Zepto Retail Inventory & Pricing Analysis

Prepared by: Risit Sahoo

Project Type: SQL-Based Exploratory Data Analysis Project

Date – 28.11.2025

1. Introduction

In today's fast-paced e-commerce environment, organizations rely heavily on data-driven insights to manage inventory, optimize pricing strategies, improve product availability, and enhance customer experience.

This project focuses on performing an end-to-end SQL-based data analysis on a real-world Zepto grocery delivery dataset. The objective is to analyze product pricing, discounts, stock availability, and category-level performance using structured query language (SQL).

The project demonstrates strong capabilities in:

➢ Data exploration
➢ Data cleaning and transformation
➢ Business-driven SQL querying
➢ Revenue and inventory analysis
➢ Translating raw data into actionable insights

This analysis closely resembles the real responsibilities of a Data Analyst working in retail or e-commerce companies.

2. Business Problem Statement

Zepto operates with thousands of SKUs across multiple product categories. Managing such a large catalog requires clear insights into:

➢ Product availability and stock issues
➢ Discount effectiveness
➢ Revenue contribution by category
➢ High-value products with poor stock availability
➢ Pricing efficiency across different product weights

Without proper analysis, the business may face:

➢ Revenue leakage due to poor pricing strategies
➢ Overstocking or understocking issues
➢ Inefficient discount allocation
➢ Low-margin product prioritization

3. Business Objectives

The primary objectives of this SQL analysis are:

➢ Understand the structure and quality of the dataset
➢ Identify missing or inconsistent data
➢ Analyze stock availability across products
➢ Evaluate discount distribution and pricing strategy
➢ Estimate revenue contribution by category
➢ Identify high-value and low-performing products
➢ Support business decision-making using SQL-driven insights

4. Dataset Overview

The dataset represents Zepto's product catalog data containing the following attributes:

Column Name	Description
sku_id	Unique product SKU identifier
product_name	Name of the product
category	Product category
mrp	Maximum Retail Price
discountPercent	Discount percentage applied
discountedSellingPrice	Final selling price
weightInGms	Product weight in grams
availableQuantity	Available inventory units
outOfStock	Stock availability indicator
quantity	Package quantity

5. Tools, Technologies & Skills Used

Tools

➢ MySQL Workbench
➢ Microsoft Excel (CSV preprocessing)

Technical Skills

➢ SQL Queries
➢ Data Exploration (EDA)
➢ Data Cleaning & Transformation
➢ Aggregations & Joins
➢ Conditional Logic (CASE statements)
➢ Business KPI Analysis

SQL Concepts Applied

➢ SELECT statements
➢ WHERE filtering
➢ GROUP BY & HAVING clauses
➢ ORDER BY

- ➢ Aggregate functions (SUM, AVG, COUNT)
- ➢ CASE expressions
- ➢ Boolean logic

6. Database Design & Table Creation

The dataset was imported into MySQL Workbench and structured into a relational table named zepto.

Key design considerations:

- ➢ Appropriate numeric data types for price columns
- ➢ Boolean values for stock availability
- ➢ Standardized naming conventions
- ➢ UTF-8 encoding to prevent import errors

7. Data Exploration

7.1. Total Number of Records

SELECT COUNT(*) FROM zepto;

Purpose:
To understand dataset size and processing scope.

7.2. Sample Data Inspection

SELECT * FROM zepto

LIMIT 10;

Purpose:
To review column structure, data types, and value patterns.

7.3. NULL Value Detection

SELECT * FROM zepto

WHERE product_name IS NULL

OR category IS NULL

OR mrp IS NULL

OR discountPercent IS NULL

OR discountedSellingPrice IS NULL

OR weightInGms IS NULL

OR availableQuantity IS NULL

OR outOfStock IS NULL

OR quantity IS NULL;

Insight:
Ensures data completeness before analysis.

7.4. Unique Product Categories

```
SELECT DISTINCT category

FROM zepto

ORDER BY category;
```

Business Use:
Helps in category-level revenue and inventory analysis.

7.5. Stock Availability Analysis

```
SELECT outOfStock, COUNT(sku_id)

FROM zepto

GROUP BY outOfStock;
```

Insight:
Identifies proportion of products currently unavailable for sale.

