
E-Commerce Inventory Analytics Report

Project: Zepto Product Listings (Scraped & Kaggle Dataset)
Analyst: Ayush

---

1. Dataset Overview

The dataset represents a real-world e-commerce inventory system, mimicking Zepto's product catalog. Each row corresponds to a unique SKU (Stock Keeping Unit), with duplicate product names appearing across different package sizes, weights, discounts, or categories — reflecting how catalog data is structured in practice.

Columns:

- **sku_id** – Synthetic primary key
- **name** – Product name (raw app listing)
- **category** – Product category (Fruits, Snacks, Beverages, etc.)
- **mrp** – Maximum Retail Price (converted from paise to ₹)
- **discountPercent** – Discount applied on MRP
- **discountedSellingPrice** – Final selling price (converted to ₹)
- **availableQuantity** – Units available in inventory
- **weightInGms** – Product weight in grams
- **outOfStock** – Boolean flag for availability
- **quantity** – Number of units per package (mixed with grams for loose produce)

---

2. Database Setup

A PostgreSQL table was created with appropriate data types to ensure scalability and accuracy:

```
CREATE TABLE zepto (
  sku_id SERIAL PRIMARY KEY,
  category VARCHAR(120),
  name VARCHAR(150) NOT NULL,
```

```
 mrp NUMERIC(8,2),
 discountPercent NUMERIC(5,2),
 availableQuantity INTEGER,
 discountedSellingPrice NUMERIC(8,2),
 weightInGms INTEGER,
 outOfStock BOOLEAN,
 quantity INTEGER
);
```

Data was imported using \copy with UTF-8 encoding fixes to handle special characters in product names.

---

3. Exploratory Data Analysis (EDA)

Key exploration steps included:

- Verified total rows imported successfully.
- Viewed random rows to understand structure and anomalies.
- Checked for missing entries across category, weightInGms, and discountPercent.
- Extracted distinct product categories.
- Compared in-stock vs out-of-stock counts.
- Detected multiple SKUs for the same product name (different sizes/discounts).

---

4. Data Cleaning

Cleaning ensured business-ready data:

- Removed rows where mrp = 0 or discountedSellingPrice = 0.
- Converted mrp and discountedSellingPrice from paise to rupees.
- Standardized null handling (e.g., missing weights replaced with averages per category).

---

5. Business Insights

SQL queries were designed to answer business-driven questions:

5.1 Best-Value Products

```
SELECT name, category, discountPercent
FROM zepto
ORDER BY discountPercent DESC
LIMIT 10;
```

Identified top 10 products offering the highest discounts.

5.2 High-MRP Out-of-Stock Products

```
SELECT name, category, mrp
FROM zepto
WHERE outOfStock = TRUE AND mrp > 500;
```

Flagged premium products unavailable for purchase — critical for supply chain alerts.

5.3 Potential Revenue by Category

```
SELECT category,
    SUM(discountedSellingPrice * availableQuantity) AS potential_revenue
FROM zepto
GROUP BY category
ORDER BY potential_revenue DESC;
```

Estimated revenue opportunity per category, useful for sales prioritization.

5.4 Expensive Products with Minimal Discount

```
SELECT name, category, mrp, discountPercent
FROM zepto
WHERE mrp > 500 AND discountPercent < 5;
```

Highlighted luxury items with negligible discounts — potential pricing strategy gap.

5.5 Top Discounted Categories

```
SELECT category,
    AVG(discountPercent) AS avg_discount
FROM zepto
GROUP BY category
ORDER BY avg_discount DESC
LIMIT 5;
```

Ranked top 5 categories offering the highest average discounts.

5.6 Price per Gram Analysis

```
SELECT name, category,
    discountedSellingPrice / weightInGms AS price_per_gram
FROM zepto
WHERE weightInGms > 0
ORDER BY price_per_gram ASC
LIMIT 10;
```

Identified value-for-money products based on unit economics.

5.7 Weight-Based Grouping

```
SELECT name, category,
    CASE
     WHEN weightInGms < 250 THEN 'Low'
     WHEN weightInGms BETWEEN 250 AND 1000 THEN 'Medium'
     ELSE 'Bulk'
    END AS weight_category
FROM zepto;
```

Grouped products into Low, Medium, Bulk categories for logistics planning.

5.8 Inventory Weight by Category

```
SELECT category,
    SUM(weightInGms * availableQuantity) AS total_inventory_weight
FROM zepto
GROUP BY category
ORDER BY total_inventory_weight DESC;
```

Measured total inventory weight per category, aiding warehouse optimization.

---

6. Key Takeaways

- Data cleaning was essential to remove invalid entries and standardize pricing.
- EDA revealed duplicates and stock inconsistencies, common in real-world catalogs.
- Business insights provided actionable intelligence:
    - High-discount products drive customer acquisition.
    - Out-of-stock premium items highlight supply chain risks.
    - Category-level revenue and discount analysis informs pricing and promotion strategies.
    - Weight-based grouping supports logistics and packaging decisions.

---

7. Conclusion

This project successfully simulates how data analysts in e-commerce use SQL to transform messy catalog data into business-ready insights. The workflow demonstrates end-to-end analytics:

- Database setup
- Data import and cleaning
- Exploratory analysis
- Business-driven SQL queries

The final outputs provide strategic visibility into pricing, inventory, and revenue — exactly the kind of analysis that supports decision-making in retail and e-commerce industries.