

zepto

Zepto Inventory Management Using SQL

By Satyabrata Khan



□ Project Overview



The goal is to simulate how real-world data analysts in e-commerce or retail operate behind the scenes using SQL to:

- ☒ Set up and structure a messy e-commerce inventory dataset just like you'd find in the wild
- ☒ Perform Exploratory Data Analysis (EDA) to uncover trends, category breakdowns, stock availability, and pricing issues
- ☒ Clean the data by handling nulls, removing invalid entries, and converting prices from paise to rupees for consistency
- ☒ Write business-focused SQL queries that deliver actionable insights around pricing strategy, inventory flow, stock-out risks, estimated revenues, and more

Dataset Overview



Columns:

- `serial_number`: Unique identifier for each product entry (Primary Key)
- `name`: Product name as it appears on the app
- `category`: Product category like Fruits, Snacks, Beverages, etc.
- `mrp`: Maximum Retail Price (originally in paise, converted to ₹)
- `discountPercent`: Discount applied on MRP
- `discountedSellingPrice`: Final price after discount (also converted to ₹)
- `availableQuantity`: Units available in inventory
- `weightInGms`: Product weight in grams
- `outOfStock`: Boolean flag indicating stock availability
- `quantity`: Number of units per package (mixed with grams for loose produce)

Created Database and Table

```
1 ● create database zepto_in;
2 ● use zepto_in;
3 ● drop table if exists zepto;
4
5
6 ● ⊖ CREATE TABLE zepto_v2 (
7     category VARCHAR(120),
8     name VARCHAR(150) NOT NULL,
9     mrp DECIMAL(8,2),
10    discountPercent DECIMAL(5,2),
11    availableQuantity INT,
12    discountedSellingPrice DECIMAL(8,2),
13    weightInGms INT,
14    outOfStock TINYINT(1) DEFAULT 0,
15    quantity INT
16 );
```

□ Data Exploration

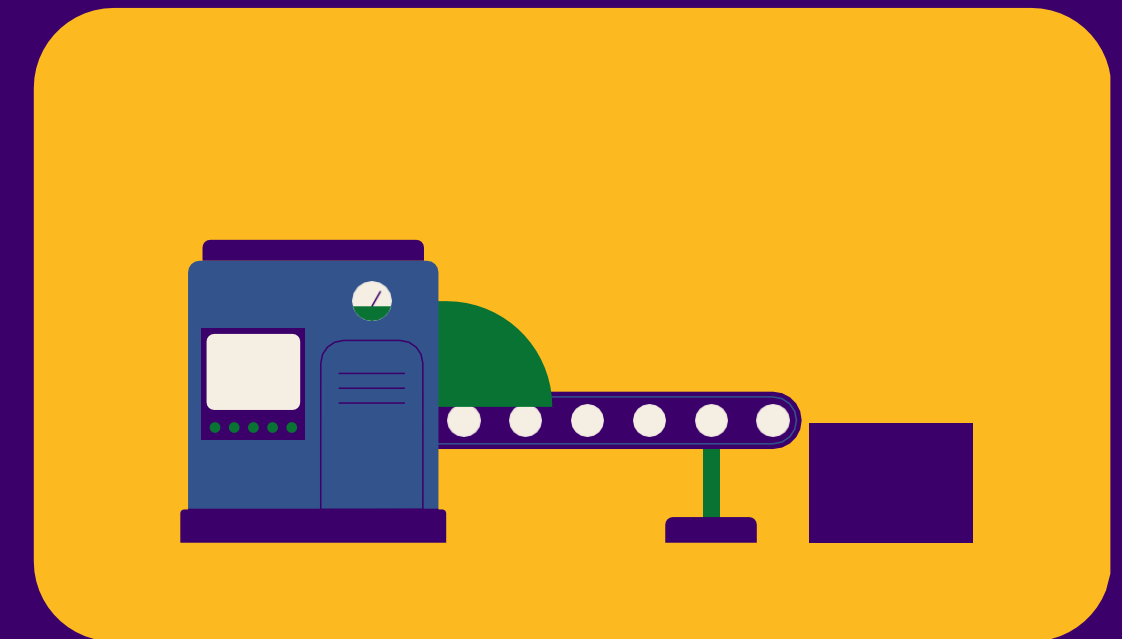
```
21 -- Count of rows
22 • select count(*) from zepto_v2;
23
24 -- Sample data
25 • select * from zepto_v2
26 limit 10;
27
```

