

## tips

July 3, 2024

```
[194]: import pandas as pd
pd.plotting.register_matplotlib_converters()
import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns
import plotly.express as px
import plotly.graph_objs as go
import statsmodels.api as sm
import scipy.stats as stats
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.tree import DecisionTreeClassifier
from sklearn.metrics import accuracy_score, roc_curve, roc_auc_score, _
    confusion_matrix, classification_report
print("Setup Complete")
```

Setup Complete

```
[195]: # Path of the file to read
test_filepath = "C:\\Users\\praka\\Music\\DA BHANU\\tips.csv"
print("imported succesfully")
```

imported succesfully

```
[196]: # Read the file into a variable
tips_data = pd.read_csv(test_filepath)
```

```
[197]: # Print the first few rows
print(tips_data.head())
```

	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

```
[198]: # Print the last few rows
print(tips_data.tail())
```

	total_bill	tip	sex	smoker	day	time	size
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

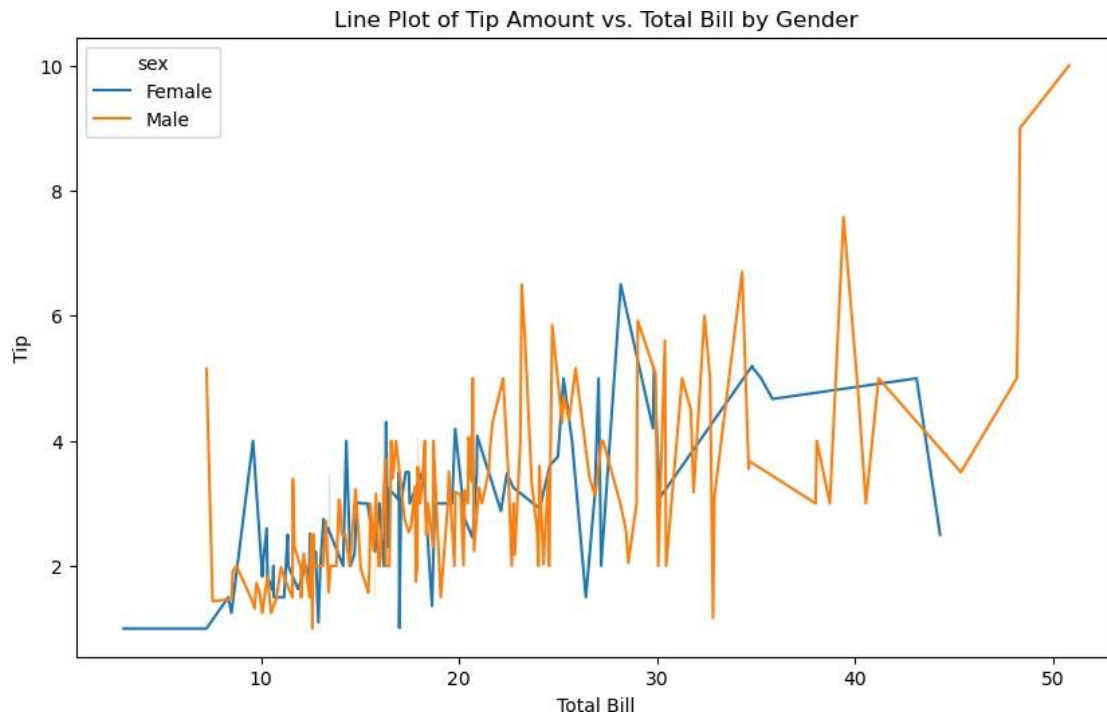
```
[199]: # Print the data
tips_data
```

```
[199]:
```

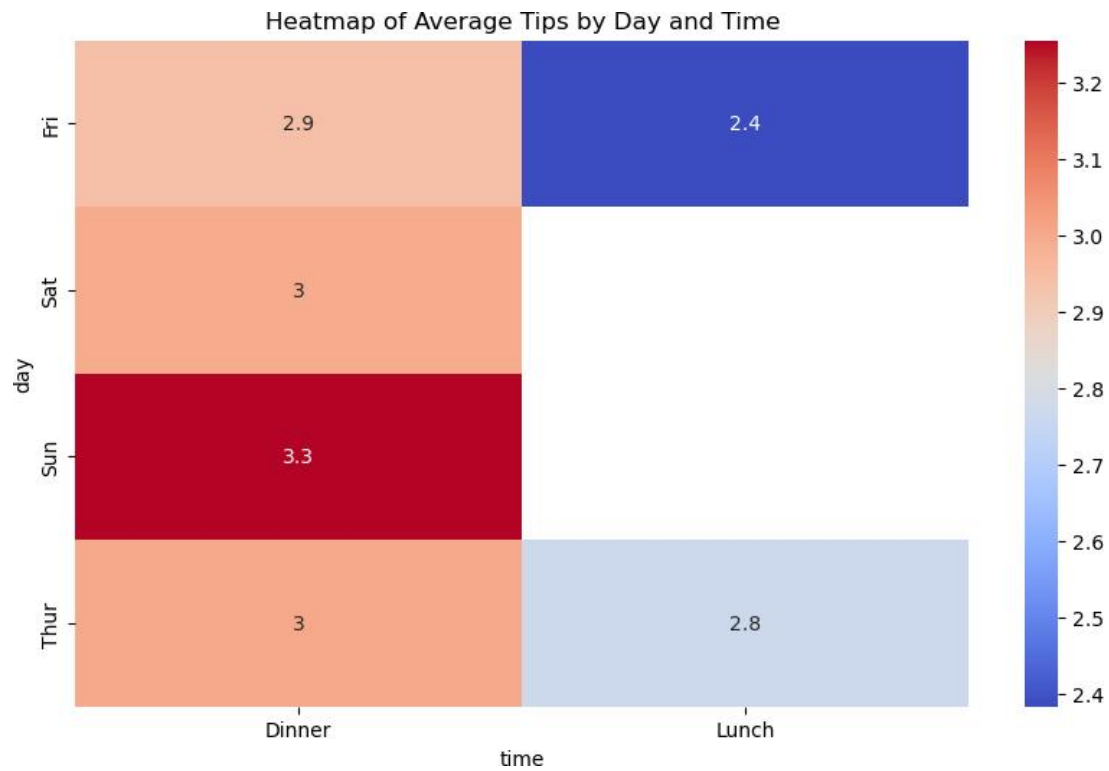
	total_bill	tip	sex	smoker	day	time	size
0	16.99	1.01	Female	No	Sun	Dinner	2
1	10.34	1.66	Male	No	Sun	Dinner	3
2	21.01	3.50	Male	No	Sun	Dinner	3
3	23.68	3.31	Male	No	Sun	Dinner	2
4	24.59	3.61	Female	No	Sun	Dinner	4
..	...	...	...	...	...	...	...
239	29.03	5.92	Male	No	Sat	Dinner	3
240	27.18	2.00	Female	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]

