

## PROJECT REPORT

### Exploratory Data Analysis on Diwali Sales Dataset

#### Objective:

The objective of this project is to perform Exploratory Data Analysis (EDA) on the Diwali Sales dataset to understand customer purchasing behavior, product popularity, and regional sales trends during the festive season.

#### Tools & Libraries Used:

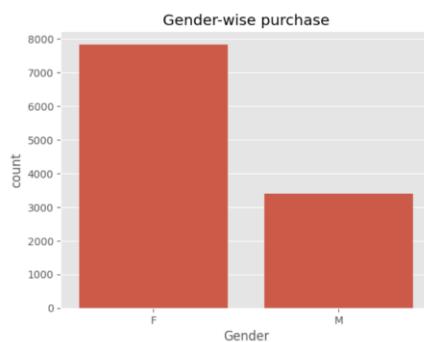
- Python
- Google Colab
- Pandas
- NumPy
- Matplotlib
- Seaborn

#### Data Cleaning & Preprocessing:

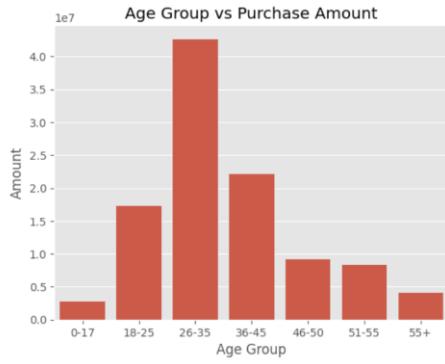
- Removed unnecessary columns
- Handled missing values by dropping null rows
- Converted Amount column into numeric format
- Removed duplicate entries
- Exported the cleaned dataset as **diwali\_cleaned.csv**

#### Visualizations Performed:

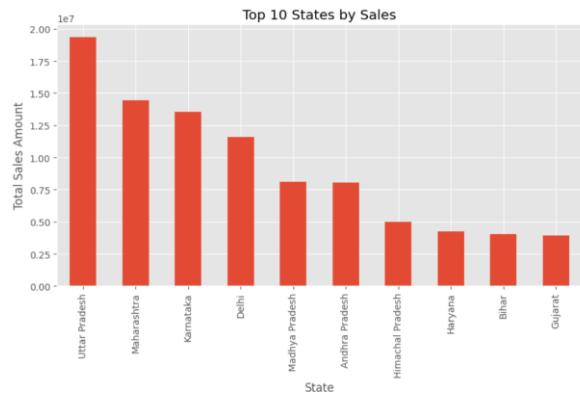
- Gender-wise purchase



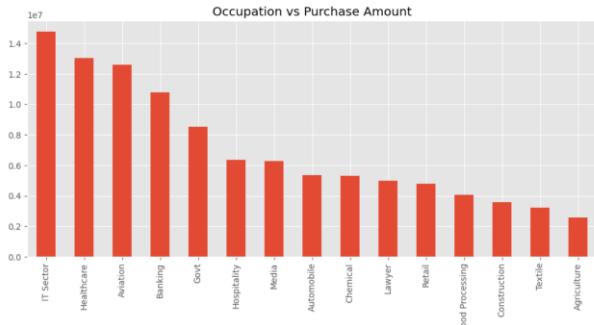
- Age group vs purchase amount



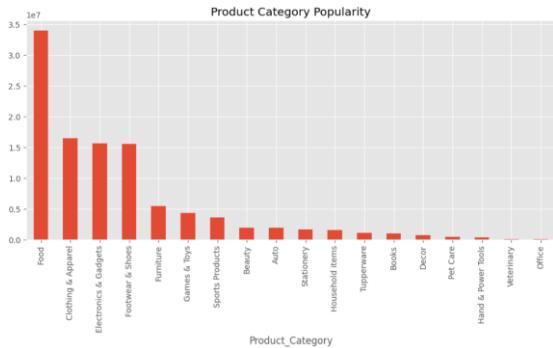
- State-wise sales (bar plot)



