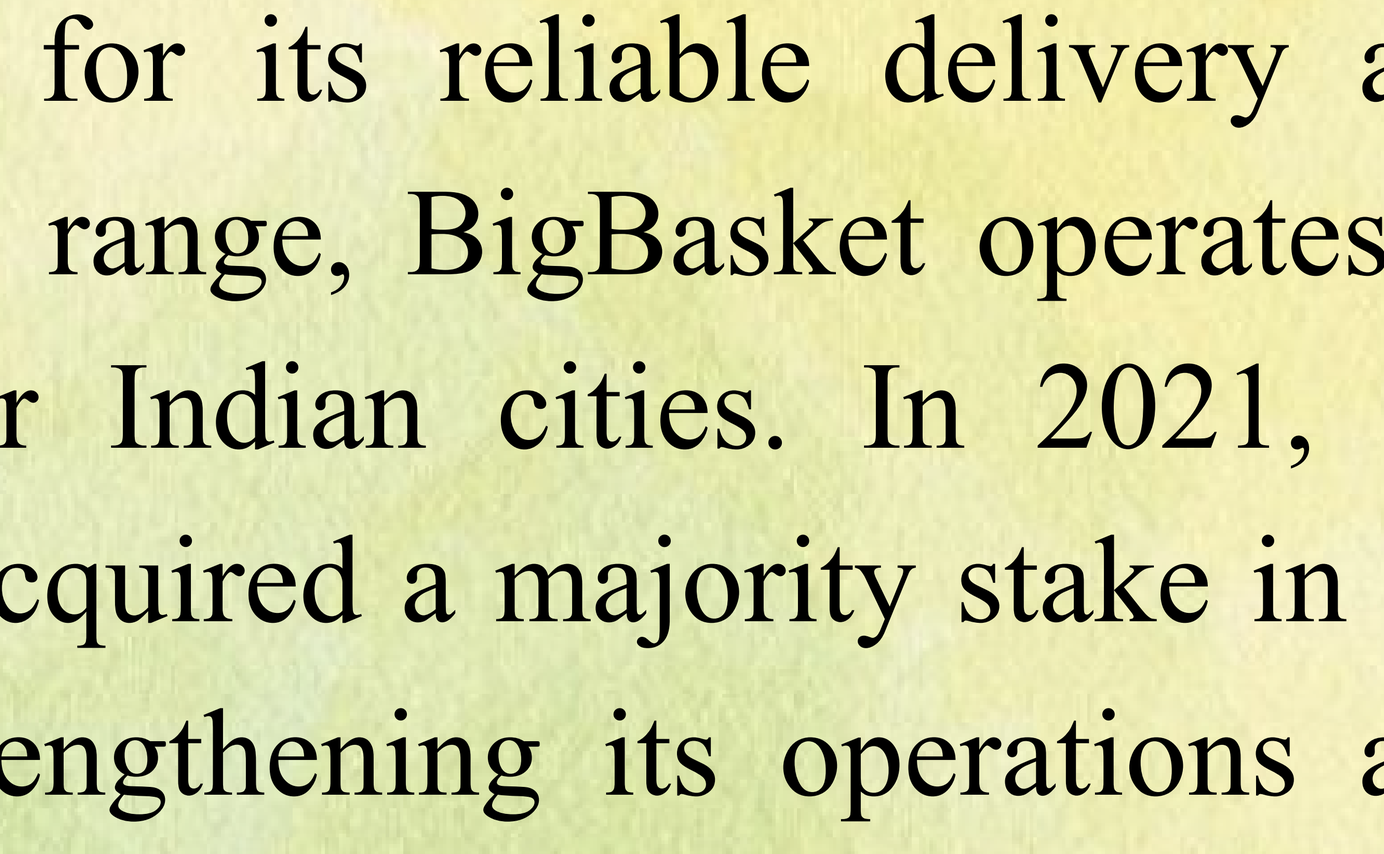
bigbasket

Big Basket Data Analysis

By - Drishti Khosla

About Bigbasket

BigBasket is one of India's largest online grocery platforms, founded in 2011 by Hari Menon and his team. It allows customers to conveniently shop for groceries, fruits, vegetables, dairy items, and household essentials through its website and mobile app. Known for its reliable delivery and wide product range, BigBasket operates in several major Indian cities. In 2021, the Tata Group acquired a majority stake in the company, strengthening its operations and supply chain.



