# PROJECT REPORT

**Project Title: Durga Puja Sale Data Analysis with Python, Numpy, Pandas, Matplotlib.**

## Project description:

**The "Durga Puja Sales Analysis" project involves analyzing sales data related to Durga Puja festivities using Python libraries such as NumPy, Pandas, Matplotlib, and Jupyter Notebook.**

**This Project Aims to analyze historical sales data, identify trends, patterns and insights and present findings through meaningful visualizations and report.**

## Key features:

1. **Data Acquisition:**

   **Raw sales data is collected from YouTube source, which include information such as sales transactions, product categories, amount, customer demographics, and state names.**

2. **Data clearing and preprocessing:**

   **The collected data is preprocessed using Pandas to clean, format, and organize it into a structured dataframe. Tasks such as handling missing values, converting data types, and removing duplicates are performed to ensure data quality and consistency.**

3. **Exploratory data analysis (EDA):**

   **Using Pandas and Matplotlib, exploratory data analysis techniques are applied to gain insights into sales patterns, trends, and correlations. Descriptive statistics, visualizations (e.g., line plots, bar charts, histograms), and statistical tests are used to analyze sales data and identify key factors influencing sales performance.**

4. **Market Analysis:**

   **Analyzing transaction data to identify frequent item associations (Food & Clothing) and patterns in customer purchasing behavior. This can be used for selling and stock inventory.**

5. **Visualization and reporting:**

   **Creating informative visualizations such as line charts, bar plots to illustrate key findings and insights. Additionally, generating summary reports to communicate results effectively.**

## Technologies used:

**Python:**

**The programming language used for data analysis, manipulation, visualization.**

**Numpy:**

**Library for numerical computing, providing support for mathematical operations and array manipulation.**

**Pandas:**

**Library for data manipulation and analysis, offering powerful data structures and functions for data processing.**

**Matplotlib**:

**Library for creating static, animated and interactive visualizations in python.**

## Project Outcome:

**The sales data analysis portfolio project delivers a comprehensive analysis of sales data, providing actionable insights, visualizations and recommendations. By utilizing the capabilities of Numpy, Pandas and Matplotlib, gained a deeper understanding of sales performance, customer behaviour and market trends, ultimately driving informed decision.**

