**Zepto E-commerce SQL Data Analysis Project**

**Project Overview**

This is a comprehensive, real-world data analyst project based on an e-commerce inventory dataset from Zepto, one of India’s fastest-growing quick-commerce startups. The project simulates the complete workflow of a data analyst, from handling raw, messy data to performing sophisticated, business-focused analysis using SQL.

The primary goal is to use SQL to set up an e-commerce inventory database, perform exploratory data analysis, clean the data, and write business-driven queries to derive insights around pricing, inventory, stock availability, and potential revenue.

**Dataset Overview**

The dataset, originally sourced from Kaggle, mimics a real-world e-commerce inventory system where each row represents a unique SKU (Stock Keeping Unit). It's common for the same product to appear multiple times with different package sizes, weights, or discounts to enhance visibility on the platform.

**Dataset Columns**

* **sku\_id:** Unique identifier for each product entry.
* **name:** The product name as it appears on the app.
* **category:** Product category (e.g., Fruits, Snacks, Beverages).
* **mrp:** Maximum Retail Price, originally in paise.
* **discountPercent:** The discount applied to the MRP.
* **discountedSellingPrice:** The final price after the discount.
* **availableQuantity:** The number of units available in inventory.
* **weightInGms:** The product's weight in grams.
* **outOfStock:** A boolean flag indicating if the product is out of stock.
* **quantity:** The number of units per package.

**Project Workflow & Analysis**

The project followed a structured workflow, moving from initial data exploration to cleaning and finally to deriving business insights.

**1. Data Exploration**

The initial phase focused on understanding the dataset's structure, content, and quality.

* Counted the total number of records to gauge the scale of the data.
* Viewed a sample of the dataset to understand its structure and content.
* Checked for null values across all columns to identify missing data.
* Identified all distinct product categories available.
* Compared the counts of in-stock vs. out-of-stock products.
* Detected products that were present multiple times, representing different SKUs.

**2. Data Cleaning**

This phase involved preparing the data for analysis by addressing inconsistencies and formatting issues.

* Identified and removed rows where both the MRP and the discounted selling price were zero, as these entries provide no value.
* Converted the mrp and discountedSellingPrice columns from paise to rupees for better readability and consistency in financial analysis.

**3. Business Insights**

With clean data, the focus shifted to answering key business questions to drive strategy.

* **Best-Value Products:** Found the top 10 products with the highest discount percentages.
* **Stock Analysis:** Identified high-MRP products that are currently out of stock, highlighting potential revenue loss.
* **Revenue Estimation:** Estimated the potential revenue for each product category to identify top-performing categories.
* **Pricing Strategy:** Filtered for expensive products (MRP > ₹500) that have minimal discounts.
* **Category Performance:** Ranked the top 5 categories that offer the highest average discounts to customers.
* **Value-for-Money Analysis:** Calculated the price per gram for products to identify which items offer the best value to consumers.
* **Inventory Segmentation:** Grouped products based on weight into "Low," "Medium," and "Bulk" categories to understand inventory distribution.
* **Logistics Planning:** Measured the total inventory weight for each product category, which is useful for warehousing and logistics planning.

**Acknowledgments**

* The dataset for this project was sourced from [**Kaggle**](https://www.kaggle.com/datasets/palvinder2006/zepto-inventory-dataset/data?select=zepto_v2.csv).
* The project structure and methodology were inspired by a public video tutorial by [**Amlan Mohanty**](https://www.youtube.com/@amlanmohanty1).
* All SQL queries for exploration, cleaning, and analysis were written independently for this project.

**Author**

* **Parveen Jalwal**
* **LinkedIn**: <https://www.linkedin.com/in/parveen-jalwal-201a2a302>