Project Overview

This project analyzes product data from BigBasket, a leading online grocery platform in India. The main aim is to study product categories, brands, prices, discounts, and ratings to understand market trends. Using Power BI, various visuals like bar charts, bubble charts, and word clouds are created to represent data insights. The analysis highlights top brands, category performance, and the relationship between price and rating. It helps in identifying BigBasket's product trends and understanding customer preferences. Overall, the project supports data-driven insights for better business decisions.

Tools Used

- Python
- Pandas
- Google Collab (for coding)
- Power BI
- Canva (for presentation)

Dataset Description

The dataset contains information about various products listed on BigBasket. Below is a description of each column:

- **product** – Name of the product listed on BigBasket.
- **category** – The main category to which the product belongs (e.g., Beverages, Snacks, Dairy, etc.).
- **sub_category** – The specific type or sub-division of the main category (e.g., Tea, Chips, Milk).
- **brand** – The name of the brand or manufacturer of the product.
- **sale_price** – The final selling price of the product after discounts.
- **market_price** – The original price of the product before any discount is applied.
- **rating** – The average customer rating of the product based on user reviews.
- **description** – A short summary of the product details and features.

Analysis

Big Basket.ipynb

File Edit View Insert Runtime Tools Help

Commands + Code + Text Run all

```
b.describe()
```

	index	sale_price	market_price	rating
count	27555.00000	27549.000000	27555.000000	18919.000000
mean	13778.00000	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

```
b.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
```

Analysis

Big Basket.ipynb

File Edit View Insert Runtime Tools Help

Commands + Code + Text Run all

```
[ ]  #here we find out the Top &amp;amp; least sold products
```

```
[ ]  #: Measuring discount on a certain item
discount = b.groupby("product").agg({'market_price':'sum', 'sale_price':"sum"})
discount['discount'] = discount['market_price'] - discount['sale_price']
display(discount)
```

product	market_price	sale_price	discount
& 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

```
[ ]  #we are going to find the 'missing values' from the dataset.
```

```
[ ]
```

Analysis

Big Basket.ipynb

File Edit View Insert Runtime Tools Help

Commands + Code + Text Run all

```
b.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

```
b['sale_price'].fillna(b['sale_price'].mean(), inplace=True)
b['market_price'].fillna(b['market_price'].mean(), inplace=True)
b['rating'].fillna(b['rating'].median(), inplace=True)
b['brand'].fillna('Unknown', inplace=True)
```

```
/tmp/ipython-input-973851093.py:1: FutureWarning: A value is trying to be set on a copy of a DataFrame or Series through chained assignment using an inplace method.
The behavior will change in pandas 3.0. This inplace method will never work because the intermediate object on which we are setting values always behaves as a copy.

For example, when doing 'df[col].method(value, inplace=True)', try using 'df.method({col: value}, inplace=True)' or df[col] = df[col].method(value) instead, to perform the operatic


  b['sale_price'].fillna(b['sale_price'].mean(), inplace=True)
/tmp/ipython-input-973851093.py:2: FutureWarning: A value is trying to be set on a copy of a DataFrame or Series through chained assignment using an inplace method.
The behavior will change in pandas 3.0. This inplace method will never work because the intermediate object on which we are setting values always behaves as a copy.
```

Big Basket.ipynb

File Edit View Insert Runtime Tools Help

Commands + Code + Text Run all

```
Q1 = b['sale_price'].quantile(0.25)
Q3 = b['sale_price'].quantile(0.75)
IQR = Q1 - Q3
lower_bound = Q1 - 1.5*IQR
upper_bound = Q3 + 1.5*IQR
outliers = b[(b['sale_price']<lower_bound) & (b['sale_price']>upper_bound)]
print(outliers[["product", "sale_price"]])


#calculate mean of sale price

mean_val= b['sale_price'].mean()

#replace outliers with mean

b["sale_price"] = b["sale_price"].apply(
    lambda x:mean_val if x in outliers else x)
display(b["sale_price"])
```