Result Grid	
	count(*)
▶	3727

Result Grid										
	Category	name	mrp	discountPercent	availableQuantity	discountedSellingPrice	weightInGms	outOfStock	quantity	serial_number
▶	Fruits & Vegetables	Onion	25	16	3	21	1000	FALSE	1	1
	Fruits & Vegetables	Tomato Hybrid	42	16	3	35	1000	FALSE	1	2
	Fruits & Vegetables	Tender Coconut	51	15	3	43	58	FALSE	1	3
	Fruits & Vegetables	Coriander Leaves	20	15	3	17	100	FALSE	100	4
	Fruits & Vegetables	Ladies Finger	14	14	3	12	250	FALSE	250	5
	Fruits & Vegetables	Murgh	25	16	2	21	300	FALSE	300	6
	Fruits & Vegetables	Potato	35	17	3	29	1000	FALSE	1	7
	Fruits & Vegetables	Lemon	75	16	3	63	200	FALSE	200	8
	Fruits & Vegetables	Watermelon	58	15	3	49	58	FALSE	1	9

□ Data Exploration

```
35  -- Null values
36  ●  SELECT
37      *
38  FROM
39      zepto_v2
40  WHERE
41      name IS NULL OR Category IS NULL
42      OR mrp IS NULL
43      OR discountPercent IS NULL
44      OR availableQuantity IS NULL
45      OR discountedSellingPrice IS NULL
46      OR weightInGms IS NULL
47      OR outOfStock IS NULL
48      OR quantity IS NULL;
```



Result Grid										
	Category	name	mrp	discountPercent	availableQuantity	discountedSellingPrice	weightInGms	outOfStock	quantity	serial_number
*	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL

Insight: As there are no null values so we don't have to handle those values .

□ Data Exploration

```
49
50      --      different product catagoiries
51 •  SELECT DISTINCT
52      Category
53  FROM
54      zepto_v2
55  ORDER BY category;

56
57  -- products in stock and outofstock
58 •  SELECT
59      outOfstock, COUNT(serial_number)
60  FROM
61      zepto_v2
62  GROUP BY outOfstock;
```

Result Grid		Filter Rows:
	Category	
▶	Beverages	
	Biscuits	
	Chocolates & Candies	
	Cooking Essentials	
	Dairy, Bread & Batter	
	Fruits & Vegetables	
	Health & Hygiene	
	Home & Cleaning	
	Ice Cream & Desserts	
	Meats, Fish & Eggs	
	Munchies	
	Paan Corner	
	Packaged Food	
	Personal Care	

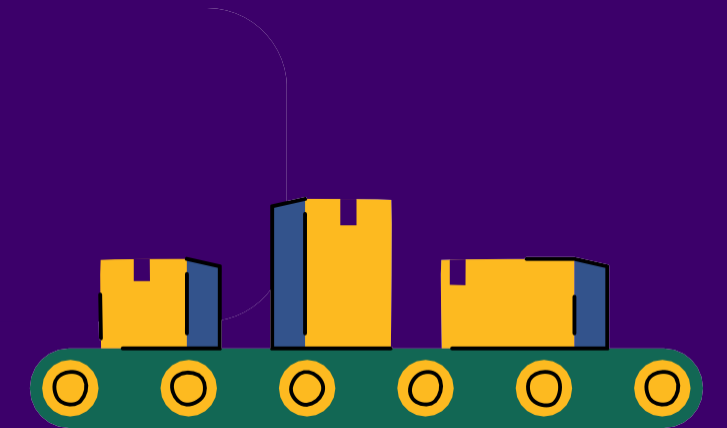
Result Grid			Filter Rows:
	outOfstock	COUNT(serial_number)	
▶	FALSE	3274	
	TRUE	453	

