zepto

SQL Data Analysis Project

PROJECT OVERVIEW

Zepto

📂 Built a structured SQL database to decode Zepto's fast-moving grocery data.

🧹 Cleaned messy data—removed duplicates, fixed pricing, and handled stock info.

💸 Spotted high-discount gems and costly products that are out of stock.

📈 Calculated revenue by category and total inventory weight for smart planning.

🧃 Analyzed price per gram and grouped products by weight to uncover value deals.

PROJECT GOALS

Zepto

🛒 Decode Zepto's Inventory

Dive into product-level data to uncover what's in stock, what's not, and why it matters.

🔍 Spot Hidden Value

Identify best-value products using discount percentages and price-per-gram analysis.

💰 Track the Money Trail

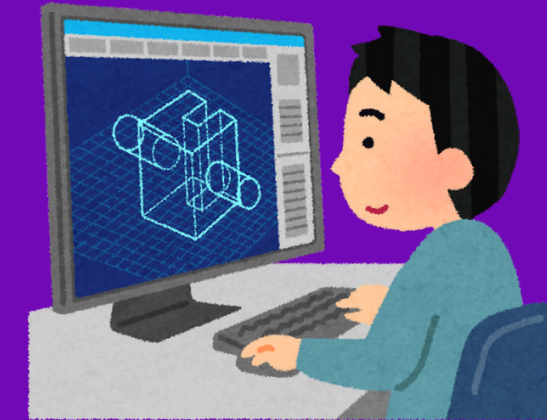Calculate estimated revenue by category to see where the profits really lie.

🚚 Simplify Storage & Delivery

Group products by weight (Low/Medium/Bulk) to optimize logistics and warehouse planning.

TOOLS AND TECHNOLOGY

MySQL

Excel/CSV

Kaggle

DATASET DESCRIPTION

Zepto

- *Source : Zepto Inventory Dataset From Kaggle*

- *Total rows : 3732*

- *Columns : 9*

- *Format : CSV*

Key Columns

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

SQL Queries 1

```sql
-- Q1.Find the top 10 best value products based on the discount percentage.
select distinct name,mrp,discountPercent
from zepto_sql
order by discountPercent desc
limit 10 ;
```

name	mrp	discountPercent
Dukes Waffy Chocolate Wafers	45.00	51.00
Dukes Waffy Orange Wafers	45.00	51.00
Dukes Waffy Strawberry Wafers	45.00	51.00
Ceres Foods Fish Mustard Instant Liquid Masala	220.00	50.00
Ceres Foods Laal Maas Instant Liquid Masala	220.00	50.00
Ceres Foods Nalli Nihari Instant Liquid Masala	220.00	50.00
Chef's Basket Durum Wheat Elbow Pasta	160.00	50.00
Chef's Basket Durum Wheat Fusilli Pasta	160.00	50.00
Chef's Basket Durum Wheat Penne Pasta	160.00	50.00
Dukes Waffy Chocolate Wafer Rolls	150.00	50.00

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

zepto

SQL Queries 2

```
-- Q2.What are the products with high MRP but out of stock.
select distinct name, mrp,outOfStock
from zepto_sql
where outOfStock=1 and mrp >300
order by mrp desc ;
```

name	mrp	outOfStock
Patanjali Cow's Ghee	565.00	1
MamyPoko Pants Standard Diapers, Extra Large…	399.00	1
Aashirvaad Atta With Mutigrains	315.00	1
Everest Kashmiri Lal Chilli Powder	310.00	1

• Insights:These are high-ticket items currently unavailable — may result in lost revenue if not restored

SQL Queries 3

```
-- Q3.Calculate estimated revenue for each category.
select category , sum(discountedSellingPrice*availableQuantity) as total_revenue
from zepto_sql
group by category
order by total_revenue;
```

category	total_revenue
Fruits & Vegetables	10846.00
Meats, Fish & Eggs	20693.00
Biscuits	25007.60
Dairy, Bread & Batter	55051.00
Beverages	55051.00
Health & Hygiene	64180.00
Home & Cleaning	122661.00
Packaged Food	224385.00
Ice Cream & Desserts	224385.00

