[Date]16|09|2023

Project Title:

Analysis and visualisation of hotel industry “tips.csv” dataset with the help of Seaborn

Name:K.Pranavi

Branch:CSE

Year:4

Project Guide

Mr. Lokesh B. Sir

Problem Statement

The organization has been provided “tips.csv” dataset including various columns such as total bills,tip amount,gender ,person per bill and days of the week.As a data science engineer my task is predict and analyse dataset using Seaborn visualization library and generate insights and explore the relationship between different columns according to the needs.

Task-1:Plot the trends between columns total_bill across tip

Task-2:Find out which day has a maximum or minimum,along with that number of customers.

Task-3:Predict which day number of customers rate should be high.

Task-4:Plot all the columns in single graph

Task-5:Find out relation between different different columns

Task-6:Find out cumulative frequency on the basis of future on the total amount

Task-7:Find out how many customers has give high number of rupees on particular day

Task-8:Use the statistics and find out which column contains maximum outliers(errors)

Task-9:Visualize tip across total bill in such a way outliers and range show in a single plot

Task-10:Visualize total bills separately for each particular day

Tools:

1. Jupyter Notebook.
2. Python Programing Language.
3. Numpy (Python Libery_)
4. Seaborn(Python Libery_)
5. Pandas(Python Libery_)
6. Matplot(Python Libery_)
7. Google Colab.

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

Tools and Technology:

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

✨Data Pre-Processing

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

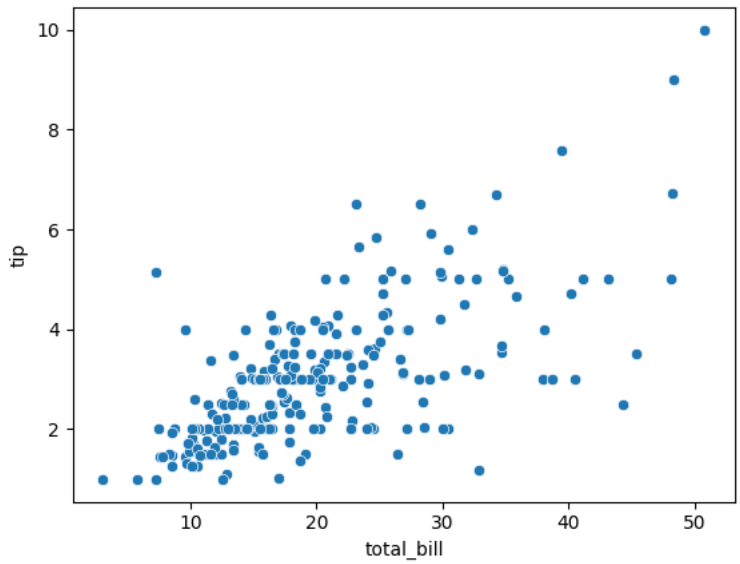
```
#load seaborn
import seaborn as sns
```

✨ Scatter Plot

Input Variable (x) --> total _bill column

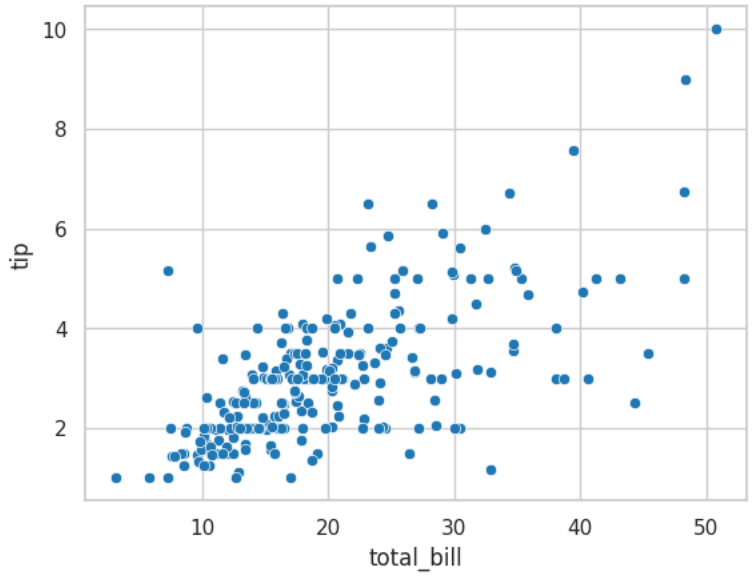
Output Variable (y) --> tip column

```
#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()
```



