EDS ASSIGNMENT NO : 05

Name : Aditya Jadhav
PRN No : 202201040175
Roll No : 671
Div : F4

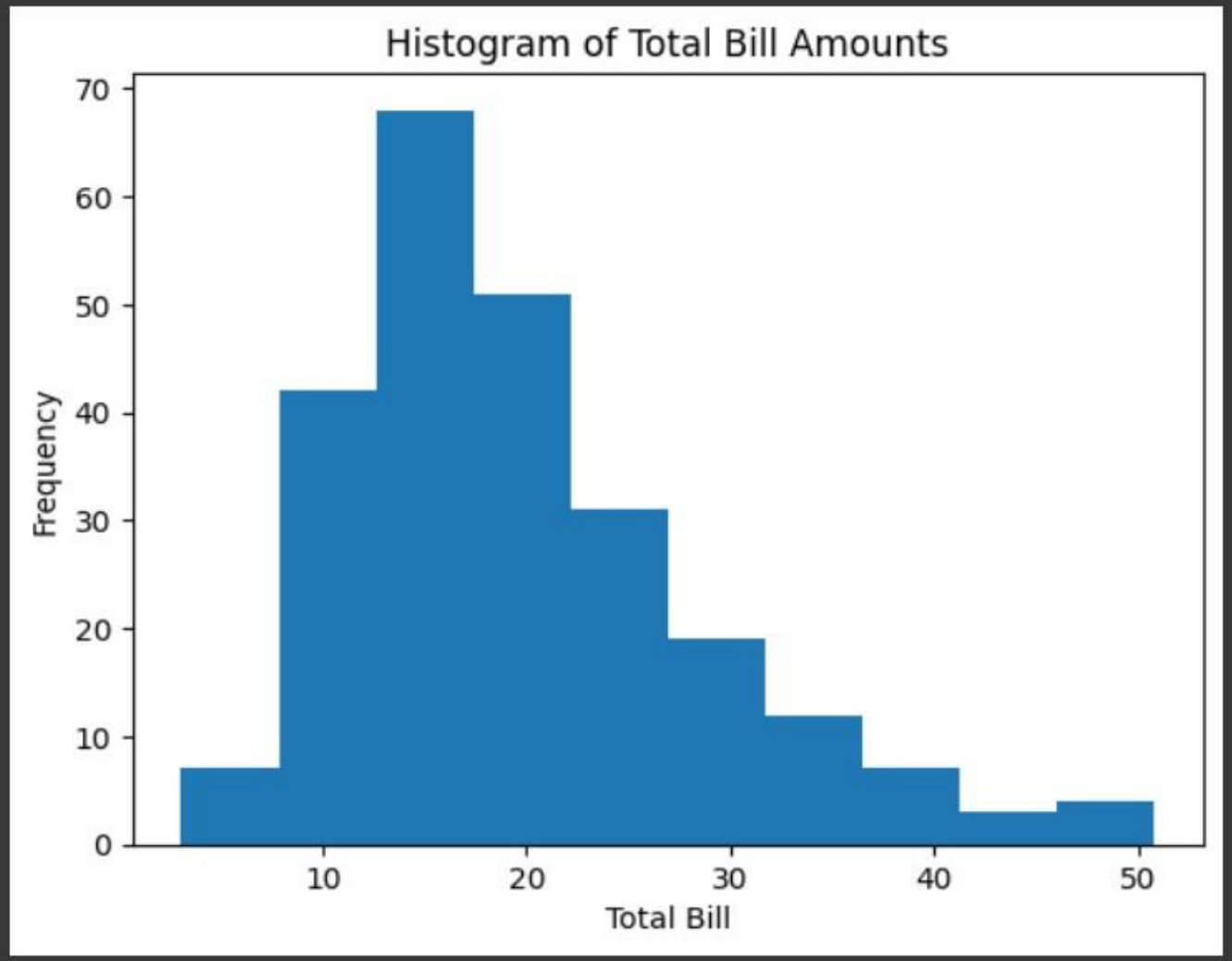
2s

```
import matplotlib.pyplot as plt
import seaborn as sns
import pandas as pd
data=pd.read_csv("/content/tips.csv")
df=pd.DataFrame(data)
print(df)
```

