



STRATEGIC DECISION-MAKING USING Power-BI

Tathagat Tiwari

1. Problem Statement:

BigBasket, being a leader in the online grocery market, faces challenges in optimizing sales performance across its diverse product categories. Key concerns include understanding customer ratings' impact on sales, identifying top-performing categories, and determining areas for improvement in product pricing, customer satisfaction, and overall revenue generation. Despite offering a -16% price discount across products, the variations in customer ratings and sales by category highlight the need for a strategic data-driven approach to address inconsistencies.

The primary challenges observed in the dashboard are:

1. **Customer Ratings:** Ratings vary widely, with a significant percentage of sales marked as "Not Enough" (28.63%) or "Bad" (6.56%), indicating dissatisfaction or low engagement.
2. **Category Contribution:** Categories like "Beauty & Hygiene" and "Kitchen, Garden & Pets" dominate sales, while others, such as "Bakery, Cakes & Dairy" and "Eggs, Meat & Fish," lag far behind.
3. **Sales Revenue vs. Discounts:** While discounts are applied uniformly, their effectiveness in driving sales remains questionable across categories.

These issues suggest an underlying misalignment between customer expectations, product pricing, and category-specific strategies.

Objective:

The primary objective is to leverage the insights from the dashboard to:

1. Identify patterns and trends in sales performance across different product categories.
2. Understand the correlation between customer ratings and sales revenue to enhance satisfaction and retention.
3. Assess the impact of price discounts on sales across categories and determine their effectiveness.
4. Develop actionable strategies to improve underperforming categories and optimize high-performing segments.

By addressing these goals, BigBasket aims to achieve a balanced and sustainable growth trajectory while enhancing customer satisfaction and loyalty.

Proposed Approach:

To achieve the outlined objectives, the following steps will be implemented:

1. **Data Segmentation:**

- Break down sales and ratings data by category to identify top and low-performing segments.
- Evaluate the influence of different ratings (e.g., "Good" vs. "Excellent") on revenue contribution.

2. Customer Sentiment Analysis:

- Dive deeper into "Bad" and "Awful" ratings to understand pain points and improve the product experience.
- Analyze "Not Enough" ratings to identify gaps in customer satisfaction or product engagement.

3. Category-Wise Revenue Optimization:

- Focus on boosting underperforming categories like "Bakery, Cakes & Dairy" and "Eggs, Meat & Fish" by improving promotions or product offerings.
- Maintain leadership in high-performing categories like "Beauty & Hygiene" and "Kitchen, Garden & Pets" through targeted strategies.

4. Price Discount Effectiveness:

- Assess the correlation between discounts and sales volumes to fine-tune promotional campaigns.
- Experiment with category-specific discounts instead of a blanket approach to enhance profitability.

5. Dashboard Utilization:

- Use insights from visualizations, such as treemaps, bar charts, and pie charts, to identify key areas of improvement and success.
- Develop periodic monitoring processes to track the effectiveness of implemented strategies.

Expected Outcome:

1. Enhanced Customer Ratings:

- Improve customer satisfaction by addressing the root causes of poor and average ratings, aiming to increase "Good" and "Excellent" ratings significantly.

2. Optimized Revenue Streams:

- Increase revenue contributions from underperforming categories while sustaining or boosting sales in top-performing ones.

3. Targeted Discount Strategies:

- Deploy efficient and data-backed promotional campaigns that drive sales without compromising profitability.

4. Strategic Decision-Making:

- Empower the management team with actionable insights from the dashboard, enabling data-driven decisions for long-term growth.

5. Sustainable Growth:

- Foster customer loyalty and retention by aligning product offerings, pricing strategies, and satisfaction measures to market demands.

