

What is Seaborn?

Seaborn is a Python data visualization library based on matplotlib. It provides a high-level interface for drawing attractive and informative statistical graphics.

How Seaborn Library is Different from Matplotlib Library?

Seaborn is great for quickly creating visually appealing plots with minimal code, while Matplotlib offers more customization options and fine-grained control over every aspect of a plot. Ultimately, the choice between Seaborn and Matplotlib will depend on the specific requirements of your project.

What are the different types covered in Seaborn?

- 1)Line Plot
- 2)Bar Plot
- 3)Histogram
- 4)Scatter Plot
- 5)Heatmap
- 6)Countplot
- 7)Violin Plot
- 8)Pairplot
- 9)Strip Plot
- 10)Box Plot
- 11)Categorical Plot/Cat Plot
- 12)Color Palletes
- 13)Multiple Plots
- 14)Relational Plots
- 15)Swarm Plot
- 16)KDE Plot (Kernel Density Estimation)

Dataset Links:

- 1)This dataset is the inbulit dataset used in Seaborn Library: <https://github.com/mwaskom/seaborn-data> (<https://github.com/mwaskom/seaborn-data>).
- 2)This dataset is my dataset provided into Github: <https://github.com/tech-dheer-98> (<https://github.com/tech-dheer-98>)

Importing Libraries

In [2]:

```
1 import pandas as pd
2 import matplotlib.pyplot as plt
3 import seaborn as sns
```

1) Line Graph:

Line graphs are a type of chart that shows how a value changes over time or in relation to another variable.

Used for:

In [176]:

1

Importing the Dataset

2

3

emp=pd.read_excel("ESD.xlsx")

4

emp.head(2)

Out[176]:

	EEID	Full Name	Job Title	Department	Business Unit	Gender	Ethnicity	Age	Hire Date	Annual Salary	Bonus %	Country	City	Exit Date
0	E02387	Emily Davis	Sr. Manger	IT	Research & Development	Female	Black	55	2016-04-08	141604	0.15	United States	Seattle	2021-10-16
1	E04105	Theodore Dinh	Technical Architect	IT	Manufacturing	Male	Asian	59	1997-11-29	99975	0.00	China	Chongqing	NaT

In [177]:

1

Reading all columns

2

emp.columns

Out[177]:

Index(['EEID', 'Full Name', 'Job Title', 'Department', 'Business Unit',
 'Gender', 'Ethnicity', 'Age', 'Hire Date', 'Annual Salary', 'Bonus %',
 'Country', 'City', 'Exit Date'],
 dtype='object')

In [179]:

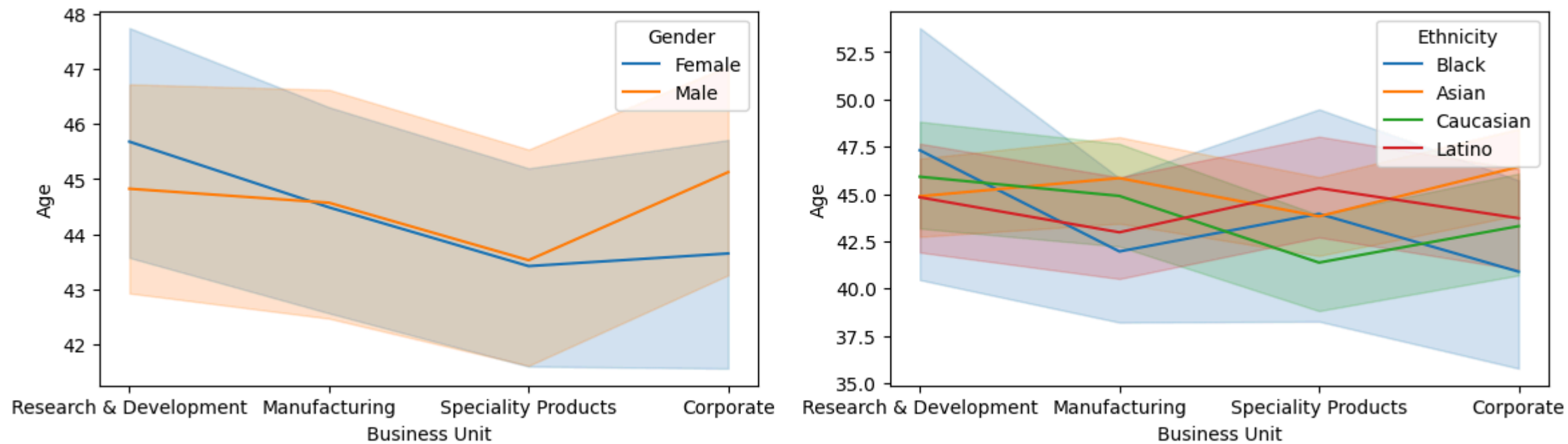
1

emp['Business Unit'].value_counts()

Out[179]:

Manufacturing 269
Speciality Products 265
Corporate 237
Research & Development 229
Name: Business Unit, dtype: int64

```
In [180]: 1 # Plotting the Line Graphs based on Gender and Ethnicity
2 plt.figure(figsize=(14,8),dpi=100)
3
4 plt.subplot(2,2,1)
5 sns.lineplot(data=emp,x=emp['Business Unit'],y=emp.Age,hue='Gender')
6
7 plt.subplot(2,2,2)
8 sns.lineplot(data=emp,x=emp['Business Unit'],y=emp.Age,hue='Ethnicity')
9
10 plt.show()
```



Visual Insights :

1. Graph 1: This depicts the age categories based on gender, distinguishing between males and females.
2. Graph 2: This illustrates the age categories based on ethnicity.

2) Bar Plot:

A bar plot, also known as a bar chart, is a type of graph that shows the relationship between a categorical and numeric variable. The bars can be plotted vertically or horizontally. A vertical bar chart is sometimes called a column chart.

Used for :

A bar chart is used when you want to show a distribution of data points or perform a comparison of metric values across different subgroups of your data.

In [186]:

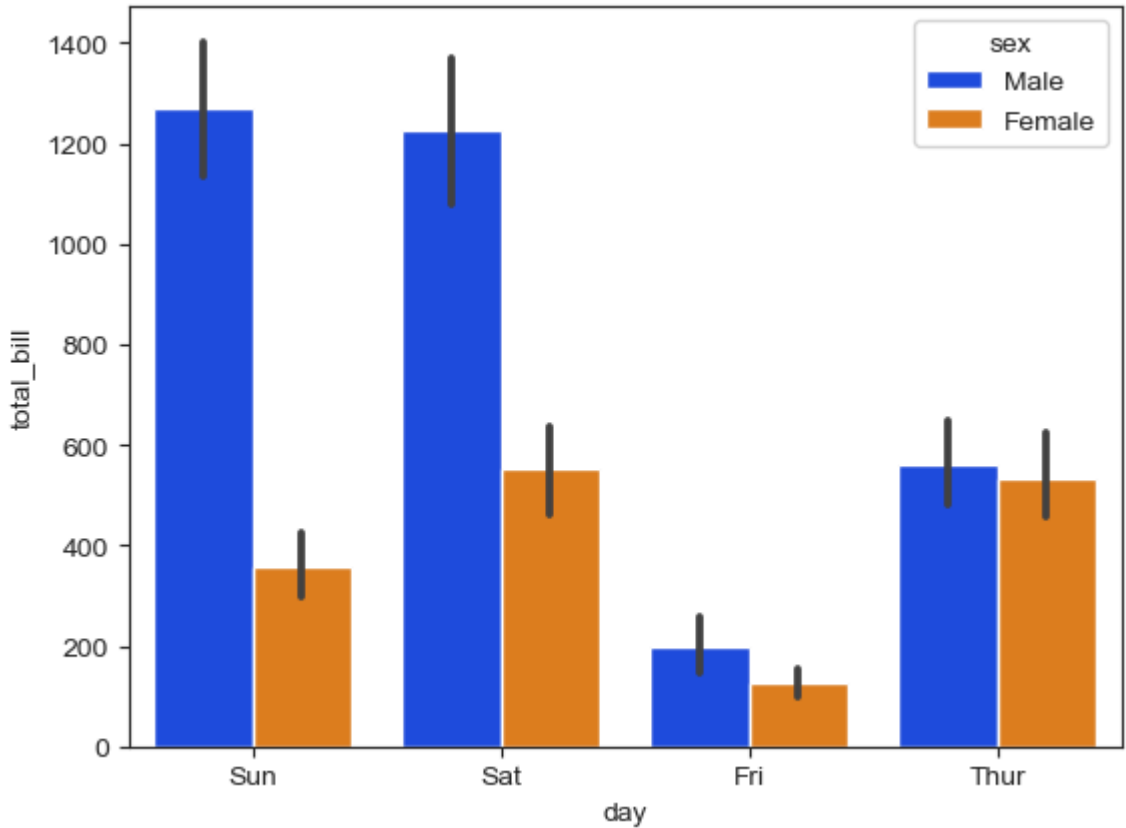
```
1 # Importing the Dataset:
2
3 data=sns.load_dataset("tips")
4 data.head(2)
```

Out[186]:

	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

In [740]:

```
1 # Plotting the Bar Graph
2
3 sns.barplot(data=data,x="day",y="total_bill",estimator="sum",hue="sex",palette='bright',order=["Sun","Sat","Fri","Thur"])
4 plt.show()
5 #estimator=Sum,Mean,Median
6 #hue= Grouping the values.
7 #order=Arranging the values as per choise , for which we are plotting (x-axis)
8 #palette=It is the List of color code we want to give Eg(Blues,spring,etc)..for more refer the seaborn color documentation.
```



Visual Insights :

Based on the graph above, it can be concluded that males spend more than females in restaurants, with the highest expenditures occurring on Saturdays and Sundays.

3) Histogram Plot :

A histogram is a graph showing frequency distributions.

