**Zepto Analysis using SQL – README**

**📌 Introduction**

This project focuses on analyzing product-level data from **Zepto**, an online grocery delivery service, using **SQL**.  
The aim was to explore product availability, pricing trends, discount patterns, and inventory distribution in order to uncover business insights that could guide pricing strategies, stock management, and customer offers.

The work involved:

* Setting up a well-structured SQL table for the dataset.
* Performing **data cleaning and transformation**.
* Running **exploratory queries** to understand product distribution and stock patterns.
* Conducting **analytical queries** to answer business-oriented questions.

**📂 Dataset Description**

The dataset includes structured product-level details with the following key attributes:

| **Column Name** | **Description** |
| --- | --- |
| sky\_id | Unique product identifier |
| category | Product category (e.g., Dairy, Snacks, Beverages) |
| name | Product name |
| mrp | Maximum Retail Price |
| discountpercent | Discount applied (%) |
| availableQuantity | Current quantity available |
| discountedSellingPrice | Price after discount |
| weightIngms | Weight of the product in grams |
| outofStock | Boolean indicating if product is out of stock |
| quantity | Quantity sold or packaged |

**🛠 Steps Performed**

**1. Table Creation & Schema Definition**

* Defined a SQL table with proper data types for each column.
* Ensured sky\_id is the **primary key**.

**2. Data Exploration**

* Counted total product entries in the dataset.
* Viewed sample records for an overview.
* Checked for missing values across columns.
* Listed unique product categories in alphabetical order.
* Counted products that are **in stock vs. out of stock**.
* Detected products listed multiple times with the same name.

**3. Data Cleaning**

* Removed products with mrp = 0 or discountedSellingPrice = 0.
* Converted prices from **paise to rupees** for accurate calculations.

**4. Analysis Queries**

* **Q1:** Top 10 products by highest discount percentage.
* **Q2:** High MRP items (> ₹300) that are out of stock.
* **Q3:** Total revenue generated per category.
* **Q4:** Products with MRP > ₹500 and discount < 10%.
* **Q5:** Top 5 categories offering highest average discount %.
* **Q6:** Price per gram for products above 100g (best value ranking).
* **Q7:** Categorized products into *low*, *medium*, *bulk* based on weight.
* **Q8:** Total inventory weight per category.

**📊 Key Insights**

* Some categories consistently provide **higher average discounts**, which could drive promotional campaigns.
* Several **high-priced products** remain out of stock — potential lost revenue opportunity.
* **Bulk products** generally offer better cost efficiency per gram.
* Inventory weight distribution varies — some categories dominate stock capacity.

**🖥 Tools Used**

* **SQL** (PostgreSQL / MySQL compatible syntax)
* SQL features used: GROUP BY, ORDER BY, DISTINCT, CASE, ROUND, aggregate functions (SUM, AVG), and conditional filtering (WHERE, HAVING).

**📖 How to Use**

1. Import the dataset into your SQL database.
2. Create the table using the provided schema in sql\_zepto\_data\_project.sql.
3. Run the **data cleaning** queries before analysis.
4. Execute analysis queries one by one to reproduce insights.
5. Modify filters (price, discount, category) to explore additional scenarios.

**👤 Author Info**

**Author:** Komal Deshmukh  
**Purpose:** To demonstrate SQL skills in **data cleaning, exploratory analysis, and insight generation** for real-world e-commerce datasets.