Project Title

Prediction of hotel industry with the help of seaborn graphics libery.

Name: T. Jaya Krishna
Roll No:20X01A05B8

Project Guide Name
Mr. Lokesh B Sir

Problem Statement

The organization has been provided “Tips.csv” data set including various including various columns such as total bins,tip amount,gender,person per give and days of the week.As a data science engineer my task is predict and analysis data set using cborn visualizaation library and generate insides and explore the reduction between different columns according to the needs.

Task1: Plot the trands between columns total_bins across tip.
TASK2: Find out which day tatal bin has a max or min , along with that no of customers.
Task3: Predict which day no of customers rate should be high.
Task4: Plot al the columns in a single plot.
Task5: Find out reduction between different diff columns.
Task6: Find out huminitive frequency on the basis of feature of total amount.
Task7: Find out how many customers has give high no of rupees on particular day.
Task8: Use the statatics and find out which column contains max outliers(error).
Task9: Visualize tip across total bill in such a way outliers and range show in a single plot.
Task10: Visualize total bill seperately for each particualr day.

Tools:
1. Jupyter Notebook.
2. Python Programing Language.
3. Numpy (Python Libery_)
4. Google Colab.

seaborn-batch-02-3-1

September 18, 2023

1 Project Title: Prediction of hotel industry with the help of seaborn graphics libery.

#problem statement : The organization has been provided "Tips.csv" data set including various including various columns such as total bins,tip amount,gender,person per give and days of the week.As a data science engineer my task is predict and analysis data set using cborn visualizaation library and generate insides and explore the reduction between different columns according to the needs. Task1: Plot the trands between columns total_bins across tip.

TASK2: Find out which day tatal bin has a max or min , along with that no of customers.

Task3: Predict which day no of customers rate should be high.

Task4: Plot al the columns in a single plot.

Task5: Find out reduction between different diff columns.

Task6: Find out huminitive frequency on the basis of feature of total amount.

Task7: Find out how many customers has give high no of rupees on particular day.

Task8: Use the statatics and find out which column contains max outliers(error).

Task9: Visualize tip across total bill in such a way outliers and range show in a single plot.

Task10: Visualize total bill seperately for each particualr day.

1.1 Tools and Technology:

1. Python
2. Numpy
3. Pandas
4. MatplotLib
5. Plotly
6. Google Colab

2 Data Pre-Processing

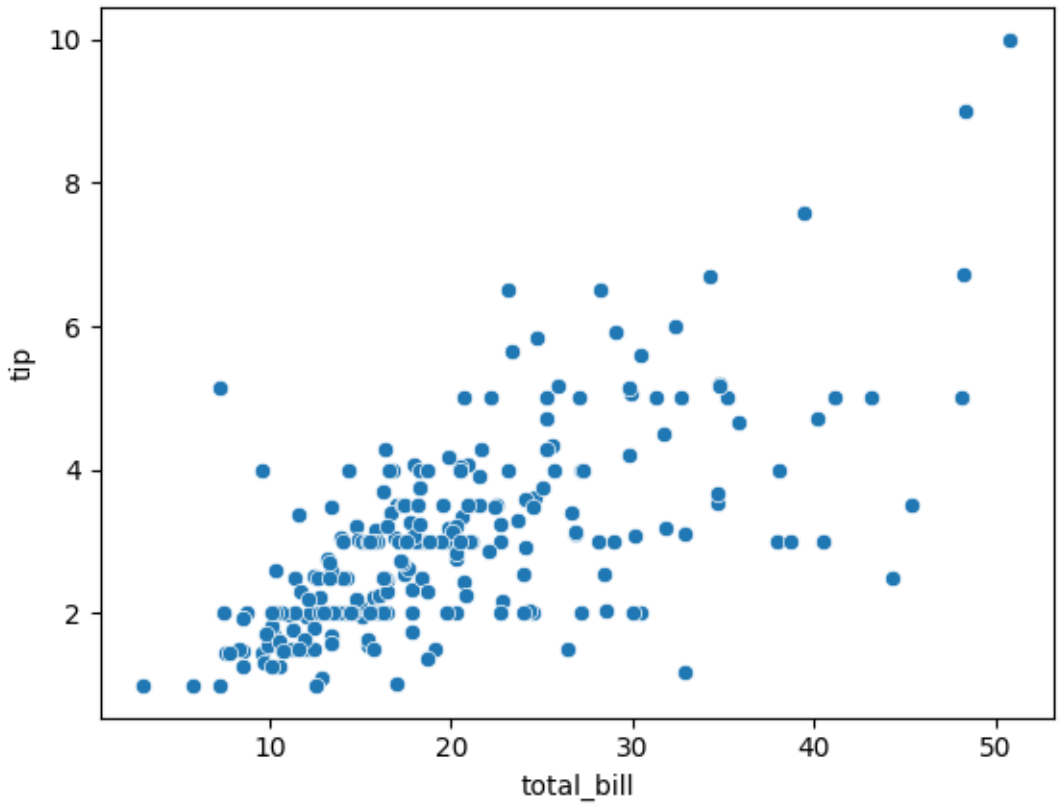
[1]:
```
#import the required libery's
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
```

[4]:
```
#load seaborn
import seaborn as sns
```

3 Scatter Plot

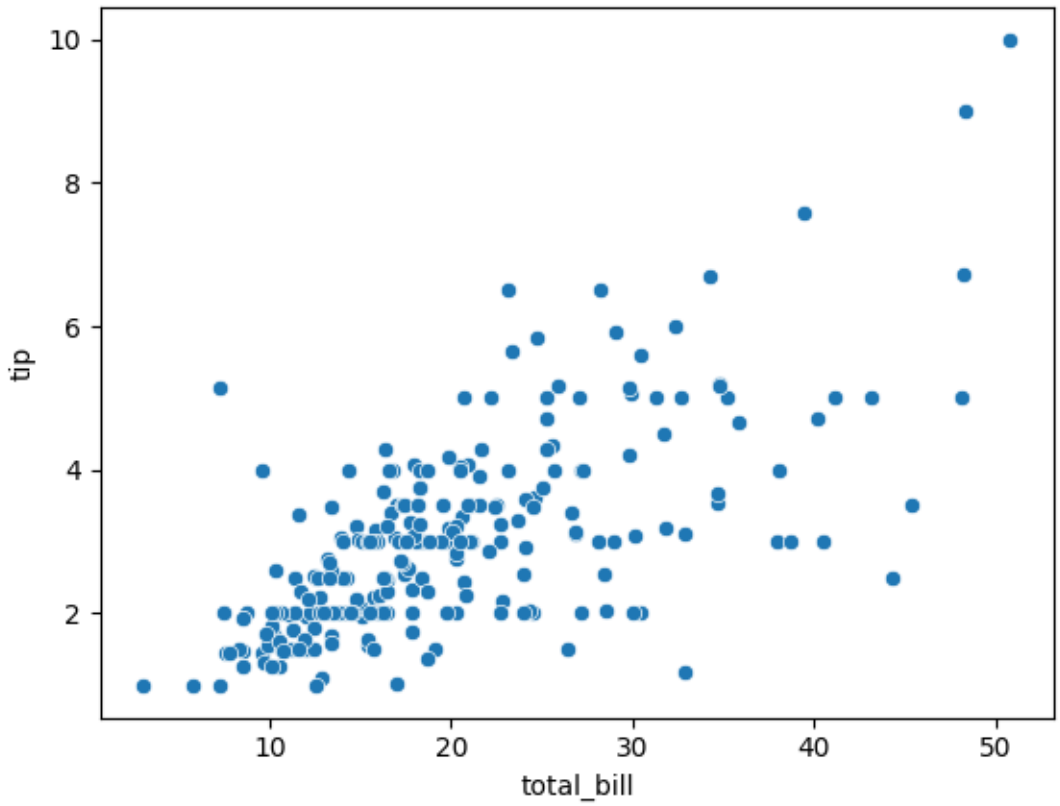
3.0.1 Input Variable (x) –> total _bill column ### Output Variable (y) –> tip column

