

Diwali Sales Analysis Project Report

Objective

This project aims to analyze Diwali sales data to uncover insights about customer purchasing patterns and sales trends. The analysis can help identify key factors that influence customer behavior during the Diwali season, allowing retailers to optimize marketing and sales strategies.

Data Overview

The dataset contains customer and sales data, including information such as:

- Customer demographics (age, gender, marital status)
- Product details (product category and type)
- Purchase information (amount spent, purchase date, and city category)

Steps and Analysis

1. Data Import and Preprocessing

- Imported necessary libraries for data analysis and visualization.
- Loaded the Diwali sales dataset, checking for any missing or duplicate values.
- Performed initial data cleaning to ensure data consistency and accuracy.

```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
```

```
df=pd.read_excel("C:\\Users\\hp\\Downloads\\Diwali Sales Data (3) 01.xlsx")
```

df

	User_ID	Cust_name	Product_ID	Gender	Age Group	Age	Marital_Status	State	Zone	Occupation	Product_Category	Orders	Amount	Status	unn
0	1002903	Sanskriti	P00125942	F	26-35	28	0	Maharashtra	Western	Healthcare	Auto	1	23952.0	NaN	
1	1000732	Kartik	P00110942	F	26-35	35	1	Andhra Pradesh	Southern	Govt	Auto	3	23934.0	NaN	
2	1001990	Bindu	P00118542	F	26-35	35	1	Uttar Pradesh	Central	Automobile	Auto	3	23924.0	NaN	
3	1001425	Sudevi	P00237842	M	0-17	16	0	Karnataka	Southern	Construction	Auto	2	23912.0	NaN	
4	1000588	Joni	P00057942	M	26-35	28	1	Gujarat	Western	Food Processing	Auto	2	23877.0	NaN	
...
11246	1000695	Manning	P00296942	M	18-25	19	1	Maharashtra	Western	Chemical	Office	4	370.0	NaN	
11247	1004089	Reichenbach	P00171342	M	26-35	33	0	Haryana	Northern	Healthcare	Veterinary	3	367.0	NaN	

```
[4]: df.describe()
```

```
[4]:
```

	User_ID	Age	Marital_Status	Orders	Amount	Status	unnamed1
count	1.125100e+04	11251.000000	11251.000000	11251.000000	11239.000000	0.0	0.0
mean	1.003004e+06	35.421207	0.420318	2.489290	9453.610858	NaN	NaN
std	1.716125e+03	12.754122	0.493632	1.115047	5222.355869	NaN	NaN
min	1.000001e+06	12.000000	0.000000	1.000000	188.000000	NaN	NaN
25%	1.001492e+06	27.000000	0.000000	1.500000	5443.000000	NaN	NaN
50%	1.003065e+06	33.000000	0.000000	2.000000	8109.000000	NaN	NaN
75%	1.004430e+06	43.000000	1.000000	3.000000	12675.000000	NaN	NaN
max	1.006040e+06	92.000000	1.000000	4.000000	23952.000000	NaN	NaN

```
[ ]:
```

```
[5]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 11251 entries, 0 to 11250
Data columns (total 15 columns):
#   Column                Non-Null Count  Dtype
---  -
0   User_ID                11251 non-null  int64
```

```
[6]: df.shape
```

```
[6]: (11251, 15)
```

```
[7]: df.drop(["Status","unnamed1"],axis=1,inplace=True)
```

```
[8]: df
```

```
[8]:
```

	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	23952.0
1	1000732	Kartik	P00110942	F	26-35	35	1	Andhra Pradesh	Southern	Govt	Auto	3	23934.0
2	1001990	Bindu	P00118542	F	26-35	35	1	Uttar Pradesh	Central	Automobile	Auto	3	23924.0
3	1001425	Sudevi	P00237842	M	0-17	16	0	Karnataka	Southern	Construction	Auto	2	23912.0
4	1000588	Joni	P00057942	M	26-35	28	1	Gujarat	Western	Food Processing	Auto	2	23877.0
...
11246	1000695	Manning	P00296942	M	18-25	19	1	Maharashtra	Western	Chemical	Office	4	370.0
11247	1004089	Reichenbach	P00171342	M	26-35	33	0	Haryana	Northern	Healthcare	Veterinary	3	367.0
11248	1001209	Oshin	P00201342	F	36-45	40	0	Madhya Pradesh	Central	Textile	Office	4	213.0
11249	1004023	Noonan	P00059442	M	36-45	37	0	Karnataka	Southern	Agriculture	Office	3	206.0
11250	1002744	Brumley	P00281742	F	18-25	19	0	Maharashtra	Western	Healthcare	Office	3	188.0

