Diwali Sales Data

```
import numpy as np
import pandas as pd
import matplotlib as mp
import seaborn as sns
%matplotlib inline
%matplotlib notebook
%matplotlib inline
```

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

```
df.shape
```

```
(11251, 15)
```

```
df.head()
```

	User_ID	Cust_name	Product_ID	Gender	Age Group	Age	Marital_Status	State	Zone	Occupation	Product_Category	Orders	Amount	Status	unnamed1
0	1002903	Sanvishi	P00239642	F	26-35	28	0	Maharashtra	Western	Healthcare	Auto	1	23953.0	NaN	NaN
1	1000732	Kartik	P00116142	M	26-35	35	1	Andhra Pradesh	Southern	Govt	Auto	3	23304.0	NaN	NaN
2	1001990	Bindu	P00114942	F	36-45	35	1	Uttar Pradesh	Central	Automobile	Auto	3	23933.0	NaN	NaN
3	1001610	Sudeer	P00099742	M	0-17	16	0	Karnataka	Southern	Construction	Auto	2	23123.0	NaN	NaN
4	1000588	joni	P00037742	M	26-35	28	1	Gujrat	Western	Food Processing	Auto	2	23877.0	NaN	NaN

```
df.info()
```

```
<class pandas.core.frame.DataFrame>
RangeIndex: 11251 entries, 0 to 11250
Data columns (total 15): 
 #   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_Category11251 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: int64, 6, float64, 3, object, 6
memory usage: 1.3 MB
```

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

```
df.shape
```

```
(11251, 13)
```

```
df.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
```

```
df.shape
```

```
(11251, 13)
```

```
df.dropna(inplace=True)
```

```
df.shape
```

```
(11239, 13)
```

```
df.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              0
dtype: int64
```

```
df.shape
```

```
(11239, 13)
```

```
df.dropna(inplace=True)
```

```
df.shape
```

```
(11239, 13)
```

```
df.info()
```

```
<class pandas.core.frame.DataFrame>
Int64Index: 11239 entries, 0 to 11250
Data columns (total 13): 
 #   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_Category11239 non-null  object 
 11  Orders          11239 non-null  int64  
 12  Amount          11239 non-null  float64
dtypes: int64, 4, float64, 1, object, 8
memory usage: 1.2 MB
```

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

```
df.info()
```

```
<class pandas.core.frame.DataFrame>
Int64Index: 11239 entries, 0 to 11250
Data columns (total 13): 
 #   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_Category11239 non-null  object 
 11  Orders          11239 non-null  int64  
 12  Amount          11239 non-null  float64
dtypes: int64, 4, float64, 1, object, 8
memory usage: 1.2 MB
```

```
df.shape
```

```
(11239, 13)
```