SESSIONS 5

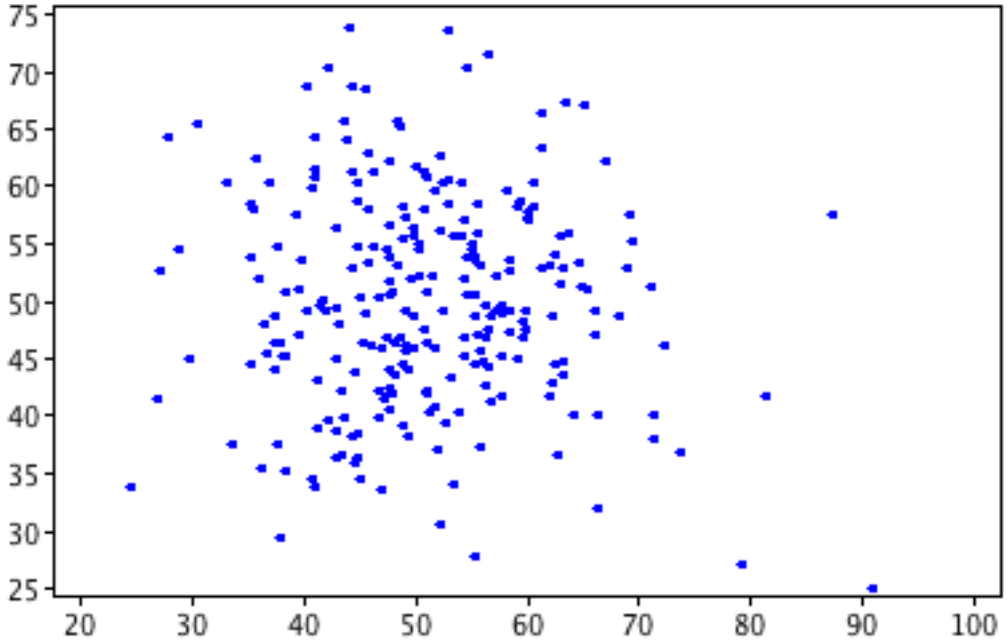
Scatter Plot

Data Science Program

Purwadhika
Startup and Coding School

Outline

- What is Scatter Plot?
- When to use Scatter Plot
- When to avoid Scatter Plot
- Create Scatter Plot using Matplotlib, Seaborn, and Pandas



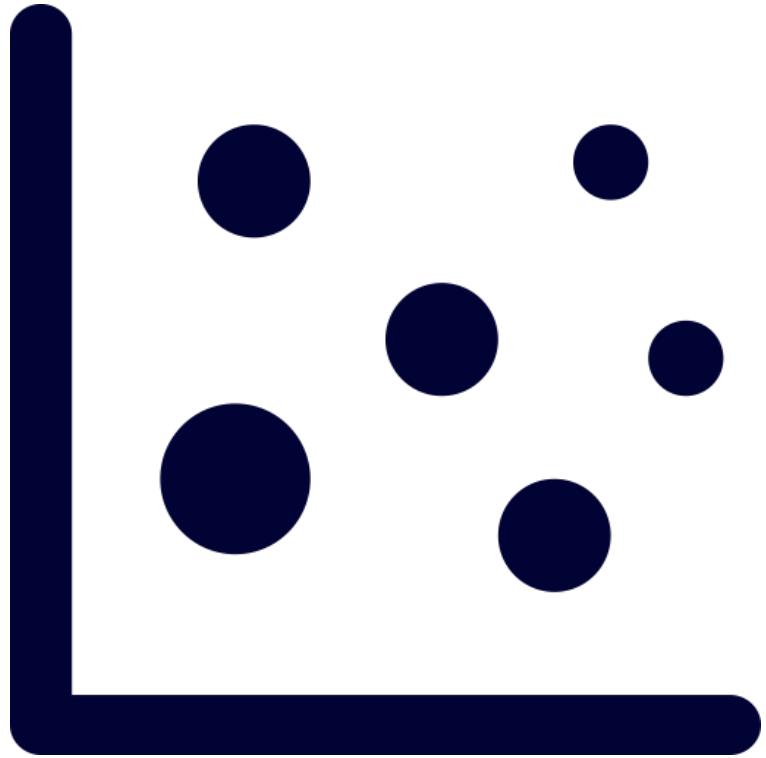
Purwadhika
Startup and Coding School

What is Scatter Plot?