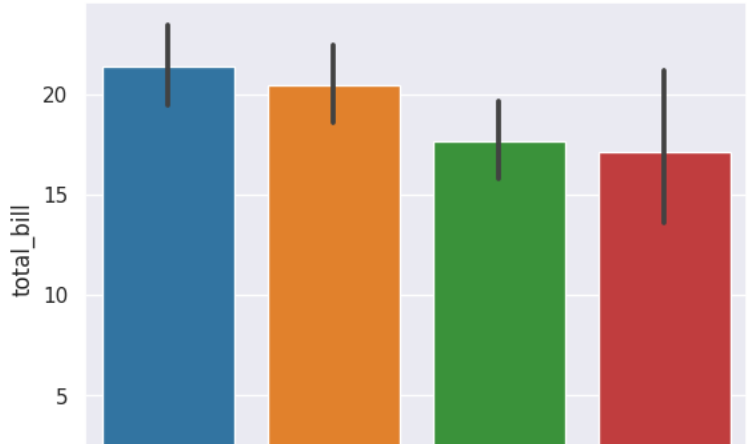
✨ Custmize Plot in Seaborn

```
# 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("darkgrid")
sns.set_context("notebook")
plt.show()
```



✨Barplot & Count Plot

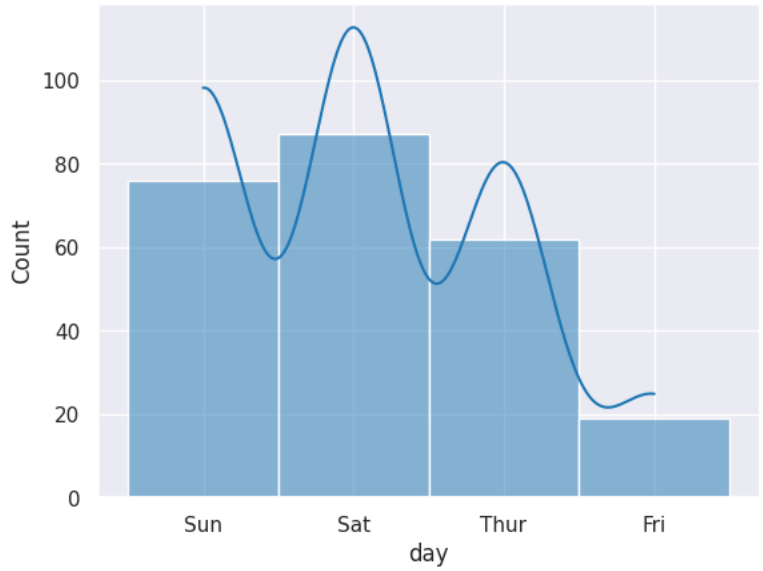
```
# Plot the bar plot and with indivisual count for seperate visual (X,Y,data) as a argument,colomn name day & total_bill
sns.barplot(x="day",y="total_bill",data=tips)
plt.show()
#count has only one argument
sns.countplot(x="day",data=tips)
plt.show()
```



