

BIG BASKET ANALYSIS

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.

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 Blinkit etc. but Big Basket has still not lost anything - thanks to ever expanding popular base and their shift to online buying.



- By AKSHAY GUJARE

Project Objectives

Data Upload

Upload the dataset in pandas for analysis and cleaning.

Data Cleaning

Identify and remove records with missing or null values for data integrity.

Exploratory Analysis

Perform EDA to Understand dataset patterns and characteristics

Bussiness Insights

Answers specific questions using pandas

Create Visualization

Provided Insights using PowerBI visualization tool

Key Business Queries

- Get Description of the data in the DataFrame.
- Find Information about the DataFrame.
- Find out Top & least sold products.
- Measuring discount on a certain item.
- Find out the Missing Values from the Dataset.
- Find out the outliers from the dataset according to the columns and fill them with the mean.
- Create visualizations(Using PowerBI)

Dataset : bigbasket

Range Index :

Data Rows : 27555 entries, 0 to 2755

Data Columns : Total 10.

The screenshot shows a Google Colab notebook titled "BigBasket.ipynb". The code cell [5] contains Python code to read an Excel file from Google Drive:

```
1 #Now we are going to fetch the file from Google Drive
2 import numpy as np           #Import Library
3 import pandas as pd
4 df=pd.read_excel("/content/BigBasket Products.xlsx")
```

The output cell [7] displays the first 12 rows of the DataFrame:

```
1 df.head(12)  #Here we are going to see First 12 Rows
```

		index	product	category	sub_category	brand	sale_price	market_price
0	1	1	Garlic Oil - Vegetarian Capsule 500 mg	Beauty & Hygiene	Hair Care	Sri Sri Ayurveda	220.0	220.0
1	2	2	Water Bottle - Orange	Kitchen, Garden & Pets	Storage & Accessories	Mastercook	180.0	180.0
2	3	3	Brass Angle Deep - Plain, No.2	Cleaning & Household	Pooja Needs	Trm	119.0	250.0
3	4	4	Cereal Flip Lid Container/Storage Jar - Assort...	Cleaning & Household	Bins & Bathroom Ware	Nakoda	149.0	176.0
4	5	5	Creme Soft Soap - For Hands & Body	Beauty & Hygiene	Bath & Hand Wash	Nivea	162.0	162.0
5	6	6	Germ - Removal Multipurpose Wipes	Cleaning & Household	All Purpose Cleaners	Nature Protect	169.0	199.0
6	7	7	Multani Mati	Beauty & Hygiene	Skin Care	Satinance	58.0	58.0
7	8	8	Hand Sanitizer - 70% Alcohol Base	Beauty & Hygiene	Bath & Hand Wash	Bionova	250.0	250.0
8	9	9	Biotin & Collagen Volumizing Hair Shampoo + Ri	Beauty & Hygiene	Hair Care	StBotanica	1098.0	1098.0

The last row (index 8, product 9) has a note: "Activate Windows Go to Settings to activate Windows."

At the bottom, the status bar shows "Variables" and "Terminal" icons, the time "10:10PM", and the Python version "Python 3".



Commands

+ Code

+ Text

▶ Run all

RAM

Disk

9	Biotin & Collagen Volumizing Hair Shampoo + Bi...	Beauty & Hygiene	Hair Care	StBotanica	1098.0	1098.0	Sharr Cond		
10	Scrub Pad - Anti-Bacterial, Regular	Cleaning & Household	Mops, Brushes & Scrubs	Scotch brite	20.0	20.0	Scrub		
11	Wheat Grass Powder - Raw	Gourmet & World Food	Cooking & Baking Needs	NUTRASHIL	261.0	290.0	Fl Pre		
12	Butter Cookies Gold Collection	Gourmet & World Food	Chocolates & Biscuits	Sapphire	600.0	600.0	Choc		

Next steps: [Generate code with df](#) [New interactive sheet](#)

[8]

```
1 df.describe() #Get Description of the data in the DataFrame.  
2 #Describe will show you only Numerical columns
```

	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	1.00000	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

[9]

```
1 df.info() # Information about the DataFrame.
```

```
2 <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
```

Activate Windows
Go to Settings to activate
Windows.





