

zepto

zepto

# Analysis SQL Project

Analyzing Real-World Inventory Data from India's Quick-Commerce Leader

# Project Overview

1. Real-world data analyst portfolio project using Zepto's online grocery catalog
2. Built using actual inventory data scraped from the Zepto platform
3. Structured for practical SQL experience with business relevance
4. Designed for aspiring analysts and e-commerce interview preparation

# Project Goals

Simulate how data analysts work with inventory data in retail/e-commerce

Build, clean, and query a relational database using SQL

Extract pricing, inventory, and revenue insights

Translate raw data into business decisions

# Business Requirements

1. Analyze discount and pricing strategies

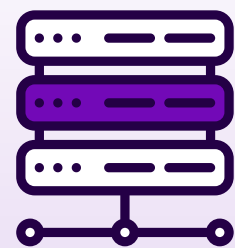
2. Identify out-of-stock high-value items

3. Estimate potential revenue by product category

4. Assess inventory distribution across SKUs

5. Classify products based on weight tiers for logistics use

# Tools & Technologies



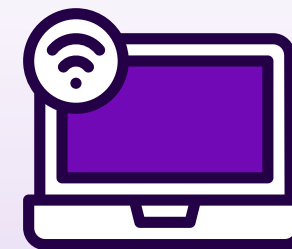
PostgreSQL

SQL database and  
queries



pgAdmin

Data import and query  
execution



Kaggle

Dataset storage &  
inspection



Excel / CSV

Data sourcing

# Dataset Summary

- **Source:** Zepto Inventory Dataset on Kaggle
- **Total Rows:** 3,732
- **Columns:** 9
- **Format:** .csv
- **Primary Use Case:** E-commerce catalog and inventory analysis

## Key Columns

name, category, mrp, discountPercent, discountedSellingPrice, availableQuantity, outOfStock, weightInGms, quantity

# SQL Queries

Find the top 10 best-value products based on the discount percentage.

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

	name character varying (150)	mrp numeric (8,2)	discountpercent numeric (5,2)
1	Dukes Waffy Orange Wafers	45.00	51.00
2	Dukes Waffy Chocolate Wafers	45.00	51.00
3	Dukes Waffy Strawberry Wafers	45.00	51.00
4	RRO Mozzarella Block Cheese	295.00	50.00
5	Ceres Foods Laal Maas Instant Liquid Masala	220.00	50.00
6	Chef's Basket Durum Wheat Elbow Pasta	160.00	50.00
Total rows: 10		Query complete 00:00:00.082	

- Insights: These products offer the highest discounts, great for promotions and banners.



# SQL Queries

## High-MRP Products Out of Stock

```
SELECT DISTINCT name, mrp
FROM zepto
WHERE outOfStock = TRUE and mrp > 300
ORDER BY mrp DESC;
```

	name character varying (150)	mrp numeric (8,2)
1	Patanjali Cow's Ghee	565.00
2	MamyPoko Pants Standard Diapers, Extra Large (12 - 17 kg)	399.00
3	Aashirvaad Atta With Mutigrains	315.00
4	Everest Kashmiri Lal Chilli Powder	310.00
Total rows: 4    Query complete 00:00:00.087		

- Insights: These are high-ticket items currently unavailable – may result in lost revenue if not restocked.



# SQL Queries

## Revenue by Category

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

	category character varying (120) 🔒	total_revenue numeric 🔒
1	Fruits & Vegetables	10846.00
2	Meats, Fish & Eggs	20693.00
3	Biscuits	25007.60
4	Beverages	55051.00
5	Dairy, Bread & Batter	55051.00
6	Health & Hygiene	64180.00
Total rows: 14		Query complete 00:00:00.117

- Insights: Reveals top revenue-generating categories – useful for product prioritization.

# SQL Queries

## Expensive Products with Low Discount

```
SELECT DISTINCT name, mrp, discountPercent
FROM zepto
WHERE mrp > 500 AND discountPercent < 10
ORDER BY mrp DESC, discountPercent DESC;
```

	name character varying (150)	mrp numeric (8,2)	discountpercent numeric (5,2)
1	Dhara Kachi Ghani Mustard Oil Jar	1250.00	8.00
2	Saffola Gold (Jar)	1240.00	0.00
3	Dhara Filtered Groundnut Oil (Jar)	1050.00	1.00
4	Fortune Rice Bran Health Oil (Jar)	1050.00	1.00
5	Dhara Filtered Groundnut Oil (Jar)	1050.00	0.00
6	Fortune Soyabean Oil	1005.00	0.00
Total rows: 39		Query complete 00:00:00.064	

- Insights: These are premium-priced items that could benefit from promotional pricing.