□ Data Exploration

```
-- product name present multiple times
SELECT
    name, COUNT(serial_number) AS Number_of_SKUs
FROM
    zepto_v2
GROUP BY name
HAVING COUNT(serial_number) > 1
ORDER BY COUNT(serial_number) DESC;
```

Result Grid			Filter Rows:	Export:	Write
	name	Number_of_SKUs			
▶	Arden Eggs White	10			
	Saffola Veggie Twist Masala Oats	10			
	Quaker Oats	10			
	Sunfeast Yippee! Pasta Treat - Sour Cream Onion	10			
	Sunfeast YIPee! Magic Masala Noodles	10			
	Mother's Recipe Tamarind Paste	10			
	Amul Delicious Fat Spread - Cholesterol Free	10			
	Kellogg's Real Almond & Honey Corn Flakes	9			
	Amul Fresh Cream	8			
	iD Idli & Dosa Batter	7			
	Everest Garam Masala	6			
	Everest Chicken Masala	6			
	Godrej Yummiez Chicken Nuggets	6			
	Everest Kitchen King Masala	6			

Data Cleaning



```
75  -- products of price 0
76
77  •  SELECT
78      *
79  FROM
80      zepto_v2
81  WHERE
82      mrp = 0 OR discountedSellingPrice = 0;
83
84  •  SET SQL_SAFE_UPDATES = 0;
85  •  DELETE FROM zepto_v2
86  WHERE
87      mrp = 0;
```

Result Grid										
Filter Rows: <input type="text"/>										
Edit:										
Export/Import:										
Wrap Cell Content:										
	Category	name	mrp	discountPercent	availableQuantity	discountedSellingPrice	weightInGms	outOfStock	quantity	serial_number
*	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL

Insight: Since there was few cells having mrp and discounted selling price as 0 which is not possible so I removed those values.



Data Transformation



```
-- convert paise to rupees

UPDATE zepto_v2
SET
    mrp = mrp / 100.0,
    discountedSellingPrice = discountedSellingPrice / 100.0;
```

Result Grid				Filter Rows:	Export:	Wrap
	name	mrp	discountedSellingPrice			
▶	Onion	25	21			
	Tomato Hybrid	42	35			
	Tender Coconut	51	43			
	Coriander Leaves	20	17			
	Ladies Finger	14	12			

Insight: As mrp and discounted selling price was in paise so converted then into rupees for my analysis.

Business Questions

-- Top 10 best-value products based on discount percentage

```
105 • SELECT DISTINCT
106     name, mrp, discountPercent
107 FROM
108     zepto_v2
109 ORDER BY discountPercent DESC
110 LIMIT 10;
```

	name	mrp	discountPercent
▶	Dukes Waffy Chocolate Wafers	45	51
	Dukes Waffy Orange Wafers	45	51
	Dukes Waffy Strawberry Wafers	45	51
	Ceres Foods Fish Mustard Instant Liquid Masala	220	50
	Ceres Foods Laal Maas Instant Liquid Masala	220	50
	Ceres Foods Nalli Nihari Instant Liquid Masala	220	50
	Chef's Basket Durum Wheat Elbow Pasta	160	50
	Chef's Basket Durum Wheat Fusilli Pasta	160	50
	Chef's Basket Durum Wheat Penne Pasta	160	50
	Dukes Waffy Chocolate Wafer Rolls	150	50

Insight: This shows the top– 10 products having highest discount reason can be low sales performance.

-High-MRP products that are currently out of stock

```
113 • SELECT DISTINCT
114     name, mrp, outOfstock
115 FROM
116     zepto_v2
117 WHERE
118     outOfstock = "TRUE" AND mrp > 300
119 ORDER BY mrp DESC;
```







Result Grid Filter Rows: <input type="text"/> Export: Wrap Cell Content:			
	name	mrp	outOfstock
▶	Patanjali Cow's Ghee	565	TRUE
	MamyPoko Pants Standard Diapers, Extra Large...	399	TRUE
	Aashirvaad Atta With Mutigrains	315	TRUE
	Everest Kashmiri Lal Chilli Powder	310	TRUE

Insight: Products with the highest mrp and out of stock shows the demand and necessity of the product.

