```python
# MINI PROJECT IN PYTHON
# GROUP MEMBERS
# 1) 745 Neha More
# 2) 749 Aditya Pahurkar
# 3) 750 Kalyani Pardeshi


import pandas as pd
import numpy as np
import matplotlib.pyplot as plt


df=pd.read_csv("/tips.csv")


print(df.to_string())
```

	total_bill	tip	sex	smoker	day	time	size	tip_percentage
0	16.99	1.01	Female	No	Sun	Dinner	2	5.944673
1	10.34	1.66	Male	No	Sun	Dinner	3	16.054159
2	21.01	3.50	Male	No	Sun	Dinner	3	16.658734
3	23.68	3.31	Male	No	Sun	Dinner	2	13.978041
4	24.59	3.61	Female	No	Sun	Dinner	4	14.680765
5	25.29	4.71	Male	No	Sun	Dinner	4	18.623962
6	8.77	2.00	Male	No	Sun	Dinner	2	22.805017
7	26.88	3.12	Male	No	Sun	Dinner	4	11.607143
8	15.04	1.96	Male	No	Sun	Dinner	2	13.031915
9	14.78	3.23	Male	No	Sun	Dinner	2	21.853857
10	10.27	1.71	Male	No	Sun	Dinner	2	16.650438
11	35.26	5.00	Female	No	Sun	Dinner	4	14.180374
12	15.42	1.57	Male	No	Sun	Dinner	2	10.181582
13	18.43	3.00	Male	No	Sun	Dinner	4	16.277808
14	14.83	3.02	Female	No	Sun	Dinner	2	20.364127
15	21.58	3.92	Male	No	Sun	Dinner	2	18.164968
16	10.33	1.67	Female	No	Sun	Dinner	3	16.166505
17	16.29	3.71	Male	No	Sun	Dinner	3	22.774708
18	16.97	3.50	Female	No	Sun	Dinner	3	20.624632
19	20.65	3.35	Male	No	Sat	Dinner	3	16.222760
20	17.92	4.08	Male	No	Sat	Dinner	2	22.767857
21	20.29	2.75	Female	No	Sat	Dinner	2	13.553475
22	15.77	2.23	Female	No	Sat	Dinner	2	14.140774
23	39.42	7.58	Male	No	Sat	Dinner	4	19.228818
24	19.82	3.18	Male	No	Sat	Dinner	2	16.044400
25	17.81	2.34	Male	No	Sat	Dinner	4	13.138686
26	13.37	2.00	Male	No	Sat	Dinner	2	14.958863
27	12.69	2.00	Male	No	Sat	Dinner	2	15.760441
28	21.70	4.30	Male	No	Sat	Dinner	2	19.815668
29	19.65	3.00	Female	No	Sat	Dinner	2	15.267176

```
30       9.55  1.45    Male    No   Sat  Dinner   2     15.183246
31      18.35  2.50    Male    No   Sat  Dinner   4     13.623978
32      15.06  3.00  Female    No   Sat  Dinner   2     19.920319
33      20.69  2.45  Female    No   Sat  Dinner   4     11.841469
34      17.78  3.27    Male    No   Sat  Dinner   2     18.391451
35      24.06  3.60    Male    No   Sat  Dinner   3     14.962594
36      16.31  2.00    Male    No   Sat  Dinner   3     12.262416
37      16.93  3.07  Female    No   Sat  Dinner   3     18.133491
38      18.69  2.31    Male    No   Sat  Dinner   3     12.359551
39      31.27  5.00    Male    No   Sat  Dinner   3     15.989767
40      16.04  2.24    Male    No   Sat  Dinner   3     13.965087
41      17.46  2.54    Male    No   Sun  Dinner   2     14.547537
42      13.94  3.06    Male    No   Sun  Dinner   2     21.951220
43       9.68  1.32    Male    No   Sun  Dinner   2     13.636364
44      30.40  5.60    Male    No   Sun  Dinner   4     18.421053
45      18.29  3.00    Male    No   Sun  Dinner   2     16.402406
46      22.23  5.00    Male    No   Sun  Dinner   2     22.492128
47      32.40  6.00    Male    No   Sun  Dinner   4     18.518519
48      28.55  2.05    Male    No   Sun  Dinner   3      7.180385
49      18.04  3.00    Male    No   Sun  Dinner   2     16.629712
50      12.54  2.50    Male    No   Sun  Dinner   2     19.936204
51      10.29  2.60  Female    No   Sun  Dinner   2     25.267250
52      34.81  5.20  Female    No   Sun  Dinner   4     14.938236
53       9.94  1.56    Male    No   Sun  Dinner   2     15.694165
54      25.56  4.34    Male    No   Sun  Dinner   4     16.979656
55      19.49  3.51    Male    No   Sun  Dinner   2     18.009236
56      38.01  3.00    Male   Yes   Sat  Dinner   4      7.892660
```

