

edunet  
foundation

Proudly supported by



## Diwali Sales Analysis

Team ID - CU\_CP\_Team\_9478

Gajendra Patil  
Komal Rathod  
Gauri Patil  
Krushank Kadam  
Heena Makwana

## Project Objectives

- Problem Statement
- Project Overview – Introduction
- End Users
- Wow Factor in Project
- Modelling/Block Diagram/Flow of Project
- Result/outcomes
- Conclusion
- Future Perspective



# Problem Statement

➤ By examining consumer demographics and product popularity, this project seeks to maximize Diwali sales performance using python. This will allow for more focused marketing, effective inventory control, and enhanced customer satisfaction.

## Diwali Sales Analysis

Using Python



# Project overview - Introduction

➤ The Diwali Sale Data Analysis project aims to leverage data analytics to improve customer experience and sales performance. By analyzing the Diwali sale data, we sought to identify potential customers across various demographics and determine the most popular product categories and products. This analysis helps in strategic decision-making, targeted marketing, and efficient inventory management.



# End User

## Individual Consumers:

These are the people directly purchasing goods and services during the Diwali season. This includes people of all ages, genders, and socioeconomic backgrounds across India.

## Businesses and Retailers:

These entities use sales data to understand consumer behavior, optimize marketing strategies, and manage inventory.



# Wow Factor in Solution

- Improve customer experience by identifying potential customers across different states, occupations, genders, and age groups.
- Improve sales by identifying the most selling product categories and products, helping to plan inventory and meet demand effectively.



# Modelling

## Data Cleaning & Manipulation

- **Handling Missing Values:** Imputed missing values using appropriate techniques to avoid skewing the analysis.
- **Outlier Detection and Removal:** Identified and removed outliers to maintain data integrity.
- **Standardizing Data Formats:** Ensured consistency in data formats for seamless analysis.

- Created new features, such as average spend per age group and product count per gender.

```
# use describe() for specific columns  
df[['Age', 'Orders', 'Amount']].describe()
```

	Age	Orders	Amount
count	11251.000000	11251.000000	11239.000000
mean	35.421207	2.489290	9453.610858
std	12.754122	1.115047	5222.355869
min	12.000000	1.000000	188.000000
25%	27.000000	1.500000	5443.000000
50%	33.000000	2.000000	8109.000000
75%	43.000000	3.000000	12675.000000
max	92.000000	4.000000	23952.000000

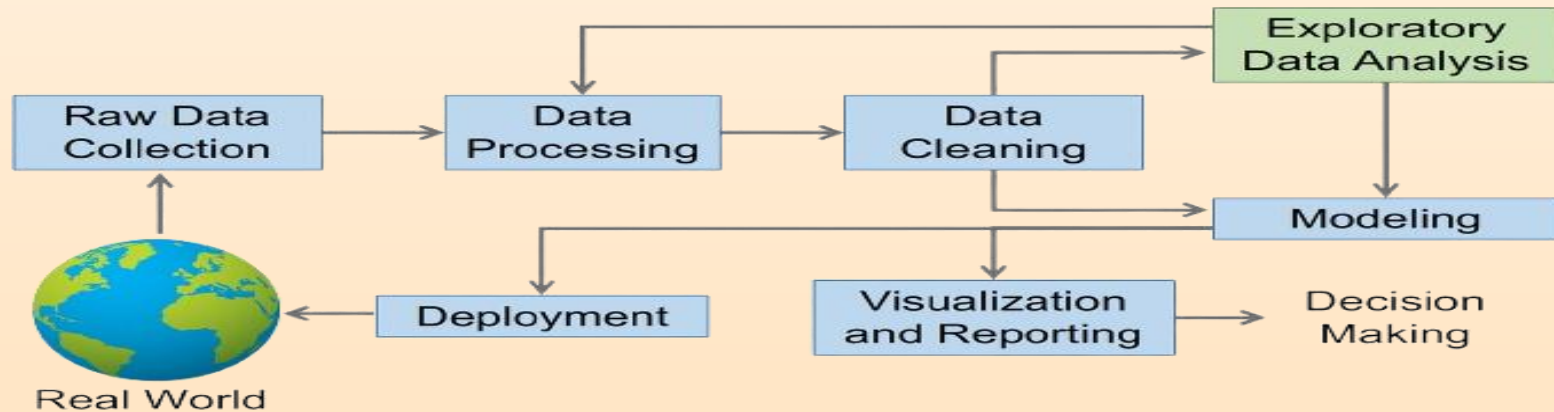
# Result / Outcomes

## Exploratory Data Analysis (EDA)

EDA was performed to uncover trends, patterns, and insights from the data. The following steps were taken:

**Data Visualization:** Created visualizations using matplotlib and seaborn to illustrate key findings.

**Descriptive Statistics:** Calculated summary statistics to understand the data distribution and central tendencies.





## Key Findings

### Customer Insights

By analyzing customer demographics, we identified potential customers across various segments:

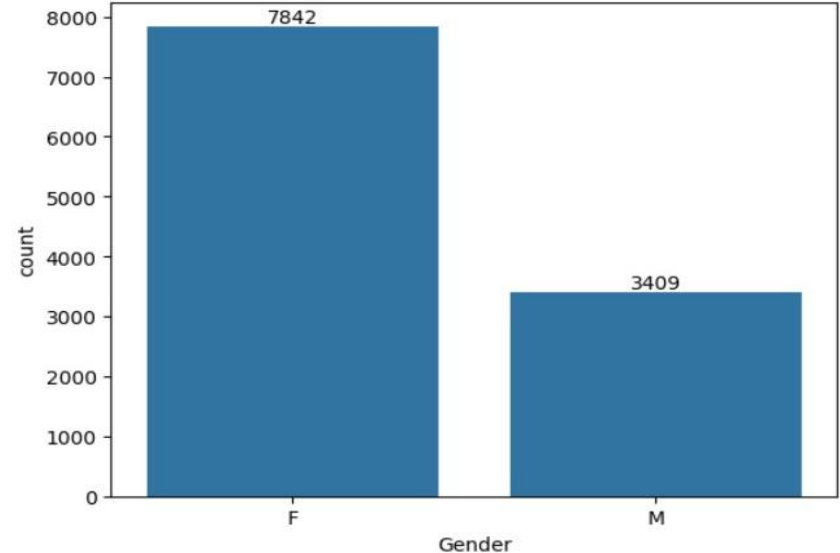
- Gender:** Examined gender-wise distribution to tailor marketing strategies.
- Age Groups:** Identified age groups with the highest purchasing potential.
- States:** Determined which states had the highest number of potential customers.

### Sales Insights

- Best-Selling Products:** Determined which specific products had the highest sales.

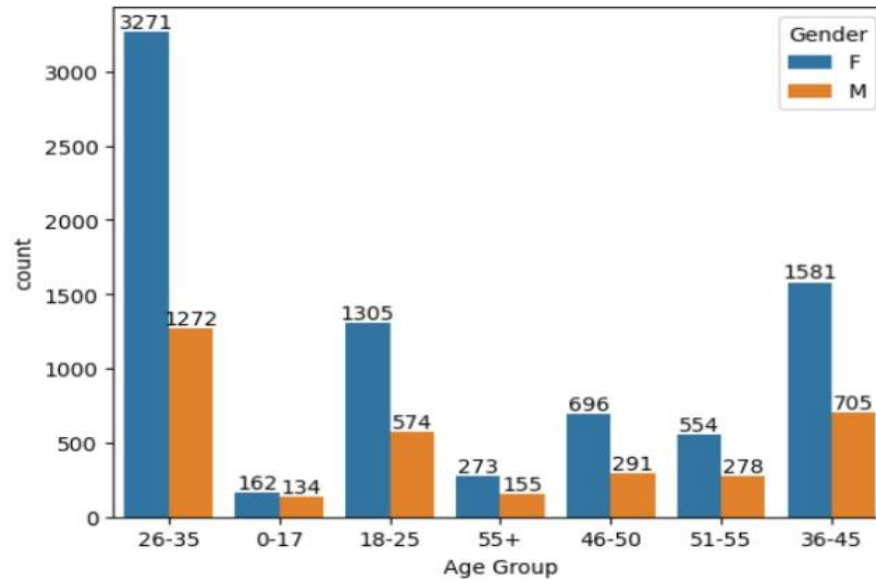
## Gender-wise Customer Distribution

```
: # plotting a bar chart for Gender and it's count  
  
ax = sns.countplot(x = 'Gender', data = df)  
  
for bars in ax.containers:  
    ax.bar_label(bars)
```



## Age Group Analysis

```
[17]: ax = sns.countplot(data = df, x = 'Age Group', hue = 'Gender')  
  
for bars in ax.containers:  
    ax.bar_label(bars)
```



*From above graphs we can see that most of the buyers are of age group between 26-35 yrs female*

