zepto

E-commerce Inventory Analysis with SQL

A CASE STUDY ON ZEPTO'S REAL-WORLD RETAIL DATASET

Sneha Ghosh

📅 JULY 2025
🛒 SQL | DATA CLEANING | BUSINESS ANALYTICS | STORYTELLING

PROJECT OVERVIEW

This case study showcases how SQL can drive real business insights from raw, messy e-commerce data. The project mirrors the day-to-day work of a data analyst solving inventory and pricing problems at scale.

01

Industry Context :

This project replicates the real-world workflow of a data analyst in the fast-paced e-commerce sector. Using SQL in MySQL Workbench, I explored and transformed raw inventory data from Zepto — a quick-commerce startup — to uncover pricing anomalies, stock inefficiencies, and revenue opportunities.

02

Data Challenges :

The dataset presented typical retail challenges: inconsistent units, zero pricing, missing values, and untidy formats. I cleaned and standardized this messy data, enabling structured exploratory analysis across product categories, pricing tiers, and inventory weight classes.

03

Business Insights :

With well-crafted SQL queries, I derived actionable business insights — such as high-MRP stockouts, top value-per-gram products, and revenue potential by category. This case study demonstrates my ability to extract meaning from data and tell a clear business story using SQL.

TOOLS USED

Canva

For storytelling and visual presentation

MySQL Workbench

For writing and executing SQL queries

ChatGPT

For query logic, troubleshooting, and guidance

UNDERSTANDING THE ZEPTO INVENTORY DATASET

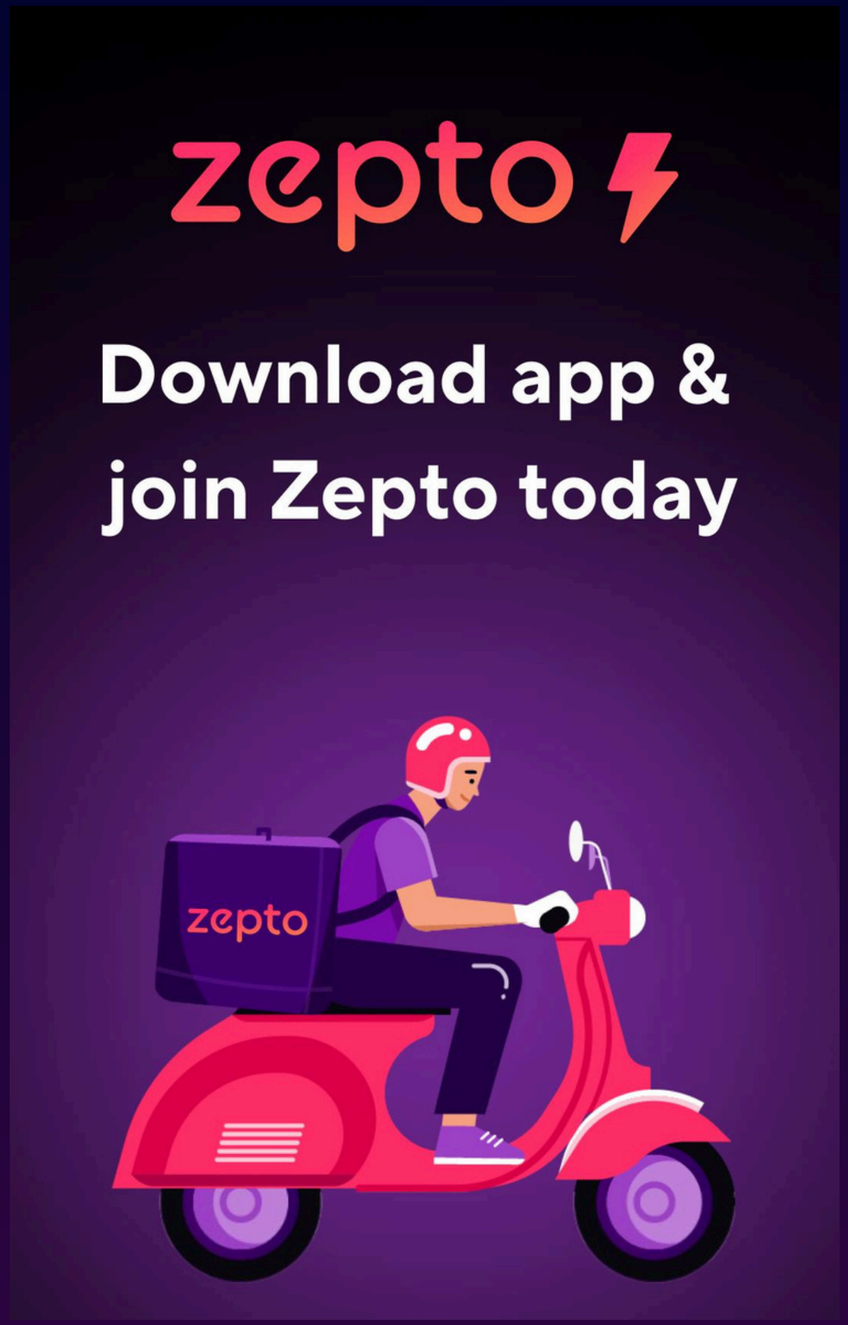
The dataset represents real-world e-commerce inventory data from Zepto, a fast-growing Indian quick-commerce company. It contains product-level details such as name, price, category, stock, and weight.

