

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

In [2]: sales_data = pd.read_csv('D:\\wall Sales Data.csv', encoding='unicode_escape')

In [3]: sales_data.shape
Out[3]: (11251, 15)

In [4]: sales_data.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
 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              11239 non-null    float64
12   Amount              11239 non-null    float64
13   Status              0 non-null        float64
14   unnamed1             0 non-null        float64
dtypes: float64(3), int64(4), object(8)
memory usage: 1.3+ MB

In [5]: sales_data.head(5)
Out[5]:
   User_ID  Cust_Name  Product_ID  Gender  Age Group  Age  Marital_Status  State  Zone  Occupation  Product_Category  Orders  Amount  Status  unnamed1
0  1002903  Sarsniti  P00125842    F      26-35    28         0  Maharashtra  Western  Healthcare  Auto          1  23952.0  NaN      NaN
1  1000732  Karkk  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  Sudov  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 [6]: sales_data['Amount'] = sales_data['Amount'].fillna(0).astype(int)

In [7]: sales_data = sales_data.drop(columns=['Status', 'unnamed1'])

In [8]: sales_data.describe()
Out[8]:
   User_ID  Cust_Name  Age  Marital_Status  Orders  Amount
count  11251.000000  11251.000000  11251.000000  11251.000000  11251.000000
mean      1.003004e+06    35.421207      0.420318      2.489290    9443.527598
std      1.716125e+03    12.754122      0.493632      1.110647    5228.663337
min       1.000001e+06    12.000000      0.000000      1.000000      0.000000
25%      1.001492e+06    27.000000      0.000000      1.000000    5440.000000
50%      1.003004e+06    33.000000      0.000000      2.000000    8108.000000
75%      1.004410e+06    43.000000      1.000000      3.000000   12571.000000
max       1.005040e+06    92.000000      1.000000      4.000000   23952.000000

In [9]: sales_data[['Age', 'Orders', 'Amount']].describe()
Out[9]:
   Age  Orders  Amount
count  11251.000000  11251.000000  11251.000000
mean     35.421207      2.489290    9443.527598
std     12.754122     1.110647    5228.663337
min      12.000000      1.000000      0.000000
25%      27.000000      1.000000    5440.000000
50%      33.000000      2.000000    8108.000000
75%      43.000000      3.000000   12571.000000
max      92.000000      4.000000   23952.000000

In [10]: sales_data.loc[sales_data['Gender'] == 'F', 'Gender'] = 'Female'
sales_data.loc[sales_data['Gender'] == 'M', 'Gender'] = 'Male'

In [11]: sales_data.loc[sales_data['Marital_Status'] == 0, 'Marital_Status'] = 'Unmarried'
sales_data.loc[sales_data['Marital_Status'] == 1, 'Marital_Status'] = 'Married'