[5]:
```
#Load the pre-defined tips dataset of catering and staffing service based␣
 ↪industry
#tips = sns.load_dataset("tips")
# Sample data

tips = pd.read_csv("tips.csv")
# Create a basic scatter plot
sns.scatterplot(x="total_bill",y="tip", data=tips)
plt.show()
```



4 Custmize Plot in Seaborn

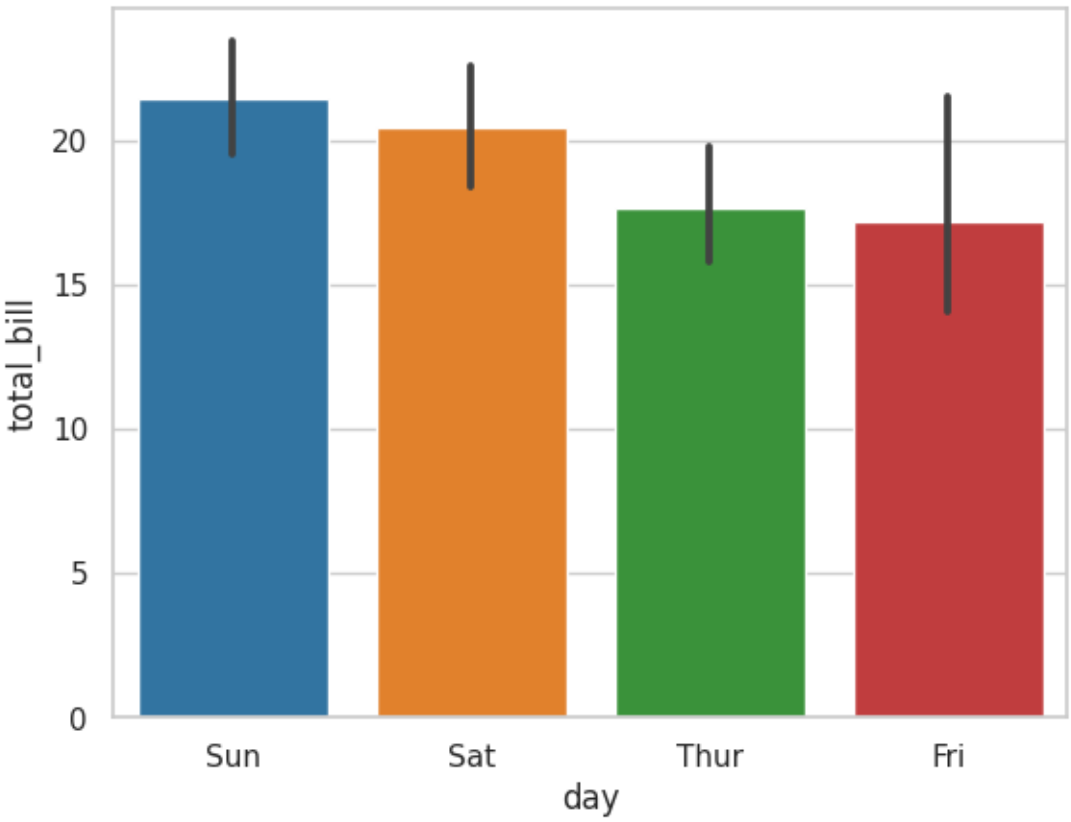
[6]:
```
# Costmize or update the scatterplot by adding context, style.
#set_style & set_context --> only for notebook
sns.scatterplot(x="total_bill",y="tip", data=tips)
sns.set_style("whitegrid")
sns.set_context("notebook")
plt.show()
```

