Ghanshyam Panchal Diwali sales data ananlysy to improove sales

```
# import python libraries
# import numpy as np
# import pandas as pd
# import matplotlib.pyplot as plt
# %matplotlib inline
# import seaborn as sns

# import csv file
df = pd.read_csv('Diwali Sales Data.csv', encoding='unicode_escape')
df.shape
```

```
(11251, 15)
```

```
df.head(3)
```

	User_ID	Cust_name	Product_ID	Gender	Age Group	Age	Marital_Status	State	Zone	Occupation	Product_Category	Orders	Amount	Unnamed1
0	1002903	Santhi	P00265242	F	26-35	28	1	Maharashtra	Western	Healthcare	Auto	1	23953	NaN
1	1000732	Kartik	P00110942	F	36-45	35	1	Andhra Pradesh	Southern	Govt	Auto	3	23404	NaN
2	1001990	Bindu	P00110842	F	26-35	35	1	Uttar Pradesh	Central	Automobile	Auto	3	23705	NaN
3	1001425	Sudevi	P00237842	M	0-17	16	0	Karnataka	Southern	Construction	Auto	2	23814	NaN
4	1000588	Joni	P00057942	M	36-45	44	0	Gujarat	Western	Food Processing	Auto	2	23799	NaN

```
df.info()
```

```
<class 'pd.DataFrame'>
RangeIndex: 11251 entries, 0 to 11250
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(5), object(7)
memory usage: 1.3 MB
```

```
# drop unused column values
df.drop(['Status', 'Unnamed1'], axis=1, inplace=True)
```

```
df.head()
```

	User_ID	Cust_name	Product_ID	Gender	Age Group	Age	Marital_Status	State	Zone	Occupation	Product_Category	Orders	Amount
0	1002903	Santhi	P00265242	F	26-35	28	1	Maharashtra	Western	Healthcare	Auto	1	23953
1	1000732	Kartik	P00110942	F	36-45	35	1	Andhra Pradesh	Southern	Govt	Auto	3	23404
2	1001990	Bindu	P00110842	F	26-35	35	1	Uttar Pradesh	Central	Automobile	Auto	3	23705
3	1001425	Sudevi	P00237842	M	0-17	16	0	Karnataka	Southern	Construction	Auto	2	23814
4	1000588	Joni	P00057942	M	36-45	44	0	Gujarat	Western	Food Processing	Auto	2	23799

```
# check for null values
pd.isnull().sum()
```

```
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
```

```
# drop null values
df.dropna(inplace=True)
```

```
# change data type
df['Amount'] = df['Amount'].astype('int')
```

```
df['Amount'].dtypes
```

```
dtype('int32')
```

```
df.columns
```

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

```
# rename column
df.rename(columns={'Marital_Status': 'Married'})
```

	User_ID	Cust_name	Product_ID	Gender	Age Group	Age	Married	State	Zone	Occupation	Product_Category	Orders	Amount
0	1002903	Santhi	P00265242	F	26-35	28	1	Maharashtra	Western	Healthcare	Auto	1	23953
1	1000732	Kartik	P00110942	F	36-45	35	1	Andhra Pradesh	Southern	Govt	Auto	3	23404
2	1001990	Bindu	P00110842	F	26-35	35	1	Uttar Pradesh	Central	Automobile	Auto	3	23705
3	1001425	Sudevi	P00237842	M	0-17	16	0	Karnataka	Southern	Construction	Auto	2	23814
4	1000588	Joni	P00057942	M	36-45	44	0	Gujarat	Western	Food Processing	Auto	2	23799

```
# describe
df.describe()
```

	User_ID	Age	Marital_Status	Orders	Amount
count	1.124900e+04	11249.000000	11249.000000	11249.000000	11249.000000
mean	1.003064e+06	35.410757	0.420391	2.489867	9453.609391
std	1.716039e+03	12.753866	0.493589	1.114967	5232.693460
min	1.000001e+06	12.000000	0.000000	1.000000	188.000000
25%	1.001492e+06	27.000000	0.000000	1.000000	5443.000000
50%	1.003064e+06	33.000000	0.000000	2.000000	8109.000000
75%	1.004546e+06	43.000000	1.000000	3.000000	12675.000000
max	1.006040e+06	92.000000	1.000000	4.000000	23953.000000

```
# get description for specific column
df['Amount', 'Orders', 'Gender', 'Age'].describe()
```

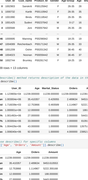
	Amount	Orders	Gender	Age
count	11239.000000	11249.000000	NaN	NaN
mean	9453.609391	2.489867	NaN	NaN
std	5232.693460	1.114967	NaN	NaN
min	188.000000	1.000000	NaN	NaN
25%	5443.000000	1.000000	NaN	NaN
50%	8109.000000	2.000000	NaN	NaN
75%	12675.000000	3.000000	NaN	NaN
max	23953.000000	4.000000	NaN	NaN