#1) What is the average total bill amount?
```python
average_total_bill = df['total_bill'].mean()
print("Average total bill amount:", average_total_bill)
```

    Average total bill amount: 19.78594262295082

#2) How many customers are smokers?
```python
smoker_count = df[df['smoker'] == 'Yes'].shape[0]
print("Number of smokers:", smoker_count)
```

    Number of smokers: 93

#3)What is the highest tip amount?
```python
highest_tip = df['tip'].max()
print("Highest tip amount:", highest_tip)
```

    Highest tip amount: 10.0

#4)What is the percentage of male and female customers?

```python
gender_percentage = df['sex'].value_counts(normalize=True) * 100
print("Percentage of male and female customers:")
print(gender_percentage)
```

```
Percentage of male and female customers:
Male      64.344262
Female    35.655738
Name: sex, dtype: float64
```

```python
#5)What is the average tip amount for each day of the week?
average_tip_per_day = df.groupby('day')['tip'].mean()
print("Average tip amount per day:")
print(average_tip_per_day)
```

```
Average tip amount per day:
day
Fri     2.734737
Sat     2.993103
Sun     3.255132
Thur    2.771452
Name: tip, dtype: float64
```

```python
#6)What is the average total bill amount for dinner and lunch?
average_bill_per_time = df.groupby('time')['total_bill'].mean()
print("Average total bill amount per time:")
print(average_bill_per_time)
```

```
Average total bill amount per time:
time
Dinner    20.797159
Lunch     17.168676
Name: total_bill, dtype: float64
```

```python
#7)What is the total number of records in the dataset?
total_records = len(df)
print("Total number of records:", total_records)
```

```
Total number of records: 244
```

```python
#8)What is the average tip percentage (tip amount as a percentage of the total bill) for each record?
df['tip_percentage'] = (df['tip'] / df['total_bill']) * 100
average_tip_percentage = df['tip_percentage'].mean()
print("Average tip percentage:", average_tip_percentage)
```

```
       Average tip percentage: 16.08025817225047
```

```
#9)How many customers are males and how many are females?
gender_counts = df['sex'].value_counts()
print("Number of males and females:")
print(gender_counts)
```

```
       Number of males and females:
       Male        157
       Female       87
       Name: sex, dtype: int64
```

```
#10)What is the total amount spent on dinners and lunches separately?
total_spent_per_time = df.groupby('time')['total_bill'].sum()
print("Total amount spent per time:")
print(total_spent_per_time)
```

```
       Total amount spent per time:
       time
       Dinner    3660.30
       Lunch     1167.47
       Name: total_bill, dtype: float64
```

```
#11)What is the highest total bill amount for smokers and non-smokers?
highest_total_bill_smokers = df[df['smoker'] == 'Yes']['total_bill'].max()
highest_total_bill_non_smokers = df[df['smoker'] == 'No']['total_bill'].max()
print("Highest total bill amount for smokers:", highest_total_bill_smokers)
print("Highest total bill amount for non-smokers:", highest_total_bill_non_smokers)
```

```
       Highest total bill amount for smokers: 50.81
       Highest total bill amount for non-smokers: 48.33
```

```
#12)What is the average tip amount for male and female customers?
average_tip_per_gender = df.groupby('sex')['tip'].mean()
print("Average tip amount per gender:")
print(average_tip_per_gender)
```

```
      Average tip amount per gender:
      sex
      Female    2.833448
      Male      3.089618
      Name: tip, dtype: float64
```

```
#13)What is the highest tip amount given on Sundays?
highest_tip_sunday = df[df['day'] == 'Sun']['tip'].max()
print("Highest tip amount on Sundays:", highest_tip_sunday)
```

```
      Highest tip amount on Sundays: 6.5
```

```
#14)What is the average total bill amount for parties of different sizes?
average_bill_per_size = df.groupby('size')['total_bill'].mean()
print("Average total bill amount per group size:")
print(average_bill_per_size)
```

```
      Average total bill amount per group size:
      size
      1     7.242500
      2    16.448013
      3    23.277632
      4    28.613514
      5    30.068000
      6    34.830000
      Name: total_bill, dtype: float64
```

```
#15)How does the total bill amount vary with party size?
import matplotlib.pyplot as plt

plt.scatter(df['size'], df['total_bill'])
plt.xlabel('Party Size')
plt.ylabel('Total Bill Amount')
plt.title('Variation of Total Bill Amount with Party Size')
plt.show()
```