11251 rows × 13 columns

```
[9]: df.isnull().sum()
```

```
[9]: User_ID      0
Cust_name     0
Product_ID    0
Gender         0
Age Group     0
Age           0
Marital_Status 0
State         0
Zone          0
Occupation    0
Product_Category 0
Orders        0
Amount       12
dtype: int64
```

```
[10]: df.dropna(inplace=True)
```

```
[11]: #change data type
df["Amount"]=df["Amount"].astype("int64")
```

```
[12]: df["Amount"].dtypes
```

```
[12]: dtype('int64')
```

```
[13]: df.columns

[13]: Index(['User_ID', 'Cust_name', 'Product_ID', 'Gender', 'Age Group', 'Age',
        'Marital_Status', 'State', 'Zone', 'Occupation', 'Product_Category',
        'Orders', 'Amount'],
        dtype='object')

[14]: #rename column
      #df.rename(columns={"Marital_Status":"shaddi"})

[15]: # to describe specific columns
      df[["Age", "Orders", "Amount"]].describe()
```

	Age	Orders	Amount
count	11239.000000	11239.000000	11239.000000
mean	35.410357	2.489634	9453.610553
std	12.753866	1.114967	5222.355168
min	12.000000	1.000000	188.000000
25%	27.000000	2.000000	5443.000000
50%	33.000000	2.000000	8109.000000
75%	43.000000	3.000000	12675.000000
max	92.000000	4.000000	23952.000000

2. Exploratory Data Analysis (EDA)

- **Demographic Analysis:**
 - Analyzed customer demographics such as age and gender distribution to understand the primary consumer segments.
- **Purchase Behavior:**
 - Examined purchase patterns by age, gender, and marital status to identify high-spending customer groups.
- **Sales Trends:**
 - Explored trends in sales over time, focusing on peak purchase periods during the Diwali season.

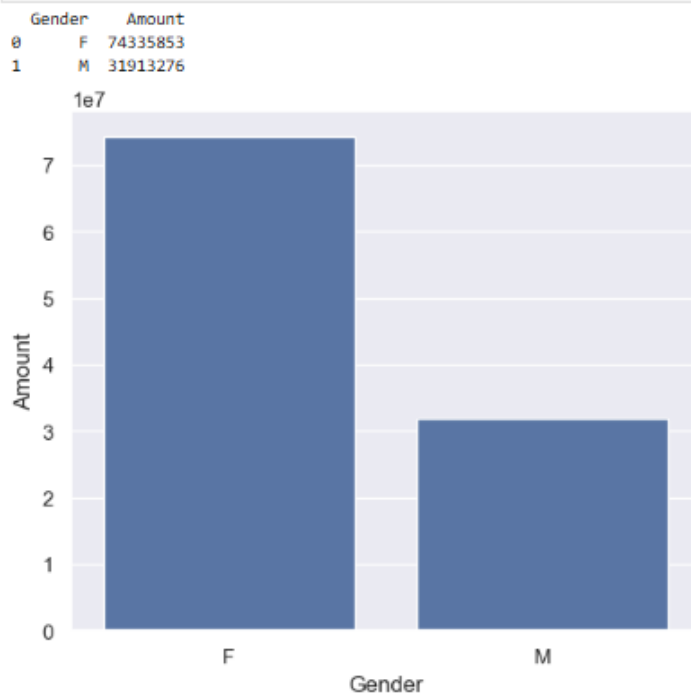
3. Key Insights

- **Gender-Based Spending:** Males tend to spend more on average compared to females during the Diwali season.
- **Age Group Preferences:** Customers in the 26-35 age group are the highest spenders, showing a strong purchasing trend in this demographic.
- **City-Wise Sales Analysis:** Tier 1 cities exhibit higher sales volume and spending compared to Tier 2 and Tier 3 cities.
- **Product Category Trends:** Certain product categories are more popular, especially during festive periods, which could guide inventory and promotional strategies.

```
[63]: #plotting a bar chart for gender vs total amount
```