Purwadhika
Startup and Coding School

What is Scatter Plot?

- A **scatter plot** is a type of chart that is often used in the fields of statistics and data science.
- It consists of multiple data points plotted across two axes. Each variable depicted in a scatter plot would have multiple observations.
- If a scatter plot includes more than two variables, then we would use different colours to signify that.
- Scatter plot helps in visualizing 2 numeric variables. It helps in identifying the relationship of the data with each variable, i.e. correlation or trend patterns.
- It also helps in detecting outliers in the plot.

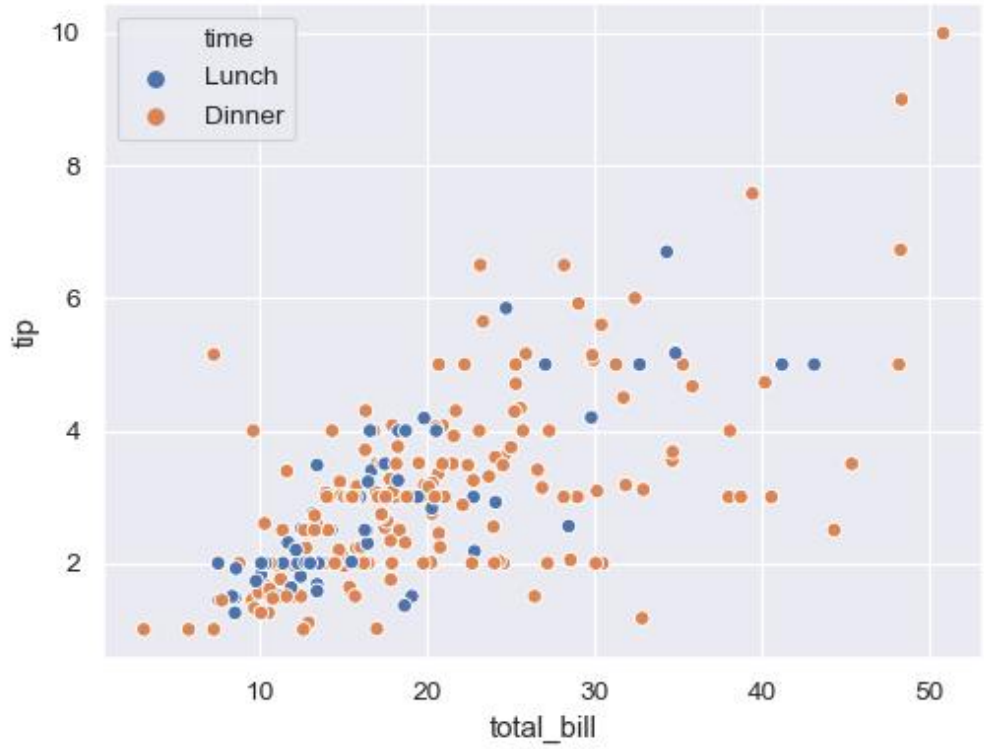


When to Use Scatter Plot?

Purwadhika
Startup and Coding School

When to Use Scatter Plot?

