Big Basket Product Data Analysis - EDA Report

Project Objective:

To perform exploratory data analysis (EDA) on Big Basket's product catalog to understand price distribution, discount patterns, brand/category performance, and product rating insights for actionable business decisions.

Dataset Overview:

bigbasket

The dataset contains 27,555 rows and 10 columns

To perform exploratory data analysis (EDA) on Big Basket's product catalog to understand price distribution, discount patterns, brand/category performance, and product rating insights for actionable business decisions.

product: Name of the product

category: Primary product category (e.g., Beverages, Snacks)

sub_category: Sub-classification of products

brand: Manufacturer or label

sale_price: Current Big Basket selling price

market_price: Standard market price

rating: Customer rating (1 to 5)

description: Text description of product

type: Product type info

Data Cleaning:
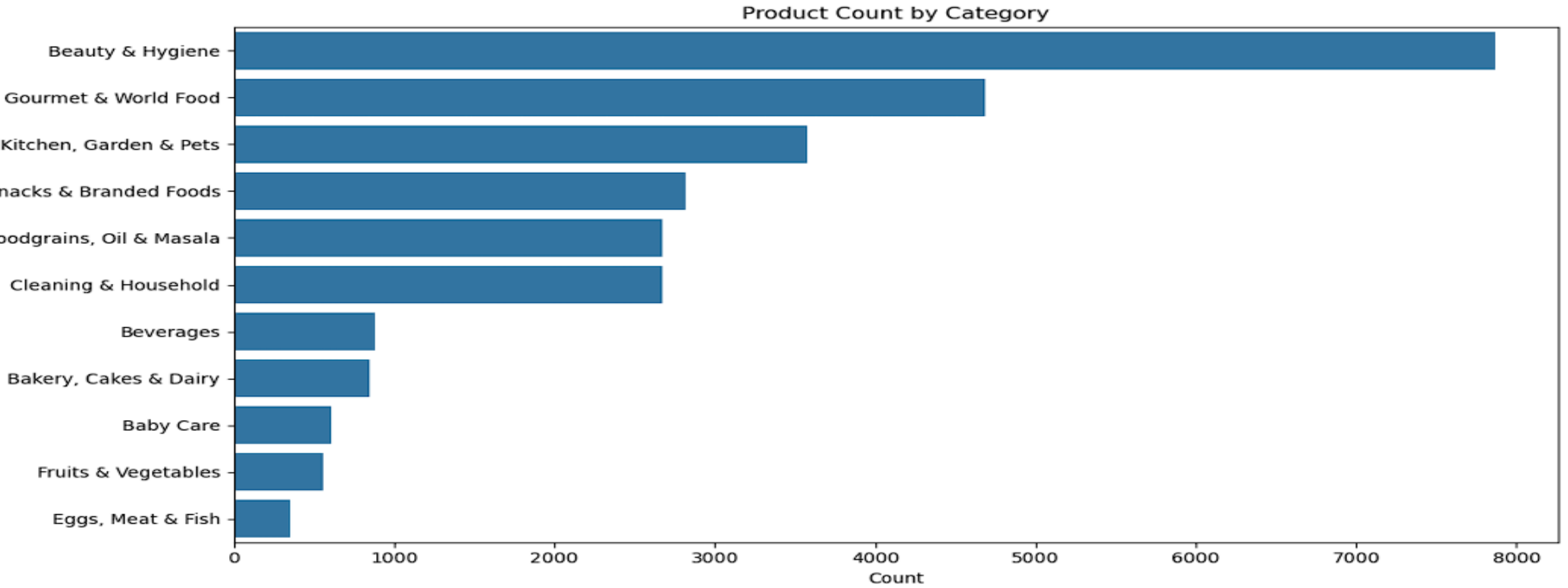
The dataset contains 27,555 rows and 10 columns

Dropped null values in product and brand.

Replaced missing values in sale_price and rating with their mean.

Filled missing description values with "Unknown".

Created new feature: discount_percent = ((market_price - sale_price) / market_price) * 100

Product Count by Category

The Beauty & Hygiene category has the highest product count, significantly ahead of other categories like Snacks and Beverages.

This dominance suggests that Big Basket places strategic emphasis on personal care and wellness products, likely due to consistent demand and higher customer retention in this segment

Distribution of Sale Price

The sale price distribution reveals that the majority of products are affordably priced under ₹500, aligning with Big Basket's mass-market strategy.

