Market Price, Rating
- Missing values in: product, brand, sale price, rating, description

df.head(12)

	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...

Data Cleaning

- Handled missing values (filled or removed)
- Calculated discount: $\text{discount} = \text{market price} - \text{sale price}$
- Treated outliers using IQR method
- Saved cleaned dataset

```
import statistics
numerical_columns = ['sale_price', 'market_price', 'rating']

for col in numerical_columns:
    Q1 = df[col].quantile(0.25)
    Q3 = df[col].quantile(0.75)
    IQR = Q3 - Q1
    lower_bound = Q1 - 1.5 * IQR
    upper_bound = Q3 + 1.5 * IQR
    mean_value = df[(df[col] >= lower_bound) & (df[col] <= upper_bound)][col].mean()
    df[col] = df[col].apply(lambda x: mean_value if x < lower_bound or x > upper_bound else x)

df.to_csv('cleaned_dataset.csv', index=False)
```

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

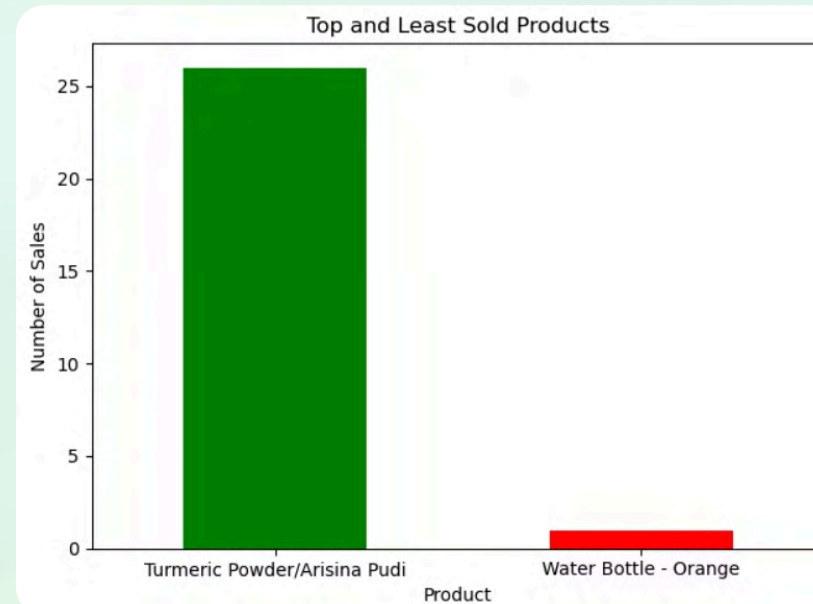
```
index      0
product    1
category   0
sub_category 0
brand      1
sale_price 6
market_price 0
type       0
rating     8636
description 115
discount   6
dtype: int64
```

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

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


Product Popularity

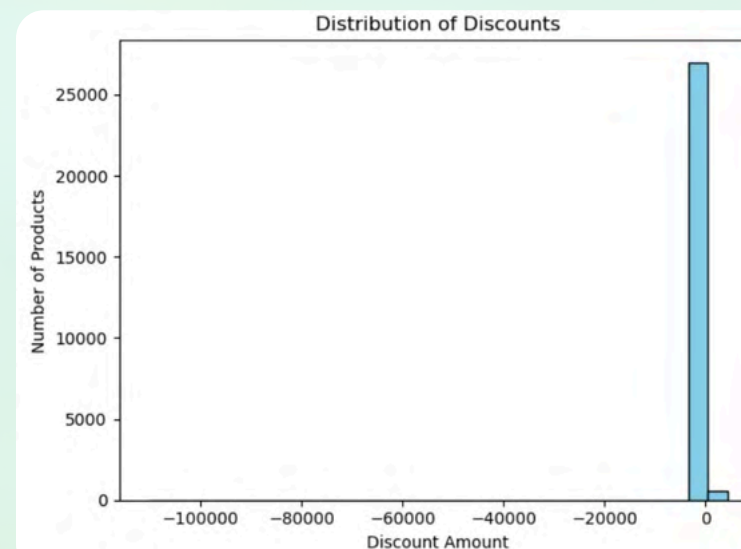
- Top Sold Product: Turmeric Powder (sold 26 times)
- Least Sold Product: Water Bottle - Orange (sold once)



```
combin = pd.concat([top_sold, least_sold])
combin.plot(kind='bar', color=['green', 'red'])
plt.title('Top and Least Sold Products')
plt.ylabel('Number of Sales')
plt.xlabel('Product')
plt.xticks(rotation=0)
plt.tight_layout()
plt.show()
```

Discount Insights

- Histogram plotted to visualize how discounts are distributed
- Many products have no discount; some have very high discounts