Commands

+ Code

+ Text

Run all

RAM

Disk

[27]
✓ 0s

```
1 #Top and Least Sold Product
2 product_counts = df.groupby('product').size().reset_index(name='product_count')
3 Final_Count=product_counts.sort_values('product_count', ascending=False)
4 display(Final_Count)
```



product product_count

22257	Turmeric Powder/Arisina Pudi	26
7194	Extra Virgin Olive Oil	15
5365	Cow Ghee/Tuppa	14
14502	Olive Oil - Extra Virgin	12
19971	Soft Drink	12
...
8416	Germ Protection Hand Wash Pump	1
8414	Germ Protection Fruit & Veggie Wash âœœ Natura...	1
8413	Germ Protection Dish Wash Liquid Gel âœœ Remov...	1
8412	Germ Protection Dish Wash Liquid Gel , Fragran...	1
8424	Get Slim Ayurvedic Tea - 7 Active Ingredients	1

23540 rows × 2 columns

Next steps: [Generate code with Final_Count](#)[New interactive sheet](#)

[]

1 #Here we Find out Top & least sold products

[34]

```
1 #Now we are Measuring discount on a certain item.
2 Discount=df.groupby('product').agg({'market_price':'sum','sale_price':'sum'})
3 Discount['Discount']=Discount['market_price']-Discount['sale_price']
4 display(Discount)
```



market_price sale_price Discount

product

& Moms - Citrus Soap	195.0	195.00	0.00
& Moms - Sandal Soap	120.0	108.00	12.00
& Moms Bathing Soap - Jasmine	120.0	108.00	12.00
& Moms Bathing Soap - Tulsi	120.0	108.00	12.00
0.0 Non Alcoholic Beer	75.0	56.25	18.75
...
oriental Instant Noodles - Curry Flavour	40.0	36.00	4.00
oriental Instant Noodles - Masala Flavour	40.0	36.00	4.00
oriental Instant Noodles - Tom Yam Flavour	40.0	36.00	4.00
oriental Instant Noodles - Tomato Flavour	40.0	36.00	4.00
pasta shell	70.0	70.00	0.00

23540 rows × 3 columns

Activate Windows
Go to Settings to activate
Windows.

BigBasket.ipynb

File Edit View Insert Runtime Tools Help

Commands + Code + Text Run all

RAM Disk

[35] ✓ 0s

```
1 # We are Going to find the "Missing Values" from the Dataset.
2 df.isnull().sum()
```

	0
index	0
product	1
category	0
sub_category	0
brand	1
sale_price	6
market_price	0
type	0
rating	8636
description	115

dtype: int64

[36]

```
1 #We are going to find Outliers
2 Q1=df["sale_price"].quantile(0.25)
3 Q3=df["sale_price"].quantile(0.75)
4 IQR=Q3-Q1
5 lower_bound=Q1-1.5*IQR
6 upper_bound=Q3+1.5*IQR
7 outliers=df[(df["sale_price"]<lower_bound)|(df["sale_price"]>upper_bound)]
8 print(outliers[["product", "sale_price"]])
```

	product	sale_price
8	Biotin & Collagen Volumizing Hair Shampoo + Bi...	1098.00
47	Colour Catcher Sheets	799.00
51	Peach Syrup	850.00
69	Padded Harness - 3/4 inch, Grey Colour	840.00
91	Hard Anodised Ezee-Pour Saucepan With Lid - L88	864.50
...
27498	Juicer - Fruit & Vegetable, Light Green	1071.00
27505	Virgin Coconut Oil	875.00
27514	Verge & Sheer Perfume For Pair	1615.50
27538	Quista Pro Advanced Whey Protein Formula forti...	4500.00
27542	Embellish Skin Lightening Serum	1525.75

[2205 rows x 2 columns]

```
[38]
1 Q1=df["sale_price"].quantile(0.25)
2 Q3=df["sale_price"].quantile(0.75)
3 IQR=Q3-Q1
4 lower_bound=Q1-1.5*IQR
5 upper_bound=Q3+1.5*IQR
6
7 # Calculate mean of sale_price
8 mean_val=df["sale_price"].mean()
9
10 # Replace outliers with mean
11 df["sale_price"] = df["sale_price"].apply(
12     lambda x: mean_val if (x < lower_bound or x > upper_bound) else x)
13 display(df["sale_price"])

[40] 7s
1 #After Manipulation our Dataset we are going to export the file
2 df.to_excel("Cleaned_data_of_BigBasket.xlsx",index=False)

Activate Windows
Go to Settings to activate
Windows.
```

Variables Terminal Python 3

Hi Data Lovers!

I Just completed an end-to-end Data Analysis Project using only Pandas!