However, the presence of a long right tail and extreme outliers (above ₹10,000) suggests either premium/bulk items or possible pricing anomalies that may need review.
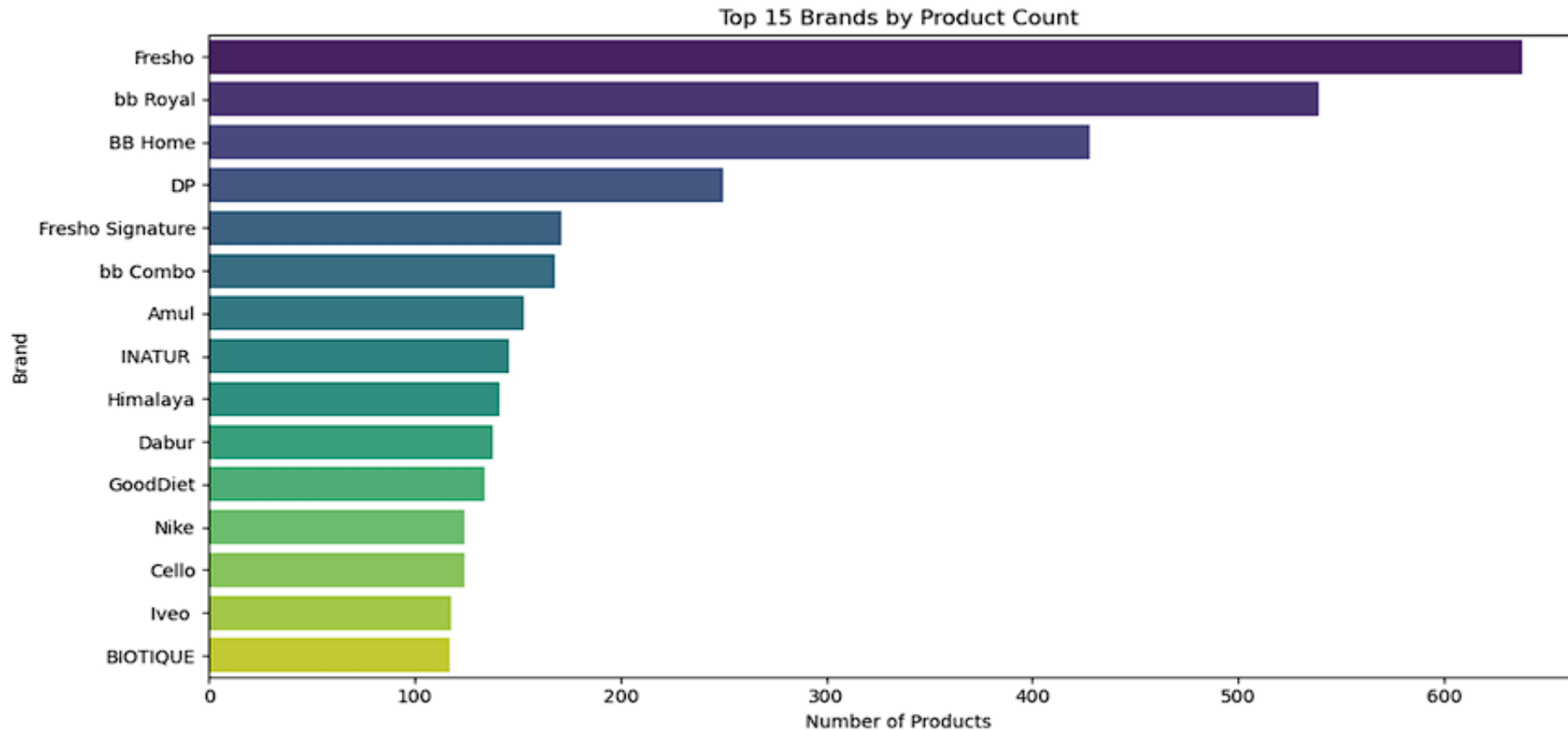
Distribution of Discount Percent

