

DIWALI SALES ANALYSIS

EXPLORING CONSUMER BEHAVIOR DURING FESTIVAL SEASON

Prepared By:

Aditya Kumar Pandey

MTech (AI & DS), 240201001

Dept. of CSE, IIIT Bhagalpur

Submitted To:

Dr. Avantika

Assistant Professor

Dept. of CSE, IIIT Bhagalpur



CASE STUDY

INTRODUCTION

- Diwali Sales Analysis – Exploring Consumer Behavior During Festival Season
- Objective: To analyze consumer purchase patterns during Diwali using a retail sales dataset.
- Goal: Derive actionable business insights through data visualization and statistical exploration.



DATASET INFORMATION

- Dataset Type: Structured CSV file
- Source: Kaggle
- Total Records: ~11,000+ rows (varies based on cleaning)
- Purpose: Captures transactional and customer demographic data during Diwali season
- Features:
- User_ID: Unique identifier for each customer
- Cust_name: Customer name
- Product_ID: Unique product identifier
- Gender: Male or Female
- Age Group: Age grouped into ranges (e.g., 18-25, 26-35)
- Age: Exact age in years
- Marital_Status: Indicates if customer is married
- State: State where purchase was made
- Zone: Geographical zone (North, South, etc.)
- Occupation: Profession of customer
- Product_Category: Category under which product falls
- Orders: Number of units ordered
- Amount: Total purchase amount
- unnamed: NA

DATA PREPROCESSING

- Dataset Loaded from CSV
- Initial Exploration
- Converted Data Types
- Checked Missing Values
- Post-Cleaning Summary



DATA PREPROCESSING

Pre-Cleaning Data Summary

```
data columns (total 15 columns):
#  Column      Non-Null Count  Dtype
--  --
0  User_ID      11251 non-null  int64
1  Cust_name    11251 non-null  object
2  Product_ID   11251 non-null  object
3  Gender       11251 non-null  object
4  Age Group    11251 non-null  object
5  Age          11251 non-null  int64
6  Marital_Status 11251 non-null  int64
7  State        11251 non-null  object
8  Zone         11251 non-null  object
9  Occupation   11251 non-null  object
10 Product_Category 11251 non-null  object
11 Orders      11251 non-null  int64
12 Amount      11239 non-null  float64
13 Status       0 non-null      float64
14 unnamed1     0 non-null      float64
dtypes: float64(3), int64(4), object(8)
memory usage: 1.3+ MB
done
```

	User_ID	Cust_name	Product_ID	Gender	Age Group	Age	Marital_Status	State	Zone	Occupation	Product_Category	Orders	Amount	Status	unnamed1
0	1002903	Sanskriti	P00125942	F	26-35	28	0	Maharashtra	Western	Healthcare	Auto	1	23952.0	NaN	NaN
1	1000732	Kartik	P00110942	F	26-35	35	1	Andhra Pradesh	Southern	Govt	Auto	3	23934.0	NaN	NaN
2	1001990	Bindu	P00118542	F	26-35	35	1	Uttar Pradesh	Central	Automobile	Auto	3	23924.0	NaN	NaN
3	1001425	Sudevi	P00237842	M	0-17	16	0	Karnataka	Southern	Construction	Auto	2	23912.0	NaN	NaN
4	1000588	Joni	P00057942	M	26-35	28	1	Gujarat	Western	Food Processing	Auto	2	23877.0	NaN	NaN



DATA PREPROCESSING

INITIAL DATA OVERVIEW

- We have 11,251 rows and 15 columns.
- `df['Amount'] = df['Amount'].astype('int')` : This converts the Amount column from float to integer
- Amount: 12 nulls
- Status: all nulls
- unnamed1: all nulls
- These two columns (Status, unnamed1) are completely empty, and Amount has 12 missing entries.
- `df.drop(['Status', 'unnamed1'], axis=1, inplace=True)`
- This line removes the two completely empty columns — they don't add any value to the analysis.

Null Values	count
User_ID	0
Cust_ID	0
Product_ID	0
Gender	0
Age Group	0
Marital Status	0
State	0
Zone	0
Occupation	0
Product_Category	0
Orders	0
Amount	12

DATA PREPROCESSING

Post Cleaning Data Summary

Index: 11239 entries, 0 to 11250

Data columns (total 13 columns):

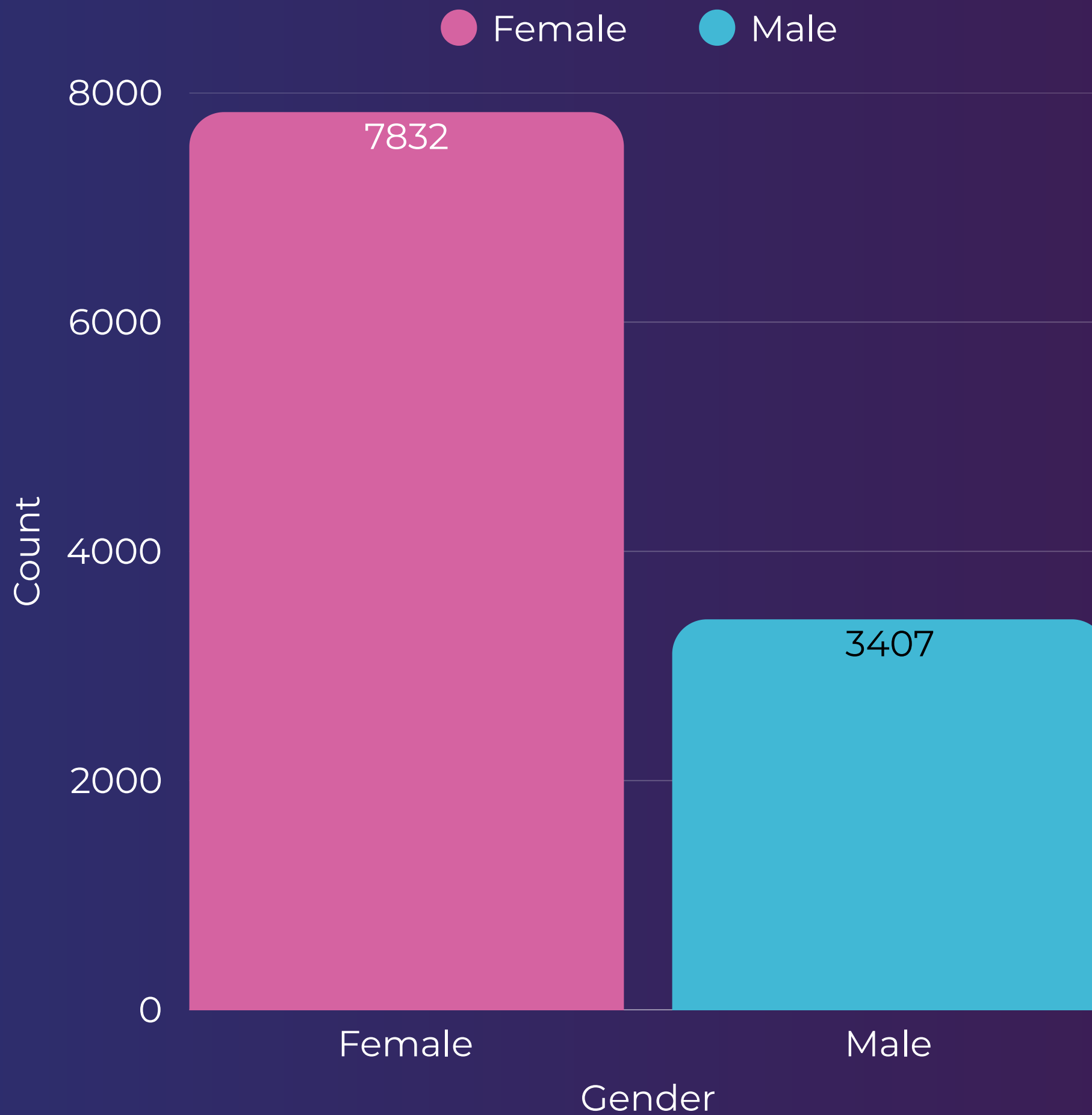
#	Column	Non-Null Count	Dtype
0	User_ID	11239 non-null	int64
1	Cust_name	11239 non-null	object
2	Product_ID	11239 non-null	object
3	Gender	11239 non-null	object
4	Age Group	11239 non-null	object
5	Age	11239 non-null	int64
6	Marital_Status	11239 non-null	int64
7	State	11239 non-null	object
8	Zone	11239 non-null	object
9	Occupation	11239 non-null	object
10	Product_Category	11239 non-null	object
11	Orders	11239 non-null	int64
12	Amount	11239 non-null	int64

