MarketMetrics – Analyzing Grocery & FMCG Pricing Trends

Team Members

• Ashutosh Rawat – Data Cleaning & Preparation

• Abhay Tiwari – Statistical Analysis & Visualization

• Yadvendra Singh Dhakad – SQL Analysis & Dashboard Development

Project Overview

The Indian online grocery and FMCG (Fast-Moving Consumer Goods) sector is booming, driven by convenience, affordability, and digital transformation. With the increasing variety of products and brands available, understanding pricing trends is crucial for online platforms aiming to enhance user experience and drive profitability.

In this project, we analyzed a dataset of over 27,000 product listings from BigBasket—India's largest online grocery platform. Our objective was to uncover actionable insights around pricing behavior using a mix of SQL, Python, and statistical techniques. The project was divided among team members for focused contributions, including data cleaning, exploration, visualization, and dashboarding.

Data Cleaning Tasks

Performed by Ashutosh Rawat:

Data cleaning is a foundational step in any data science project. This dataset contained various inconsistencies such as embedded symbols (₹) in price columns and irregular pack size formats. The key tasks included:

- Removing special characters and ensuring sale_price and market_price were numeric

- Standardizing brand names and sub-category labels for grouping

- Cleaning inconsistent or overly verbose product names

- Ensuring no missing values disrupted SQL aggregation or statistical tests

These steps ensured the dataset was ready for smooth exploratory and analytical processing.

Exploratory Data Analysis (EDA)

Conducted by Abhay Tiwari:

EDA focused on discovering patterns and outliers in the pricing data. Using Python libraries like Seaborn, Matplotlib, and Plotly, several visualizations were developed:

- **Grouped Bar Charts**: Displayed average prices across categories and brands

- **Box Plots**: Highlighted price variability within categories

- **Distribution Plots**: Showed price distribution skewness

- **Heatmaps**: Displayed weak-to-moderate correlations between sale_price, market_price, and rating

Key findings:

- Beauty & Hygiene and Gourmet products had the highest average prices

- Staples and Bakery products were the most affordable on average

- Price variance was greatest in Snacks & Branded Foods, indicating a mix of premium and budget offerings

- Ratings had minimal correlation with prices, indicating other factors drive product value perception

Statistical Analysis

Also by Abhay Tiwari:

- t-Test: Compared prices of branded vs unbranded items — branded items were significantly more expensive.
- ANOVA Test: Confirmed significant price differences between product categories.
- Descriptive Stats: Computed mean, median, standard deviation for each category and sub-category.

SQL-Based Insights

Conducted by Yadvendra Singh Dhakad:

Queries Performed:

1. Average Price per Category

```
SELECT category, ROUND(AVG(sale_price), 2) AS avg_price
FROM products
GROUP BY category
ORDER BY avg_price DESC;
```

2. Top 5 Expensive Items in Each Category

```
SELECT *
FROM (
    SELECT *, RANK() OVER (PARTITION BY category ORDER BY sale_price DESC) AS rank
    FROM products
) ranked
WHERE rank <= 5;
```

3. Categories with Highest Price Variation

```
SELECT category, STDDEV(sale_price) AS price_std_dev
FROM products
GROUP BY category
ORDER BY price_std_dev DESC;
```

4. Brand-wise Price Analysis

```
SELECT brand, COUNT(*) AS products, AVG(sale_price) AS avg_price
FROM products
GROUP BY brand
ORDER BY products DESC
LIMIT 10;
```
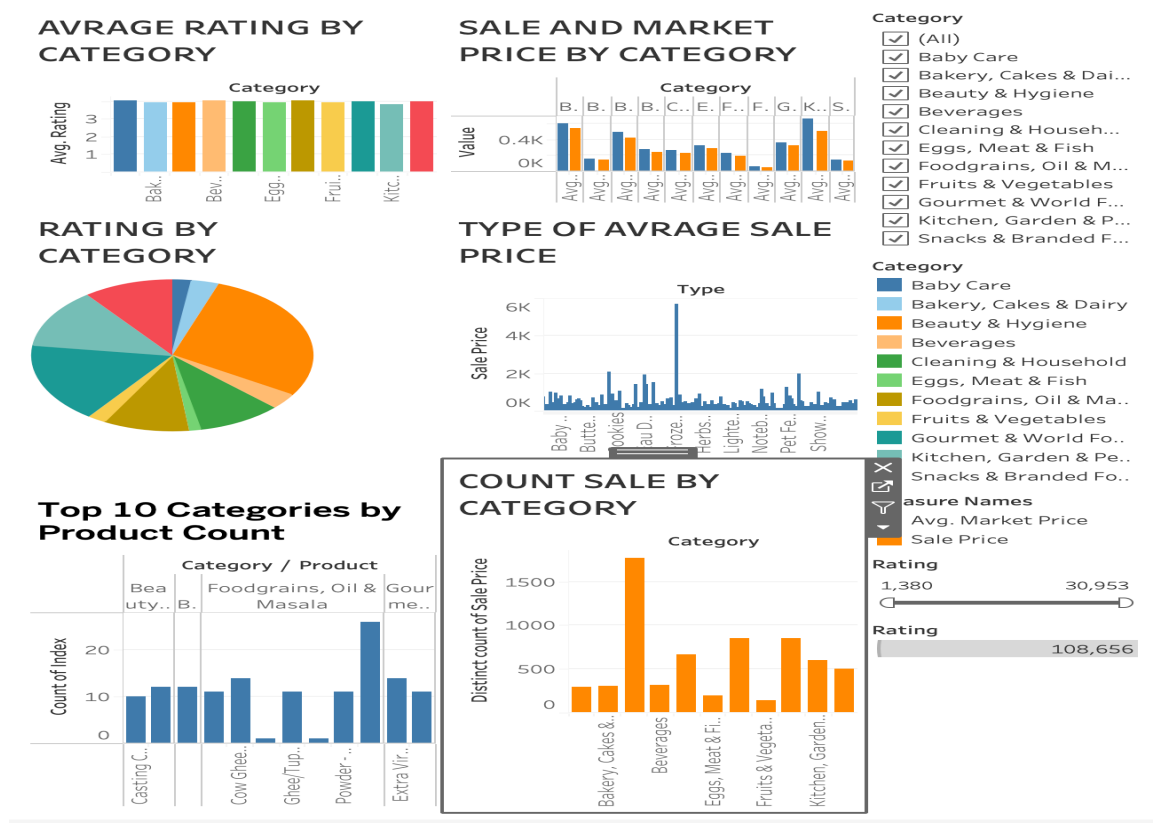
Dashboard & KPI Cards

Built by Yadvendra Singh Dhakad using Tableau:

Highlights:
- Average Price Across All Products: ₹174.36
- Total Product Categories: 17
- Top Category by Average Price: Beauty & Hygiene

- Most Represented Brands: Nivea, Surf Excel, Haldiram's
- Popular Pack Sizes: 1 kg, 500 ml
- Visual Filters: Category, Brand, Price Range

Conclusion

The MarketMetrics project demonstrates how data science tools can provide critical insight into FMCG pricing trends. From understanding brand value to identifying outlier products, our findings can guide online platforms in optimizing product listings, promotional strategies, and pricing models.

Using SQL and statistical rigor ensured accurate analysis, while dashboards helped communicate insights effectively. Going forward, integrating user behaviour and sales data could further strengthen market segmentation and predictive pricing models.