**Zepto Inventory Analysis - Business Report**

**Prepared by: Hitanshi Shroff**

**Role:** Data Analyst   
**Tools Used:** PostgreSQL, pgAdmin, Excel

**Executive Summary**

This report provides a detailed analysis of Zepto’s inventory dataset, simulating a real world ecommerce data analyst's workflow. The objective was to draw meaningful insights from raw product data, improve business decision making, and highlight areas with growth or optimization potential. The focus areas include pricing trends, stock availability, discount strategies, and category wise product performance.

**Dataset Overview**

The dataset was originally scraped from Zepto’s official product listings and mimics what a real retail inventory system looks like:

* Each row represents a unique product SKU.
* Duplicate product names exist due to variations in weight, packaging, or discount.
* The dataset reflects actual business scenarios like missing data, inconsistent formats, and mixed units.

Key columns included:

* **Product Name**
* **Category**
* **MRP & Discounted Selling Price**
* **Stock Availability**
* **Weight**
* **Discount %**

**Step-by-Step Breakdown**

**1. Data Cleaning**

Before diving into analysis, the dataset was cleaned to reflect business ready data:

* **Removed invalid entries**: Products with MRP or discounted price marked as ₹0 were excluded.
* **Handled missing values**: Nulls were identified and logically imputed where required.
* **Standardized pricing**: All prices were converted from paise to rupees for readability and consistency.

**Outcome:** A clean, reliable dataset ready for exploration and insights.

**2. Exploratory Data Analysis (EDA)**

Key questions explored:

* What categories have the most products?
* How many products are out of stock?
* What’s the average discount across categories?
* Are there products listed multiple times?

**Findings:**

* **Snacks** and **Beverages** were the most listed categories.
* Over **20%** of products were marked **out of stock**, suggesting either high demand or supply chain inefficiencies.
* A significant number of products had multiple SKUs due to varied pack sizes or weights a common e commerce tactic for better visibility.

**3. Business Insights**

**Top Discounted Products**

The top 10 products with the highest discounts were identified useful for promotional targeting and upselling.

**High MRP But Out-of-Stock**

Several expensive products (MRP > ₹500) were unavailable. This could indicate missed revenue opportunities or logistic challenges.

**Revenue Potential by Category**

By estimating revenue as discounted price × available quantity, we found:

* **Beverages** and **Household items** showed high revenue potential.
* **Personal Care** had lower inventory but higher per-unit prices.

**Value-for-Money Analysis**

Calculated price per gram to find the best value Product important for pricing strategy and customer satisfaction.

**Weight-Based Grouping**

Products were categorized into:

* **Low weight (0–250g)**
* **Medium weight (251g–500g)**
* **Bulk (500g+)**

Bulk items had fewer listings but more stock indicating less frequent purchase but higher basket size per order.

**Recommendations**

Based on the insights, here are actionable suggestions for business stakeholders:

1. **Focus on replenishing out-of-stock, high value items** to prevent lost sales.
2. **Review discount strategy** for expensive items some high MRP products had very minimal discounts.
3. **Optimize product visibility** by prioritizing high revenue SKUs within top performing categories.
4. **Introduce more SKUs in the bulk category**, especially for frequently purchased products like rice, flour, or beverages.
5. **Monitor overstocked low-performing SKUs** to avoid storage costs and waste.

**Analyst's Reflection**

Working on this project allowed me to understand how raw, messy data can hold incredible value when structured and analyzed effectively. From identifying gaps in stock availability to unlocking patterns in discounting, SQL proved to be a powerful tool for turning data into decisions.

This hands on experience enhanced my confidence in:

* Writing clean, optimized SQL queries
* Building a logical analysis pipeline
* Translating technical data into business value