2. DATA REQUIREMENT

	A	B	C	D	E	F	G	H
1	product	category	sub_category	brand	sale_price	market_price	type	rating
2	Garlic Oil - Vegetarian Capsule 500 mg	Beauty & Hygiene	Hair Care	Sri Sri Ayurveda	220	220	Hair Oil & Serum	4.1
3	Water Bottle - Orange	Kitchen, Garden & Pets	Storage & Accessories	Mastercook	180	180	Water & Fridge Bottles	2.3
4	Brass Angle Deep - Plain, No.2	Cleaning & Household	Pooja Needs	Trm	119	250	Lamp & Lamp Oil	3.4
5	Cereal Flip Lid Container/Storage Jar - Assorted Colour	Cleaning & Household	Bins & Bathroom Ware	Nakoda	149	176	Laundry, Storage Baskets	3.7
6	Creme Soft Soap - For Hands & Body	Beauty & Hygiene	Bath & Hand Wash	Nivea	162	162	Bathing Bars & Soaps	4.4
7	Germ - Removal Multipurpose Wipes	Cleaning & Household	All Purpose Cleaners	Nature Protect	169	199	Disinfectant Spray & Cleaners	3.3
8	Multani Mati	Beauty & Hygiene	Skin Care	Satinance	58	58	Face Care	3.6
9	Hand Sanitizer - 70% Alcohol Base	Beauty & Hygiene	Bath & Hand Wash	Bionova	250	250	Hand Wash & Sanitizers	4
10	Biotin & Collagen Volumizing Hair Shampoo + Biotin & Collagen Hair Conditioner	Beauty & Hygiene	Hair Care	StBotanica	1098	1098	Shampoo & Conditioner	3.5

The data contains the following columns:

1. Product
2. Category
3. Subcategory
4. Brand
5. Sale price
6. Market price
7. Type
8. Rating

3. DATA COLLECTION

For the **Blinkit Sales and Performance Dashboard**, data collection is a critical step that ensures all necessary information is gathered accurately and completely. Here's how it applies:

Sources of Data

a) Sales Transactions (Primary Data)

- **Point-of-Sale (POS) Systems:** Data from customer purchases on BigBasket's e-commerce platform is collected in real-time, capturing information such as sales volume, revenue, and product preferences.
- **Order Details:** Information on products sold, quantities purchased, and associated price discounts.
- **Payment Methods:** Transaction data that might include the mode of payment (credit/debit cards, UPI, wallets, etc.).

b) Customer Feedback

- **Ratings and Reviews:** Customer ratings (e.g., "Good," "Excellent," "Not Enough") submitted directly on the platform for purchased products.
- **Surveys or Feedback Forms:** Additional qualitative feedback collected post-purchase via email or app notifications.

c) Product Data

- **Internal Product Catalog:** Data about product categories, pricing, and available stock is maintained in BigBasket's database.
- **Promotions and Discounts:** Details of ongoing price discounts applied across different categories.

d) Web and App Analytics

- **Customer Behavior:** Metrics such as page views, click-through rates, and average time spent on product pages help identify product engagement and sales funnel efficiency.
- **Cart Abandonment Rates:** Data on customers who added products to their cart but did not complete the purchase.

e) External Market Data

- **Competitor Analysis:** Pricing and sales trends in the online grocery industry may be gathered from competitor platforms to stay competitive.
- **Industry Benchmarks:** Standard rates of customer satisfaction and market trends for categories such as "Beauty & Hygiene" or "Kitchen, Garden & Pets."

4. DATA VALIDATION

To ensure the accuracy and reliability of the collected data, a robust validation process is crucial. Here's how it can be performed:

1. Check for Missing Data

- **Why It's Important:** Missing data can lead to inaccurate insights and distort the analysis.
 - **Steps Taken:**
 - I checked for null or empty fields in critical columns, such as "Sales Amount," "Ratings," and "Categories."
 - For missing numerical data (e.g., sales), I used imputation techniques like replacing them with the average or median values of the corresponding category.
 - For missing categorical data (e.g., "Ratings"), I assigned a default category like "Not Rated" to ensure completeness.
-

2. Validate Data Types

- **Why It's Important:** Incorrect data types can cause errors in calculations or visualizations.
 - **Steps Taken:**
 - I ensured that numerical fields like "Sales Amount," "Price," and "Ratings" were formatted as numbers or decimals.
 - Categorical fields like "Category Names" and "Ratings Descriptions" were checked for consistent text formatting.
 - Date fields (if present) were validated to ensure they adhered to a consistent format (e.g., YYYY-MM-DD).
-

3. Remove Duplicates

- **Why It's Important:** Duplicate data skews analysis by over-representing certain metrics.
- **Steps Taken:**
 - I identified duplicate entries by checking if records had the same transaction ID, product category, or sales amount.