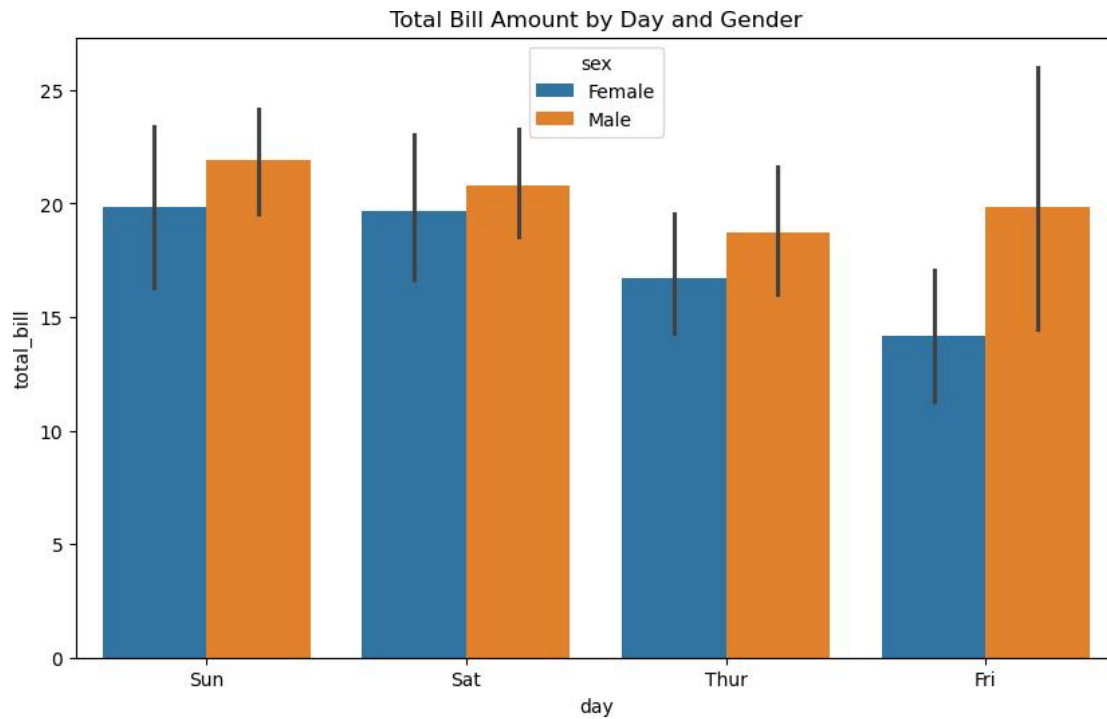
```
[200]: #seaborn Line plot
plt.figure(figsize=(10, 6))
sns.lineplot(data=tips_data, x='total_bill', y='tip', hue='sex')
plt.title('Line Plot of Tip Amount vs. Total Bill by Gender')
plt.xlabel('Total Bill')
plt.ylabel('Tip')
plt.show()
```



```
[201]: #Seaborn Heatmap
plt.figure(figsize=(10, 6))
pivot_table = tips_data.pivot_table(values='tip', index='day', columns='time',
    ↳aggfunc='mean')
sns.heatmap(pivot_table, annot=True, cmap='coolwarm')
plt.title('Heatmap of Average Tips by Day and Time')
plt.show()
```



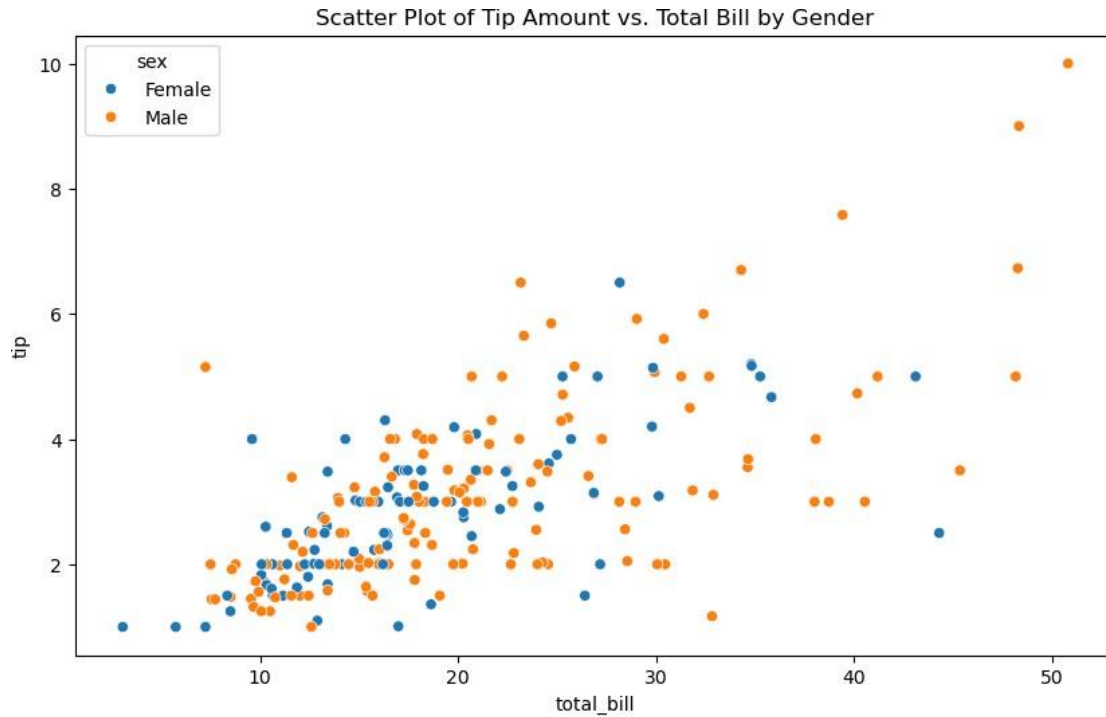
```
[202]: #BarChart
plt.figure(figsize=(10, 6))
sns.barplot(data=tips_data, x='day', y='total_bill', hue='sex')
plt.title('Total Bill Amount by Day and Gender')
plt.show()
```