Used for :

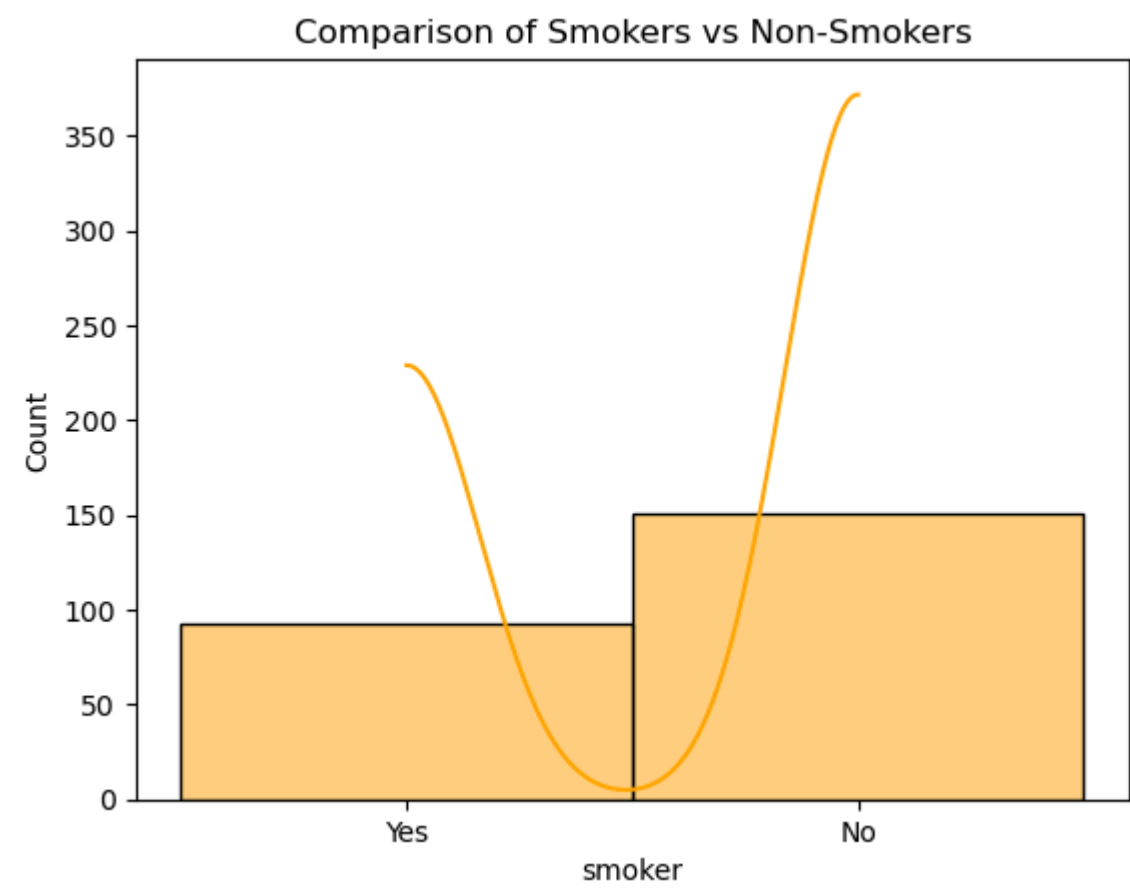
To check the frequency of any dataset.

```
In [231]: 1 # Importing the inbuilt datasets of Seaborn Library:
          2 # Example1:
          3 data=sns.load_dataset("tips")
          4 data.head(2)
```

Out[231]:

	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

```
In [253]: 1 # Plotting the Histogram Graph :
          2
          3 sns.histplot(data=data,x="smoker",kde=True,color='orange')
          4 plt.title("Comparison of Smokers vs Non-Smokers")
          5
          6 # kde=It will Distribution of the graph
          7
          8 plt.show()
```



Visual Inisght:

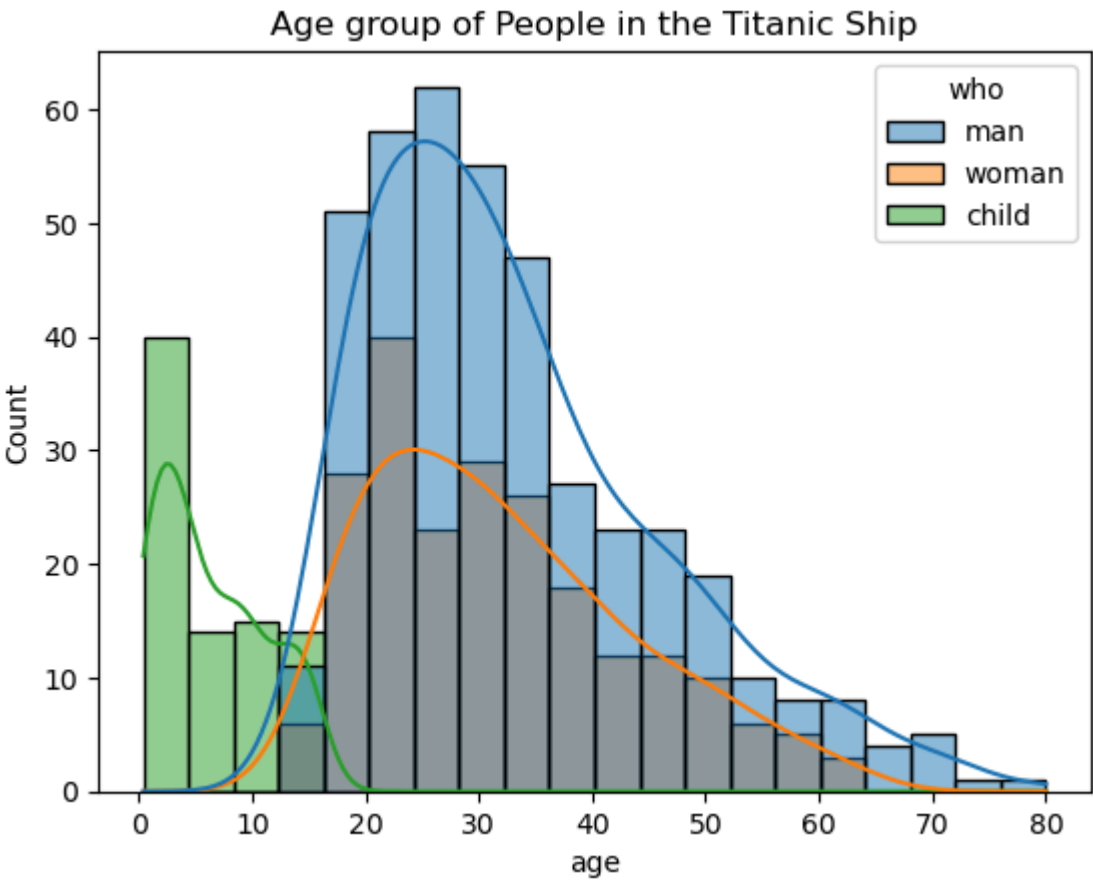
From the above bar graph we can conclude that the count of Non-Smoker is greater then Smoker .

```
In [255]: 1 # Importing the dataset
          2 df=sns.load_dataset("titanic")
          3 df.head(2)
```

Out[255]:

	survived	pclass	sex	age	sibsp	parch	fare	embarked	class	who	adult_male	deck	embark_town	alive	alone
0	0	3	male	22.0	1	0	7.2500	S	Third	man	True	NaN	Southampton	no	False
1	1	1	female	38.0	1	0	71.2833	C	First	woman	False	C	Cherbourg	yes	False

```
In [258]: 1 # Plotting the histogram Graph
          2
          3 plt.title("Age group of People in the Titanic Ship")
          4 sns.histplot(data=df,x="age",kde=True,hue="who")
          5
          6 plt.show()
```



Visual Insights:

- 1)Count of Age group of Child between 0-5 years , is highesh
- 2)Count of Age Group of Women between 20-25years , is highest
- 3)Count of Age Group of Men between 15-35years ,is highest

4) Scatter Plot:

A scatter plot is a diagram where each value in the data set is represented by a dot. It is created by plotting the values of one variable on the x-axis and the values of the other variable on the y-axis. Each point on the plot represents a single data point.

Used for:

Scatter plots' primary uses are to observe and show relationships between 2 numeric variables.

In [261]:

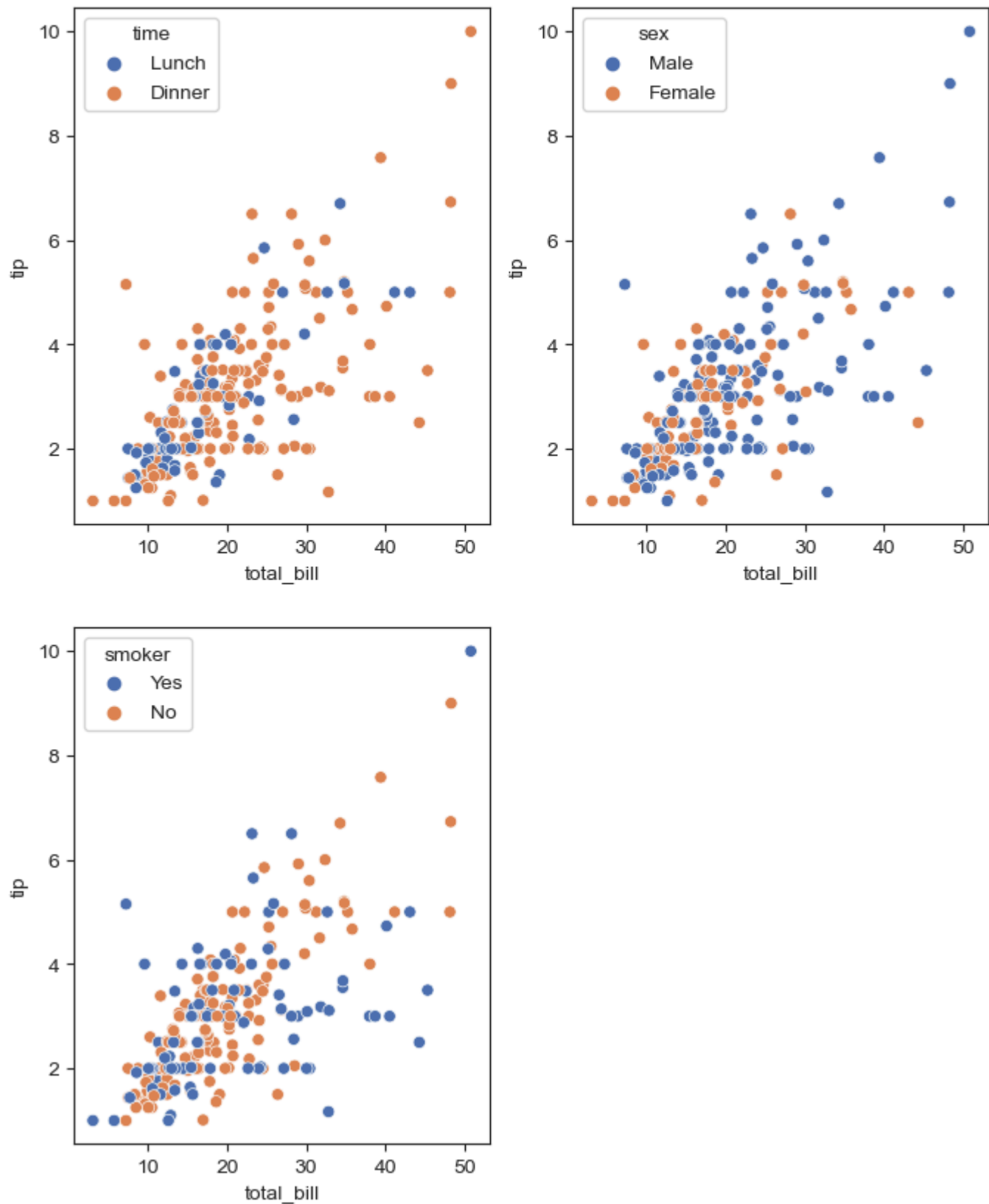
```
1 # Importing the Dataset:
2 df=sns.load_dataset("tips")
3 df.head(2)
```

