SESSIONS 5

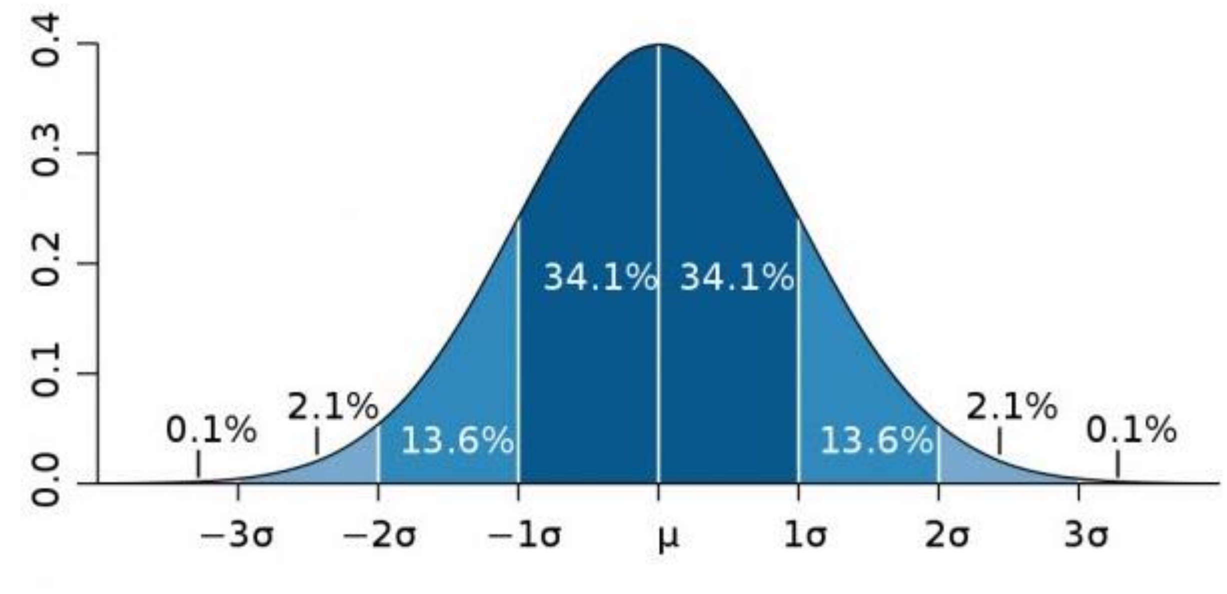
# Histogram

## Data Science Program

Purwadhika
Startup and Coding School

# Outline

- What is Histogram?
- Part of a Histogram
- Difference between a bar graph & a histogram
- Create Histogram using Matplotlib, Seaborn, and Pandas



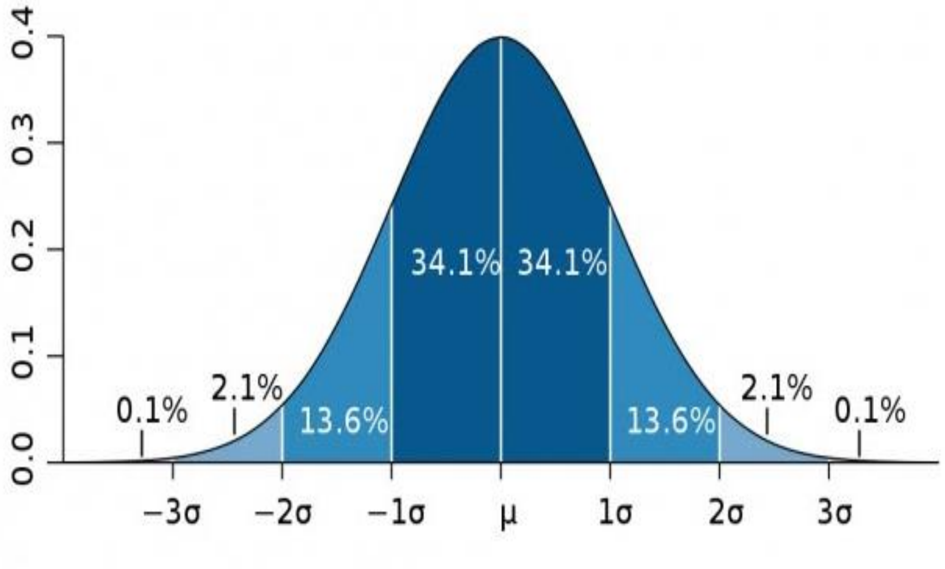
Purwadhika
Startup and Coding School

# What is Histogram?

Purwadhika
Startup and Coding School

# What is Histogram?

- A histogram is an accurate graphical representation of the distribution of numerical data.
- It is an estimate of the probability distribution of a continuous variable (quantitative variable) and was first introduced by Karl Pearson.
- It is a kind of bar graph.
- To construct a histogram, the first step is to “bin” the range of values. Bin has divided the entire range of values into a series of intervals. Then, count how many values fall into each interval.