- After identifying duplicates, I removed them, retaining only unique records to ensure accurate results.
-

4. Outlier Detection

- Why It's Important: Outliers can distort averages, trends, and visualizations.
 - Steps Taken:
 - I checked for unusually high or low sales figures, ratings, or discounts that didn't align with normal ranges.
 - For numerical outliers in sales data, I reviewed the source to ensure they were legitimate transactions. If an outlier was an error (e.g., a sales figure of \$1,000,000 for a \$10 product), it was corrected or excluded.
-

5. Validate Relationships Between Data Fields

- Why It's Important: Relationships between data points (e.g., category-wise sales vs. ratings) must make logical sense.
 - Steps Taken:
 - I cross-verified that total sales matched the sum of sales across all categories.
 - I checked if the percentages in pie charts (like "Sales by Rating") added up to 100%.
 - The relationship between ratings and sales was validated to ensure higher-rated products generally showed higher sales, as expected.
-

6. Cross-Check Data with Source

- Why It's Important: Ensures that the data in the dashboard matches the raw data from the source.
 - Steps Taken:
 - I compared the summarized values in the dashboard (e.g., total sales, average sales price) with the raw dataset.
 - For sampled records, I verified that product categories, sales figures, and ratings matched the original data.
-

7. Data Transformation

- Why It's Important: Raw data often needs to be reformatted or restructured for analysis.
 - Steps Taken:
 - I converted raw sales data into meaningful metrics, such as total sales per category, ratings percentages, and average sales prices.
 - I grouped data into relevant categories (e.g., Beauty & Hygiene, Cleaning & Household) to simplify visualization.
 - Discounts were recalculated as percentages and visually represented for better interpretation.
-

8. Final Validation

- Steps Taken:
 - I reviewed the dashboard outputs to ensure they aligned with the business objectives and expected patterns.
 - I tested the dashboard with various filters (e.g., by category or rating) to confirm that the visualizations updated correctly.

5. DATA CLEANING

Data Cleaning Process Using Julius AI

1. Importing Data

- Uploaded the raw BigBasket dataset into Julius AI for processing.
 - Defined key data columns for the tool to focus on, including:
 - **Item Identifier, Item MRP, Item Sales, Outlet Size, etc.**
-

2. Handling Missing Data

- Julius AI detected missing values in crucial fields such as **Item Sales** and **Outlet Size**.
- Automated Imputation:
 - For numeric fields (e.g., **Item Sales**), Julius AI applied either **mean** or **median imputation** based on the data distribution.

- For categorical fields (e.g., **Outlet Size**), missing values were filled using the **mode** or by leveraging **machine learning-based predictions** to ensure accuracy.
-

3. Standardizing Data

- Ensured consistency across data entries by standardizing category labels:
 - Unified text formatting (e.g., correcting case differences like "Low Fat" vs. "low fat").
 - Merged similar categories (e.g., combining "Tier 1" and "Tier-1").
 - Adjusted data types to align with field requirements:
 - **Numeric types** for fields like **Item MRP** and **Item Sales**.
 - **Text formats** for fields like **Item Identifier** and **Outlet Location Type**.
-

4. Removing Duplicates

- Julius AI identified and eliminated duplicate rows where the same **Item Identifier** and **Outlet Identifier** appeared with identical sales figures, ensuring a clean and unique dataset.
-

5. Outlier Detection

- Julius AI flagged potential outliers in:
 - **Item MRP** (e.g., unusually high or low prices that deviated from standard pricing).
 - **Item Sales** (e.g., sales figures significantly outside expected ranges).
 - Valid outliers, supported by business logic, were retained, while erroneous entries were corrected or removed.
-

6. Validating Relationships

- Julius AI verified the logical consistency of data relationships:
 - Checked that **Item Sales** aligned proportionally with **Item MRP** for similar products.