Variation of Total Bill Amount with Party Size

#16)How does the average tip percentage vary based on the day of the week?

```
import numpy as np

average_tip_percentage_per_day = df.groupby('day')['tip_percentage'].mean()

plt.bar(average_tip_percentage_per_day.index, average_tip_percentage_per_day.values)
plt.xlabel('Day of the Week')
plt.ylabel('Average Tip Percentage')
plt.title('Average Tip Percentage per Day')
plt.xticks(np.arange(len(average_tip_percentage_per_day)), average_tip_percentage_per_day.index)
plt.show()
```
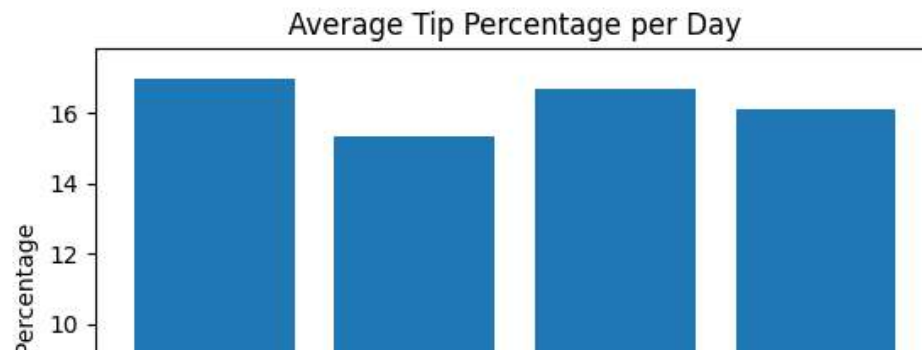
Object `week` not found.

Average Tip Percentage per Day

#17)What is the gender distribution of customers in the dataset?
```
gender_counts = df['sex'].value_counts()

plt.pie(gender_counts, labels=gender_counts.index, autopct='%1.1f%%')
plt.title('Gender Distribution of Customers')
plt.show()
```
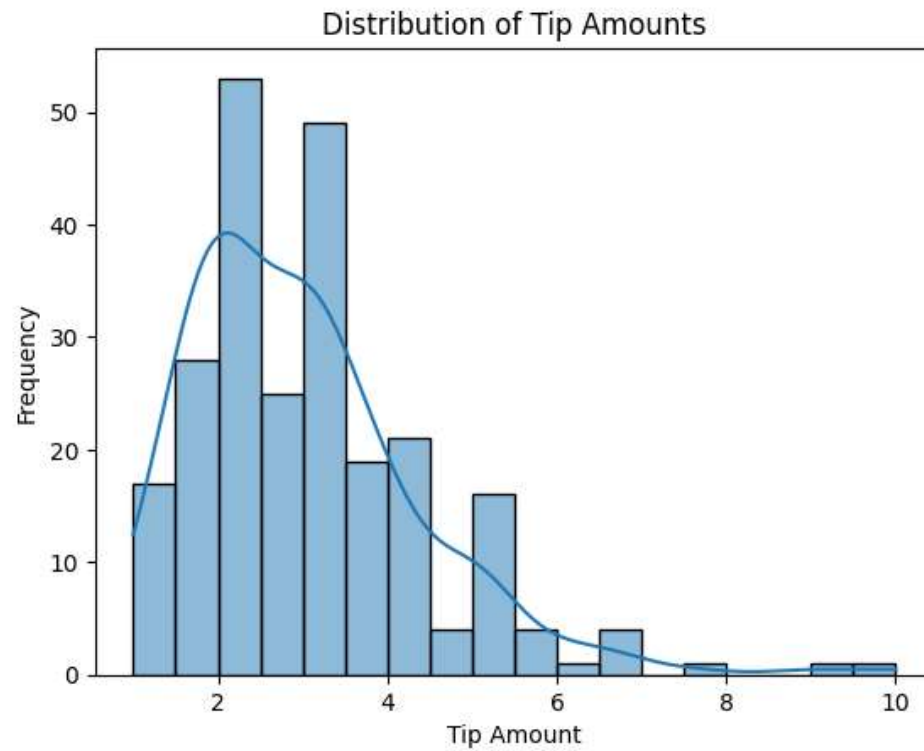
Gender Distribution of Customers
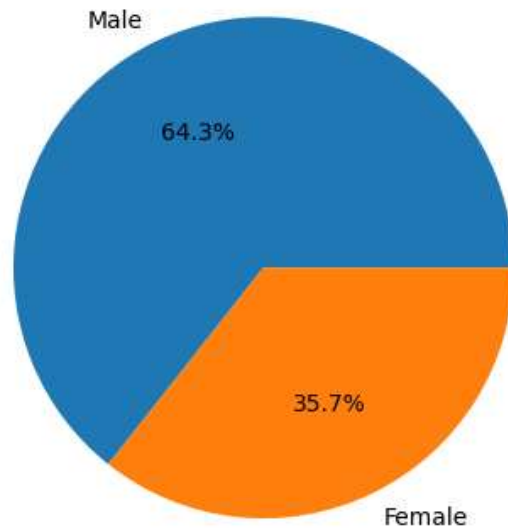
#18)What is the distribution of tip amounts?
```
import seaborn as sns
```

```python
sns.histplot(df['tip'], kde=True)
plt.xlabel('Tip Amount')
plt.ylabel('Frequency')
plt.title('Distribution of Tip Amounts')
plt.show()
```


Distribution of Tip Amounts

#19)How does the total bill amount differ between smokers and non-smokers?
```python
import seaborn as sns

sns.boxplot(x='smoker', y='total_bill', data=df)
plt.xlabel('Smoker')
plt.ylabel('Total Bill Amount')
plt.title('Comparison of Total Bill Amount between Smokers and Non-Smokers')
plt.show()
```