dtypes: int64(5), object(8)

memory usage: 1.2+ MB

None

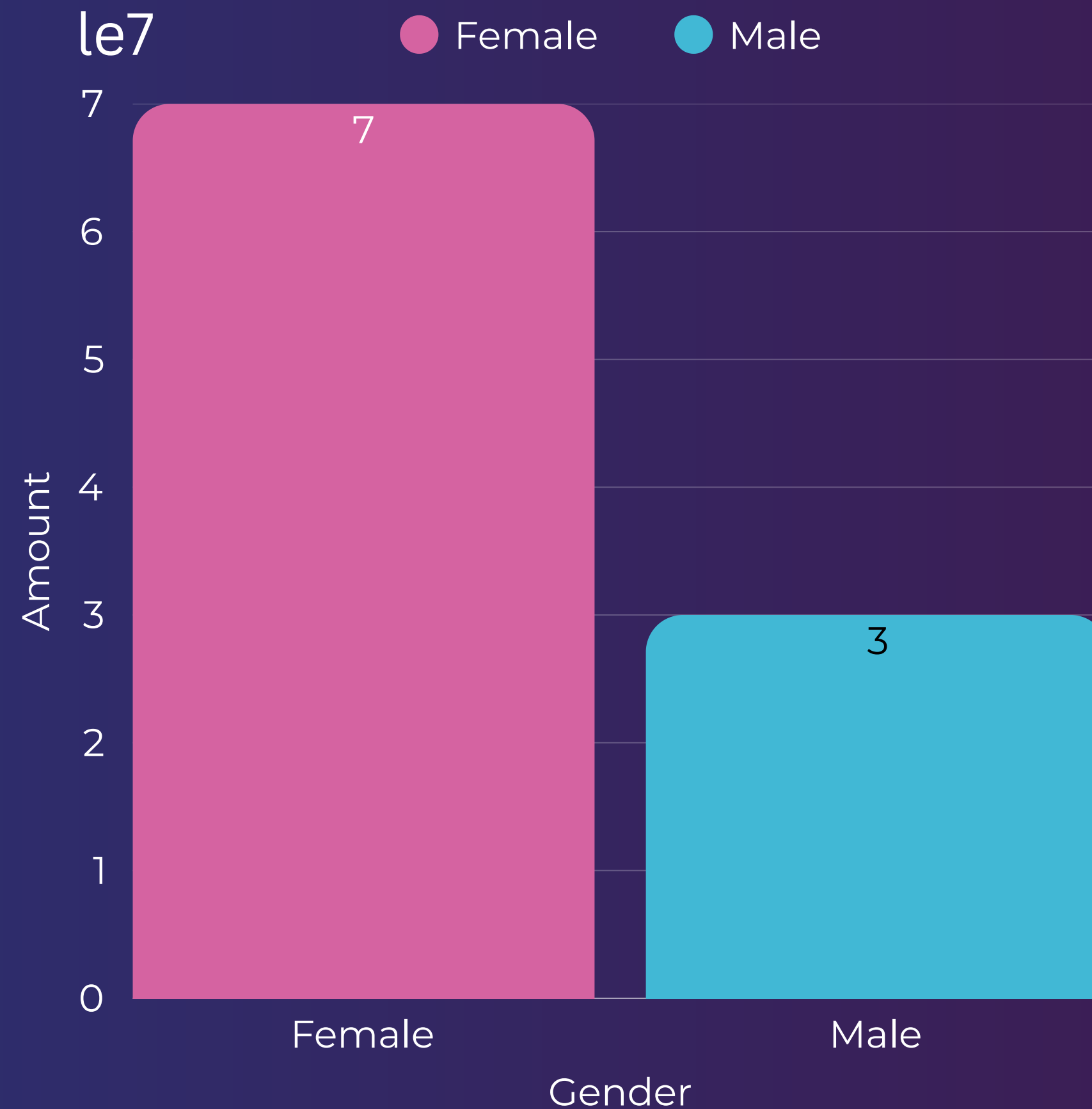
	User_ID	Cust_name	Product_ID	Gender	Age Group	Age	Marital_Status	State	Zone	Occupation	Product_Category	Orders	Amount
0	1002903	Sanskriti	P00125942	F	26-35	28	0	Maharashtra	Western	Healthcare	Auto	1	23952
1	1000732	Kartik	P00110942	F	26-35	35	1	Andhra Pradesh	Southern	Govt	Auto	3	23934
2	1001990	Bindu	P00118542	F	26-35	35	1	Uttar Pradesh	Central	Automobile	Auto	3	23924
3	1001425	Sudevi	P00237842	M	0-17	16	0	Karnataka	Southern	Construction	Auto	2	23912
4	1000588	Joni	P00057942	M	26-35	28	1	Gujarat	Western	Food Processing	Auto	2	23877