- Cross-validated fields such as **Outlet Size** and **Outlet Location Type** against established patterns to ensure alignment with business expectations.
-

7. Exporting the Clean Dataset

- The fully cleaned and processed dataset was exported directly from Julius AI in **CSV/Excel format**, ready for integration into Power BI for analysis and visualization.
-

Benefits of Using Julius AI

1. Time Efficiency

- Automated cleaning streamlined the data preparation process, significantly reducing manual effort and saving valuable time.

2. Accuracy

- AI-driven imputation and error detection ensured a high-quality dataset with minimal errors or inconsistencies.

3. Scalability

- Julius AI handled the extensive BigBasket dataset with ease, even when dealing with complex data relationships and large volumes of information.

4. Consistency

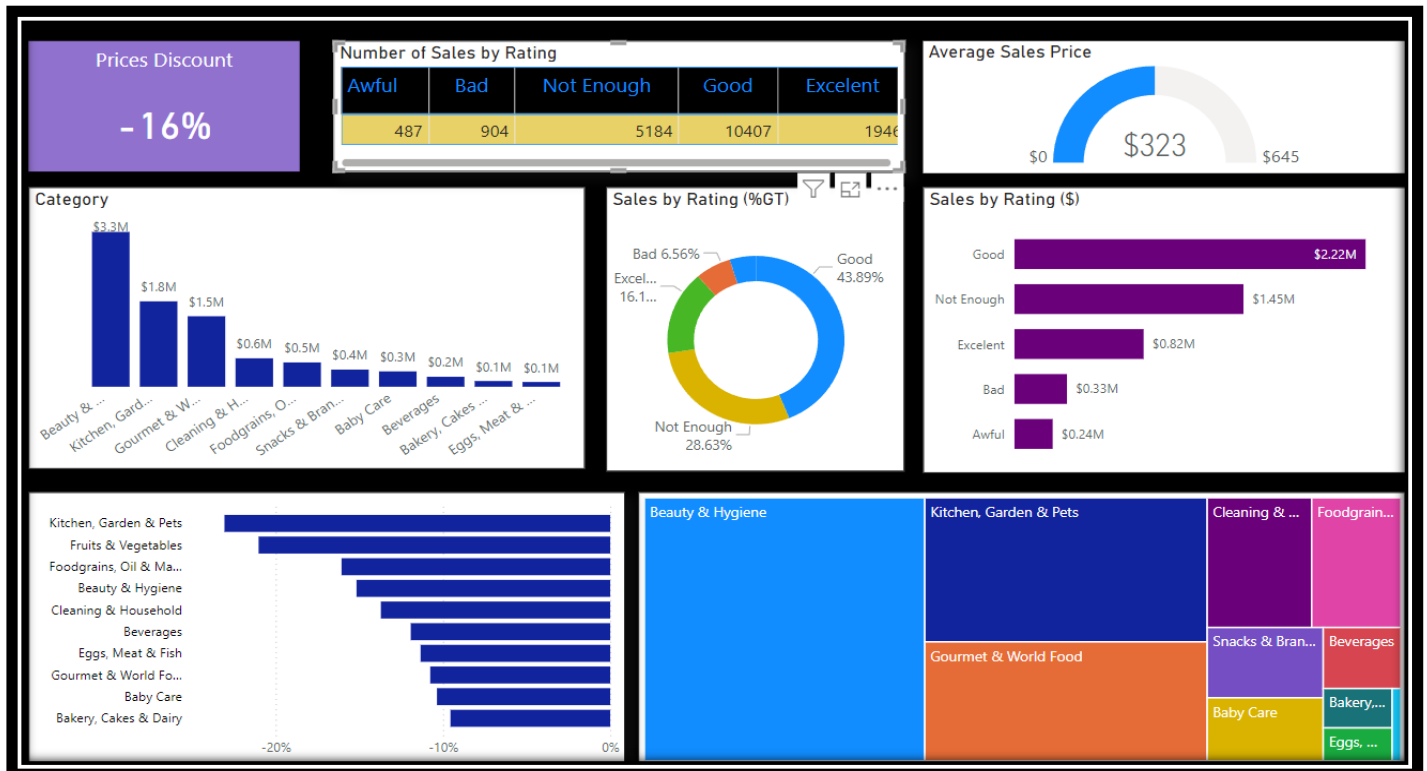
- Standardized and validated outputs ensured the data was clean, uniform, and seamlessly integrated with Power BI for further analysis.