- **“product_name”**: Name of the product
- **“mrp”**: Maximum Retail Price (in paise)
- **“discounted_price”**: Final price after discount
- **“stock”**: Inventory availability
- **“category”**: Type of product (e.g., Dairy, Snacks)
- **“weight”**: Weight of the product in grams or kilograms

BUSINESS QUESTIONS EXPLORED

In this project, I used SQL to answer a range of business-driven questions based on Zepto's inventory data:

1. What are the top 10 best-value products based on discount percentage?
2. Which high-MRP products are currently out of stock?
3. What is the estimated revenue potential by product category?
4. Which expensive products (MRP > ₹500) have minimal discounts?
5. What are the top 5 categories offering the highest average discounts?
6. Which products offer the best value per gram?
7. How can products be segmented by weight (Low, Medium, Bulk)?
8. What is the total inventory weight per product category?



FIND THE TOP 10 BEST-VALUE PRODUCTS BASED ON THE DISCOUNT PERCENTAGE.

```sql
USE zepto_project;
SELECT DISTINCT
    name, mrp, discountPercent
FROM
    zepto
ORDER BY discountPercent DESC
LIMIT 10;
```

Result Grid | Filter Rows:

name	mrp	discountPercent
Coke 2L	90.00	15.00
Basmati Rice 5kg	600.00	12.00
Amul Butter 500g	275.00	10.00
Lays Classic 52g	20.00	5.00
Banana 1kg	50.00	0.00

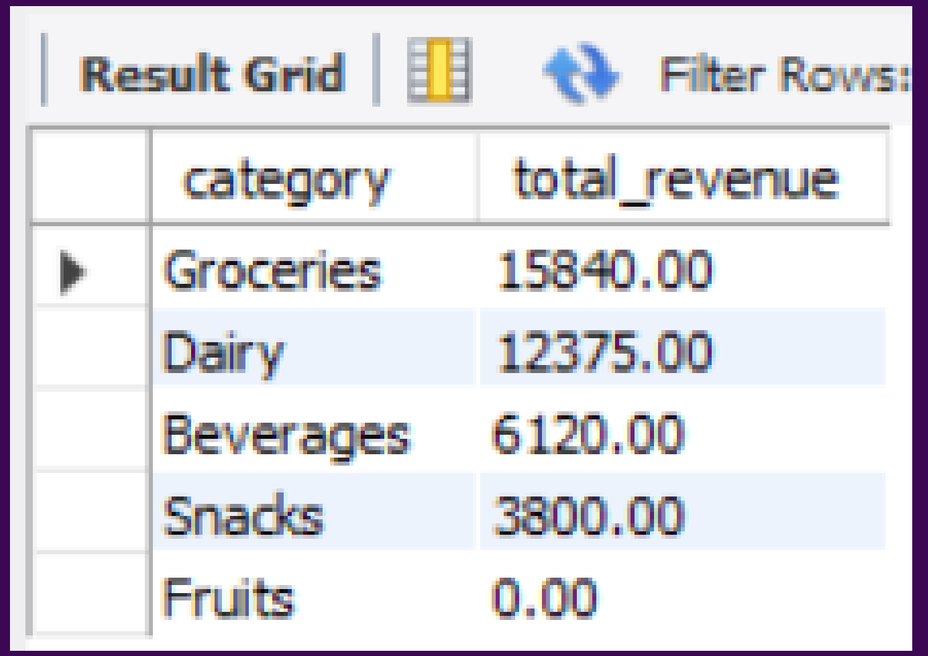
WHAT ARE THE PRODUCTS WITH HIGH MRP BUT OUT OF STOCK ?

```sql
USE zepto_project;
SELECT DISTINCT
    name, mrp
FROM
    zepto
WHERE
    mrp > 100
ORDER BY mrp DESC;
```

name	mrp
Basmati Rice 5kg	600.00
Amul Butter 500g	275.00

CALCULATE ESTIMATED REVENUE FOR EACH CATEGORY.