-- Potential revenue for each product category

```
SELECT
    Category,
    SUM(discountedSellingPrice * quantity) AS estimated_revenue_from_each_Category
FROM
    zepto_v2
GROUP BY Category
ORDER BY estimated_revenue_from_each_Category DESC;
```

Result Grid   Filter Rows: <input type="text"/> Export:  Wrap Cell Content: 		
	Category	estimated_revenue_from_each_Category
▶	Packaged Food	17349917
	Ice Cream & Desserts	17349917
	Chocolates & Candies	17349917
	Cooking Essentials	13613234
	Munchies	13613234
	Personal Care	12614825
	Paan Corner	12614825
	Meats, Fish & Eggs	3693976
	Home & Cleaning	3678343
	Dairy, Bread & Batter	3330483
	Beverages	3330483

Insight: Packaged Food, IceCream & Desserts and Chocolates & Candies has shown the highest revenue among all the categories representing the demand.

-- Expensive products (MRP > ₹500) and discount is less than 10%

```
131 • select avg(discountPercent)
132   from zepto_v2;
133 • SELECT DISTINCT
134     name, mrp, discountPercent
135   FROM
136     zepto_v2
137   WHERE
138     mrp > 500 AND discountPercent < 10
139   ORDER BY mrp DESC;
```

name	mrp	discountPercent
Dhara Kachi Ghani Mustard Oil Jar	1250	8
Saffola Gold (Jar)	1240	0
Dhara Filtered Groundnut Oil (Jar)	1050	0
Dhara Filtered Groundnut Oil (Jar)	1050	1
Fortune Rice Bran Health Oil (Jar)	1050	1
Fortune Soyabean Oil	1005	0
Fortune Sunlite Refined Sunflower (Jar)	925	0
Surf Excel Matic Powder Front Load	810	7
Surf Excel Matic Top Load	720	9
Pedigree Puppy Dry Dog Food Food Chicken & Milk	690	6
Pedigree Dog Food Adult Meat & Rice	660	7

Insight: High-MRP products with minimal discounting (<10%) may reflect premium pricing strategies or low promotional focus, offering limited value to deal-seeking customers.



-- Top 5 categories offering highest average discounts

```
41 • SELECT
42     category, AVG(discountPercent) avg_discount
43 FROM
44     zepto_v2
45 GROUP BY category
46 ORDER BY avg_discount DESC
47 LIMIT 5;
```