Purwadhika
Startup and Coding School

# What is Histogram?

- The bins are usually specified as consecutive, non-overlapping intervals of a variable. The bins (intervals) must be adjacent and are often (but are not required to be) of equal size.
- Basically, histograms are used to represent data given in form of some groups.
- The X-axis is about bin ranges where Y-axis talks about frequency.
- So, if you want to represent an age-wise population in form of the graph then histogram suits well as it tells you how many exist in certain group range or bin.

# Parts of Histogram

Purwadhika
Startup and Coding School

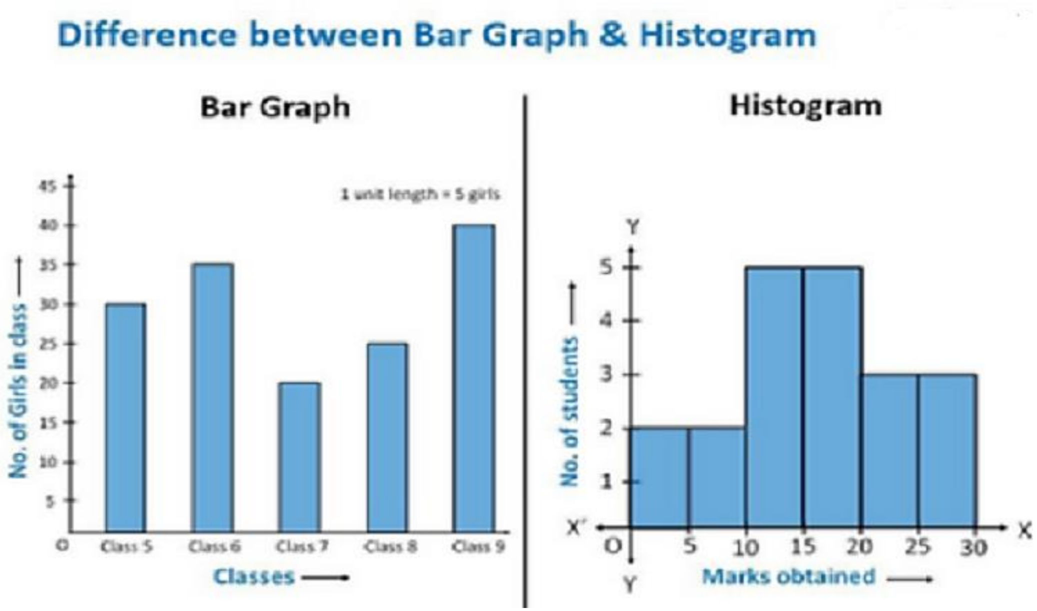
# Parts of a Histogram

- **The title**: The title describes the information included in the histogram.
- **X-axis**: The X-axis are intervals that show the scale of values which the measurements fall under.
- **Y-axis**: The Y-axis shows the number of times that the values occurred within the intervals set by the X-axis.
- **The bars**: The height of the bar shows the number of times that the values occurred within the interval, while the width of the bar shows the interval that is covered. For a histogram with equal bins, the width should be the same across all bars

# Difference between a Bar Graph & a Histogram

Purwadhika
Startup and Coding School

# Difference between a Bar Graph & a Histogram



- The major difference is that a histogram is only used to plot the frequency of score occurrences in a continuous data set that has been divided into classes, called bins.
- Bar charts, on the other hand, can be used for a lot of other types of variables, including ordinal and nominal data sets.

### In Bar Graph

- Bars have equal space
- On the y-axis, we have numbers & on the x-axis, we have data which can be anything.