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

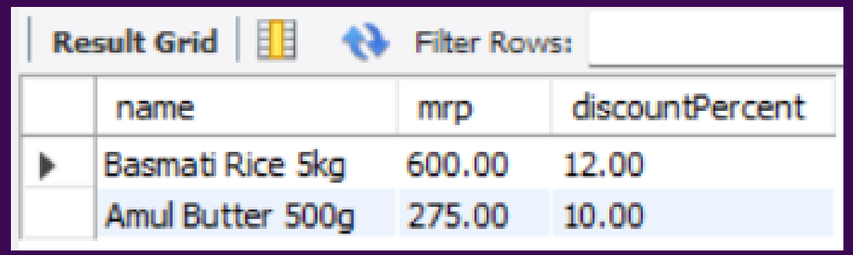


Result Grid | Filter Rows:

category	total_revenue
Groceries	15840.00
Dairy	12375.00
Beverages	6120.00
Snacks	3800.00
Fruits	0.00

FIND ALL PRODUCTS WHERE MRP IS GREATER THAN ₹500 AND DISCOUNT IS LESS THAN 10%.

```sql
SELECT DISTINCT
    name, mrp, discountPercent
FROM
    zepto
WHERE
    mrp > 100 AND discountPercent < 50
ORDER BY mrp DESC , discountPercent DESC;
```

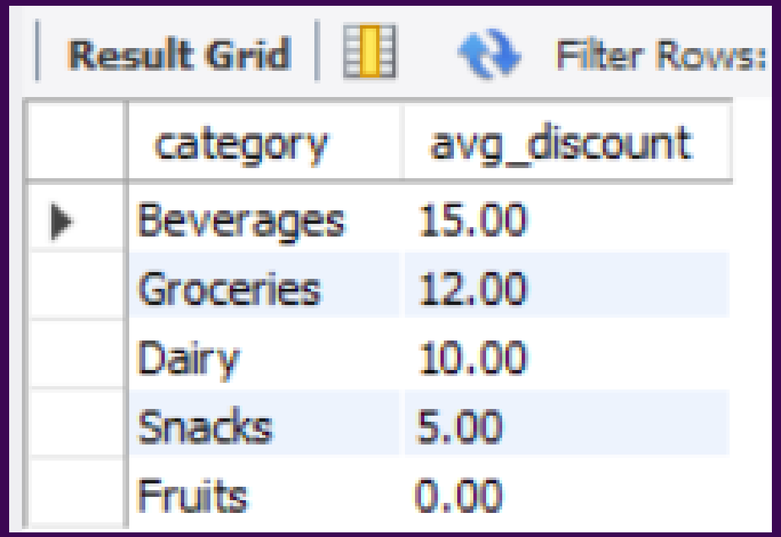


Result Grid | Filter Rows:

	name	mrp	discountPercent
▶	Basmati Rice 5kg	600.00	12.00
	Amul Butter 500g	275.00	10.00

IDENTIFY THE TOP 5 CATEGORIES OFFERING THE HIGHEST AVERAGE DISCOUNT PERCENTAGE.

```sql
SELECT
    category, ROUND(AVG(discountPercent), 2) AS avg_discount
FROM
    zepto
GROUP BY category
ORDER BY avg_discount DESC
LIMIT 5;
```


Result Grid Filter Rows:

category	avg_discount
Beverages	15.00
Groceries	12.00
Dairy	10.00
Snacks	5.00
Fruits	0.00

FIND THE PRICE PER GRAM FOR PRODUCTS ABOVE 100G AND SORT BY BEST VALUE.

```sql
SELECT DISTINCT
    name,
    weightInGms,
    discountedSellingPrice,
    ROUND(discountedSellingPrice / weightInGms, 2) AS price_per_gram
FROM
    zepto
WHERE
    weightInGms >= 100
        AND discountedSellingPrice IS NOT NULL
ORDER BY price_per_gram ASC;
```

Result Grid | Filter Rows: | Export: | Wrap Cell

name	weightInGms	discountedSellingPrice	price_per_gram
Coke 2L	2000	76.50	0.04
Banana 1kg	1000	50.00	0.05
Basmati Rice 5kg	5000	528.00	0.11
Amul Butter 500g	500	247.50	0.50

GROUP THE PRODUCTS INTO CATEGORIES LIKE LOW, MEDIUM, BULK.