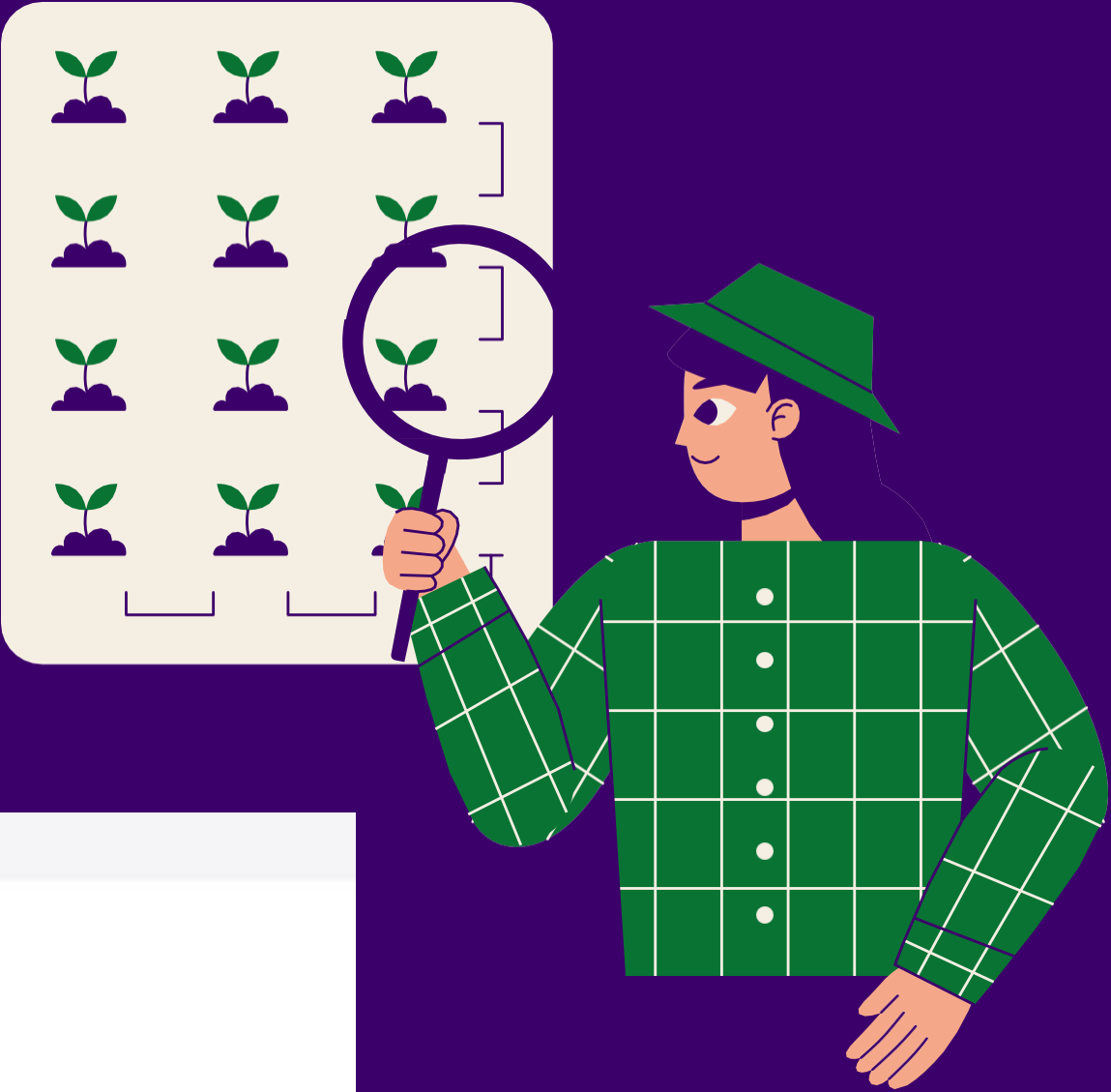
Result Grid			Filter Rows:	Export
	category	avg_discount		
▶	Fruits & Vegetables	15.4624		
	Meats, Fish & Eggs	11.0317		
	Packaged Food	8.3247		
	Ice Cream & Desserts	8.3247		
	Chocolates & Candies	8.3247		



Insight: Top – 5 Categories that has highest average discount means the product has less life cycle compared to the other products.

-- Calculated price per gram products above 100 g and sort by best value

```
150 • SELECT DISTINCT
151     name,
152     mrp,
153     weightInGms,
154     discountedSellingPrice,
155     (discountedSellingPrice / weightInGms) AS Price_per_gm
156 FROM
157     zepto_v2
158 WHERE
159     weightInGms >= 100
160 ORDER BY Price_per_gm desc;
```



Result Grid Filter Rows: Export: Wrap Cell Content:					
	name	mrp	weightInGms	discountedSellingPrice	Price_per_gm
▶	Indulekha Bhringa Hair Oil	432	100	367	3.6700
	L'Oreal Paris Excellence Creme Hair Color, 1 Black	620	172	620	3.6047
	L'Oreal Paris Excellence Creme Hair Color, 3 Dar...	620	172	620	3.6047
	L'Oreal Paris Excellence Creme Hair Color, 4 Nat...	620	172	620	3.6047
	L'Oreal Paris Excellence Creme Hair Color, 4.25 ...	620	172	620	3.6047
	L'Oreal Casting Creme Ebony Black Hair Color	550	159	550	3.4591
	L'Oreal Paris Casting Creme Gloss Hair Color (Da...	550	159	550	3.4591
	L'Oreal Paris Excellence Creme Hair Color, 3.16 ...	550	159	550	3.4591
	L'Oreal Paris Casting Creme Gloss Hair Color, M...	550	160	550	3.4375
	Bombay Shaving Company Sensitive Cartridges	375	116	375	3.2328
	Organic India Wheat Grass	395	100	316	3.1600
	Mamaearth Ubtan Face Scrub Turmeric & Walnut	349	100	314	3.1400