EXPLORATORY DATA ANALYSIS

GENDER DISTRIBUTION

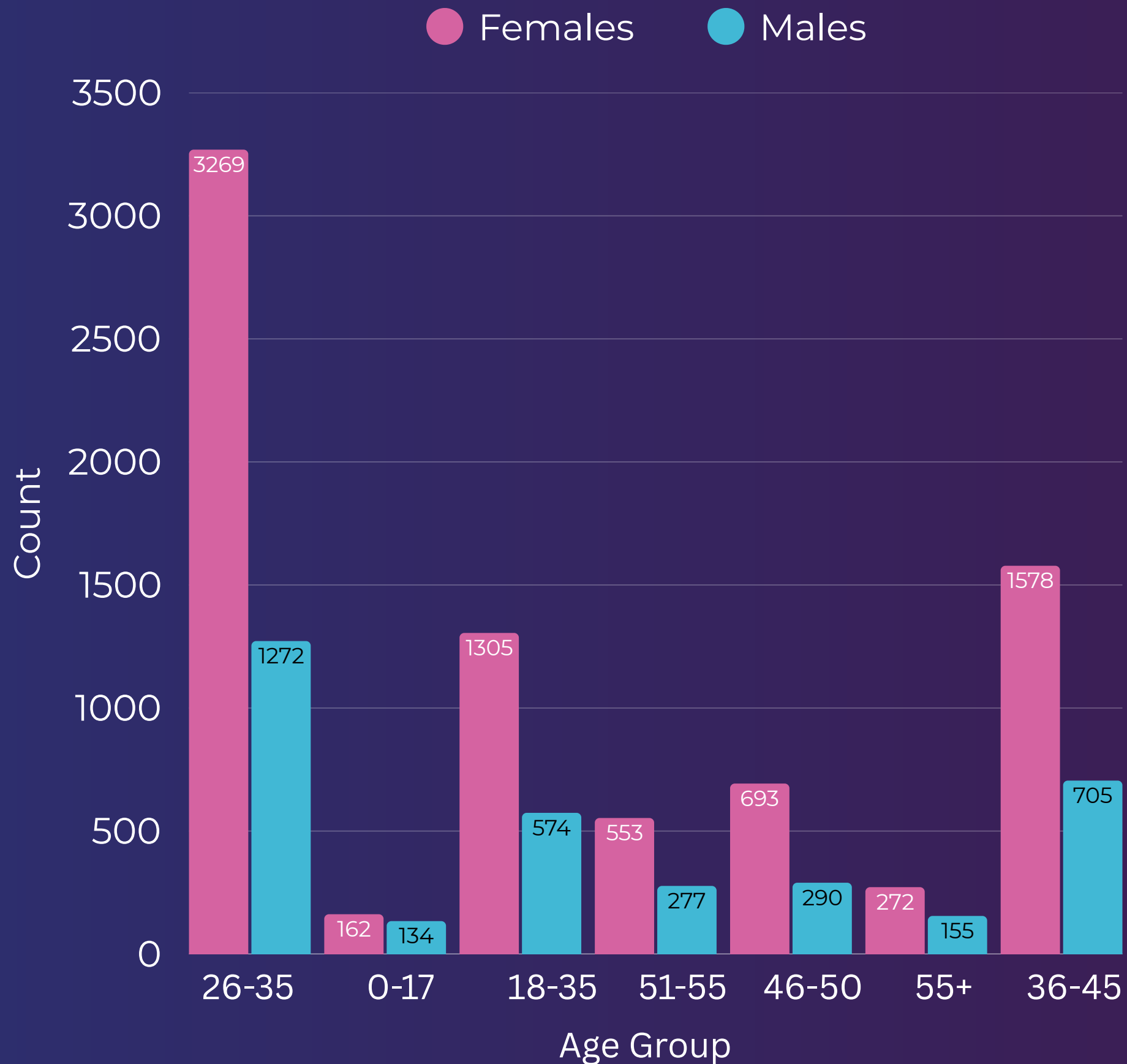
- Counted male vs female customers using a bar chart.
- Majority of shoppers were Female.



EXPLORATORY DATA ANALYSIS

GENDER VS TOTAL AMOUNT SPENT

- Sales data is grouped by gender to analyze spending patterns.
- Total amount spent by females (F) is significantly higher than males (M).
- Suggests that female customers were more active shoppers during Diwali..
- $1e7$ means the values are in millions.



EXPLORATORY DATA ANALYSIS

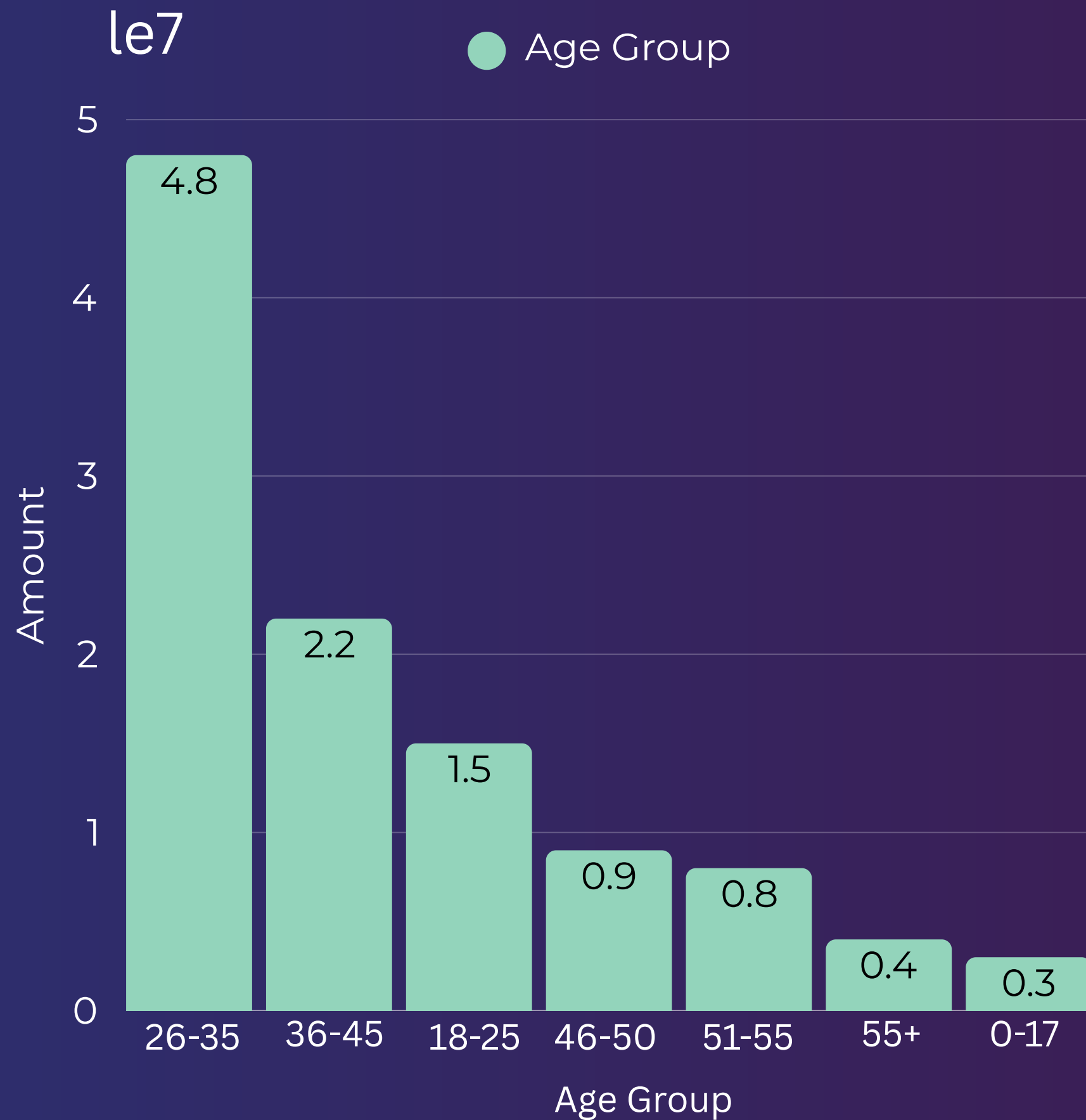
AGE GROUP DISTRIBUTION BY GENDER

This chart shows:

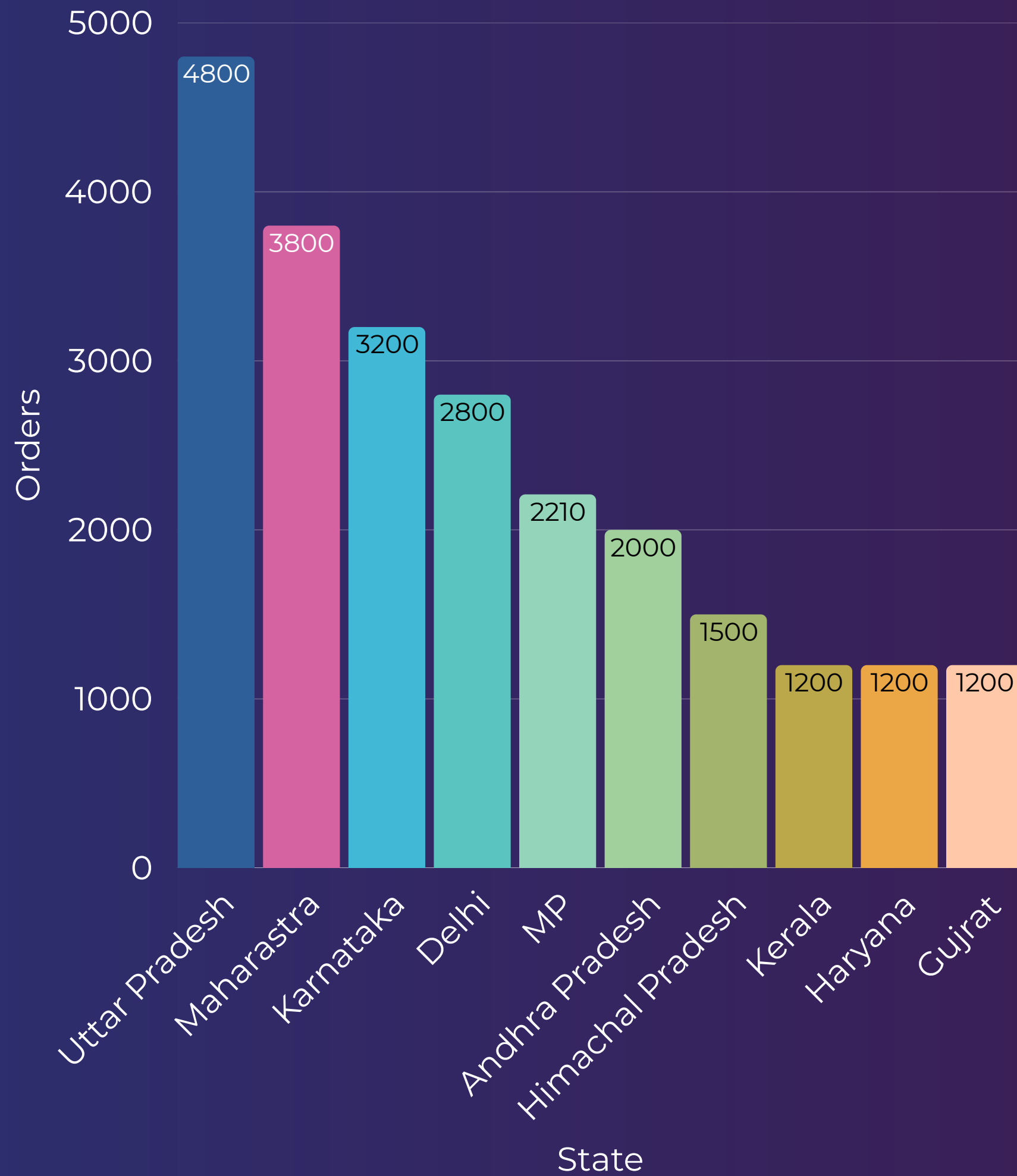
- Which age groups are most active.
- Gender distribution in each age group.
- Most customers might be in the 26-35 age group.
- Both males and females are well represented, but females dominate in some groups.

EXPLORATORY DATA ANALYSIS

TOTAL AMOUNT SPENT BY AGE GROUP



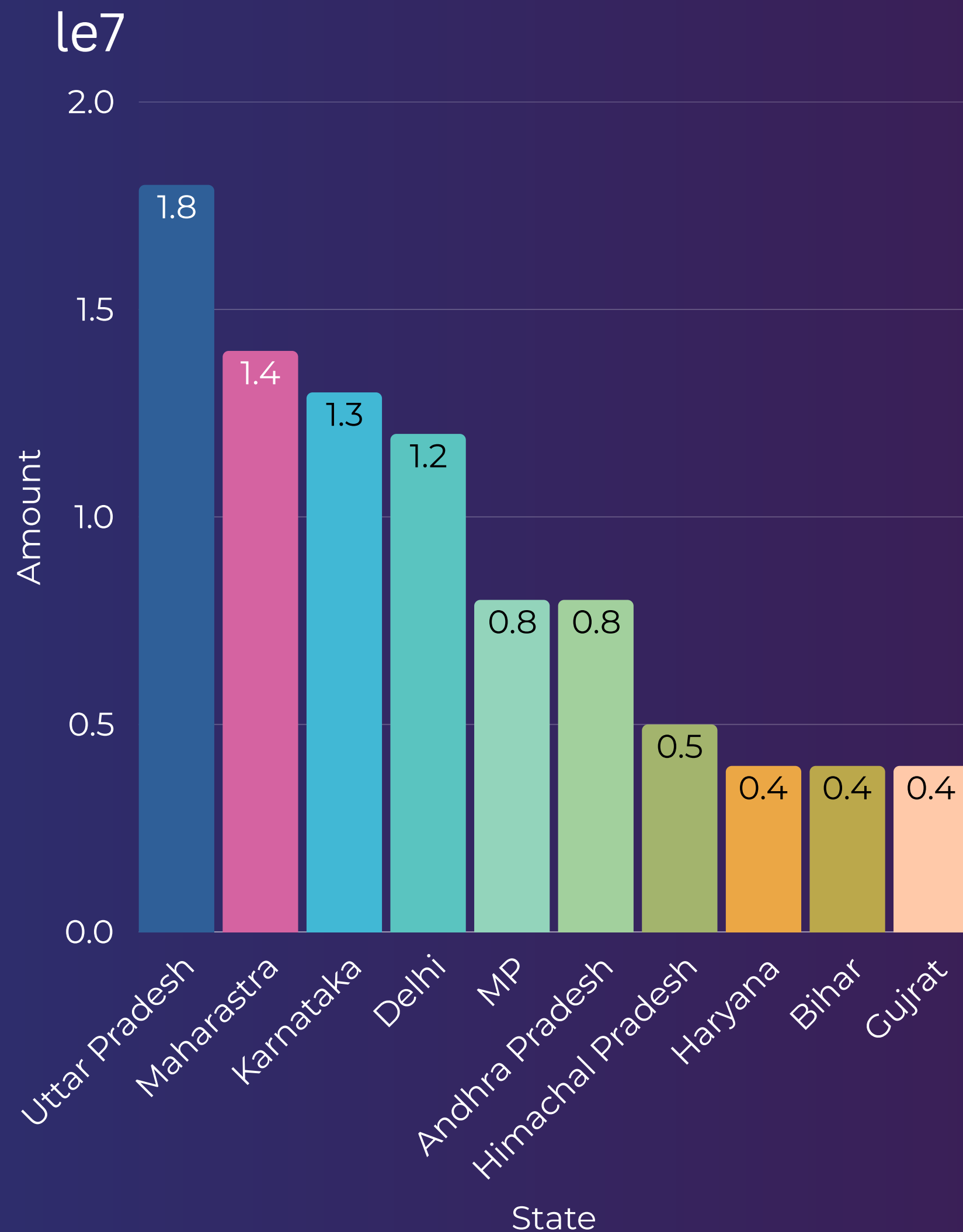
- Shows spending power of different age groups.
- Likely, the 26-35 or 36-45 groups spend the most — prime working age.
- Helps target marketing for high-spending age demographics.



