ABDELGHAFOR'S VIRTUAL INTERNSHIP

DATA ANALYSIS PROGRAM

SESSION (3)

PREPARED BY : MARK KOSTANTINE

Agenda Overview

WHAT IS SEABORN?

- Built on top of Matplotlib, Seaborn is a well-known Python library for data visualization that offers a user-friendly interface for producing visually appealing and informative statistical graphics. It is designed to work with Pandas dataframes, making it easy to visualize and explore data quickly and effectively.
- Seaborn offers a variety of powerful tools for visualizing data, including scatter plots, line plots, bar plots, heat maps, and many more. It also provides support for advanced statistical analysis, such as regression analysis, distribution plots, and categorical plots.
- Seaborn's key benefit lies in its capability to generate attractive plots with minimal coding efforts. It provides a range of default themes and color palettes, which you can easily customize to suit your preferences. Additionally, Seaborn offers a range of built-in statistical functions, allowing users to easily perform complex statistical analysis with their visualizations.
- Another notable feature of Seaborn is its ability to create complex multi-plot visualizations. With Seaborn, users can create grids of plots that allow for easy comparison between multiple variables or subsets of data. This makes it an ideal tool for exploratory data analysis and presentation.

SEABORN VS. MATPLOTLIB

- Python's two most widely used data visualization libraries are Matplotlib and Seaborn. While both libraries are designed to create high-quality graphics and visualizations, they have several key differences that make them better suited for different use cases.
- One of the main differences between Matplotlib and Seaborn is their focus. Matplotlib is a **low-level** plotting library that provides a wide range of tools for creating highly customizable visualizations. It is a highly flexible library, allowing users to create almost any type of plot they can imagine. This flexibility comes at the cost of a steeper learning curve and more verbose code.
- Seaborn, on the other hand, is a **high-level** interface for creating statistical graphics. It is built **on top of** Matplotlib and provides a simpler, more intuitive interface for creating common statistical plots. Seaborn is designed to work with Pandas dataframes, making it easy to create visualizations with minimal code. It also offers a range of built-in statistical functions, allowing users to easily perform complex statistical analyses with their visualizations.
- Another key difference between Matplotlib and Seaborn is their default styles and color palettes. Matplotlib provides a limited set of default styles and color palettes, requiring users to customize their plots manually to achieve a desired look. Seaborn, on the other hand, offers a range of default styles and color palettes that are optimized for different types of data and visualizations.