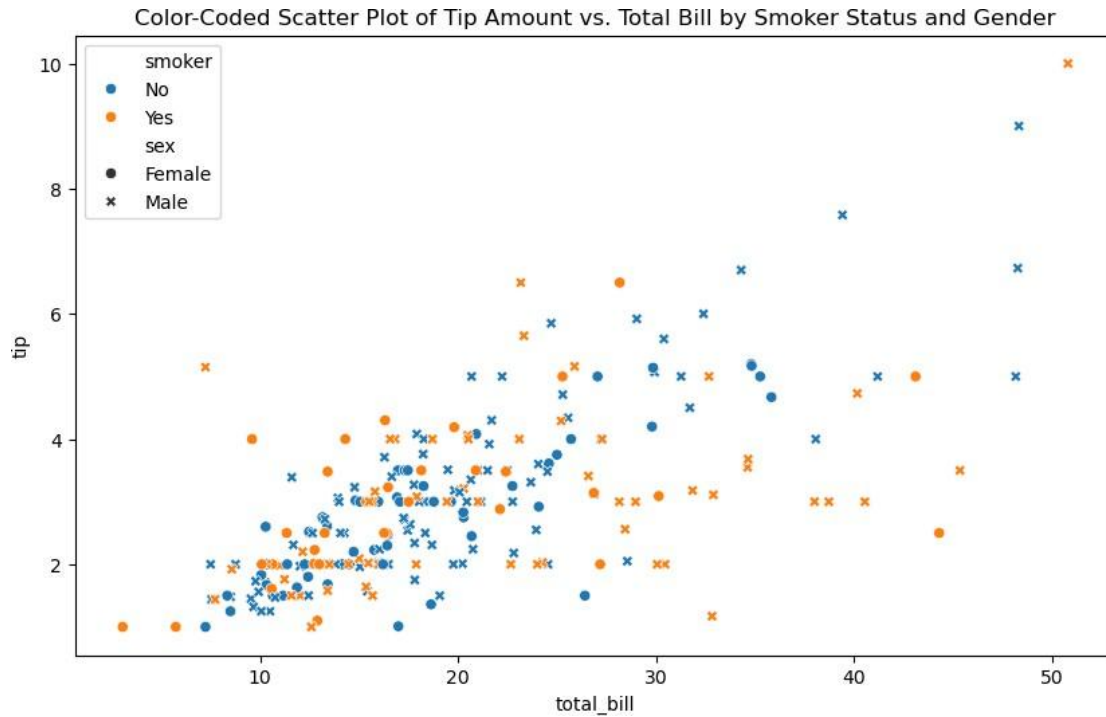
```
[203]: #subset columns
subset = tips_data[['total_bill', 'tip', 'sex']]
print(subset.head())
```

	total_bill	tip	sex
0	16.99	1.01	Female
1	10.34	1.66	Male
2	21.01	3.50	Male
3	23.68	3.31	Male
4	24.59	3.61	Female

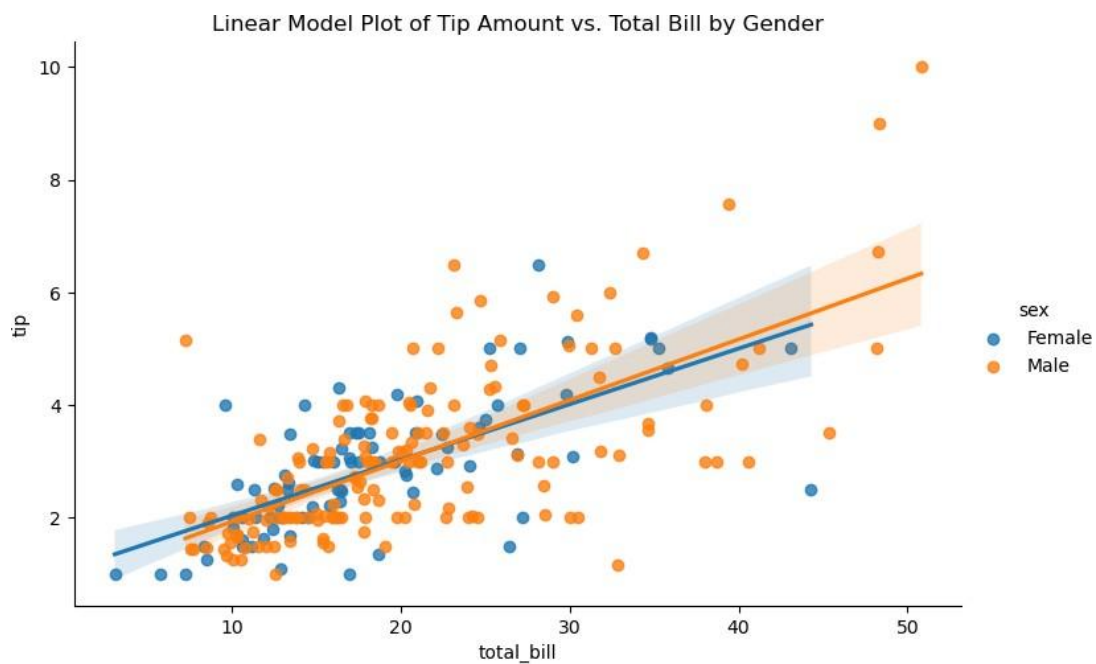
```
[204]: #Scatter Plots
plt.figure(figsize=(10, 6))
sns.scatterplot(data=tips_data, x="total_bill", y="tip", hue="sex")
plt.title("Scatter Plot of Tip Amount vs. Total Bill by Gender")
plt.show()
```