✨Histogram

day

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

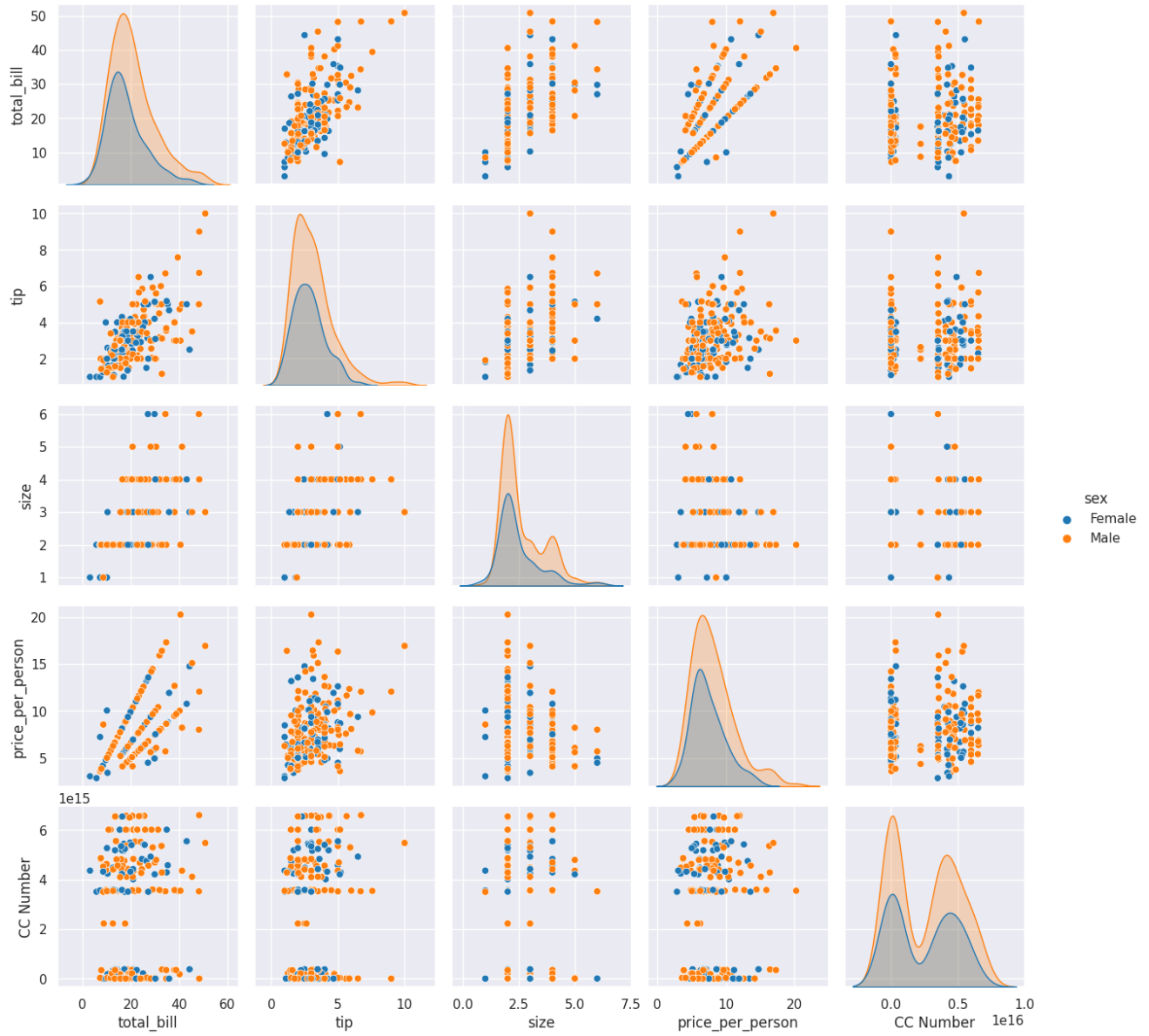



```
#plot the histogram which contents maximun 10 bins and known deviation as True
sns.histplot(x="total_bill",bins=10,kde=True,data=tips)#kde=kernaldensityestimation
plt.show()
```

✨ Pair Plot

```
#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")
```

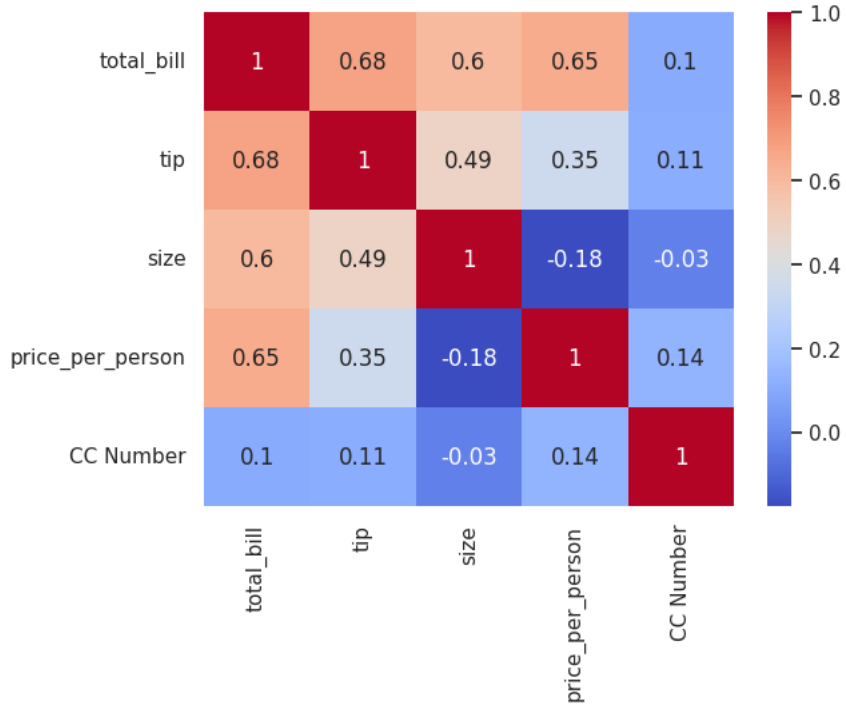
```
<seaborn.axisgrid.PairGrid at 0x7dada87e9780>
```