Out[261]:

	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

```
In [741]: 1 #Plotting the Scatter Plot
2
3 plt.figure(figsize=(8,10))
4 plt.suptitle("Tips Analysis",size=30,color='green')
5 plt.subplot(2,2,1)
6 sns.scatterplot(data=df,x='total_bill',y='tip',hue='time',marker="o",palette='deep')
7
8 plt.subplot(2,2,2)
9 sns.scatterplot(data=df,x='total_bill',y='tip',hue='sex',marker="o",palette='deep')
10
11
12 plt.subplot(2,2,3)
13 sns.scatterplot(data=df,x='total_bill',y='tip',hue='smoker',marker="o",palette='deep')
14
15
16
17 plt.show()
```


Tips Analysis



Visual Insights:

- 1. Graph1: We conclude from this Scatter chart that ,Maximum Tips received was during dinner time.

- 2. Garph2: We conclude from this Scatter chart that , Maximum Tips received was from Males.
- 3. Graph3: We conclude from this Scatter chart that, Maximum Tips received was from Non-Smokers.

5) Heat Map :

A heatmap is a graphical representation of data where the individual values contained in a matrix are represented as colors

Used for :

Heatmaps are a powerful tool for visualizing data. They can be used to identify patterns and trends in data, and to compare different datasets.

```
In [307]: 1 # Importing the Dataset:
          2
          3 data=sns.load_dataset("tips")
          4 data.head(2)
```

Out[307]:

	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

```
In [319]: 1 # Method 1
          2 gb=data.groupby("day")["tip"].mean()
          3 gb
```

Out[319]:

day	
Thur	2.771452
Fri	2.734737
Sat	2.993103
Sun	3.255132
Name: tip, dtype: float64	

```
In [320]: 1 # Method 2: using the Aggregate Function
          2 agb=data.groupby("day").agg({"tip": "mean"})
          3 agb
```

Out[320]:

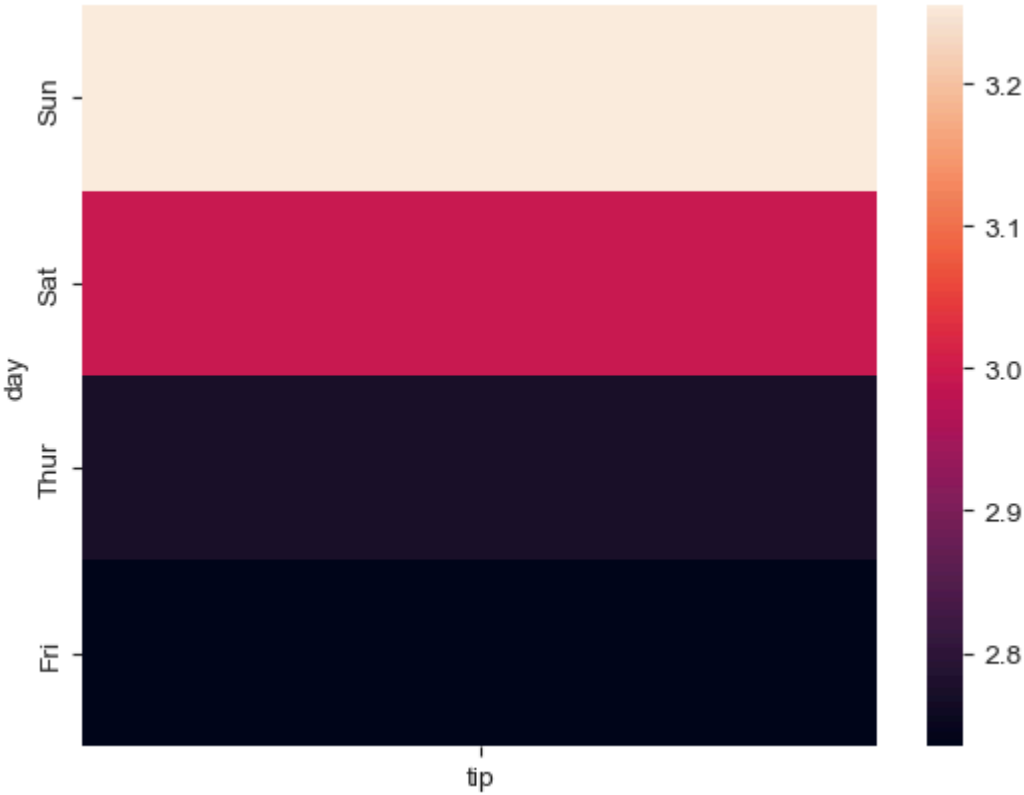
	tip
day	
Thur	2.771452
Fri	2.734737
Sat	2.993103
Sun	3.255132

```
In [743]: 1 # Sorting the values into descending order:
          2 agb_sort=agb.sort_values("tip",ascending=False)
          3 agb_sort
```

Out[743]:

	tip
day	
Sun	3.255132
Sat	2.993103
Thur	2.771452
Fri	2.734737

```
In [745]: 1 # Plotting the Heatmap
          2
          3 sns.heatmap(data=agb_sort)
          4
          5 plt.show()
```



Visual Inisghts :

From the above Heatmap we can conclude that Darker the color , lesser the tip received .

In [328]:

```
1 # Example 2: Taking the Employee Dataset
2
3 emp.head(2)
```

Out[328]:

	EEID	Full Name	Job Title	Department	Business Unit	Gender	Ethnicity	Age	Hire Date	Annual Salary	Bonus %	Country	City	Exit Date
0	E02387	Emily Davis	Sr. Manger	IT	Research & Development	Female	Black	55	2016-04-08	141604	0.15	United States	Seattle	2021-10-16
1	E04105	Theodore Dinh	Technical Architect	IT	Manufacturing	Male	Asian	59	1997-11-29	99975	0.00	China	Chongqing	NaT

In [340]:

```
1 # Using the Aggregate Function
2 egb=emp.groupby("Business Unit").agg({"Annual Salary":"mean"})
3 egb
4
```

Out[340]:

	Annual Salary
Business Unit	
Corporate	120776.151899
Manufacturing	103651.371747
Research & Development	112490.205240
Speciality Products	116796.003774

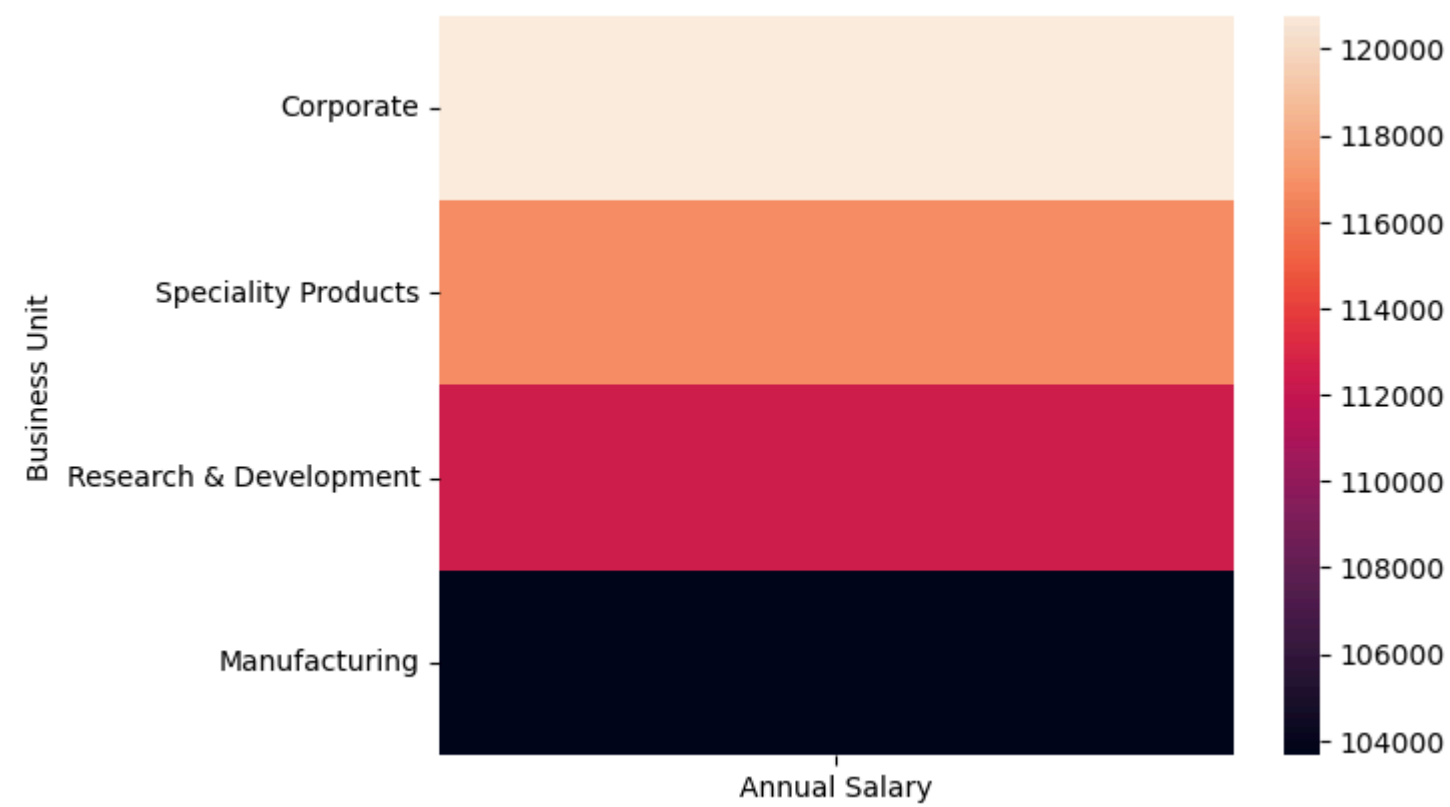
In [342]:

```
1 #Sorting the values in descending order:
2 egb_sorted=egb.sort_values("Annual Salary",ascending=False)
3 egb_sorted
```