EXPLORATORY DATA ANALYSIS

TOP 10 STATES BY NUMBER OF ORDERS

- Visualizes the states with the highest number of purchases.
- Likely leaders: Uttar Pradesh, Maharashtra, Karnataka, etc.
- Useful for identifying geographic hot spots in terms of customer activity.



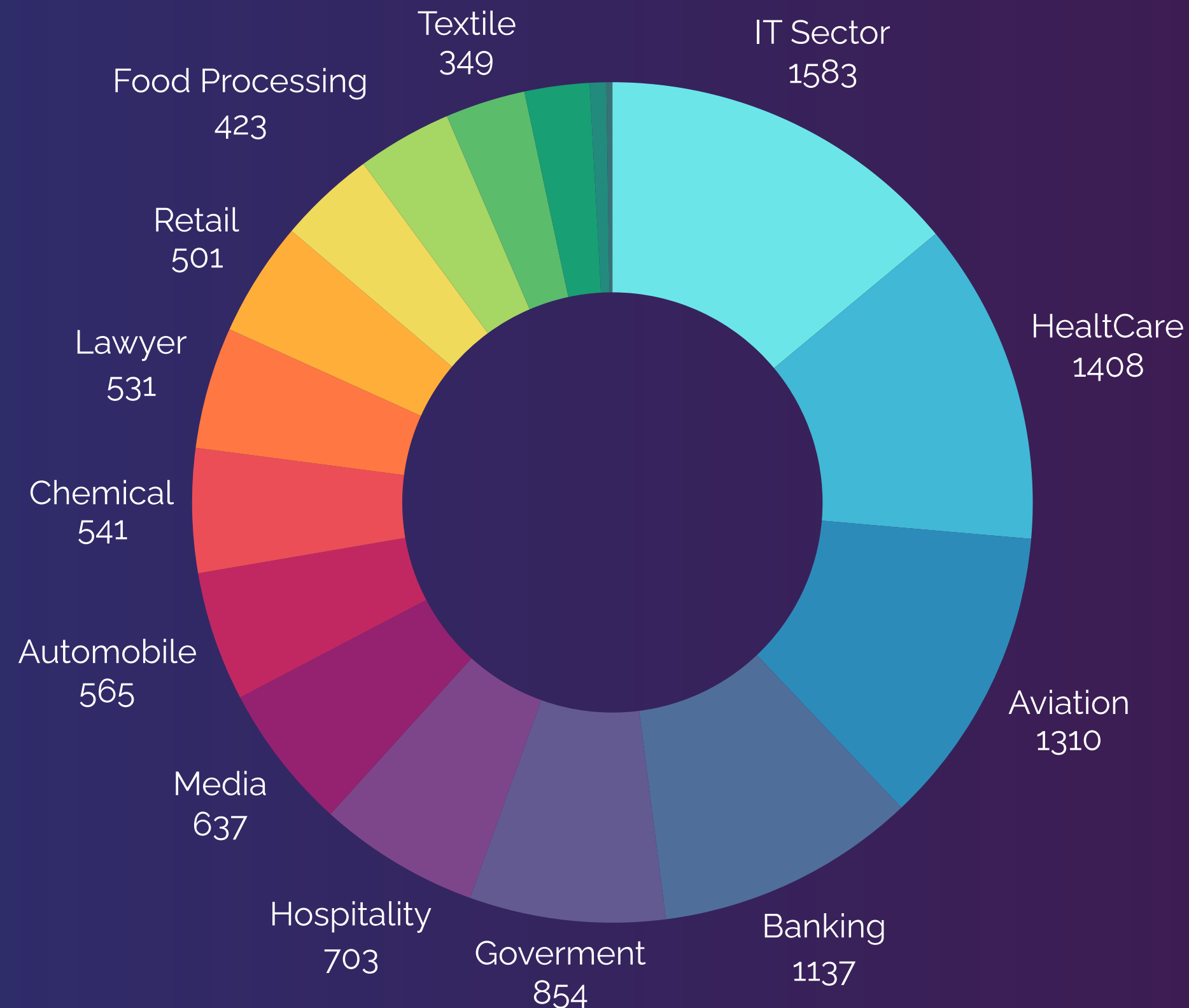
EXPLORATORY DATA ANALYSIS

TOP 10 STATES BY SALES AMOUNT

- Reveals top contributing states in revenue.
- Some states might have fewer orders but higher value per order.
- Helps in prioritizing marketing, logistics, and sales efforts geographically.

EXPLORATORY DATA ANALYSIS

OCCUPATION-WISE DISTRIBUTION OF DIWALI SHOPPERS



Top Spenders:

- IT Sector (1583) and Healthcare (1408) lead due to high income and tech adoption.
- Aviation (1310) and Banking (1137) also show strong festive buying power.

Moderate Participation:

- Govt (854), Hospitality (703), and Media (637) balance tradition with modern spending.

Low Engagement:

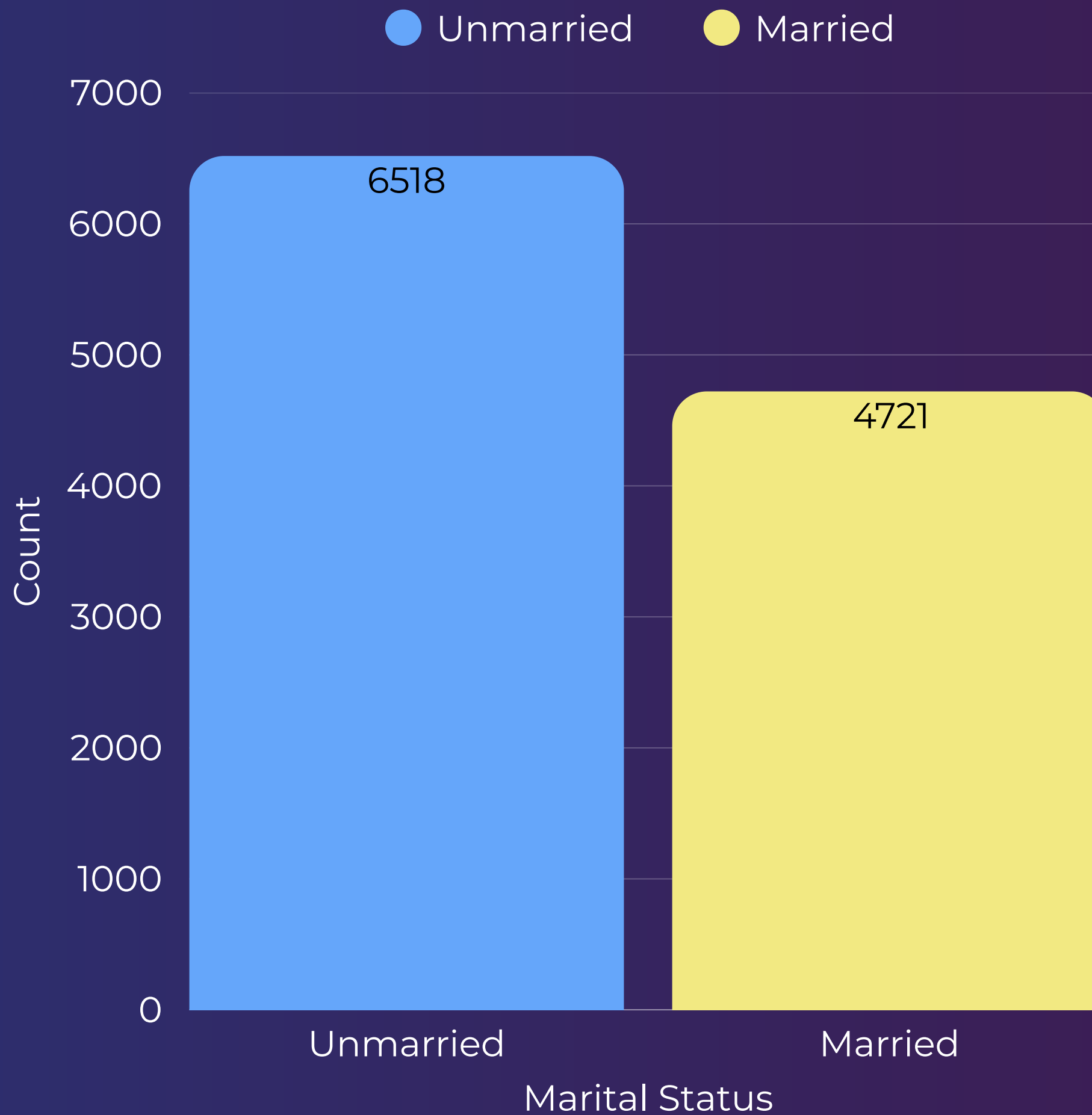
- Agriculture (283), Textile (349), and Construction (414) may face income or access constraints.