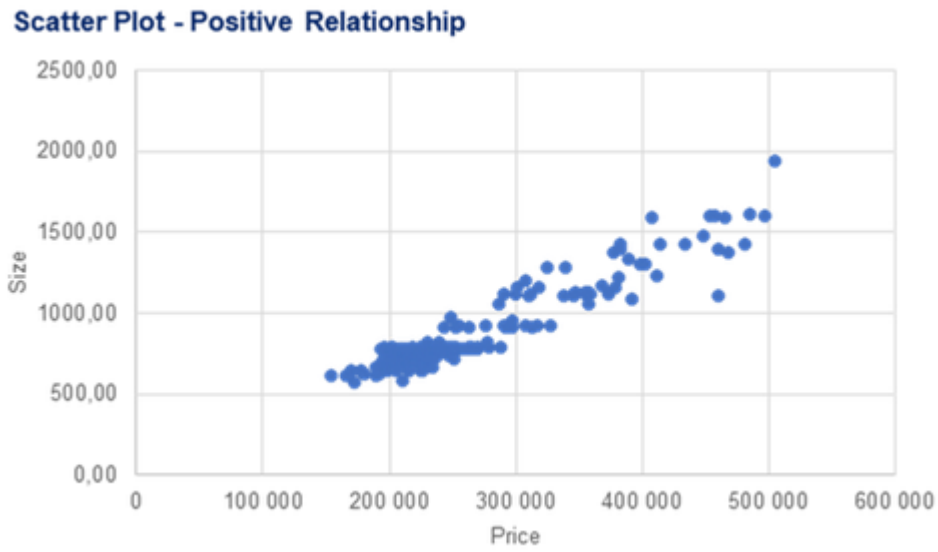
- **When to use**: A scatter plot chart is a great indicator that allows us to see whether there is a pattern to be found between two variables.
- It is used in Machine learning concepts like regression, where x and y are continuous variables.
- It is also used in clustering scatters or outlier detection.

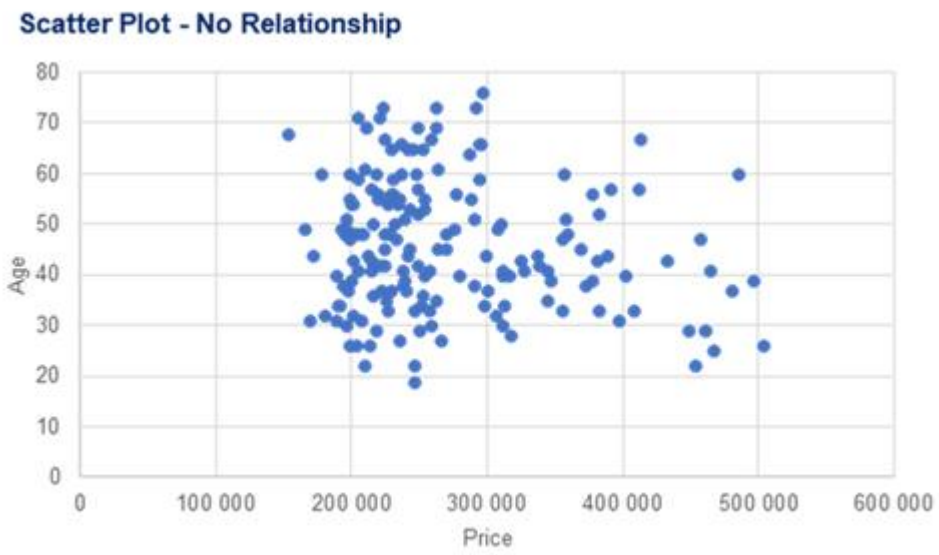


Purwadhika
Startup and Coding School

When to Use Scatter Plot?

This can be a very useful chart type whenever we would like to see if there is any **relationship** between two sets of data.