# SQL Queries

## Categories with Highest Average Discounts

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

	category character varying (120)	avg_discount numeric
1	Fruits & Vegetables	15.46
2	Meats, Fish & Eggs	11.03
3	Ice Cream & Desserts	8.32
4	Chocolates & Candies	8.32
5	Packaged Food	8.32
Total rows: 5      Query complete 00:00:00.066		

- Insights: Identifies high-discount categories – ideal for discount marketing or sales strategy.

# SQL Queries

## Price Per Gram

```
SELECT DISTINCT name, weightInGms, discountedSellingPrice,  
ROUND(discountedSellingPrice/weightInGms,2) AS price_per_gram  
FROM zepto  
WHERE weightInGms >= 100  
ORDER BY price_per_gram;
```

	name character varying (150)	weightingms integer	discountedsellingprice numeric (8,2)	price_per_gram numeric
1	Tata Salt	1000	24.00	0.02
2	Onion	1000	21.00	0.02
3	Onion	3000	57.00	0.02
4	Shubh kart - Nirmal sugandhi mogra wet dhoop zipper 20 sticks	1160	28.00	0.02
5	Vicks Cough Drops Menthol	1160	20.00	0.02
6	Aashirvaad Iodised Salt	1000	19.00	0.02
Total rows: 1395		Query complete 00:00:00.179		

- Insights: Helps surface best value-for-money items in terms of quantity per ₹.

# SQL Queries

## Weight-Based Product Segmentation

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

	name <span>🔒</span> character varying (150)	weightInGms <span>🔒</span> integer	weight_category <span>🔒</span> text
1	Popular Essentials Poppy Seeds (Gasagase)	100	Low
2	Origami Kitchen Towels 2 In 1 Quilted 2ply	116	Low
3	Johnson's Baby Top To Toe Bath Body Wash	210	Low
4	Harpic Hygienic Toilet Cleaner Rim Block Citrus	26	Low
5	Lizol Disinfectant Surface & Floor Cleaner Liquid Jasmine	500	Low
6	Sapota	348	Low
Total rows: 1783		Query complete 00:00:00.170	

- Insights: Enables packaging strategy and inventory planning based on product size.

# SQL Queries

## Total Inventory Weight by Category

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

	category character varying (120) 🔒	total_weight bigint 🔒
1	Meats, Fish & Eggs	48016
2	Biscuits	84431
3	Fruits & Vegetables	91794
4	Health & Hygiene	142904
5	Dairy, Bread & Batter	143735
6	Beverages	143735
Total rows: 14		Query complete 00:00:00.207

- Insights: Understand physical storage demands by category – relevant for logistics and warehouse teams.

# Business Insights from SQL

## Top Discounts

Identified highest-discounted products for marketing pushes

## Stock Gaps

High-MRP products out of stock = missed revenue opportunities

## Revenue Trends

Category-wise revenue shows where profit concentration lies

## Pricing Logic

Calculated price per gram to find value-for-money products

## Weight Tiers

Products segmented into Low, Medium, Bulk for logistics use

## Discount Focus

Top 5 categories with the highest average discounts



# Challenges and Solutions

## Challenges

- Encoding errors during CSV import
- Zero MRP and price entries
- Inconsistent units in quantity field
- Repetitive product names

## Solutions

- Saved using UTF-8 format in Excel
- Filtered and deleted invalid rows
- Focused analysis on weight-based insights
- Used name + weight + price for uniqueness

# Key Outcomes

1. Built a clean SQL database from semi-structured data
2. Delivered 8+ business-driven queries with actionable results
3. Simulated real analyst workflows: import → clean → analyze → present
4. Gained deep insight into inventory strategy, pricing, and availability

# Conclusion

Demonstrated the power of SQL in analyzing e-commerce inventory

Addressed real-world data issues and extracted business insights

Perfect showcase project for aspiring Data Analysts

Flexible for expansion into dashboards, NLP, and Python analysis

# Thank You

"From Raw Data to Real Impact – Turning Inventory into Insights"



[bhaskarpal.official@gmail.com](mailto:bhaskarpal.official@gmail.com)