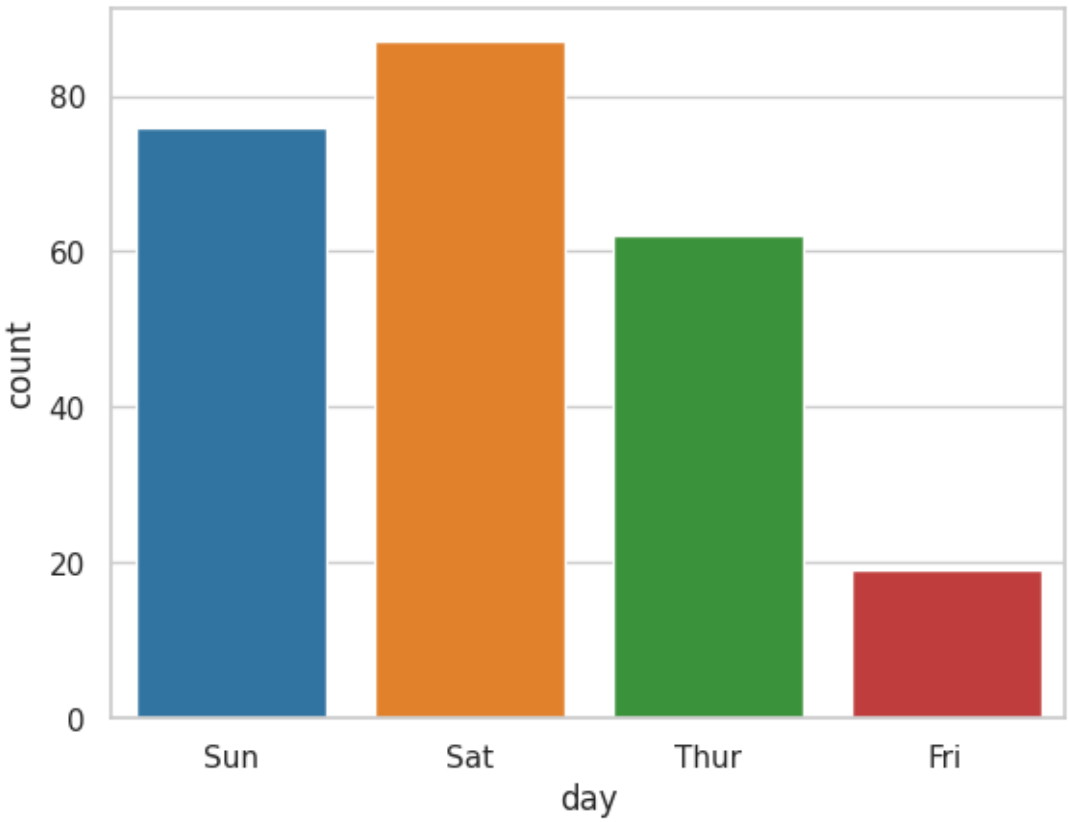


5 Barplot & Count Plot

[7]:
```
# Plot the bar plot and with indivisual count for seperate visual (X,Y,data) as␣
 ↪a argument,colomn name day & total_bill
sns.barplot(x="day", y="total_bill",data=tips)
plt.show()
#count has only one argument
sns.countplot(x="day",data=tips)
```

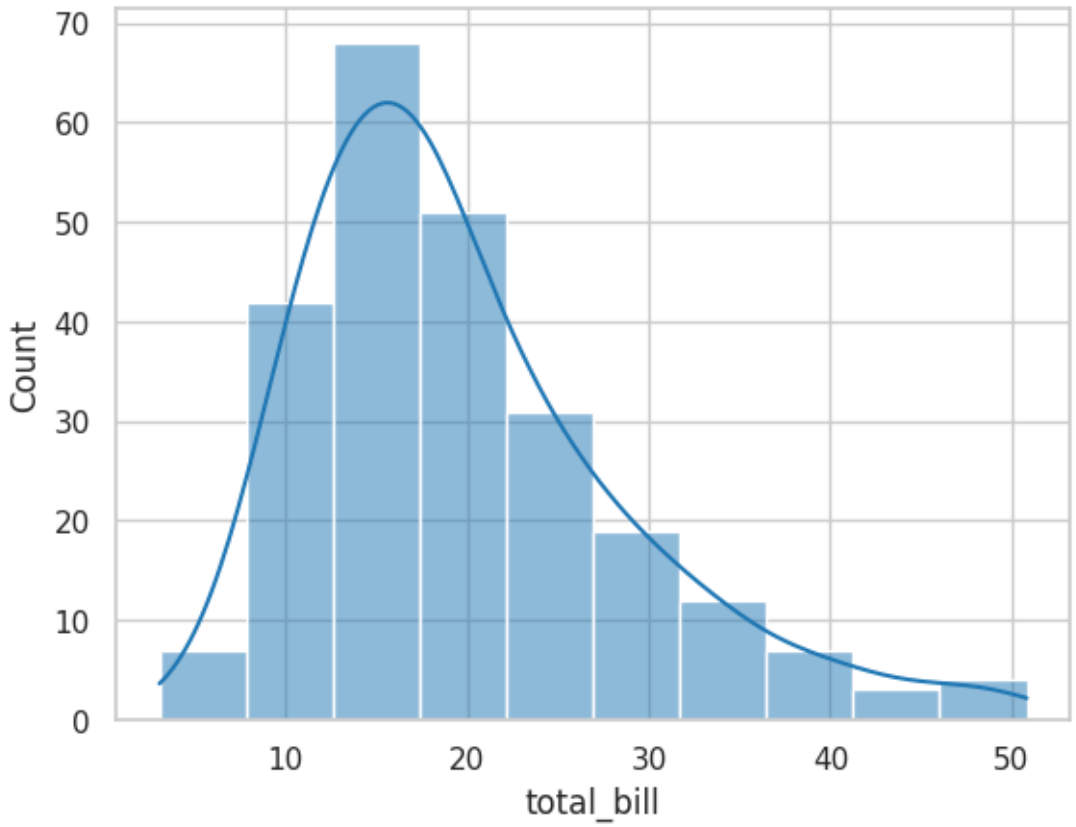
```
plt.show()
```





6 Histogram

[8]:
```
#plot the histogram which contents maximun 10 bins and known deviation as True
sns.histplot(x="total_bill", bins=10, kde=True , data=tips )
plt.show()
```



7 Pair Plot

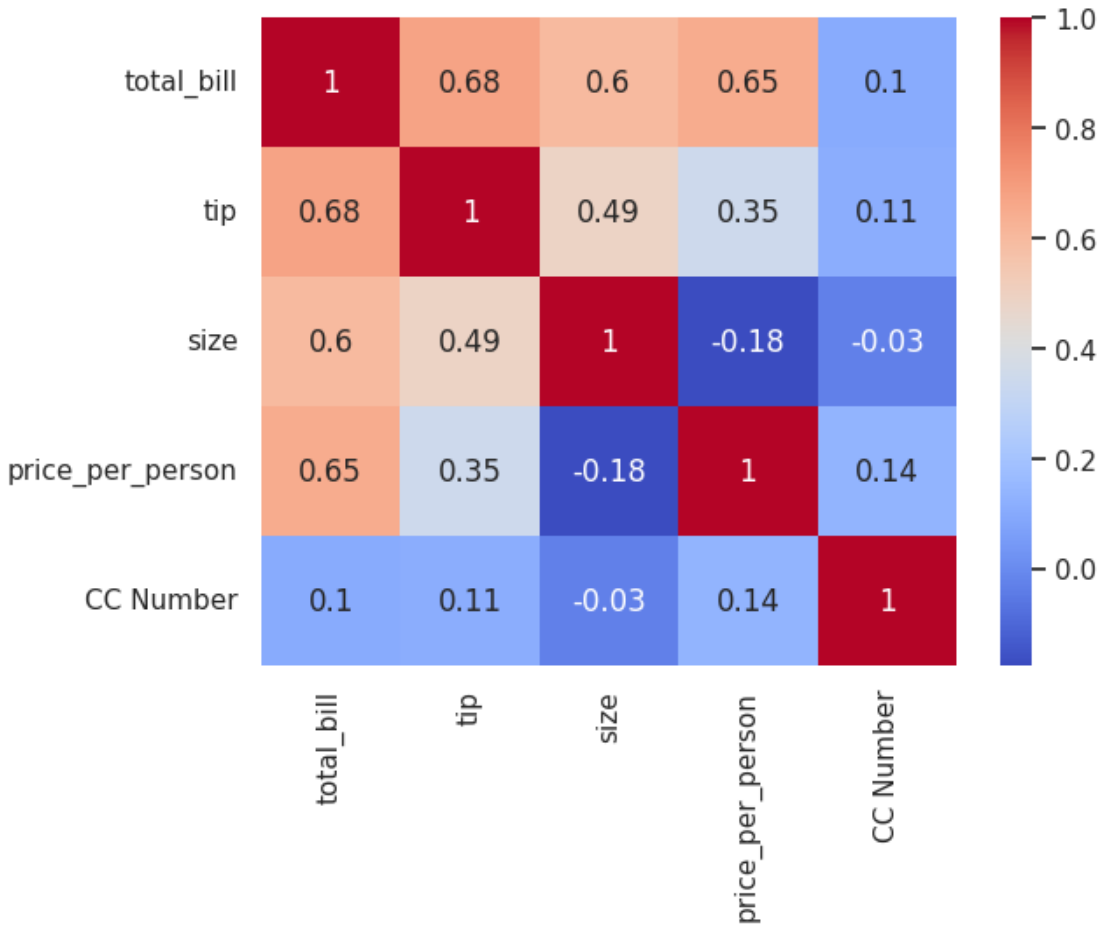
[9]:
```
#plot the diffrent types of categories in single plot as easy to be visual. for␣
 ↪coloum "sex" -->Gender
#hue is built in function is used
sns.pairplot(tips, hue="sex")
plt.show()
```