My Approach:

- Cleaned & preprocessed the dataset
- Explored the data with 8 Pandas Q&A.
- Derived key insights only through data manipulation, grouping, aggregation & correlation.
- Finally created dashboard and provided few insights

Deliverables:

- Cleaned dataset
- Pandas-only Q&A solutions
- Dashboard
- Business use case recommendations

This project strengthened my data wrangling, exploratory analysis, and business interpretation skills—all with the power of Pandas.

Data Exploration Results

Key Insights:

1. Brand Analysis:

The bar chart compares Market Price, Sale Price, and Discount across major brands like Prestige, BB Home, bb Royal, DP, Ajmal, and Borges, helping identify top-performing and discounted brands.

2. Product Analysis:

Shows pricing trends for product types such as Olive Oil, Organic, Non-organic, etc., indicating which products have the highest market and sale prices.

3. Category Analysis:

Compares Kitchen & Garden, Gourmet & World Food, Foodgrains & Masala, Snacks, Eggs, and Fruits & Vegetables

highlighting the total market value, sales, and discount distribution across these categories.

4. KPIs & Summary Metrics:

Total Market Price: ₹10.52M ,Total Sale Price: ₹9.22M ,Total Discount: ₹9.22M .These KPI gauges visually track overall sales performance and discount impact relative to the total potential value.

5. Interactive Filters:

Users can filter data by Brand, Category, Product, and adjust sliders for Market Price, Sale Price, and Discount to explore specific data ranges dynamically.

Conclusion:

The exploration offers a clear visual insight into pricing strategy, discount distribution, and category-wise sales performance, enabling data-driven decisions to improve profitability and inventory management.

Big Basket Analysis Dashboard

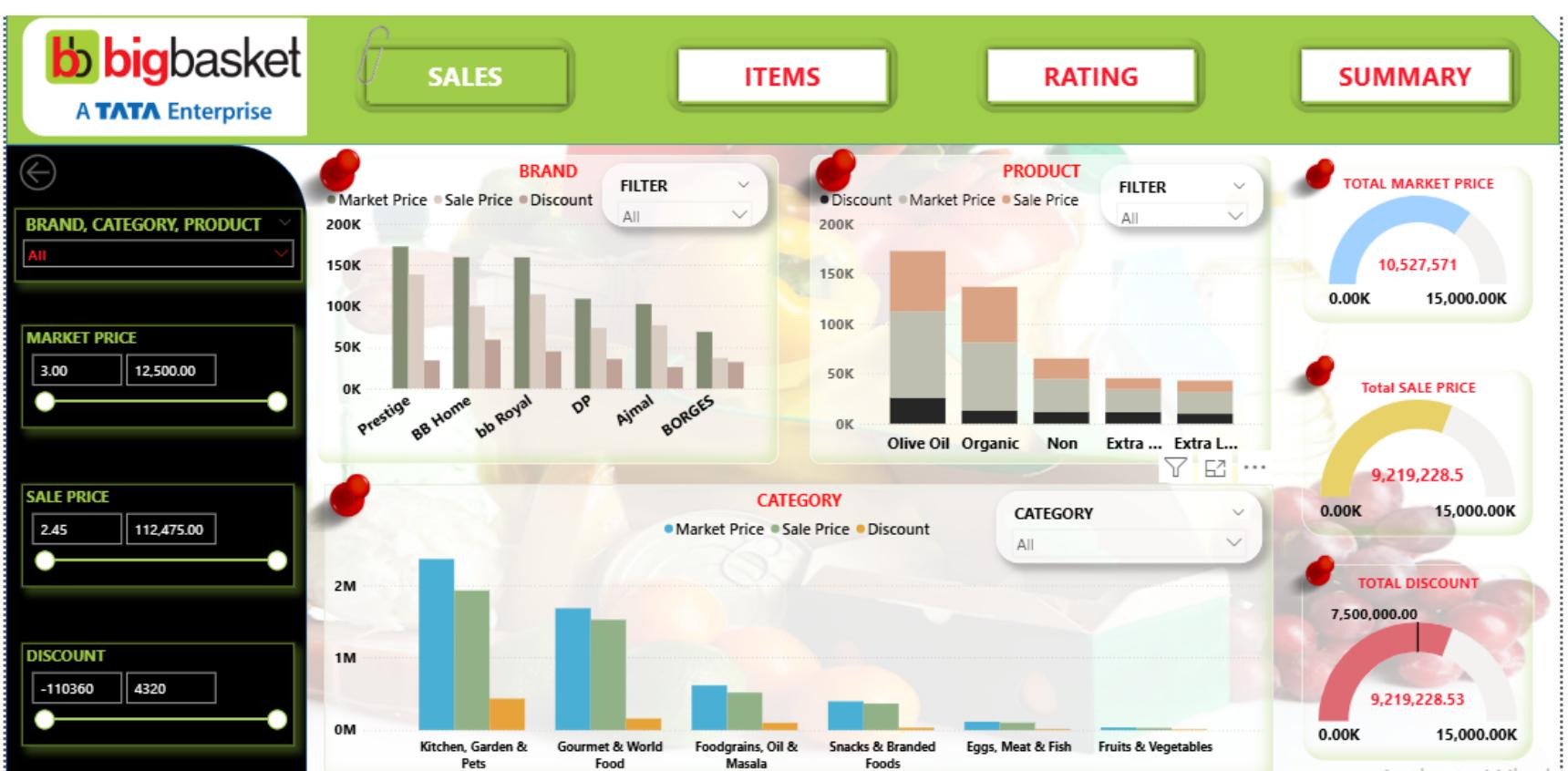
Main Dashboard:



Continue

Activate Window

Sales Dashboard:



Activate Window

Item Dashboard:



SALES **ITEMS** **RATING** **SUMMARY**

11
Total Categories

90
Total Sub Category

2313
Total Brands

15719
TOTAL PRODUCT

SEARCH BRAND		SEARCH PRODUCT			
All		All			
BRAND	CATEGORY	SUB CATEGORY	PRODUCT	APPLICATION	TYPE
1st Bites	Baby Care	Baby Food & Formula	Ragi	Stage-1, 6-24 Months	Organic Baby Food
1st Bites	Baby Care	Baby Food & Formula	Ragi & Dal	Stage-2, 8- 24 Months	Organic Baby Food
Amul	Baby Care	Baby Food & Formula	Spray Infant Milk Food/Substitute	None	Infant Formula
Aptamil	Baby Care	Baby Food & Formula	Follow Up Infant Formula Powder	Stage 2, From 6 To 12 Months	Infant Formula
Aptamil	Baby Care	Baby Food & Formula	Follow Up Infant Formula Powder	Stage-3, From 12 Months Onwards	Infant Formula
Aptamil	Baby Care	Baby Food & Formula	Infant Formula Powder With GOS & FOS	Stage 1, From Birth To 6 Months	Infant Formula
Aveeno	Baby Care	Baby Bath & Hygiene	Baby Bath	Daily Moisturising	Baby Oil & Shampoo
Aveeno	Baby Care	Baby Bath & Hygiene	Baby Creamy Wash	Soothing Relief	Baby Oil & Shampoo
Aveeno	Baby Care	Baby Bath & Hygiene	Baby Moisture Cream Soothing Relief	None	Baby Creams & Lotions
Aveeno	Baby Care	Baby Bath & Hygiene	Baby Moisturising Wash Cleansing Therapy	None	Baby Oil & Shampoo
Aveeno	Baby Care	Baby Bath & Hygiene	Daily Moisturising Baby Lotion	None	Baby Creams & Lotions
Baby Dove	Baby Care	Baby Bath & Hygiene	Rich Moisture Baby Shampoo	None	Baby Oil & Shampoo
Baby Dove	Baby Care	Baby Bath & Hygiene	Rich Moisture Bar	None	Baby Bath
Baby Dove	Baby Care	Diapers & Wipes	Rich Moisture Baby Wipes	None	Baby Wipes
bb Combo	Baby Care	Baby Bath & Hygiene	Pampers Pants Diapers	Medium Size 80pcs + Sebamed Baby Wash - Extra Soft 400ml	Diapers
bb Combo	Baby Care	Diapers & Wipes	Pampers Pants Diapers	Medium Size 80pcs + Sebamed Baby Wash - Extra Soft 400ml	Diapers

Rating Dashboard:



SALES **ITEMS** **RATING** **SUMMARY**

18.92K
RATING COUNT

TYPE		BRAND		CATEGORY											
All		All		All											
bb Royal	DP	Britannia	Iveo	Claycraft	Soulflow...	Ratan	Eas...	La...	Ve...	Saf...	Tru...	Bella	Nu...		
1,125	424	381	342	302	274	247	236	236	228	225	223	223	222		
	Keya	Colgate	Loreal P...	Aroma ...	Nutty ...	Park a...	Lotus ...	Nike	Ph...	Rit...	M...	Se...	O...		
BB Home	414	381	330	300	266	246	246	218	Gala	203	203	202	201	200	197
	BIOTIQUE	Khadi Na...	Sunfeast	Gramin...	Vlcc	Himal...	Dettol	216	Engage	196	MeraK...	Ga...	To...	Al...	Bo...
711	413	359	314	299	257	246	213	The M...	190	189	188	187			
	Cello	24 Mantra	Dabur	Princew...	Pampers	245	245	213	Gillette	196	PRIS...	Ke...	Ar...	Pr...	
Amul	403	357	308	284	255	242	242	Huggies	185	Lakme	183	182	182		
	Organic Ta...	Nivea	Mamae...	Nakoda	Asian	Epiga...	240	Fogg	212	Yardle...	184	Ponds	183	182	Ka...
618	401	351	307	284	252	240	240	Tata S...	204	Parle	183	Saffola	183	Sirona	179
	Himalaya	MTR	INATUR	Anjali	Organic ...	StBota...	248	248	247	236	236	204	Catch	183	
576	394	347	307	275	248	240	240	Tata S...	204	Parle	183				

Summary Dashboard:

bigbasket A TATA Enterprise		SALES	ITEMS	RATING	SUMMARY				
TYPE		BRAND	CATEGORY						
BRAND	CATEGORY	SUB CATEGORY	PRODUCT	APPLICATION	RATING	MARKET PRICE	SALE PRICE	DISCOUNT	TYPE
Cleaning & Household	Disposables, Garbage Bag	Food Package	Medium	50.00	50.00	0	Aluminium Foil, Clingwrap		
&Me Beverages	Health Drink, Supplement	Skin Women's Health Drink	Watermelon & Rose	3.80	85.00	85.00	0	Men & Women	
&Me Beverages	Tea	PCOS Green Tea	Kashmiri Kahwa	4.70	330.00	330.00	0	Green Tea	
&Me Beverages	Tea	PCOS Green Tea	Lavender & Chamomile	4.80	330.00	330.00	0	Green Tea	
&Stirred Beverages	Energy & Soft Drinks	Cosmopolitan	Cranberry & Orange	4.10	50.00	50.00	0	Soda & Cocktail Mix	
&Stirred Beverages	Energy & Soft Drinks	Kamikaze Shots	None	4.10	250.00	250.00	0	Soda & Cocktail Mix	
&Stirred Beverages	Energy & Soft Drinks	Margarita	Lemon & Lime	4.00	50.00	50.00	0	Soda & Cocktail Mix	
&Stirred Beverages	Energy & Soft Drinks	Pina Colada	None	4.10	180.00	180.00	0	Soda & Cocktail Mix	
&Stirred Beverages	Energy & Soft Drinks	Red Sangria	None	3.90	480.00	480.00	0	Soda & Cocktail Mix	
&Stirred Gourmet & World Food	Drinks & Beverages	Cocktail Mix	Whisky Sour	240.00	240.00	0	Cocktail Mixes		
&Stirred Gourmet & World Food	Drinks & Beverages	Hibiscus Cosmo	None	80.00	80.00	0	Cocktail Mixes		
&Stirred Gourmet & World Food	Drinks & Beverages	Mocktails	Classic Colada	80.00	80.00	0	Cocktail Mixes		
&Stirred Gourmet & World Food	Drinks & Beverages	Mocktails	Mexican Mule	80.00	80.00	0	Cocktail Mixes		
&Stirred Gourmet & World Food	Drinks & Beverages	Mocktails	Mountain Mojito	80.00	80.00	0	Cocktail Mixes		
109°F Beauty & Hygiene	Health & Medicine	Face Mask For Men & Women	Washable, Black	3.20	199.00	99.00	100	Face Masks & Safety Gears	
137 Degree Gourmet & World Food	Drinks & Beverages	Milk	Walnut, Original	3.50	445.00	445.00	0	Health Drinks	
18 Herbs Beauty & Hygiene	Hair Care	Organics Hair Oil	Karippan Thailam	4.70	1,010.00	1,010.00	0	Hair Oil & Serum	
18 Herbs Beauty & Hygiene	Hair Care	Organics Stress Relief Oil	Asuvagendhi Thailam	580.00	580.00	0	Hair Oil & Serum		
18 Herbs Beauty & Hygiene	Skin Care	Organics Skin Therapy Oil	Ponnanganni Thailam	240.00	240.00	0	Body Care		
18 Herbs Beverages	Fruit Juices & Drinks	Organics Noni Fruit Fresh	None	500.00	500.00	0	Juices		
18 Herbs Beverages	Health Drink, Supplement	Organics Herbo Protein	None	3.00	300.00	300.00	0	Men & Women	