**Lab Session #7**

**Visualization using python**

**Aim:** To visualize data and series using matplotlib and seaborn packages in python.

**Problem Definition**: Develop python programs for the following:

1. Implement Taylor Series for 5 terms and visualize the same using matplotlilb.
2. Download the Penguin dataset from <https://www.kaggle.com/code/parulpandey/penguin-dataset-the-new-iris/notebook> and plot the following visualizations:
   1. Scatter plot to study relationship between flipper length and bill length.
   2. Histogram plot for bill depth.
   3. Bar plot for the body mass of the penguin species by sex.
   4. Box plot for distribution of flipper length by species.
   5. Pair plot for pairwise visualization of the attributes.
3. Infer any 3 observations from each of the plot.

**Theory:** Seaborn is a Python library for data visualization built on Matplotlib. Matplotlib is used to plot 2D and 3D graphs, while Seaborn is used to plot statistical graphs. Because Seaborn builds on Matplotlib, you can use these two libraries together to create very powerful visualizations.

You can install the Seaborn with the following command: **pip install seaborn**

When you install the Anaconda, Seaborn is installed automatically. After installing Seaborn, we need to import this library to use it. Let’s import Seaborn: **import seaborn as sns**

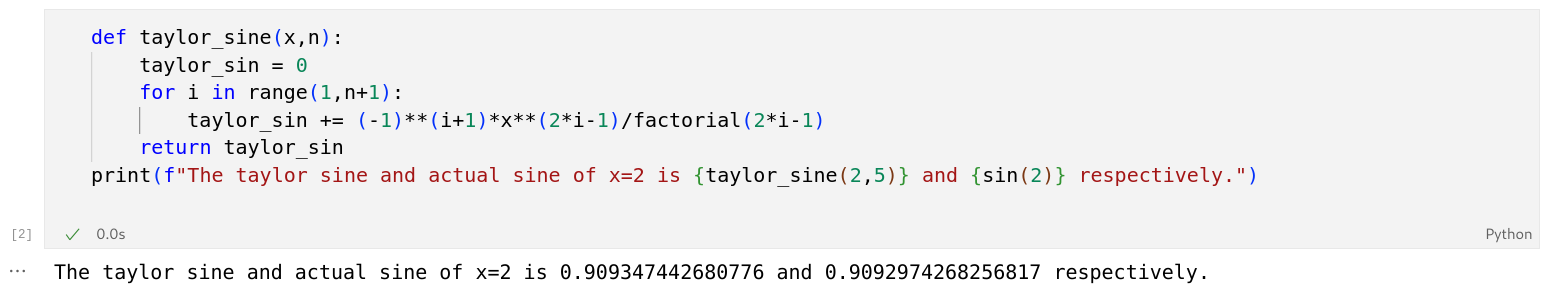
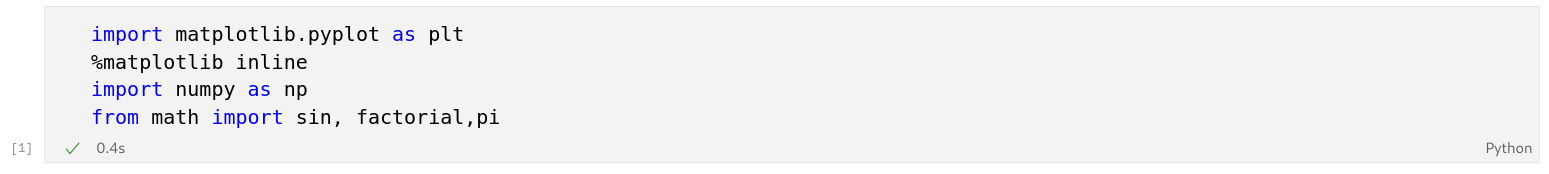
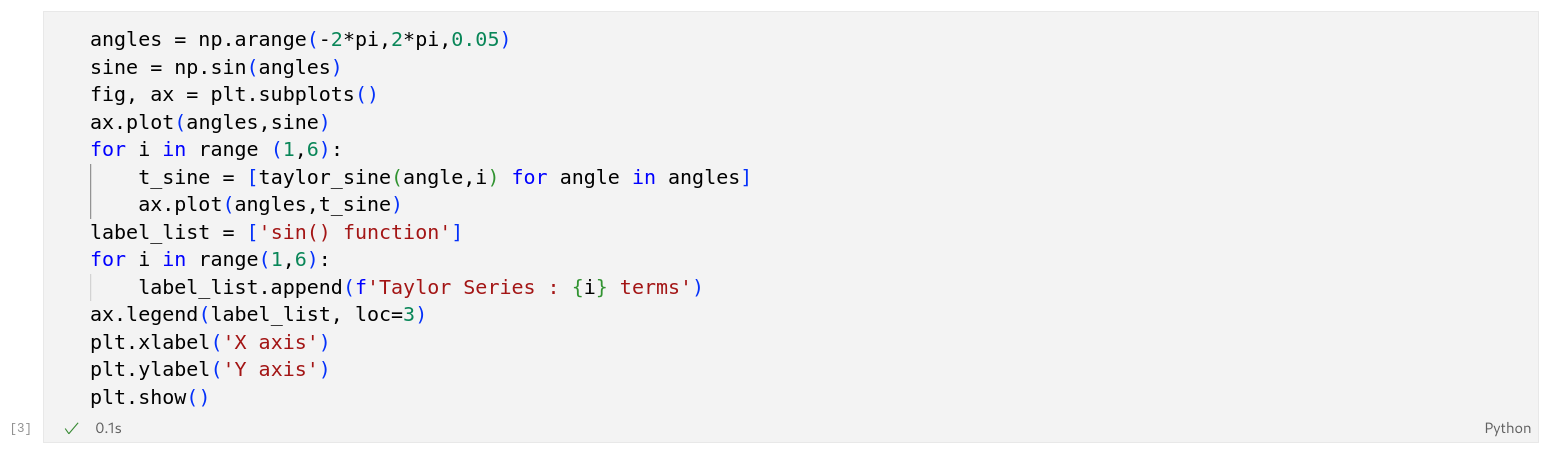
With Seaborn, you can easily load some famous datasets used for data science.