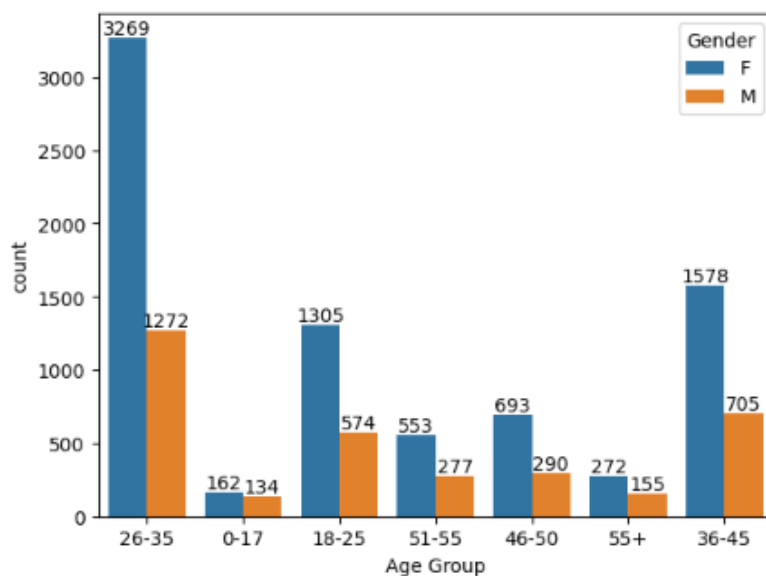
```
sales_gen=df.groupby(["Gender"],as_index=False)["Amount"].sum().sort_values(by="Amount",ascending =False)
sales_gen1=sns.barplot(x="Gender",y="Amount",data=sales_gen)

print(sales_gen)
```



from aboove graphs we can see that most of the buyers are female and even the purchasing power of females are greater than men

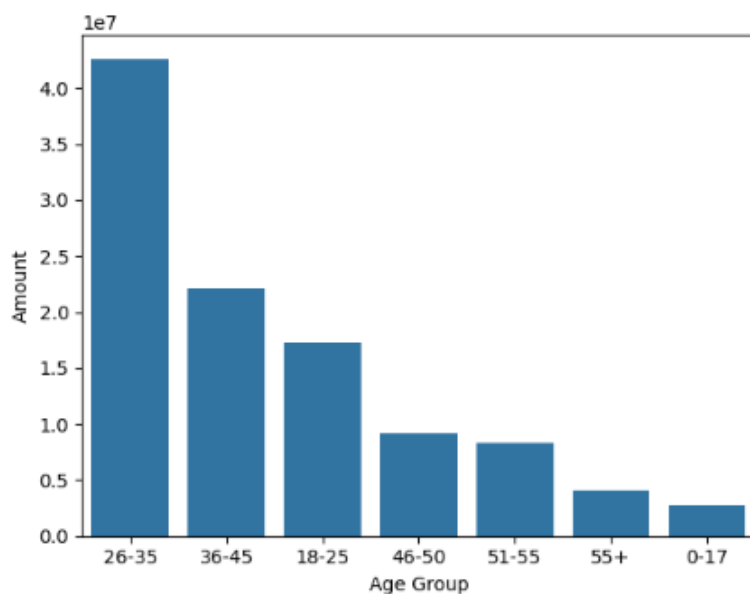
```
[19]: ax=sns.countplot(data=df,x="Age Group",hue="Gender")
for bars in ax.containers:
    ax.bar_label(bars)
```



Age Group

```
[20]: # Total Amount vs age group
sales_age=df.groupby(["Age Group"],as_index=False)["Amount"].sum().sort_values(by="Amount",ascending=False)
sns.barplot(x="Age Group",y="Amount",data=sales_age)
```

```
[20]: <Axes: xlabel='Age Group', ylabel='Amount'>
```