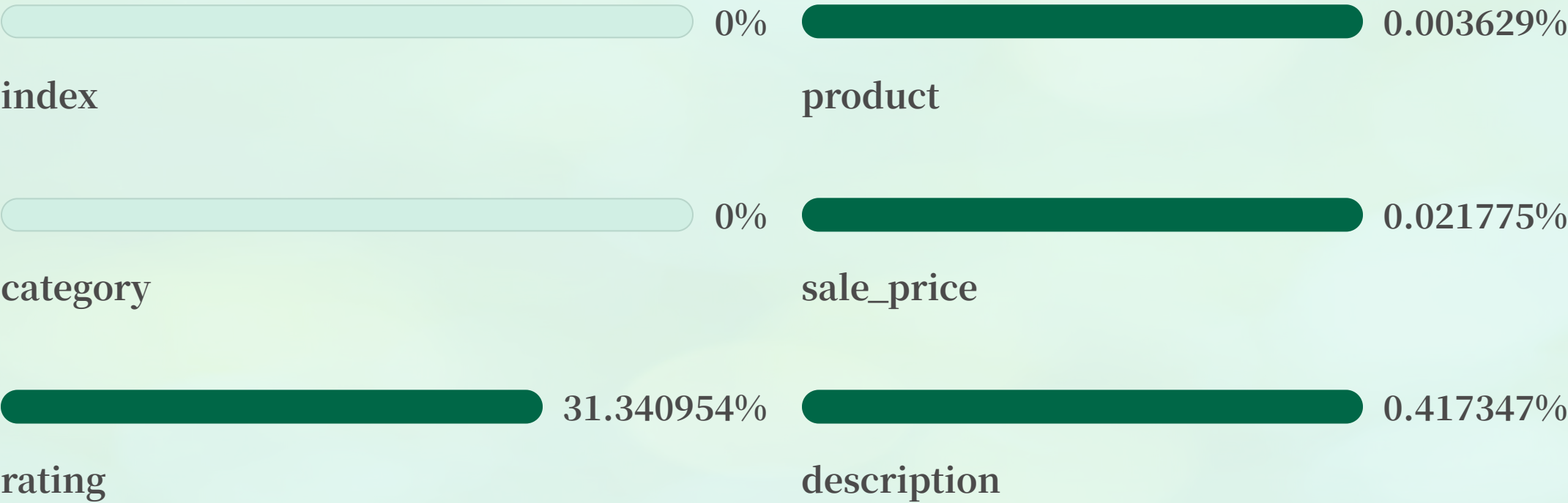


```
df['discount'] = df['market_price'] - df['sale_price']

plt.hist(df['discount'].dropna(), bins=30, color='skyblue', edgecolor='black')
plt.title('Distribution of Discounts')
plt.xlabel('Discount Amount')
plt.ylabel('Number of Products')
plt.tight_layout()
plt.show()
```

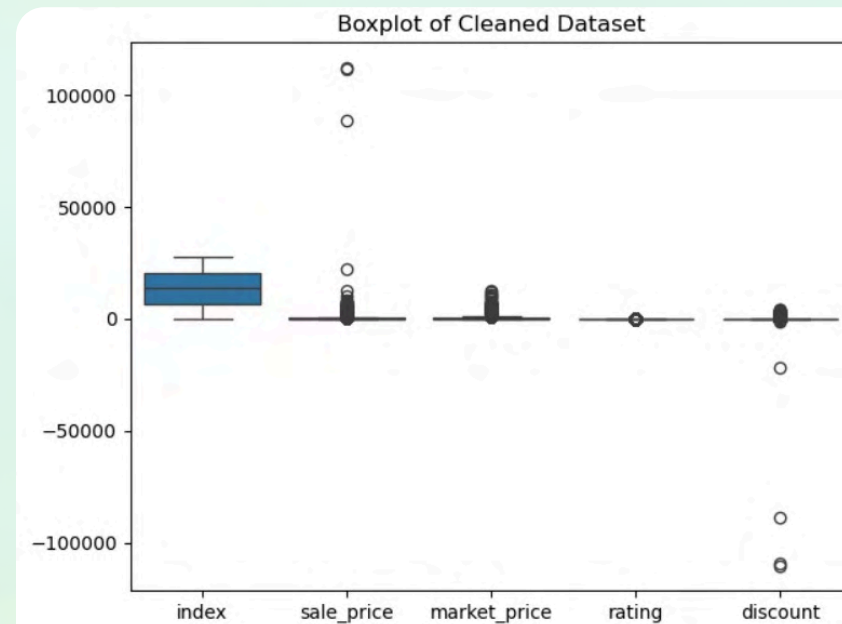
Missing Values Analysis

An overview of missing data percentages for each column, highlighting areas for data cleaning.



Outlier Detection

- Boxplot shows outliers in pricing and ratings.
- Helped identify extreme values before data cleaning.



```
sns.boxplot(data=df)
plt.title("Boxplot of Cleaned Dataset")
plt.xticks(rotation=0)
plt.tight_layout()
plt.show()
```


Cleaned Dataset Sample

The dataset after cleaning, showing the first few rows with adjusted values.

0	Garlic Oil - Vegetarian Capsule 500 mg	Beauty & Hygiene	Hair Care	220.0	220.0	Hair Oil & Serum	4.1
1	Water Bottle - Orange	Kitchen, Garden & Pets	Storage & Accessories	180.0	180.0	Water & Fridge Bottles	2.3
2	Brass Angle Deep - Plain, No.2	Cleaning & Household	Pooja Needs	119.0	250.0	Lamp & Lamp Oil	3.4

Conclusion & Next Step

This analysis provides a foundational understanding of the Big Basket product data. Future steps could involve more in-depth categorical analysis and predictive modeling.

Key Takeaways

- Most products have low ratings missing.
- Top-selling products can help drive marketing focus.
- Discounts vary widely and influence price positioning.

Future Work

- Deep dive into category-wise sales trends.
- Analyze customer ratings impact on sales.
- Develop a price optimization model.