✨ Corelational Maxtix / Heatmap

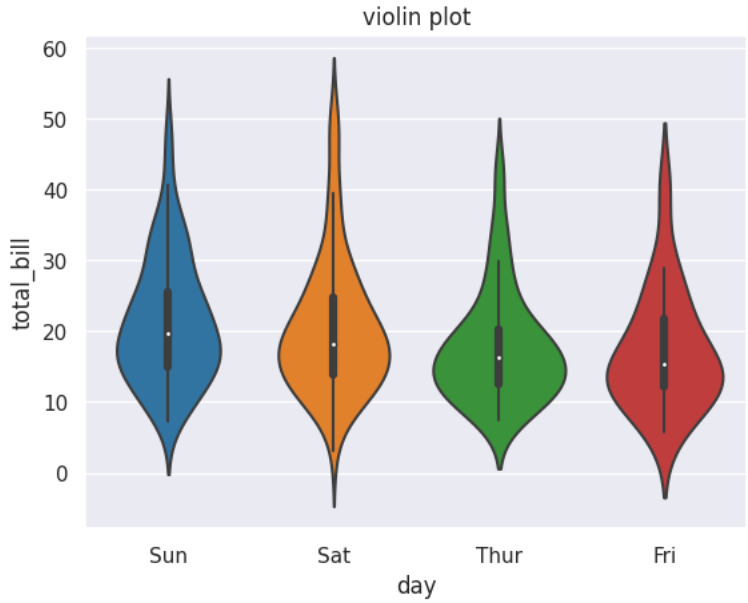
```
#Plot the corelational matrix
sns.heatmap(tips.corr(),annot=True,cmap="coolwarm")
plt.show()
```

```
<ipython-input-21-e4ef9dcd5b94>:2: FutureWarning: The default value of numeric_only in DataFrame.corr is deprecated. In a future
  sns.heatmap(tips.corr(),annot=True,cmap="coolwarm")
```



✨ Violin Plot

```
sns.violinplot(x="day",y="total_bill",data=tips)
# Customize the plot (optional).
plt.title("violin plot")

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

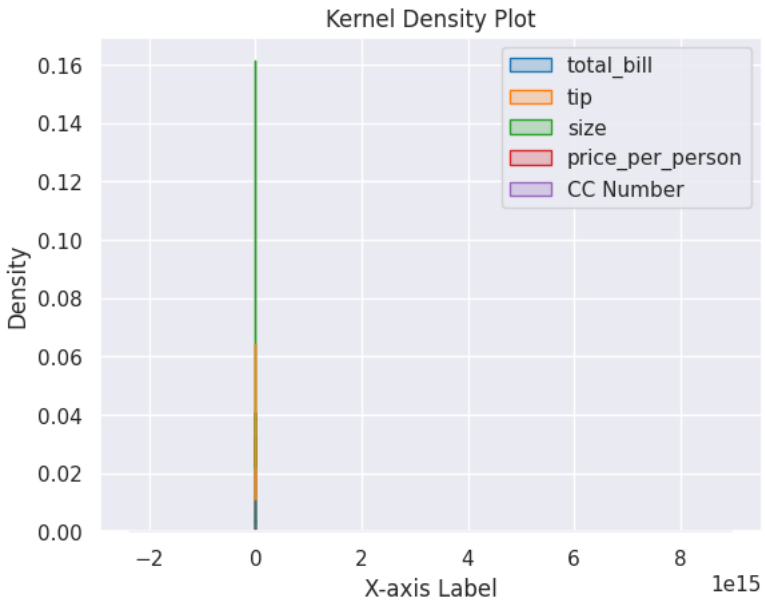


✨ Kernal Density Estimate Plot