[ ]:
```
#customize from more plot to in single scatter plot with proper title and␣
↪labels.
```

8 Corelational Maxtix / Heatmap

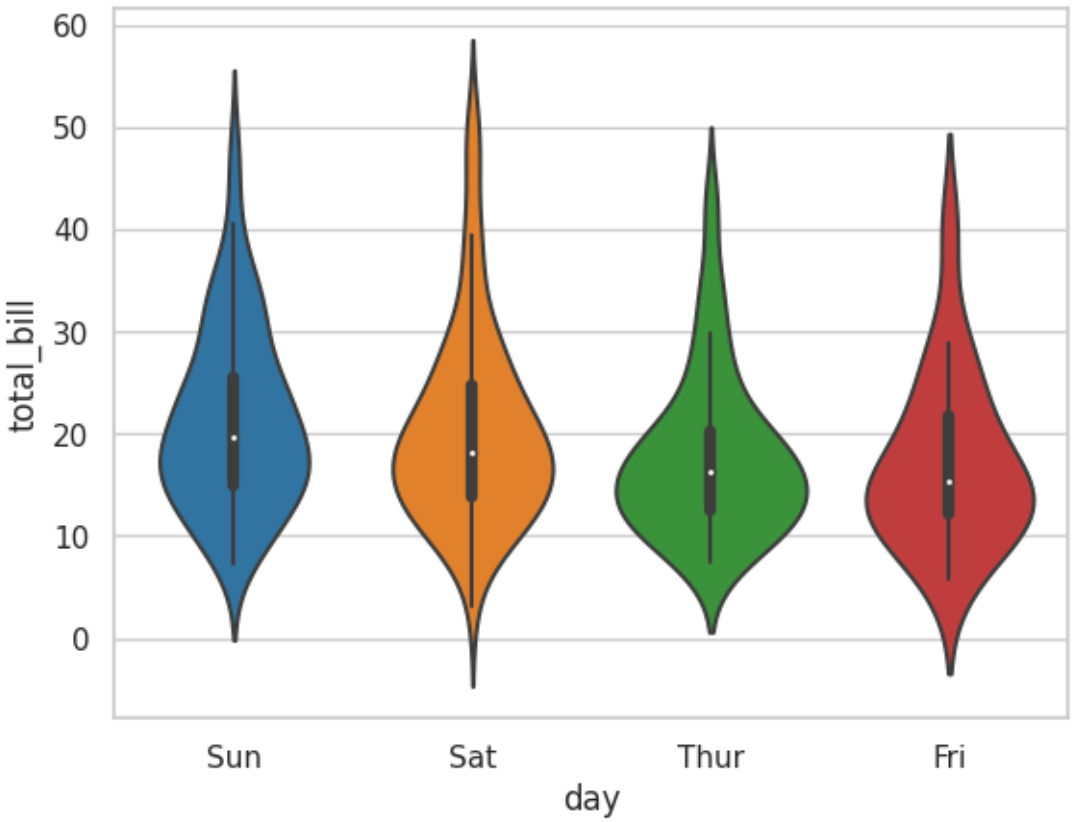
[10]:
```
#Plot the corelational matrix
sns.heatmap(tips.corr(), annot= True, cmap="coolwarm")
plt.show()
```

```
<ipython-input-10-e63d077a3c93>:2: FutureWarning: The default value of
numeric_only in DataFrame.corr is deprecated. In a future version, it will
default to False. Select only valid columns or specify the value of numeric_only
to silence this warning.
  sns.heatmap(tips.corr(), annot= True, cmap="coolwarm")
```



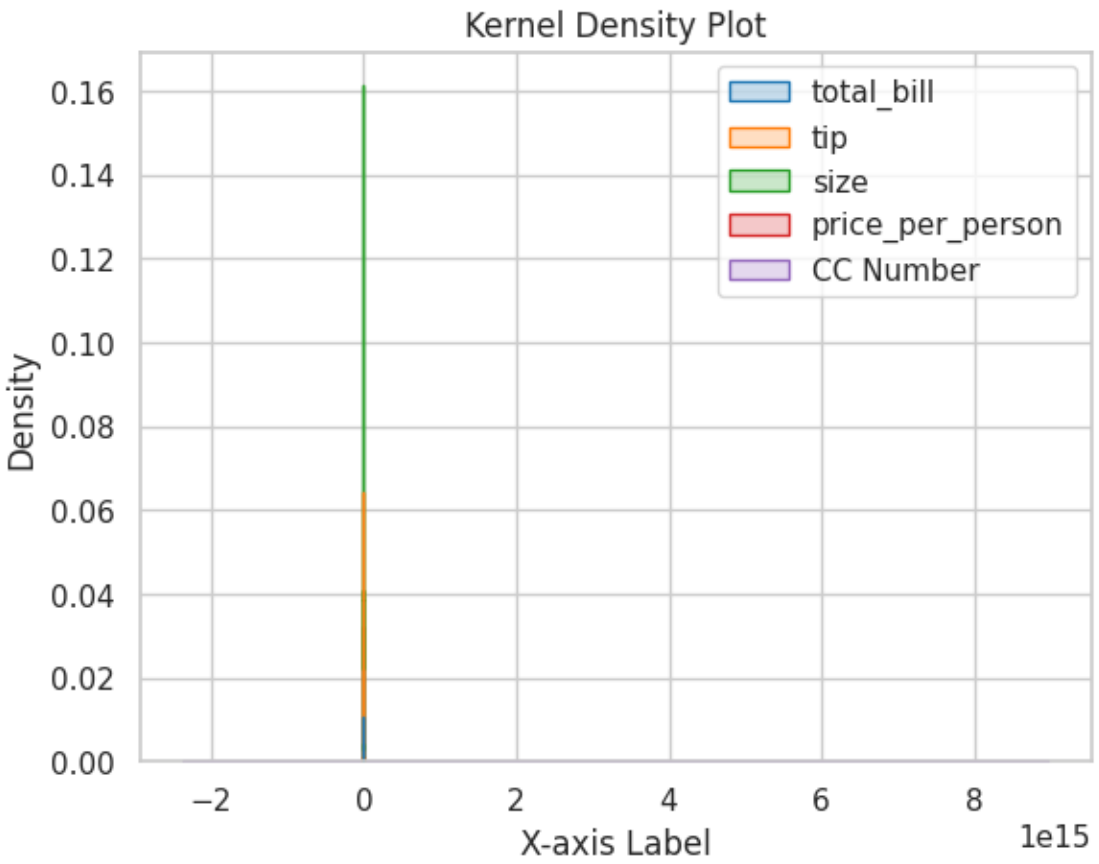
9 Violin Plot

[11]:
```
sns.violinplot(x="day",y="total_bill",data=tips)
# Customize the plot (optional)


# Display the plot
plt.show()
```



