Zepto Database Exploration and Querying

🏷️ 1. Introduction

Zepto is one of India's fastest-growing instant grocery delivery platforms. It offers a wide range of products—from groceries and personal care to household essentials—delivered within minutes.

In this SQL project, we explored a dataset named **zepto_v2**, which contains product-level information such as:

- Product names
- Categories
- Prices and discounts
- Availability and stock levels
- Product weight and estimated revenue

The goal of this project was to perform **data exploration, cleaning, and analysis** using SQL to uncover valuable insights into Zepto's product inventory and sales dynamics.

⚙️ 2. Project Objectives

The main objectives of this project were:

1. To explore and understand the structure of the zepto_v2 dataset.
2. To perform **data cleaning** and handle missing or invalid values.
3. To derive **business insights** through SQL-based analytical queries.
4. To summarize key metrics such as discounts, stock levels, and category performance.

🧩 3. Dataset Overview

The dataset zepto_v2 includes the following key columns:

Column Name	Description
name	Product name
category	Product category (e.g., Beverages, Snacks, Dairy)
MRP	Maximum Retail Price
Discount_Percentage	Discount offered on product
discounted_selling_price	Final selling price after discount
availability_Quantity	Available quantity in stock
Out_of_stock	Indicates whether the product is out of stock (1 = Yes, 0 = No)
weight_in_grms	Product weight in grams
quantity	Pack or unit count

🧹 4. Data Cleaning and Preparation

Several data quality checks and transformations were applied before analysis:

📜 Steps Performed:

1. **Added a Serial Number (SKU ID)**
    - Added a new column sku_id using AUTO_INCREMENT to uniquely identify each product.
2. ALTER TABLE zepto_v2 ADD COLUMN sku_id INT AUTO_INCREMENT PRIMARY KEY FIRST;
3. **Counted Total Records**
   Verified total number of rows in the dataset.
4. SELECT COUNT(*) FROM zepto_v2;

5. **Null Value Analysis**
   Checked for missing data in key fields like name, category, MRP, discount, etc.
6. **Removed Invalid Entries**
   Deleted rows where MRP = 0, as they represent incomplete or incorrect data.
7. DELETE FROM zepto_v2 WHERE MRP = 0;
8. **Data Formatting**
   Converted price from **paise to rupees** for better readability:
9. UPDATE zepto_v2 SET MRP = MRP / 100.0;

📊 5. Data Exploration and Insights

Q1️⃣ Top 10 Best-Value Products Based on Discount Percentage

```
SELECT name, category, MRP, Discount_Percentage
FROM zepto_v2
ORDER BY Discount_Percentage DESC
LIMIT 10;
```

Insight: Identifies products offering the highest discounts—useful for promotional strategies.

Q2️⃣ Products with High MRP but Out of Stock

```
SELECT name, category, MRP, Out_of_stock
FROM zepto_v2
WHERE MRP > 500 AND Out_of_stock = 1;
```

Insight: Highlights premium products that are currently unavailable, indicating potential restocking opportunities.

Q 3 Estimated Revenue by Category

```
SELECT category,
    SUM(discounted_selling_price * availability_Quantity) AS estimated_revenue
FROM zepto_v2
GROUP BY category
ORDER BY estimated_revenue DESC;
```

Insight: Estimates potential revenue per category to identify top-performing product lines.

Q 4 High MRP but Low Discount (<10%)

```
SELECT name, category, MRP, Discount_Percentage
FROM zepto_v2
WHERE MRP > 500 AND Discount_Percentage < 10;
```

Insight: Finds expensive products offering minimal discounts—helpful for pricing strategy optimization.

Q 5 Top 5 Categories Offering Highest Average Discounts

```
SELECT category,
    AVG(Discount_Percentage) AS avg_discount
FROM zepto_v2
GROUP BY category
ORDER BY avg_discount DESC
LIMIT 5;
```

Insight: Reveals which categories attract customers through heavier discounts.

Q 6 Price per Gram for Products Above 100g

```
SELECT name, category, discounted_selling_price, weight_in_grms,
    (discounted_selling_price / weight_in_grms) AS price_per_gram
FROM zepto_v2
WHERE weight_in_grms > 100
ORDER BY price_per_gram ASC;
```

Insight: Helps determine which products provide the best value per gram.

Q 7 Categorizing Products by Weight Range

```
SELECT name, category, weight_in_grms,
    CASE
     WHEN weight_in_grms < 100 THEN 'Low'
     WHEN weight_in_grms BETWEEN 100 AND 500 THEN 'Medium'
     ELSE 'Bulk'
    END AS weight_category
FROM zepto_v2;
```

Insight: Groups products by size (Low, Medium, Bulk) for better inventory planning.

Q 8 Total Inventory Weight Per Category

```
SELECT category,
    SUM(weight_in_grms * availability_Quantity) AS total_inventory_weight
FROM zepto_v2
GROUP BY category
```

```
ORDER BY total_inventory_weight DESC;
```

Insight: Shows which product categories occupy the most space in storage or logistics.

📈 6. Key Findings

- **Top Discount Categories:** Certain categories consistently offer higher discounts—useful for marketing.
- **Stock Management:** Several high-value items are **out of stock**, suggesting better inventory forecasting is needed.
- **Revenue Drivers:** A small number of categories contribute disproportionately to total revenue.
- **Data Cleaning Impact:** Removing zero-MRP items improved accuracy of financial calculations.

🚀 7. Conclusion

This project demonstrates how SQL can be effectively used for **data exploration, cleaning, and business analysis** in a retail/e-commerce context. By leveraging queries on the Zepto dataset, we gained insights into:

- Discount strategies
- Inventory health
- Revenue estimation
- Product categorization and pricing

These insights can support data-driven decision-making for inventory optimization, promotions, and supply chain planning.

💡 8. Future Scope

- Integrate real-time sales data to estimate live revenue.
- Build dashboards using Power BI or Tableau for visualization.
- Extend analysis to customer behavior and order frequency.