

Name : Pranit Pramod Dere.

SQL Project

Zepto SQL Data Analysis Project

This project focuses on analyzing Zepto's product dataset using SQL to extract meaningful business insights. The analysis helps understand product pricing, discounts, inventory distribution, category performance, and overall business trends. The project demonstrates practical SQL skills such as data exploration, data cleaning, aggregation, and business-driven query analysis.

Objectives of the Project :

- 1] To analyze product pricing and discount patterns
- 2] To identify best-value products for customers
- 3] To perform category-wise business analysis
- 4] To understand inventory distribution and weight-based grouping
- 5] To generate meaningful insights using SQL queries for decision-making

Entity Relationship Diagram (ERD) :-

The project uses a single-table database design.

zepto	
sku_id	BIGINT
category	VARCHAR(120)
name	VARCHAR(150)
mrp	DECIMAL(8,2)
discountPercent	DECIMAL(5,2)
availableQuantity	INT
discountedSellingprice	DECIMAL(8,...)
weightInGram	INT
outOfStock	TINYINT(1)
quantity	INT
Indexes	

Database :

```
create database zepto;  
use zepto;
```

Creat Table:

```
create table zepto(  
    sku_id SERIAL PRIMARY KEY,  
    category varchar(120),  
    name VARCHAR(150) NOT NULL,  
    mrp NUMERIC(8, 2),  
    discountPercent NUMERIC(5, 2),  
    availableQuantity INTEGER,  
    discountedSellingprice NUMERIC(8, 2),  
    weightInGram INTEGER,  
    outOfStock BOOLEAN,  
    quantity INTEGER  
)
```

Data Import :-

After creating the table, the Zepto dataset was imported into the database using MySQL import functionality. The dataset was loaded from an Excel/CSV file and verified to ensure all records were inserted correctly for further analysis.

Data Cleaning and Preparation:-

The following data cleaning steps were performed before analysis:-

1]Checked for null values in important columns:-

```
select * from zepto  
where name is null  
or  
category is null  
or
```

mrp is null
or
discountPercent is null
or
availableQuantity is null
or
discountedSellingprice is null
or
weightInGram is null
or
outOfStock is null
or
quantity is null;

2]Removed products with zero MRP :-

select * from zepto
where mrp = 0 or discountedSellingprice = 0;

delete from zepto
where mrp = 0;

3] Converted price values from paise to rupees :-

update zepto

set mrp = mrp / 100.0,

discountedSellingprice = discountedSellingprice / 100.0;

select mrp, discountedSellingprice from zepto;

-- OUTPUT :

mrp	discountedSellingprice
25.00	21.00
42.00	35.00
51.00	43.00
20.00	17.00
14.00	12.00
35.00	29.00
75.00	63.00
58.00	49.00
23.00	19.00
19.00	16.00
29.00	24.00
11.00	9.00
26.00	22.00
14.00	12.00
19.00	16.00
42.00	35.00
15.00	13.00
30.00	25.00
36.00	30.00

These steps helped improve data quality and accuracy of analysis.

Business Insight Questions and SQL Analysis:-

Q1. Find the top 10 best-value products based on the discount percentage.

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

-- OUTPUT :

name	discountPercent
Dukes Waffy Orange Wafers	51.00
Dukes Waffy Chocolate Wafers	51.00
Dukes Waffy Strawberry Wafers	51.00
Ceres Foods Fish Mustard Instant Liquid Masala	50.00
Chef's Basket Durum Wheat Elbow Pasta	50.00
Ceres Foods Laal Maas Instant Liquid Masala	50.00
Ceres Foods Fish Mustard Instant Liquid Masala	50.00
Chef's Basket Durum Wheat Elbow Pasta	50.00
Chef's Basket Durum Wheat Penne Pasta	50.00
Chef's Basket Durum Wheat Fusilli Pasta	50.00

