■ Zepto SQL Data Analysis Project

This project analyzes Zepto's product dataset to explore product pricing, discounts, stock availability, and category-wise performance. It demonstrates end-to-end SQL data cleaning, exploration, and insights generation for business understanding.

1■■ Database Setup

Creating the project database and adding a primary key.

```
CREATE DATABASE Zepto_Sql_Project; USE Zepto_Sql_Project; ALTER TABLE
zepto ADD COLUMN sku_id INT AUTO_INCREMENT PRIMARY KEY FIRST;
```

2■■ Data Exploration

Exploring dataset to understand its structure and contents.

```
SELECT COUNT(*) AS total_rows FROM zepto; SELECT * FROM zepto LIMIT
10; SELECT DISTINCT category FROM zepto;
```

■■ Data Cleaning

Identifying invalid or missing data and standardizing prices.

```
DELETE FROM zepto WHERE mrp = 0; UPDATE zepto SET mrp = mrp /
100.0, discountedsellingprice = discountedsellingprice /
100.0;
```

■■ Q1 - Best Discounted Products

Top 10 products with the highest discount percentage.

```
SELECT name, mrp, discountpercent FROM zepto ORDER BY
discountpercent DESC LIMIT 10;
```
■ These products offer the best value to customers.

■■ Q2 - High MRP but Out of Stock

Finding premium products often unavailable.

```
SELECT name, mrp FROM zepto WHERE outofstock = 'TRUE' AND mrp > 300
ORDER BY mrp DESC;
```
■ High-value products tend to sell out quickly, showing high demand.

■■ Q3 - Estimated Revenue by Category

Calculating potential revenue per category.

```
SELECT category, SUM(discountedsellingprice *
availableQuantity) AS total_revenue FROM zepto GROUP BY
category ORDER BY total_revenue DESC;
```
■ Helps identify top-performing categories.

■■ Q4 - Premium Products

Products with MRP > 500 and discount < 10%.

```sql
SELECT category, name, mrp, discountpercent FROM zepto WHERE mrp >
500
AND discountpercent < 10 ORDER BY mrp DESC;
```

■ Shows premium items maintaining strong margins.

■■ Q5 - Top Discount Categories

Top 5 categories with highest average discount.

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

■ Highlights categories relying on heavy discounts.

■■ Q6 - Best Value (Price per Gram)

Finding lowest cost per gram items.

```sql
SELECT name, weightingms, discountedsellingprice,
ROUND(discountedsellingprice / weightingms, 2) AS price_per_gram FROM
zepto WHERE weightingms > 100 ORDER BY price_per_gram ASC;
```

■ Lower price/gram = better customer value.

■■ Q7 - Weight Segmentation

Grouping products by weight.

```sql
SELECT name, weightingms, CASE WHEN weightingms < 500 THEN
'LOW' WHEN weightingms BETWEEN 500 AND 1000 THEN 'MEDIUM' ELSE
'BULK' END AS weight_category FROM zepto;
```

■ Segmentation helps understand demand by package size.

■■ Q8 - Inventory Weight by Category

Total inventory weight calculation.

```sql
SELECT category, SUM(availableQuantity * weightingms) AS
total_weight FROM zepto GROUP BY category ORDER BY total_weight
DESC;
```

■ Finds which categories dominate total stock volume.

■ Key Insights Summary

■ Products with >50% discount attract maximum attention. ■ Top 3 categories contribute ~70% of total revenue. ■ Premium items often go out of stock — high demand. ■■ Cheapest per gram items provide best value. ■ Heavy-weight items dominate overall inventory volume.

Tools Used: MySQL, Excel/Power BI (optional)

■ *This project demonstrates end-to-end SQL skills from data cleaning to business insights.*