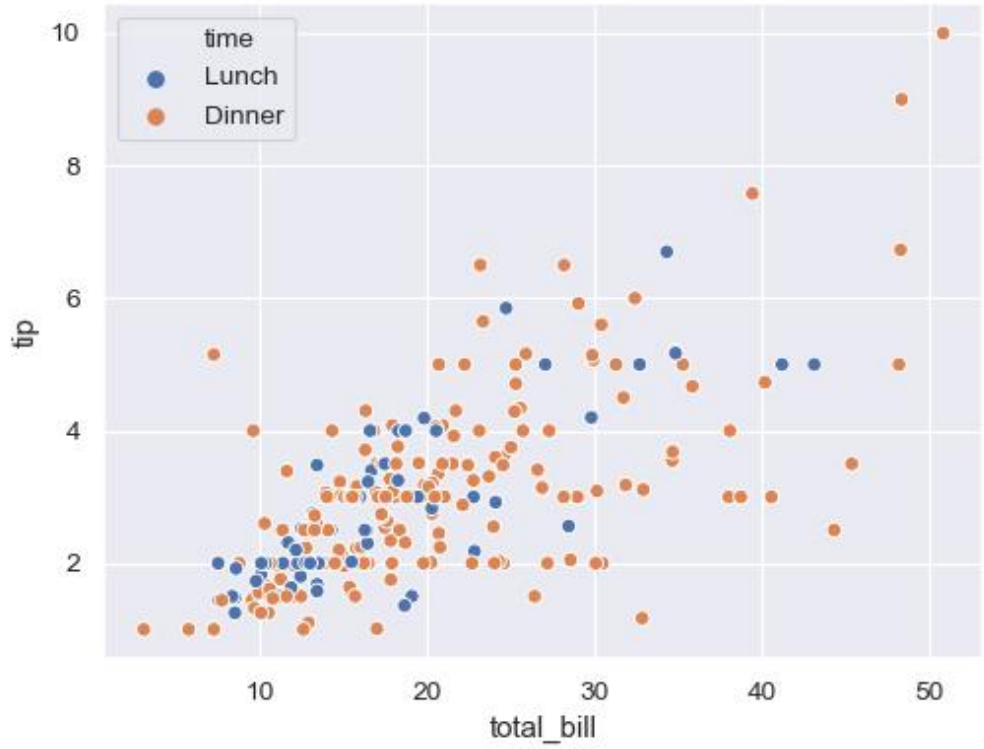
Purwadhika
Startup and Coding School

When to Avoid Scatter Plot?

Purwadhika
Startup and Coding School

When to Avoid Scatter Plot?

- We can’t use scatter plots **when we don’t have bi-dimensional data**.
- A scatter plot requires at least two dimensions for our data.
- In addition, scatter plots are **not suitable** if we are interested in observing **time patterns**.
- Finally, a scatter plot is used with numerical data or numbers. If we have categories such as 3 divisions, 5 products, and so on, a scatter plot would not reveal much.



Purwadhika
Startup and Coding School

Create Scatter Plot using Matplotlib

Purwadhika
Startup and Coding School

Create Scatter Plot using Matplotlib

Matplotlib is a comprehensive library for creating static, animated, and interactive visualizations in Python.

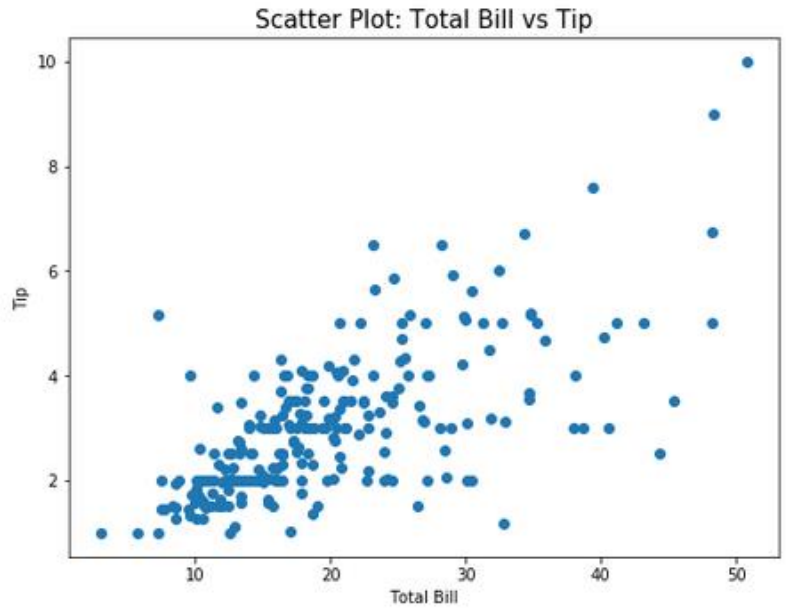
[2]:
```
# Import Matplotlib & Seaborn
import matplotlib.pyplot as plt
import seaborn as sns

# Import Tips Dataset from seaborn
tips = sns.load_dataset("tips")
tips.head(3)
```

[2]:

	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

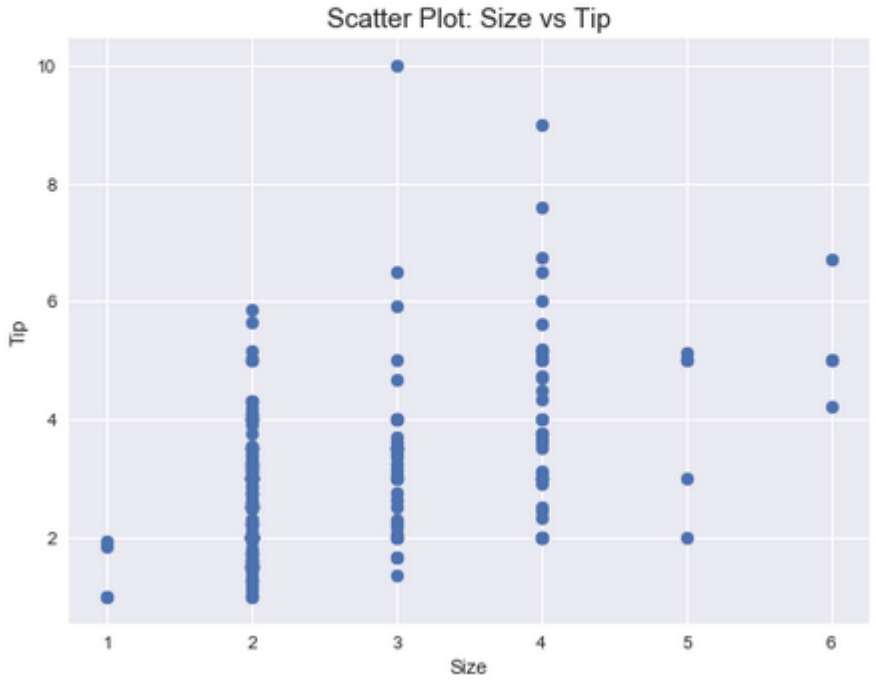
[3]:
```
plt.figure(figsize=(8,6))                                  # figure size
plt.scatter(tips['total_bill'], tips['tip'])               # create scatterplot in matplotlib
plt.title('Scatter Plot: Total Bill vs Tip', size=15)      # Title
plt.xlabel('Total Bill')                                   # X label
plt.ylabel('Tip')                                          # Y label
plt.show()
```



Purwadhika
Startup and Coding School

Create Scatter Plot using Matplotlib

[6]:
```
plt.style.use('seaborn')                          # change style
plt.figure(figsize=(8,6))                         # figure size
plt.scatter(tips['size'], tips['tip'])            # create scatterplot in matplotlib
plt.title('Scatter Plot: Size vs Tip', size=15) # Title
plt.xlabel('Size')                                # X label
plt.ylabel('Tip')                                 # Y label
plt.grid(True)                                    # add grid
plt.savefig('SizeTip_ScatterPlot.png')            # saving plot
plt.show()
```