```
                                              product  sale_price
0                 Garlic Oil - Vegetarian Capsule 500 mg      220.00
1                                 Water Bottle - Orange      180.00
2                          Brass Angle Deep - Plain, No.2      119.00
3      Cereal Flip Lid Container/Storage Jar - Assort...      149.00
4                       Creme Soft Soap - For Hands & Body      162.00
...                                                   ...         ...
27550         Wottagirl! Perfume Spray - Heaven, Classic      199.20
27551                                           Rosemary       67.50
27552                       Peri-Peri Sweet Potato Chips      200.00
27553                          Green Tea - Pure Original      396.00
27554                       United Dreams Go Far Deodorant      214.53

[23117 rows x 2 columns]
```

	sale_price
0	220.00
1	180.00

Variables Terminal

Analysis

Big Basket.ipynb

File Edit View Insert Runtime Tools Help

Commands + Code + Text Run all

```
...                                                 ...         ...
27550         Wottagirl! Perfume Spray - Heaven, Classic      199.20
27551                                           Rosemary       67.50
27552                             Peri-Peri Sweet Potato Chips  200.00
27553                                Green Tea - Pure Original  396.00
27554                           United Dreams Go Far Deodorant  214.53

[23117 rows x 2 columns]
```

	sale_price
0	220.00
1	180.00
2	119.00
3	149.00
4	162.00
...	...
27550	199.20
27551	67.50
27552	200.00
27553	396.00
27554	214.53

27555 rows × 1 columns

dtype: float64

[ ]

```
sc#after manipulating our dataset we'll import the file
b.to_csv("Cleaned_data_Bigbasket.csv", index = False)
```

[ ]

```
from google.colab import files
files.download('Cleaned_data_Bigbasket.csv')
```

Power BI Dashboard

Overview Dashboard

BRAND
All

CATEGORY
All

PRODUCT
All

MARKET PRICE
3.00 12,500.00

SALE PRICE
2.45 112,475.00

11
Count of category

23.51K
Count of product

Rating by category
category: Beauty & Hy..., Gourmet & ..., Kitchen, Gar..., Snacks & Bra..., Foodgrains, O..., Cleaning & H..., Beverages, Bakery, Cake..., Baby Care, Fruits & Vege...
0K 20K 40K
Rating

Market price of product
Market Price: 30K 20K 10K 0K
Extra Vir... Olive Oil ... Olive Oil ... Extra Lig... Olive Po... Eau De To... Olive Oil ... Bravura C... Gas Stov... Deodora... Epilator S...

Product count by category
Category: Beaut... Gour... Kitch... Snack... Food...
7.87K (28...)
4.69K (17...)
3.58K (12.99%)

Average rating
0.00 3.99

Sum of sale price
0.00M 9.22M 18.44M

Brand Performance

Count of product by brand
Brand: Fresho, bb Royal, BB Home, DP, Fresho Signature, bb Combo, Amul, INATUR, Himalaya
0 500
Product count

Average Rating & Price by Brand and Category
category: Baby Care, Bakery, ..., Beauty ..., Beverages, Cleaning ..., Eggs, M..., Foodgrai..., Fruits & ..., Gourmet ...
Average of sale_price: 10K 5K 0K
1 2 3 4
Average of rating

Average of rating by brand
Rating: 5 0
1mg Am... Annai Aur... B2B BA... Ber... Car... Che... Chi... Cor... Cre... De... DIV... ENE Epis... Floh Go...
Brand

Subcategories by Category
category: Beauty & ..., Gourmet ..., Kitchen, ..., Snacks & ..., Foodgrai...
0.56K (2.02%)
7.87K (28.55%)
2.68K (9.7...)
2.... (...)
2.81K (10...)
4.69K (17.02%)
3.58K (12.99%)

Total Rating by Category
Sum of rating: 40K 20K 0K
Beaut... Gour... Kitch... Snack... Food... Cleani... Beve... Baker... Fruits... Eggs,...
category

Conclusion

The analysis of BigBasket's product dataset provides valuable insights into the platform's pricing, brand performance, and customer preferences. The study highlights how discounts and ratings vary across different categories and brands, helping to identify top-performing products and competitive pricing strategies. It also shows that certain brands dominate specific categories due to better pricing and customer satisfaction. By visualizing data through Power BI, meaningful patterns and relationships were discovered, supporting better business and marketing decisions. Overall, this project demonstrates the importance of data analysis in understanding e-commerce trends and improving customer experience on platforms like BigBasket.

Thank you!!!