

QUESTIONS

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

Analysis: The best-value products are those with the highest percentage discount applied to their Maximum Retail Price (MRP). The query should sort the products in descending order of `discountPercent` and limit the results to the top 10.

Approach

1. **Select Required Columns:** Select the product name, MRP, and discount percentage (`name`, `mrp`, `discountPercent`).
2. **Filter for Unique Products:** Use `DISTINCT` to ensure each unique product is considered once, even if there are duplicate entries.
3. **Sort for Best Value:** Order the results using `ORDER BY discountPercent DESC` to put the highest discount at the top.
4. **Limit Results:** Apply `LIMIT 10` to retrieve only the top 10 products.

```
SQL

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

Key Insights

- **Maximum Discount:** The highest discount percentage shown is **51.00%**, which applies to multiple products, notably the **Dukes Waffly Orange/Chocolate/Strawberry Wafers**.
- **Common Discount Tier:** A **50.00%** discount is a common promotional tier, applying to a diverse set of products including **RRO Mozzarella Block Cheese (₹295.00)** and various **Ceres Foods Instant Liquid Masalas (₹220.00)**.
- **Value vs. Price:** Products with the highest discount percentage are not necessarily the most expensive. For example, **Dukes Wafers (₹45.00)** have a higher discount percentage (51.00) than **RRO Cheese (₹295.00)** (50.00).
- **Category Focus:** The high-discount products span across snacks (Wafers), dairy (Cheese), ready-to-cook meals (Liquid Masalas), and staples (Pasta), indicating broad promotional activity.

name character varying (150)	mrp numeric (8,2)	discountpercent numeric (5,2)
Dukes Waffy Orange Wafers	45.00	51.00
Dukes Waffy Chocolate Wafers	45.00	51.00
Dukes Waffy Strawberry Wafers	45.00	51.00
RRO Mozzarella Block Cheese	295.00	50.00
Ceres Foods Laal Maas Instant Liquid Mas...	220.00	50.00
Chef's Basket Durum Wheat Elbow Pasta	160.00	50.00
RRO Mozzarella Pizza Cheese	275.00	50.00
Ceres Foods Fish Mustard Instant Liquid M...	220.00	50.00
Dukes Waffy Strawberry Roll	150.00	50.00
Ceres Foods Nalli Nihari Instant Liquid Mas...	220.00	50.00

2) What are the Products with High MRP but Out of Stock

Analysis: This query requires a combination of filtering and sorting. We need to filter for items that are both out of stock and have a high MRP, then sort them by price to see the most expensive items first.

Approach

1. **Select Required Columns:** Select the product name and MRP (**name**, **mrp**).
2. **Filter for Stock and Price:** Use the **WHERE** clause to filter the results. The first condition is **outOfStock = TRUE** and the second is **mrp > 300**. Both conditions must be met for a product to be included.
3. **Sort by Price:** Order the results using **ORDER BY mrp DESC** to list the most expensive products first.
4. **Filter for Uniqueness:** Use **DISTINCT** to ensure each product appears only once in the final list.

SQL

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

Key Insights

- **High-Value Staples:** The out-of-stock items priced over ₹300 include daily household staples such as **Patanjali Cow's Ghee (₹565.00)** and **Aashirvaad Atta With Mutigrains (₹315.00)**.
- **Essential/Bulk Items Out of Stock:** Key household necessities like **MamyPoko Diapers (₹399.00)** are listed, suggesting a potential high-demand/fast-selling scenario for critical items.
- **Variety in Categories:** The out-of-stock list is not limited to one category; it includes dairy/staples (Ghee, Atta), baby products (Diapers), and spices (Everest Chilli Powder), indicating a broad inventory issue for high-value goods.
- **Most Expensive Outlier: Patanjali Cow's Ghee** is the most expensive item on the list at ₹565.00, highlighting it as a top priority for restocking due to its high MRP and customer value.

name character varying (150)	mrp numeric (8,2)
Patanjali Cow's Ghee	565.00
MamyPoko Pants Standard Diapers, Extra Large (12 - 17 ...	399.00
Aashirvaad Atta With Mutigrains	315.00
Everest Kashmiri Lal Chilli Powder	310.00

3). Calculate Estimated Revenue for each category

Analysis: To calculate total revenue, we must multiply the price at which the product is sold (**discountedSellingPrice**) by the number of units available for sale (**availableQuantity**) and then aggregate this sum for all products within each **category**.

Approach

1. **Calculate Product Revenue:** For each product, multiply its selling price by its available quantity (**discountedSellingPrice * availableQuantity**).