Out[342]:

	Annual Salary
Business Unit	
Corporate	120776.151899
Speciality Products	116796.003774
Research & Development	112490.205240
Manufacturing	103651.371747

```
In [343]: 1 # Plotting the Heatmap for the given value:
          2
          3 sns.heatmap(egb_sorted)
          4
          5 plt.show()
```



Visual Insights:

From the above Heatmap we can conclude that Corporate Department receive the Highest Salary, whereas Manufacturing Department received the lowest Salary.

6) Countplot Graph:

The countplot is used to represent the occurrence(counts) of the observation present in the categorical variable

Used for:

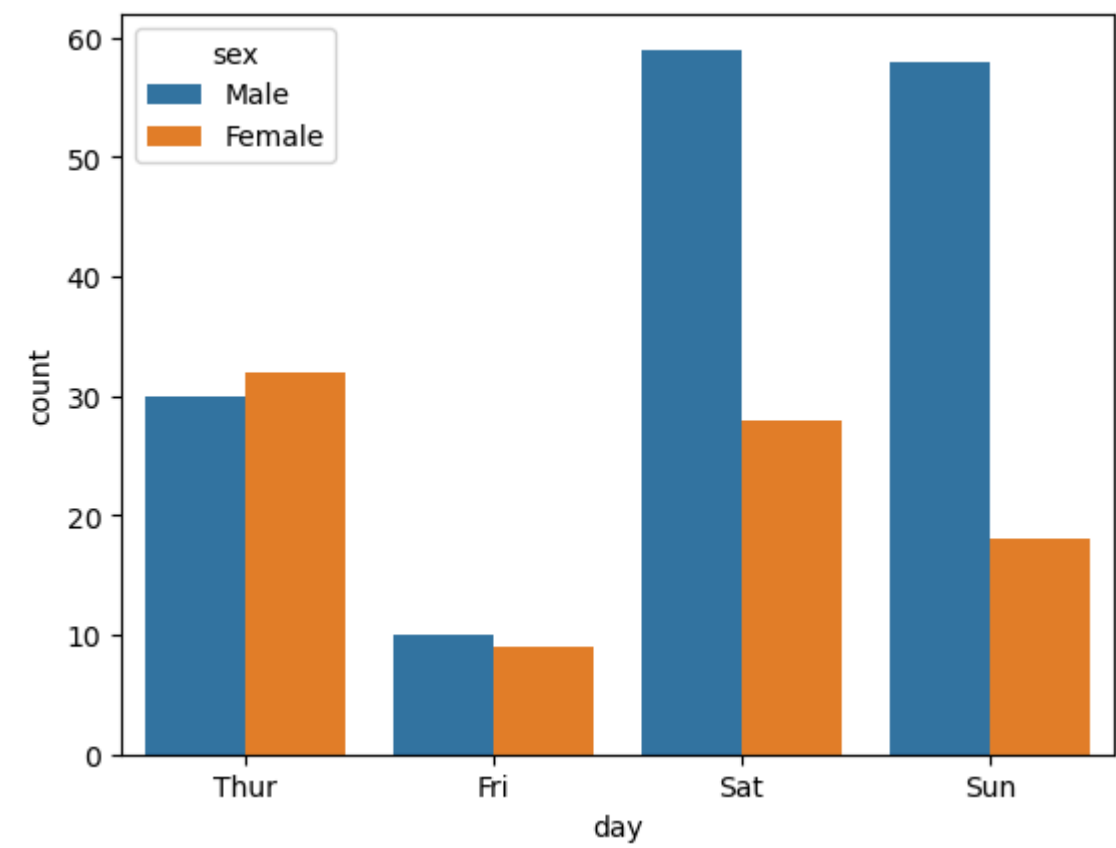
It uses the concept of a bar chart for the visual depiction.

```
In [347]: 1 # Using the inbuilt dataset
          2 data=sns.load_dataset("tips")
          3 data.head(2)
```

Out[347]:

	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

```
In [423]: 1 # Plotting the Countplot
          2 sns.countplot(data=data,x='day',hue='sex')
          3
          4 plt.show()
```



Visual Insights:

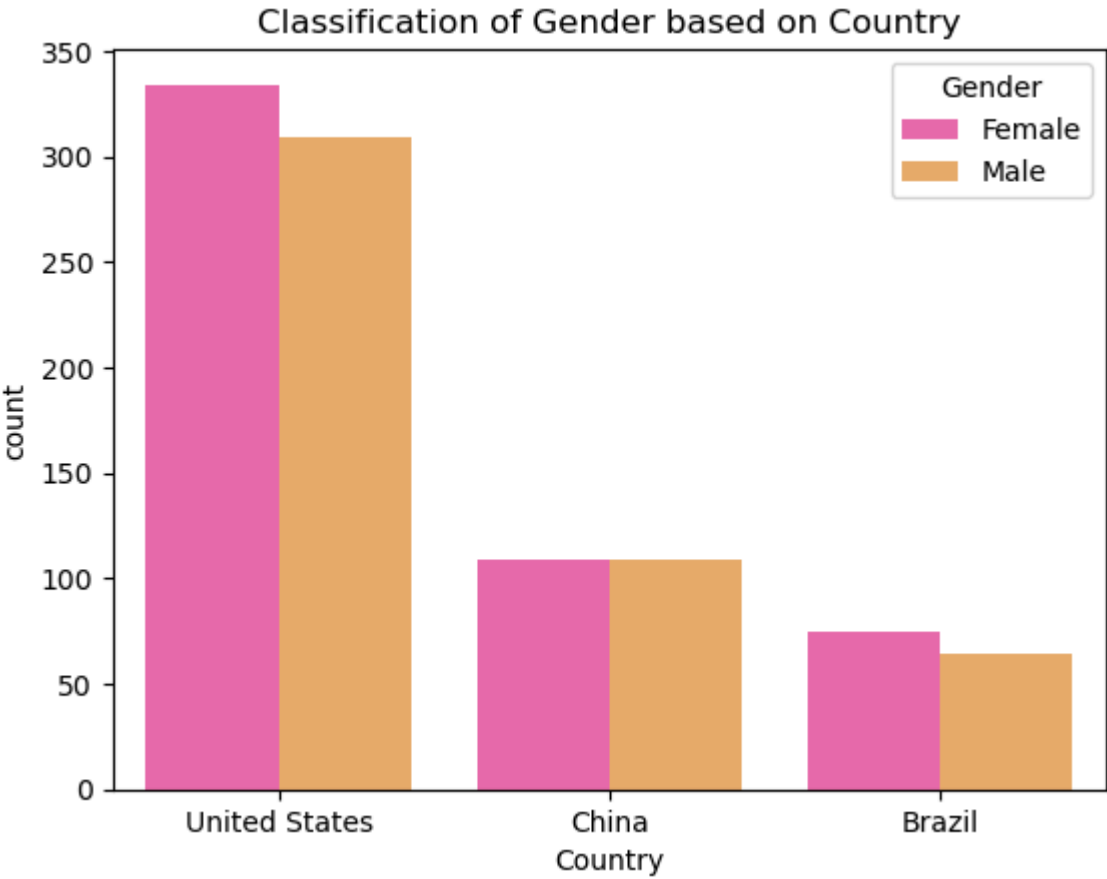
From the above Countplot we can conclude that,Females visit the resturant less on the weekends compared to Males.

```
In [356]: 1 # Import the Employee dataset
          2 emp.head(2)
```

Out[356]:

	EEID	Full Name	Job Title	Department	Business Unit	Gender	Ethnicity	Age	Hire Date	Annual Salary	Bonus %	Country	City	Exit Date
0	E02387	Emily Davis	Sr. Manger	IT	Research & Development	Female	Black	55	2016-04-08	141604	0.15	United States	Seattle	2021-10-16
1	E04105	Theodore Dinh	Technical Architect	IT	Manufacturing	Male	Asian	59	1997-11-29	99975	0.00	China	Chongqing	NaT

```
In [422]: 1 sns.countplot(data=emp,x='Country',hue='Gender',palette="spring")
          2 plt.title("Classification of Gender based on Country")
          3
          4 plt.show()
```



Viusual Insights :

- 1)From the above countplot graph we can conclude that United States rank the highest number of working Employees.
- 2)Also it can be concluded that China ranks equal gender working professionals.