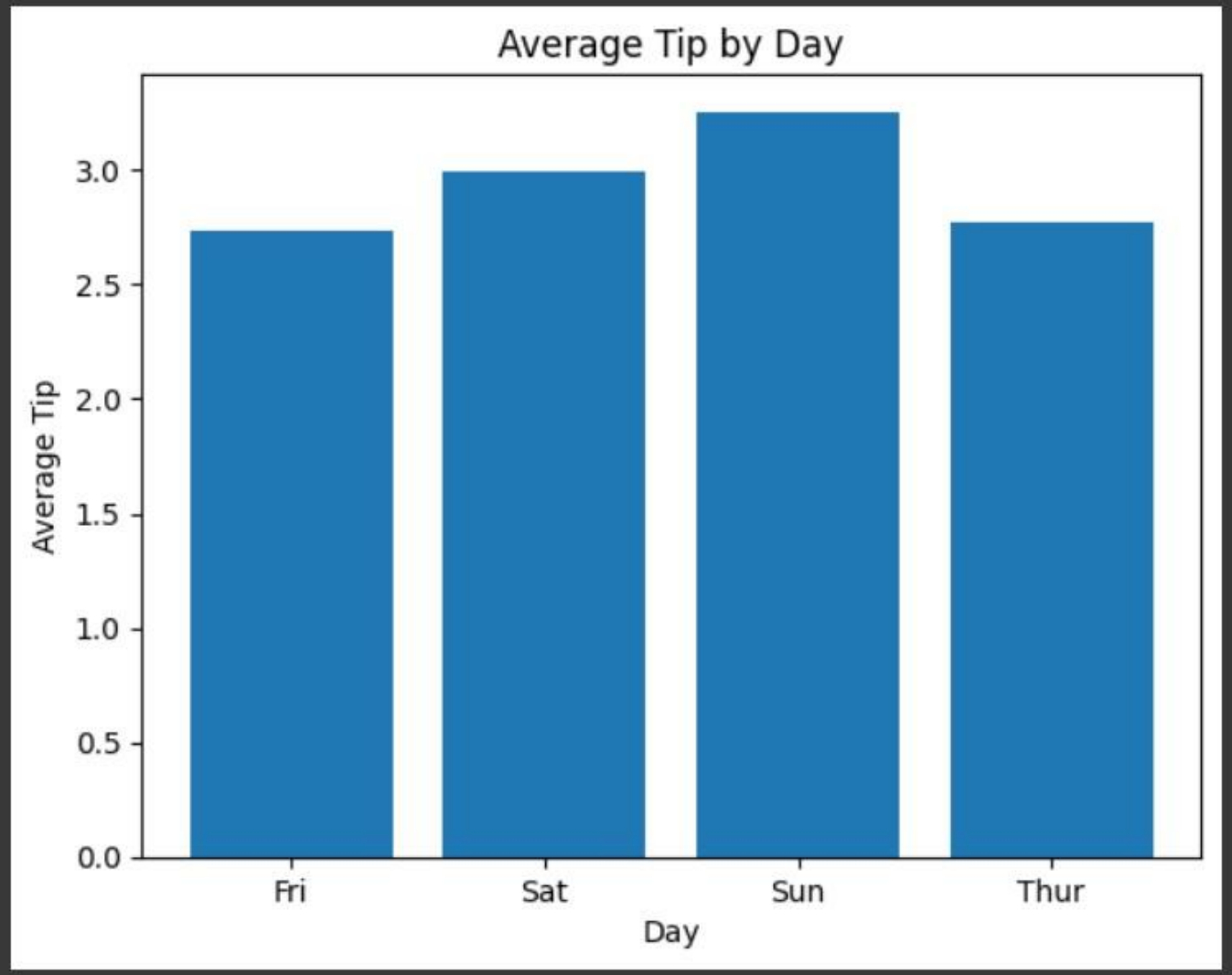
```
     total_bill   tip     sex smoker   day    time  size
0         16.99  1.01  Female     No   Sun  Dinner     2
1         10.34  1.66    Male     No   Sun  Dinner     3
2         21.01  3.50    Male     No   Sun  Dinner     3
3         23.68  3.31    Male     No   Sun  Dinner     2
4         24.59  3.61  Female     No   Sun  Dinner     4
..          ...   ...     ...    ...   ...     ...   ...
239       29.03  5.92    Male     No   Sat  Dinner     3
240       27.18  2.00  Female    Yes   Sat  Dinner     2
241       22.67  2.00    Male    Yes   Sat  Dinner     2
242       17.82  1.75    Male     No   Sat  Dinner     2
243       18.78  3.00  Female     No  Thur  Dinner     2

[244 rows x 7 columns]
```

```
plt.hist(df['total_bill'])
plt.xlabel('Total Bill')
plt.ylabel('Frequency')
plt.title('Histogram of Total Bill Amounts')
plt.show()
```