```
sns.kdeplot(tips,fill=True)

# Add labels and title
```

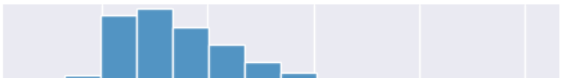
```
plt.xlabel("X-axis Label")
plt.ylabel("Density")
plt.title("Kernel Density Plot")
```

```
Text(0.5, 1.0, 'Kernel Density Plot')
```



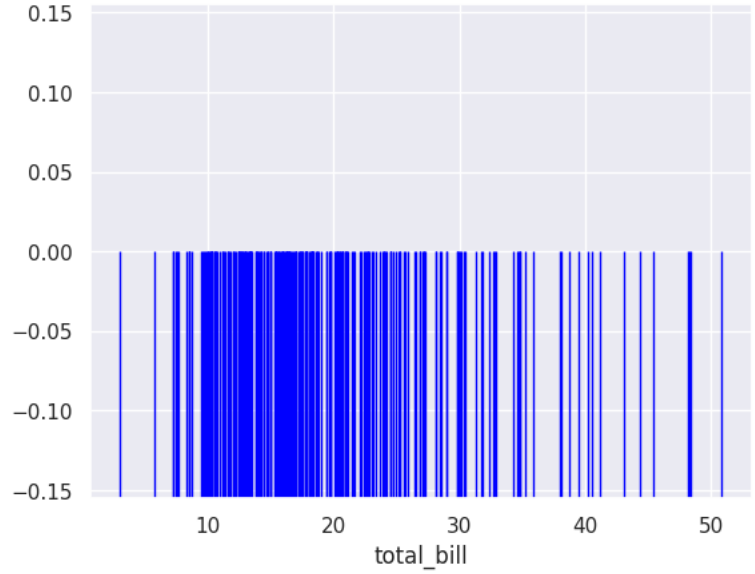
✨ Joint Plot

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



✨ Rug Plot

```
# 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()
```

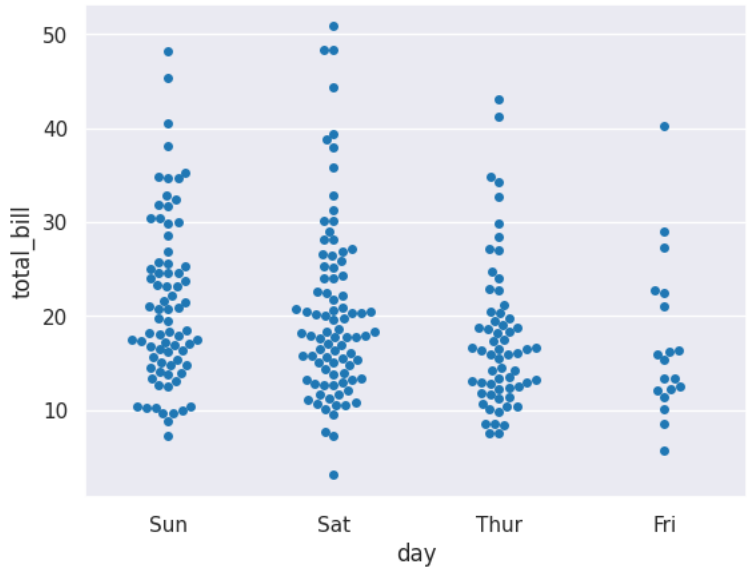