7) Violin Plot:

A violin plot is a statistical graph that shows the distribution of numerical data for one or more groups

Used for :

Violin plots are used to visualize the distribution of numerical data.

```
In [425]: 1 # Importing the Inbuit seaborn dataset
          2 data.head(2)
```

Out[425]:

	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

```
In [461]: 1 # Plotting the violin plot
          2
          3 plt.figure(figsize=(12,6))
          4
          5 plt.subplot(2,2,1)
          6 sns.violinplot(data=data,x="total_bill")
          7
          8 plt.subplot(2,2,2)
          9 sns.violinplot(data=data,x="tip")
         10
         11 plt.suptitle("Distribution of Tip and Total Bills ")
         12
         13
         14 plt.show()
```



Visual Insights:

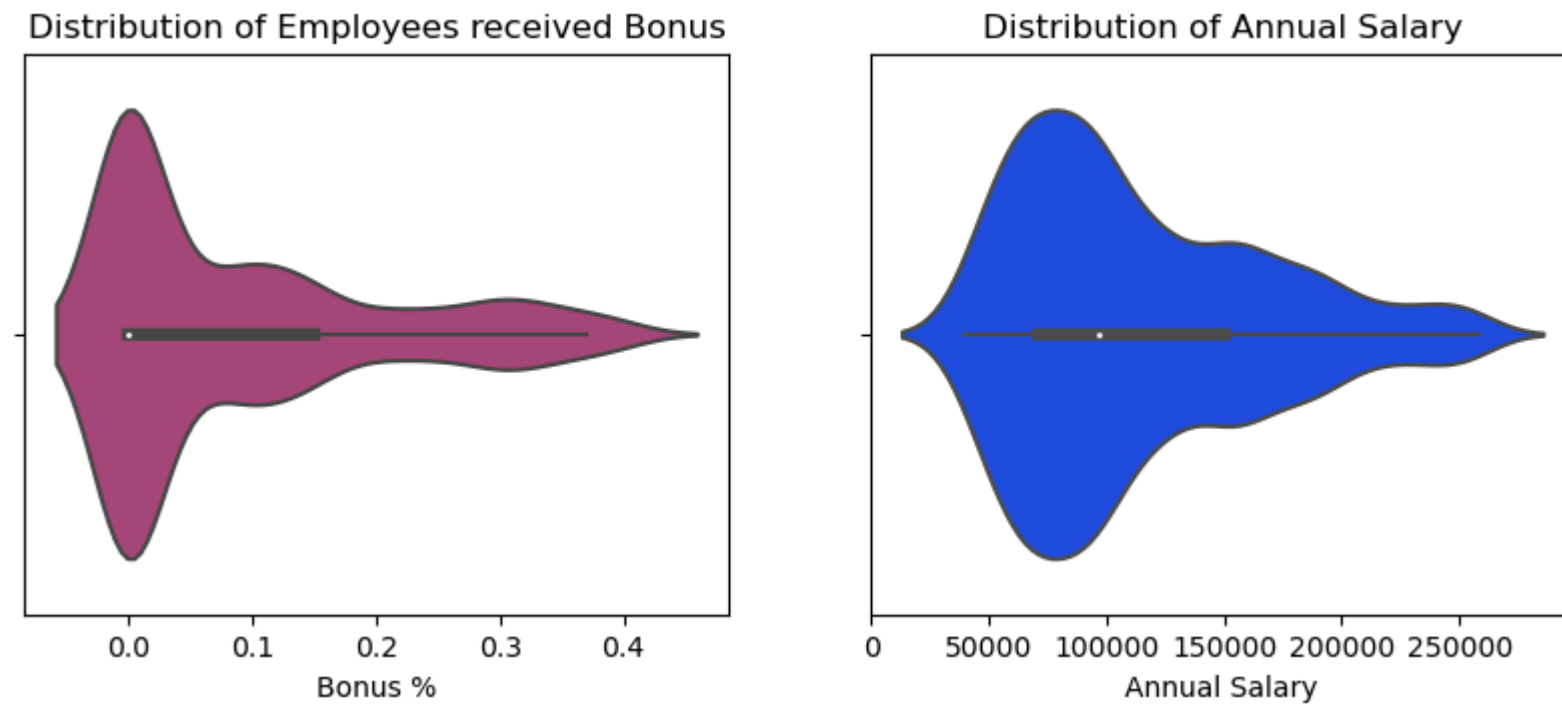
- 1)The maximum amounts of bill is between 10 to 20(dollars).
- 2)The maximum tips received is 2 (dollars).

```
In [462]: 1 # Importing the employee dataset
          2 emp.head(2)
```

Out[462]:

	EEID	Full Name	Job Title	Department	Business Unit	Gender	Ethnicity	Age	Hire Date	Annual Salary	Bonus %	Country	City	Exit Date
0	E02387	Emily Davis	Sr. Manger	IT	Research & Development	Female	Black	55	2016-04-08	141604	0.15	United States	Seattle	2021-10-16
1	E04105	Theodore Dinh	Technical Architect	IT	Manufacturing	Male	Asian	59	1997-11-29	99975	0.00	China	Chongqing	NaT


```
In [484]: 1 # Plotting the Violin Plot
2 plt.figure(figsize=(10,8))
3 plt.subplot(2,2,1)
4 sns.violinplot(data=emp,x='Bonus %',palette="magma")
5 plt.title("Distribution of Employees received Bonus")
6
7 plt.subplot(2,2,2)
8 sns.violinplot(data=emp,x='Annual Salary',palette="bright")
9 plt.title("Distribution of Annual Salary")
10
11 plt.show()
```



Visual Insights:

- 1) From the Graph we can conclude that maximum employees haven't received any Bonues.
- 2) From the Graph we can conclude that Annual Average Salary of Employees is under 1 Lakh.

8) Pairplot:

A pair plot, also called a scatterplot matrix, is a matrix of graphs that shows the relationship between each pair of variables in a dataset. It combines scatter plots and histograms to provide a unique overview of the dataset's distributions and correlations

Used for:

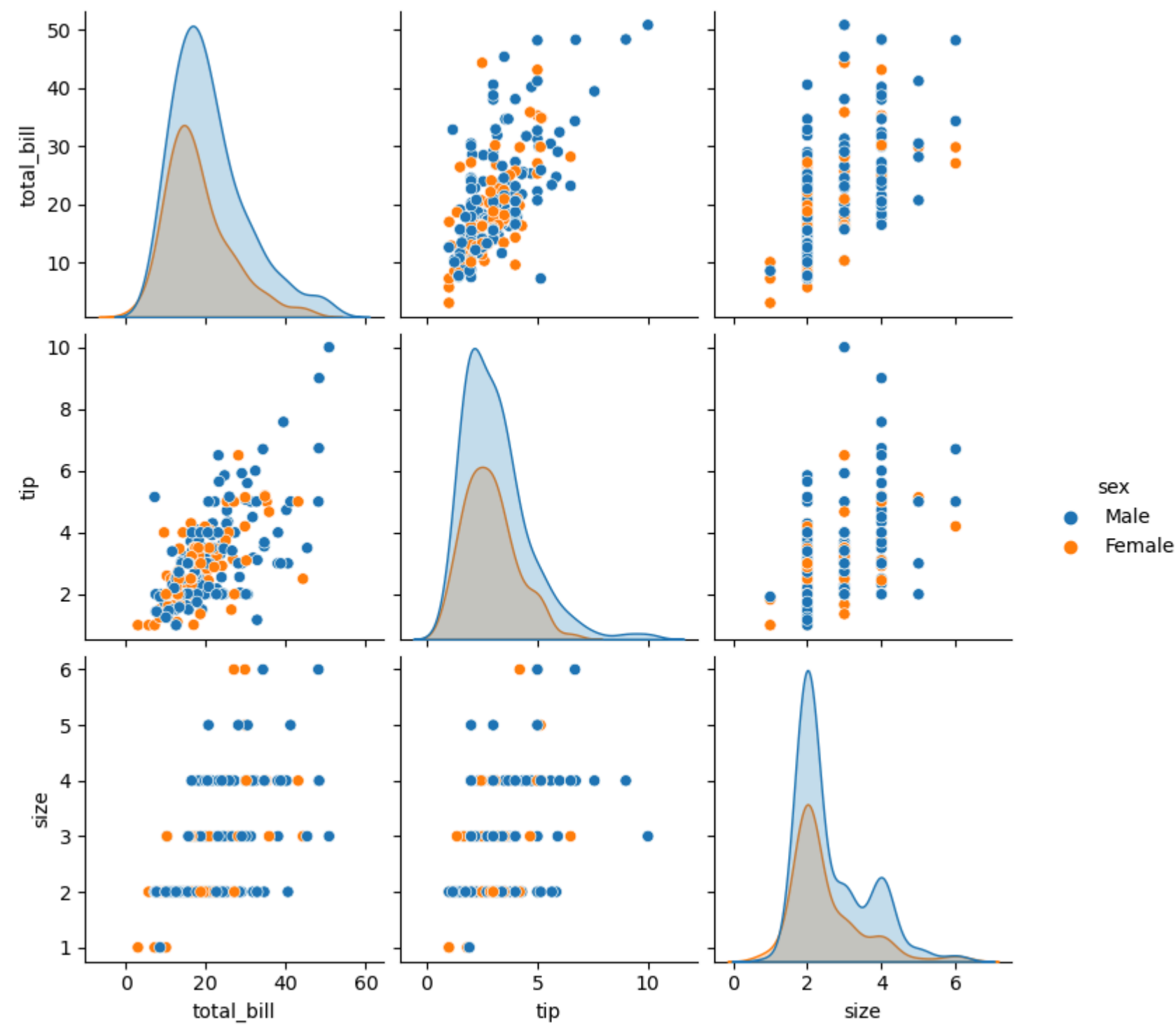
It's a tool used for exploring relationships and distributions in a dataset

```
In [493]: 1 data.head(2)
```

Out[493]:

	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

```
In [498]: 1 # Plotting the Pair Plot
2
3 sns.pairplot(data=data,hue='sex')
4
5 plt.show()
```



9) Strip Plot:

A strip plot is a type of data visualization that shows the distribution of a single-dimensional value. It's also known as a dot plot, individual value plot, or single-axis scatter plot.