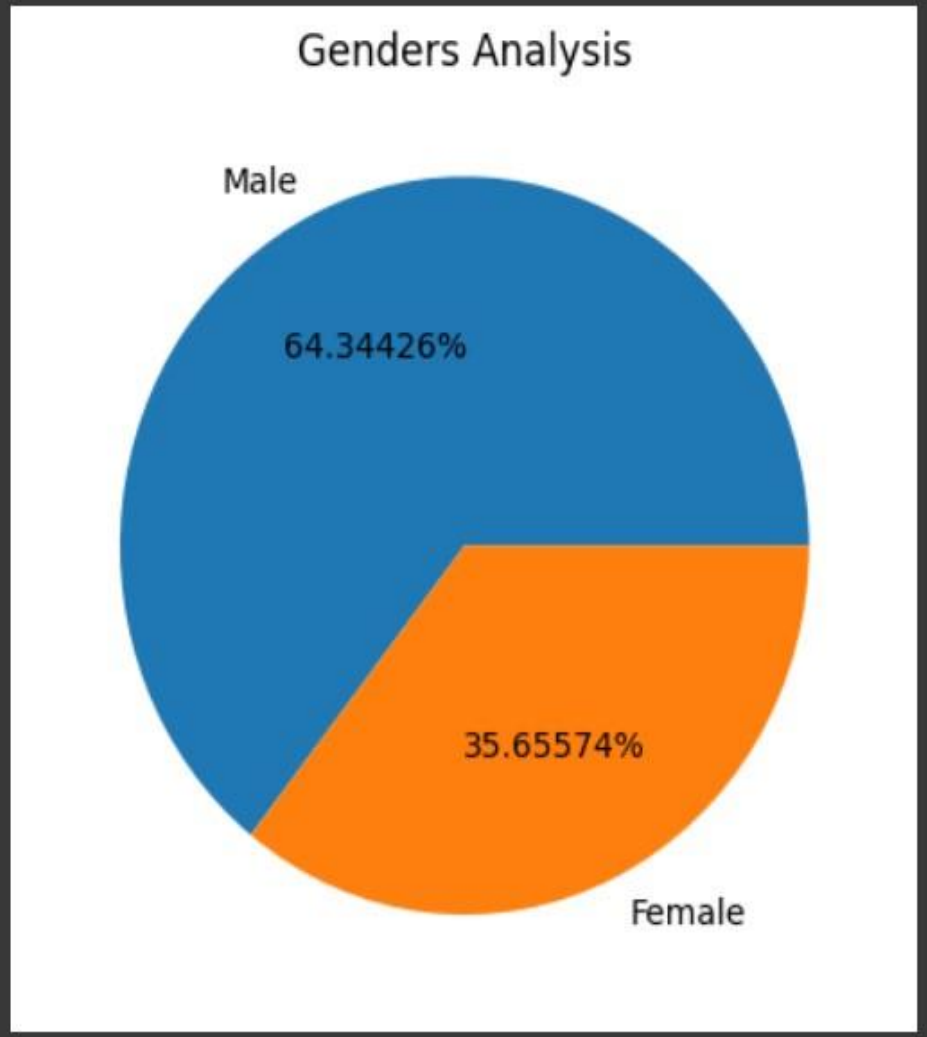


```
avg_tip_by_day = df.groupby('day')['tip'].mean()
plt.bar(avg_tip_by_day.index, avg_tip_by_day)
plt.xlabel('Day')
plt.ylabel('Average Tip')
plt.title('Average Tip by Day')
plt.show()
```