Comparison of Total Bill Amount between Smokers and Non-Smokers
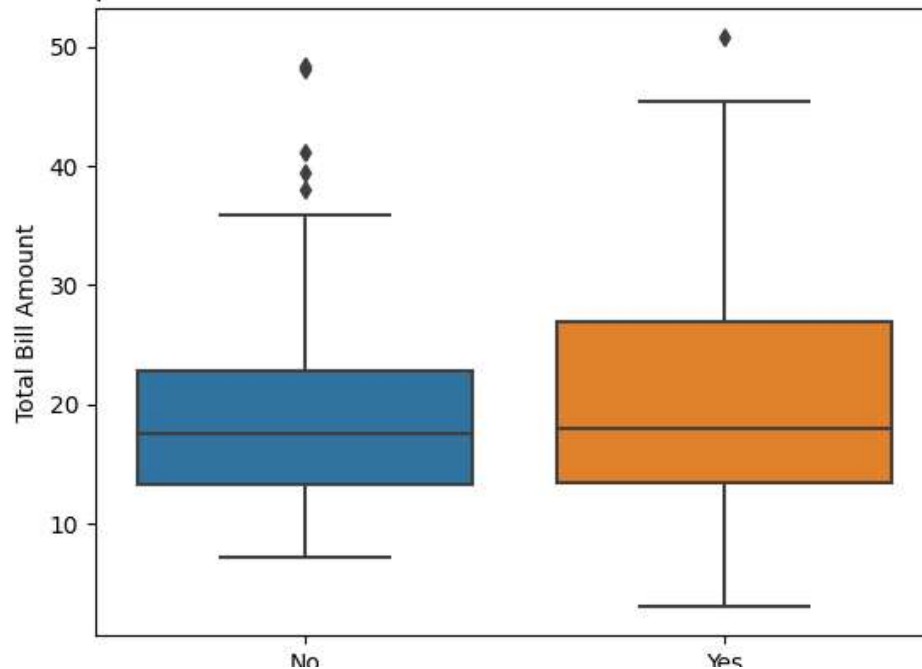

```
#20) What is the relationship between the total bill amount and the tip amount?
#Can we predict the tip amount based on the total bill using linear regression?
from sklearn.linear_model import LinearRegression
X = df[['total_bill']]
y = df['tip']
model = LinearRegression()
model.fit(X, y)
slope = model.coef_[0]
intercept = model.intercept_
new_total_bill = 30.0
predicted_tip = model.predict([[new_total_bill]])
plt.scatter(X, y)
plt.plot(X, model.predict(X), color='red', linewidth=2)
plt.xlabel('Total Bill Amount')
plt.ylabel('Tip Amount')
plt.title('Linear Regression: Total Bill vs Tip Amount')
plt.show()
print("Slope:", slope)
print("Intercept:", intercept)
print("Predicted tip amount for total bill $30.0:", predicted_tip)
```
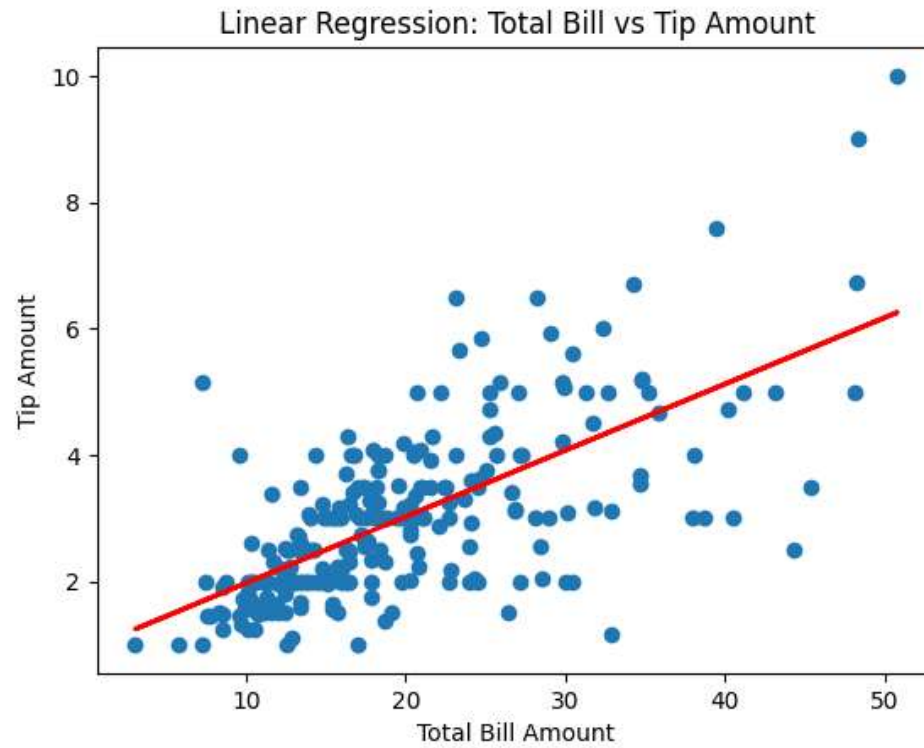
Linear Regression: Total Bill vs Tip Amount

Slope: 0.10502451738435337
Intercept: 0.9202696135546731
Predicted tip amount for total bill $30.0: [4.07100514]

✓ 0s   completed at 8:07 PM   ● ✕