Used for:

It is used to visualise the distribution of many individual one-dimensional values.

In [500]:

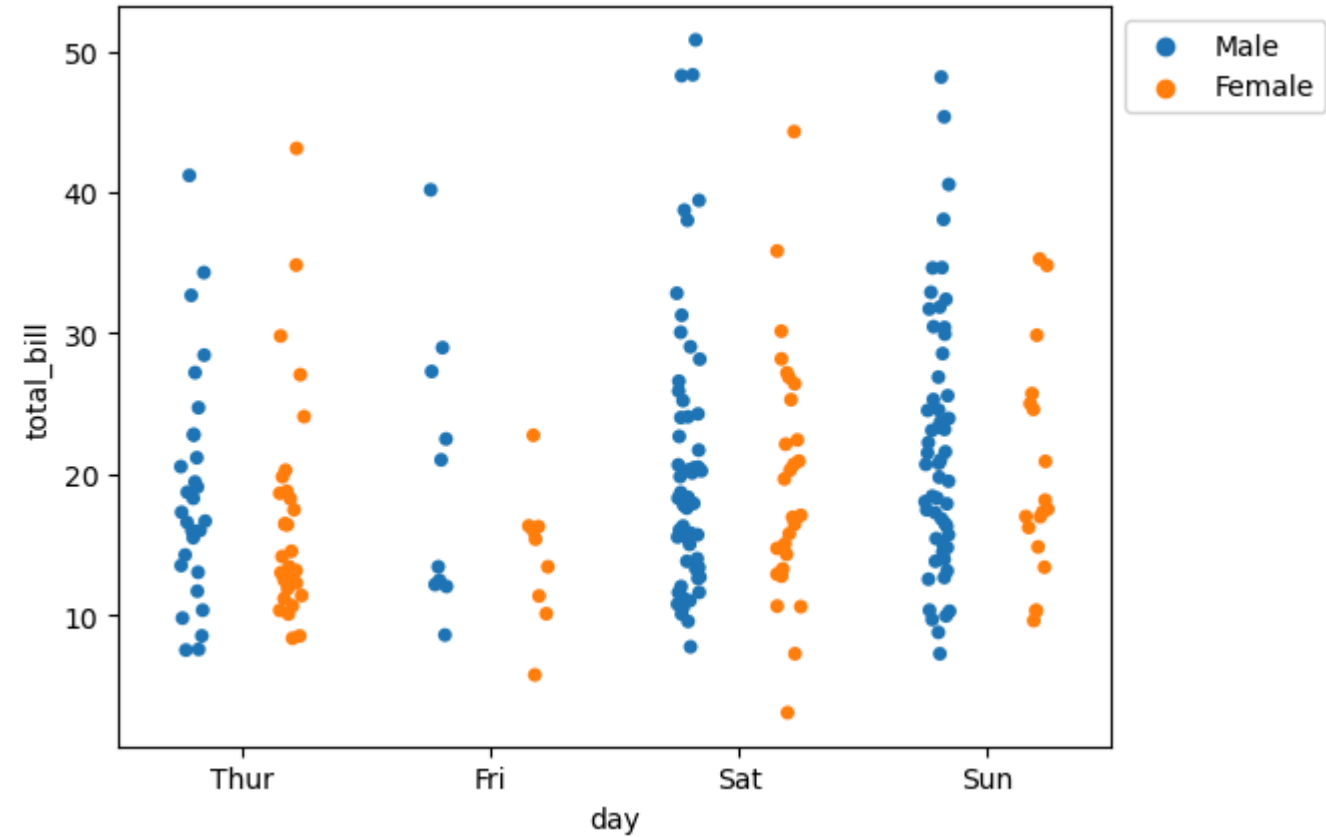
```
1 # Importing the Inbult dataset
2
3 data.head(2)
```

Out[500]:

	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

In [527]:

```
1 # Plotting the Strip Plot
2
3 sns.stripplot(data=data,x='day',y='total_bill',hue='sex',dodge=True,jitter=0.1)
4
5 plt.legend(bbox_to_anchor=(1,1))
6
7 plt.show()
```



Visual Insights:

From the above Strip Plot it can be concluded that,on Thru and Friday resturant made the less sales compared to weekend.

10) Box Plot :

A box plot, also known as a box and whisker plot, is a graph that summarizes a set of data.

Used for:

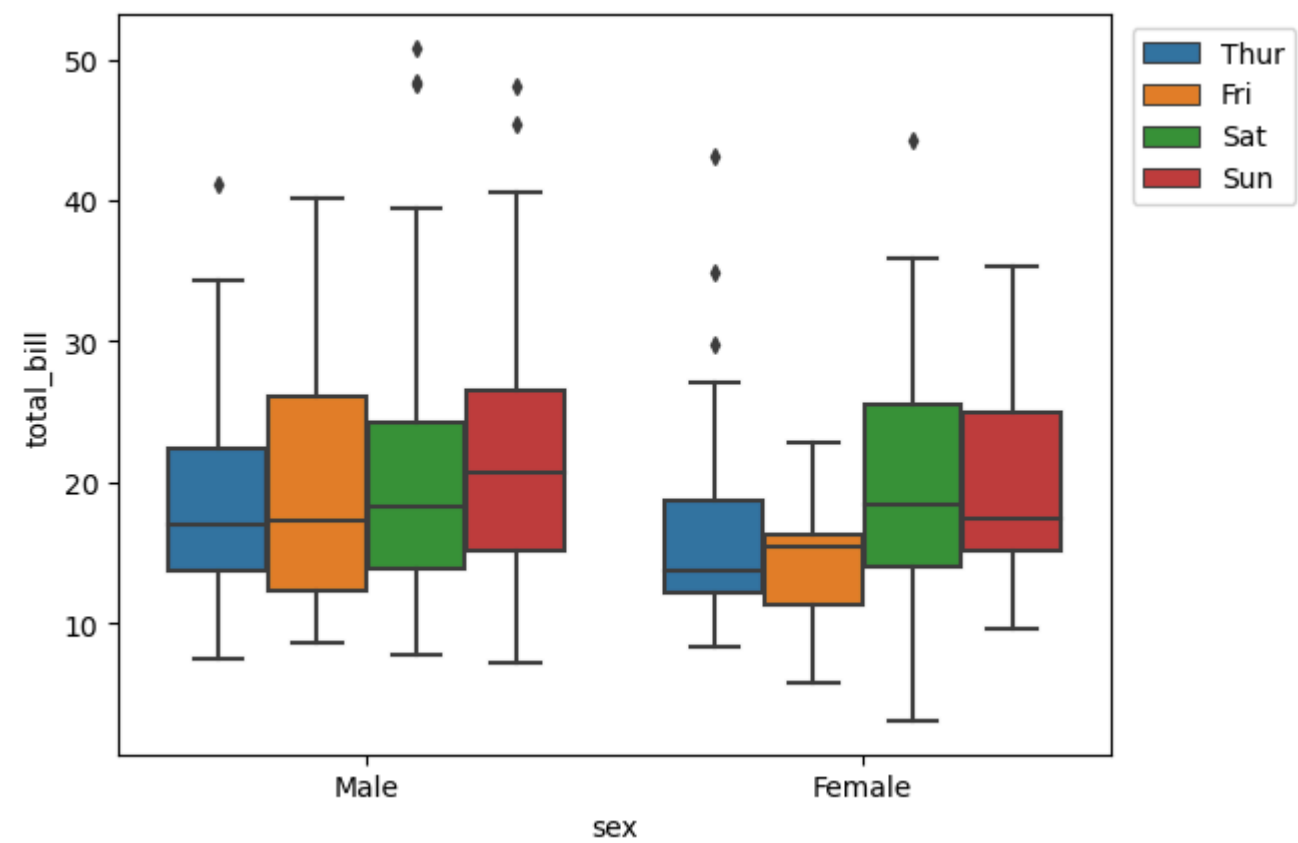
Box plots are used to show distributions of numeric data values, especially when you want to compare them between multiple groups

```
In [529]: 1 # Importing the Dataset
          2 data.head(2)
```

Out[529]:

	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

```
In [558]: 1 # Plotting the Box plot
          2
          3 sns.boxplot(data=data,x='sex',y='total_bill',hue='day',fliersize=4) #fliersize= adjust the size of the outliers.
          4 plt.legend(bbox_to_anchor=(1.2,1)) #This adjust the positon of the label
          5 plt.show()
```



Visual Insights:

From the above Box Plot we can conclude that, Males and Female's Average count of Bill is under 20(Dollars).

11) Cat Plot(Categorical Plot):

These kinds of plots allow us to choose a numerical variable, like age, and plot the distribution of age for each category in a selected categorical variable.

Used for:

Categorical plots are used to visualize the distribution of categories within a dataset. They can also be used to compare the representations of data across different plots.