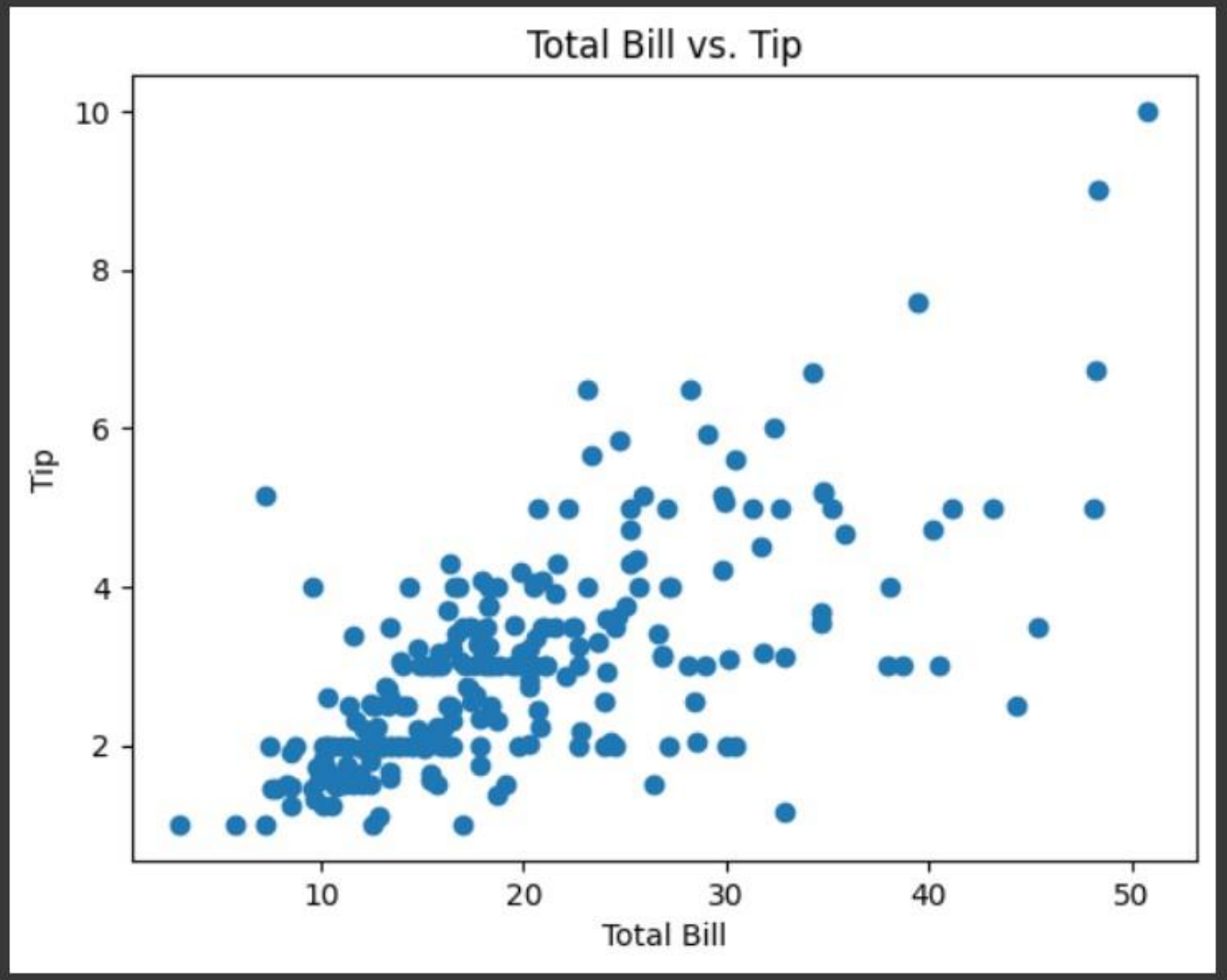


0s

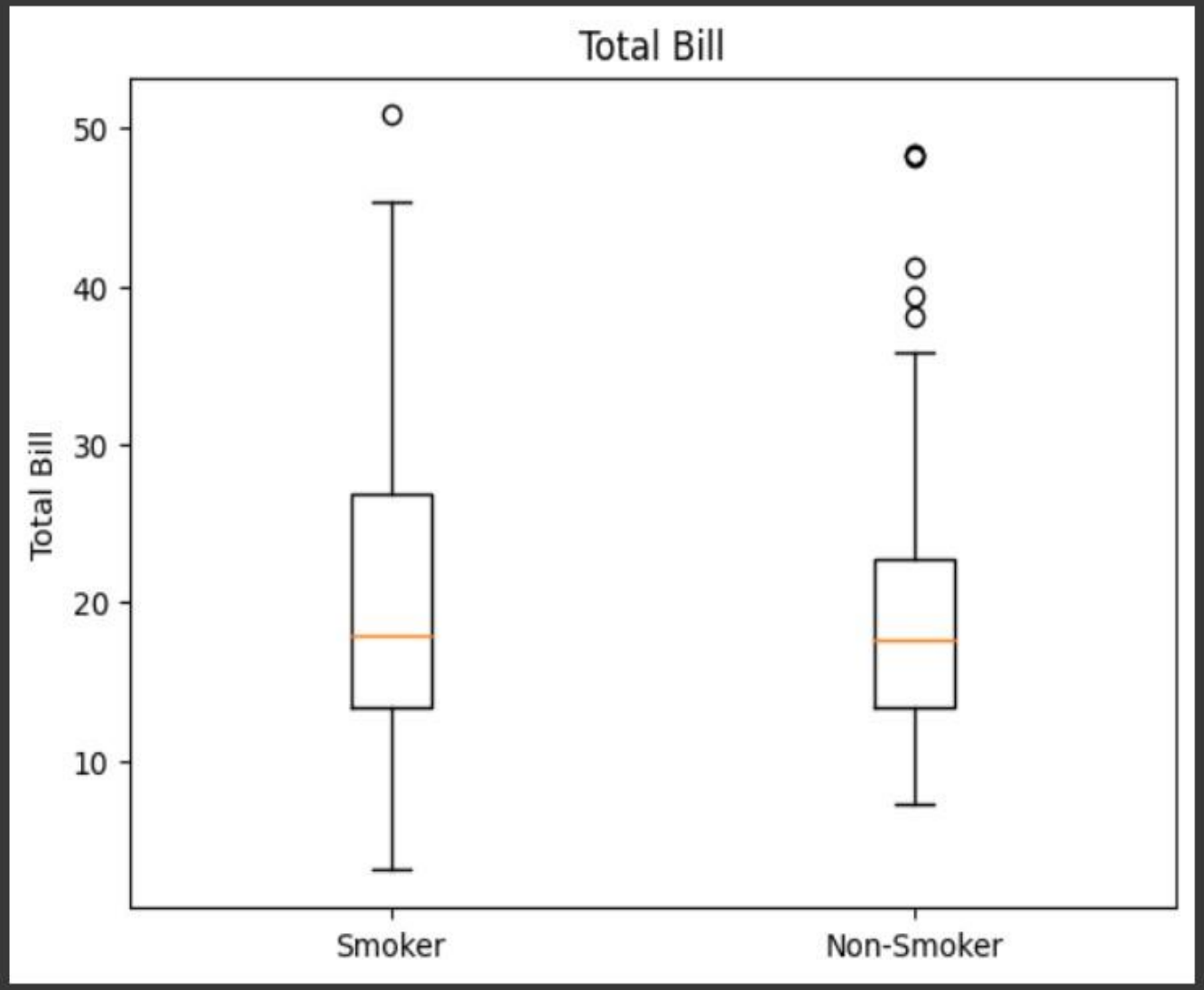
```
gender_counts = df['sex'].value_counts()
plt.pie(gender_counts, labels=gender_counts.index, autopct='%1.5f%%')
plt.title('Genders Analysis')
plt.show()
```