Q2. Find the average MRP for each product category.

```
SELECT category, AVG(mrp) AS avg_mrp  
FROM zepto  
GROUP BY category; -- OUTPUT :
```

category	avg_mrp
Fruits & Vegetables	47.010753
Cooking Essentials	157.009766
Munchies	157.009766
Dairy, Bread & Batter	118.705426
Beverages	118.705426
Packaged Food	155.726804
Ice Cream & Desserts	155.726804
Chocolates & Candies	155.726804
Meats, Fish & Eggs	187.285714
Biscuits	59.489796
Personal Care	207.119186
Paan Corner	207.119186
Home & Cleaning	152.943005
Health & Hygiene	159.175258

Q3. Find the total estimated revenue for each product category.

```
SELECT category, SUM(discountedSellingprice * quantity) AS revenue  
FROM zepto  
GROUP BY category;
```

-- OUTPUT :

category	revenue
Fruits & Vegetables	440997.00
Cooking Essentials	13613234.00
Munchies	13613234.00
Dairy, Bread & Batter	3330483.00
Beverages	3330483.00
Packaged Food	17349917.00
Ice Cream & Desserts	17349917.00
Chocolates & Candies	17349917.00
Meats, Fish & Eggs	3693976.00
Biscuits	1792076.00
Personal Care	12614825.00
Paan Corner	12614825.00
Home & Cleaning	3678343.00
Health & Hygiene	1496117.00

Q4. Find all products where MRP is greater than ₹500 and discount is less than 10%.

```
SELECT name, mrp, discountPercent  
FROM zepto  
WHERE mrp > 500 AND discountPercent < 10;
```

-- OUTPUT :

name	mrp	discountPercent
Fortune Sunlite Refined Sunflower (Jar)	925.00	0.00
Amul Pure Ghee (Pouch)	505.00	1.00
Amul Ghee (Tin)	520.00	2.00
Fortune Soyabean Oil	1005.00	0.00
Dhara Kachi Ghani Mustard Oil Jar	1250.00	8.00
Gowardhan Ghee Jar	598.00	3.00

Q5. Identify the top 5 categories offering the highest average discount percentage.

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

-- OUTPUT :

category	avg_discount
Fruits & Vegetables	15.462366
Meats, Fish & Eggs	11.031746
Chocolates & Candies	8.324742
Ice Cream & Desserts	8.324742
Packaged Food	8.324742

Q6. Find the price per gram for products above 100g.

```
SELECT name, discountedSellingprice / weightInGram AS price_per_gram  
FROM zepto  
WHERE weightInGram > 100;
```

-- OUTPUT :

name	price_per_gram
Vicks Cough Drops Menthol	0.017241
Onion	0.019000
Aashirvaad Iodised Salt	0.019000
Aashirvaad Iodised Salt	0.019000
Onion	0.021000
Tata Salt	0.024000
Tata Salt	0.024000
Shubh kart - Nirmal sugandhi mogra wet dhoop ...	0.024138
Beetroot	0.026000
Potato	0.028000

Q7. Group products into Low, Medium, and Bulk based on weight.

```
SELECT name,  
CASE  
    WHEN weightInGrams < 500 THEN 'Low'  
    WHEN weightInGrams <= 2000 THEN 'Medium'  
    ELSE 'Bulk'  
END AS weight_category  
FROM zepto;
```

-- OUTPUT :

name	weight_category
Onion	Medium
Tomato Hybrid	Medium
Tender Coconut	Low
Coriander Leaves	Low
Ladies Finger	Low
Potato	Medium
Lemon	Low
Watermelon	Low
Capsicum Green	Low
Chilli Green	Low

Q8. Find the total inventory weight for each category.

```
SELECT category, SUM(weightInGrams * quantity) AS total_weight  
FROM zepto  
GROUP BY category; -- OUTPUT :
```

category	total_weight
Fruits & Vegetables	5544710
Cooking Essentials	43865863
Munchies	43865863
Dairy, Bread & Batter	9247315
Beverages	9247315
Packaged Food	42891488
Ice Cream & Desserts	42891488
Chocolates & Candies	42891488
Meats, Fish & Eggs	7430828
Biscuits	7317165