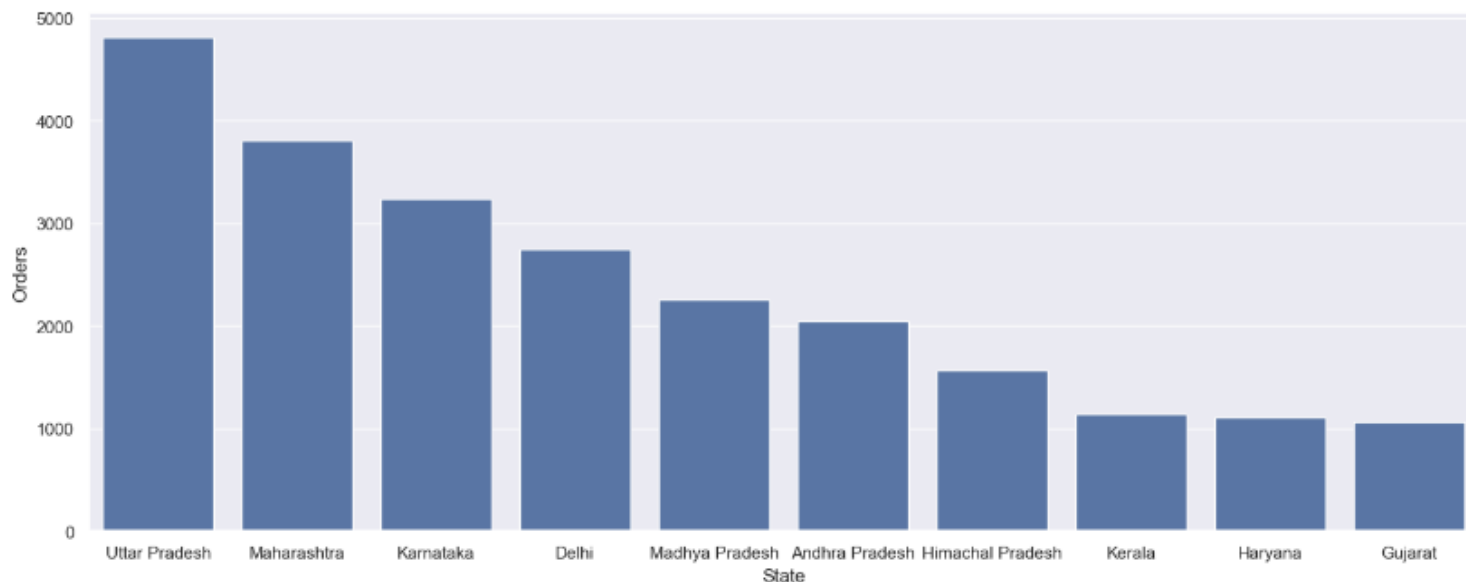


from above graphs we can see that most of the buyers are of age group between 26-35 yr females

State

```
[21]: # total number of orders from top 10 states
sales_state=df.groupby(["State"],as_index=False)["Orders"].sum().sort_values(by="Orders",ascending=False).head(10)
sns.set(rc={"figure.figsize":(16,6)})
sns.barplot(data=sales_state,x="State",y="Orders")
```