In [559]:

1

importing the Dataset"

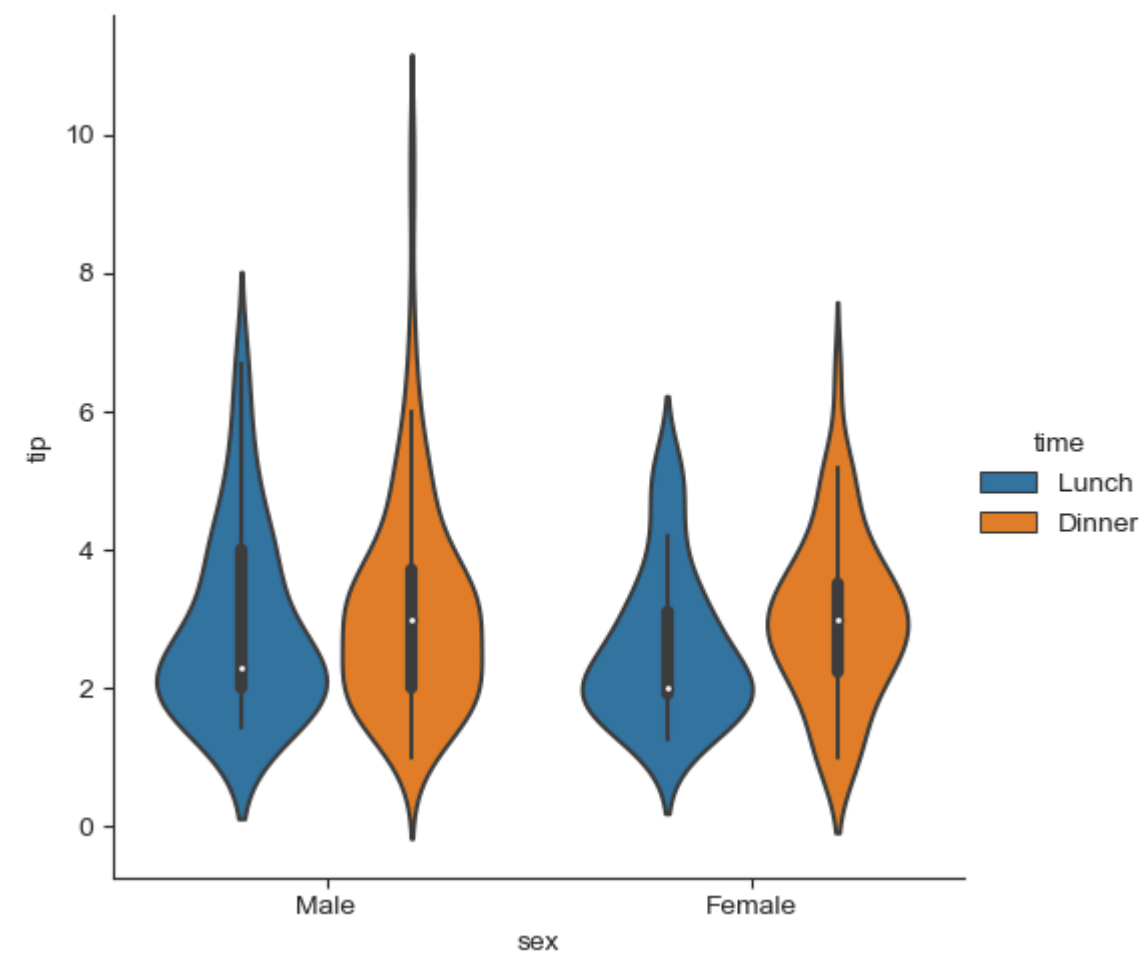
2

data.head(2)

Out[559]:

	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

```
In [749]: 1 # Plotting the Cat Plot:
2
3 sns.catplot(data=data,x='sex',y='tip',hue='time',jitter=0.2,dodge=True,kind='violin')
4 #jitter= it means the dataset points are shifted by the number assigned to jitter
5 #dodge:It is used to rearrang the dataset values.
6 #kind: It is used to plot the type of graph
7 plt.show()
```



Visual Insights:

From the above Catplot it can be concluded that, Female and Male contributes more during Dinner.

```
In [610]: 1 # Importing the Dataset from Inbuilt :
2
3 new_data=sns.load_dataset("exercise")
4 new_data.head()
```

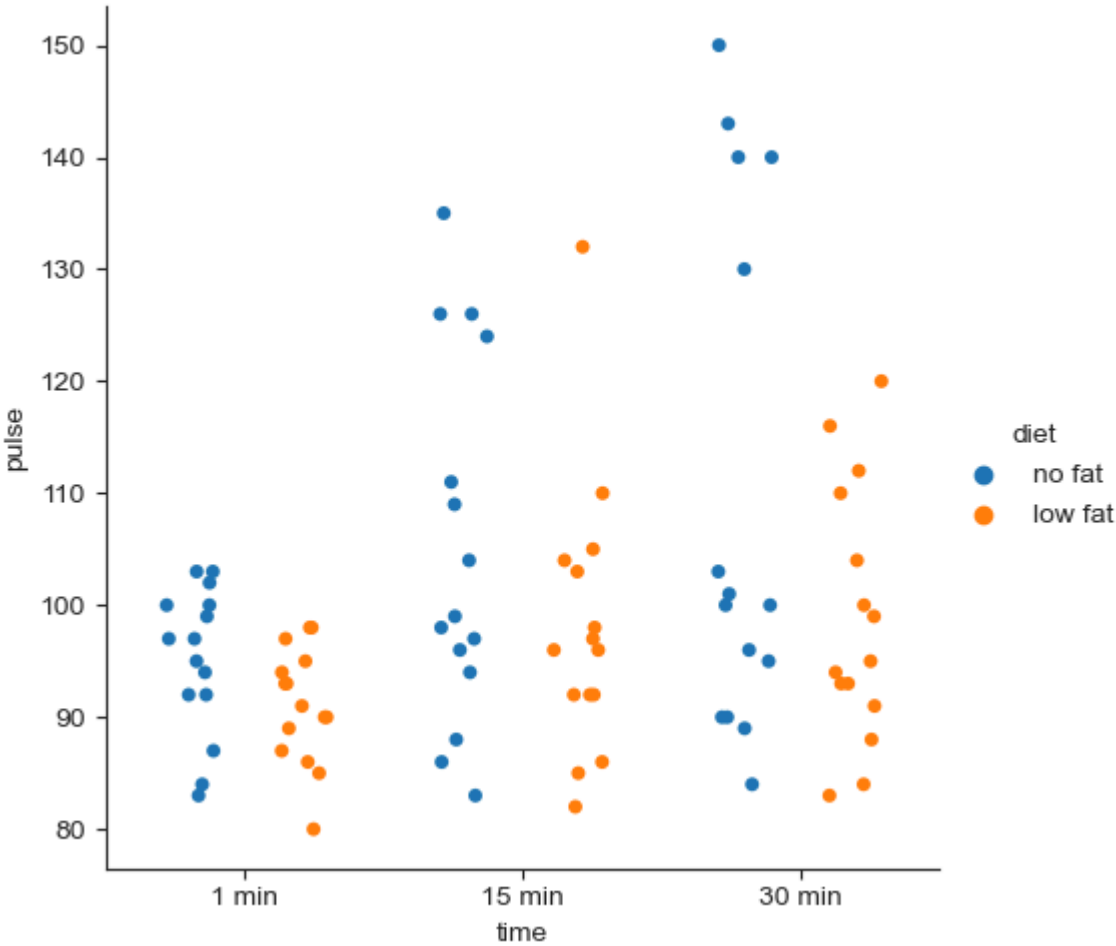
Out[610]:

	Unnamed: 0	id	diet	pulse	time	kind
0	0	1	low fat	85	1 min	rest
1	1	1	low fat	85	15 min	rest
2	2	1	low fat	88	30 min	rest
3	3	2	low fat	90	1 min	rest
4	4	2	low fat	92	15 min	rest

```
In [611]: 1 new_data.dtypes
```

Out[611]: Unnamed: 0 int64
id int64
diet category
pulse int64
time category
kind category
dtype: object

```
In [639]: 1 sns.catplot(data=new_data,x='time',y='pulse',dodge=True,jitter=0.2,hue='diet')  
2 sns.set_style(style='ticks')  
3 plt.show()
```



Visual Insights:

It can be concluded from the above Catplot that ,more the time spend over at workout , it increses the pulse.

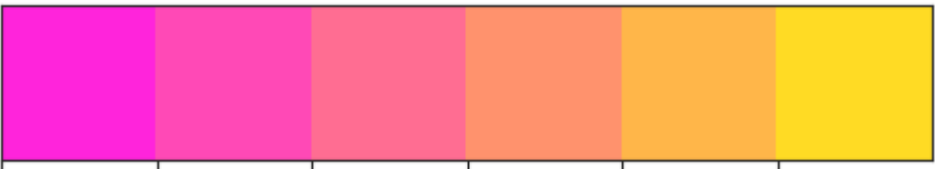
12) Color Palletes:

It is used to select diffferent types of color shades.

```
In [651]: 1 sns.palplot(sns.color_palette('viridis'))
          2 plt.show()
```



```
In [652]: 1 sns.palplot(sns.color_palette('spring'))
          2 plt.show()
```



```
In [654]: 1 sns.palplot(sns.color_palette('Blues'))
          2 plt.show()
```



```
In [656]: 1 sns.palplot(sns.color_palette("deep"))
```



13) Multiple Plots :

It is used to create different plots at a time .

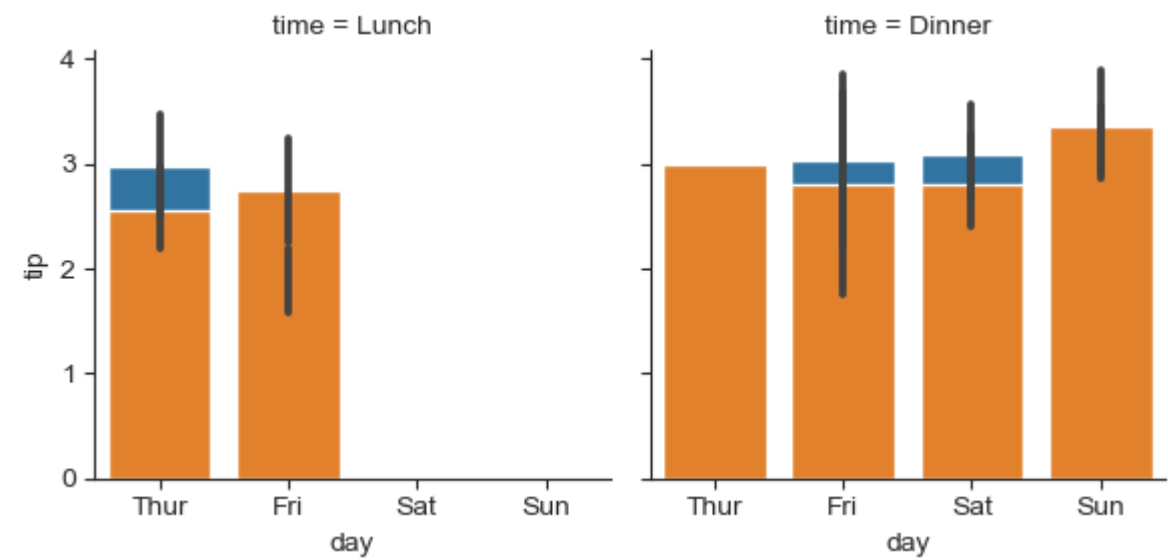
```
In [658]: 1 # Importing the Dataset
          2
          3 data.head(2)
```