**Additional Resources:**

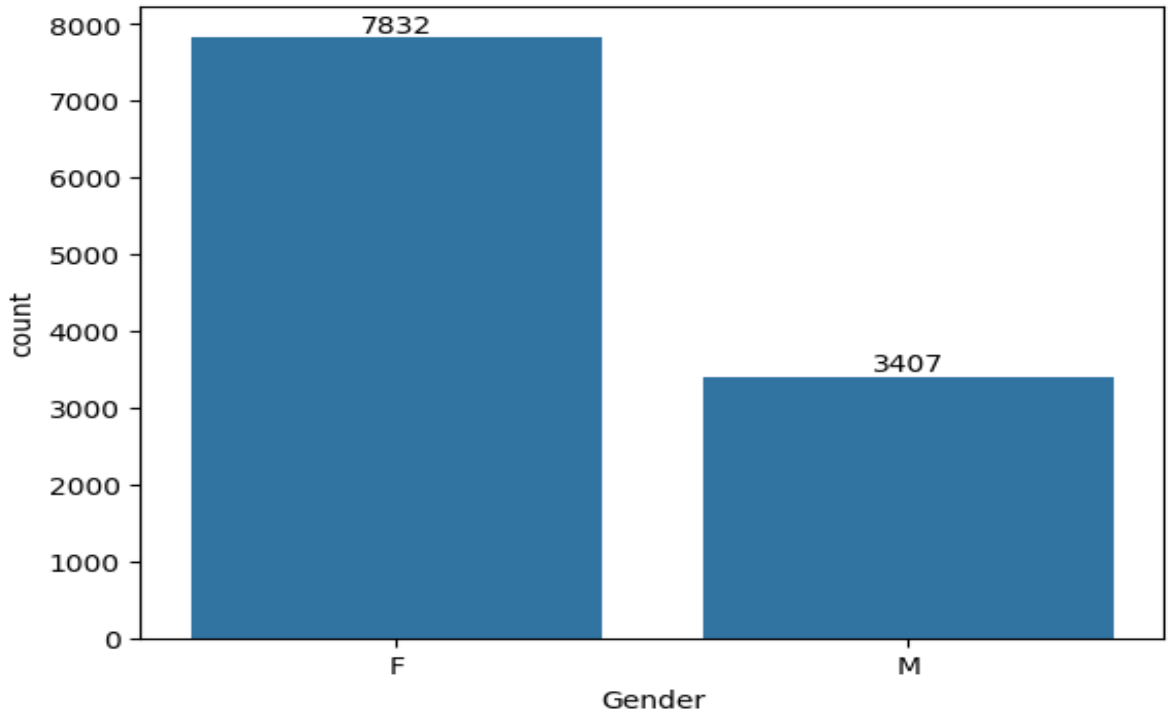
1. **Jupyter-Notebook:**

   **Integrated development environment (IDE) commonly used for data analysis and visualization tasks.**

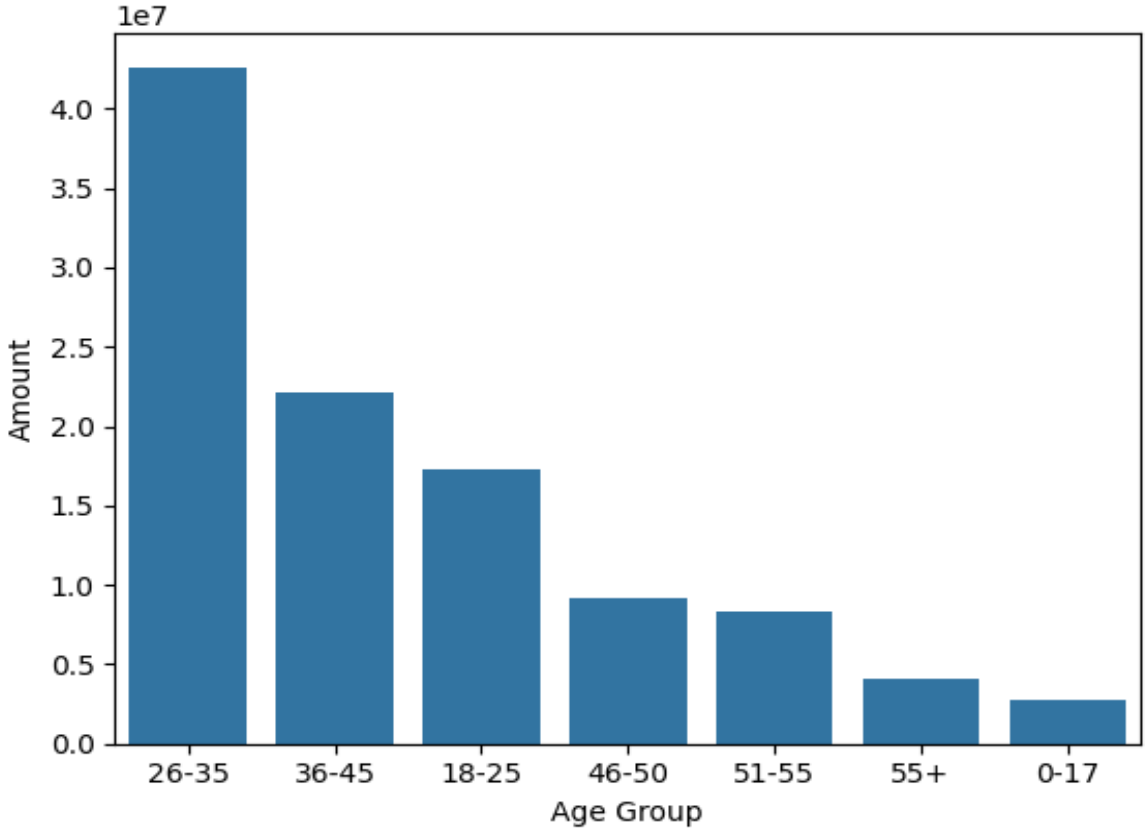
2. **Seaborn:**

   **Statistical data visualization library that