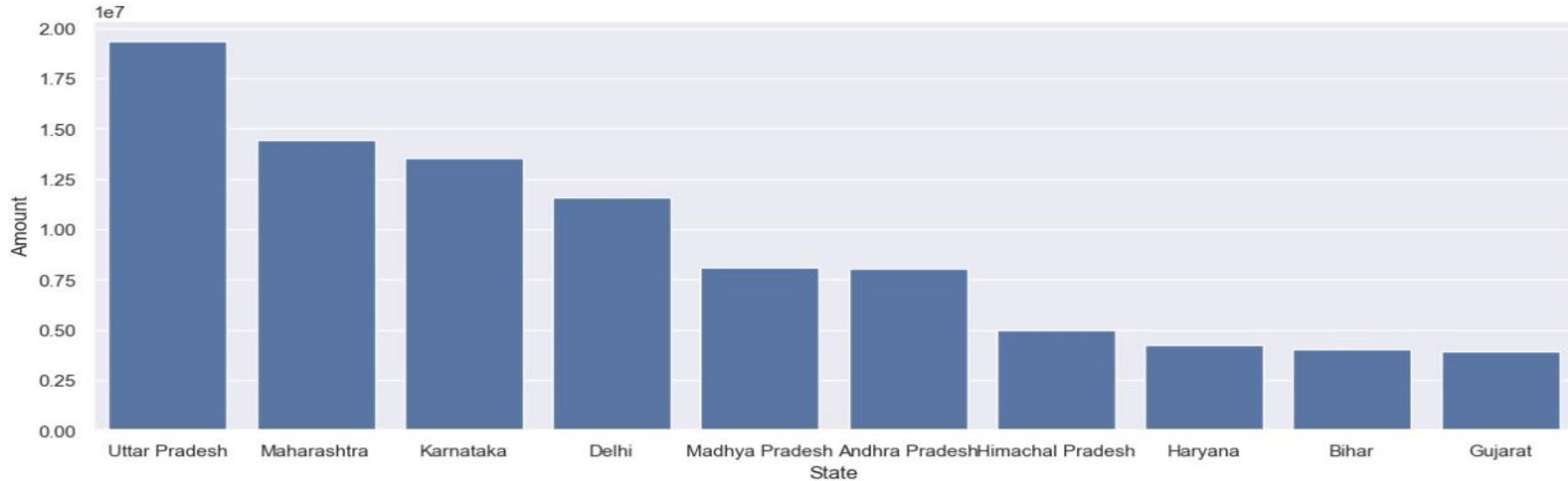
## Customer Distribution by State

```
In [ ]: # total amount/sales from top 10 states

sales_state = df.groupby(['State'], as_index=False)['Amount'].sum().sort_values(by='Amount', ascending=False).head(10)

sns.set(rc={'figure.figsize':(15,5)})
sns.barplot(data = sales_state, x = 'State',y= 'Amount')

Out[ ]: <Axes: xlabel='State', ylabel='Amount'>
```