Out[658]:

	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


```
In [750]: 1 # Plotting the Graphs
2
3 a=sns.FacetGrid(data,col='time',hue='sex') #col: It is used to create subplots of the graph
4 a.map(sns.barplot,"day","tip")
5
6 plt.show()
```

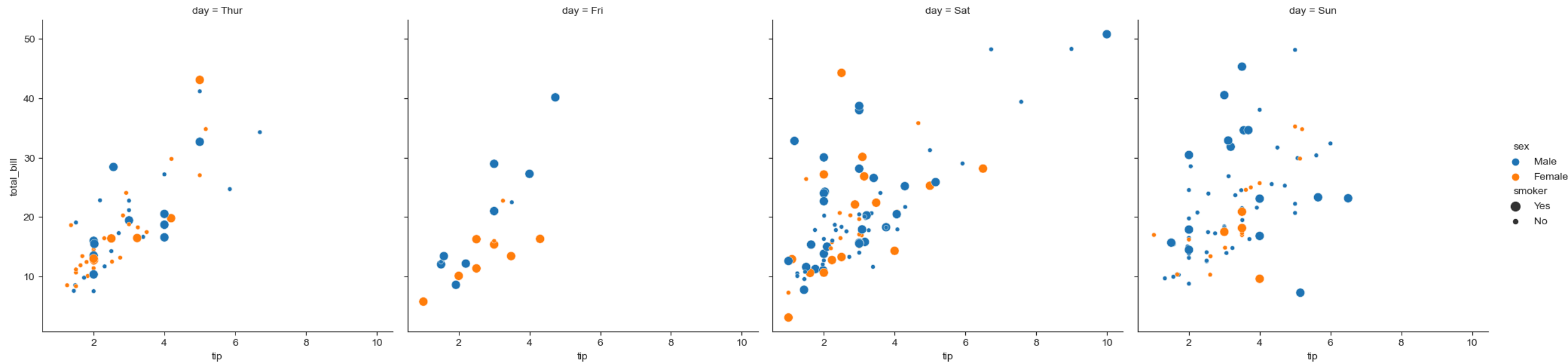
C:\Users\HP\anaconda3\Lib\site-packages\seaborn\axisgrid.py:712: UserWarning: Using the barplot function without specifying `order` is likely to produce an incorrect plot.
warnings.warn(warning)



14) Relational Plots :

A relational plot is a graphical technique that visualizes the statistical relationship between data points.

```
In [691]: 1 # Importing the dataset
2
3 sns.relplot(data=data,x='tip',y='total_bill',col='day',hue='sex',size='smoker')
4
5 plt.show()
```



15) Swarm Plot:

The Data does not overlap on each other ,like Scatter plot

In [692]:

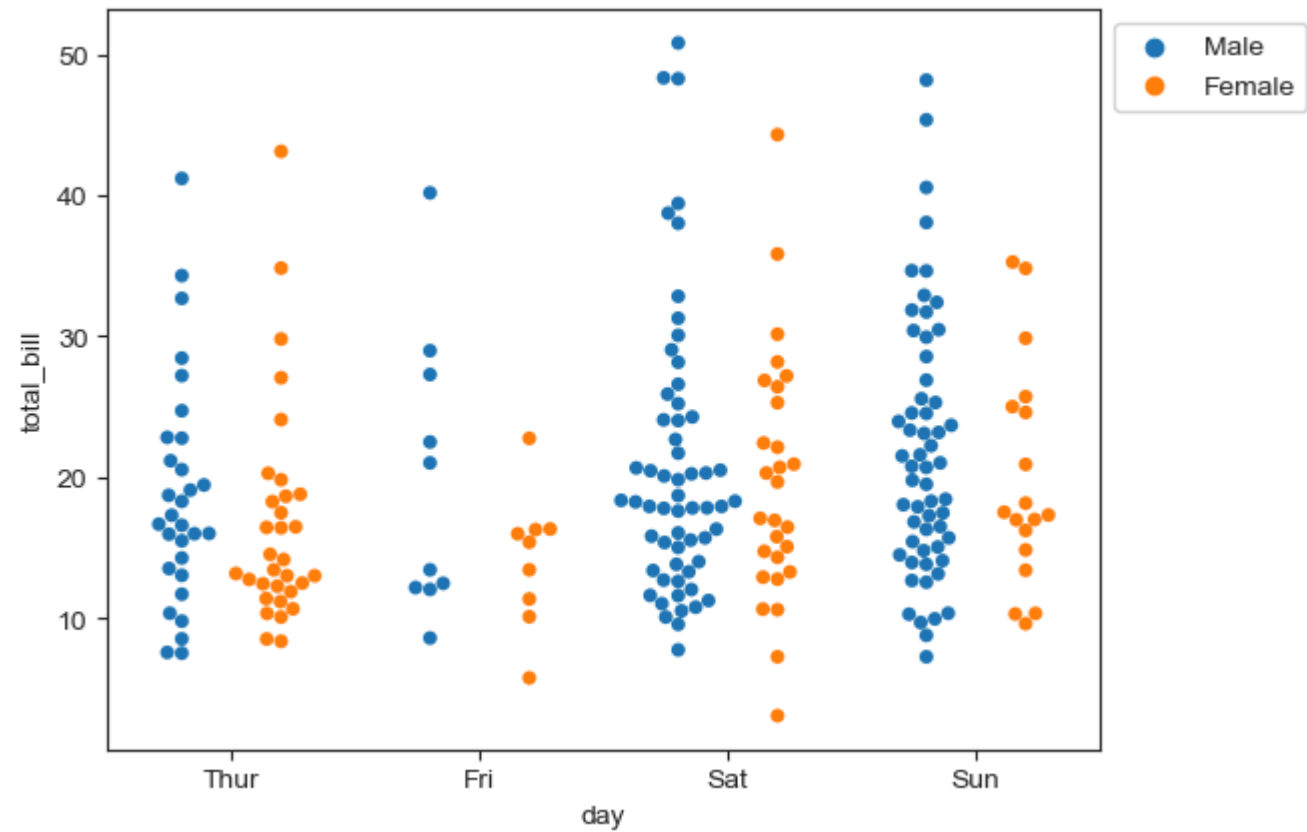
```
1 # Importing the data
2 data.head(2)
```

Out[692]:

	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

In [698]:

```
1 # Plotting the Swarm Plot:
2
3 sns.swarmplot(data=data,x='day',y='total_bill',hue='sex',dodge=True)
4 plt.legend(bbox_to_anchor=(1,1))
5 plt.show()
```



16) KDE Plot (Kernel Density Estimation):

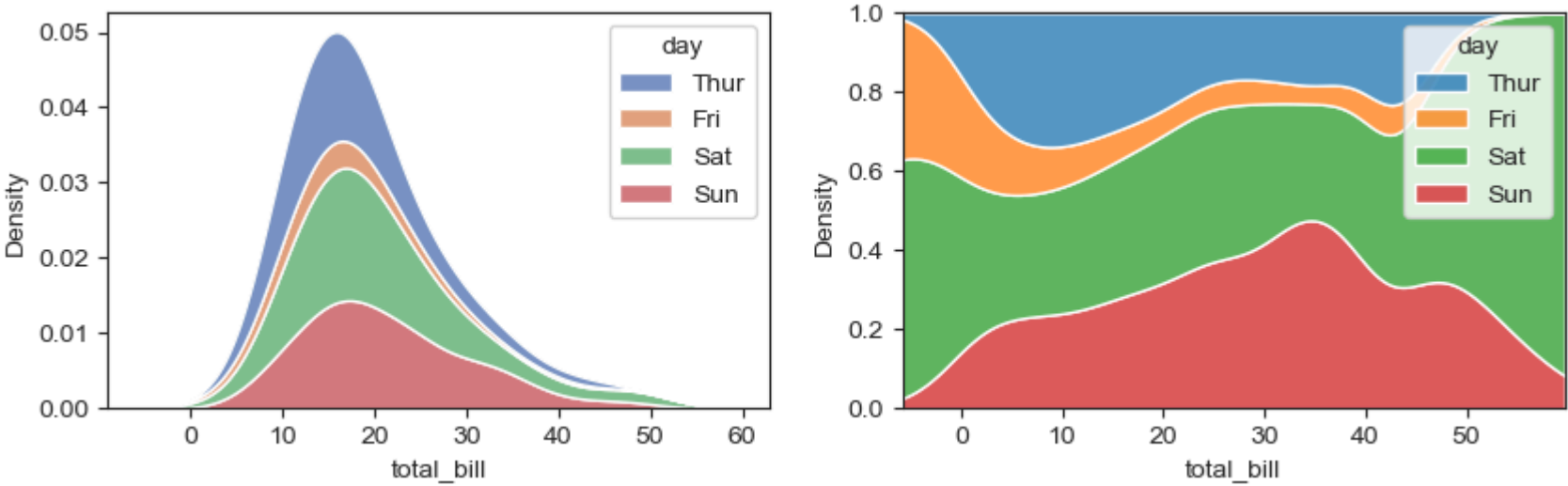
Here we get the frequency of data in the form of wave.

```
In [727]: 1 # Importing the data
          2 data.head(2)
```

Out[727]:

	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

```
In [725]: 1 plt.figure(figsize=(10,6))
          2
          3 plt.subplot(2,2,1)
          4 sns.kdeplot(data=data,x='total_bill',hue='day',multiple='stack',palette='deep')
          5
          6 plt.subplot(2,2,2)
          7 sns.kdeplot(data=data,x='total_bill',hue='day',multiple='fill')
          8
          9 plt.show()
         10
```



Visual Insights:

- 1)From the above graphs we can conclude that Sunday contributes the larger area and Thursday contributes the lesser Area.
- 2)Also we can see that Friday contributes the less bill amount.