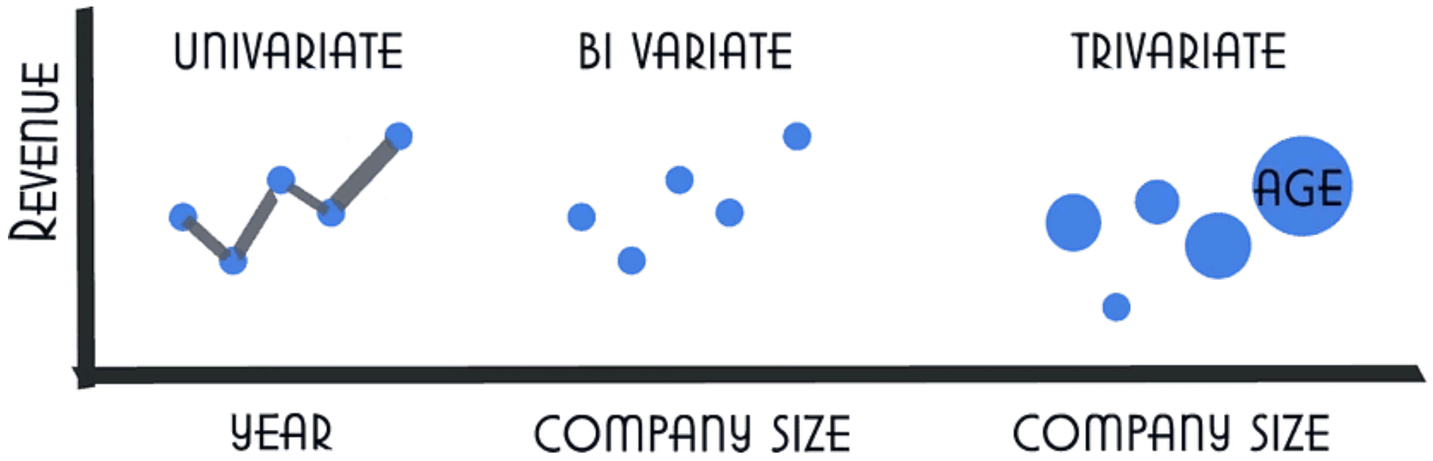
CATEGORIES OF PLOTS

- Plots are generally used to make visualization of the relationships between the given variables. These variables can either be a category like a group, division, or class or can be completely numerical variables. There are various different categories of plots that we can create using the seaborn library.
- In the seaborn library, the plot that we create is divided into the following various categories:
  - **Distribution plots:** This type of plot is used for examining both types of distributions, i.e., univariate and bivariate distribution.
  - **Relational plots:** This type of plot is used to understand the relation between the two given variables.
  - **Regression plots:** Regression plots in the seaborn library are primarily intended to add an additional visual guide that will help to emphasize dataset patterns during the analysis of exploratory data.
  - **Categorical plots:** The categorical plots are used to deals with categories of variables and how we can visualize them.
  - **Multi-plot grids:** The multi-plot grids are also a type of plot that is a useful approach is to draw multiple instances for the same plot with different subsets of a single dataset.
  - **Matrix plots:** The matrix plots are a type of arrays of the scatterplots.

SEABORN PLOT TYPES

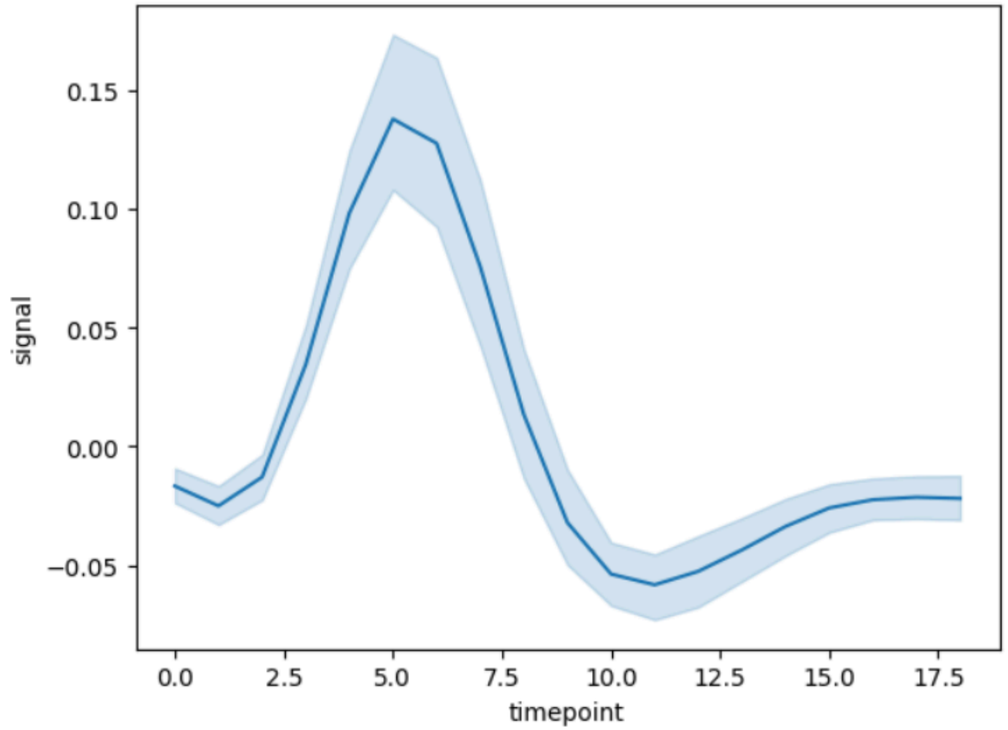
Seaborn provides a wide range of plot types that can be used for data visualization and exploratory data analysis. Broadly speaking, any visualization can fall into one of the three categories.