In [12]: sales_data.head(5)
Out[12]:
   User_ID  Cust_Name  Product_ID  Gender  Age Group  Age  Marital_Status  State  Zone  Occupation  Product_Category  Orders  Amount
0  1002903  Sarsniti  P00125842  Female  26-35    28      Unmarried  Maharashtra  Western  Healthcare  Auto          1  23952
1  1000732  Karkk  P00110942  Female  26-35    35      Married  Andhra Pradesh  Southern  Govt          Auto          3  23934
2  1001990  Bindu  P00118542  Female  26-35    35      Married  Uttar Pradesh  Central  Automobile  Auto          3  23924
3  1001425  Sudov  P00237842  Male  0-17    16      Unmarried  Karnataka  Southern  Construction  Auto          2  23912
4  1000588  Joni  P00057942  Male  26-35    28      Married  Gujarat  Western  Food Processing  Auto          2  23877
```

Exploratory Data Analysis

1. Gender-Based Analysis of Order Volume and Spending Patterns

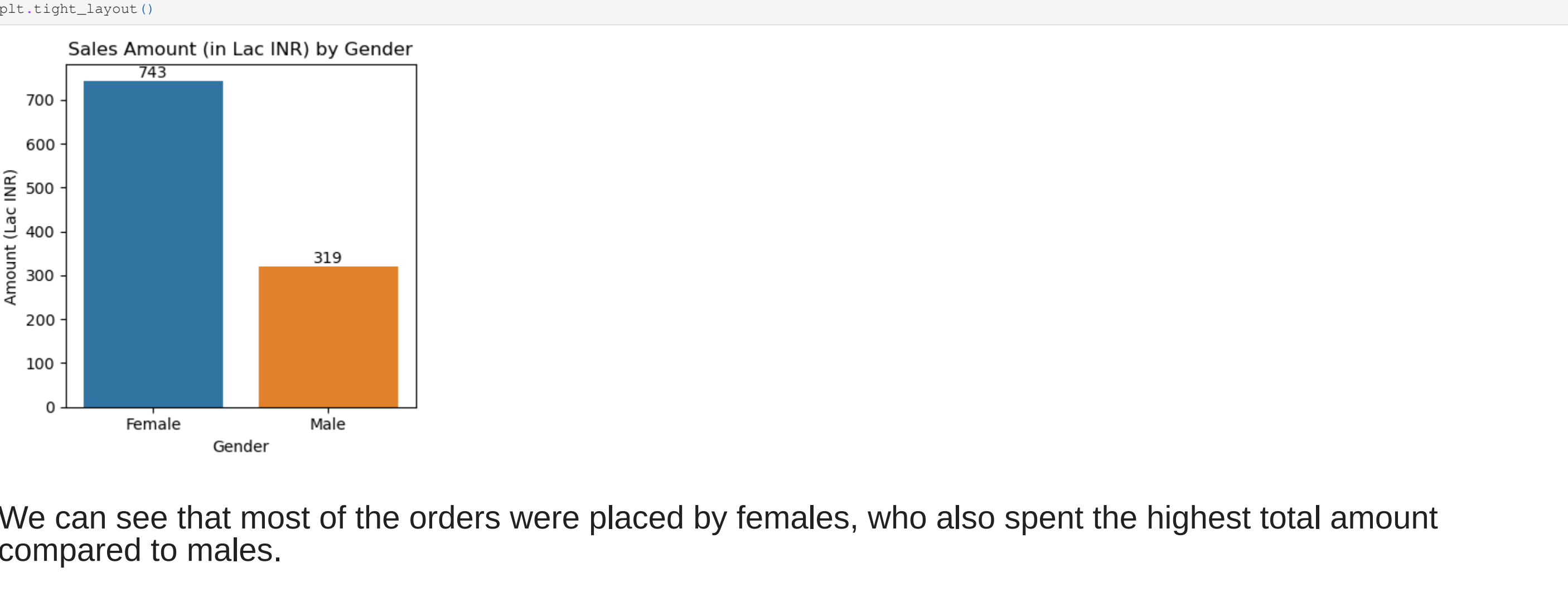
```
In [67]: plt.figure(figsize=(4, 4))
plot = sns.countplot(x='Gender', data=sales_data)
plt.title('Total orders count by Gender')

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

In [68]: sales_gen = sales_data.groupby(['Gender'], as_index=False)['Amount'].sum().sort_values(by='Amount', ascending=False)
sales_gen['Amount'] = sales_gen['Amount'] / 1e5
sales_gen['Amount'] = sales_gen['Amount'].astype(int)

In [68]: plt.figure(figsize=(4,4))
ax = sns.barplot(data=sales_gen, x='Gender', y='Amount')
plt.title('Sales Amount (in Lac INR) by Gender')
ax.bar_label(ax.containers(0), fmt='%0.1f') # '%0.1f' removes decimal places
plt.tight_layout()

In [68]:
Total orders count by Gender
count
Gender
Female 7842
Male 3409
```



We can see that most of the orders were placed by females, who also spent the highest total amount compared to males.

2. Age-group Based Analysis of Order Volume and Spending Patterns