```
<Figure size 800x400 with 0 Axes>
```

✨ Strip Plot

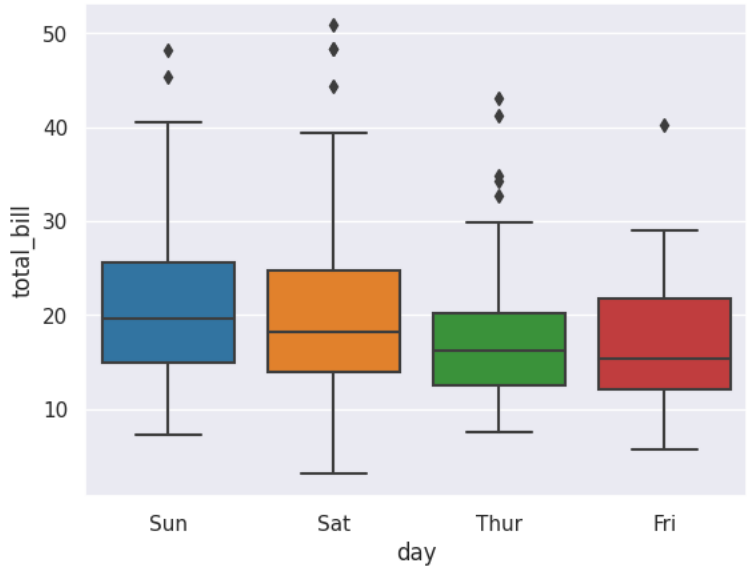
```
sns.stripplot(x="day", y="total_bill", data=tips)
plt.show()
```

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



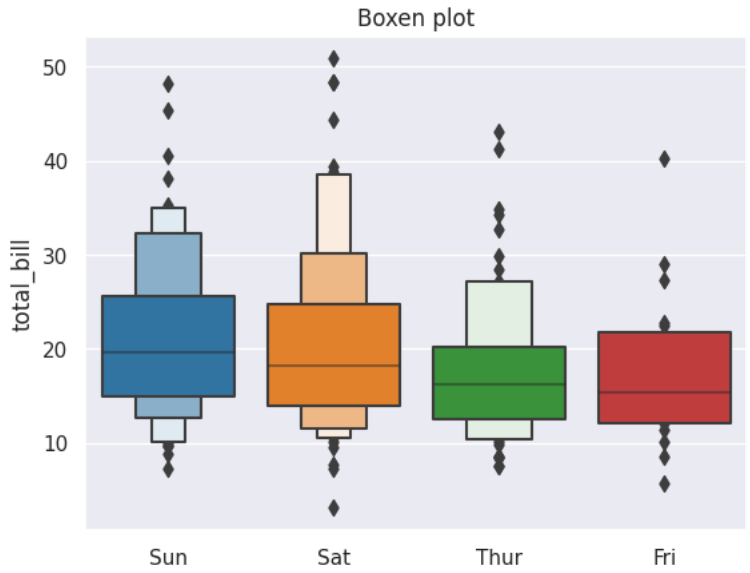
✨ Box Plot

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



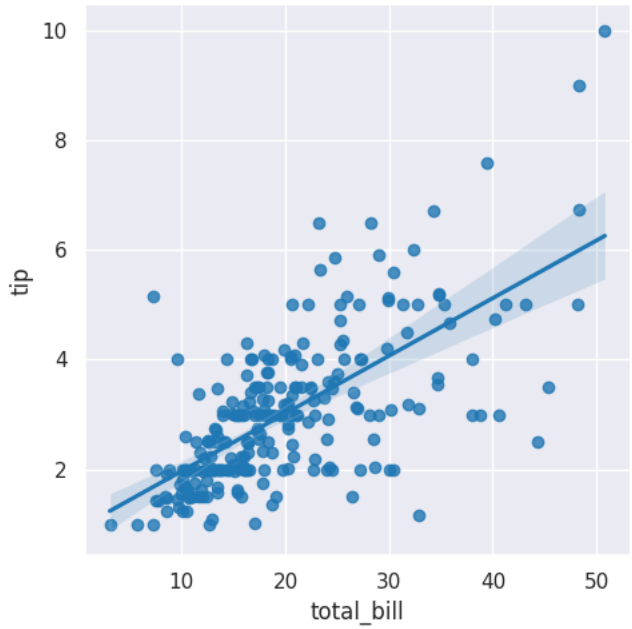
✨ Boxen Plot

```
sns.boxenplot(x="day",y="total_bill",data=tips)
# Add labels and a title
plt.title("Boxen plot")

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



✨ Regression Plot

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

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



🏆 Final Result

1. The trends between total bill versus tip has shown scatter plot,which show maximum tip as 10 rupees on amount 50 rupees
2. The maximum total bill is 21 rupees on sunday and minimum 17 rupees on Friday.
3. Saturday has highest number of customers visited to the hotel.The average count is 85
4. All the columns are visualized in pair plot
5. The relationship between different kinds of columns are shown in heatmap such as total bill and tip,tip and size etc
6. The cummulative frequency has count 61 related to total amount
7. According to the violin plot in sunday highest amount is 25,saturday highest amount is 28 rupees,thursday highest amont is 21 rupees ,friday highest amount is 22 rupees.
8. The size column contains maximum number of outliers