Exploratory Data Analysis

Gender

```
# plotting a bar chart for gender vs orders
sns.set(rc={'figure.figure.size':(7,5)})
ax = sns.barplot('Gender','Orders',data = df)
```

```
for bar in ax.patches:
    bar.set_edgecolor('')
```




```
# plotting a bar chart for gender vs total amount
df2 = df.groupby(['Gender','Amount'])['Amount'].sum().sort_values(ftr='Amount', ascending=False)
```

```
sns.barplot('x','Amount','y',data = df2,palette='Set1')
ax.set_xlabel('Gender',fontsize = 20)
```

```
df1 = sns.countplot(hue = df,x = 'Age Group', hue = 'Gender')
for bar in ax.patches:
    bar.set_edgecolor('')
```

```
# total amount vs age group
ax1=df.groupby('Age Group').sum()[['Amount']].sort_values(ftr='Amount', ascending=False)
```

```
sns.barplot('Age Group','Amount',data = ax1,palette='Set1')
ax.set_xlabel('Age Group',fontsize = 20)
```

```
# total number of orders from top 10 states
ax2=df.groupby('State').sum()[['Orders']].sum().sort_values(ftr='Orders', ascending=False).head(10)
```

```
sns.barplot('State','Orders',data = ax2,palette='Set1')
ax.set_xlabel('State',fontsize = 20)
```

```
# total amount from top 10 states
ax3=df.groupby('State').sum()[['Amount']].sum().sort_values(ftr='Amount', ascending=False).head(10)
```

```
sns.barplot('State','Amount',data = ax3,palette='Set1')
ax.set_xlabel('State',fontsize = 20)
```

Marital Status

```
df_1 = df.sort_values('Marital_Status')
```

```
sns.barplot('Marital_Status','Amount',data = df_1,palette='Set1')
ax.set_xlabel('Marital_Status',fontsize = 20)
```



```
df_2 = df.groupby('Marital_Status', 'Gender').sum()[['Amount']].sum().sort_values(ftr='Amount', ascending=False)
```

```
sns.barplot('Marital_Status','Amount',data = df_2,palette='Set1',hue='Gender')
ax.set_xlabel('Marital_Status',fontsize = 20)
```


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

Occupation

```
sns.set(rc={'figure.figure.size':(26,6)})
df1 = df.sort_values('Amount', ascending=False)
```

```
for bar in ax.patches:
    bar.set_edgecolor('')
```



```
df2=df.groupby('Occupation').sum()[['Amount']].sum().sort_values(ftr='Amount', ascending=False)
```

```
sns.barplot('Occupation','Amount',data = df2,palette='Set1')
ax.set_xlabel('Occupation',fontsize = 20)
```


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

Product Category¶

```
sns.set(rc={'figure.figure.size':(26,6)})
df1 = df.sort_values('Amount', ascending=False)
```

```
for bar in ax.patches:
    bar.set_edgecolor('')
```



```
sns.set(rc={'figure.figure.size':(26,6)})
df1 = df.sort_values('Amount', ascending=False)
```

```
for bar in ax.patches:
    bar.set_edgecolor('')
```



```
# top 10 best sold products (sum this up above)
df.groupby('Product_ID').sum()[['Amount']].sort_values(ftr='Amount', ascending=False).head(10)
```

```
df2=df.groupby('Product_ID').sum()[['Amount']].sort_values(ftr='Amount', ascending=False).head(10)
```

```
sns.barplot('x','Amount','y',data = df2,palette='Set1')
ax.set_xlabel('Product_ID',fontsize = 20)
```


Conclusion:

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

```
!pip install -U notebook ipywidgets
Collecting notebook ipywidgets
  Downloading notebook-7.0.6-py3-none-any.whl (5.1 MB)
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 5.1/5.1 MB 9.7 MB/s eta 0:00:00
Collecting ipywidgets
  Downloading ipywidgets-8.1.2-py3-none-any.whl (139 kB)
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 139.4/139.4 kB 7.9 MB/s eta 0:00:00
Requirement already satisfied: jupyter-lab-server~=2.19 in /usr/local/lib/python3.10/dist-packages (from notebook) (2.27.1)
Requirement already satisfied: jupyterlab~=4.0.6 in /usr/local/lib/python3.10/dist-packages (from notebook) (4.0.7)
Requirement already satisfied: nbformat~=5.8.0 in /usr/local/lib/python3.10/dist-packages (from notebook) (5.9.2)
Requirement already satisfied: traitlets~=5.6 in /usr/local/lib/python3.10/dist-packages (from notebook) (5.7.1)
Requirement already satisfied: jupyterlab_server~=2.28 in /usr/local/lib/python3.10/dist-packages (from notebook) (2.28.0)
Requirement already satisfied: jupyterlab_pygments~=0.3 in /usr/local/lib/python3.10/dist-packages (from nbconvert -nbsphinx -rtd theme -sphinxcontrib.spelling) (0.3.0)
Requirement already satisfied: MarkupSafe~=2.1 in /usr/local/lib/python3.10/dist-packages (from nbconvert -nbsphinx -rtd theme -sphinxcontrib.spelling) (2.1.5)
Requirement already satisfied: mistune~=3.0.1 in /usr/local/lib/python3.10/dist-packages (from nbconvert -nbsphinx -rtd theme -sphinxcontrib.spelling) (3.0.2)
Requirement already satisfied: nbclient~=0.5.0 in /usr/local/lib/python3.10/dist-packages (from nbconvert -nbsphinx -rtd theme -sphinxcontrib.spelling) (0.5.13)
Requirement already satisfied: nbconvert~=6.5.1 in /usr/local/lib/python3.10/dist-packages (from nbconvert -nbsphinx -rtd theme -sphinxcontrib.spelling) (6.5.4)
Requirement already satisfied: pandocfilters~=1.4.1 in /usr/local/lib/python3.10/dist-packages (from nbconvert -nbsphinx -rtd theme -sphinxcontrib.spelling) (1.4.3)
Requirement already satisfied: tinycss2~=1.1.1 in /usr/local/lib/python3.10/dist-packages (from nbconvert -nbsphinx -rtd theme -sphinxcontrib.spelling) (1.1.1)
Requirement already satisfied: comm~=0.1.3 in /usr/local/lib/python3.10/dist-packages (from ipywidgets) (0.1.3)
Requirement already satisfied: ipykernel~=6.19.3 in /usr/local/lib/python3.10/dist-packages (from ipywidgets) (6.19.3)
Requirement already satisfied: jupyterlab_widgets~=3.0.9,~=3.0.10 in /usr/local/lib/python3.10/dist-packages (from ipywidgets) (3.0.10)
Requirement already satisfied: widgetsnbextension~=4.0.9,~=4.0.10 in /usr/local/lib/python3.10/dist-packages (from ipywidgets) (4.0.10)
Requirement already satisfied: ipywidgets~=8.1.2 in /usr/local/lib/python3.10/dist-packages (from ipywidgets) (8.1.2)
Requirement already satisfied: jupyter-lab-server~=2.19 in /usr/local/lib/python3.10/dist-packages (from notebook) (2.27.1)
Requirement already satisfied: jupyterlab~=4.0.6 in /usr/local/lib/python3.10/dist-packages (from notebook) (4.0.7)
Requirement already satisfied: nbformat~=5.8.0 in /usr/local/lib/python3.10/dist-packages (from notebook) (5.9.2)
Requirement already satisfied: traitlets~=5.6 in /usr/local/lib/python3.10/dist-packages (from notebook) (5.7.1)
Requirement already satisfied: jupyterlab_server~=2.28 in /usr/local/lib/python3.10/dist-packages (from notebook) (2.28.0)
Requirement already satisfied: jupyterlab_pygments~=0.3 in /usr/local/lib/python3.10/dist-packages (from jupyter-client~=5.3.4/6.1.12/6.2.0/6.3.0/6.5.0/7.0.0+5.5.0/6.0.0+5.6.0/6.1.0+5.6.1/6.2.0+5.7.0/6.3.0+5.7.1/6.4.0+5.7.2/6.5.0+5.7.3/6.6.0+5.7.4/7.0.0b1+5.7.5b1 -jupyter-console -qtconsole -sphinx -nbsphinx -rtd theme -sphinxcontrib.spelling) (0.3.0)
Requirement already satisfied: MarkupSafe~=2.1 in /usr/local/lib/python3.10/dist-packages (from nbconvert -nbsphinx -rtd theme -sphinxcontrib.spelling) (2.1.5)
Requirement already satisfied: mistune~=3.0.1 in /usr/local/lib/python3.10/dist-packages (from nbconvert -nbsphinx -rtd theme -sphinxcontrib.spelling) (3.0.2)
Requirement already satisfied: nbclient~=0.5.0 in /usr/local/lib/python3.10/dist-packages (from nbconvert -nbsphinx -rtd theme -sphinxcontrib.spelling) (0.5.13)
Requirement already satisfied: nbconvert~=6.5.1 in /usr/local/lib/python3.10/dist-packages (from nbconvert -nbsphinx -rtd theme -sphinxcontrib.spelling) (6.5.4)
Requirement already satisfied: pandocfilters~=1.4.1 in /usr/local/lib/python3.10/dist-packages (from nbconvert -nbsphinx -rtd theme -sphinxcontrib.spelling) (1.4.3)
Requirement already satisfied: tinycss2~=1.1.1 in /usr/local/lib/python3.10/dist-packages (from nbconvert -nbsphinx -rtd theme -sphinxcontrib.spelling) (1.1.1)
Requirement already satisfied: comm~=0.1.3 in /usr/local/lib/python3.10/dist-packages (from ipywidgets) (0.1.3)
Requirement already satisfied: ipykernel~=6.19.3 in /usr/local/lib/python3.10/dist-packages (from ipywidgets) (6.19.3)
Requirement already satisfied: jupyterlab_widgets~=3.0.9,~=3.0.10 in /usr/local/lib/python3.10/dist-packages (from ipywidgets) (3.0.10)
Requirement already satisfied: widgetsnbextension~=4.0.9,~=4.0.10 in /usr/local/lib/python3.10/dist-packages (from ipywidgets) (4.0.10)
Requirement already satisfied: ipywidgets~=8.1.2 in /usr/local/lib/python3.10/dist-packages (from ipywidgets) (8.1.2
```