### In Histogram

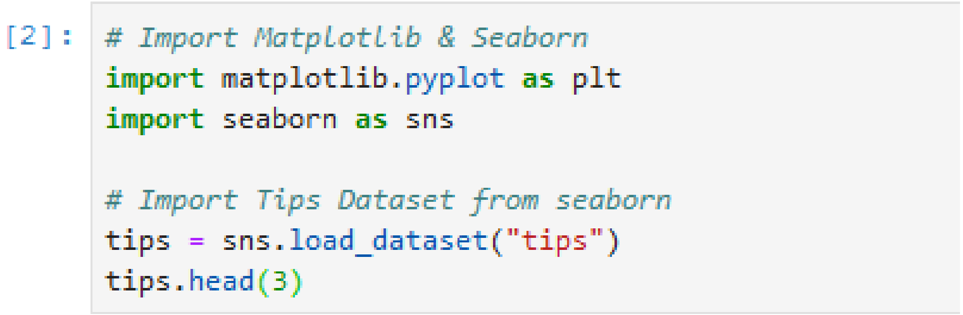
- Bars are fixed
- On the y-axis, we have numbers & on the x-axis, we have data which is continuous & will always be number

Purwadhika
Startup and Coding School

# Create Histogram using Matplotlib

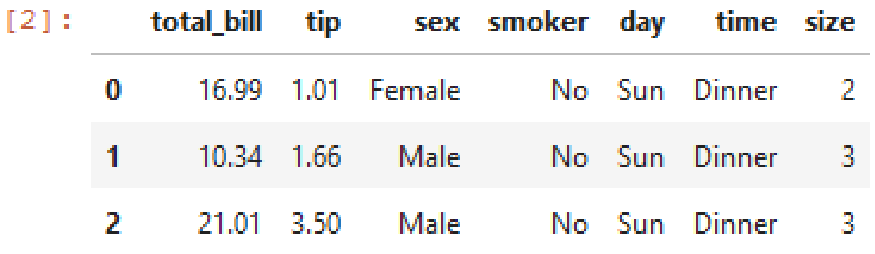
# Create Histogram using Matplotlib

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



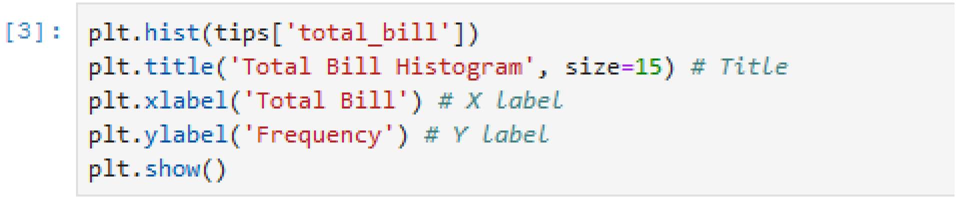
```
[2]: # Import Matplotlib & Seaborn