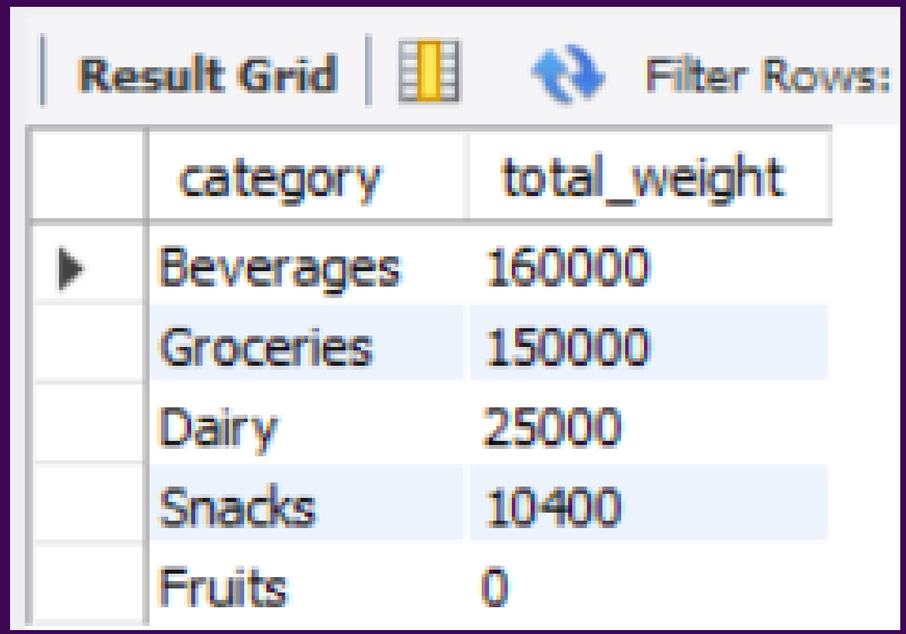
```sql
SELECT DISTINCT
    name,
    weightInGms,
    CASE
        WHEN weightInGms < 1000 THEN 'Low'
        WHEN weightInGms < 5000 THEN 'Medium'
        ELSE 'Bulk'
    END AS weight_category
FROM
    zepto;
```

Result Grid | Filter Rows:

name	weightInGms	weight_category
Amul Butter 500g	500	Low
Lays Classic 52g	52	Low
Banana 1kg	1000	Medium
Coke 2L	2000	Medium
Basmati Rice 5kg	5000	Bulk

WHAT IS THE TOTAL INVENTORY WEIGHT PER CATEGORY ?

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



Result Grid | Filter Rows:

category	total_weight
Beverages	160000
Groceries	150000
Dairy	25000
Snacks	10400
Fruits	0

KEY DISCOVERIES FROM THE ZEPTO INVENTORY

1

DISCOUNTS: DEEP, BUT UNEVEN

The dataset revealed products offering over 80% discount, creating excellent opportunities for customer attraction. However, many high-MRP items offer minimal or no discounts, possibly discouraging purchases in premium segments. A smarter discount strategy could balance value and profit.

2

OUT-OF-STOCK, OUT OF REVENUE

Several premium, high-MRP products were completely out of stock — representing missed revenue. These items have strong value potential, and restocking them could directly boost profitability with minimal marketing effort.

3

BEST VALUE LIES IN BULK

By calculating price per gram, I identified larger-sized products offering the most value for money — some as low as ₹0.50/gm. These products can be positioned as “budget-friendly bundles” for cost-conscious buyers.

4

INVENTORY WEIGHT IMBALANCE

Certain product categories dominate total stock weight, which may affect logistics and warehouse efficiency. Segmenting items by weight (Low, Medium, Bulk) uncovered skewed inventory patterns that companies can optimize for better storage and delivery planning.