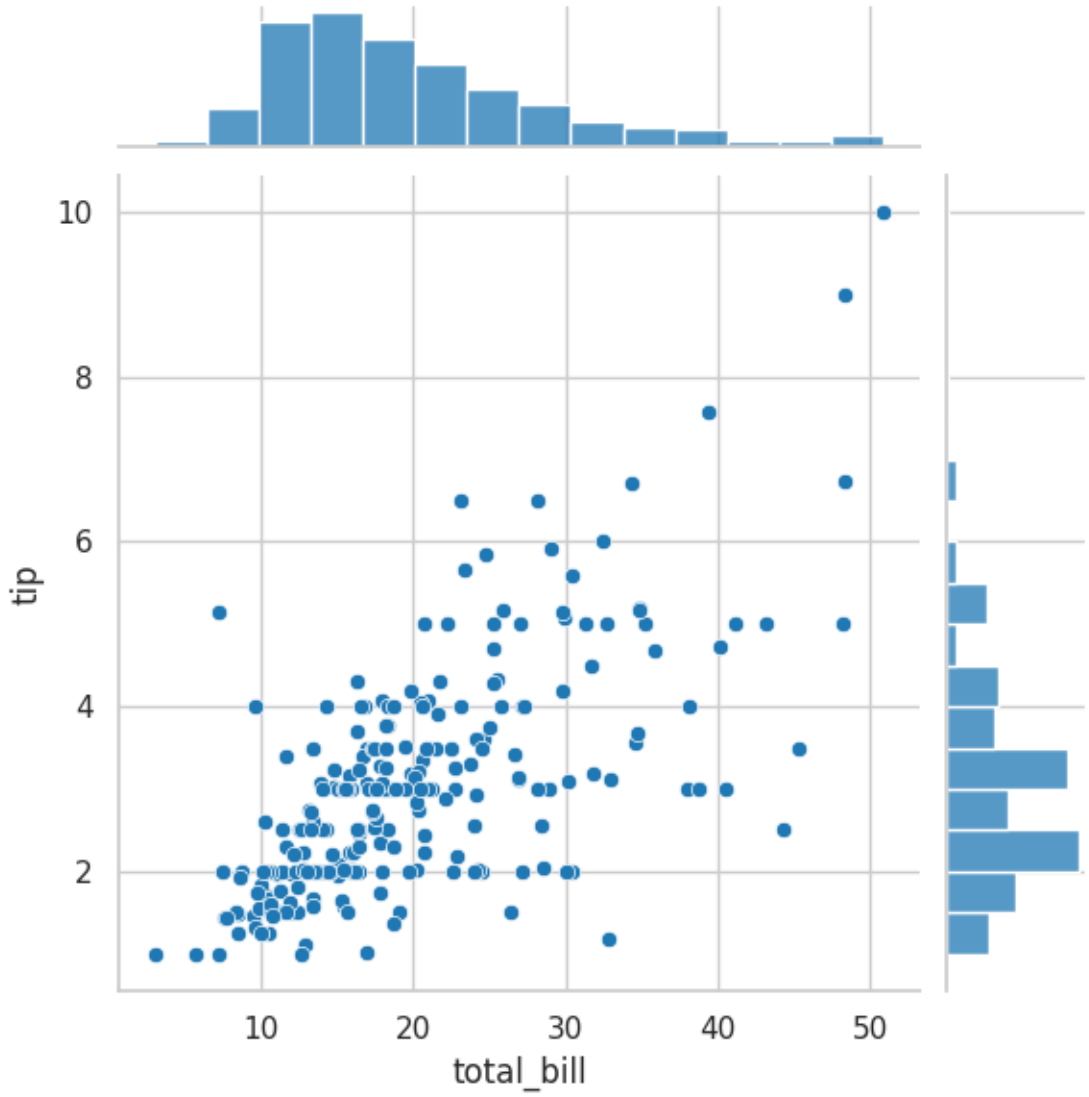
10 Kernal Density Estimate Plot

[12]:
```
sns.kdeplot(tips, fill=True)

# Add labels and title
plt.xlabel("X-axis Label")
plt.ylabel("Density")
plt.title("Kernel Density Plot")

plt.show()
```



11 Joint Plot

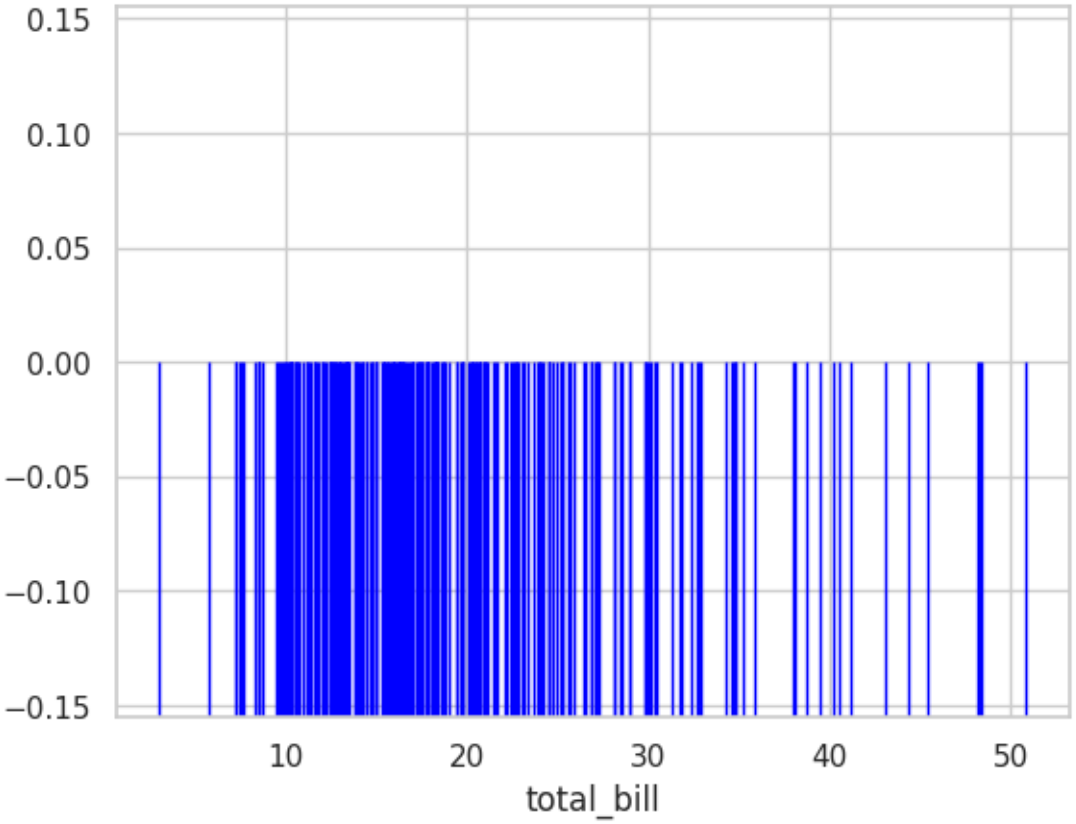
[13]:
```
# Create a joint plot
sns.jointplot(x="total_bill", y="tip", data= tips, kind="scatter")
# Show the plot
plt.show()
```