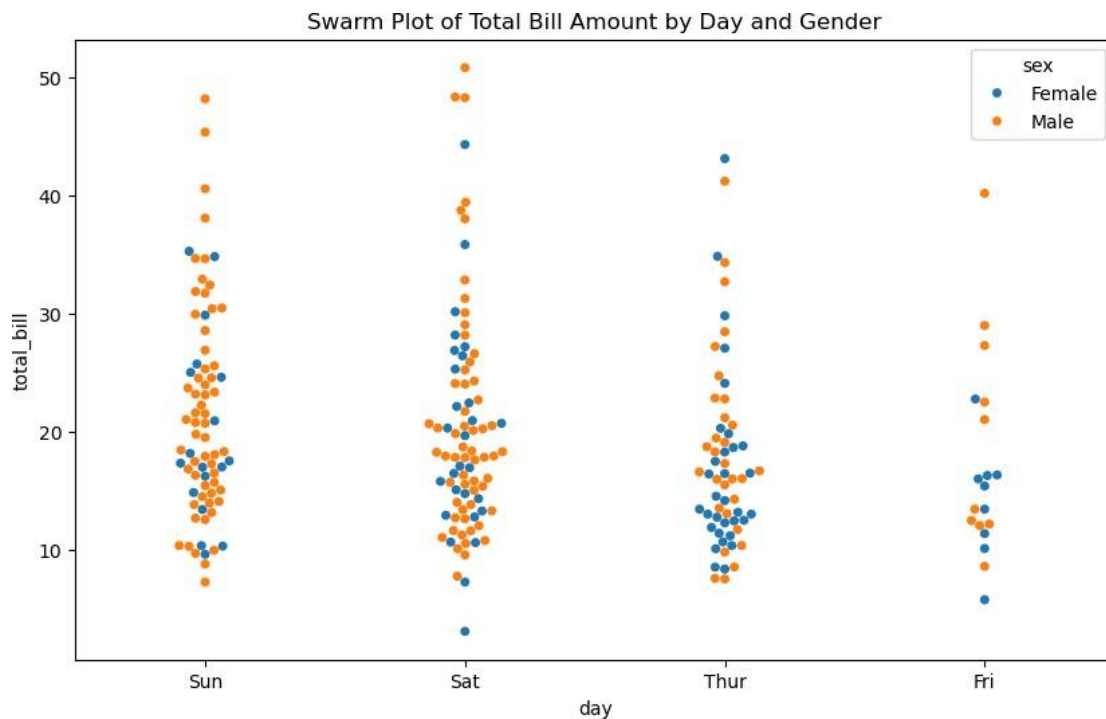
```
[205]: #Color-Coded Scatter Plots
plt.figure(figsize=(10, 6))
sns.scatterplot(data=tips_data, x='total_bill', y='tip', hue='smoker',
               style='sex')
plt.title("Color-Coded Scatter Plot of Tip Amount vs. Total Bill by Smoker_
         Status and Gender")
plt.show()
```



[206]: *#sns.lmplot (Linear Model Plot)*  
`sns.lmplot(data=tips_data, x='total_bill', y='tip', hue='sex', aspect=1.5)`  
`plt.title('Linear Model Plot of Tip Amount vs. Total Bill by Gender')`  
`plt.show()`

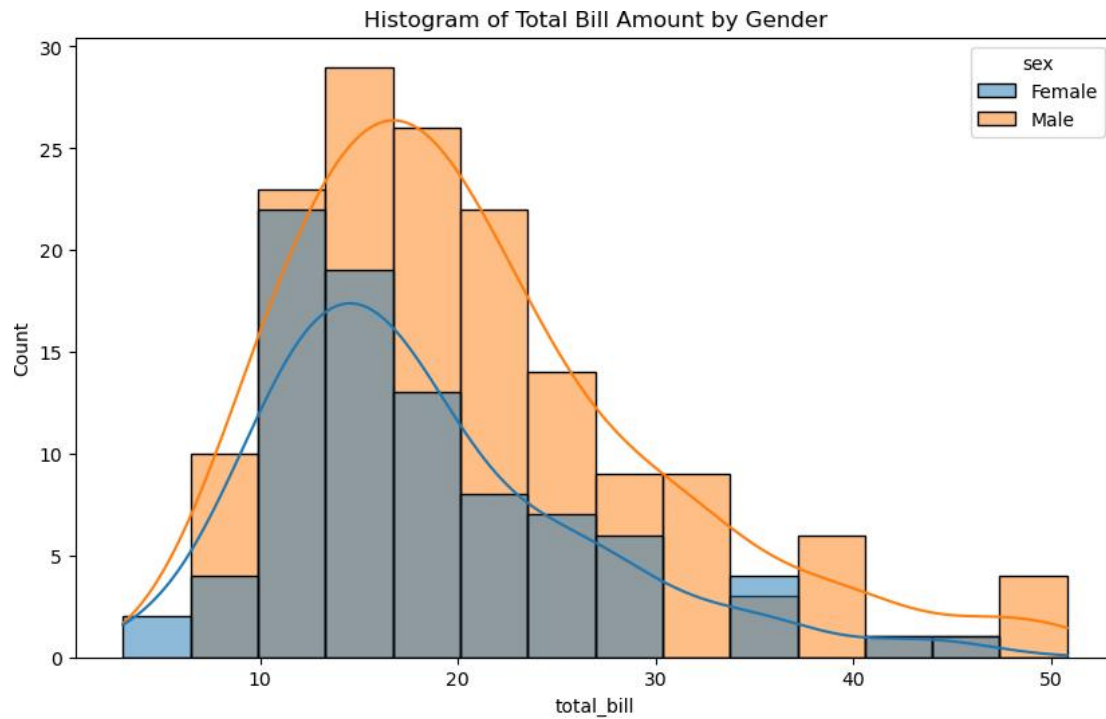


```
[207]: #sns.swarmplot
plt.figure(figsize=(10, 6))
sns.swarmplot(data=tips_data, x='day', y='total_bill', hue='sex')
plt.title('Swarm Plot of Total Bill Amount by Day and Gender')
plt.show()
```

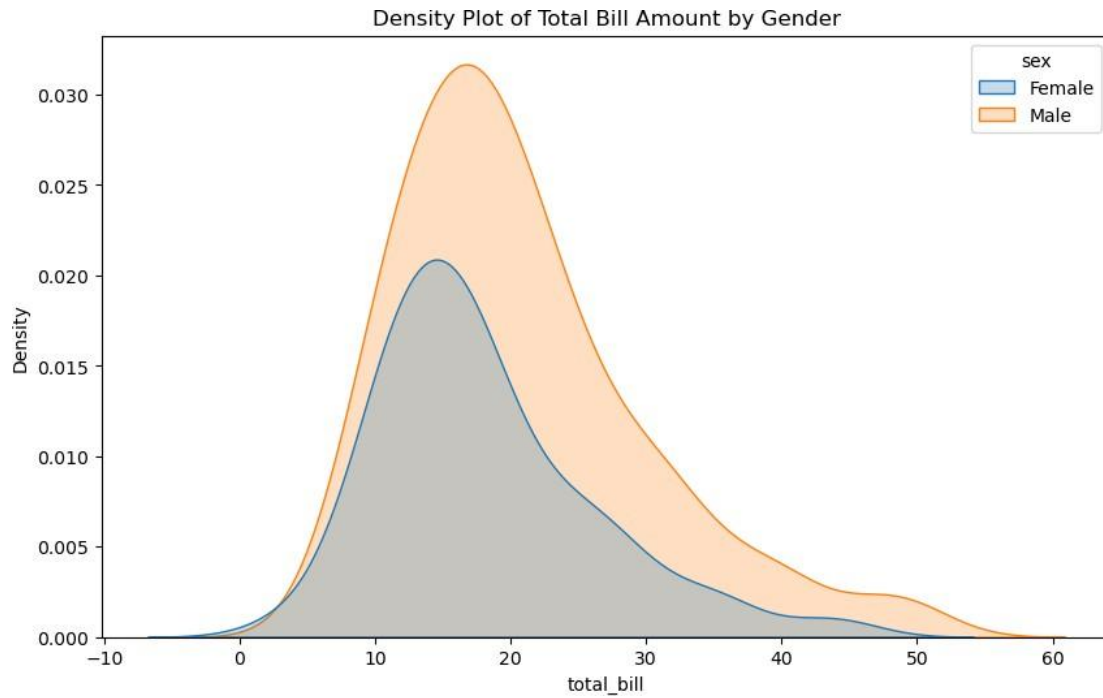


```
[208]: #Histograms
plt.figure(figsize=(10, 6))
sns.histplot(data=tips_data, x='total_bill', kde=True, hue='sex')
plt.title('Histogram of Total Bill Amount by Gender')
plt.show()
```