     import matplotlib.pyplot as plt
     import seaborn as sns

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

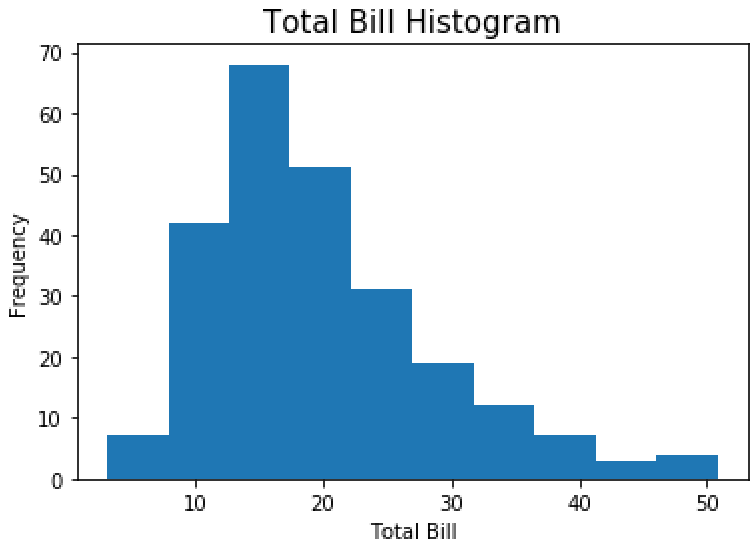


[2]:

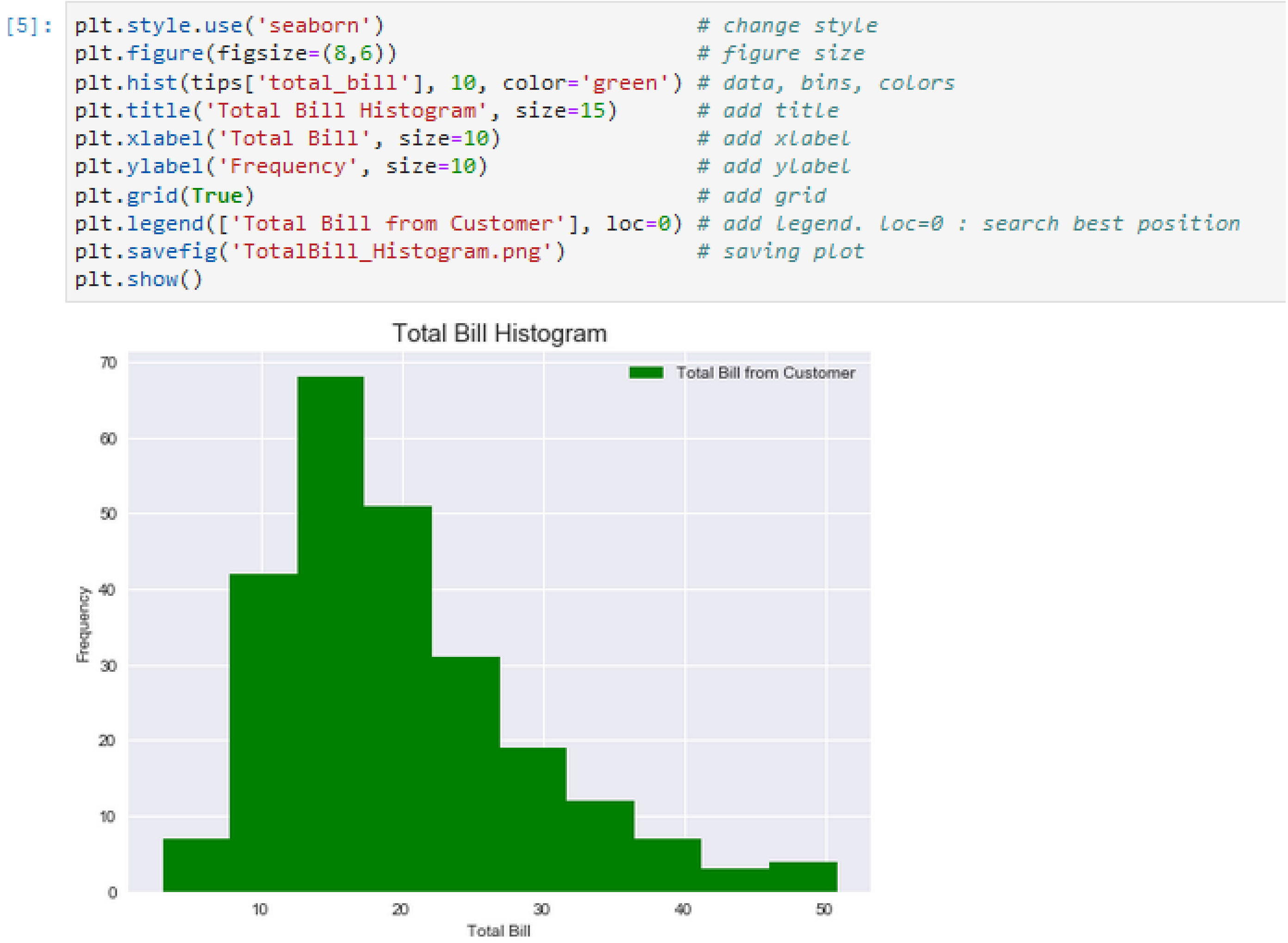
|   | total_bill | tip | sex | smoker | day | time | size |
|---|---|---|---|---|---|---|---|
| **0** | 16.99 | 1.01 | Female | No | Sun | Dinner | 2 |
| **1** | 10.34 | 1.66 | Male | No | Sun | Dinner | 3 |
| **2** | 21.01 | 3.50 | Male | No | Sun | Dinner | 3 |