12 Rug Plot

[14]:
```
# Create a Rug Plot for the "total_bill" column
sns.rugplot(x="total_bill",data=tips, height= 0.5, color="blue")
# Set the style (optional)
plt.figure(figsize=(8,4))
# Set the figure size (optional)


# Add labels and a title (optional)
```

```
# Show the plot
plt.show()
```

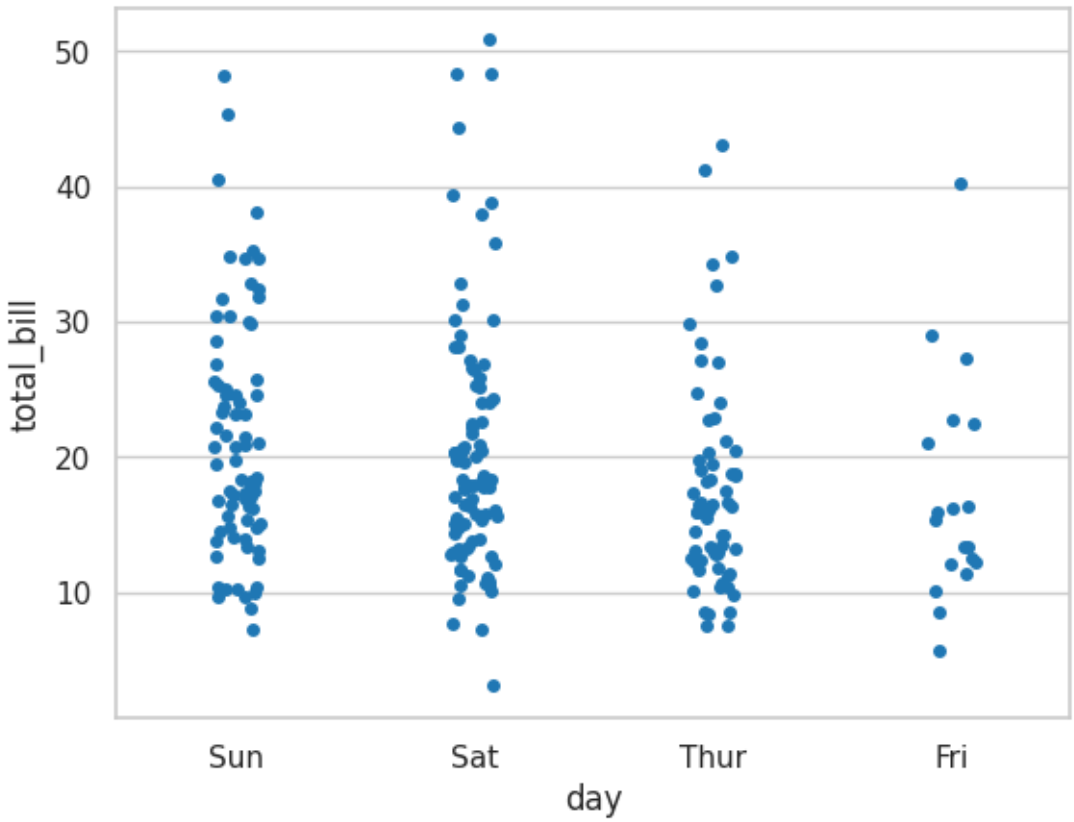



```
<Figure size 800x400 with 0 Axes>
```

