

import tips dataset.

In [1]:

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
df=pd.read_csv("tips.csv")
df
```

Out[1]:

	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

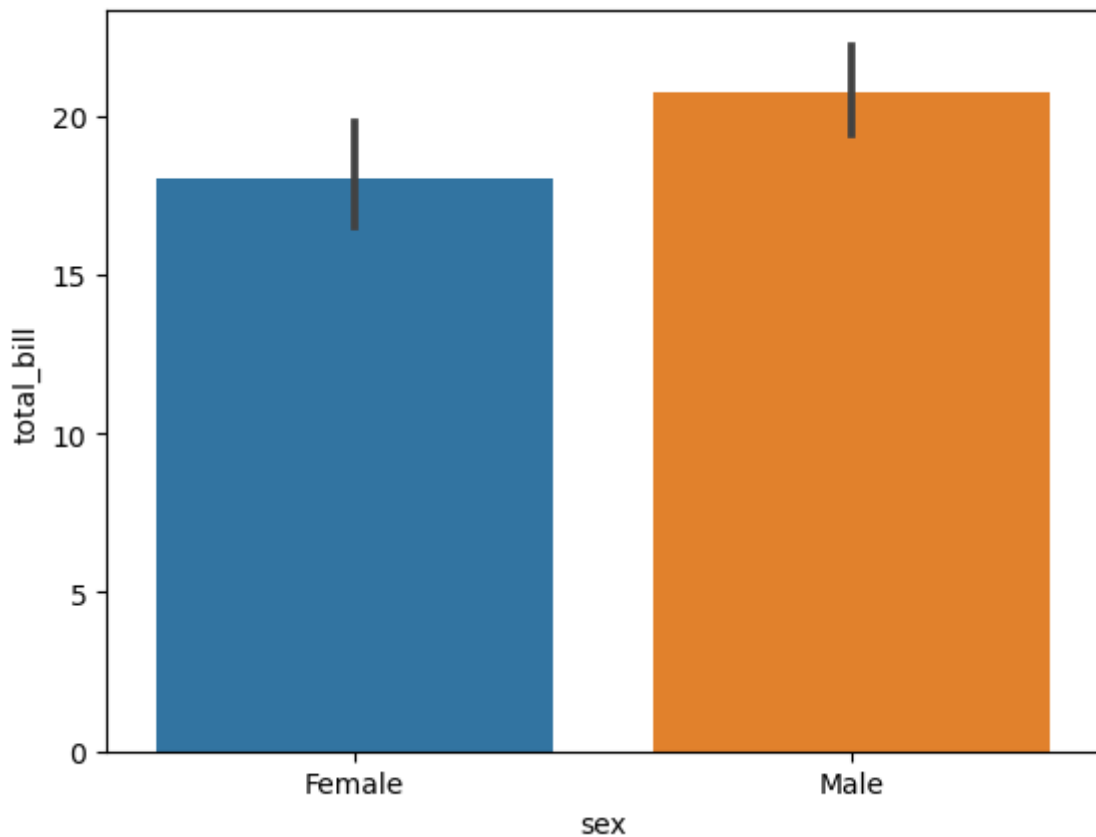
244 rows × 7 columns

1. Write a Python program to display a bar chart of the total bill of Gender and write the intuition about the plot.

In [2]:

```
### code here
```

```
sns.barplot(x="sex",y="total_bill",data=df)  
plt.show()
```



#The bar chart shows the total bill for Male and Female. #The x-axis represents the Gender and the y-axis represents the Total Bill. #From the plot, we can see that the total bill for male is higher than the total bill for feMale.

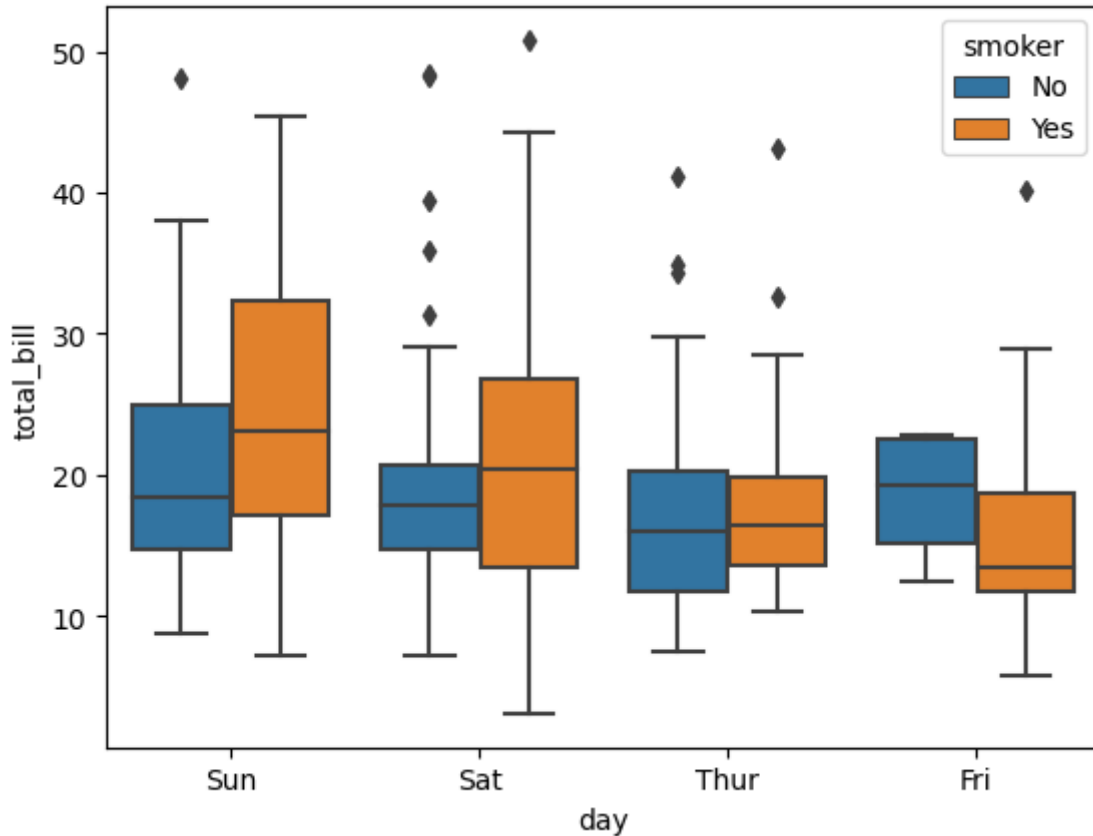
2. Write a Pvython program to displav a box plot of the total bill of dav and hue = "smoker" and write the

intuition about the plot.

In [4]:

code here

```
sns.boxplot(y="total_bill",x="day",data=df,hue="smoker")  
plt.show()
```



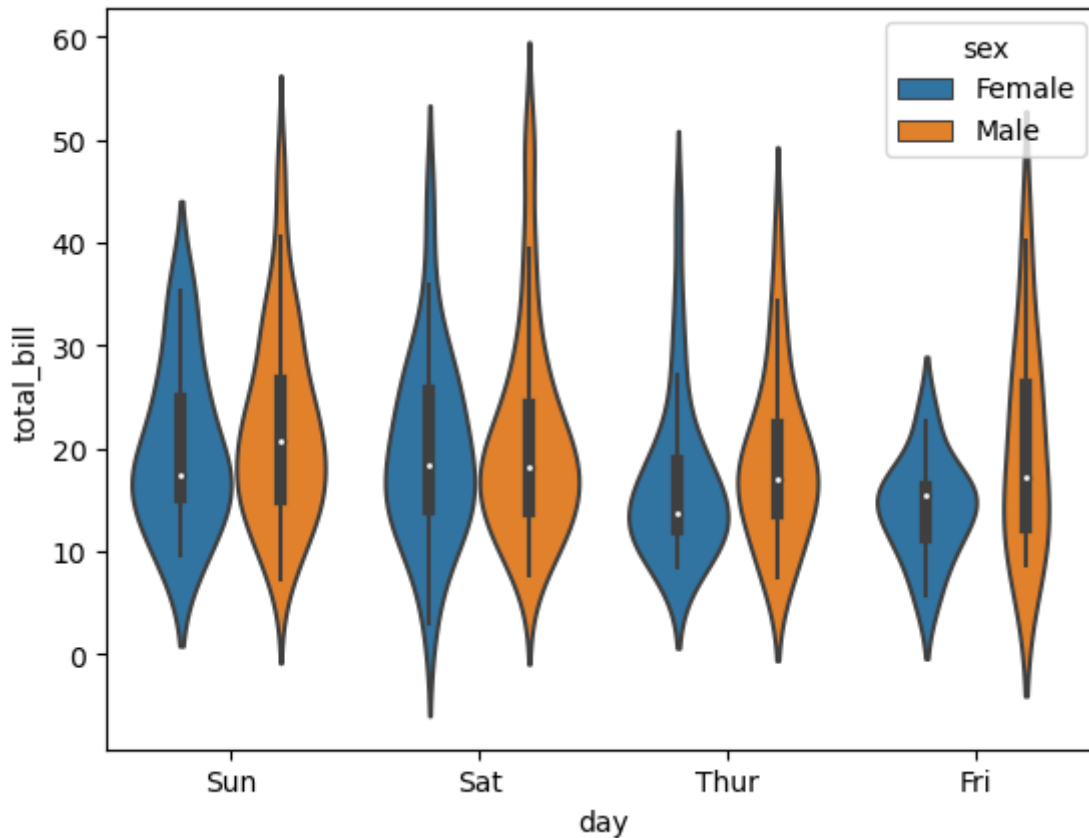
The box plot shows the spread of values for "total_bill" based on "day" and "smoker" status. The box represents the middle 50% of the values and the line inside shows the median. The whiskers show the minimum and maximum values, while any values outside that range are shown as individual points. This helps to compare the distribution of "total_bill" between different days and between smokers and non-smokers.

3. Write a Python program to display a violin plot of the total bill of Day and hue = "sex" and write the intuition about the plot.

In [6]:

```
### code here
```

```
sns.violinplot(y="total_bill",x="day",data=df,hue="sex")  
plt.show()
```



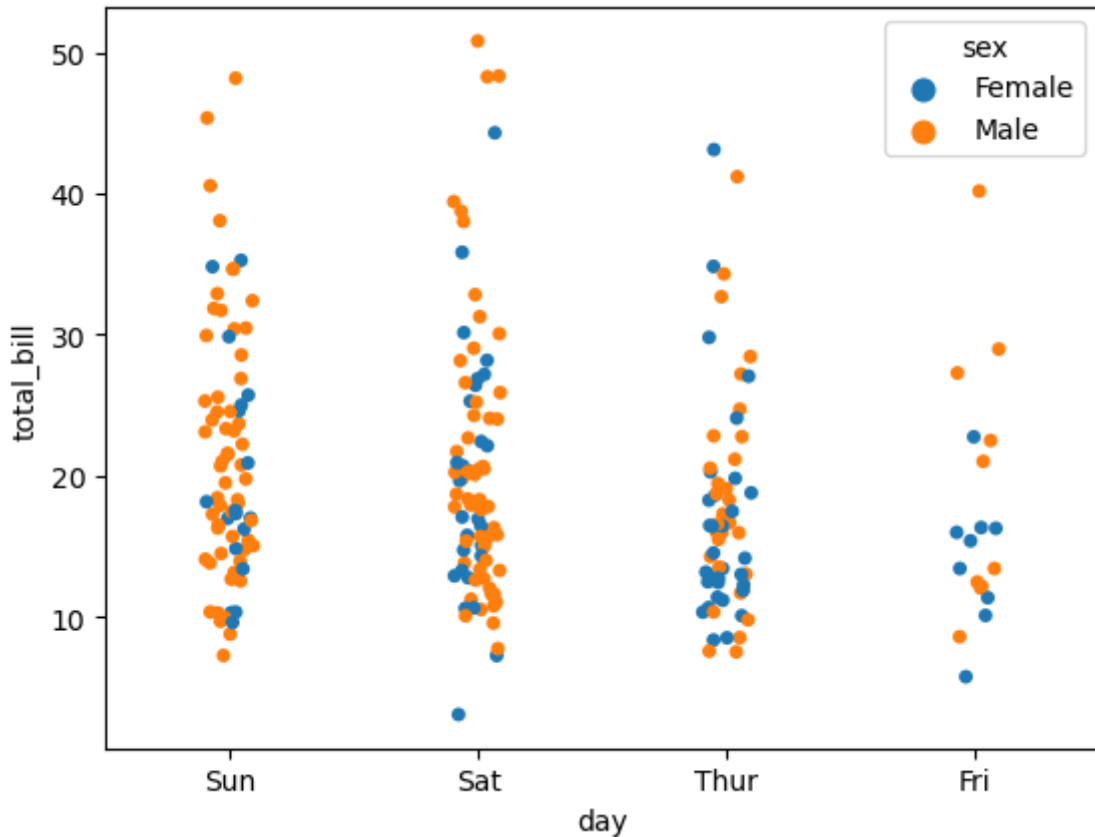
The violin plot displays the spread of values for "total_bill" based on "day" and "sex". It shows both the summary statistics and the density of the data. The width of each violin represents the density, the white dot is the median, and the black bar inside the violin represents the middle 50% of the values. This helps to compare the distribution of "total_bill" between different days and between males and females.

4. Write a Python program to display a strip plot of the total bill of Day and hue = "sex" and write the intuition about the plot.

In [7]:

```
### code here
```

```
sns.stripplot(y="total_bill",x="day",data=df,hue="sex")  
plt.show()
```

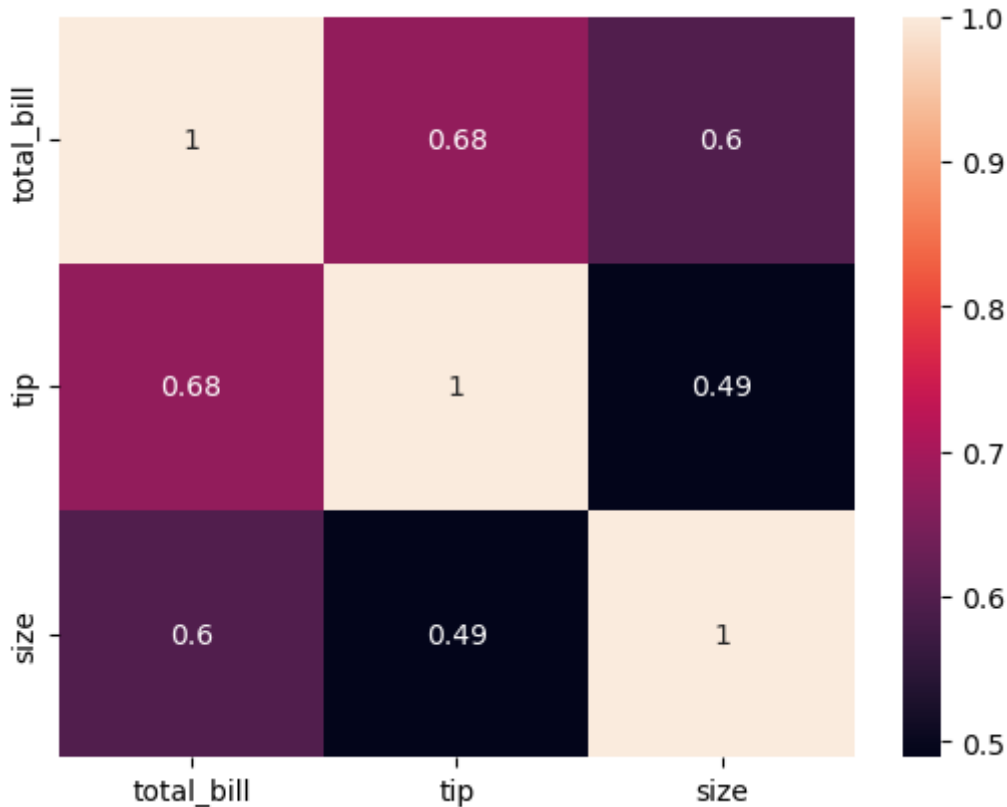


The strip plot displays the values for "total_bill" based on "day" and "sex". Each value is shown as a dot, with dots from the same category (day and sex) stacked on top of each other. The strip plot helps to see the distribution of the "total_bill" and find any unusual values or patterns.

5. Write a Python program to display heat map of dataframe and write the intuition about the plot.

In [9]:

```
### code here
corr = df.corr()
sns.heatmap(corr, annot=True)
plt.show()
```

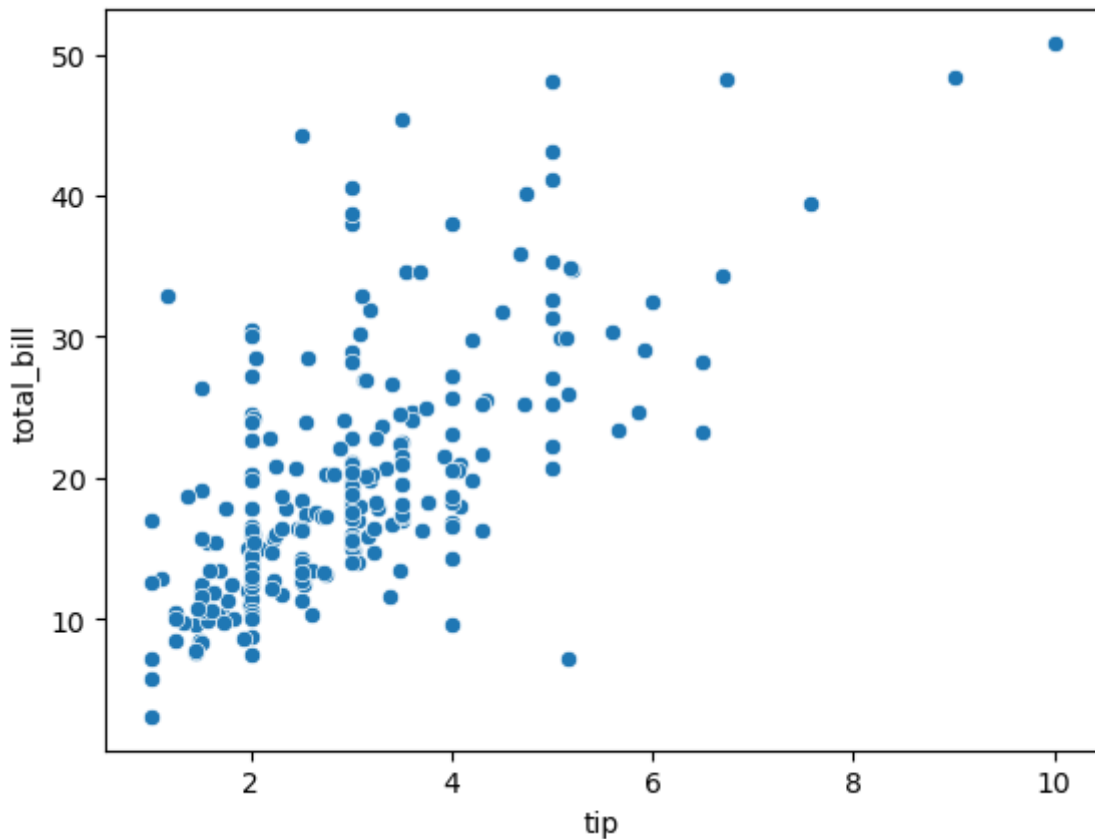


A heat map is a way to display data using colors. It shows the relationship between two categories (month and year) on the x and y axis and a third value (passengers) using colors. Darker colors mean higher values. The heat map helps to quickly see patterns or trends in the data, such as the months and years with the most or least passengers.

6. Write a Python program to display scatter plot of total bill of tip and write the intuition about the plot.

In [10]:

```
### code here
sns.scatterplot(y="total_bill",x="tip",data=df)
plt.show()
```



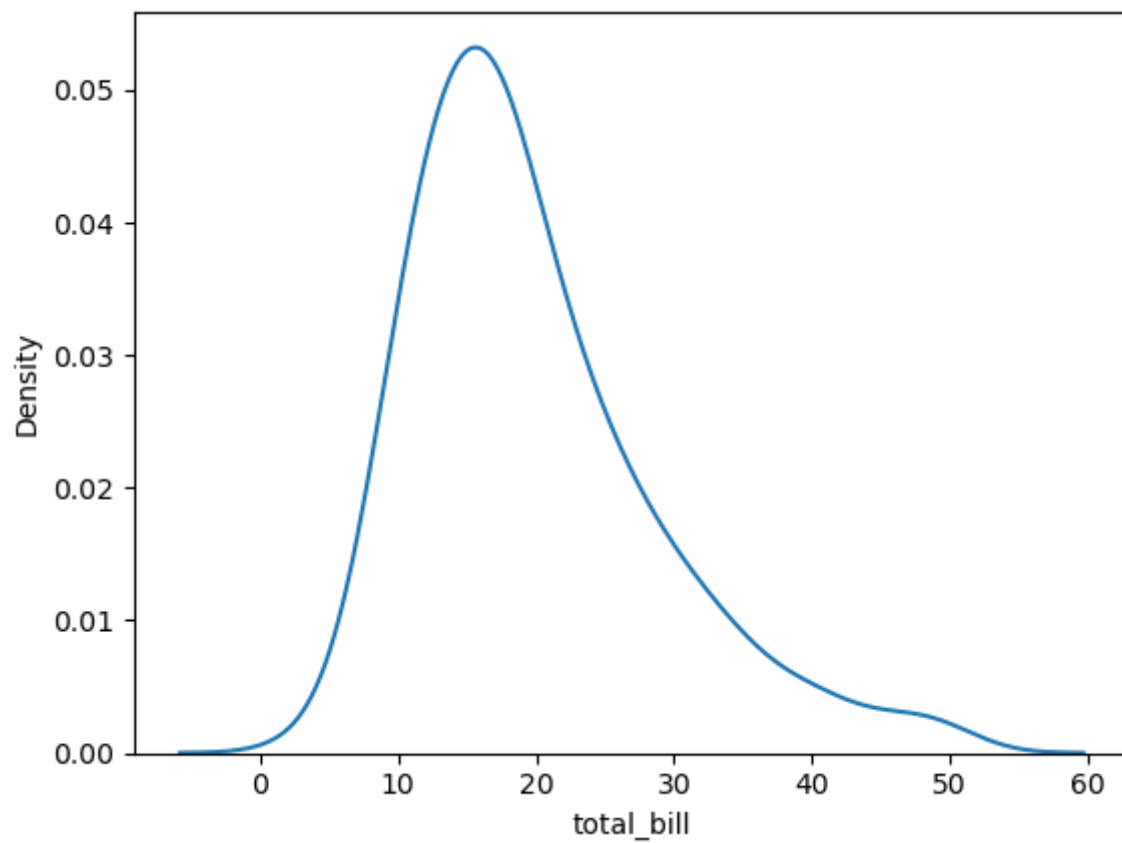
The scatter plot shows how two numbers, "total_bill" and "tip", are related to each other. Each dot on the plot represents one data point, with the "total_bill" being shown on the x-axis and the "tip" being shown on the y-axis. This plot helps us see if there is a pattern between the two numbers.

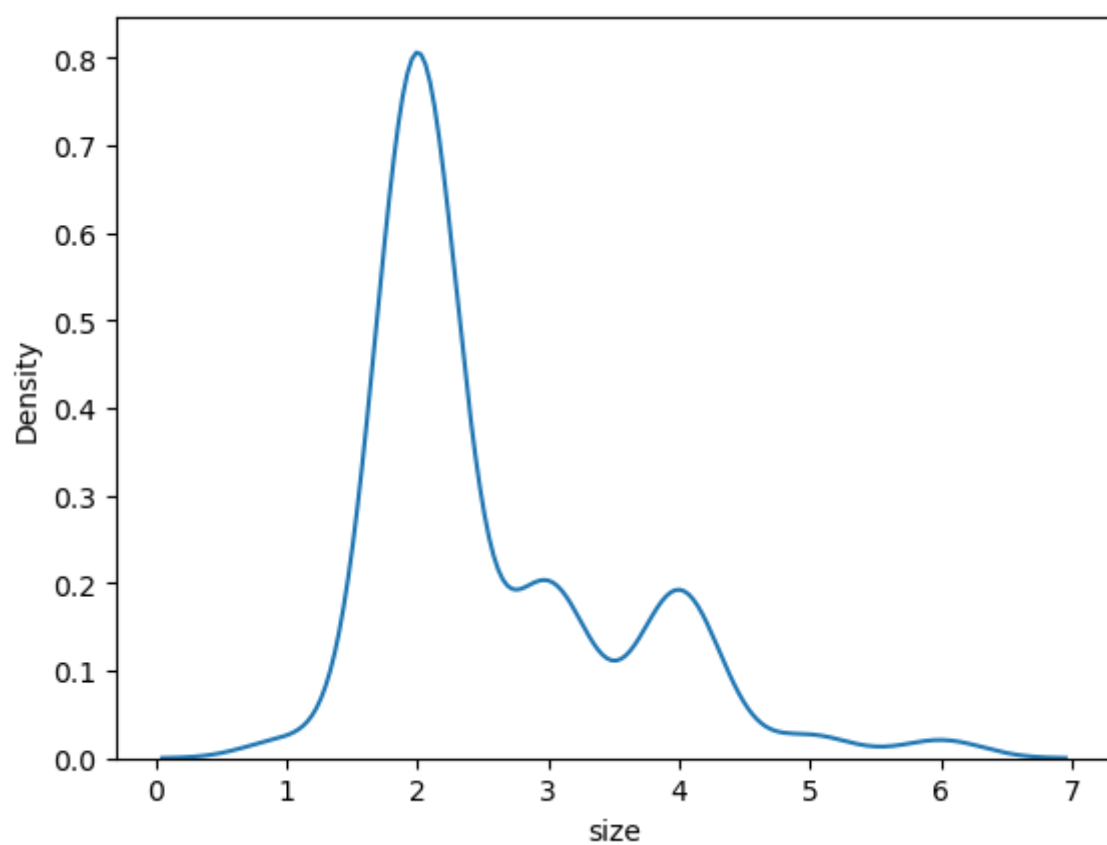
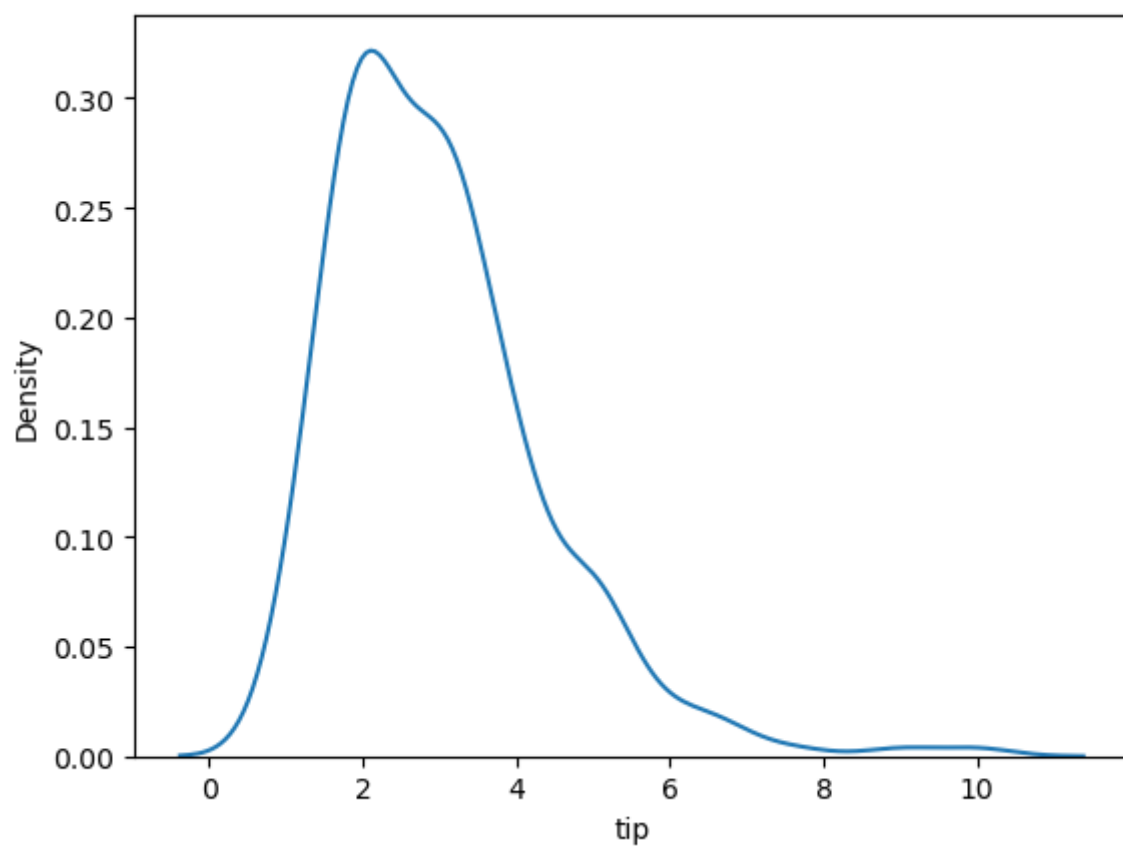
7. Write a Python program to display density plot of all numerical columns from the dataframe and write the intuition about plot.

In [12]:

```
### code here
```

```
for column in df.select_dtypes(include='number').columns:  
    sns.kdeplot(df[column].dropna())  
    plt.show()
```



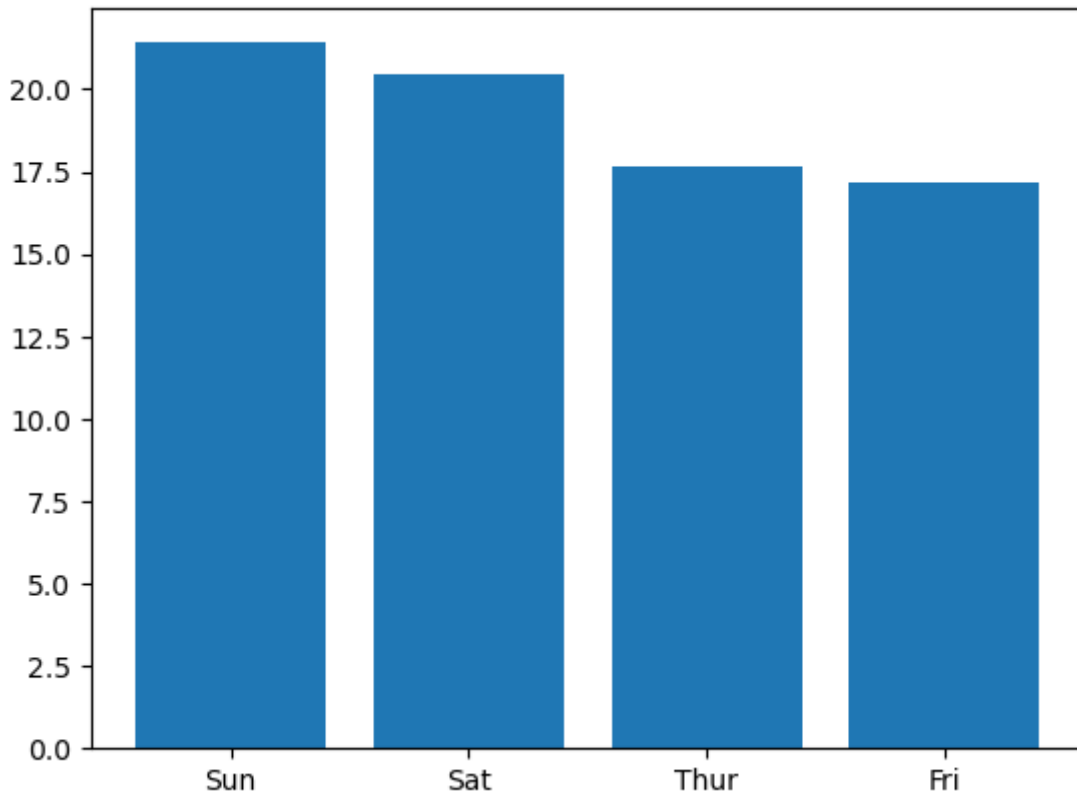


The density plot is a graph that displays how the values of a numerical variable are distributed. It helps us understand the shape of the distribution, including the height and spread of the values. Each plot shows the density of one numerical column in a dataframe. By comparing the density plots of different columns, we can see how their distributions differ.

8. Write a Python program to calculate average total bills by days and display top 4 days by plot.

In [13]:

```
### code here
df_day_avg = df.groupby("day")["total_bill"].mean().reset_index()
df_day_avg = df_day_avg.sort_values("total_bill", ascending=False)
df_top_4 = df_day_avg.head(4)
plt.bar(df_top_4["day"], df_top_4["total_bill"])
plt.show()
```



Bonus Question

1. What is Histogram chart ?

write here

A histogram chart is a graphical representation of the distribution of a dataset.

It is an estimate of the probability distribution of a continuous variable.

It is a way to show the frequency of different values that fall within a set of ranges or bins.

The x-axis represents the bins or ranges of values, and the y-axis represents the frequency of observations that fall within each bin.

Histograms are useful for visualizing the distribution of a dataset and identifying patterns such as skewness and outliers.

2. What is box plot ?

write here

A box plot, also known as a box-and-whisker plot, is a way to display the distribution of a dataset based on five number summary. The five number summary includes the minimum value, first quartile, median, third quartile, and maximum value. The box plot is composed of a box, which represents the interquartile range (IQR), and whiskers, which represent the minimum and maximum values. The box plot is useful for identifying outliers and comparing the distribution of multiple datasets.

3. What are the plots we can use for univariate analysis.

write here

Plots that can be used for univariate analysis are:

Histograms

Density plots

Kernel Density Estimation (KDE) plots

Box plots

Violin plots

Strip plots

4. What are the plots we can use for bivariate analysis.

code here

Plots that can be used for bivariate analysis are:

Scatter plots

Line plots

Hexbin plots

Contour plots

Violin plots

Box plots

5. What are plot we can use for multivariate analysis.

code here

Plots that can be used for multivariate analysis are:

Scatter plot matrix

Parallel Coordinates

Heatmaps

3D Scatter Plots

3D Surface Plots