# import python libraries

import numpy as np

import pandas as pd

import matplotlib.pyplot as plt # visualizing data

%matplotlib inline

import seaborn as sns

# import csv file

df = pd.read\_csv('Diwali Sales Data.csv', encoding= 'unicode\_escape')

df.shape

df.head()

df.info()

#drop unrelated/blank columns

df.drop(['Status', 'unnamed1'], axis=1, inplace=True)

#check for null values

pd.isnull(df).sum()

# drop null values

df.dropna(inplace=True)

# change data type

df['Amount'] = df['Amount'].astype('int')

df['Amount'].dtypes

df.columns

#rename column

df.rename(columns= {'Marital\_Status':'Shaadi'})

# describe() method returns description of the data in the DataFrame (i.e. count, mean, std, etc)

df.describe()

# use describe() for specific columns

df[['Age', 'Orders', 'Amount']].describe()

# 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)

sns.barplot(x = 'Gender',y= 'Amount' ,data = sales\_gen)

ax = sns.countplot(data = df, x = 'Age Group', hue = 'Gender')

for bars in ax.containers:

ax.bar\_label(bars)

# Total Amount vs Age Group

sales\_age = df.groupby(['Age Group'], as\_index=False)['Amount'].sum().sort\_values(by='Amount', ascending=False)

# 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':(15,5)})

sns.barplot(data = sales\_state, x = 'State',y= 'Orders')

# 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':(15,5)})

sns.barplot(data = sales\_state, x = 'State',y= 'Orders')

# 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')

ax = sns.countplot(data = df, x = 'Marital\_Status')

sns.set(rc={'figure.figsize':(7,5)})

for bars in ax.containers:

ax.bar\_label(bars)

sales\_state = df.groupby(['Marital\_Status', 'Gender'], as\_index=False)['Amount'].sum().sort\_values(by='Amount', ascending=False)

sns.set(rc={'figure.figsize':(6,5)})

sns.barplot(data = sales\_state, x = 'Marital\_Status',y= 'Amount', hue='Gender')

sns.set(rc={'figure.figsize':(20,5)})

ax = sns.countplot(data = df, x = 'Occupation')

for bars in ax.containers:

ax.bar\_label(bars)

sns.barplot(x = 'Age Group',y= 'Amount' ,data = sales\_age)

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')

sns.set(rc={'figure.figsize':(20,5)})

ax = sns.countplot(data = df, x = 'Product\_Category')

for bars in ax.containers:

ax.bar\_label(bars)

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')

sales\_state = df.groupby(['Product\_ID'], as\_index=False)['Orders'].sum().sort\_values(by='Orders', ascending=False).head(10)

sns.set(rc={'figure.figsize':(20,5)})

sns.barplot(data = sales\_state, x = 'Product\_ID',y= 'Orders')

# 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')