```
In [69]: plt.figure(figsize=(10,6))
plot = sns.countplot(x='Age Group', data=sales_data, hue='Gender')
plt.title('Total orders count by Gender and their age group')

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

In [70]: sales_age = sales_data.groupby(['Age Group', 'Gender'], as_index=False)['Amount'].sum().sort_values(by='Amount', ascending=False)
sales_age['Amount'] = sales_age['Amount'] / 1e5
sales_age['Amount'] = sales_age['Amount'].astype(int)

In [70]: plt.figure(figsize=(12, 6))
ax = sns.barplot(data=sales_age, x='Age Group', y='Amount', hue='Gender')
plt.title('Sales Amount (in Lac INR) by Age Group and Gender')
ax.bar_label(ax.containers(0), fmt='%0.1f') # '%0.1f' removes decimal places
plt.tight_layout()

In [70]:
Total orders count by Gender and their age group
count
Age Group Gender
26-35 Female 3271
26-35 Male 1272
0-17 Female 162
0-17 Male 134
18-25 Female 1305
18-25 Male 574
55+ Female 273
55+ Male 155
46-50 Female 696
46-50 Male 291
51-55 Female 554
51-55 Male 278
36-45 Female 1581
36-45 Male 705
```



We can see that most of the orders were placed by females of age group 26-35, who also spent the highest total amount compared to males.

3. Top 10 Cities by Order Volume and Spending Patterns

```
In [71]: sales_state_order = sales_data.groupby(['State'], as_index=False)['Orders'].sum().sort_values(by='Orders', ascending=False).head(10)

In [71]: sales_state_amount = sales_data.groupby(['State'], as_index=False)['Amount'].sum().sort_values(by='Amount', ascending=False).head(10)
sales_state_amount['Amount'] = sales_state_amount['Amount'] / 1e5 # Convert to Lac INR
sales_state_amount['Amount'] = sales_state_amount['Amount'].astype(int)