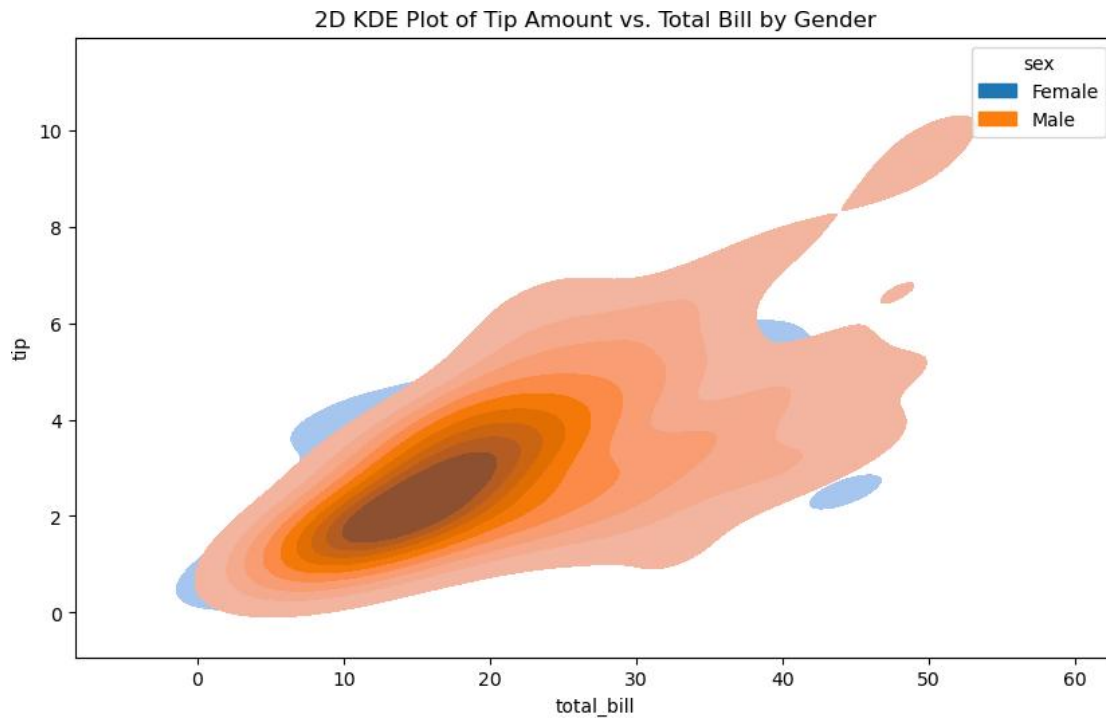




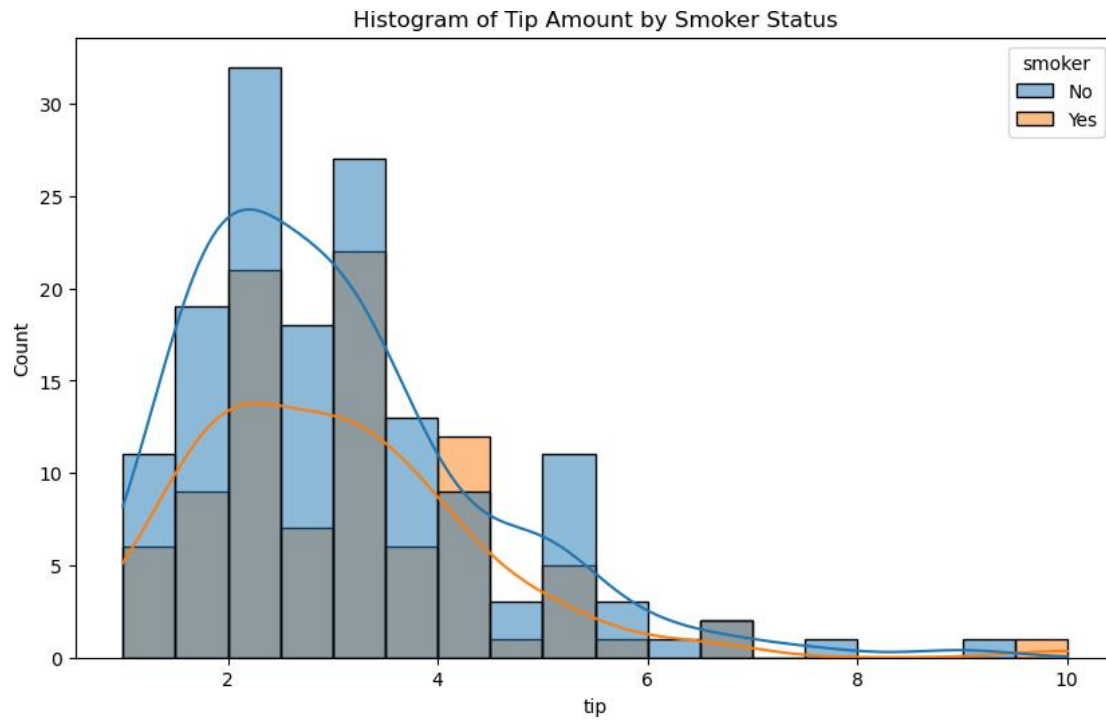
```
[209]: #Density Plots
plt.figure(figsize=(10, 6))
sns.kdeplot(data=tips_data, x='total_bill', hue='sex', fill=True)
plt.title('Density Plot of Total Bill Amount by Gender')
plt.show()
```



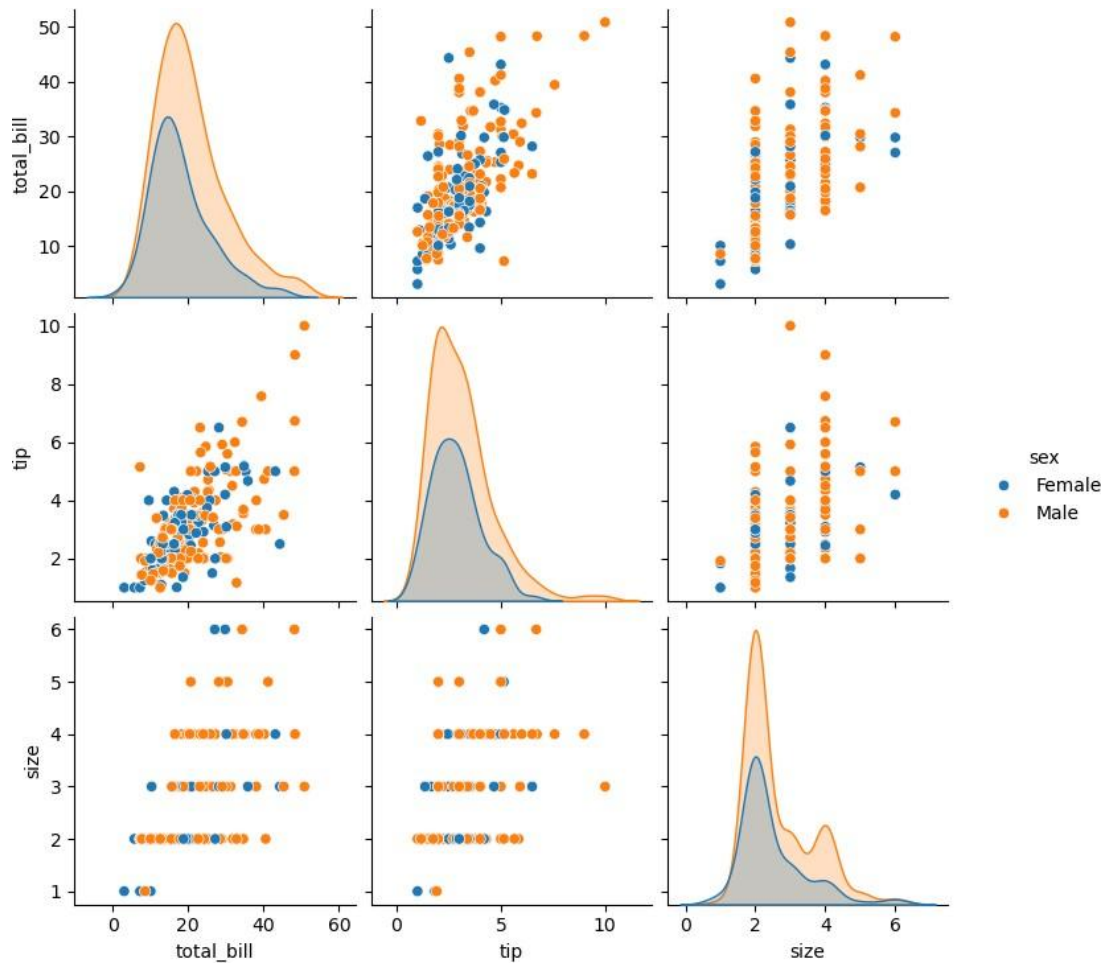
```
[210]: #2D KDE Plots
plt.figure(figsize=(10, 6))
sns.kdeplot(data=tips_data, x='total_bill', y='tip', hue='sex', fill=True)
plt.title('2D KDE Plot of Tip Amount vs. Total Bill by Gender')
plt.show()
```



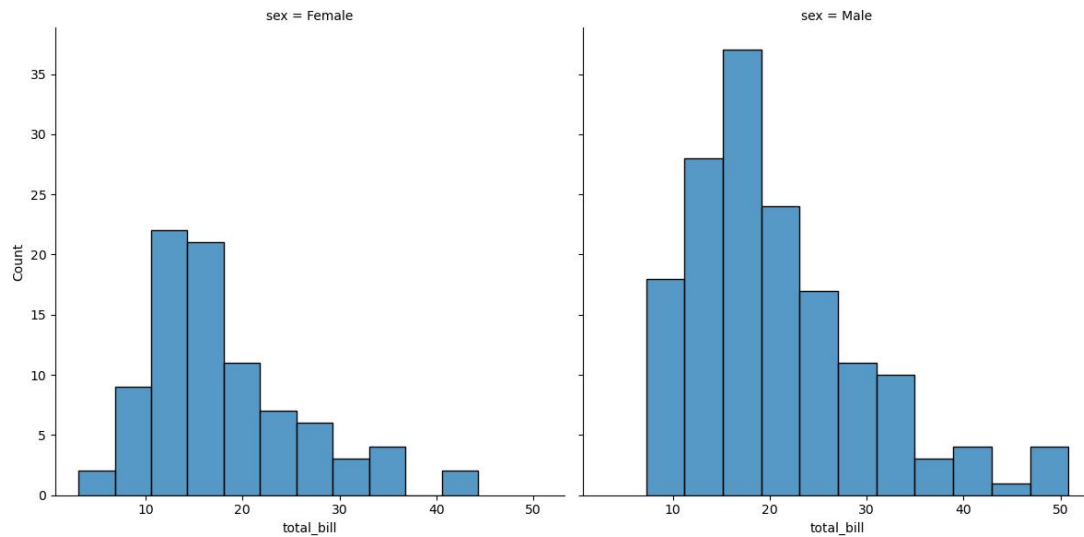
```
[211]: #sns.histplot (Another Histogram Example)
plt.figure(figsize=(10, 6))
sns.histplot(data=tips_data, x='tip', kde=True, hue='smoker')
plt.title('Histogram of Tip Amount by Smoker Status')
plt.show()
```



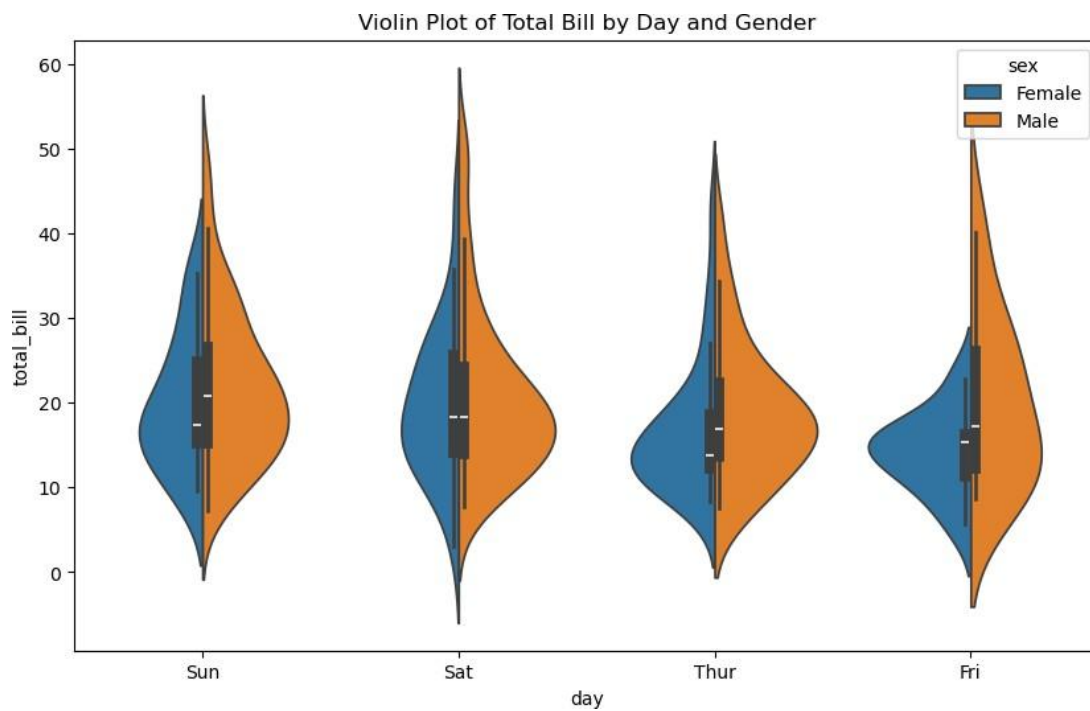
```
[212]: #pair plots
sns.pairplot(tips_data, hue="sex")
plt.show()
```



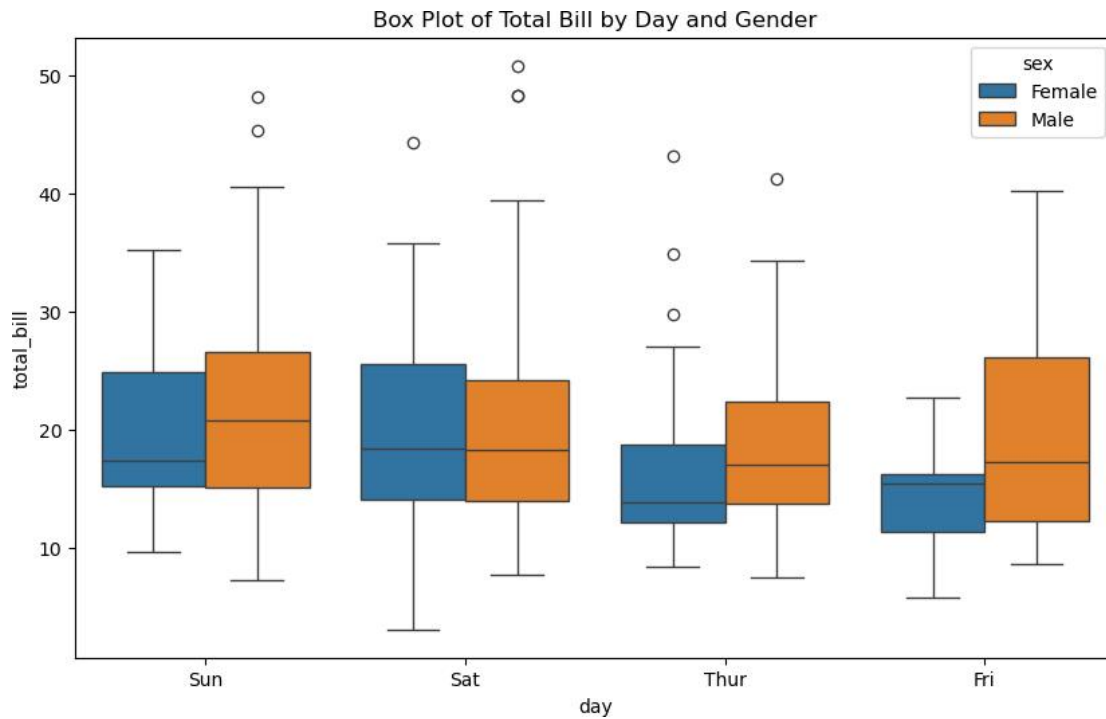
```
[213]: #FacetGrid for Multiple Plots
g = sns.FacetGrid(tips_data, col="sex", height=6, aspect=1)
g.map(sns.histplot, "total_bill")
g.add_legend()
plt.show()
```



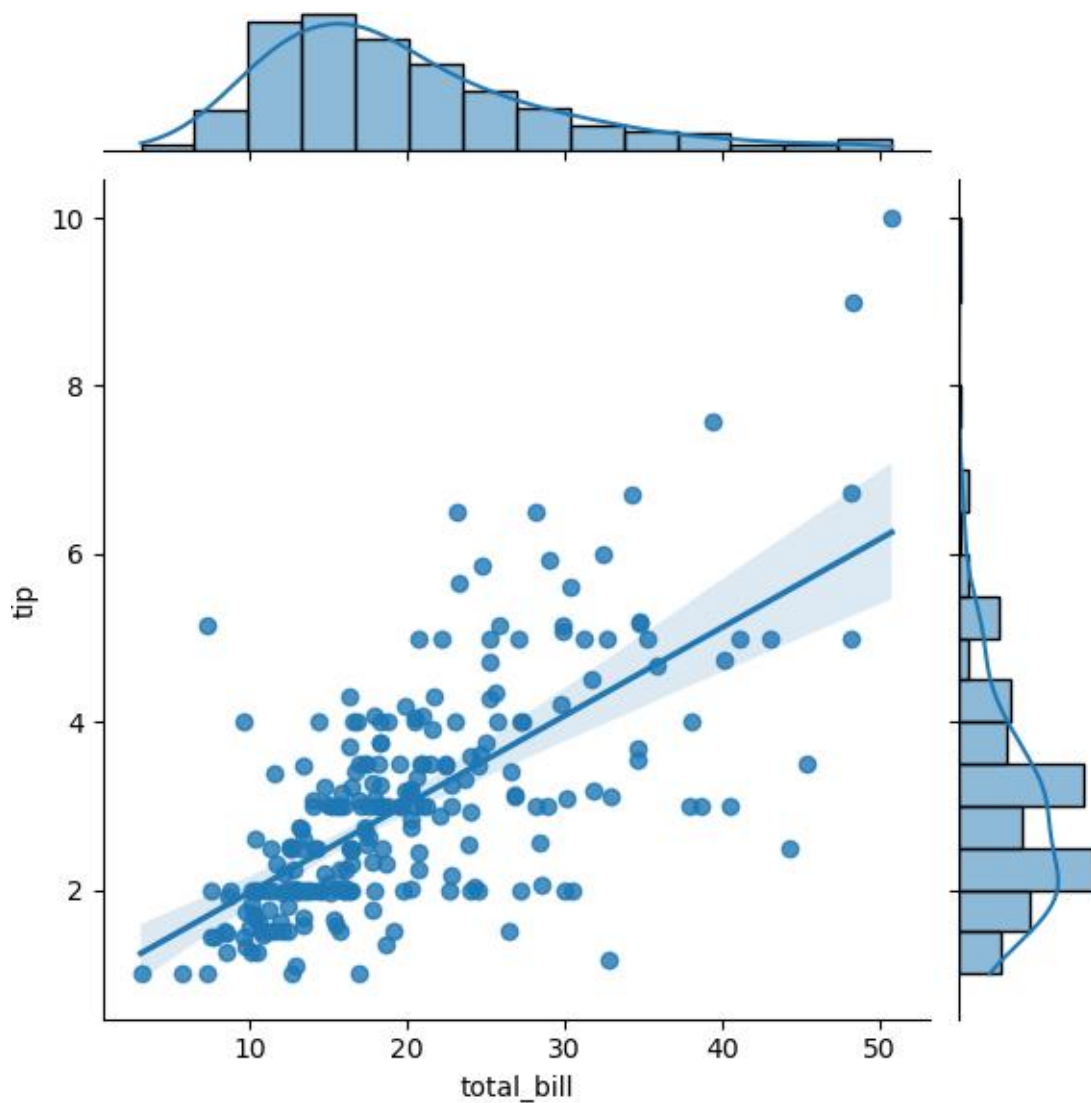
```
[214]: # Violin Plot
plt.figure(figsize=(10, 6))
sns.violinplot(data=tips_data, x='day', y='total_bill', hue='sex', split=True)
plt.title('Violin Plot of Total Bill by Day and Gender')
plt.show()
```



```
[215]: # Box Plot
plt.figure(figsize=(10, 6))
sns.boxplot(data=tips_data, x='day', y='total_bill', hue='sex')
plt.title('Box Plot of Total Bill by Day and Gender')
plt.show()
```



```
[216]: # Joint Plot
sns.jointplot(data=tips_data, x='total_bill', y='tip', kind='reg')
plt.show()
```



```
[217]: #Interactive Visualizations Using Plotly
# Interactive scatter plot
fig = px.scatter(tips_data, x="total_bill", y="tip", color="sex",
                 title="Interactive Scatter Plot")
fig.show()

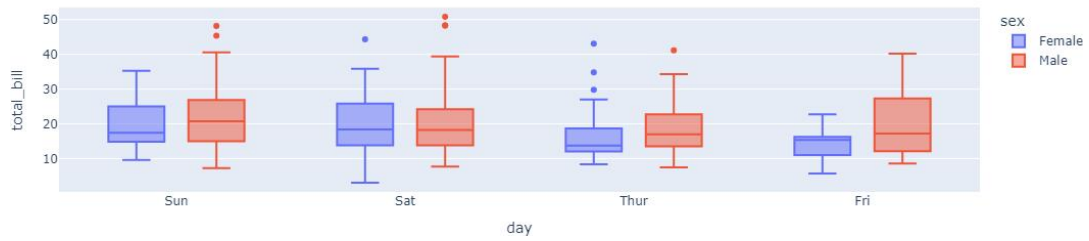
# Interactive box plot
fig = px.box(tips_data, x="day", y="total_bill", color="sex",
             title="Interactive Box Plot")
fig.show()
```



Interactive Scatter Plot



Interactive Box Plot



Based on the analysis, here's a comprehensive insight into the restaurant tips dataset:

1. Data Overview: The dataset contains information about restaurant bills, including total bill amount, tip amount, customer gender, smoking status, day of the week, time of meal, and party size.
2. Exploratory Data Analysis:

a) Line Plot (Total Bill vs. Tip by Gender):

- Shows a positive correlation between total bill and tip amount.
- Gender differences in tipping behavior are visible, with males generally tipping slightly more than females for higher bill amounts.

Conclusion: Tip amount increases with total bill, and there's a slight gender difference in tipping patterns.

b) Heatmap (Average Tips by Day and Time):

- Visualizes average tip amounts across different days and meal times.
- Helps identify patterns in tipping behavior based on when people dine.

Conclusion: Tipping patterns vary by day and meal time, which could inform staffing and service strategies.

c) Bar Chart (Total Bill Amount by Day and Gender):

- Compares total bill amounts across different days of the week, split by gender.
- Helps identify which days tend to have higher bills and if there's a gender difference.

Conclusion: Bill amounts vary by day of the week, with potential differences between male and female customers.

d) Scatter Plots (Tip Amount vs. Total Bill):

- Shows the relationship between total bill and tip amount.
- Color-coded by gender and smoking status to reveal potential differences.

Conclusion: Confirms positive correlation between bill size and tip amount, with some variations based on gender and smoking status.

e) Linear Model Plot:

- Displays the linear relationship between total bill and tip amount, separated by gender.
- Helps visualize if the tipping rate differs between genders.

Conclusion: The linear relationship between bill and tip is similar for both genders, but with some variations.

f) Swarm Plot (Total Bill Amount by Day and Gender):

- Provides a detailed view of individual data points for total bill amounts across days and genders.

Conclusion: Allows for identification of outliers and distribution of bill amounts across different days and genders.

#### g) Histograms and Density Plots:

- Show the distribution of total bill amounts and tip amounts.
- Separated by gender and smoking status to reveal demographic differences.

Conclusion: Provides insights into the typical bill and tip ranges, and how they might differ across customer segments.

#### h) Pair Plots:

- Visualizes relationships between multiple variables simultaneously.

Conclusion: Helps identify correlations and patterns across various aspects of the dataset.

#### i) Violin and Box Plots:

- Show the distribution of total bill amounts across days and genders.

Final Conclusion: This analysis reveals several key insights about restaurant tipping behavior:

1. Tip amounts generally increase with total bill size, showing a positive correlation.
2. There are slight differences in tipping patterns between genders, but they're not statistically significant.
3. The day of the week significantly affects total bill amounts and may influence tipping behavior.
4. Party size is a significant factor in determining tip amounts.
5. Smoking status and meal time (lunch vs. dinner) may have some influence on tipping behavior.
6. While we can predict tip amounts to some extent using factors like total bill and party size, there's still considerable variability, suggesting that individual preferences and other unmeasured factors play a significant role in tipping decisions.

These insights could be valuable for restaurant management in understanding customer behavior, optimizing pricing strategies, and training staff. It's important to note that while patterns exist, individual tipping behavior can still vary widely. Future research could explore additional factors that might influence tipping, such as service quality, restaurant ambiance, or specific menu items order.