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

SQL Queries 4

```sql
-- Q4.Find all products where MRP is greater than 500 rupee and discount is less than 10%.
select distinct name,mrp,discountPercent
from zepto_sql
where mrp>500 and discountPercent<10
order by mrp desc,discountPercent desc;
```

name	mrp	discountPercent
Dhara Kachi Ghani Mustard Oil Jar	1250.00	8.00
Saffola Gold (Jar)	1240.00	0.00
Dhara Filtered Groundnut Oil (Jar)	1050.00	1.00
Fortune Rice Bran Health Oil (Jar)	1050.00	1.00
Dhara Filtered Groundnut Oil (Jar)	1050.00	0.00
Fortune Soyabean Oil	1005.00	0.00
Fortune Sunlite Refined Sunflower (Jar)	925.00	0.00
Surf Excel Matic Powder Front Load	810.00	7.00
Surf Excel Matic Top Load	720.00	9.00

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

SQL Queries 5

```sql
-- Q5.Identify the top 5 categories offering the highest average discount percentage.
select category,Round(avg(discountPercent),2) as avg_discount
from zepto_sql
group by category
order by avg_discount desc
limit 5;
```

category	avg_discount
▶ Fruits & Vegetables	15.46
Meats, Fish & Eggs	11.03
Chocolates & Candies	8.32
Ice Cream & Desserts	8.32
Packaged Food	8.32

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

SQL Queries 6

```
-- Q6.Find the price per gram for products above 100g and sort by best value.
select distinct name,weightInGms,discountedSellingPrice,
discountedSellingPrice/weightInGms As price_per_gram
from zepto_sql
where weightInGms >=100
order by price_per_gram desc;
```

name	weightInGms	discountedSellingPrice	price_per_gram
▶ Indulekha Bhringa Hair Oil	100	367.00	3.670000
L'Oreal Paris Excellence Creme Hair Color, 1 Black	172	620.00	3.604651
L'Oreal Paris Excellence Creme Hair Color, 3 Dar...	172	620.00	3.604651
L'Oreal Paris Excellence Creme Hair Color, 4 Nat...	172	620.00	3.604651
L'Oreal Paris Excellence Creme Hair Color, 4.25 ...	172	620.00	3.604651
L'Oreal Casting Creme Ebony Black Hair Color	159	550.00	3.459119
L'Oreal Paris Casting Creme Gloss Hair Color (Da...	159	550.00	3.459119
L'Oreal Paris Excellence Creme Hair Color, 3.16 ...	159	550.00	3.459119
L'Oreal Paris Casting Creme Gloss Hair Color, M...	160	550.00	3.437500
Bombay Shaving Company Sensitive Cartridges	116	375.00	3.232759
Organic India Wheat Grass	100	316.00	3.160000
Mamaearth Ubtan Face Scrub Turmeric & Walnut	100	314.00	3.140000
Delight Nuts Dried Blueberries	150	432.00	2.880000

₹.

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

SQL Queries 7

```
-- Q7.Group the products into categories like low,medium,bulk.
select distinct name,weightInGms,
Case
     when weightInGms<1000 then 'Low'
     when weightInGms<5000 then 'Medium'
     else 'Bulk'
     end as weight_category
from zepto_sql;
```

name	weightInGms	weight_category
Onion	1000	Medium
Tomato Hybrid	1000	Medium
Tender Coconut	58	Low
Coriander Leaves	100	Low
Ladies Finger	250	Low
Potato	1000	Medium
Lemon	200	Low
Watermelon	58	Low
Capsicum Green	250	Low
Chilli Green	100	Low

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

SQL Queries 8

```sql
-- Q8.What is the total inventory weight per category
select category,sum(weightInGms * availableQuantity) as total_weight
from zepto_sql
group by category
order by total_weight;
```

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

• Insights:Understand physical storage demands by category — relevant for logistics and warehouse teams.

Zepto

CHALLENGES AND SOLUTIONS

Hurdles Faced :

Encoding errors during CSV import

Format changes in Mrp column

Error in Outofstock column (Datatype error)

Zero valued column.

Solutions :

Saved using UTF-8 format in Excel

Converted paisa to Rupee.

Replace the Boolean value TRUE and FALSE with 1 and 0

Drop the zero valued column