bigbasket
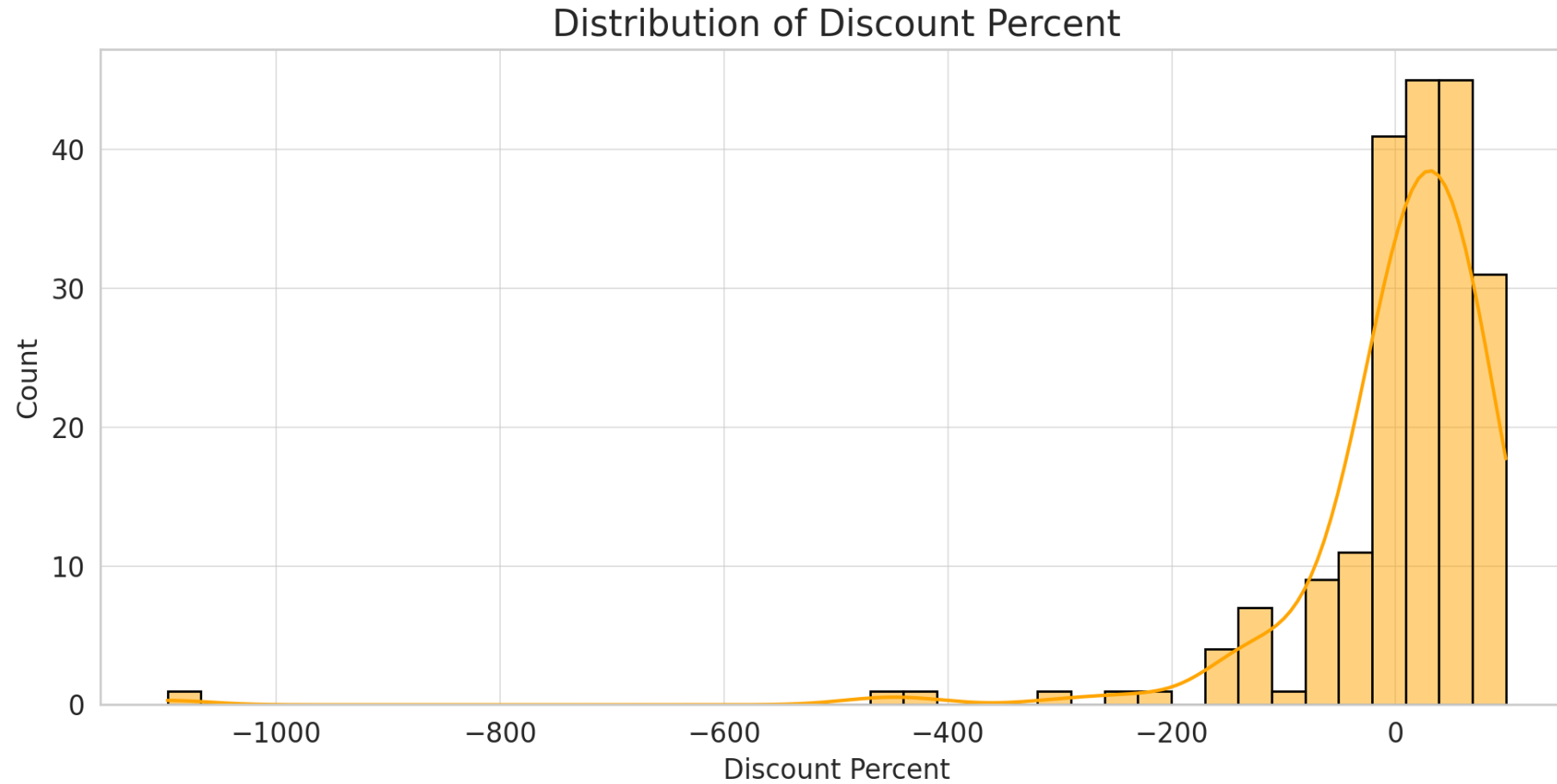
Most products offer 10% to 30% discounts.

Discount patterns suggest aggressive pricing on high-MRP items.

Top 10 Brands by Product Count

Top 15 Brands by Product Count

Brands like Fresho, Nivea, and others appear most frequently in the product catalog, suggesting strong supply chain partnerships and consistent customer demand.