Q9. Find the top 10 products with the highest discount value (₹).

```
SELECT name, (mrp - discountedSellingprice) AS discount_value  
FROM zepto  
ORDER BY discount_value DESC  
LIMIT 10;
```

-- OUTPUT :

name	discount_value
Borges Extra Light Olive Oil Bottle	1201.00
Borges Extra Light Olive Oil Bottle	1201.00
Pampers Pants - Large	700.00
Pampers Pants - Large	700.00
Kelloggs Chocos - Moons & Stars	264.00
Kelloggs Chocos - Moons & Stars	264.00
Kelloggs Chocos - Moons & Stars	264.00
Pampers Active Baby Diapers New Born Extra S...	250.00
Pampers Active Baby Diapers New Born Extra S...	250.00
Popular Essentials Californian Almond	225.00

Q10. Find the total number of products in each category.

```
SELECT category, COUNT(*) AS total_products  
FROM zepto  
GROUP BY category;
```

-- OUTPUT :

outOfStock	total_products
0	3274
1	453

Q11. Find products with low inventory quantity (less than 10).

```
SELECT name, quantity
```

```
FROM zepto
```

```
WHERE quantity < 10;
```

-- OUTPUT :

name	qaunty
Onion	1
Tomato Hybrid	1
Tender Coconut	1
Potato	1
Watermelon	1
Banana Robusta	6
Cauliflower	1
Muskmelon	1
Cabbage	1

Q12. Find the average selling price for each category.

```
SELECT category, AVG(discountedSellingprice) AS avg_price
```

```
FROM zepto
```

```
GROUP BY category;
```

-- OUTPUT :

category	avg_price
Fruits & Vegetables	39.838710
Cooking Essentials	141.244141
Munchies	141.244141
Dairy, Bread & Batter	103.728682
Beverages	103.728682
Packaged Food	141.077320
Ice Cream & Desserts	141.077320
Chocolates & Candies	141.077320
Meats, Fish & Eggs	164.095238

Q13. List products that are sold without any discount.

```
SELECT name, mrp  
FROM zepto  
WHERE discountPercent = 0;
```

-- OUTPUT :

name	mrp
Maggi 2 Minute Masala Noodles	12.00
Tata Salt	24.00
Arden Eggs White	97.00
Saffola Gold Oil (Pouch)	215.00
Fortune Sunlite Refined Sunflower (Jar)	925.00
Saffola Active Oil	210.00
Fortune Soyabean Oil	1005.00
Popular Essentials Sunflower Seeds	25.00
Tata Sampann High Protien Kabuli Chana	94.00

Q14. Find the average product weight for each category.

```
SELECT category, AVG(weightInGrams) AS avg_weight  
FROM zepto  
GROUP BY category;
```

-- OUTPUT :

category	total_products
Cooking Essentials	512
Munchies	512
Packaged Food	388
Ice Cream & Desserts	388
Chocolates & Candies	388
Personal Care	344
Paan Corner	344
Home & Cleaning	193
Biscuits	147

Q15. Find the total inventory value for each category.

```
SELECT category, SUM(mrp * quantity) AS inventory_value  
FROM zepto  
GROUP BY category;
```

-- OUTPUT :

category	inventory_value
Fruits & Vegetables	523444.00
Cooking Essentials	15103295.00
Munchies	15103295.00
Dairy, Bread & Batter	3686036.00
Beverages	3686036.00
Packaged Food	19029649.00
Ice Cream & Desserts	19029649.00
Chocolates & Candies	19029649.00
Meats, Fish & Eggs	4178086.00

Conclusion:-

This SQL project successfully demonstrates how structured queries can be used to analyze real-world business data. The insights generated from this project help understand pricing strategies, discount effectiveness, inventory distribution, and category-wise performance. The project showcases practical SQL skills that are relevant for data analyst and business analyst roles.

Tools & Technologies Used :-

- MySQL
- MySQL Workbench
- Excel (Dataset source)