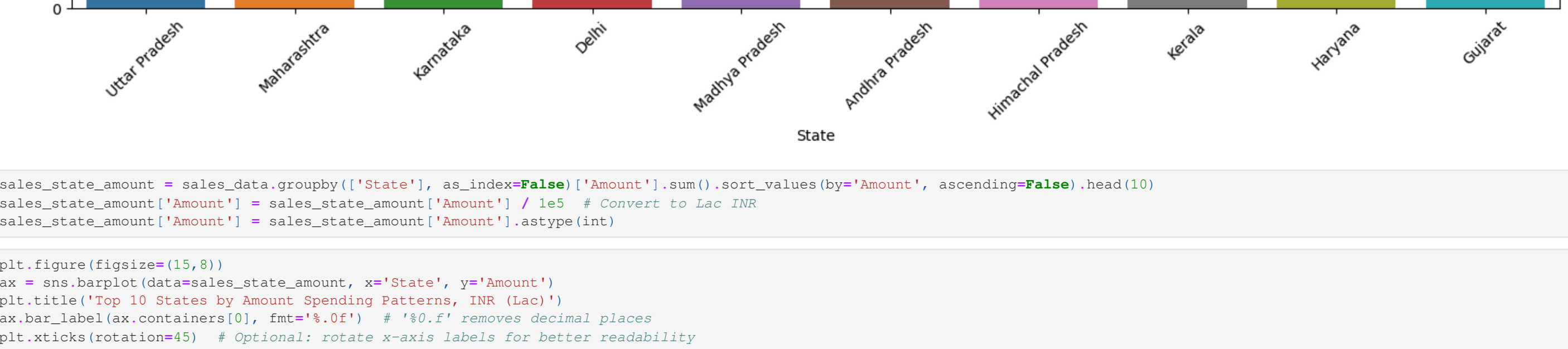
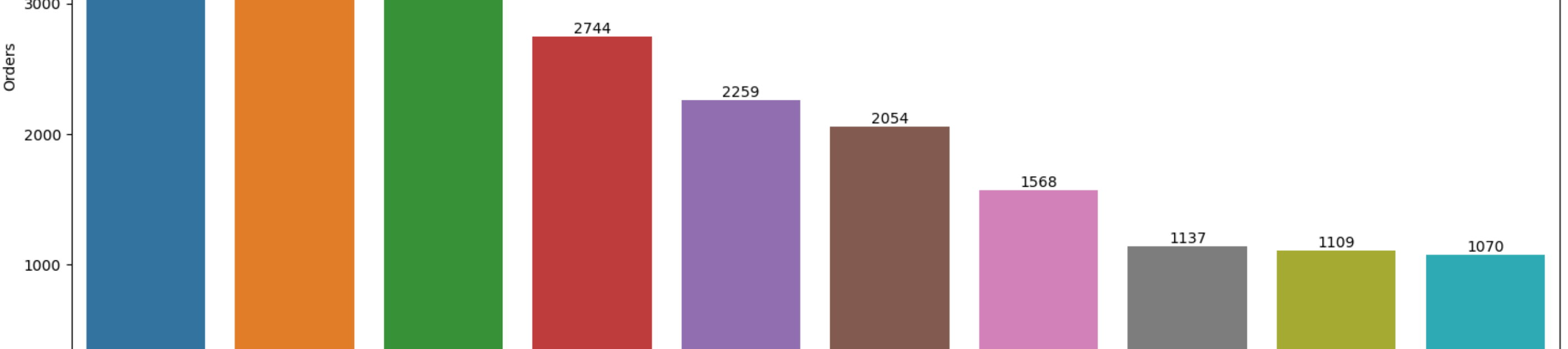
In [71]: plt.figure(figsize=(15,8))
ax = sns.barplot(data=sales_state_order, x='State', y='Orders')
plt.title('Top 10 States by Number of Orders')
ax.bar_label(ax.containers(0), fmt='%0.1f') # '%0.1f' removes decimal places
plt.xticks(rotation=45) # Optional: rotate x-axis labels for better readability
plt.tight_layout()

In [71]:
Top 10 States by Number of Orders
Orders
State
Uttar Pradesh 4813
Maharashtra 3811
Karnataka 3241
Delhi 2744
Madhya Pradesh 2259
Andhra Pradesh 2054
Himachal Pradesh 1568
Kerala 1137
Haryana 1109
Gujarat 1070

In [72]: sales_state_amount = sales_data.groupby(['State'], as_index=False)['Amount'].sum().sort_values(by='Amount', ascending=False).head(10)
sales_state_amount['Amount'] = sales_state_amount['Amount'] / 1e5 # Convert to Lac INR
sales_state_amount['Amount'] = sales_state_amount['Amount'].astype(int)

In [72]: plt.figure(figsize=(15,8))
ax = sns.barplot(data=sales_state_amount, x='State', y='Amount')
plt.title('Top 10 States by Amount Spending Patterns, INR (Lac)')
ax.bar_label(ax.containers(0), fmt='%0.1f') # '%0.1f' removes decimal places
plt.xticks(rotation=45) # Optional: rotate x-axis labels for better readability
plt.tight_layout()

In [72]:
Top 10 States by Amount Spending Patterns, INR (Lac)
Amount
State
Uttar Pradesh 193
Maharashtra 144
Karnataka 135
Delhi 116
Madhya Pradesh 81
Andhra Pradesh 80
Himachal Pradesh 49
Haryana 42
Bihar 40
Gujarat 39
```



From above graphs we can see that most of the orders & total sales/amount are from Uttar Pradesh, Maharashtra and Karnataka respectively

4. Marital-Based Analysis of Order Volume and Spending Patterns

```
In [73]: plt.figure(figsize=(5,4))
ax = sns.countplot(x='Marital_Status', data=sales_data)
ax.bar_label(ax.containers(0), fmt='%0.1f') # '%0.1f' removes decimal places
plt.title('Order Volume by Marital Status')
plt.tight_layout()

