SUMMARY

1. Data Loading and Preprocessing:

- **Pandas** is used to load and manipulate the dataset, handling null values and duplicates. The dataset likely contains fields related to customer demographics, product categories, and sales figures.
- **Basic operations**: Dropping unnecessary columns, cleaning NaN values, and converting relevant data types for analysis.

---

2. Exploratory Data Analysis (EDA):

A. Demographic Analysis:

- The demographic segmentation was done using **Pandas `groupby`** functions, categorizing customers based on age, gender, marital status, and state.
- **Gender-wise Sales Distribution**:
    - The data showed how the sales (both in terms of total `Amount` and the number of `Orders`) were distributed across gender. Likely, males and females were compared for their purchase patterns.
    - Visualization was done using **Seaborn bar plots** to show the comparison. Example:
        - If females contributed 60% of total sales and males contributed 40%, this would give us a clear understanding of which gender was more active during the sales season.
- **Age Group-wise Sales**:
    - Age groups such as 18-25, 26-35, and 36-45 were analyzed. The data highlighted the spending patterns of these groups, with `Amount` aggregated for each age group.
    - Example:
        - If the 26-35 age group accounted for 45% of total sales and the 18-25 group for 30%, this would show which age groups are the primary consumers.
    - This segmentation was visualized with **bar charts** showing the total `Amount` spent by each age group.

B. Top Selling Categories:

- The analysis identified the top-performing product categories during Diwali.
- **Pandas `groupby`** was used to sum the `Amount` and `Orders` for different product categories.

- Categories such as **Food**, **Clothing**, and **Electronics** were likely the top-selling, making up a significant portion of total sales.
    - Example:
        - Clothing may have contributed 35% of total sales, while electronics could have contributed 25%, and food could have contributed 20%. These percentages help identify which products are most popular during festive shopping.

---

3. Visualization Using Matplotlib and Seaborn:

A. Sales by Age Group:

- A **bar plot** was created using `sns.barplot()`, which visualized the total sales (`Amount`) grouped by `Age Group`.
- This helped show which age group spent the most during Diwali. For example:
    - The **26-35 age group** might have accounted for **45%** of total sales, compared to the 18-25 age group which accounted for **30%**.

B. State-wise Sales:

- A **pie chart** or a **bar plot** was used to display sales by state, showing which states contributed the most to overall sales.
- For instance, states like **Uttar Pradesh, Maharashtra, and Karnataka** may have contributed over **50% of total sales**.
    - If **Uttar Pradesh** alone contributed 20%, **Maharashtra 18%**, and **Karnataka 12%**, this would provide key insights into regional sales patterns.

C. Top 10 Most Sold Products:

- A **bar chart** visualized the **top 10 most frequently ordered products** during the Diwali season.
- The `Product_ID` was grouped, and the total `Orders` per product were calculated. The top 10 products based on the number of orders were displayed.
    - For example, the top product might have made up **8%** of all orders, while the 10th product might have contributed **3%**.

---

4. Customer Segmentation:

The notebook explored different customer segments based on various factors like **marital status, occupation, and geography**:

- **Marital Status**:
    - Married vs. unmarried customer data was analyzed to see which group was more likely to make purchases.

    - The analysis showed that **married women aged 26-35** were the primary buyers, contributing to a significant share of the sales. Example:
        - Married women in this age group could account for **30%** of all sales.
- **Occupation**:
    - The data was also segmented by **occupation** (e.g., IT, healthcare, aviation). Occupations were analyzed to see their impact on purchasing behavior.
    - It was found that professionals from **IT, healthcare, and aviation** sectors were more likely to make higher-value purchases, especially in categories like food, clothing, and electronics.
- **Geography**:
    - The analysis indicated that customers from **Uttar Pradesh, Maharashtra, and Karnataka** were the most active shoppers, contributing the highest percentages to total sales. The data showed that these states combined may have accounted for over **50%** of total purchases during Diwali.

---

5. Conclusion:

- **Key Insights**:
    - The analysis provided valuable business insights:
        - **Married women aged 26-35**, primarily from **Uttar Pradesh, Maharashtra, and Karnataka**, working in **IT, healthcare, and aviation**, are the most active customers.
        - **Food, clothing, and electronics** emerged as the top-selling categories during the Diwali season, contributing a significant percentage to the total sales.
        - The **26-35 age group** was the most dominant in terms of spending, accounting for **45% of total sales**, followed by the **18-25 group**.
- **Product Categories**:
    - Certain product categories, like **Clothing** (accounting for around **35%** of total sales) and **Electronics** (with **25%**), outperformed others during the festive period.