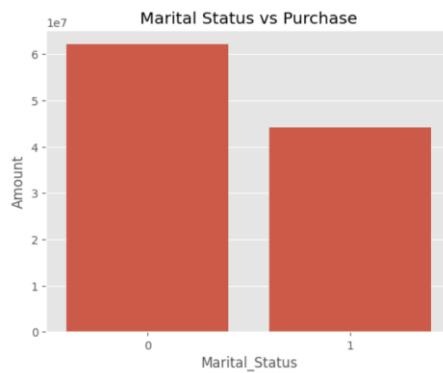
- Occupation vs purchase



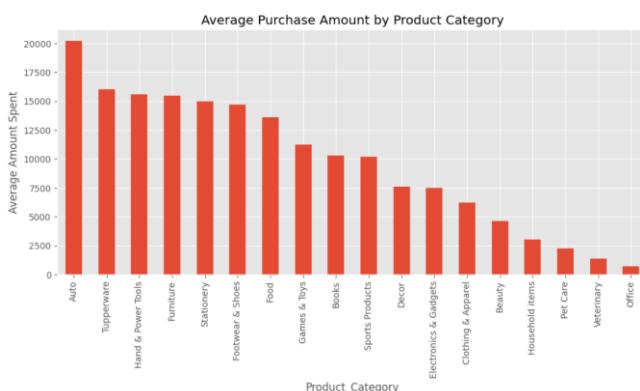
- Product category popularity



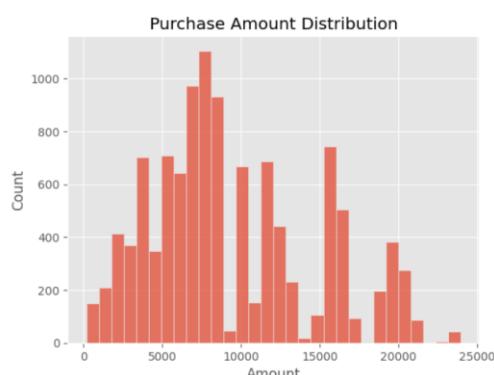
- Marital status vs purchase



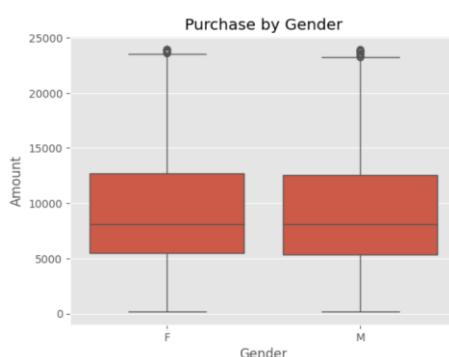
- Average Purchase Amount by Product Category



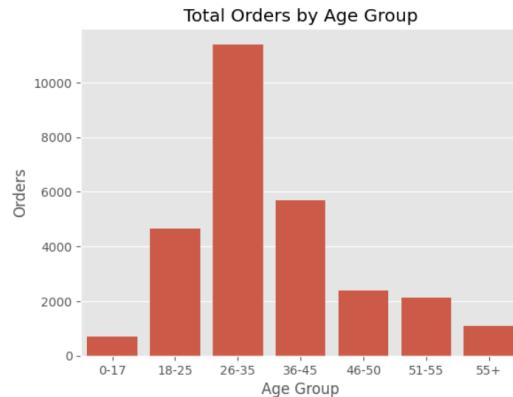
- Purchase amount distribution (histogram)



- Boxplot: Purchase by gender



- Total Orders by Age Group



### Key Insights:

- Female customers contribute more to total sales.
- Age group **26–35 years** shows the highest purchase amount and order frequency.
- **Uttar Pradesh, Maharashtra, and Karnataka** are top-performing states.
- Married customers spend more compared to unmarried customers.
- IT, Healthcare, and Aviation professionals are major buyers.
- Clothing and Food product categories dominate festive sales.

### Challenges Faced & Solutions:

Challenge	Solution
Missing values	Removed null rows
Incorrect data types	Converted Amount column to numeric
Column name mismatch	Verified column names before grouping