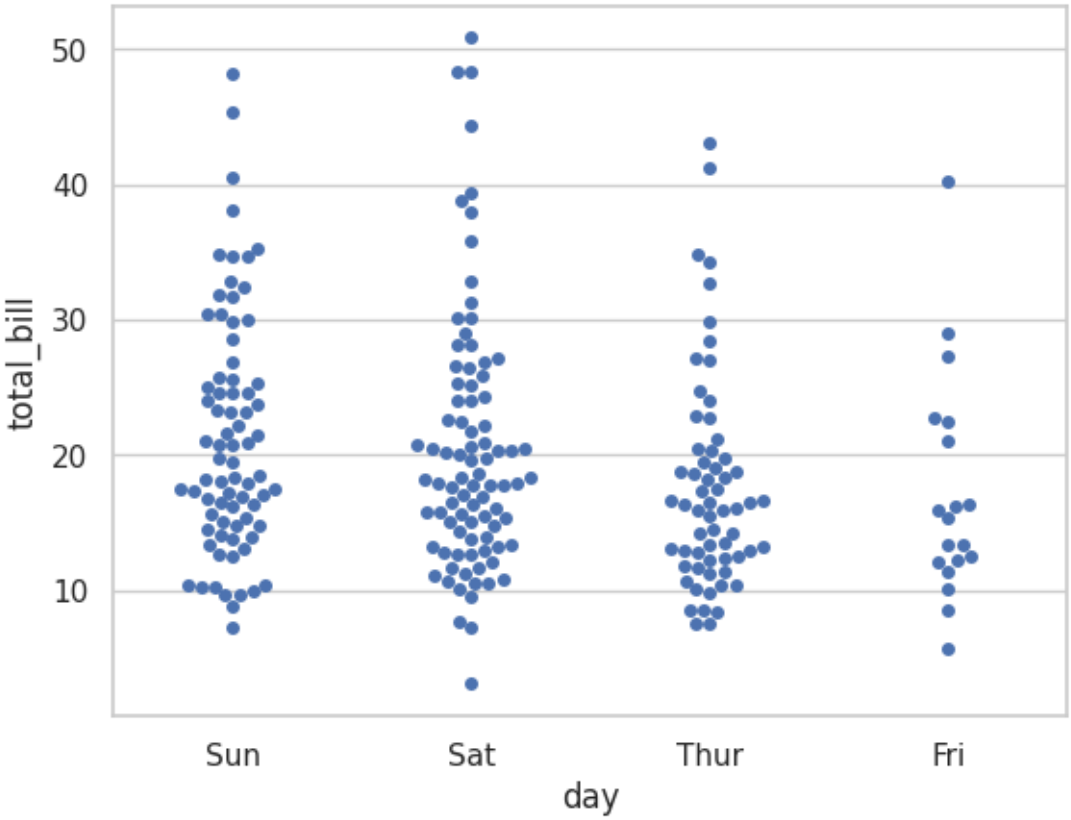
13 Strip Plot

[15]:
```
sns.stripplot(x="day",y="total_bill",data=tips)
```

[15]:
```
<Axes: xlabel='day', ylabel='total_bill'>
```

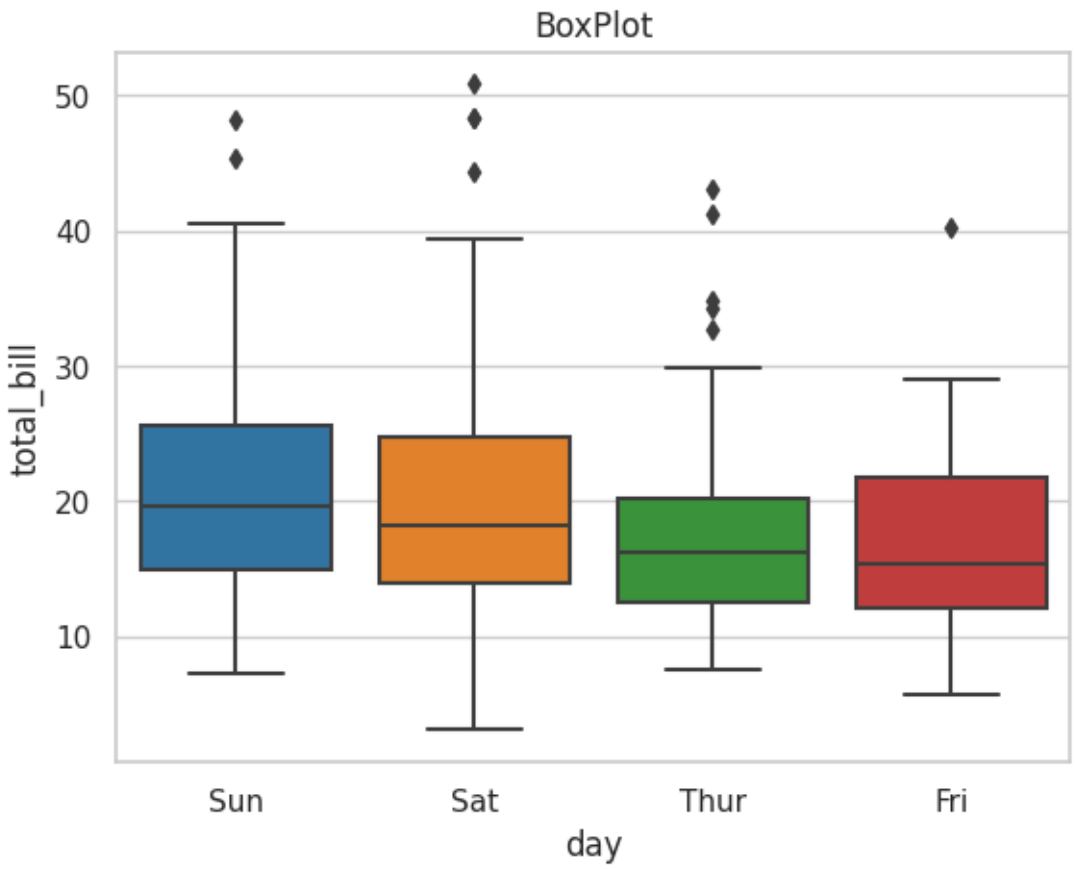


[ ]:
```
sns.swarmplot(x="day", y="total_bill", data=tips)
plt.show()
```

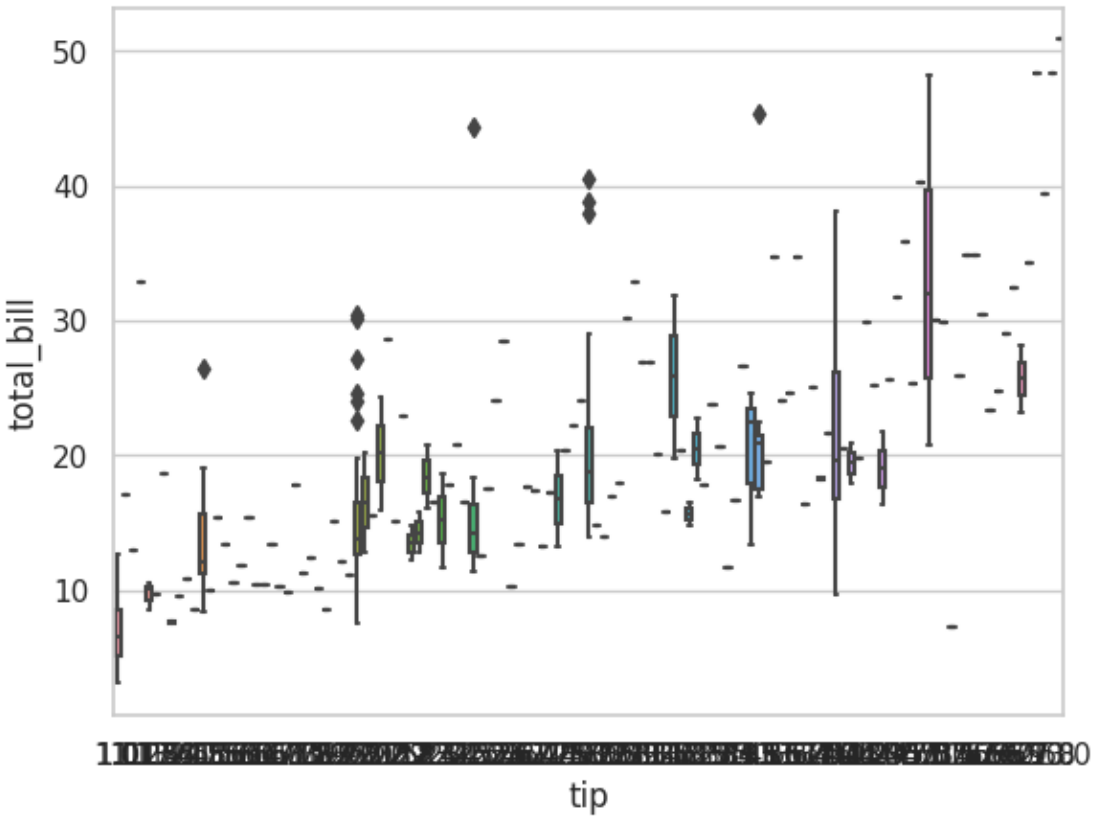


14 Box Plot

[16]:
```
sns.boxplot(x="day",y="total_bill",data=tips)
plt.title("BoxPlot")
plt.show()
```

