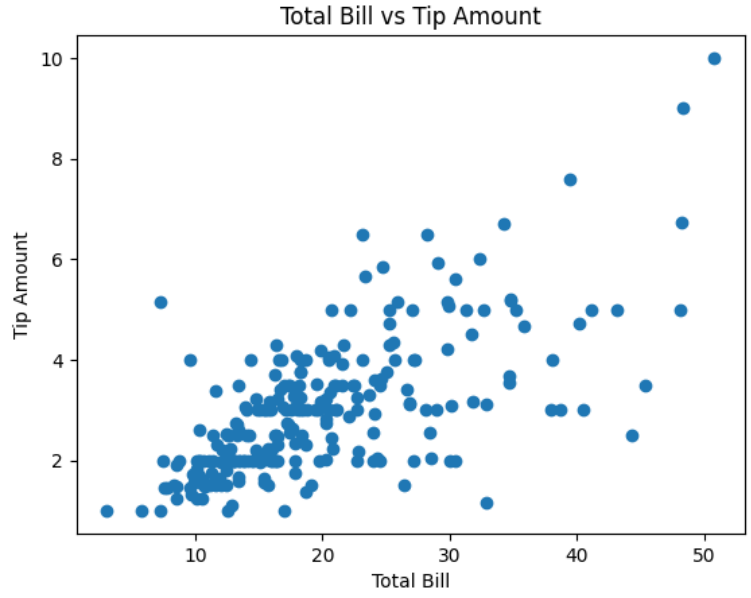
Name:-Darshan Ramagade
Roll NO.:- 548
EDS Practical 5

Here I have chosen the tips.csv file which is available on the moodle itself.

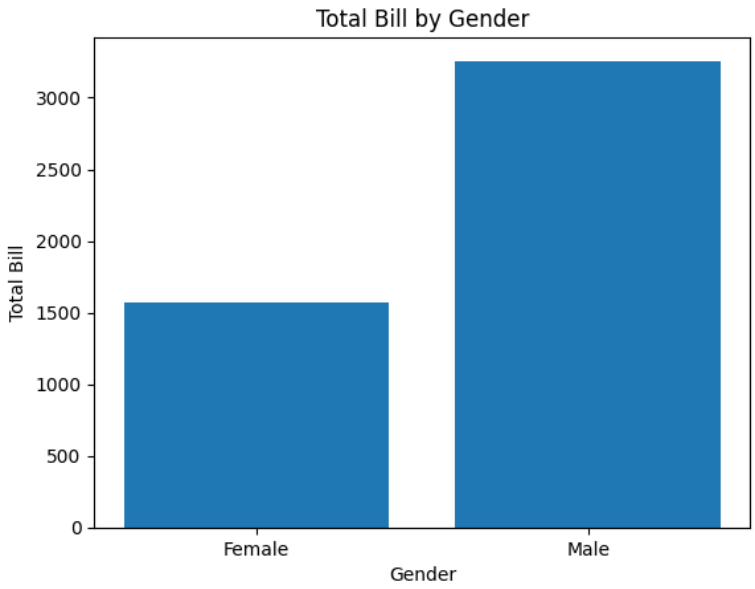
```
import pandas as pd
import matplotlib.pyplot as plt
```

```
df = pd.read_csv('/content/tips.csv')
```

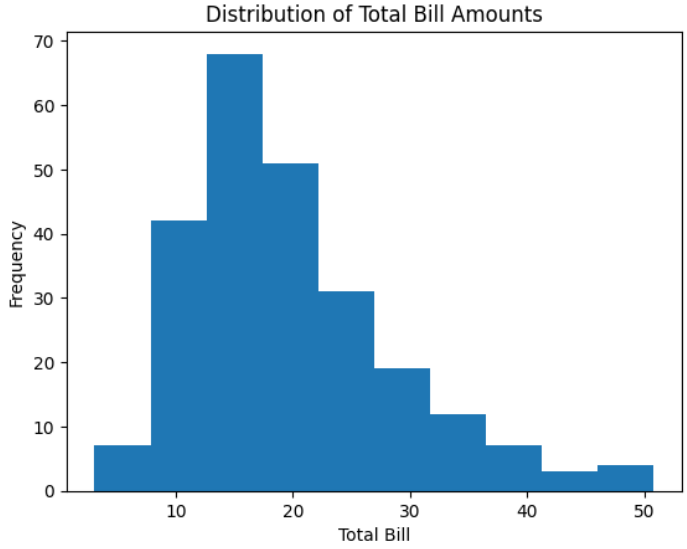
```
import matplotlib.pyplot as plt

plt.scatter(df['total_bill'], df['tip'])
plt.title('Total Bill vs Tip Amount')
plt.xlabel('Total Bill')
plt.ylabel('Tip Amount')
plt.show()
```



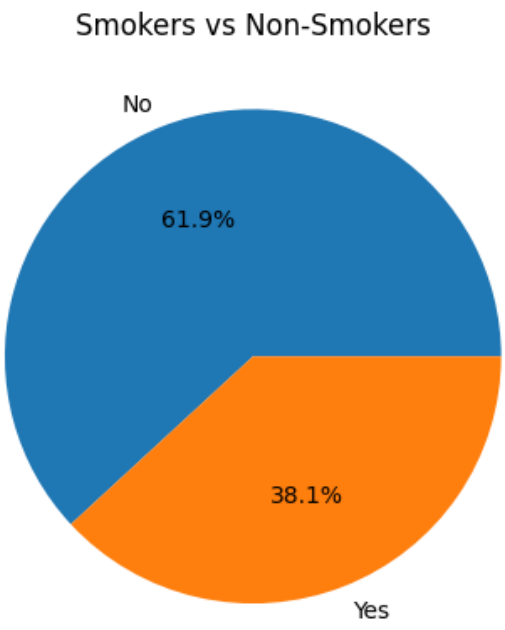

```
gender_totals = df.groupby('sex')['total_bill'].sum()
plt.bar(gender_totals.index, gender_totals.values)
plt.title('Total Bill by Gender')
plt.xlabel('Gender')
plt.ylabel('Total Bill')
plt.show()
```



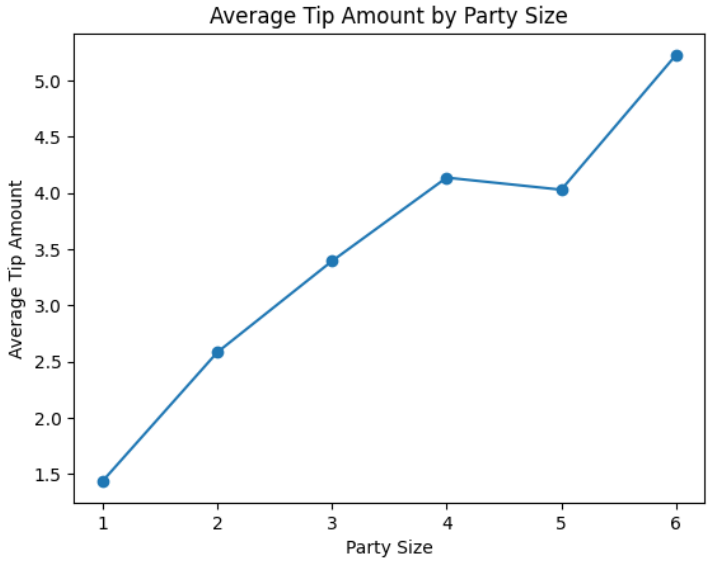

```
plt.hist(df['total_bill'], bins=10)
plt.title('Distribution of Total Bill Amounts')
plt.xlabel('Total Bill')
plt.ylabel('Frequency')
plt.show()
```




```
smoker_counts = df['smoker'].value_counts()
plt.pie(smoker_counts, labels=smoker_counts.index, autopct='%1.1f%%')
plt.title('Smokers vs Non-Smokers')
plt.show()
```