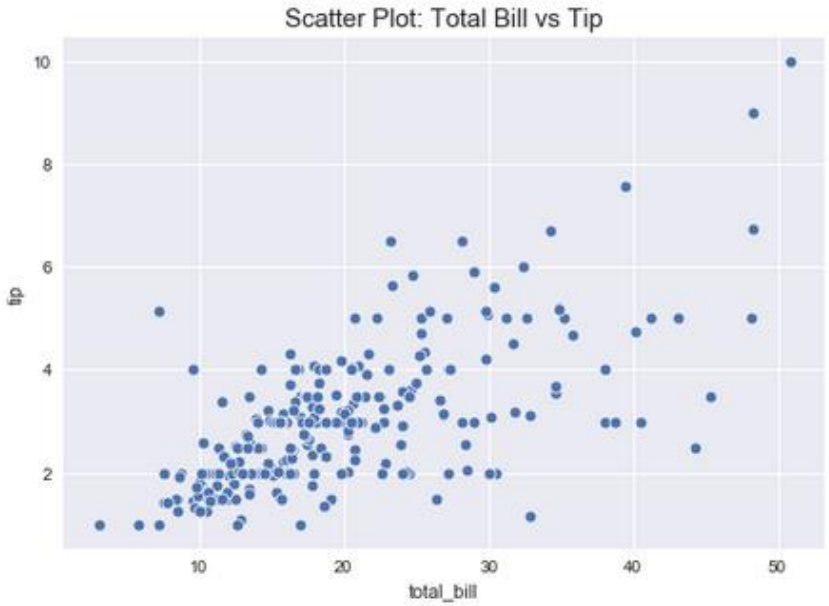
Purwadhika
Startup and Coding School

Create Scatter Plot using Seaborn

Purwadhika
Startup and Coding School

Create Scatter Plot using Seaborn

```
[7]: sns.scatterplot(data=tips, x="total_bill", y="tip")      # create scatter plot using seaborn
     plt.title('Scatter Plot: Total Bill vs Tip', size=15)   # add title
     plt.show()
```



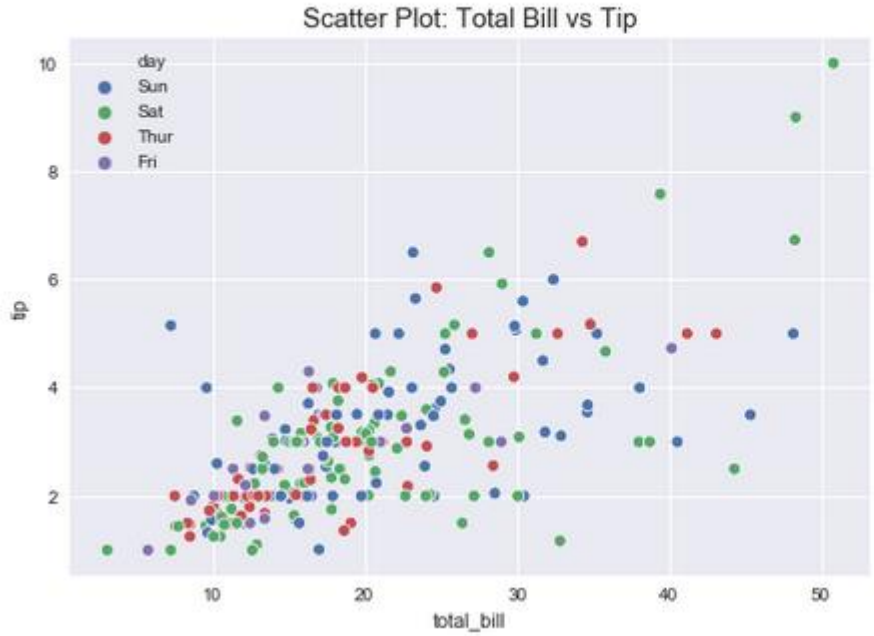
Seaborn is a Python data visualization library based on matplotlib.

It provides a high-level interface for drawing attractive and informative statistical graphics.

Purwadhika
Startup and Coding School

Create Scatter Plot using Seaborn

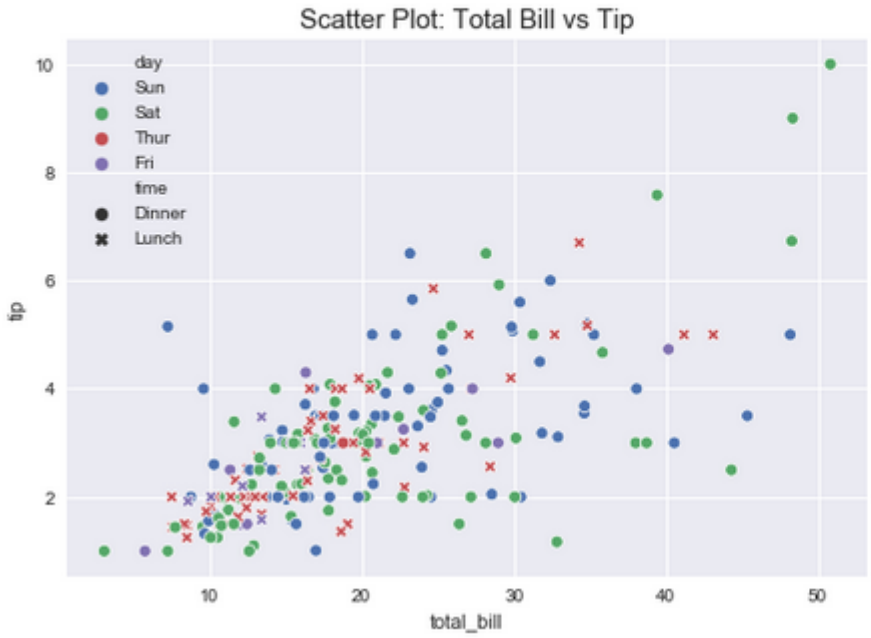
```
[8]: sns.scatterplot(data=tips, x="total_bill", y="tip", hue='day')  # create scatter plot using seaborn
     plt.title('Scatter Plot: Total Bill vs Tip', size=15)          # add title
     plt.show()
```



