

Strategic Inventory & Revenue Analysis: An End-to-End SQL Study of Zepto Retail Data

Executive Summary

The Zepto dataset was analyzed to identify operational efficiencies and growth opportunities. Key findings indicate that while certain categories like **Cooking Essentials** and **Munchies** dominate revenue potential, there are significant missed opportunities in high-value products that are currently out of stock. Additionally, aggressive discounting is concentrated in specific fresh food and snack categories.

1. Data Cleaning & Integrity

Before analysis, the dataset underwent a rigorous cleaning process to ensure accuracy:

- **Price Normalization:** Original price values were adjusted to standard currency units (Rupees) by dividing by 100.0.
- **Zero-Price Removal:** Products listed with an MRP of 0 were deleted to prevent skewed averages and inaccurate revenue projections.
- **Null Handling:** Verification was performed to ensure no critical fields (name, category, MRP) contained null values .

2. Pricing and Promotional Insights

The analysis explored how discounts are distributed across the inventory:

- **Top Value Products:** Heavily promoted items, such as **Dukes Waffly Orange Wafers**, are offered at discounts as high as **51%**.
- **Category Discounts:** **Fruits & Vegetables** offer the highest average discount at **15.46%**, followed by **Meats, Fish & Eggs** at **11.03%**.
- **Premium Pricing:** There are **39 products** with an MRP above 500 Rupees that have less than a 10% discount, including several varieties of Mustard and Groundnut oil.

3. Inventory & Operations Analysis

Inventory was segmented to understand the physical and financial load on the delivery network:

- **Weight Segmentation:** Products were categorized into **Low** (<1kg), **Medium** (<5kg), and **Bulk** (>=5kg) classes to assist in logistics and delivery planning .

- **Revenue at Risk:** High-value items such as **Patanjali Cow's Ghee** (MRP 565) and **MamyPoko Diapers** (MRP 399) are currently **out of stock**, representing a direct loss of potential revenue.
- **Logistics Load:** **Cooking Essentials** and **Munchies** contribute the highest total inventory weight, each exceeding 1.4 million grams, requiring more robust transportation resources.

4. Revenue Potential

By calculating the total value of available stock (selling price × quantity), the categories with the highest financial impact were identified:

- **Top Contributors:** **Cooking Essentials** and **Munchies** lead the platform, each contributing **15.04%** of the total revenue potential.
- **Category Leaders:** Within individual categories, specific items drive the majority of value. For example, **Borges Extra Light Olive Oil** is the top revenue-potential product within the Cooking Essentials category.

5. Strategic Recommendations

- **Restock High-Value Items:** Prioritize restocking the high-MRP items identified in the "out of stock" query to capture missed revenue.
- **Optimize Logistics:** Assign larger delivery vehicles or specialized handling for the **Cooking Essentials** category due to its high total weight and revenue share.
- **Refine Discounting:** Evaluate if the 15%+ average discount on **Fruits & Vegetables** is driving volume or hurting margins, and consider aligning other categories like **Biscuits** (currently 1.11% revenue share) with more aggressive promotions.

Technical Skills Demonstrated

1. Advanced SQL & Database Management (PostgreSQL)

- **Complex Data Cleaning:** Performed schema creation, data type definition, and sanitization by removing zero-price records and normalizing currency units .
- **Data Segmentation:** Utilized CASE WHEN statements to create custom business logic for product weight categorization (Low, Medium, Bulk) .
- **Window Functions:** Implemented OVER() clauses to calculate percentage shares of total revenue without collapsing the dataset .

- **Advanced Ranking:** Employed DENSE_RANK() with PARTITION BY to identify the top 3 revenue-generating products within each specific category .
- **Common Table Expressions (CTEs):** Structured complex multi-step queries using WITH clauses to ensure code readability and maintainability .

2. Data Exploration & Quality Assurance

- **Integrity Checks:** Wrote comprehensive scripts to detect null values across all critical columns (MRP, Category, Stock status) .
- **Descriptive Statistics:** Performed frequency analysis and distinct counts to understand the distribution of product categories and stock availability