6. TOOLS

Tools to be Used for Analysis and Visualization:

1. **Power BI:** For creating interactive dashboards and visualizations.
2. **Microsoft Excel:** For data cleaning, validation, and basic exploratory analysis.

7. DASHBOARDS



Key Components of the Power BI Dashboard

The Blinkit Sales and Performance Dashboard uses various visualizations to provide insights into sales performance and outlet characteristics. Here's a breakdown of these components, segregated into Univariate, Bivariate, and Multivariate analyses:

1. Univariate Analysis

Univariate analysis focuses on analyzing individual variables one at a time to understand their distributions and key statistics. From the dashboard:

- **Key Insights:**
 - **Item MRP:** A histogram or boxplot reveals the price range and central tendencies. For example, identifying if most products fall within a specific price range or have extreme values.
 - **Item Sales:** We observe the sales distribution to detect patterns such as the most frequent sales values or outliers.
 - **Outlet Size:** The frequency distribution shows how many outlets are small, medium, or large.

- **Purpose:** Helps to summarize each variable and prepare for further analysis by understanding its behavior individually.
-

2. Bivariate Analysis

Bivariate analysis explores the relationships between two variables to uncover associations or dependencies. From the dashboard:

- **Key Insights:**
 - **Item MRP vs. Item Sales:** A scatter plot or correlation analysis highlights whether higher MRP items sell more or less, showing potential relationships.
 - **Outlet Size vs. Outlet Sales:** Analyzing outlet size categories against sales can reveal whether larger outlets contribute more to overall sales.
 - **Item Fat Content vs. Item Sales:** Using bar charts or group means, we see how sales vary for items labeled as "Low Fat" versus "Regular."
 - **Purpose:** Identifies trends and correlations to inform decision-making and strategy.
-

3. Multivariate Analysis

Multivariate analysis evaluates the interplay of three or more variables simultaneously, helping to uncover more complex patterns. From the dashboard:

- **Key Insights:**
 - **Item MRP, Outlet Size, and Item Sales:** A multivariate plot (e.g., bubble chart) can reveal how outlet size impacts sales for products with varying MRPs.
 - **Outlet Type, Item Category, and Sales:** Exploring how sales differ across outlet types for various product categories helps identify which combinations perform best.
 - **Item Visibility, Item Sales, and MRP:** Multivariate regression or clustering could show how visibility influences sales across different price points.
 - **Purpose:** Provides a holistic view, capturing insights that individual or paired analyses might miss.
-

Summary Table of Analysis

Analysis Type	Variables Used	Techniques	Purpose/Insights
Univariate	Item MRP, Item Sales, Outlet Size	Histograms, Bar Charts, Summary Statistics	Understand distributions and detect outliers for individual variables.
Bivariate	Item MRP vs. Sales, Outlet Size vs. Sales	Scatter Plots, Correlation, Bar Charts	Examine relationships and trends between two variables (e.g., pricing vs. sales).
Multivariate	Item MRP, Outlet Size, Item Sales	Bubble Charts, Regression, Cluster Analysis	Explore complex relationships involving multiple variables to uncover deeper patterns.