```
[3]: plt.hist(tips['total_bill'])
     plt.title('Total Bill Histogram', size=15) # Title
     plt.xlabel('Total Bill') # X label
     plt.ylabel('Frequency') # Y label
     plt.show()
```



# Create Histogram using Matplotlib

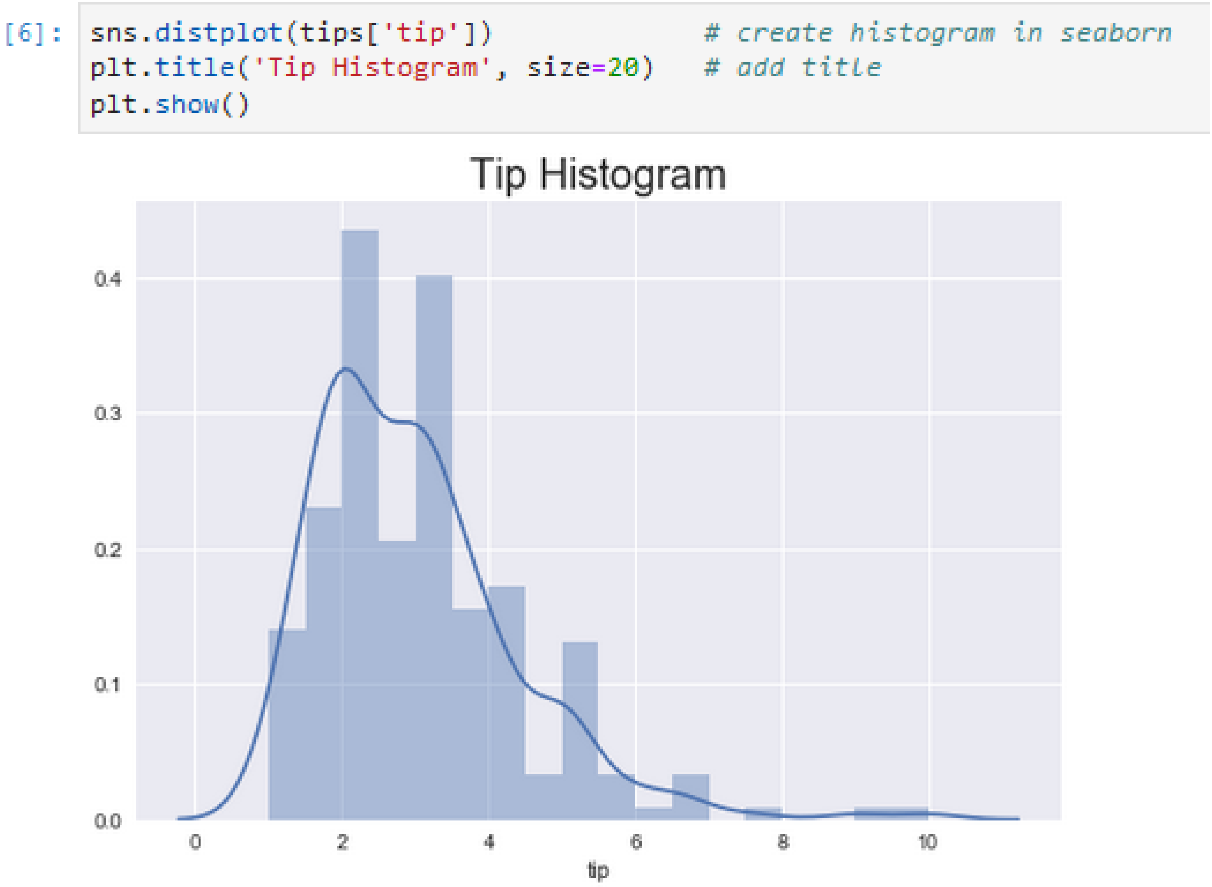


```
[5]: plt.style.use('seaborn')                            # change style
     plt.figure(figsize=(8,6))                           # figure size
     plt.hist(tips['total_bill'], 10, color='green') # data, bins, colors
     plt.title('Total Bill Histogram', size=15)       # add title
     plt.xlabel('Total Bill', size=10)                # add xlabel
     plt.ylabel('Frequency', size=10)                 # add ylabel
     plt.grid(True)                                   # add grid
     plt.legend(['Total Bill from Customer'], loc=0) # add legend. loc=0 : search best position
     plt.savefig('TotalBill_Histogram.png')           # saving plot
     plt.show()
```

# Create Histogram using Seaborn

Purwadhika
Startup and Coding School

# Create Histogram using Seaborn