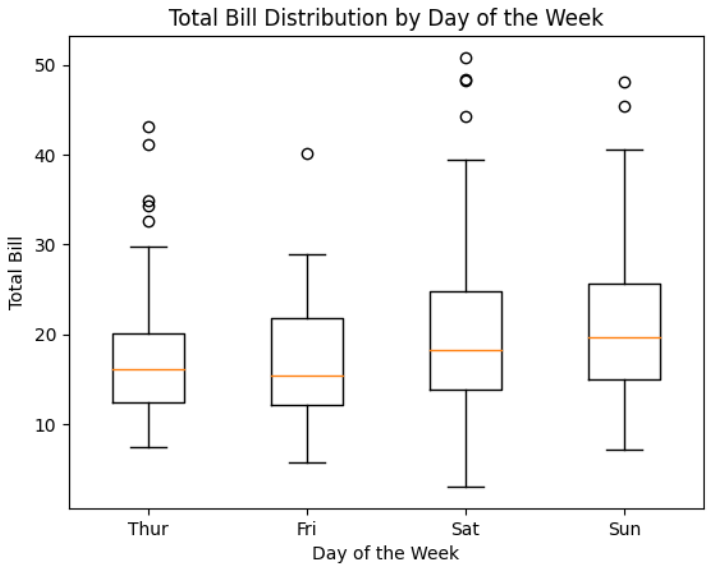



```
avg_tip_by_size = df.groupby('size')['tip'].mean()
plt.plot(avg_tip_by_size.index, avg_tip_by_size.values, marker='o')
plt.title('Average Tip Amount by Party Size')
plt.xlabel('Party Size')
plt.ylabel('Average Tip Amount')
plt.show()
```

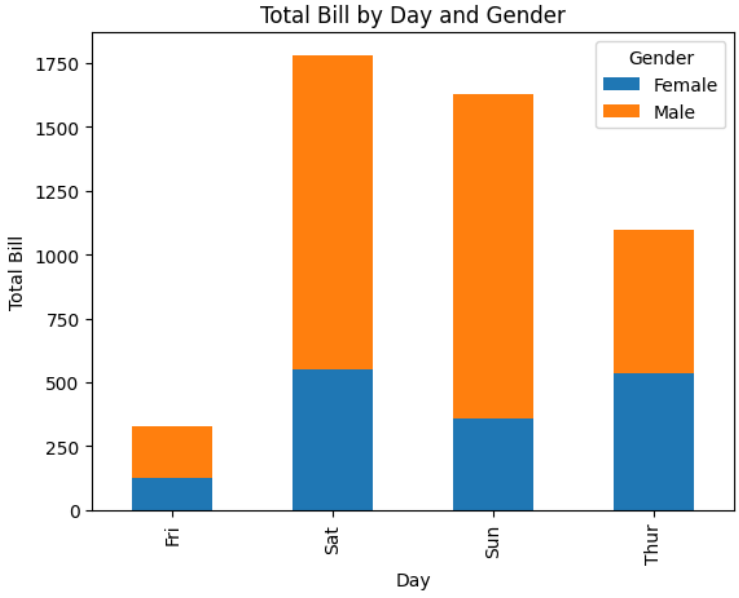



```
plt.boxplot([df[df['day'] == 'Thur']['total_bill'], df[df['day'] ==
'Fri']['total_bill'], df[df['day'] == 'Sat']['total_bill'],
df[df['day'] == 'Sun']['total_bill']])
plt.title('Total Bill Distribution by Day of the Week')
plt.xlabel('Day of the Week')
plt.ylabel('Total Bill')
```

```
plt.xticks([1, 2, 3, 4], ['Thur', 'Fri', 'Sat', 'Sun'])
plt.show()
```