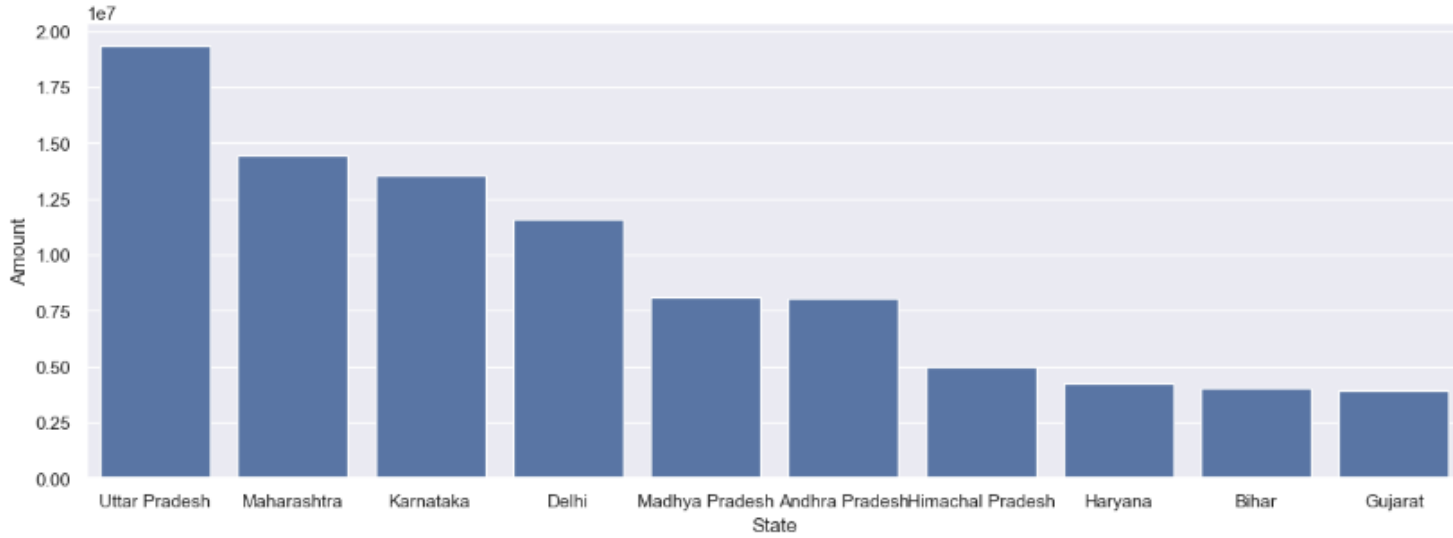
```
[21]: <Axes: xlabel='State', ylabel='Orders'>
```



State

```
[23]: # 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")
```

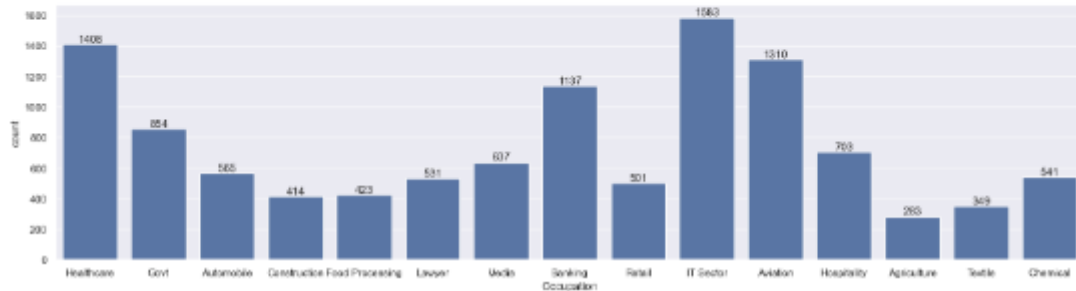
```
[23]: <Axes: xlabel='State', ylabel='Amount'>
```



from above graphs we can see that most of the ordrs and total sales /amount are from uttar pradesh maharashtra and karnataka respectively

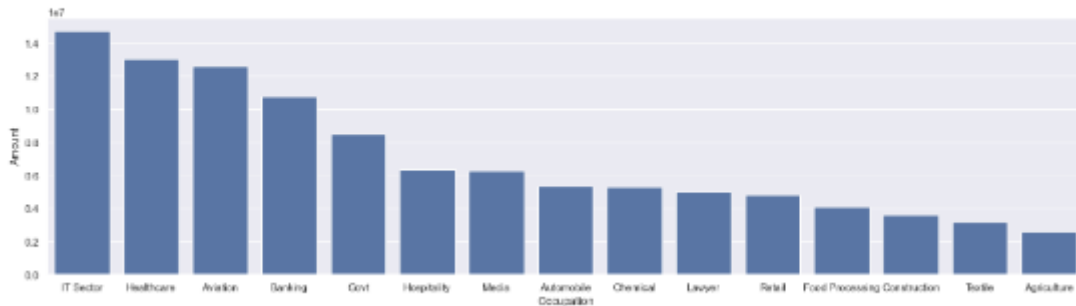
Occupation

```
[68]: sns.set(rc={"figure.figsize":(20,5)})
ax=sns.countplot(data=df,x="Occupation")
for bars in ax.containers:
    ax.bar_label(bars)
```



```
[74]: sales_state=df.groupby(["Occupation"],as_index=False)["Amount"].sum().sort_values(by="Amount",ascending=False)
sns.set(rc={"figure.figsize":(20,5)})
sns.barplot(data=sales_state,x="Occupation",y="Amount")
```