Purwadhika
Startup and Coding School

Create Scatter Plot using Seaborn

```
[9]: sns.scatterplot(data=tips, x="total_bill", y="tip", hue='day', style='time')  # create scatter plot using seaborn
     plt.title('Scatter Plot: Total Bill vs Tip', size=15)                         # add title
     plt.show()
```



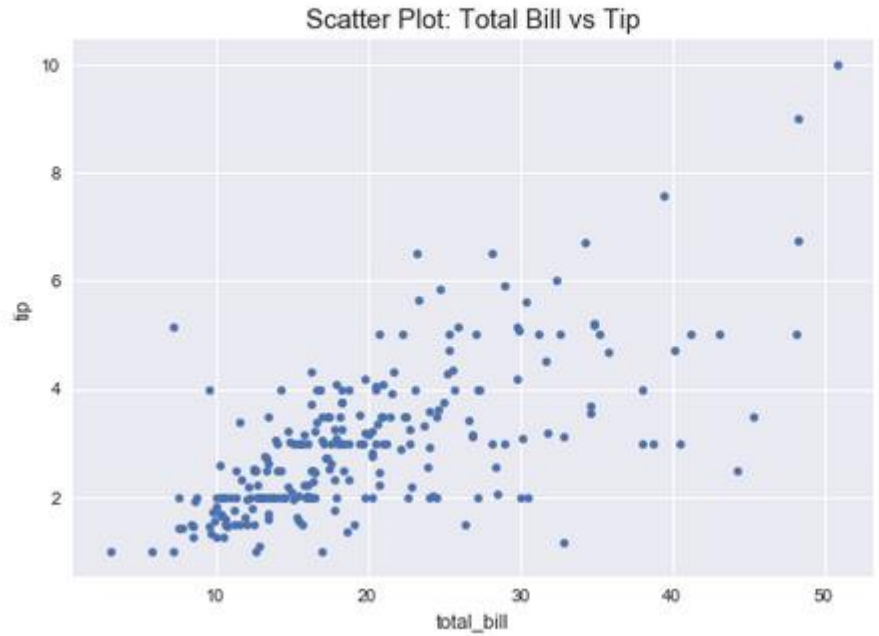
Purwadhika
Startup and Coding School

Create Scatter Plot using Pandas

Purwadhika
Startup and Coding School

Create Scatter Plot using Pandas

```
[10]: tips.plot.scatter(x='total_bill', y='tip')              # create scatter plo using pandas
      plt.title('Scatter Plot: Total Bill vs Tip', size=15)   # add title
      plt.show()
```



Pandas is a fast, powerful, flexible and easy to use open source data analysis and manipulation tool, built on top of the Python programming language.

Purwadhika
Startup and Coding School

Reference

- Veekesh Dhununjoy, "Advanced Visualization for Data Scientists with Matplotlib", https://medium.com/sfu-cspmp/advanced-visualization-for-data-scientists-with-matplotlib-15c28863c41c
- 365datascience, "Choosing the right chart: Selecting among 14 chart types", https://365datascience.com/chart-types-and-how-to-select-the-right-one/
- Badreesh Shetty, "Data Visualization using Matplotlib", https://towardsdatascience.com/data-visualization-using-matplotlib-16f1aae5ce70
- Pandas.Series.diff, https://pandas.pydata.org/docs/reference/api/pandas.Series.diff.html
- Amanda Iglesias Moreno, "Simple and multiple linear regression with Python", https://towardsdatascience.com/simple-and-multiple-linear-regression-with-python-c9ab422ec29c