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

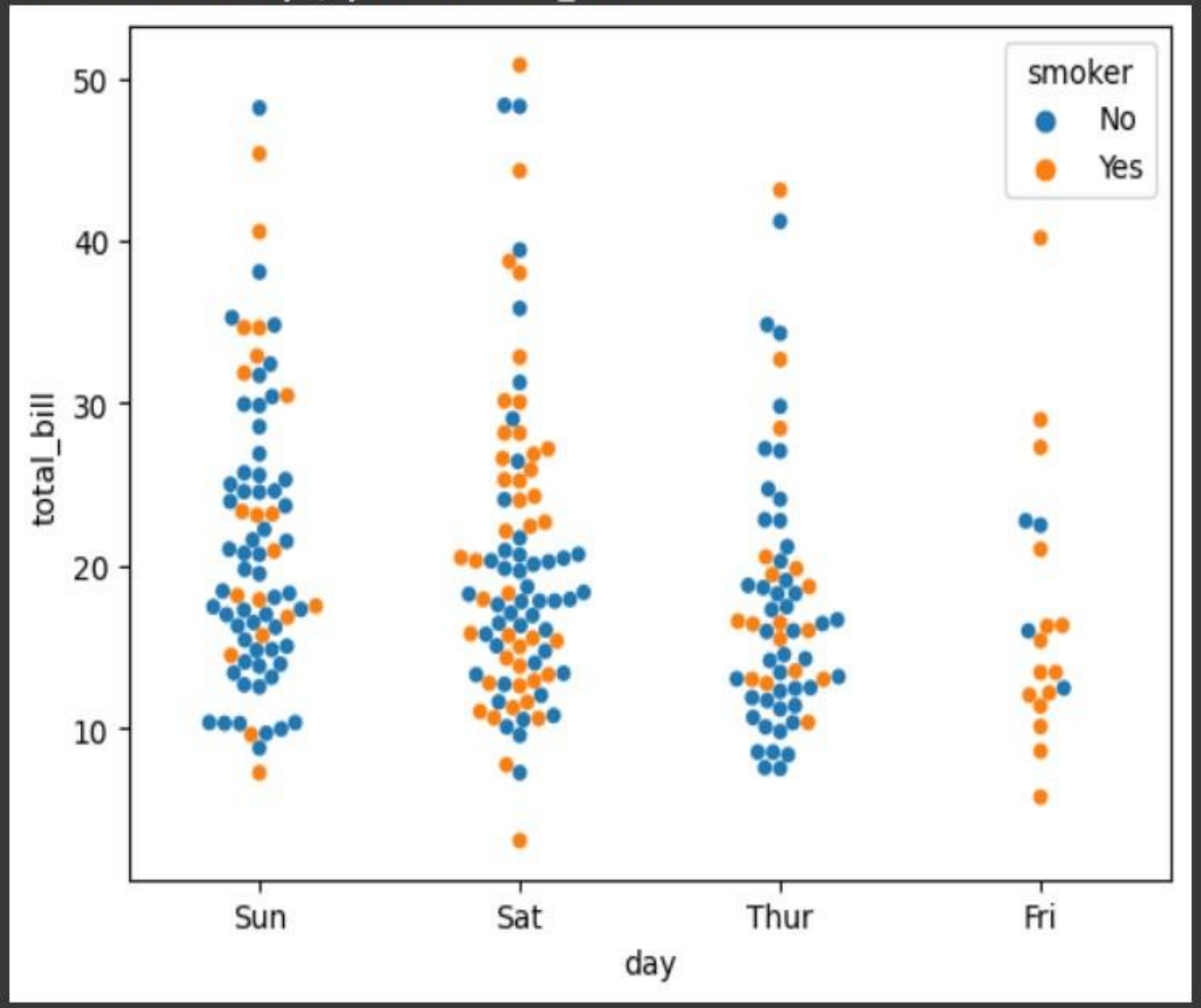


```
plt.boxplot([df[df['smoker'] == 'Yes']['total_bill'], df[df['smoker'] == 'No']['total_bill']], labels=['Smoker', 'Non-Smoker'])
plt.ylabel('Total Bill')
plt.title('Total Bill')
plt.show()
```



```
sns.swarmplot(data=df, x='day', y='total_bill', hue='smoker')
```