2. **Aggregate Revenue:** Use the aggregate function **SUM()** to total the calculated revenue for all products.
3. **Group by Category:** Use the **GROUP BY category** clause to calculate a separate total for each distinct category.
4. **Order Results:** Use **ORDER BY total_revenue** (default ascending) to list categories from the lowest to the highest revenue.

SQL



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

Key Insights

- **Low Revenue Categories:** **Fruits & Vegetables** has the lowest total revenue at ₹10,846.00, followed closely by **Meats, Fish & Eggs** (₹20,693.00). This suggests either low stock, low selling prices, or high perishability impacting inventory levels.
- **Top Revenue Categories:** **Munchies** and **Cooking Essentials** share the highest revenue at ₹337,369.00, indicating these are the most significant contributors to the total estimated sales.
- **Tied Revenue:** Several pairs of categories have identical revenue figures, such as **Chocolates & Candies** and **Ice Cream & Desserts** (₹224,385.00), and **Personal Care** and **Paan Corner** (₹270,849.00). This could imply linked promotional sales or similar inventory values.
- **Strategic Stocking:** Categories with high revenue (like Munchies, Cooking Essentials) should be prioritized for stocking to maximize potential sales, while low-revenue categories may need focused strategies to improve their contribution.

category character varying (120)	total_revenue numeric
Fruits & Vegetables	10846.00
Meats, Fish & Eggs	20693.00
Biscuits	25007.60
Beverages	55051.00
Dairy, Bread & Batter	55051.00
Health & Hygiene	64180.00
Home & Cleaning	122661.00
Ice Cream & Desserts	224385.00
Chocolates & Candies	224385.00
Packaged Food	224385.00
Personal Care	270849.00
Paan Corner	270849.00
Munchies	337369.00
Cooking Essentials	337369.00

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

Analysis: This query isolates high-value products (`mrp > 500`) that offer very little discount (`discountPercent < 10`). The sorting prioritizes the most expensive items and then the highest discount within that price tier.

Approach

1. **Select Required Columns:** Select the product name, MRP, and discount percentage (`name, mrp, discountPercent`).
2. **Filter for Price:** Use the `WHERE` clause to select only products with an MRP greater than ₹500 (`mrp > 500`).
3. **Filter for Discount:** Use the `AND` operator to combine the first condition with the discount requirement (`discountPercent < 10`).
4. **Order Results:** Use a multi-column sort: `ORDER BY mrp DESC` (most expensive first) and then `discountPercent DESC` (highest discount first among tied MRPs).
5. **Filter for Uniqueness:** Use `DISTINCT` to ensure each product appears only once.

SQL

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

Key Insights

- **Premium Commodities:** The list is dominated by essential high-value commodities, particularly various brands of **Cooking Oils** (e.g., Dhara, Saffola, Fortune), with the highest MRPs (up to ₹1250.00).
- **Low Discount Persistence:** A significant number of these high-value products are offered with **0.00%** discount (e.g., several Hair Colors, Follow-Up Formulas, and various oils), indicating they are sold near full price.
- **Highest Discount Rate:** The maximum discount percentage in this filtered list is **9.00%** (**Surf Excel Matic Top Load**) and **8.00%** (Dhara Mustard Oil, Lizol Disinfectant), which is just under the 10% cutoff.
- **Brand Loyalty/Essential Demand:** The low discount on essential household items (oils, detergent, pet food) suggests these are non-discretionary purchases where demand is relatively inelastic, allowing the retailer to maintain high margins.

name character varying (150)	mrp numeric (8,2)	discountpercent numeric (5,2)
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
Pedigree Puppy Dry Dog Food Food Chicken & Milk	690.00	6.00
Pedigree Dog Food Adult Meat & Rice	660.00	7.00
Lizol Double Concentrate Disinfectant Floor Cleaner Citrus	650.00	8.00
Nestle Nestle Nan Pro Follow-Up Formula	650.00	0.00
Dove Daily Shine Shampoo	645.00	5.00
L'Oreal Paris Excellence Creme Hair Color, 1 Black	620.00	0.00
L'Oreal Paris Excellence Creme Hair Color, 3 Dark Brown/Natural Darkest Br...	620.00	0.00
L'Oreal Paris Excellence Creme Hair Color, 4 Natural Brown/Natural Dark Br...	620.00	0.00
L'Oreal Paris Excellence Creme Hair Color, 4.25 Aishwarya's Brown	620.00	0.00
Nestle Nan Pro 4 Follow Up Formula Powder For Older Infants	620.00	0.00
Pediasure Premium Chocolate Powdered Health Drink	610.00	0.00

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

Analysis: To find the top categories by average discount, we need to calculate the mean of the `discountPercent` for all products within each category, round the result, and then sort and limit the output.