- **Univariate** – x only (contains only one axis of information)
- **Bivariate** – x and y (contains two axis of information)
- **Trivariate** – x, y, z (contains three axis of information)



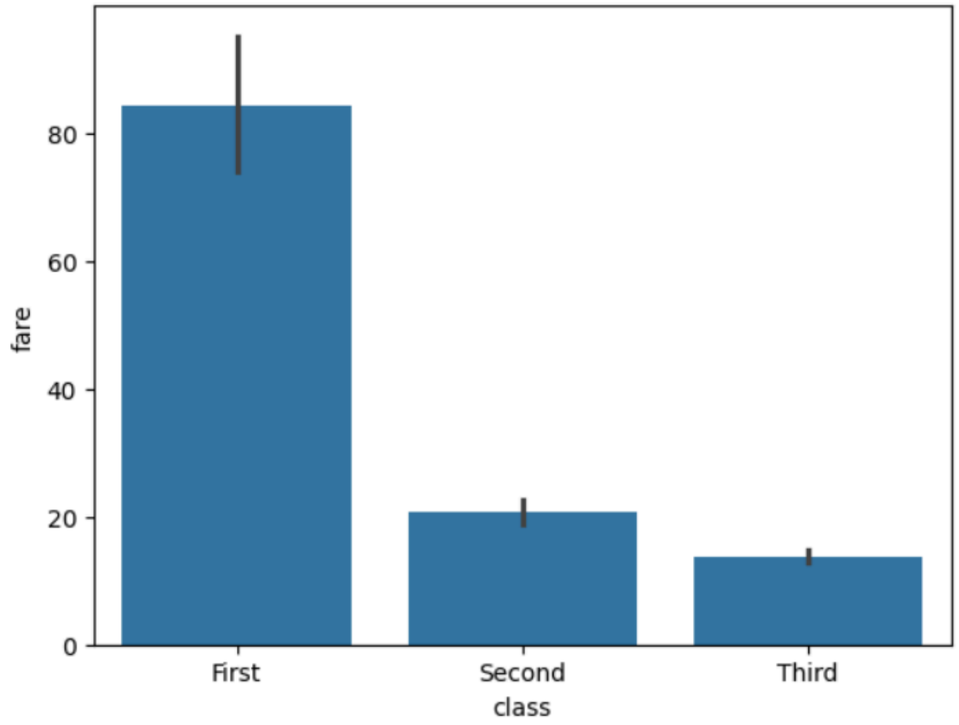
LINE PLOT

Line plots are used to visualize trends in data over time or other continuous variables. In a line plot, each data point is connected by a line, creating a smooth curve. In Seaborn, line plots can be created using the **lineplot()** function.