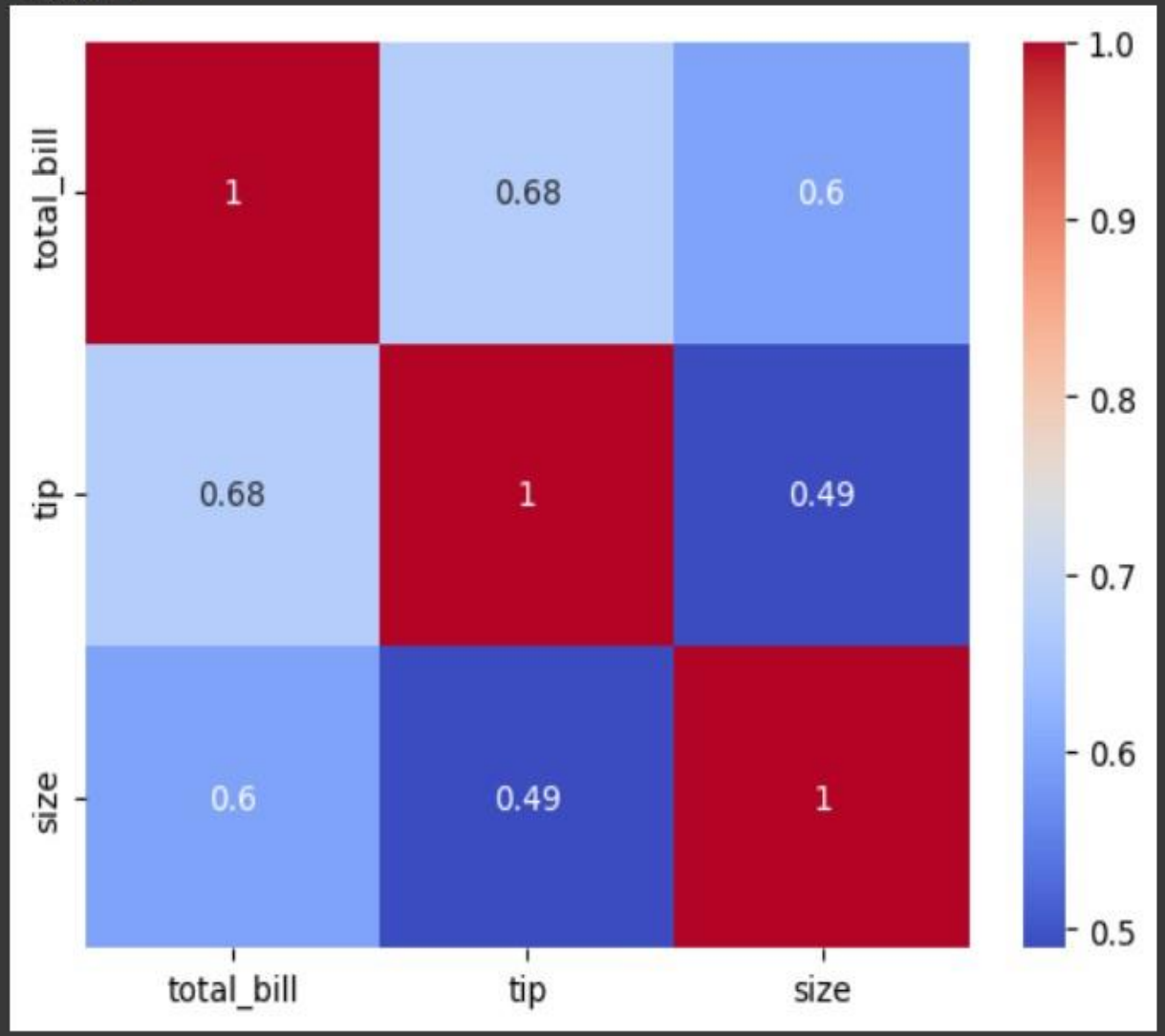
```
<Axes: xlabel='day', ylabel='total_bill'>
```