Growth Potential:

- Automobile, Retail, Food Processing, and Law show scope for targeted promotions.

Marketing Tip:

- Focus premium campaigns on top sectors, and offer EMI deals or festive combos for low-engagement groups.

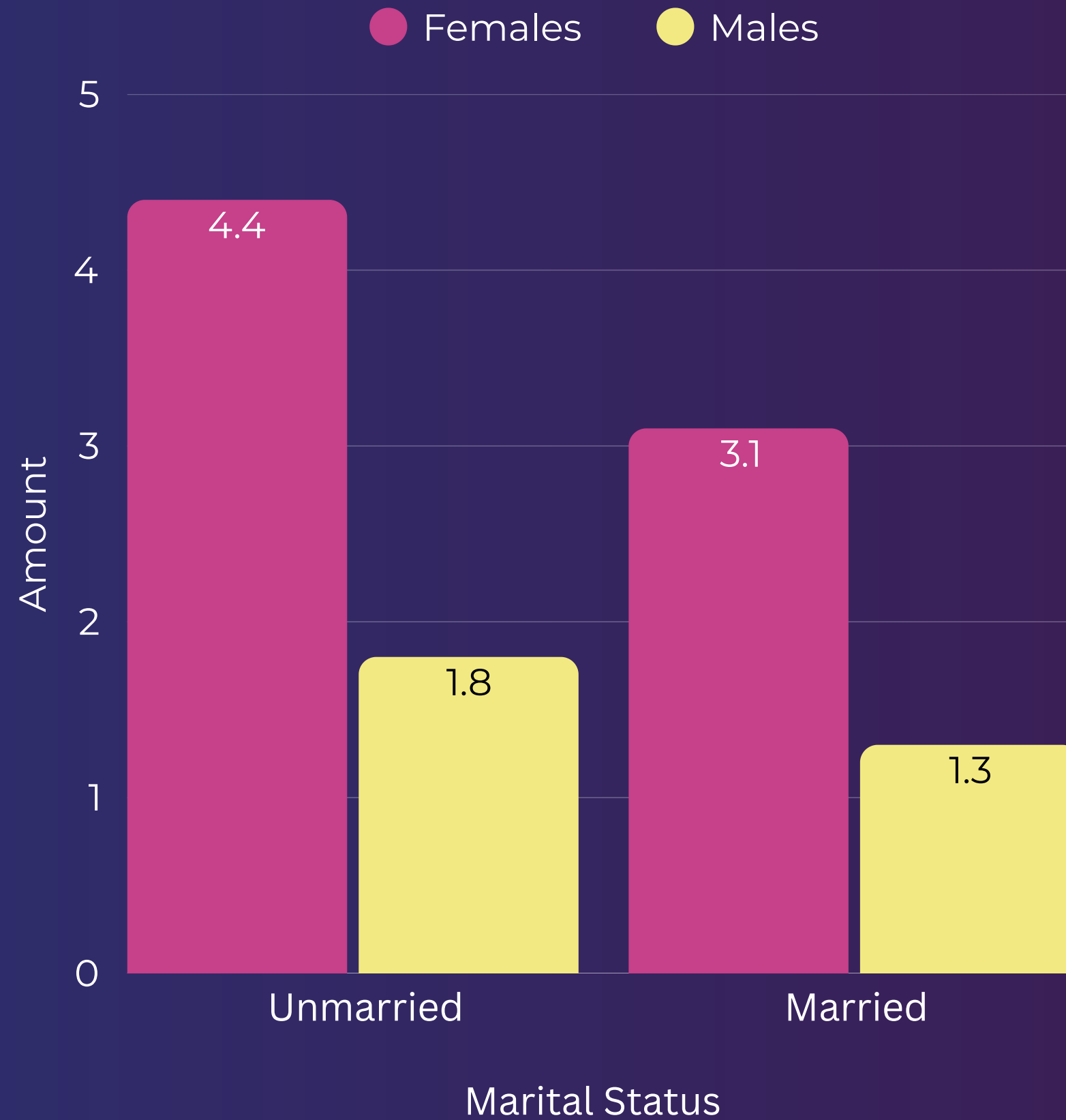


EXPLORATORY DATA ANALYSIS

MARITAL STATUS COUNT

- This chart shows the number of customers based on marital status.
- The dataset has more unmarried customers (6518) compared to married ones (4721).