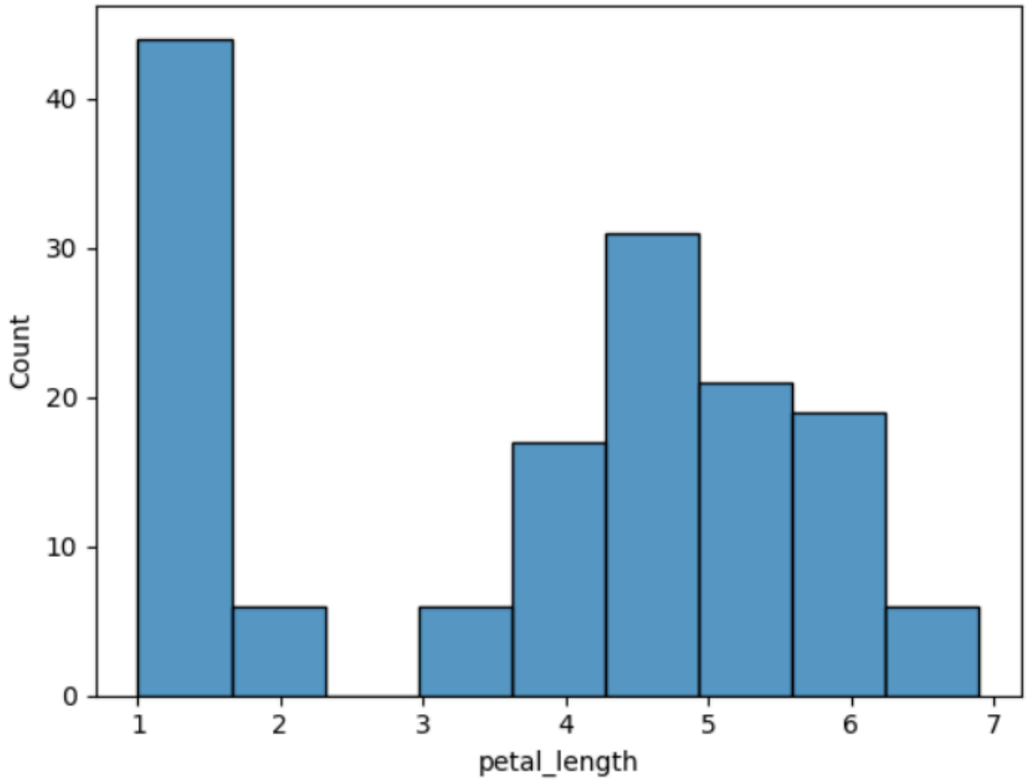
BAR PLOT

Bar plots are used to visualize the relationship between a categorical variable and a continuous variable. In a bar plot, each bar represents the mean or median (or any aggregation) of the continuous variable for each category. In Seaborn, bar plots can be created using the **barplot()** function.