8. STORYTELLING

Bigbasket, a prominent player in the online grocery sector, seeks to optimize sales performance and boost customer satisfaction through data-driven insights. This Power BI dashboard utilizes sales and outlet data to reveal trends, pinpoint challenges, and offer actionable recommendations for enhancing business operations.

1. Challenges Identified

- **Challenge 1: Sales Variability Across Different Outlets**
 - **Explanation:** There are noticeable inconsistencies in the sales data across different outlets. While larger outlets generally show higher sales, some smaller outlets are underperforming, even with similar product assortments.
 - **Impact:** This variability indicates that sales are not only driven by outlet size, but other factors could be affecting performance, such as location, promotional activities, or customer traffic.
- **Challenge 2: Outliers in Pricing and Sales**
 - **Explanation:** The data shows certain items with exceptionally high or low prices and sales numbers, which could be due to data entry errors, special promotions, or unique customer demand.
 - **Impact:** These outliers skew the overall analysis, making it harder to predict average sales and prices accurately.
- **Challenge 3: Missing Data and Inconsistent Category Labels**

- **Explanation:** Missing values in critical fields like "Item Outlet Sales" and inconsistent category labels (e.g., "Low Fat" vs "low fat") make the dataset less reliable and harder to analyze.
 - **Impact:** Incomplete or inconsistent data could lead to incorrect conclusions or biased decisions when performing deeper analysis or building predictive models.
-

2. Key Insights

- **Insight 1: Pricing Strategy Impact on Sales**

- **Explanation:** Analysis revealed that items with a higher MRP generally showed better sales in larger outlets. However, for smaller outlets, high MRP products performed poorly. This suggests that customers in smaller outlets are more price-sensitive.
- **Actionable Insight:** Tailor pricing strategies by outlet size. For smaller outlets, consider offering promotions or discounts to make higher-priced items more attractive.

- **Insight 2: Outlet Size is Not the Sole Factor**

- **Explanation:** While larger outlets tend to have higher sales, some smaller outlets are outperforming their larger counterparts in certain regions. This suggests that other factors, such as location, staff training, or local demand, might be influencing the sales performance.
- **Actionable Insight:** Investigate further into the underperforming smaller outlets. Focus on improving customer experience, staff engagement, or local marketing efforts to boost their performance.

- **Insight 3: Need for Data Standardization**

- **Explanation:** The analysis highlighted the presence of inconsistent labeling, especially in product categories like "Low Fat" and "Regular." Standardizing these labels and imputing missing data using AI will improve the overall quality of the analysis.
 - **Actionable Insight:** Ensure that data entry processes are standardized across all platforms, and fill missing values using AI-based imputation methods to avoid inaccuracies in analysis.
-

3. Recommendations

- **Recommendation 1: Refine Pricing Strategy Based on Outlet Size**

- **Action:** Develop differentiated pricing strategies tailored to the needs of each outlet type. Implement discounts or bundling offers in smaller outlets where price sensitivity is higher, and retain premium pricing for larger outlets.

- **Expected Outcome:** More consistent sales across all outlet types, with a particular focus on increasing sales in smaller outlets.
 - **Recommendation 2: Investigate Underperforming Smaller Outlets**
 - **Action:** Conduct qualitative research (e.g., customer surveys, staff interviews) to understand why certain small outlets are outperforming others. Look into factors like store layout, staff engagement, and location-specific marketing.
 - **Expected Outcome:** Identify successful strategies from high-performing smaller outlets and replicate them in other outlets to improve overall performance.
 - **Recommendation 3: Improve Data Quality**
 - **Action:** Use advanced AI and machine learning tools to clean and standardize the data. Implement regular audits to check for missing values and inconsistencies, and enhance the data collection processes to ensure uniformity.
 - **Expected Outcome:** More accurate and reliable data for future analysis, leading to better decision-making and predictive modeling.
-

Summary:

Through this analysis, we identified key challenges such as sales variability, outliers, and data quality issues. By uncovering valuable insights on pricing strategies, outlet performance, and the need for better data consistency, we are able to make strategic recommendations that could lead to improved sales, optimized pricing, and enhanced decision-making across the business.