Key Insights

- The dataset contains **500,000 sales transactions** with details on **customers, purchases, payments, discounts, returns, delivery times, and feedback**, providing insights into **sales performance, customer behavior, and operational efficiency**.

- The transaction data spans from **January 1, 2022, to December 14, 2022**, covering nearly a full year. This allows for a comprehensive analysis of seasonal trends, customer behavior, and sales performance over time.

- The dataset contains significant missing values, particularly in **CustomerID, TransactionDate, PaymentMethod, StoreType, CustomerAge, CustomerGender, ProductName, and Region** columns.
    - **50,000 missing CustomerID & TransactionDate**: This suggests a substantial portion of transactions lack customer tracking, impacting customer behavior analysis.
    - **50,000 missing PaymentMethod & StoreType**: Could indicate incomplete data collection or system issues in recording payment/store details.
    - **50,000 missing CustomerAge & Gender**: Prevents effective demographic insights and customer segmentation.
    - **42,633 missing Region**: Limits geographic-based sales analysis.

- The dataset includes **500,000 transactions**, with:
    - **Gross Sales:** ₹10,206.39 crore
    - **Net Sales (after discounts & returns):** ₹10,202.66 crore
    - **Average Transaction Value:** ₹20,412.79

- Sales are **fairly balanced between online and in-store purchases** across regions, with no single channel dominating.
    - **South region** leads in transactions, with both In-Store (70,488) and Online (70,604) channels generating high sales (~₹1,590 crore & ₹1,580 crore, respectively).
    - **East region** follows, with Online (58,358 transactions, ₹1,334 crore) slightly ahead of In-Store (58,217 transactions, ₹1,319 crore).
    - **North and West** regions have relatively lower transaction volumes but still contribute significantly to revenue (~₹1,080–₹1,090 crore per channel).

- Sales and transaction volumes are **evenly distributed across major Indian cities**, with no single city dominating.
    - **Kolkata** leads in net sales (₹1,027 crore) and transactions (50,141), followed closely by Ahmedabad, Bangalore, and Pune.

- ○ **Hyderabad** has the lowest net sales (₹1,008 crore) and transactions (49,727) among the top 10 cities, but the difference is marginal.

- The dataset has **three identified genders (Male, Female, Other)**, with a fairly **balanced customer distribution** across them.

- **Debit Card (113,015 customers) is the most used**, followed closely by **UPI (112,517 customers)**, reflecting India's strong adoption of digital payments.

- The **25-64 age group dominates** the customer base, with each segment (25-34, 35-44, 45-54, 55-64) having around **79,000 customers**.

- **Young adults (18-24) make up the smallest identified group (54,976 customers)**, possibly indicating lower purchasing power or different shopping preferences.

- The **median delivery time is 4 days**, indicating that most customers receive their orders relatively quickly.

- The **median shipping cost is ₹48.04**, which suggests that shipping charges are fairly consistent across transactions, with the majority of customers paying a similar amount for delivery.

- The dataset shows a **high concentration of repeat customers (48,953)** compared to **new customers (41)**. This indicates that the business has a **strong base of returning customers**, and acquisition of new customers seems relatively low.

- Returns are fairly **evenly distributed across cities**, with **Delhi** having the highest return volume (**25,155**), closely followed by **Bangalore (25,126)** and **Ahmedabad (25,081)**. This suggests that the **return rate is consistent across key cities**, potentially indicating similar customer expectations or product quality issues.

- **Discounts ranging from 15-35%** contribute consistently to high **net sales**, suggesting that customers are particularly responsive to discounts in this range.

- The **marginal difference** in sales between promotional and non-promotional transactions suggests that **promotions are not significantly altering overall sales volume** but may be influencing purchasing behavior in other ways (e.g., increased volume or customer loyalty).

- **Returned and non-returned transactions** have the **same average feedback score of 3.0**, suggesting that **returns may not be solely driven by customer dissatisfaction**.

- **Feedback scores remain consistent at around 3.0** across all delivery times, with **minor fluctuations (2.99–3.01)**.