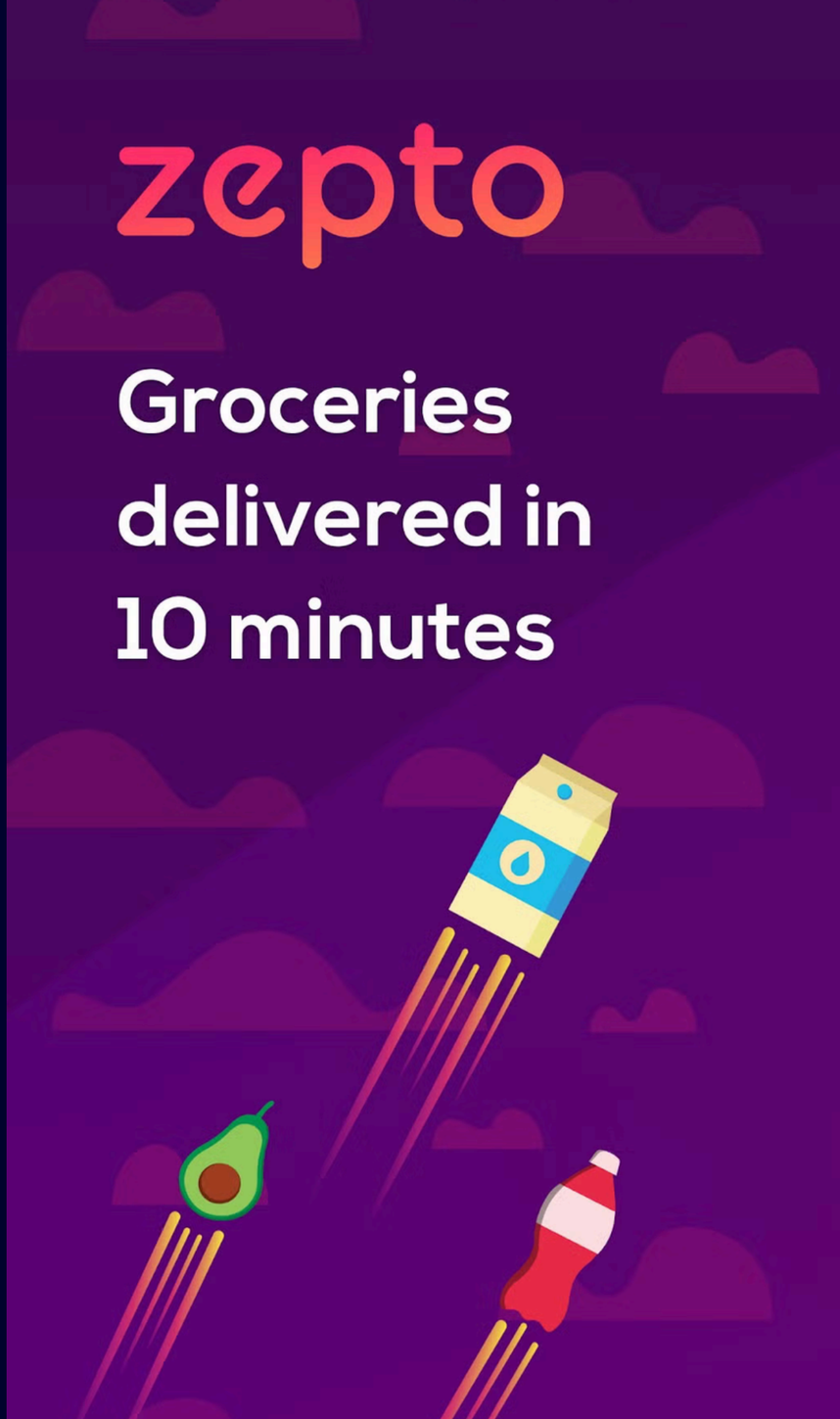
THE TAKEAWAY

01 Messy inventory data can be transformed into clear insights using SQL

02 Pricing, stock, and weight all carry hidden business risks and opportunities

03 Small tweaks — like restocking or rebalancing discounts — can drive big gains

04 Data storytelling turns raw queries into decision - making power



LET’S CONNECT!”

https://www.linkedin.com/in/sneha-ghosh-98aaa9337

"From pricing pitfalls to restocking blind spots, this dataset revealed more than just numbers — it told a story of business priorities hiding in plain sight."

THANK YOU !