complements Matplotlib with high level interface and aesthetic appeal.**

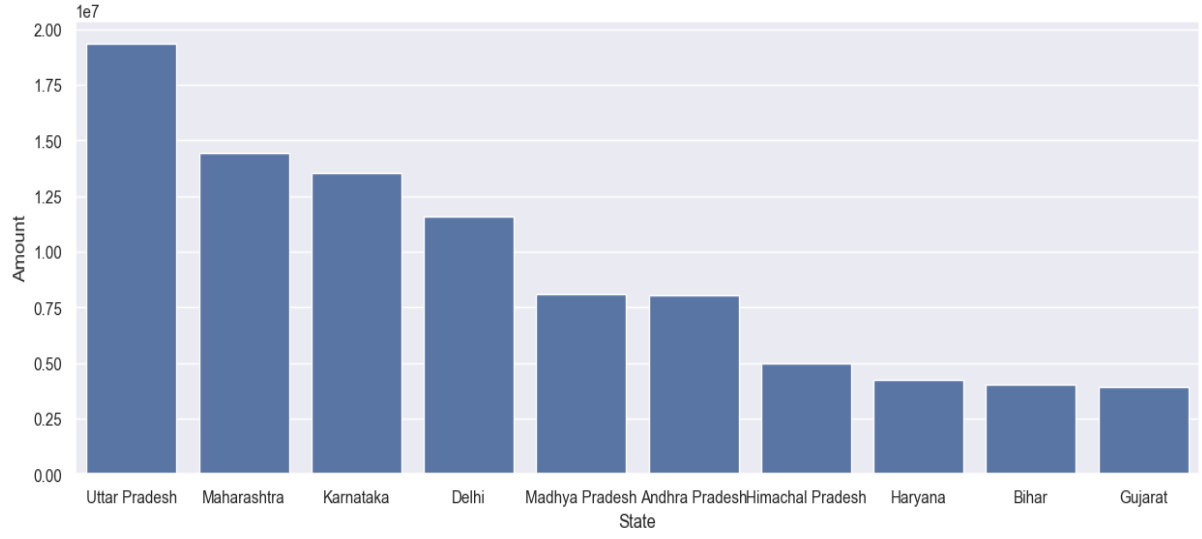
## Project Results:



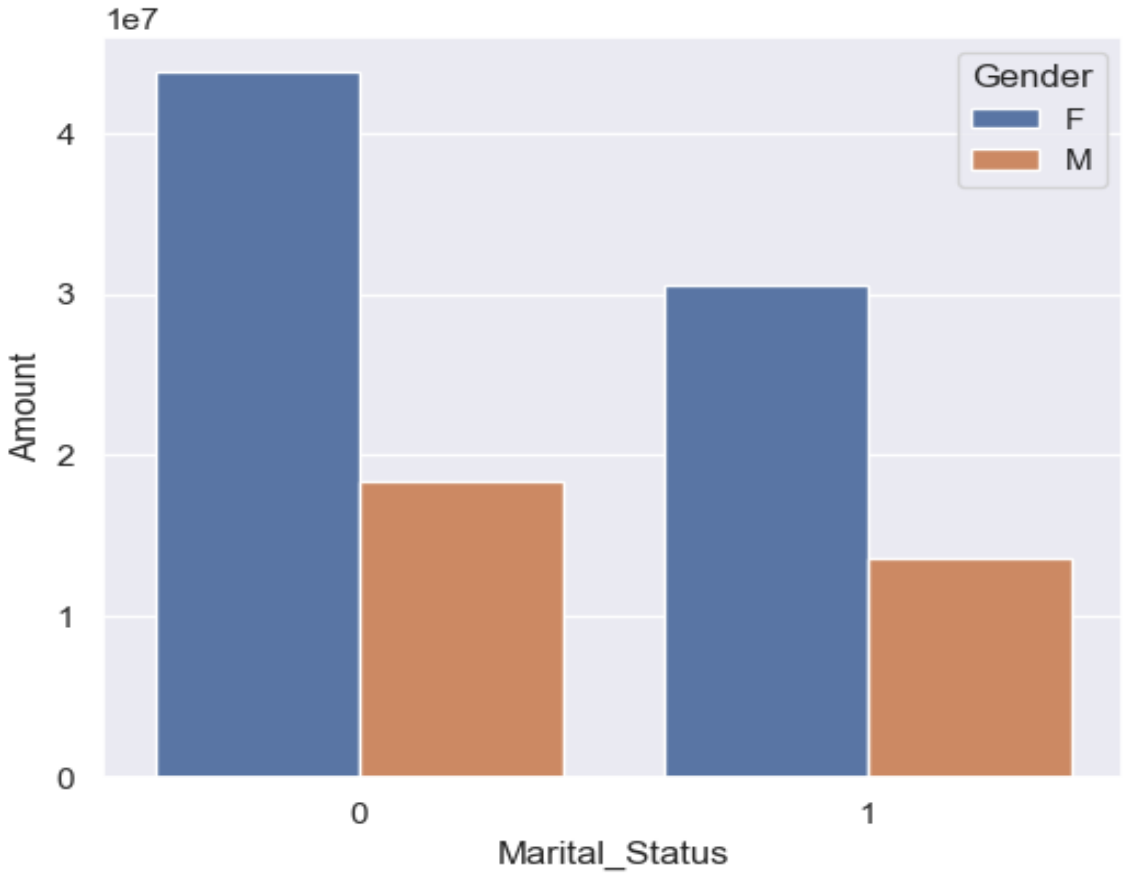
1. From above graphs we can see that most of the buyers are females and even the purchasing power of females are greater than men.