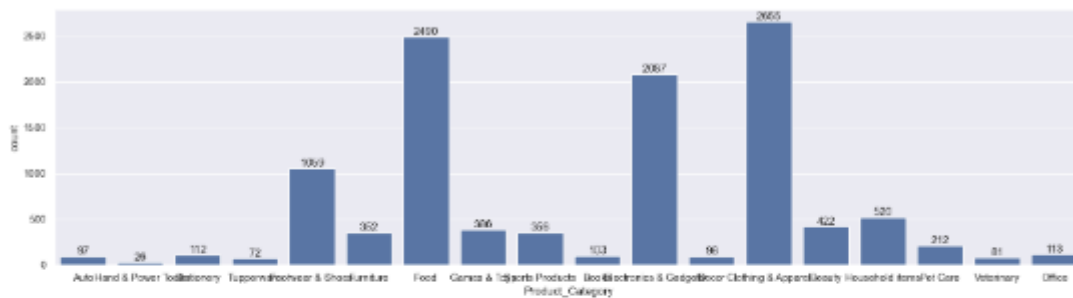
```
[74]: <Axes: xlabel='Occupation', ylabel='Amount'>
```



from above graphs we can see that most of the buyers are working in it ,healthcare and aviation sector

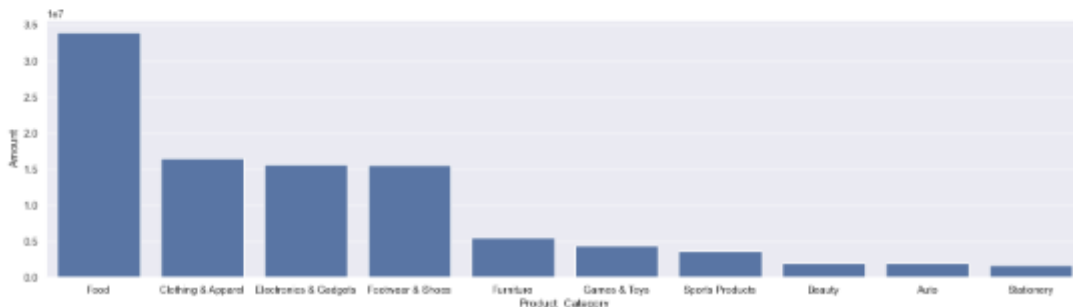
product category

```
[83]: sns.set(rc={"figure.figsize":(20,5)})
      ax=sns.countplot(data=df,x="Product_Category")
      for bars in ax.containers:
          ax.bar_label(bars)
```



```
[85]: sales_state=df.groupby(["Product_Category"],as_index=False)["Amount"].sum().sort_values(by="Amount",ascending=False).head(10)
      sns.set(rc={"figure.figsize":(20,5)})
      sns.barplot(data=sales_state,x="Product_Category",y="Amount")
```

```
[85]: <Axes: xlabel='Product_Category', ylabel='Amount'>
```



from above graphs we can see that most of the sold products are from food ,clothing and electronics category

5. Visualization

- Generated multiple visualizations to illustrate findings:
 - Bar charts and histograms for demographic breakdowns.
 - Line plots to show sales trends over time.
 - Heatmaps and scatter plots for correlations between purchase amount and other variables.

6. Recommendations

- **Target Marketing:** Focus marketing efforts on the 26-35 age group and males, as these segments show higher spending behavior.
- **Regional Promotions:** Increase promotions in Tier 1 cities, where spending volume is highest.
- **Product Stocking:** Prioritize popular product categories identified during the analysis to ensure adequate stock during high-demand periods.

Conclusion

The Diwali sales analysis provides valuable insights into consumer behavior, helping retailers to better understand their customers and optimize their sales and marketing strategies. By focusing on high-spending demographics and popular product categories, retailers can enhance customer satisfaction and maximize revenue during the festive season.