In [73]:
Order Volume by Marital Status
count
Marital_Status
Unmarried 6522
Married 4729

In [74]: sales_status = sales_data.groupby(['Marital_Status'], as_index=False)['Amount'].sum().sort_values(by='Amount', ascending=False)
sales_status['Amount'] = sales_status['Amount'] / 1e5
sales_status['Amount'] = sales_status['Amount'].astype(int)

In [74]: plt.figure(figsize=(5,4))
ax = sns.barplot(data=sales_status, x='Marital_Status', y='Amount')
plt.title('Sales Amount (in Lac INR) by Marital_Status')
ax.bar_label(ax.containers(0), fmt='%0.1f') # '%0.1f' removes decimal places
plt.tight_layout()

In [74]:
Sales Amount (in Lac INR) by Marital_Status
Amount
Marital_Status
Unmarried 621
Married 441
```

We can see that most of the orders were placed by Unmarried, who also spent the highest total amount compared to Married.

5. Occupation-Based Analysis of Order Volume and Spending Patterns

```
In [74]: plt.figure(figsize=(20,7))
ax = sns.countplot(x='Occupation', data=sales_data)
ax.bar_label(ax.containers(0), fmt='%0.1f') # '%0.1f' removes decimal places
plt.title('Order Volume by Occupation')
plt.tight_layout()

In [74]:
Order Volume by Occupation
count
Occupation
Healthcare 1408
Govt 854
Automobile 566
Construction 414
Food Processing 423
Lawyer 531
IT Sector 1588
Media 637
Banking 1139
Retail 501
Hospitality 705
Aviation 1210
Agriculture 283
Textile 350
Chemical 542

In [75]: sales_occ = sales_data.groupby(['Occupation'], as_index=False)['Amount'].sum().sort_values(by='Amount', ascending=False)
sales_occ['Amount'] = sales_occ['Amount'] / 1e5
sales_occ['Amount'] = sales_occ['Amount'].astype(int)