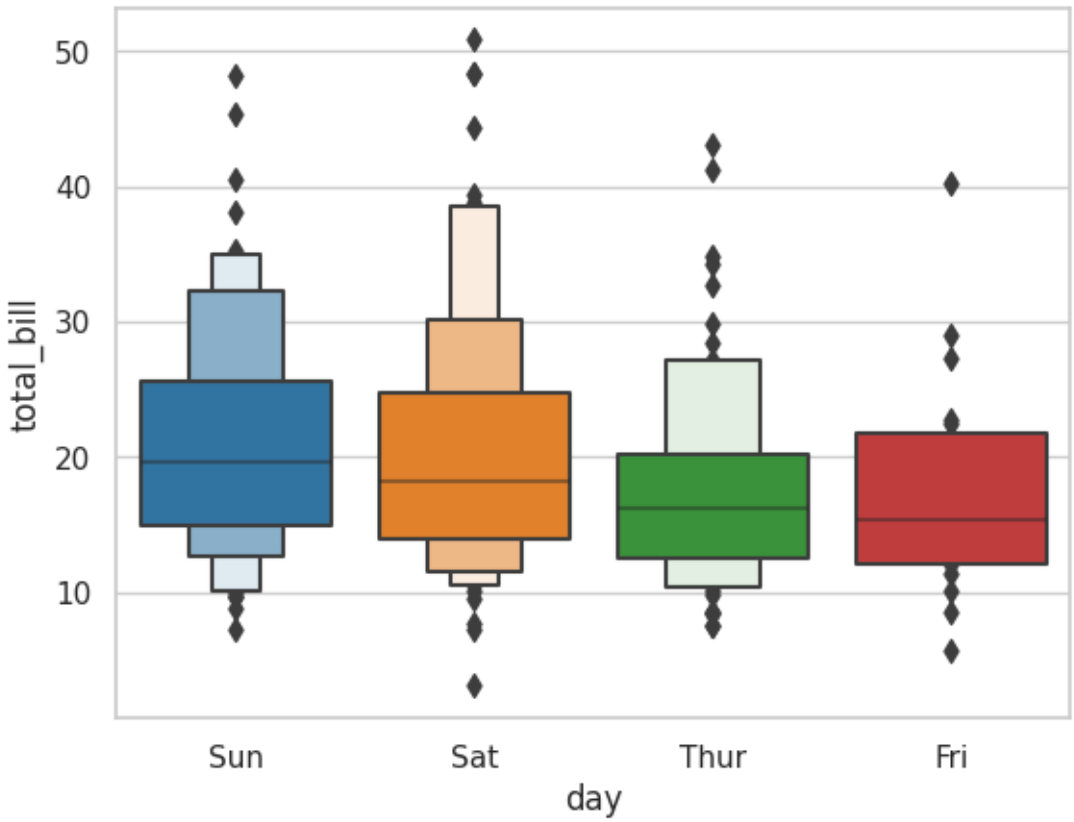


[17]:
```
sns.boxplot(x="tip",y="total_bill",data=tips)
plt.show()
```



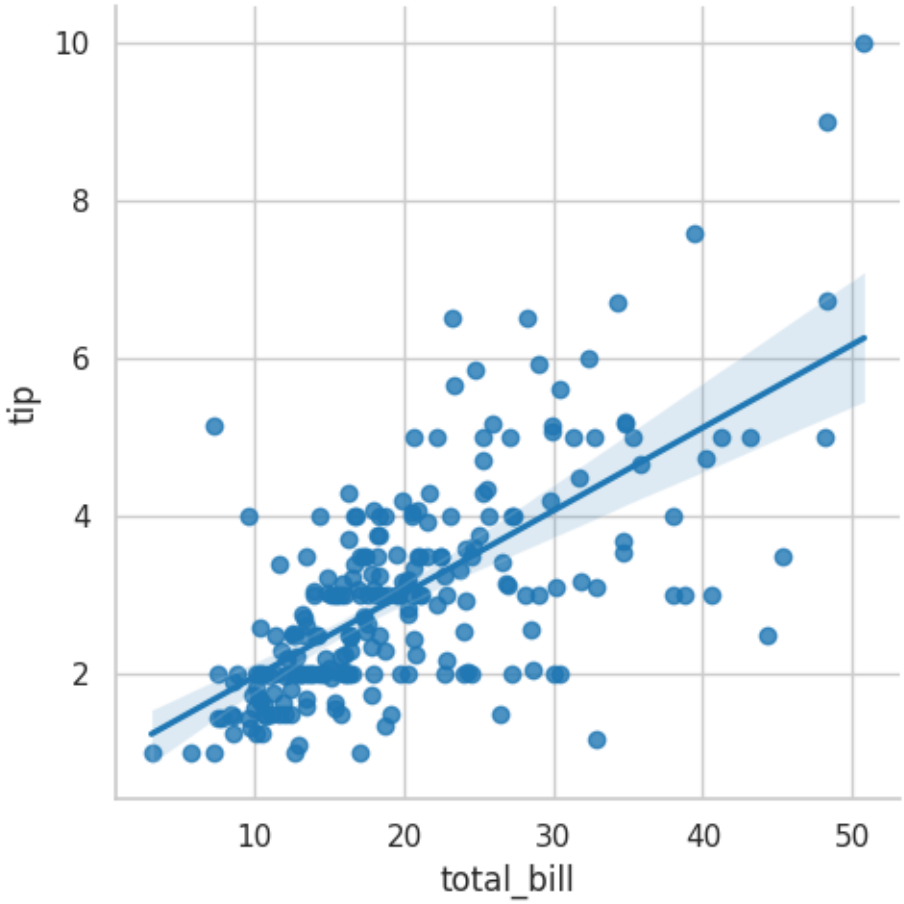
15 Boxen Plot

[18]:
```
sns.boxenplot(x="day",y="total_bill",data=tips)
# Add labels and a title


# Show the plot
plt.show()
```



16 Regression Plot

[19]:
```
# Create a regression plot using lmplot
sns.lmplot(x="total_bill",y="tip",data=tips)

# Show the plot
plt.show()
```



17 Fianl Result

1.The trends between total bill vs tip as shown scatter plot,which show maximum tip has a 10 rupees or amount 50 rupees

2.The maximum total bills 21 rupees on sunday and minimum 17 rupees on friday

3.Saturday has height number of customers visited to the hotel.The average count is 85.

4.All the columns are visualize in pair plot.

5.The relationship between different kinds of columns as shown in heatmap such as total bill and tip , tip and size etc…,

6.The cummulative frequency as found 61 related to total amount.

7.According to Violin plot in sunday height among this 25, Saturday = 28 rupees , Thursday =20 rupees , Friday = 22 rupees.

8.Size column as contains maximum number of outliers

[ ]: