

## Expanded Explanation of the Kaggle Dataset

### Overview

This dataset, sourced from Kaggle, is designed for retail analytics, covering a vast range of data points relevant to understanding customer behavior, sales trends, product performance, and promotional effectiveness. With over a million rows and 100+ columns, it provides insights for predictive modeling, including total sales (regression) and customer churn (classification). The dataset is structured into nine tables, each serving a unique analytical purpose.

---

### 1. Customer Information

This table provides details about the customers, helping businesses segment their audience and understand customer demographics.

#### Columns:

- **Customer\_ID:** Unique identifier for each customer.
- **Age:** Age of the customer, which can help in age-based segmentation.
- **Gender:** Customer's gender (Male, Female, Other).
- **Income Bracket:** Categorization of customer income (Low, Medium, High).
- **Loyalty Program:** Indicates if a customer is part of a loyalty program (Yes/No).
- **Membership Years:** The duration of a customer's loyalty program membership.
- **Churned:** Whether the customer has stopped purchasing (Yes/No). This is the target variable for churn prediction.
- **Marital Status:** Indicates if the customer is Single, Married, or Divorced.
- **Number\_of\_Children:** Number of children in the customer's household.
- **Education Level:** Highest education attained (High School, Bachelor's, Master's, etc.).
- **Occupation:** The profession of the customer.

**Use Cases:** Helps in customer segmentation, churn analysis, and targeted marketing campaigns.

---

### 2. Transactional Data

This table captures all customer purchases, forming the backbone of sales analysis.

#### Columns:

- **Transaction\_ID:** Unique identifier for each transaction.

- **Transaction Date:** Date on which the transaction took place.
- **Product\_ID:** Identifier linking the transaction to a specific product.
- **Product Category:** Type of product (Electronics, Clothing, Groceries, etc.).
- **Quantity:** Number of units purchased in the transaction.
- **Unit Price:** Price per unit of the product.
- **Discount Applied:** The discount percentage or amount applied.
- **Payment Method:** How the customer paid (Credit Card, Debit Card, Cash, etc.).
- **Store Location:** The branch or location where the purchase was made.

**Use Cases:** Analyzing purchase patterns, pricing strategy evaluation, discount effectiveness, and payment preferences.

---

### 3. Customer Behavior Metrics

These metrics offer insights into how customers interact with the store and their purchasing tendencies.

#### Columns:

- **Avg\_Purchase\_Value:** The average value of purchases made by the customer.
- **Purchase Frequency:** Frequency of purchases (Daily, Weekly, Monthly, Yearly).
- **Last\_Purchase\_Date:** Date of the most recent purchase.
- **Avg\_Discount\_Used:** The average discount percentage utilized.
- **Preferred Store:** The store location most frequently visited by the customer.
- **Online Purchases:** The number of purchases made through the online store.
- **In-Store Purchases:** The number of purchases made at physical stores.
- **Avg\_Items\_Per\_Transaction:** The average number of items bought per transaction.
- **Avg\_Transaction\_Value:** The average monetary value of transactions.
- **Total\_Returned\_Items:** The total number of returned products.
- **Total\_Returned\_Value:** The total value of returned products.

**Use Cases:** Helps in identifying customer loyalty, predicting high-value customers, and optimizing store inventory.

---

### 4. Sales Data

This table consolidates overall customer sales performance.

**Columns:**

- **Total Sales:** Cumulative sales per customer over the last year (Target for regression analysis).
- **Total Transactions:** The total number of transactions by each customer.
- **Total\_Items\_Purchased:** Total quantity of items bought by the customer.
- **Total\_Discounts\_Received:** Total value of discounts received.
- **Avg\_Spent\_Per\_Category:** Average expenditure per product category.
- **Max\_Single\_Purchase\_Value:** The highest amount spent in a single transaction.
- **Min\_Single\_Purchase\_Value:** The lowest amount spent in a single transaction.

**Use Cases:** Helps in sales forecasting, customer lifetime value prediction, and identifying high-spending customers.

---

## 5. Product Information

This table details product attributes, which are useful for inventory and demand analysis.

**Columns:**

- **Product Name:** Name of the product.
- **Product Brand:** Manufacturer or brand of the product.
- **Product Rating:** Average customer rating.
- **Product Review Count:** Number of reviews received.
- **Product Stock:** Number of units available.
- **Product Return Rate:** Percentage of returned units.
- **Product Size/Weight/Color/Material:** Physical attributes of the product.
- **Manufacture Date & Expiry Date:** Manufacturing and expiration details.
- **Shelf Life:** Expected usability duration.

**Use Cases:** Helps in demand planning, restocking decisions, and identifying best-selling products.

---

## 6. Promotional Data

Analyzes the impact of different marketing campaigns.

#### Columns:

- **Promotion\_ID:** Unique identifier for each promotion.
- **Promotion Type:** Type of promotion (e.g., Buy One Get One Free, 20% Off).
- **Promotion Start & End Date:** Duration of the promotion.
- **Promotion Effectiveness:** Effectiveness rating (High, Medium, Low).
- **Promotion Channel:** Platform where promotion was advertised (Online, In-store, Social Media).
- **Promotion Target Audience:** Intended customer group (New, Returning Customers).

**Use Cases:** Determines the most effective promotional strategies and ROI calculations.

---

### 7. Geographical Data

Understanding customer and store locations.

#### Columns:

- **Customer Zip Code, City, State:** Location of the customer.
- **Store Zip Code, City, State:** Location of the store.

**Use Cases:** Helps in regional sales analysis and location-based marketing.

---

### 8. Seasonal and Temporal Data

Evaluates how time-based factors affect sales.

#### Columns:

- **Holiday Season:** Whether the transaction happened during a holiday (Yes/No).
- **Season:** The season (Winter, Spring, Summer, Fall).
- **Weekend:** Whether the transaction happened on a weekend (Yes/No).

**Use Cases:** Helps in seasonal demand forecasting and holiday-based marketing strategies.

---

### 9. Customer Interaction Data

Captures how customers engage with the business beyond transactions.

#### Columns:

- **Customer Support Calls:** Number of times a customer contacted support.
- **Email Subscriptions:** Whether the customer subscribes to marketing emails.

- **App Usage:** Frequency of mobile app usage.
- **Website Visits:** Number of visits to the online store.
- **Social Media Engagement:** Likes, comments, shares on social media.

**Use Cases:** Helps improve customer service, engagement strategies, and digital marketing performance.