Line plot of the cumulative sum of total bill by party size

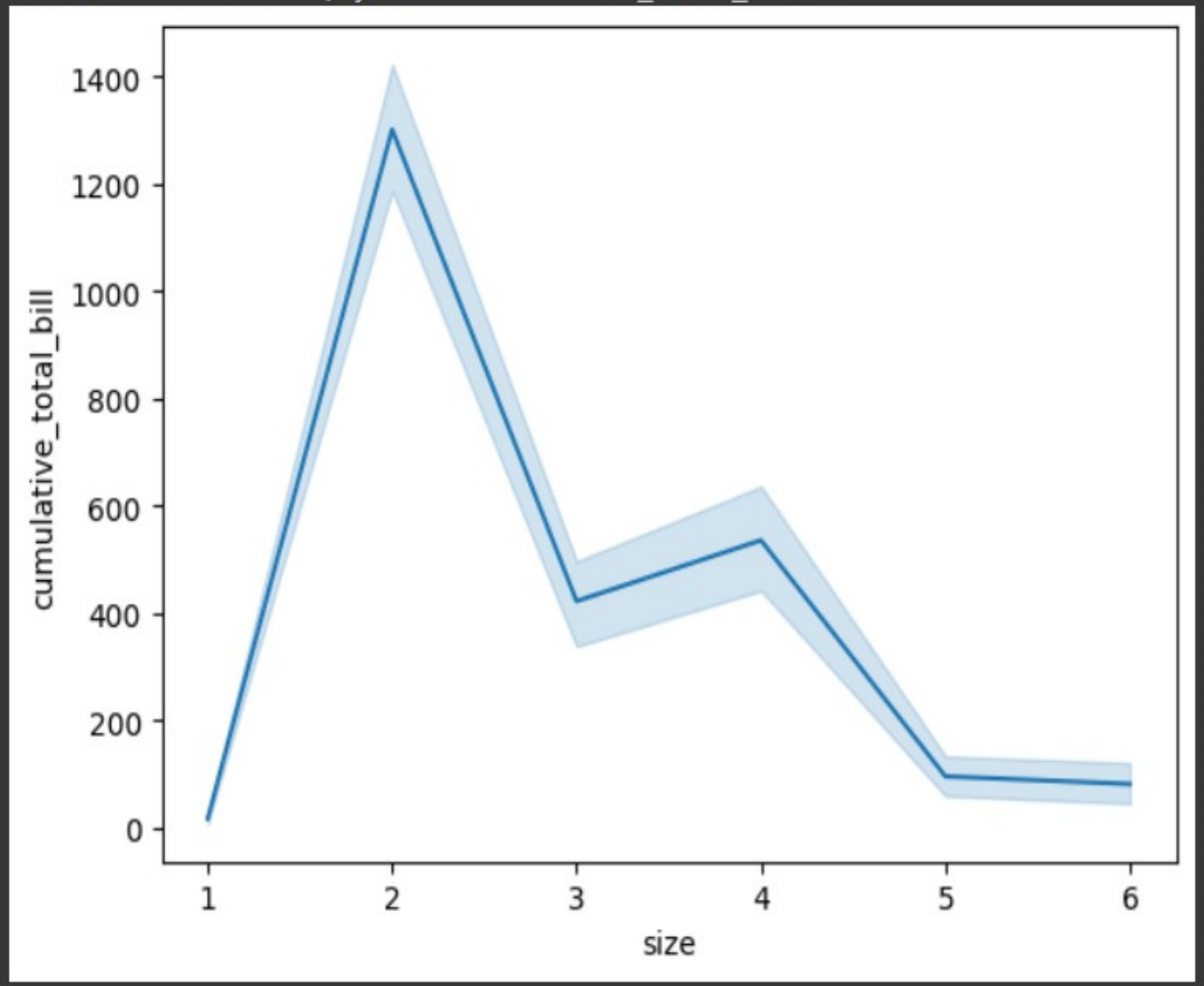
```
numeric_columns = ['total_bill', 'tip', 'size']
correlation_matrix = df[numeric_columns].corr()
sns.heatmap(data=correlation_matrix, annot=True, cmap='coolwarm')
```

<Axes: >



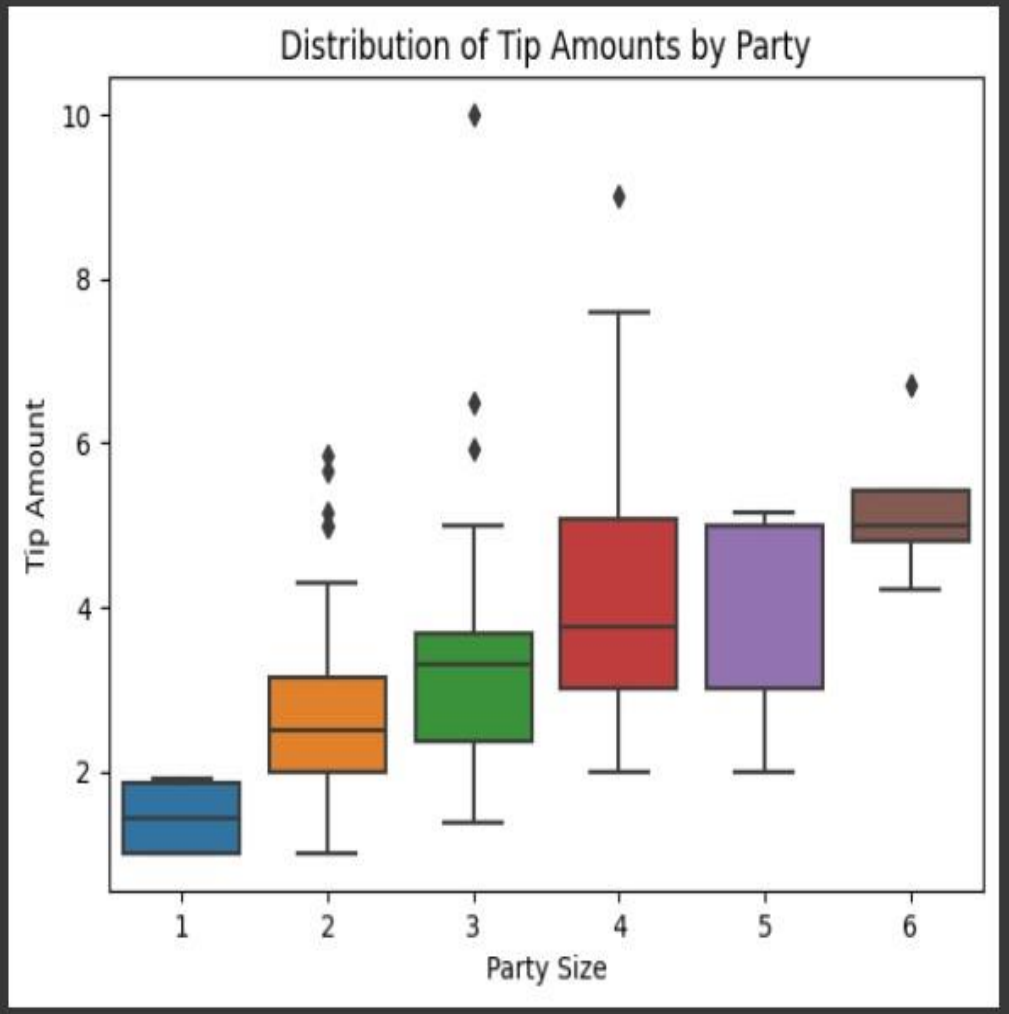
```
cumulative_total_bill = df.groupby('size')['total_bill'].cumsum()
df['cumulative_total_bill'] = cumulative_total_bill
sns.lineplot(data=df, x='size', y='cumulative_total_bill')
```

```
<Axes: xlabel='size', ylabel='cumulative_total_bill'>
```