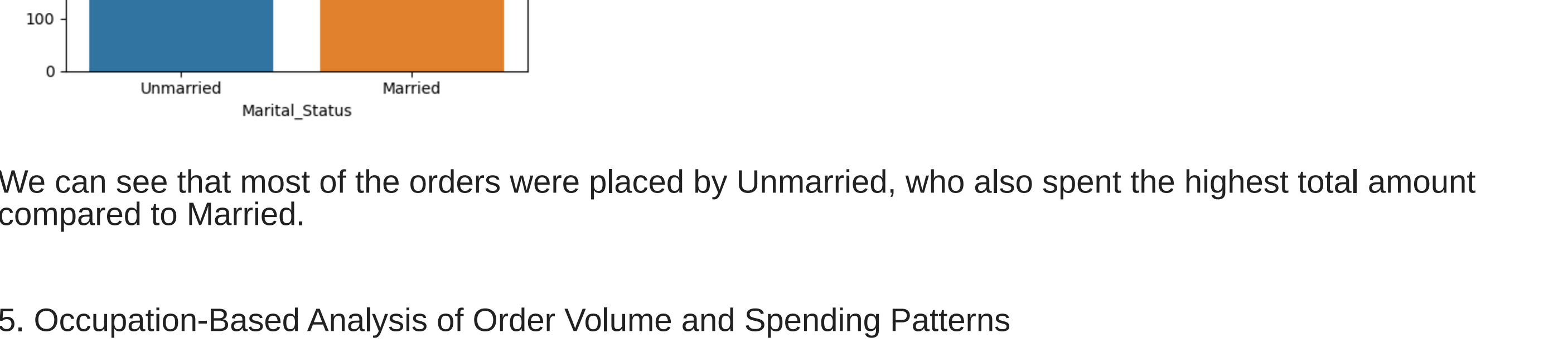
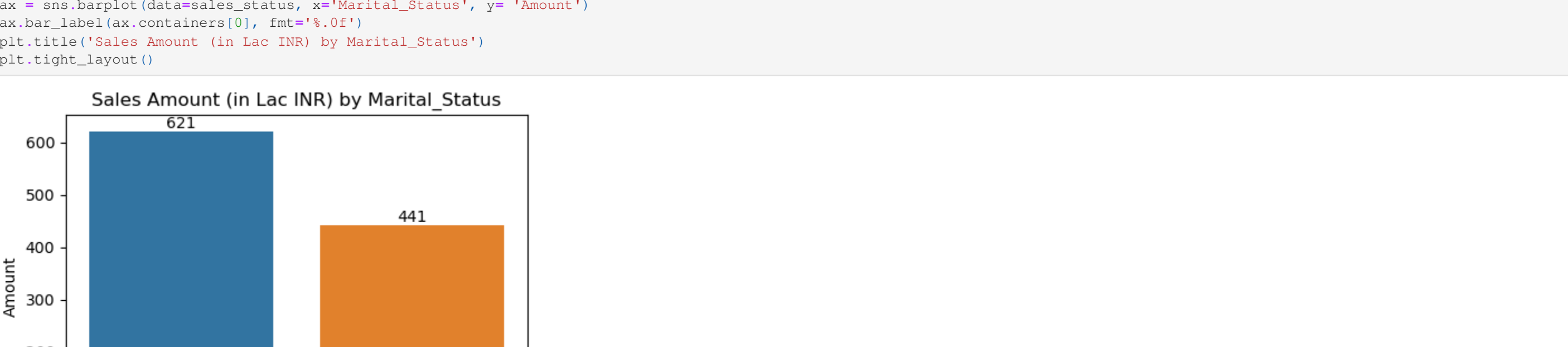
In [75]: plt.figure(figsize=(20,7))
ax = sns.barplot(data=sales_occ, x='Occupation', y='Amount')
ax.bar_label(ax.containers(0), fmt='%0.1f') # '%0.1f' removes decimal places
plt.title('Sales Amount (in Lac INR) by Occupation')
plt.tight_layout()

In [75]:
Sales Amount (in Lac INR) by Occupation
Amount
Occupation
IT Sector 147
Healthcare 130
Aviation 126
Banking 107
Govt 85
Hospitality 63
Media 62
Automobile 53
Chemical 52
Lawyer 49
Retail 47
Food Processing 40
Construction 35
Textile 32
Agriculture 25

In [76]: sales_catg = sales_data.groupby(['Product_Category'], as_index=False)['Amount'].sum().sort_values(by='Amount', ascending=False).head(10)
sales_catg['Amount'] = sales_catg['Amount'] / 1e5
sales_catg['Amount'] = sales_catg['Amount'].astype(int)

In [76]: plt.figure(figsize=(20,8))
ax = sns.barplot(data=sales_catg, x='Product_Category', y='Amount')
ax.bar_label(ax.containers(0), fmt='%0.1f') # '%0.1f' removes decimal places
plt.title('Sales Amount (in Lac INR) by Product Category')
plt.tight_layout()

In [76]:
Sales Amount (in Lac INR) by Product Category
Amount
Product_Category
Food 339
Clothing & Apparel 164
Electronics & Gadgets 156
Footwear & Shoes 155
Furniture 34
Games & Toys 41
Sports Products 36
Beauty 19
Auto 19
Stationery 16
```



we can see that as per above graphs that most of the orders placed in 'Clothing & Apparel', 'Food' and 'Electronics & Gadgets' categories However, they are spending highest amount in 'Food', 'Clothing & Apparel' and 'Electronics & Gadgets' categories

Conclusion:

Based on the Dwall sales data, most purchases were made by women aged 26-35, especially from states like Uttar Pradesh, Maharashtra, and Karnataka. Unmarried people and those working in IT and healthcare were among the top spenders.