This could be due to vendor tie-ups, exclusivity, or brand popularity, making them strategic assets in Big Basket's inventory.
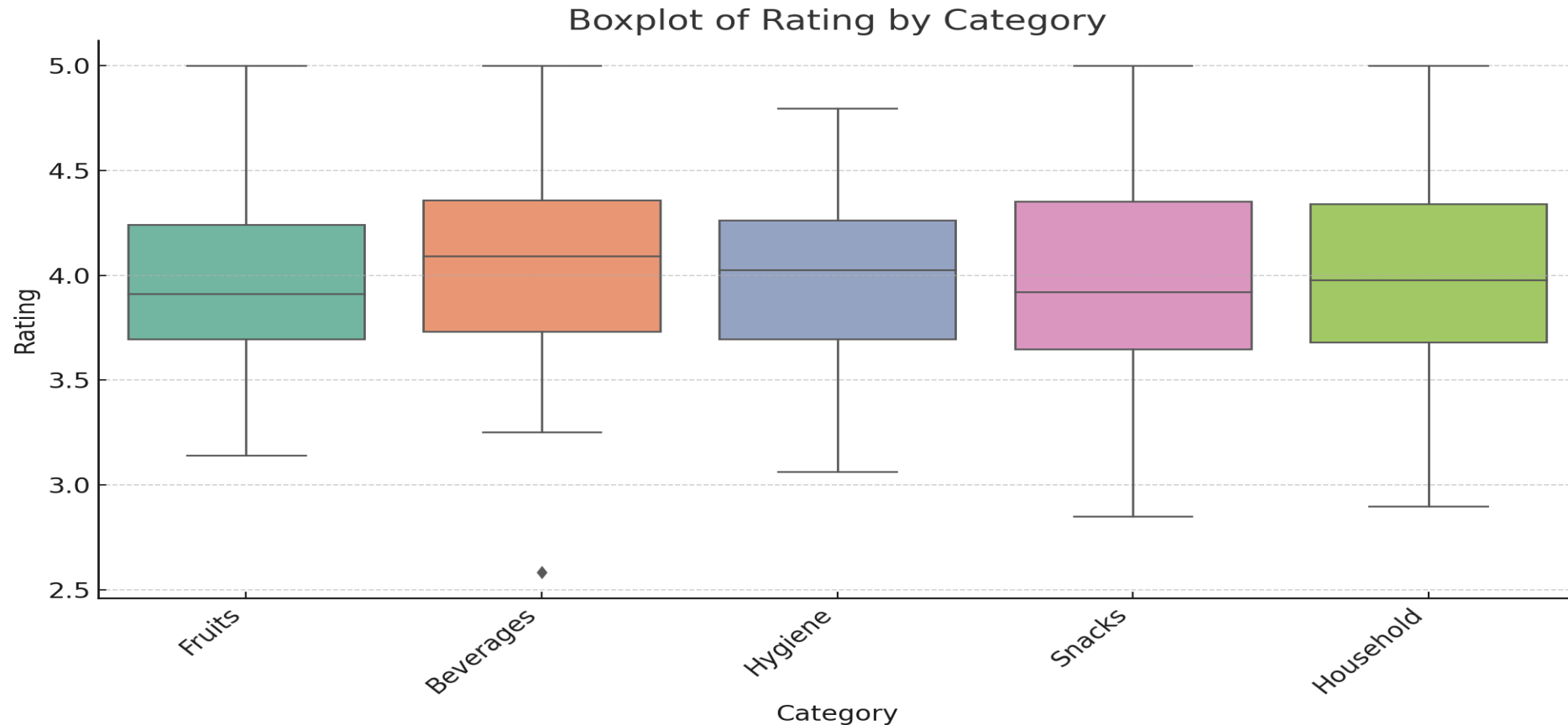
Boxplot of Sale Price

Boxplot of Sale Price
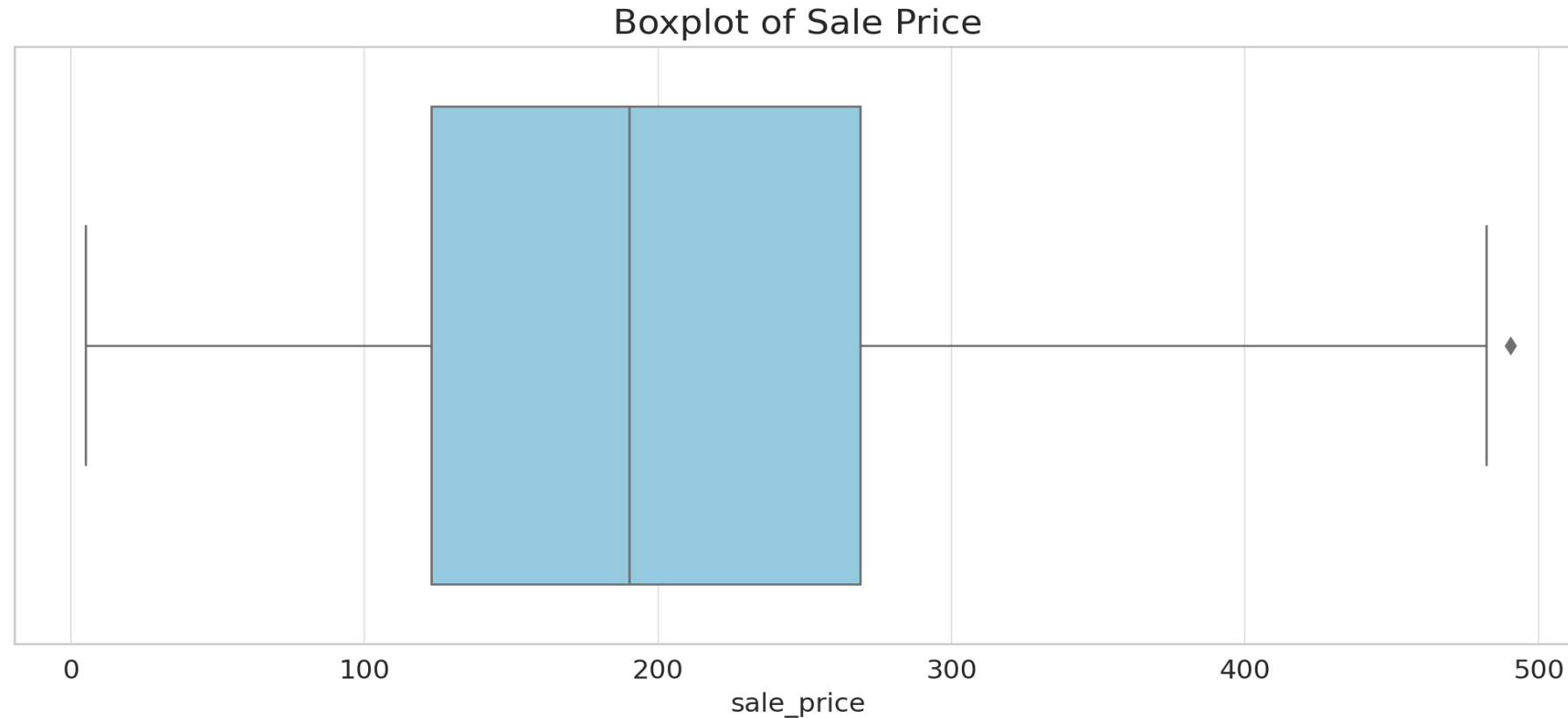
The boxplot of sale_price shows a broad price range across products.

Several extreme outliers above ₹10,000 may represent bulk product listings, premium items, or data entry errors, which may warrant further validation during catalog audits.

Boxplot of Rating by Category

Boxplot of Rating by Category

The boxplot reveals that Fruits and Beverages consistently receive higher median ratings, reflecting customer satisfaction in freshness and quality.

In contrast, categories like Household and Hygiene show wider rating spread, indicating inconsistent product performance or varied customer expectations.

Discount vs Rating (Scatter Plot)

Discount Percent vs Rating
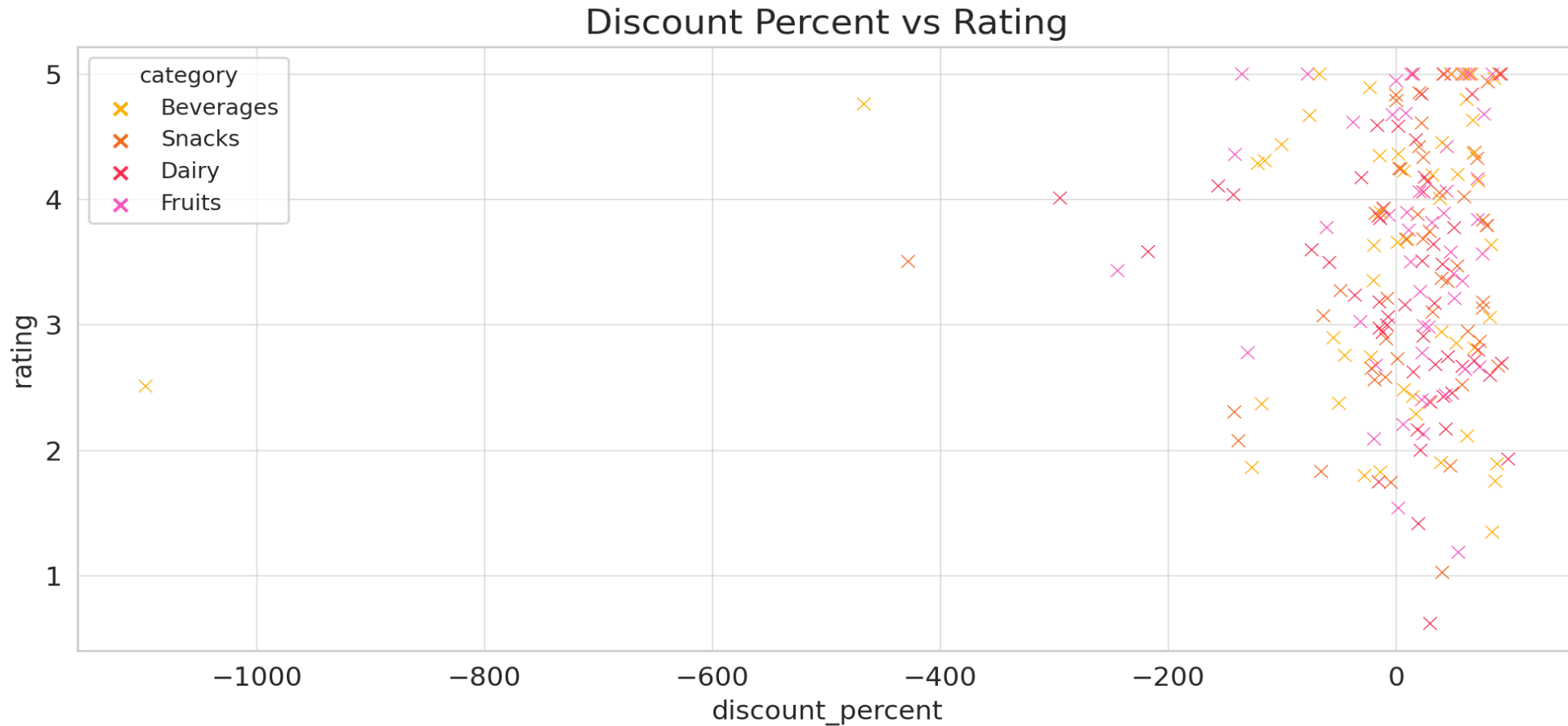
The scatter plot shows that products with higher or lower discounts receive similar rating distributions, indicating that discounts do not significantly influence customer satisfaction.

Products across all categories—regardless of discount level—generally hover between 3.5 to 5 rating range.

Top 5 Products By Maximum Discount
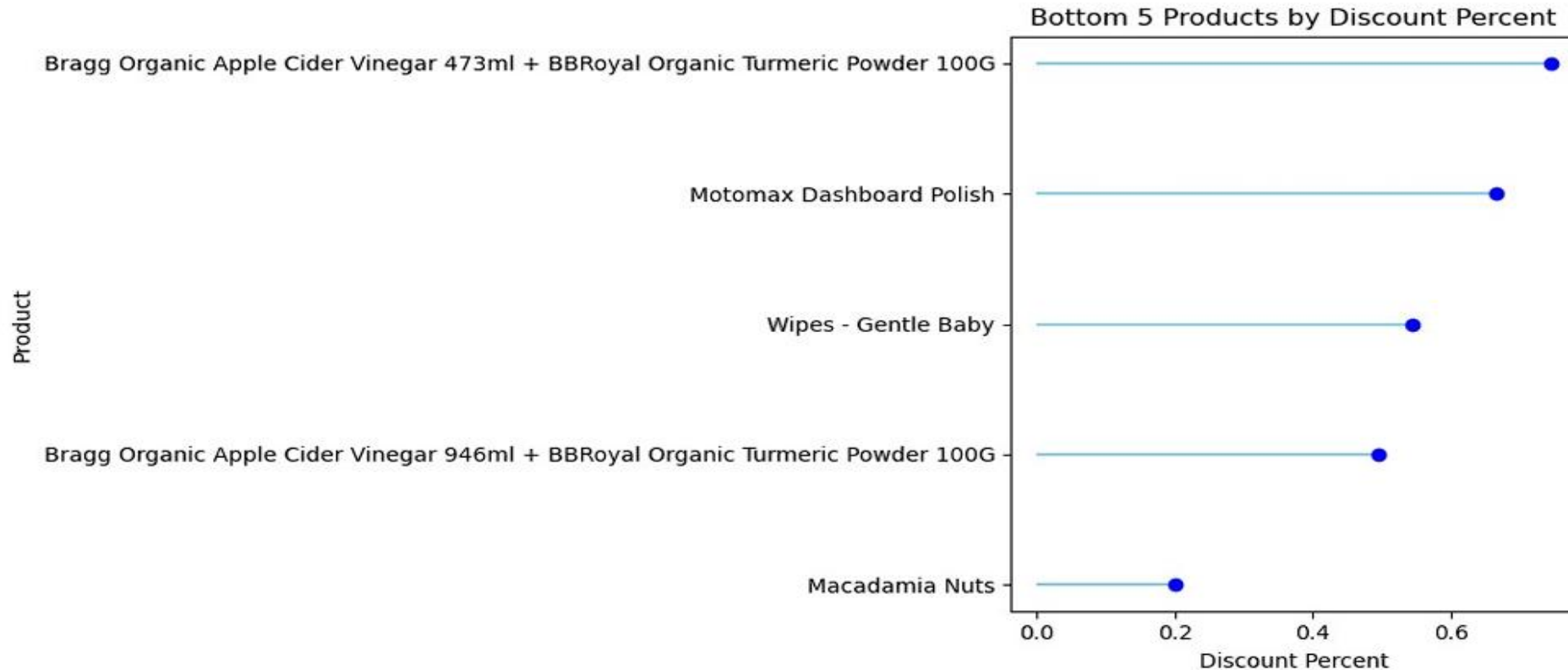
The top 5 products offer discounts ranging from ~80% to 90%, suggesting they are either overstocked, seasonal clearance items, or part of promotional campaigns.

Such deep discounting is rare and usually applied to non-essential or decorative items.
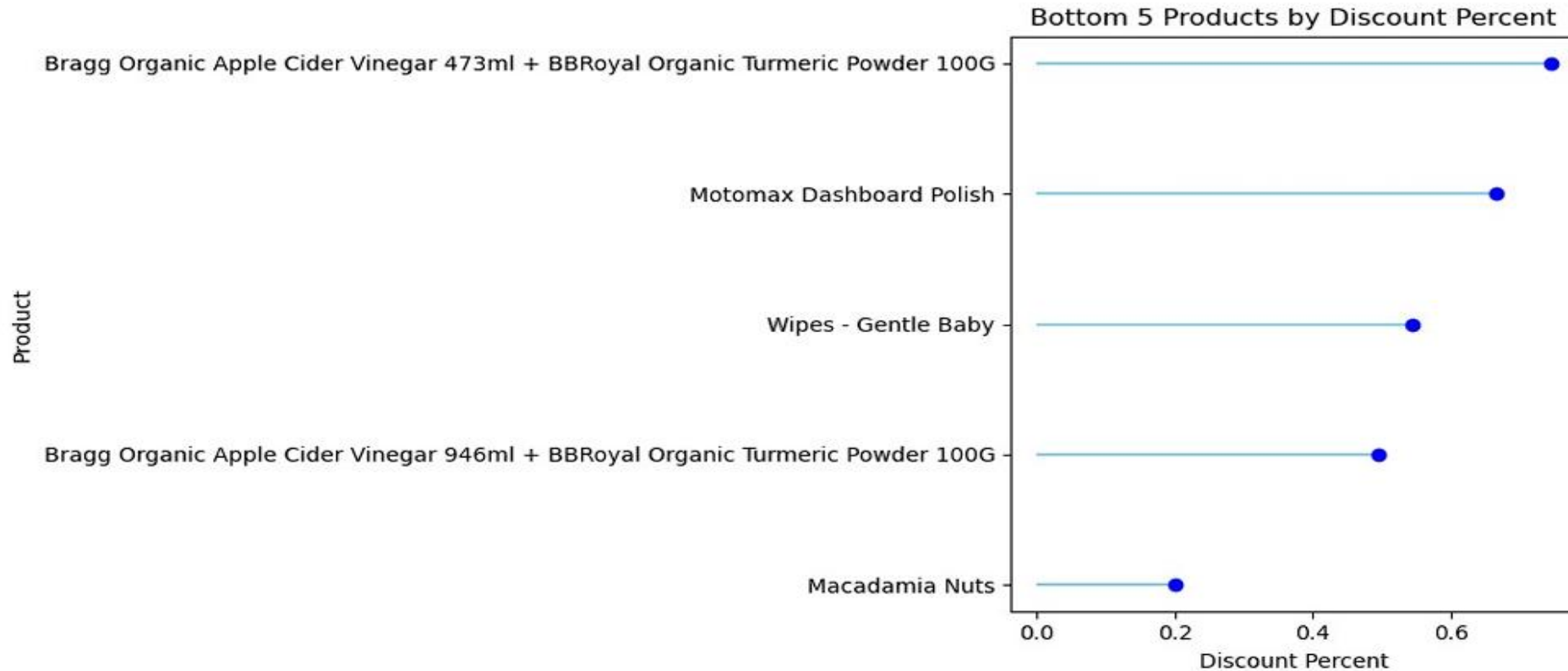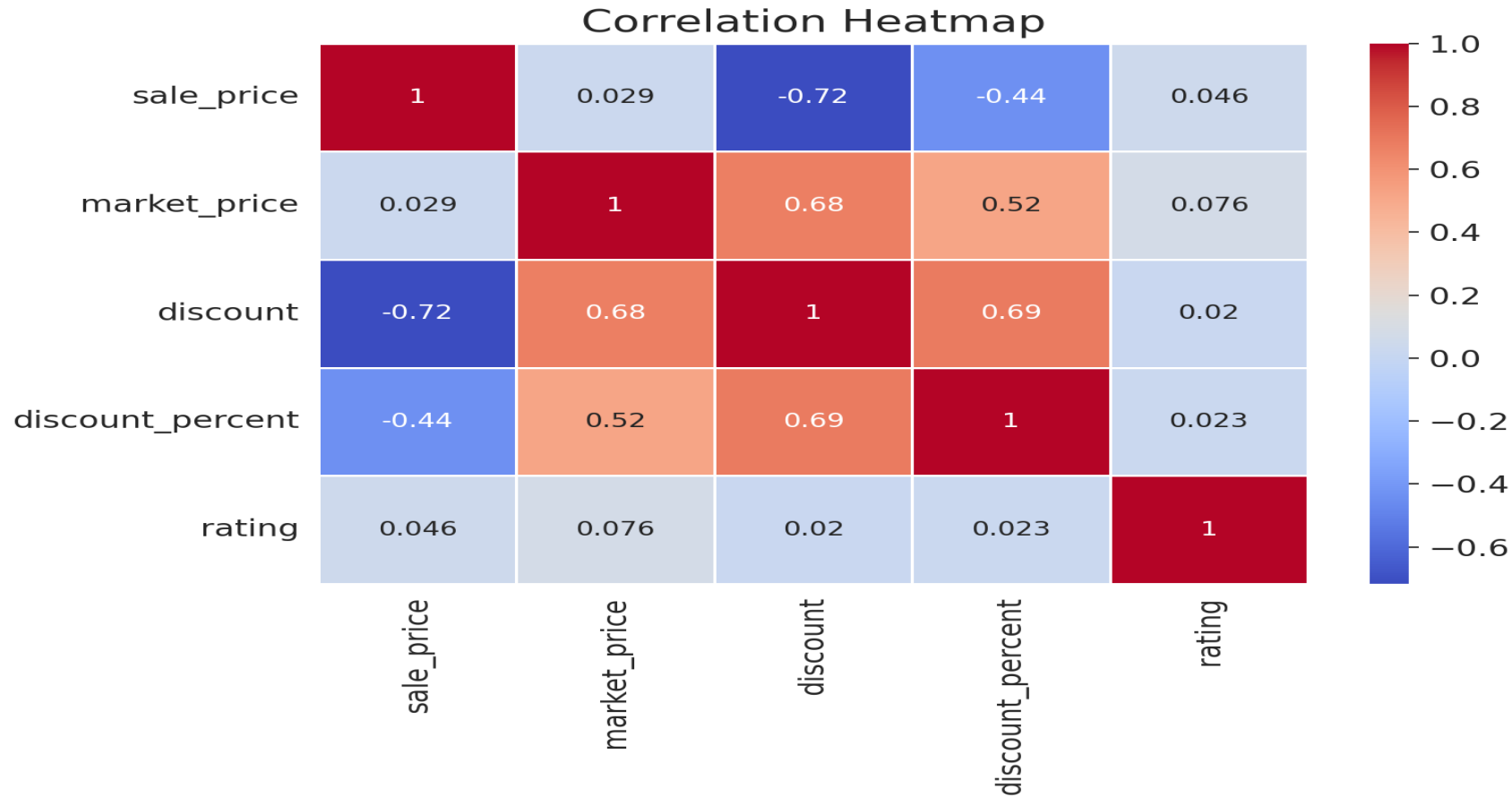
Top 5 Products By Maximum Discount

The top 5 products offer discounts ranging from ~80% to 90%, suggesting they are either overstocked, seasonal clearance items, or part of promotional campaigns.

Such deep discounting is rare and usually applied to non-essential or decorative items.

Correlation Heatmap

There's a strong positive correlation between market_price and discount, which is expected—higher priced items often carry larger absolute discounts.

Rating shows weak or negligible correlation with price or discount variables, indicating that customer satisfaction is not directly tied to pricing.

Recommendations & Strategic Suggestions

Prioritize Top Categories

Focus on "Beauty & Hygiene" and "Fruits & Beverages" due to high product volume and consistent customer satisfaction.

Optimize Pricing Structure

Most products fall below ₹500. Review and correct extreme outliers (₹10,000+) to ensure catalog accuracy.

Refine Discounting Strategy

Discounts above 80% suggest overstock or clearance. Introduce a structured discount model based on product type and category performance.

Strengthen Brand Partnerships

Brands like Fresho and Nivea dominate listings. Consider exclusive vendor relationships or Big Basket private labels for top segments.

Improve Low-Rated Segments

Categories with high rating variance (e.g., Household, Snacks) need quality checks. Investigate poor-performing SKUs/vendors.

Discount ≠ Satisfaction

Weak correlation between discounts and ratings. Customer satisfaction is driven more by product quality than price cuts.

Conclusion

The EDA reveals key trends in Big Basket's catalog related to pricing, discounting, and category performance.

It highlights:

High-performing categories to focus on

Excessive discounting patterns that need optimization

Discrepancies in product quality across categories

Strong vendor relationships driving product volume

By using these insights, Big Basket can enhance profitability, ensure catalog consistency, and deliver a better customer experience.

Thank You

Thank you for taking the time to review this analysis.

This exploratory data report aims to support better decision-making through data-driven insights.

Looking forward to feedback, collaboration, or further exploration opportunities.

Contact Information

Prepared by: Aditya Anand

Email: adityaanand5155@gmail.com

Phone: +91 91229 22306

LinkedIn: linkedin.com/in/aditya-anand-4130262a3

GitHub: github.com/adityaanand515