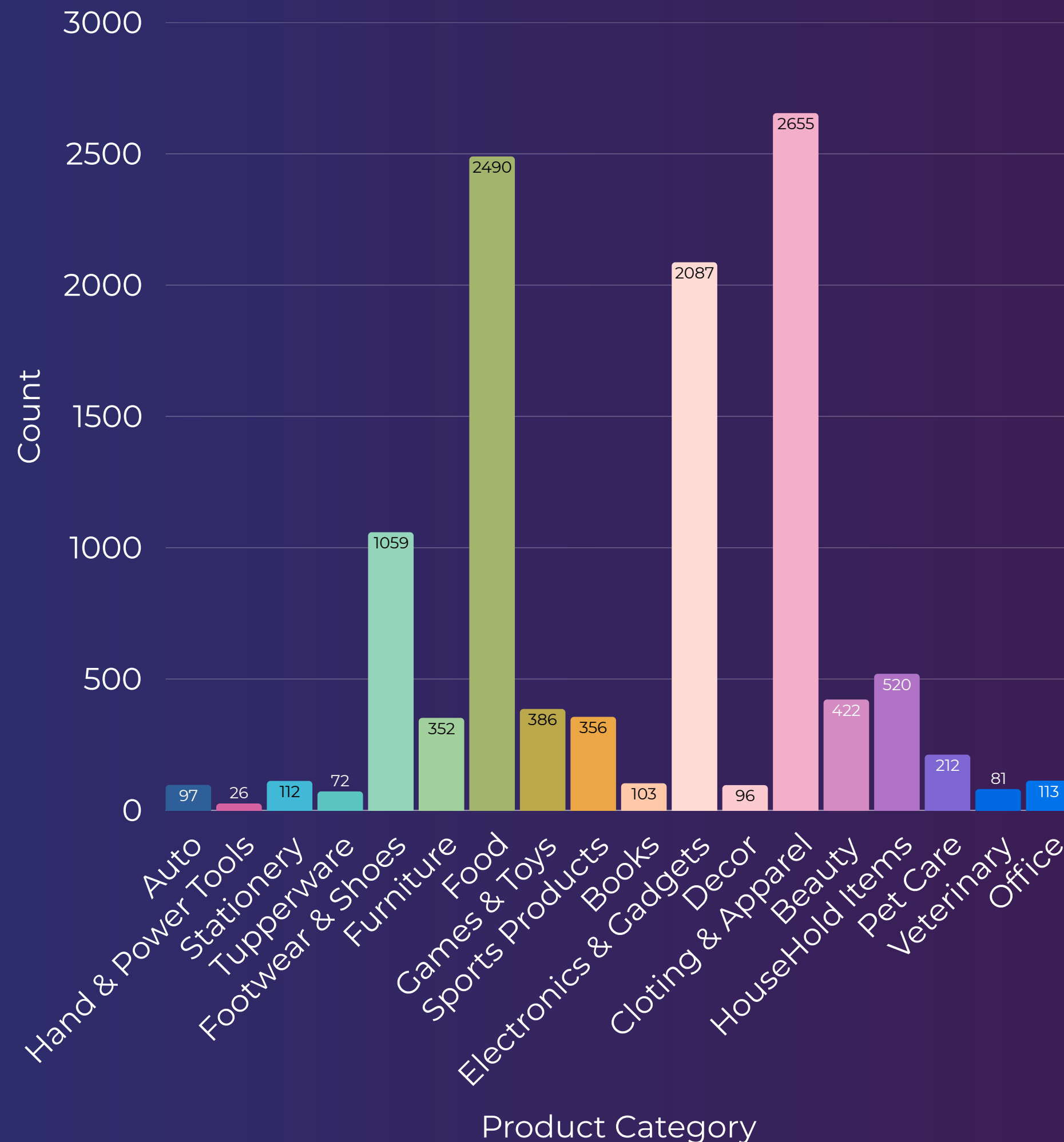
le7



EXPLORATORY DATA ANALYSIS

MARITAL STATUS VS AMOUNT

- Unmarried Females (Marital Status = Unmarried, Gender=F) spent the most, with the bar reaching over ₹4 crores.
- Unmarried Males also spent significantly, but less than females in the same group.
- Among the married group (Marital Status = Married), females again spent more than males, but both spent less compared to the unmarried group.



EXPLORATORY DATA ANALYSIS

PRODUCT CATEGORY DISTRIBUTION

- Clothing & Apparel is the most popular category with 2,655 sales, followed by:
- Food – 2,490 sales
- Electronics & Gadgets – 2,087 sales

Moderate performance seen in:

- Footwear & Shoes – 1,059 sales
- Household items – 520 sales
- Beauty – 422 sales
- Games & Toys – 386 sales
- Sports Products – 356 sales

Niche categories with lower sales include:

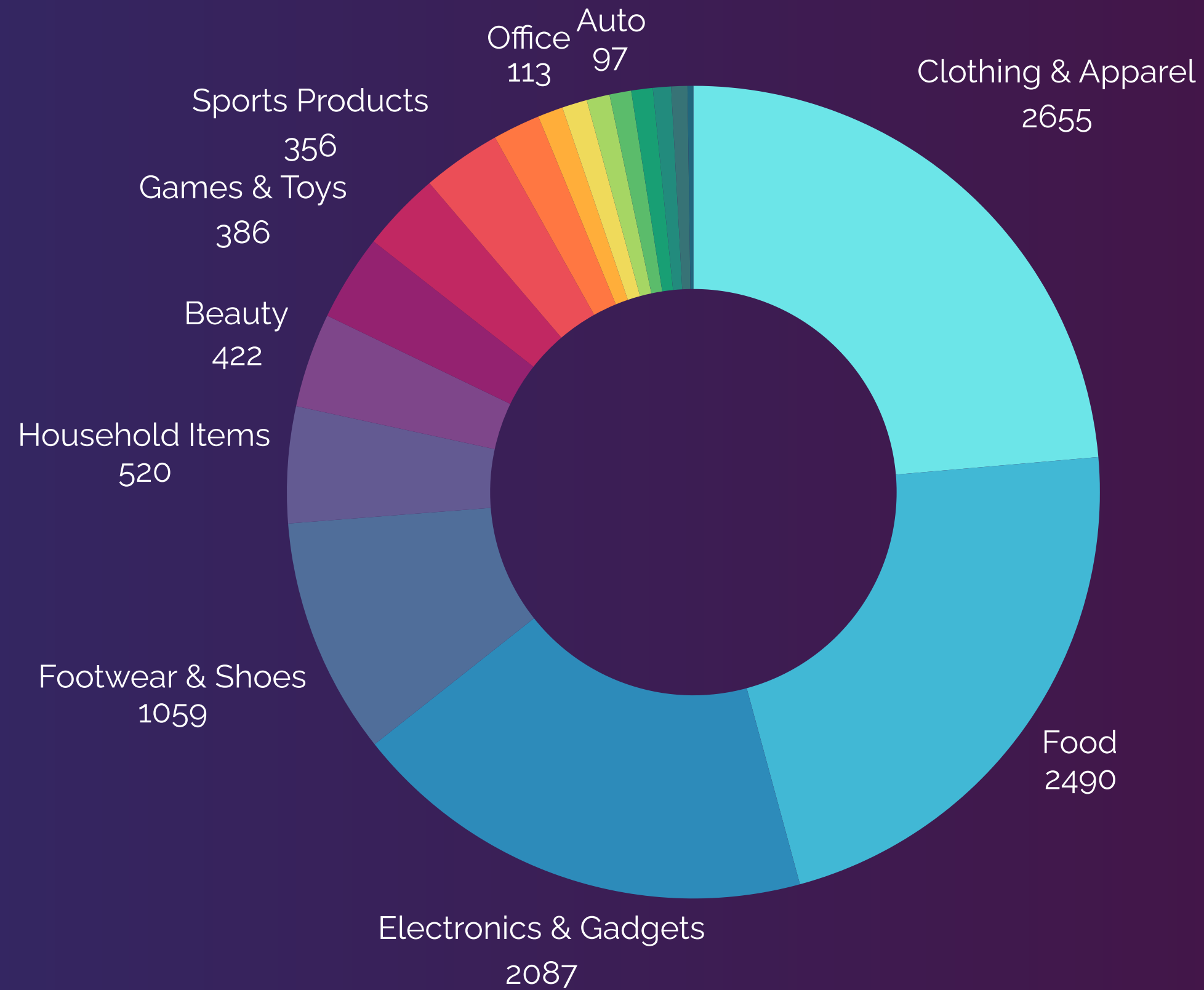
- Furniture – 352
- Pet Care – 212
- Office Supplies – 113
- Stationery – 112

Least engaged categories:

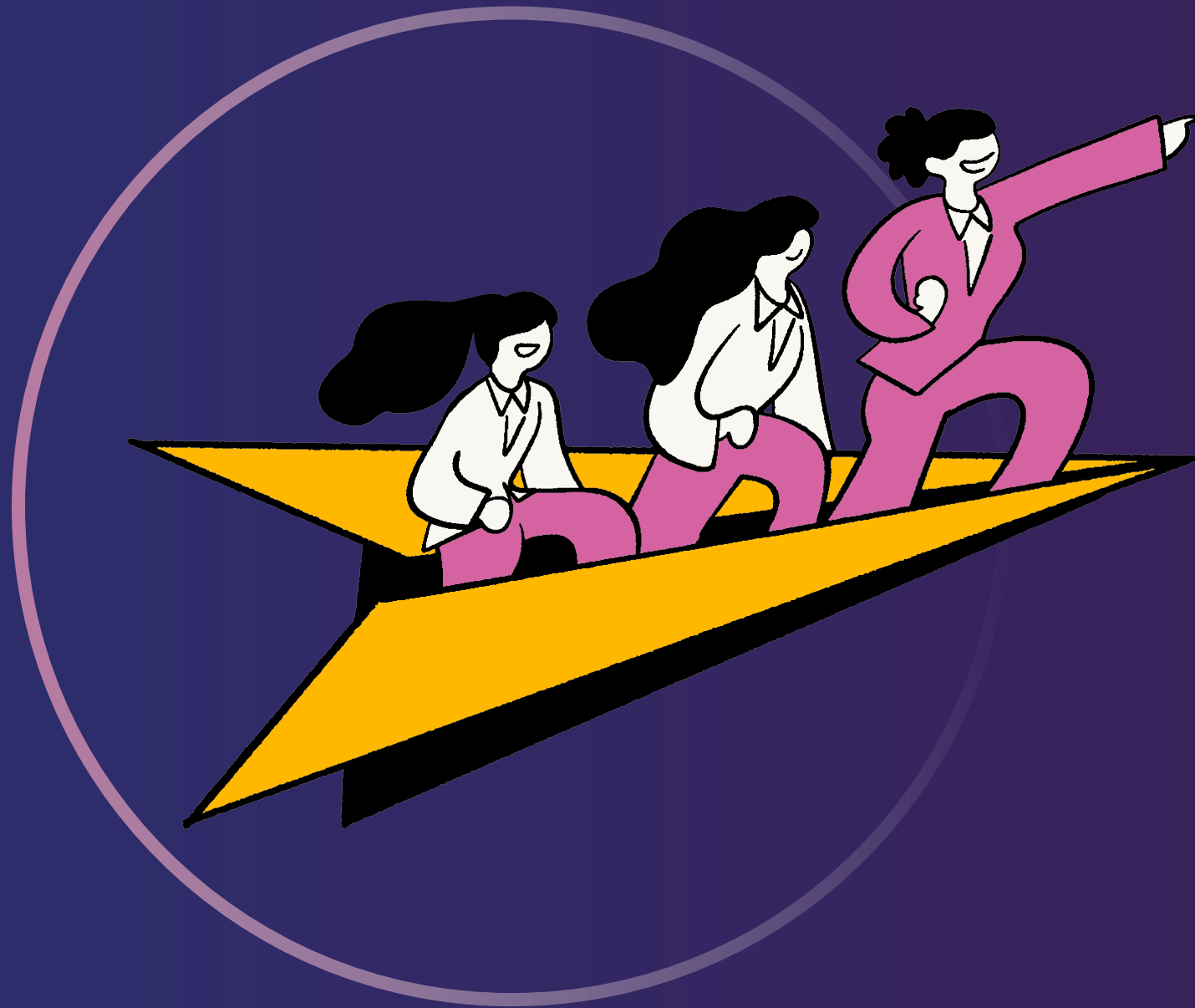
- Books – 103
- Auto – 97
- Decor – 96
- Veterinary – 81
- Tupperware – 72
- Hand & Power Tools – 26

EXPLORATORY DATA ANALYSIS

PRODUCT CATEGORY DISTRIBUTION - DONUT CHART



REAL-WORLD APPLICATIONS



Customer Segmentation

- Identify high-value customers (e.g., IT, Healthcare) to tailor premium offers.

Targeted Marketing Campaigns

- Use spending patterns by occupation and region to personalize festive or seasonal promotions.

Product Performance Insights

- Understand which product categories (e.g., Auto, Electronics) drive revenue to optimize inventory.

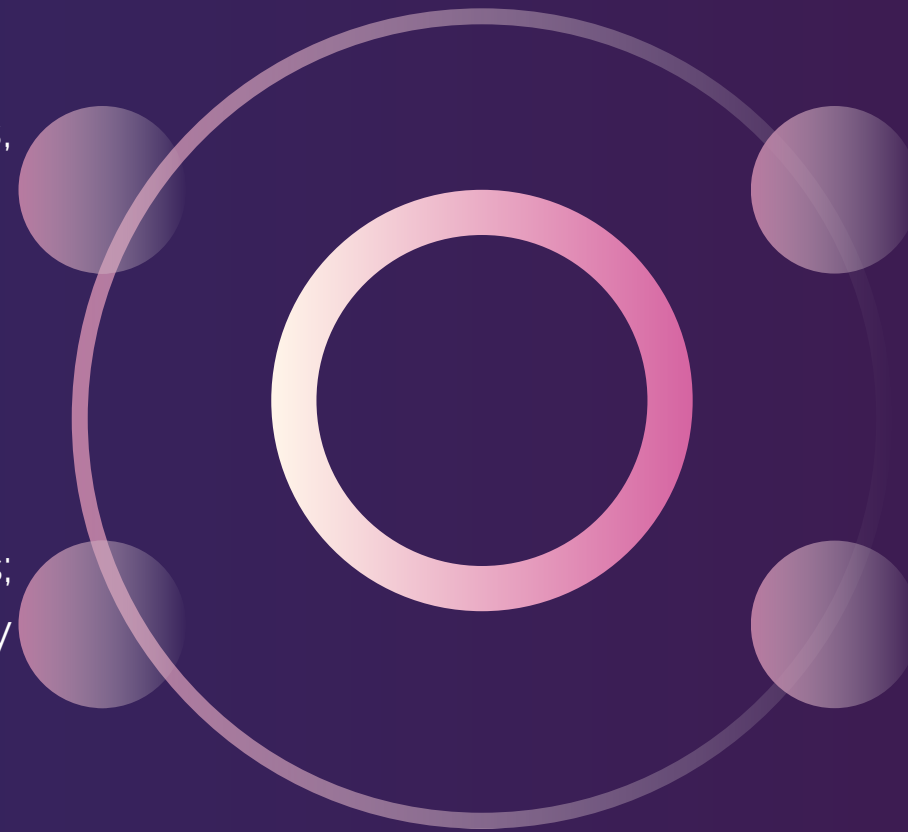
Sales Forecasting

- Predict demand based on past purchase behavior for efficient stock management.

BUSINESS RECOMMENDATIONS

Launch campaigns specifically for Govt Employees, Hospitality, or Students with relevant perks.

Younger customers (18-35) may prefer tech deals; older ones may go for healthcare or utility products.



Prioritize IT, Healthcare, Banking, and Aviation with loyalty programs or premium bundles.

Design budget-friendly bundles, festive sales, or EMI options for Construction, Agriculture, etc.



THANK YOU!

DATA ANALYSIS IS KEY TO BUSINESS
GROWTH AND SUCCESS!



ADITYA KUMAR PANDEY
MTECH (AI & DS), IIIT BHAGALPUR
LINKEDIN