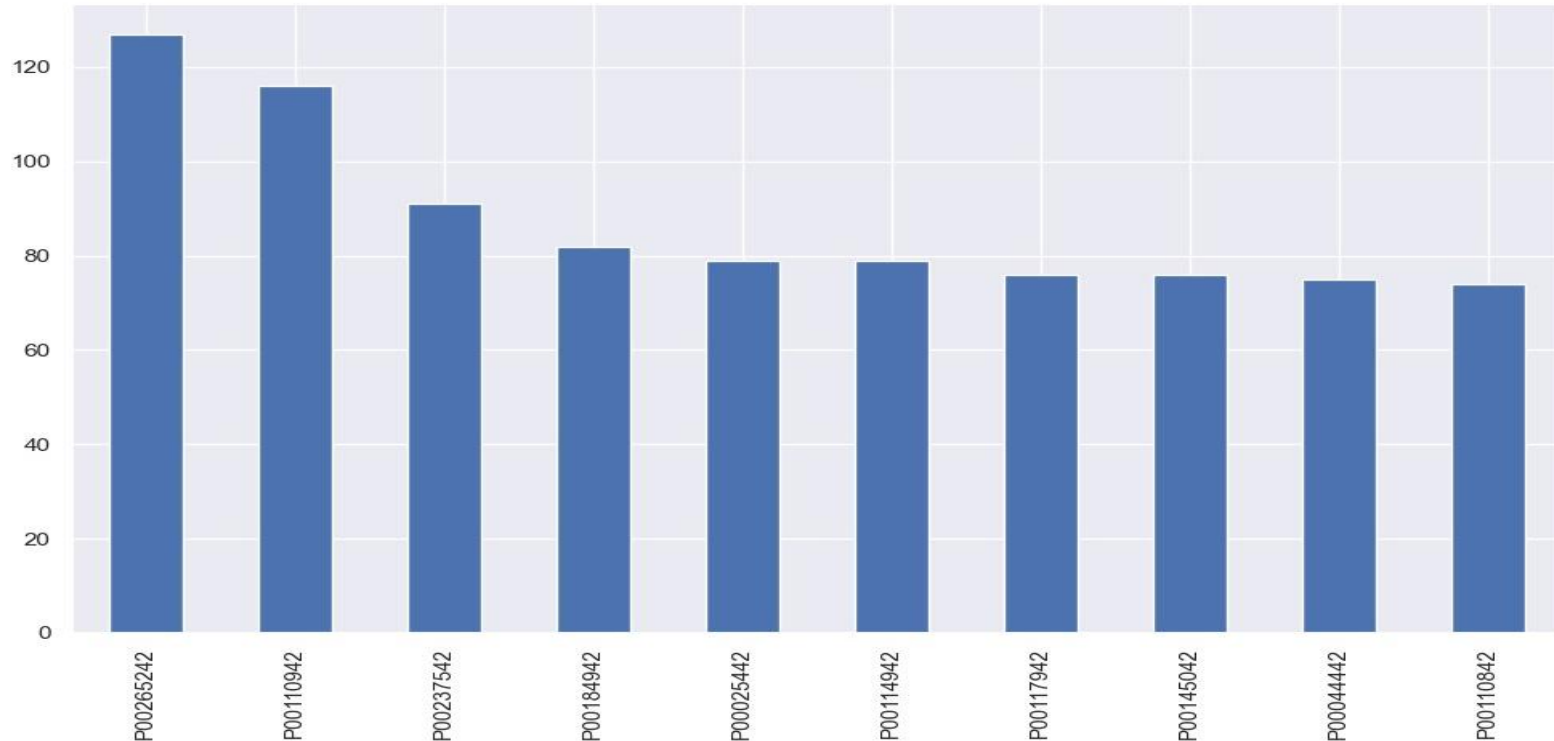
From above graphs we can see that most of the orders & total sales/amount are from Uttar Pradesh, Maharashtra and Karnataka respectively

# Best-Selling Products Count Analysis

```
]# top 10 most sold products (same thing as above)

fig1, ax1 = plt.subplots(figsize=(12,7))
df.groupby('Product_ID')['Orders'].sum().nlargest(10).sort_values(ascending=False).plot(kind='bar')

]: <Axes: xlabel='Product_ID'>
```



# Conclusion

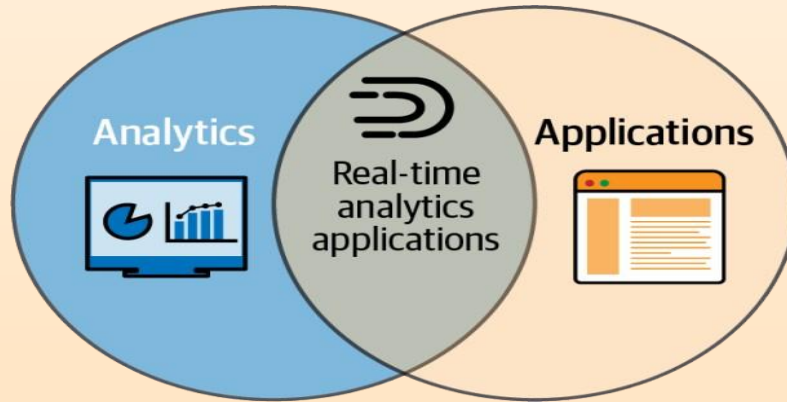
The Diwali Sale Data Analysis project successfully provided actionable insights to improve customer experience and sales. By identifying potential customers and popular products, businesses can tailor their marketing strategies and optimize inventory management to meet demand effectively

- **Women age group 26-35 yrs from UP, Maharastra and Karnataka are more likely to buy products.**



# Future Perspective

- **Predictive Analysis:** Implement machine learning models to predict future sales trends.
- **Customer Segmentation:** Perform deeper customer segmentation for more targeted marketing campaigns.
- **Real-Time Data Analysis:** Develop real-time dashboards for ongoing monitoring and decision-making.



**THANK  
YOU!**