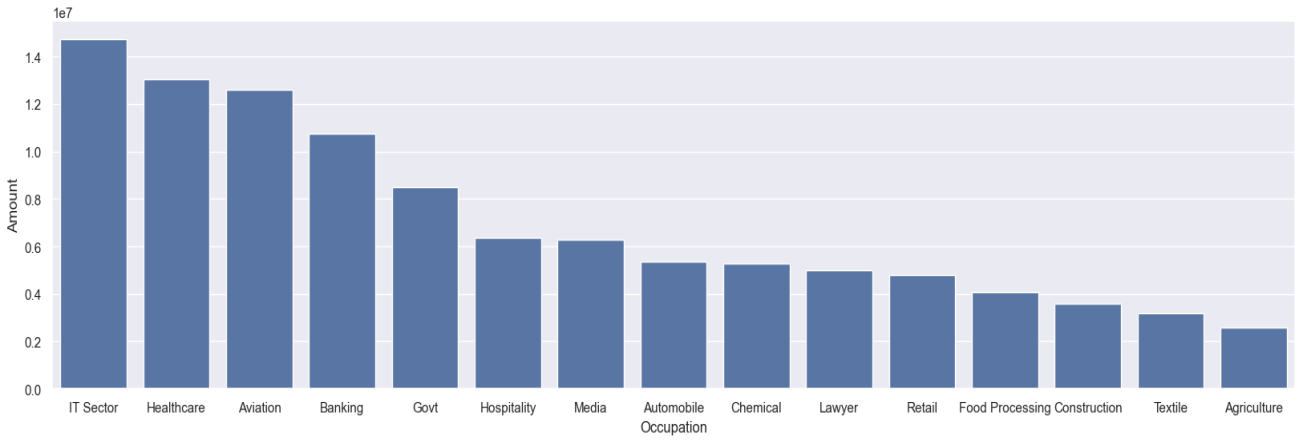

2. From above graphs we can see that most of the buyers are of age group between 26-35 year females.



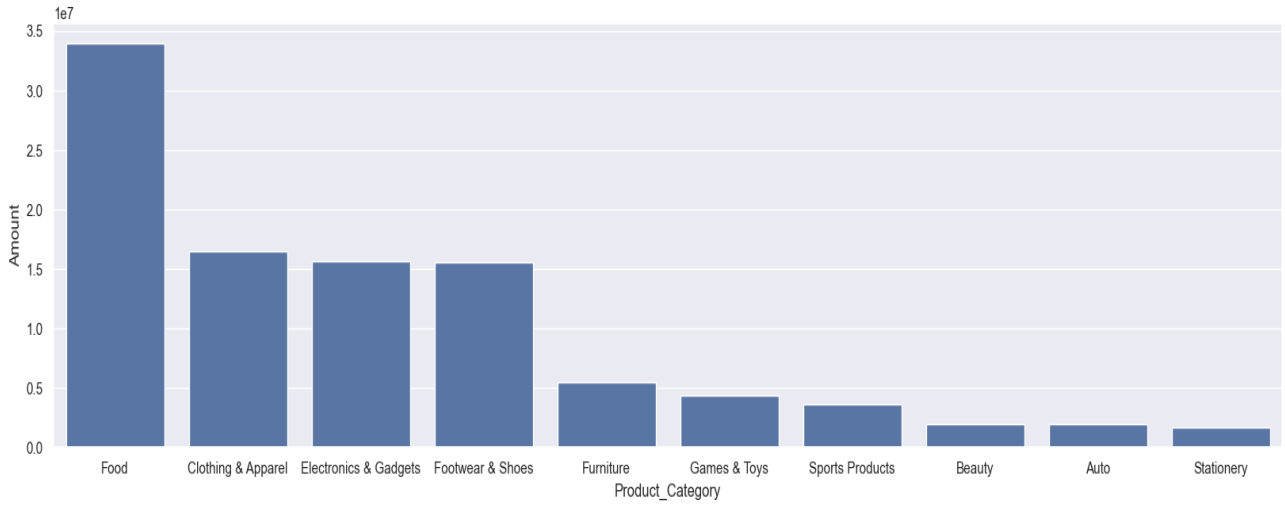
3. From above graphs we can see that most of the orders & total sales amount are from Uttar Pradesh, Maharashtra and Karnataka respectively.




4. From above graphs we can see that most of the buyers are married (women) and they have high purchasing rate.



5. From above graphs we can see that most of the buyers are working in IT, Healthcare and Aviation sector.




6. From above graphs we can see that most of the sold products are from Food, Clothing and Electronics category.

**Married women age group 26-35 years from UP, Maharashtra and Karnataka working in IT, Healthcare and Aviation are more likely to buy products from Food, Clothing and Electronics category.**

**Thank You!!**