```
[6]: sns.distplot(tips['tip'])          # create histogram in seaborn
     plt.title('Tip Histogram', size=20)   # add title
     plt.show()
```

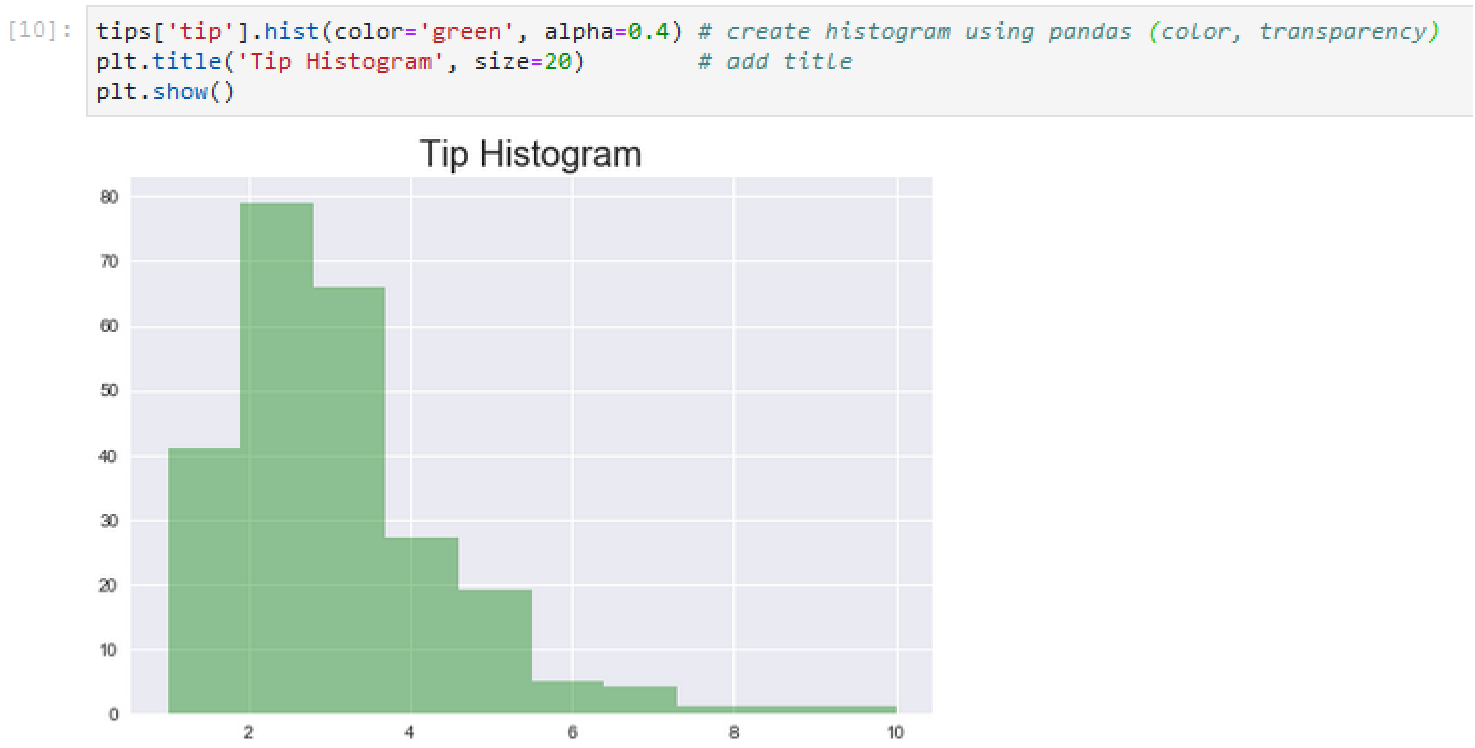
**Seaborn** is a Python data visualization library based on matplotlib.

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

# Create Histogram using Pandas

Purwadhika
Startup and Coding School

# Create Histogram using Pandas



```
[10]: tips['tip'].hist(color='green', alpha=0.4) # create histogram using pandas (color, transparency)
      plt.title('Tip Histogram', size=20)        # 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

- Seaborn, "Visualizing the distribution of a dataset", https://seaborn.pydata.org/tutorial/distributions.html
- David Gladson, "Matplotlib — Histograms Explained from Scratch | Python", https://medium.com/towards-artificial-intelligence/matplotlib-histograms-explained-from-scratch-python-6fe3e9d26de3
- cmdline, "How To Make Histogram in Python with Pandas and Seaborn?", https://cmdlinetips.com/2019/02/how-to-make-histogram-in-python-with-pandas-and-seaborn/
- Matplotlib Documentation, https://matplotlib.org/3.2.1/api/_as_gen/matplotlib.pyplot.hist.html
- CFI, "Histogram", https://corporatefinanceinstitute.com/resources/excel/study/histogram/
- AERD Statistics, ""Histograms", https://statistics.laerd.com/statistical-guides/understanding-histograms.php
- Jim Frost, "Using Histograms to Understand Your Data", https://statisticsbyjim.com/basics/histograms/