# EDA Report: Online Shopping Dataset

## 1. Dataset Overview

- \*\*Rows:\*\* 100

- \*\*Columns:\*\* 10

- The dataset consists of customer information including demographics, spending behavior, and satisfaction.

### Columns:

- `Customer\_ID`: Unique identifier for each customer

- `Age`: Age of the customer

- `Gender`: Gender (Male/Female)

- `Region`: Geographical region

- `Total\_Spent`: Total amount spent by the customer

- `No\_of\_Orders`: Number of orders placed

- `Average\_Order\_Value`: Average value per order

- `Payment\_Method`: Mode of payment used

- `Returning\_Customer`: Indicates whether the customer is returning (Yes/No)

- `Satisfaction\_Rating`: Rating out of 5

---

## 2. Data Quality Checks

- \*\*Missing Values:\*\* None

- \*\*Data Types:\*\* Appropriate for all columns

- \*\*Outliers:\*\* Possible in `Total\_Spent` and `No\_of\_Orders`, examined using boxplots

---

## 3. Univariate Analysis

- \*\*Gender Distribution:\*\* Balanced mix of Male and Female customers

- \*\*Region:\*\* All regions (East, West, North, South) are represented

- \*\*Total Spent:\*\* Right-skewed distribution; some customers spend significantly more

- \*\*Satisfaction Rating:\*\* Normally distributed with a slight skew

---

## 4. Bivariate and Grouped Analysis

- \*\*Spending by Gender:\*\*

- Males and females show similar spending behavior on average.

- \*\*Spending by Payment Method:\*\*

- Boxplots indicate certain methods (e.g., Credit Card) have higher median spending.

- \*\*Returning vs New Customers:\*\*

- Countplot shows a balanced mix.

- \*\*Orders vs Total Spending:\*\*

- Positive correlation—customers who place more orders spend more.

- \*\*Average Order Value by Payment Method:\*\*

- Some methods have higher average values, suggesting preference for high-ticket items.

---

## 5. Correlation Analysis

- \*\*Positive Correlations:\*\*

- `No\_of\_Orders` and `Total\_Spent`

- `Average\_Order\_Value` and `Total\_Spent`

- \*\*Weak Correlation:\*\*

- Age and Satisfaction/Spending are weakly correlated

---

## 6. Age Group Analysis

- Customers were bucketed into age groups:

- `<20`, `20–29`, `30–39`, `40–49`, `50–59`, `60+`

- \*\*Highest Spending Age Group:\*\* Typically in the 40–49 and 50–59 ranges.

---

## 7. Regional Satisfaction

- Calculated average satisfaction rating by region.

- Identified the region with the \*\*highest customer satisfaction\*\*.

---

## 8. Summary of Insights

- Younger customers place fewer but more valuable orders.

- Certain payment methods correlate with higher spending.

- Returning customers are not drastically more valuable than new ones, but trends may differ by region.

- The dataset is clean and well-suited for predictive modeling or customer segmentation.

---

### ✅ Recommended Next Steps

- Segment customers using clustering (e.g., KMeans)

- Predict satisfaction or churn using classification models

- Deploy dashboard for stakeholder insights