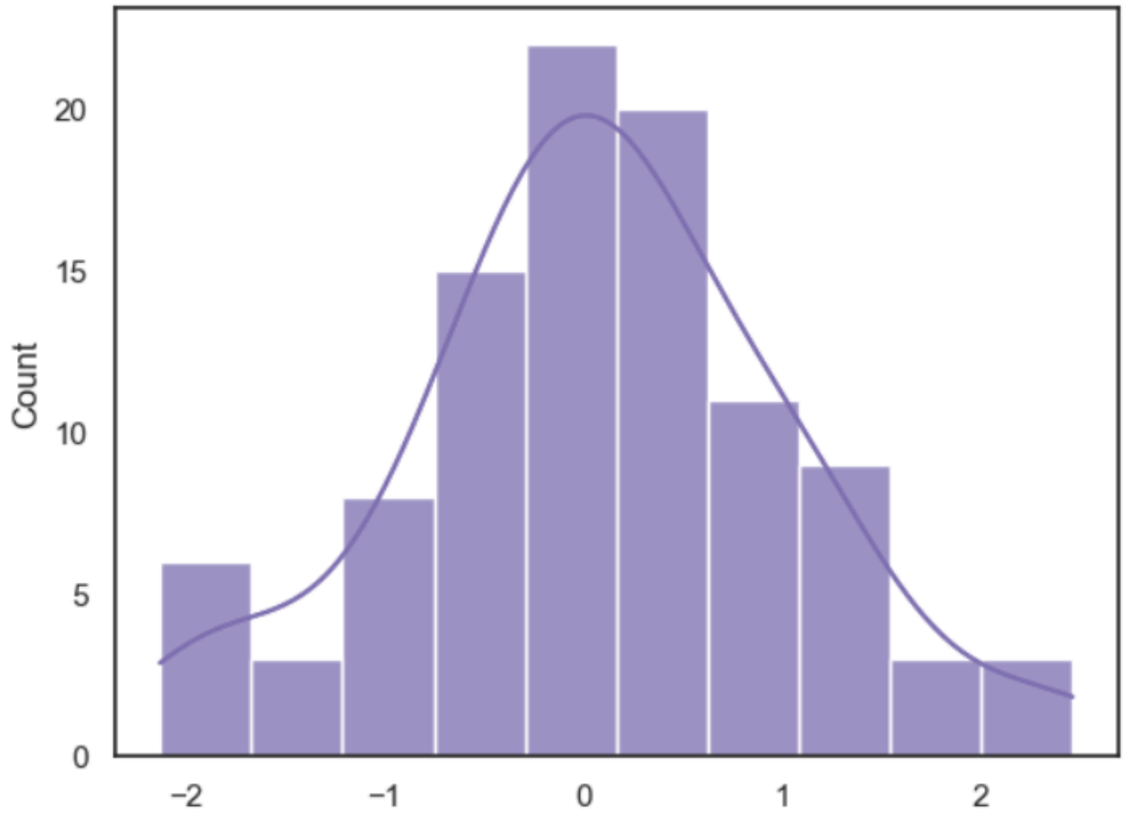
HISTOGRAMS

Histograms visualize the distribution of a continuous variable. In a histogram, the data is divided into bins and the height of each bin represents the frequency or count of data points within that bin. In Seaborn, histograms can be created using the **histplot()** function.



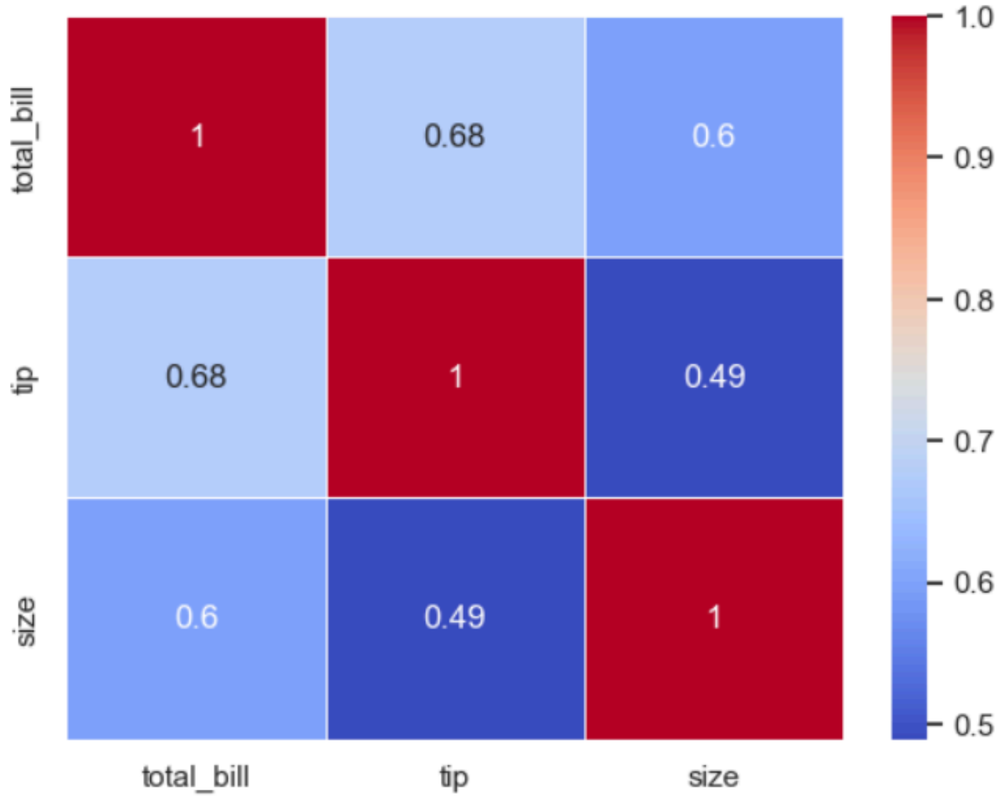
DIST PLOT

We use the seaborn dist plots to plot histograms with the given variables and data as a result. We can plot histograms with some other variations such as **rugplot and kdeplot** using a dist plot.



HEATMAPS

A heatmap is a graphical representation of data that uses colors to depict the value of a variable in a two-dimensional space. Heatmaps are commonly used to visualize the correlation between different variables in a dataset.



ANY QUESTIONS ?

DATA ANALYSIS PROGRAM

THANK YOU

UPCOMING NEXT WEEK : SESSION (4)