Approach

1. **Calculate Average Discount:** Use the aggregate function `AVG(discountPercent)` to find the average discount within each group.
2. **Round the Result:** Use `ROUND(..., 2)` to ensure the average discount is presented neatly with two decimal places.
3. **Group by Category:** Use the `GROUP BY category` clause to calculate a separate average for each distinct category.
4. **Order and Limit:** Use `ORDER BY avg_discount DESC` to put the highest average discount first, and then use `LIMIT 5` to restrict the result to the top 5 categories.

SQL

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

Key Insights

- **Highest Discount Category: Fruits & Vegetables** has the highest average discount at **15.46%**, suggesting these perishable goods are frequently discounted to move stock quickly.
- **Second Highest Discount: Meats, Fish & Eggs** follows with an average discount of **11.03%**, which is also typical for high-value perishable or temperature-sensitive goods.
- **Common Discount Tier:** A significant tier of categories, including **Ice Cream & Desserts, Chocolates & Candies**, and **Packaged Food**, share an identical average discount of **8.32%**.
- **Promotional Strategy:** The data indicates that discounts are disproportionately higher for perishable and fresh food categories compared to processed or long shelf-life items.

category character varying (120) 🔒	avg_discount numeric 🔒
Fruits & Vegetables	15.46
Meats, Fish & Eggs	11.03
Ice Cream & Desserts	8.32
Chocolates & Candies	8.32
Packaged Food	8.32

6. Find the price per gram for products above 100g and sort by best value.

Analysis: To find the "best value" (lowest price per gram), we must calculate the ratio of `discountedSellingPrice` to `weightInGms`. The query needs to filter out very light items and then sort by the calculated ratio in ascending order.

Approach

1. **Calculate Price Per Gram:** Compute the ratio of the discounted selling price to the weight (`discountedSellingPrice / weightInGms`).
2. **Round the Result:** Use `ROUND(..., 2)` to present the `price_per_gram` neatly with two decimal places.
3. **Filter by Weight:** Use the `WHERE` clause to select only products that weigh 100g or more (`weightInGms >= 100`).
4. **Order by Value:** Use `ORDER BY price_per_gram` (default ascending) to list the products from best value (lowest price per gram) to worst.
5. **Filter for Uniqueness:** Use `DISTINCT` to ensure each product is listed once

SQL

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

Key Insights

- **Best Value Items:** Several large-weight basic commodities like **Onion (3000g)**, **Aashirvaad Iodised Salt (1000g)**, and **Vicks Cough Drops (1160g)** all share the absolute lowest calculated price per gram of **₹0.02**.
- **Price and Weight Relationship:** Products with the ₹0.02 price per gram generally have significantly larger weights (1000g to 3000g), confirming that bulk purchasing offers the best value.
- **Relative Best Value:** **Carrot (500g)**, **Beetroot (500g)**, and **Raw Banana (500g)** offer a price per gram of **₹0.03**, representing the next tier of best value for mid-weight items.
- **Higher Tier Value:** Products with smaller weights within the filter (e.g., **Beetroot 250g** at **₹0.04** and **Radish White 500g** at **₹0.04**) indicate a slightly higher cost per gram compared to the very large bulk items.

name character varying (150)	weightInGms integer	discountedSellingPrice numeric (8,2)	price_per_gram numeric
Onion	3000	57.00	0.02
Aashirvaad Iodised Salt	1000	19.00	0.02
Shubh kart - Nirmal sugandhi mogra wet dhoop zipper 20 sticks	1160	28.00	0.02
Onion	1000	21.00	0.02
Vicks Cough Drops Menthol	1160	20.00	0.02
Tata Salt	1000	24.00	0.02
Carrot	500	15.00	0.03
Beetroot	500	13.00	0.03
Shubh kart - Tejas Twisted Cotton Wicks 1000n	1000	28.00	0.03
Potato	1000	29.00	0.03
Raw Banana	500	17.00	0.03
Potato	3000	84.00	0.03
Baby Potato	500	16.00	0.03
Beetroot	250	9.00	0.04
Radish White	500	19.00	0.04

7. Group the products into categories like Low, Medium, Bulk.

Analysis: This task requires conditional logic to assign a text label to each product based on its `weightInGms` value. The `CASE` statement is the ideal tool for applying multiple sequential criteria.

Approach

1. **Select Required Columns:** Select the product name and weight in grams (`name`, `weightInGms`).
2. **Apply Conditional Logic:** Use a `CASE` statement to define the categories:
 - **'Low':** For weights less than 1000g.
 - **'Medium':** For weights less than 5000g (which implicitly covers 1000g up to 4999g because the conditions are checked sequentially).
 - **'Bulk':** For all remaining weights (5000g and above).
3. **Alias the New Column:** Name the resulting column `weight_category`.
4. **Filter for Uniqueness:** Use `DISTINCT` to ensure each product appears only once

SQL

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

Key Insights

- **Dominant Category:** Based on the sample output, the majority of listed everyday products fall into the 'Low' category ($< 1000g$), including items like towels, body wash, spices, and biscuits.
- **Weight Granularity:** The 'Low' category includes a wide range of products, from very light items (e.g., **Harpic Cleaner Rim Block** at 26g) to items close to the threshold (e.g., **Lizol Disinfectant** at 500g).
- **Threshold Implications:** The **CASE** logic establishes clear break points (1kg and 5kg) for inventory management and shipping costs, distinguishing between small consumer goods, mid-sized packages, and commercial/very large items.
- **Query Consistency:** The **CASE** statement is structured correctly: conditions are checked sequentially, ensuring that a product weighing 900g is classified as 'Low' (first condition met) and not 'Medium'.

name	weightInGms	weight_category
character varying (150)	integer	text
Popular Essentials Poppy Seeds (Gasagase)	100	Low
Origami Kitchen Towels 2 In 1 Quilted 2ply	116	Low
Origami Good Karma Paper Serviettes - 100 napkins (1 ply)	58	Low
Johnson's Baby Top To Toe Bath Body Wash	210	Low
Harpic Hygienic Toilet Cleaner Rim Block Citrus	26	Low
Lizol Disinfectant Surface & Floor Cleaner Liquid Jasmine	500	Low
Sapota	348	Low
Britannia Treart Creme Wafers Chocolate	60	Low
Popular Essentials Dhania (Coriander Seeds)	200	Low
Popular Essentials Californian Almond	250	Low
Everest Shahi Biryani Masala	50	Low
Keya Italian Pizza Oregano	80	Low
Figaro Plain Green Olive	450	Low
Methi	250	Low
Sunfeast Mom's Magic Cashew & Almond Biscuits	600	Low
Kellogg's Almonds And Cranberries Granola	460	Low
Britannia 50-50 Maska Chaska Biscuits	120	Low
Pantai Pad Thai Sauce (Non-Veg)	200	Low
Hershey's Hot Chocolate Powder	250	Low
Gits Jalebi with Maker Dessert Mix	100	Low

8.What is the Total Inventory Weight Per Category

Analysis: To find the total inventory weight, we must multiply the unit weight (**weightInGms**) by the number of units available (**availableQuantity**) for every product, and then use the **SUM()** function to aggregate these totals for each **category**.

Approach

1. **Calculate Product Inventory Weight:** For each product, multiply its weight by its available quantity (`weightInGms * availableQuantity`).
2. **Aggregate Total Weight:** Use the aggregate function `SUM()` to total the calculated inventory weight.
3. **Group by Category:** Use the `GROUP BY category` clause to calculate a separate total weight for each distinct category.
4. **Order Results:** Use `ORDER BY total_weight` (default ascending) to list categories from the lightest total inventory to the heaviest.

SQL

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

Key Insights

- **Lightest Inventory: Meats, Fish & Eggs** has the lowest total inventory weight at **48,016g**. This is likely due to low stock levels necessitated by high perishability and high cost.
- **Heaviest Inventory: Munchies and Cooking Essentials** share the highest total inventory weight at **1,404,654g**, indicating the largest physical volume of stock is dedicated to these categories (suggesting large unit sizes or high unit volume).
- **Inventory Parity:** Similar to the revenue data, multiple pairs of categories have identical total weights, such as **Dairy, Bread & Batter** and **Beverages** (143,735g), and **Ice Cream & Desserts, Chocolates & Candies**, and **Packaged Food** (490,797g).
- **Logistical Impact:** The extreme difference in weight (from $\approx 48\text{kg}$ for Meats to $\approx 1400\text{kg}$ for Munchies/Cooking Essentials) highlights the varying logistical demands (storage space, transportation) required for different product categories.

category character varying (120) 	total_weight bigint 
Meats, Fish & Eggs	48016
Biscuits	84431
Fruits & Vegetables	91794
Health & Hygiene	142904
Dairy, Bread & Batter	143735
Beverages	143735
Paan Corner	348187
Personal Care	348187
Home & Cleaning	373161
Ice Cream & Desserts	490797
Chocolates & Candies	490797
Packaged Food	490797
Munchies	1404654
Cooking Essentials	1404654

