

```
In [1]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns

In [2]: data = pd.read_csv('C:\Users\kumar\Downloads\Python proj\01\wall Sales Data.csv', encoding='unicode_escape')

In [3]: data.head(5)

Out[3]:
   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  Karik  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

In [4]: df = pd.DataFrame(data)

In [5]: df.count()

Out[5]:
User_ID      11251
Cust_name    11251
Product_ID   11251
Gender        11251
Age Group     11251
Age           11251
Marital_Status 11251
State         11251
Zone          11251
Occupation    11251
Product_Category 11251
Orders        11239
Amount        11239
Status        0
unnamed1      0
dtype: int64

In [6]: dfa=df.drop(['Status', 'unnamed1'], axis=1)

In [7]: dfa.count()

Out[7]:
User_ID      11251
Cust_name    11251
Product_ID   11251
Gender        11251
Age Group     11251
Age           11251
Marital_Status 11251
State         11251
Zone          11251
Occupation    11251
Product_Category 11251
Orders        11239
Amount        11239
dtype: int64

In [8]: dfa.head()

Out[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  Karik  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

In [9]: dfa.isnull().sum()

Out[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

In [10]: dfa.dropna(inplace=True)

In [11]: dfa.isnull().sum()

Out[11]:
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
```

Exploratory data analysis

```
In [12]: vw = sns.countplot(x='Gender', data=dfa)

for bars in vw.containers:
    vw.bar_label(bars)

Out[12]:
count
7832
3407
Gender
F
M

In [13]: 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)

Out[13]:
<Axes: xlabel='Gender', ylabel='Amount'>

Amount
7e7
7
6
5
4
3
2
1
0
Gender
F
M

In [14]: ax=sns.countplot(x='Age Group', data=dfa, hue='Gender')
for bars in ax.containers:
    ax.bar_label(bars)

Out[14]:
count
3269
1272
162 134
1305
574
553
277
693
290
272 155
1578
705
Age Group
26-35 0-17 18-25 51-55 46-50 55+ 36-45
Gender
F
M

In [15]: sales_age = dfa.groupby(['Age Group'], as_index=False)['Amount'].sum().sort_values(by='Amount',ascending=False)
sns.barplot(x='Age Group', y='Amount', data=sales_age)

Out[15]:
<Axes: xlabel='Age Group', ylabel='Amount'>

Amount
1e7
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Age Group
26-35 36-45 18-25 46-50 51-55 55+ 0-17

In [16]: sns.set(rc={'figure.figsize':(15,5)})
orders_state = dfa.groupby(['State'], as_index=False)['Orders'].sum().sort_values(by='Orders', ascending=False).head(10)
sns.barplot(x='State', y='Orders', data=orders_state)

Out[16]:
<Axes: xlabel='State', ylabel='Orders'>

Orders
5000
4000
3000
2000
1000
0
Uttar Pradesh Maharashtra Karnataka Delhi Madhya Pradesh Andhra PradeshHimachal Pradesh Kerala Haryana Gujarat

In [17]: sns.set(rc={'figure.figsize':(15,5)})
sales_state = dfa.groupby(['State'], as_index=False)['Amount'].sum().sort_values(by='Amount', ascending=False).head(10)
sns.barplot(x='State', y='Amount', data=sales_state)

Out[17]:
<Axes: xlabel='State', ylabel='Amount'>

Amount
1e7
2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
Uttar Pradesh Maharashtra Karnataka Delhi Madhya Pradesh Andhra PradeshHimachal Pradesh Haryana Bihar Gujarat

In [18]: mr = sns.countplot(x='Marital_Status', data=dfa)
for bars in mr.containers:
    mr.bar_label(bars)

Out[18]:
count
6518
4721
Marital_Status
0
1

In [19]: mar_sales = dfa.groupby(['Marital_Status', 'Gender'], as_index=False)['Amount'].sum().sort_values(by='Amount', ascending=False)
sns.barplot(x='Marital_Status', y='Amount', data= mar_sales, hue='Gender')

Out[19]:
<Axes: xlabel='Marital_Status', ylabel='Amount'>

Amount
1e7
4
3
2
1
0
Marital_Status
0
1
Gender
F
M

In [20]: sns.set(rc={'figure.figsize':(20,5)})
pz = sns.countplot(x='Occupation', data=dfa)
for bars in pz.containers:
    pz.bar_label(bars)

Out[20]:
count
1408
854
585
414
423
531
637
1137
501
1583
1310
703
283
349
541
Occupation
Healthcare Govt Automobile Construction Food Processing Lawyer Media Banking Retail IT Sector Aviation Hospitality Agriculture Textile Chemical

In [21]: sns.set(rc={'figure.figsize':(20,5)})
occ_sales=dfa.groupby(['Occupation'], as_index=False)['Amount'].sum().sort_values(by='Amount', ascending=False)
sns.barplot(x='Occupation', y='Amount', data=occ_sales)

Out[21]:
<Axes: xlabel='Occupation', ylabel='Amount'>

Amount
1e7
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
IT Sector Healthcare Aviation Banking Govt Hospitality Media Automobile Occupation Chemical Lawyer Retail Food Processing Construction Textile Agriculture

In [22]: sns.set(rc={'figure.figsize':(25,5)})
sd = sns.countplot(x='Product_Category', data=dfa)
for bars in sd.containers:
    sd.bar_label(bars)

Out[22]:
count
2655
2087
1099
362
2680
360
358
103
98
423
520
212
81
113
Product_Category
Food Clothing & Apparel Electronics & Gadgets Footwear & Shoes Furniture Games & Toys Sports Products Beauty Household Items Pet Care Veterinary Office

In [23]: sns.set(rc={'figure.figsize':(20,5)})
pro_sales=dfa.groupby(['Product_Category'], as_index=False)['Amount'].sum().sort_values(by='Amount', ascending=False).head(10)
sns.barplot(x='Product_Category', y='Amount', data=pro_sales)

Out[23]:
<Axes: xlabel='Product_Category', ylabel='Amount'>

Amount
1e7
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Food Clothing & Apparel Electronics & Gadgets Footwear & Shoes Furniture Games & Toys Sports Products Beauty Auto Stationery

In [24]: pro_ord = 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 = pro_ord, x = 'Product_ID',y= 'Orders')

Out[24]:
<Axes: xlabel='Product_ID', ylabel='Orders'>

Orders
120
100
90
80
70
60
50
40
30
20
10
0
P00265242 P00110942 P00237542 P00184942 P00114942 P00025442 P00145042 P00117942 P00044442 P00110842

In [25]: pro_am = 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 = pro_am, x = 'Product_Category',y= 'Amount')

Out[25]:
<Axes: xlabel='Product_Category', ylabel='Amount'>

Amount
1e7
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Food Clothing & Apparel Electronics & Gadgets Footwear & Shoes Furniture Games & Toys Sports Products Beauty Auto Stationery
```

Conclusion

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

```
In [ ]:
In [ ]:
```