Insight: Products with weight over 100g offering lower price per gram represent the best value buys, making them ideal for cost-conscious inventory planning

-- Grouped products based on weight into Low, Medium, and Bulk categories

```
164 • SELECT DISTINCT
165     name,
166     weightInGms,
167     CASE
168         WHEN weightInGms < 1000 THEN 'Low'
169         WHEN weightInGms < 5000 THEN 'Medium'
170         ELSE 'Bulk'
171     END AS weight_category
172 FROM
173     zepto_v2
174 ORDER BY weightInGms DESC;
```

Result Grid	Filter Rows:	Export:	Wrap Cell Content:
	name	weightInGms	weight_category
▶	Fortune Sunlite Refined Sunflower (Jar)	5000	Bulk
	Madhur Pure And Hygienic Sugar	5000	Bulk
	Pillsbury Atta With Multigrains Wheat G...	5000	Bulk
	Pillsbury Atta/Godihittu - Chakki Fresh	5000	Bulk
	Popular Essentials Classic Sona Masuri ...	5000	Bulk
	Popular Essentials HMT Kolam Steam	5000	Bulk
	Popular Essentials Refined Sugar	5000	Bulk
	Popular Essentials Rozana Sona Masuri ...	5000	Bulk
	Saffola Gold (Jar)	5000	Bulk
	Super Saver Kolam Rice Sortex Clean	5000	Bulk
	Surf Excel Matic Powder Front Load	4000	Medium
	Surf Excel Matic Top Load	4000	Medium
	Surf Excel Quick Wash	4000	Medium
	Ariel Matic Front Load Detergent Washi...	3000	Medium
	Onion	3000	Medium
	Pedigree Dog Food Adult Chicken & Veg	3000	Medium
	Pedigree Dog Food Adult Meat & Rice	3000	Medium
	Pedigree Puppy Dry Dog Food Food Chi...	3000	Medium



Insight: Categorizing products by weight highlights that heavier items dominate the inventory, offering strategic bulk options that could drive higher volume-based revenue.

-- Measured total inventory weight per product category

```
77 • SELECT
78     category,
79     SUM(weightInGms * availableQuantity) AS total_weight
80 FROM
81     zepto_v2
82 GROUP BY category
83 ORDER BY total_weight;
84
```

	category	total_weight
►	Meats, Fish & Eggs	48016
	Biscuits	84431
	Fruits & Vegetables	91794
	Health & Hygiene	142904
	Dairy, Bread & Batter	143735
	Beverages	143735
	Personal Care	348187
	Paan Corner	348187
	Home & Cleaning	373161
	Packaged Food	490797
	Ice Cream & Desserts	490797
	Chocolates & Candies	490797
	Cooking Essentials	1404326
	Munchies	1404326



Insight: "Chocolates & Candies" lead the inventory in total weight, signaling a high-volume stocking strategy for impulse or seasonal purchases.



My Recommendations and Actionable Insights

- Restock high-MRP out-of-stock products to meet demand and avoid revenue loss.
- Introduce discounts on premium products ($\text{MRP} > ₹500$) to boost sales.
- Focus on top revenue categories like Packaged Food, Ice Cream & Desserts, and Chocolates & Candies.
- Optimize storage for heavy categories like Chocolates & Candies to manage inventory space.
- Promote bulk products ($>100\text{g}$) with low price per gram as high-value deals.
- Clear over-discounted items in top 5 discounted categories to reduce excess inventory.
- Segment inventory by weight (Low, Medium, Bulk) to enhance logistics and packaging.
- Highlight top discounted products in marketing campaigns and flash sales.
- Remove products with zero MRP and price to maintain data integrity.
- Implement regular data audits to avoid anomalies in future datasets.



Thank You

Satyabrata.K

6691 satyabrata.k@gmail.com



+123-456-789