7.6. Duplicate Product Names Across SKUs

```
SELECT product_name, COUNT(sku_id) AS Number_of_SKUs

FROM zepto

GROUP BY product_name

HAVING COUNT(sku_id) > 1

ORDER BY COUNT(sku_id) DESC;
```

Business Insight:
Same product may exist in different package sizes or weights.

8. Data Cleaning & Transformation

8.1. Price Conversion (Paise → Rupees)

```
UPDATE zepto

SET mrp = mrp / 100.0,

discountedSellingPrice = discountedSellingPrice / 100.0;
```

Reason:
Prices were stored in paise and converted to rupees for accurate financial reporting.

8.2 Validation After Conversion

```
SELECT mrp, discountedSellingPrice FROM zepto;
```

9. Data Analysis & Business Queries

9.1 Top 10 Best-Value Products

```
SELECT DISTINCT product_name, mrp, discountPercent

FROM zepto
```

```
ORDER BY discountPercent DESC
```

```
LIMIT 10;
```

Insight:
Helps identify aggressive discount strategies.

9.2. High-MRP Products Out of Stock

```
SELECT DISTINCT product_name, mrp
```

```
FROM zepto
```

```
WHERE outOfStock = TRUE
```

```
AND mrp > 300
```

```
ORDER BY mrp DESC;
```

Business Impact:
Potential revenue loss due to unavailable premium products.

9.3 Estimated Revenue by Category

```
SELECT category,
```

```
SUM(discountedSellingPrice * availableQuantity) AS total_revenue
```

```
FROM zepto
```

```
GROUP BY category
```

```
ORDER BY total_revenue DESC;
```

Key KPI:
Category-level revenue contribution.

9.4 High-Priced Products With Low Discounts

```
SELECT DISTINCT product_name, mrp, discountPercent
```

```
FROM zepto
```

```
WHERE mrp > 500
```

```
AND discountPercent < 10
```

```
ORDER BY mrp DESC;
```

Insight:
Identifies premium products with limited promotional offers.

9.5 Top 5 Categories With Highest Average Discount

```
SELECT category,
```

```
ROUND(AVG(discountPercent), 2) AS avg_discount
```

```
FROM zepto
```

GROUP BY category

ORDER BY avg_discount DESC

LIMIT 5;

Use Case:
Evaluates category-wise discount strategy.

9.6 Price Per Gram Analysis

SELECT DISTINCT product_name, weightInGms, discountedSellingPrice,

ROUND(discountedSellingPrice / weightInGms, 2) AS price_per_gram

FROM zepto

WHERE weightInGms >= 100

ORDER BY price_per_gram;

Insight:
Identifies best-value products based on unit economics.

9.7 Product Weight Classification

SELECT DISTINCT product_name, weightInGms,

CASE

   WHEN weightInGms < 1000 THEN 'Low'

   WHEN weightInGms < 5000 THEN 'Medium'

   ELSE 'Bulk'

END AS weight_category

FROM zepto;

Purpose:
Supports packaging and logistics segmentation.

9.8 Total Inventory Weight by Category

SELECT category,

SUM(weightInGms * availableQuantity) AS total_weight

FROM zepto

GROUP BY category

ORDER BY total_weight DESC;

Business Use:
Warehouse capacity and supply-chain planning.

1O. Key Business Insights

➢ Several high-value SKUs are out of stock, indicating potential revenue loss
➢ Certain categories contribute disproportionately higher revenue
➢ Discounts are unevenly distributed across categories
➢ Bulk-weight products show better price efficiency
➢ Duplicate SKUs indicate multiple pack-size strategies
➢ Inventory weight concentration varies significantly by category

11. Conclusion

This SQL project demonstrates a complete data analyst workflow, including:

➢ Structured data exploration
➢ Data cleaning and normalization
➢ Advanced SQL querying
➢ Business KPI evaluation
➢ insight generation aligned with real e-commerce operations

The analysis provides valuable insights that can help businesses:

➢ Improve inventory availability
➢ Optimize pricing and discount strategy
➢ Focus on high-revenue categories
➢ Reduce stock-out related losses

This project reflects real-world analytical responsibilities expected from a Data Analyst in retail, FMCG, or e-commerce organizations.