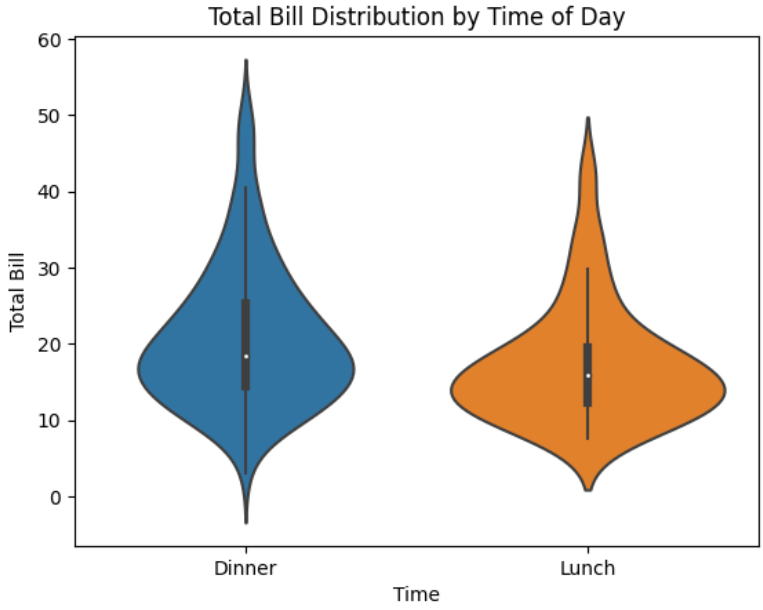



```
day_gender_totals = df.groupby(['day',
'sex'])['total_bill'].sum().unstack()
day_gender_totals.plot(kind='bar', stacked=True)
plt.title('Total Bill by Day and Gender')
plt.xlabel('Day')
plt.ylabel('Total Bill')
plt.legend(title='Gender')
plt.show()
```

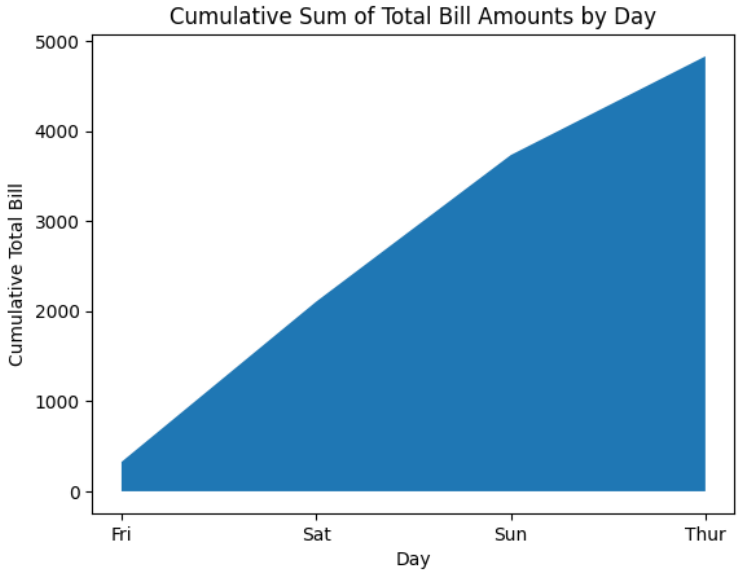



```
import seaborn as sns

sns.violinplot(x=df['time'], y=df['total_bill'])
```

```
plt.title('Total Bill Distribution by Time of Day')
plt.xlabel('Time')
plt.ylabel('Total Bill')
plt.show()
```




```
day_totals = df.groupby('day')['total_bill'].sum()
cumulative_totals = day_totals.cumsum()
plt.fill_between(cumulative_totals.index, cumulative_totals.values)
plt.title('Cumulative Sum of Total Bill Amounts by Day')
plt.xlabel('Day')
plt.ylabel('Cumulative Total Bill')
plt.show()
```




```
corr_matrix = df[['total_bill', 'tip', 'size']].corr()
sns.heatmap(corr_matrix, annot=True)
```

```
plt.title('Correlation Heatmap')
plt.show()
```