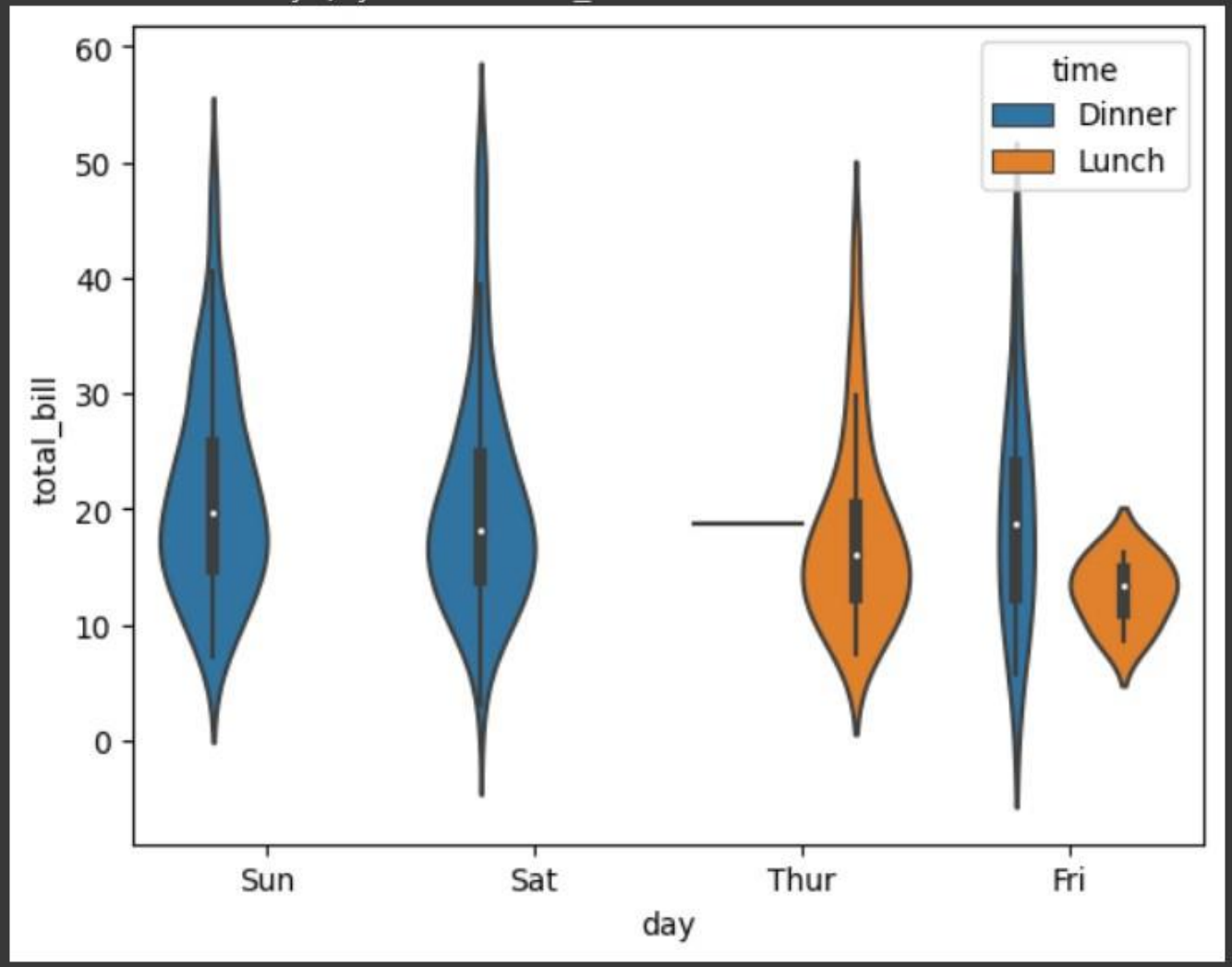
```
import matplotlib.pyplot as plt

sns.boxplot(x='size', y='tip', data=data)
plt.xlabel('Party Size')
plt.ylabel('Tip Amount')
plt.title('Distribution of Tip Amounts by Party')
plt.show()
```



```
import seaborn as sns
sns.violinplot(data=df, x='day', y='total_bill', hue='time')
```

```
<Axes: xlabel='day', ylabel='total_bill'>
```



Heatmap of the correlation matrix between numerical columns