Zepto E-Commerce Data Analysis using SQL

Project Documentation | **2025**

Prepared by: **Sada Asmita**

1. Project Overview

This project focuses on analyzing Zepto's e-commerce inventory dataset using SQL. The objective is to uncover actionable business insights related to pricing, discounts, stock availability, and revenue generation. By leveraging SQL for data cleaning and analysis, this project demonstrates how raw transactional data can be transformed into strategic insights to guide decision-making.

2. Dataset Source

The dataset has been sourced from Kaggle:
https://www.kaggle.com/datasets/palvinder2006/zepto-inventory-dataset

It contains product-level information including Category, Product Name, MRP, Discount Percentage, Available Quantity, Discounted Selling Price, Weight in Grams, Stock Status, and Order Quantity. This diversity of attributes makes the dataset suitable for in-depth e-commerce business analysis.

3. Data Preparation

To ensure accuracy and consistency, several data preparation steps were applied:

- Removal of invalid records such as products with zero MRP or selling price.
- Conversion of pricing values from Paise to Rupees for uniformity.
- Identification and removal of duplicate product entries.
- Verification of null values in critical fields like Category, Price, and Stock Status.

4. Business Insights

The SQL analysis of the dataset yielded several important business insights. Each insight is explained below along with its relevance to Zepto's operations.

Insight 1: Top 10 Best-Value Products by Discount %

This insight highlights products with the highest discount percentages. These products are ideal candidates for marketing promotions as they attract price-sensitive customers.

Insight 2: High MRP Products That Are Out of Stock

By identifying premium products that are frequently unavailable, this insight points to strong demand. Restocking them ensures customer satisfaction and prevents high-value sales loss.

Insight 3: Estimated Revenue by Category

This insight estimates total potential revenue across categories based on available stock. It helps prioritize categories that contribute the most to the company's overall revenue potential.

Insight 4: High MRP but Low Discount Products

These are luxury or premium products that Zepto does not heavily discount. The insight supports differentiated pricing strategies for premium customer segments.

Insight 5: Top 5 Categories by Average Discount %

This insight highlights categories with the most aggressive discounting strategies. It enables Zepto to evaluate the effectiveness of promotions and understand how discounts stimulate customer demand and purchasing behavior.

Insight 6: Price per Gram for Products above 100g

By analyzing the price per gram of bulk items, Zepto can evaluate packaging efficiency and offer value-based pricing strategies for cost-conscious customers.

Insight 7: Product Weight Segmentation

Products are segmented into weight categories such as Low, Medium, and Bulk. This classification is particularly useful for logistics planning and optimizing delivery costs.

Insight 8: Total Inventory Weight by Category

This insight calculates the overall weight of inventory for each category. It assists in supply chain planning, warehousing, and category-level resource allocation.

Insight 9: Most Expensive Product per Category

This identifies the highest-priced product in each category. It provides visibility into premium products that can be strategically marketed to high-value customers.

5. Conclusion

Through SQL-based analysis, this project transformed raw product data into actionable insights. The findings highlight opportunities in inventory management, discount strategies, and revenue optimization. By understanding demand patterns, pricing efficiency, and stock availability, Zepto can make informed decisions that improve both business performance and customer satisfaction.