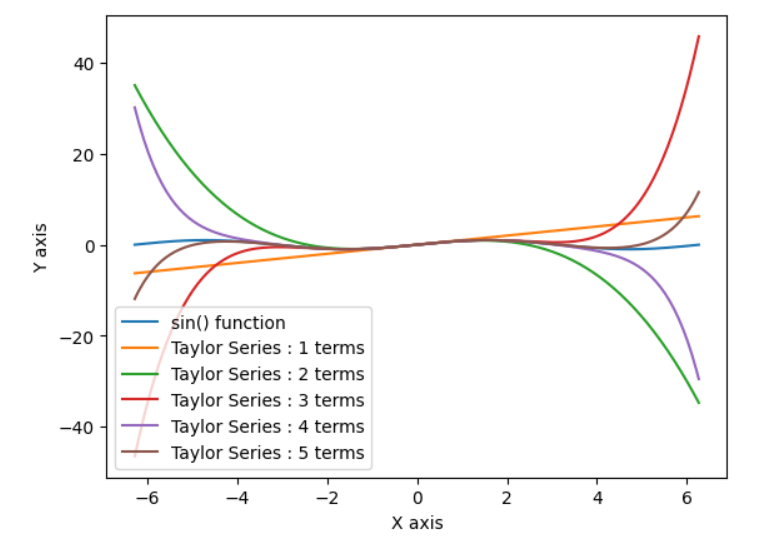
1. Scatter Plot: The best technique for understanding data is the scatter plot. The scatter plot is used to display the relationship between variables.
2. Histogram: The second type of plot I’m going to show is histogram. Histogram shows the distribution of the data. You can use the histogram plot to see the distribution of one or more variables.
3. Bar Plot: A bar plot represents an estimate of the central tendency for a numeric variable with the height of each rectangle.
4. Box Plot: The box plot is used to compare the distribution of numerical data between levels of a categorical variable.
5. Violin Plot: You can think of the violin plot as a box plot. This plot is used to compare the distribution of numerical values ​​among categorical variables.
6. Facet Grid: You can use a facet grid to see a grid graph of the different subsets in your dataset. For example, let me draw the histogram plot of the penguins’ flipper length according to the island and sex variables. Let’s assign column and row variables to add more subplots to the figure.
7. Pair Plot: Seeing the pair relationship between the variables in the dataset is one of the important steps of data analysis. You can use the pairplot method to see the pair relations of the variables. This function creates cross-plots of each numeric variable in the dataset.
8. Heatmap: Finally, let’s look at the heatmap. Heatmap is one of a very useful visualization techniques. You can use this technique to see correlations between numerical variables.

Matplotlib is an amazing visualization library in Python for 2D plots of arrays. Matplotlib is a multi-platform data visualization library built on NumPy arrays and designed to work with the broader SciPy stack. It was introduced by John Hunter in the year 2002. One of the greatest benefits of visualization is that it allows us visual access to huge amounts of data in easily digestible visuals. Matplotlib consists of several plots like line, bar, scatter, histogram etc. Installation : Windows, Linux and macOS distributions have matplotlib and most of its dependencies as wheel packages.

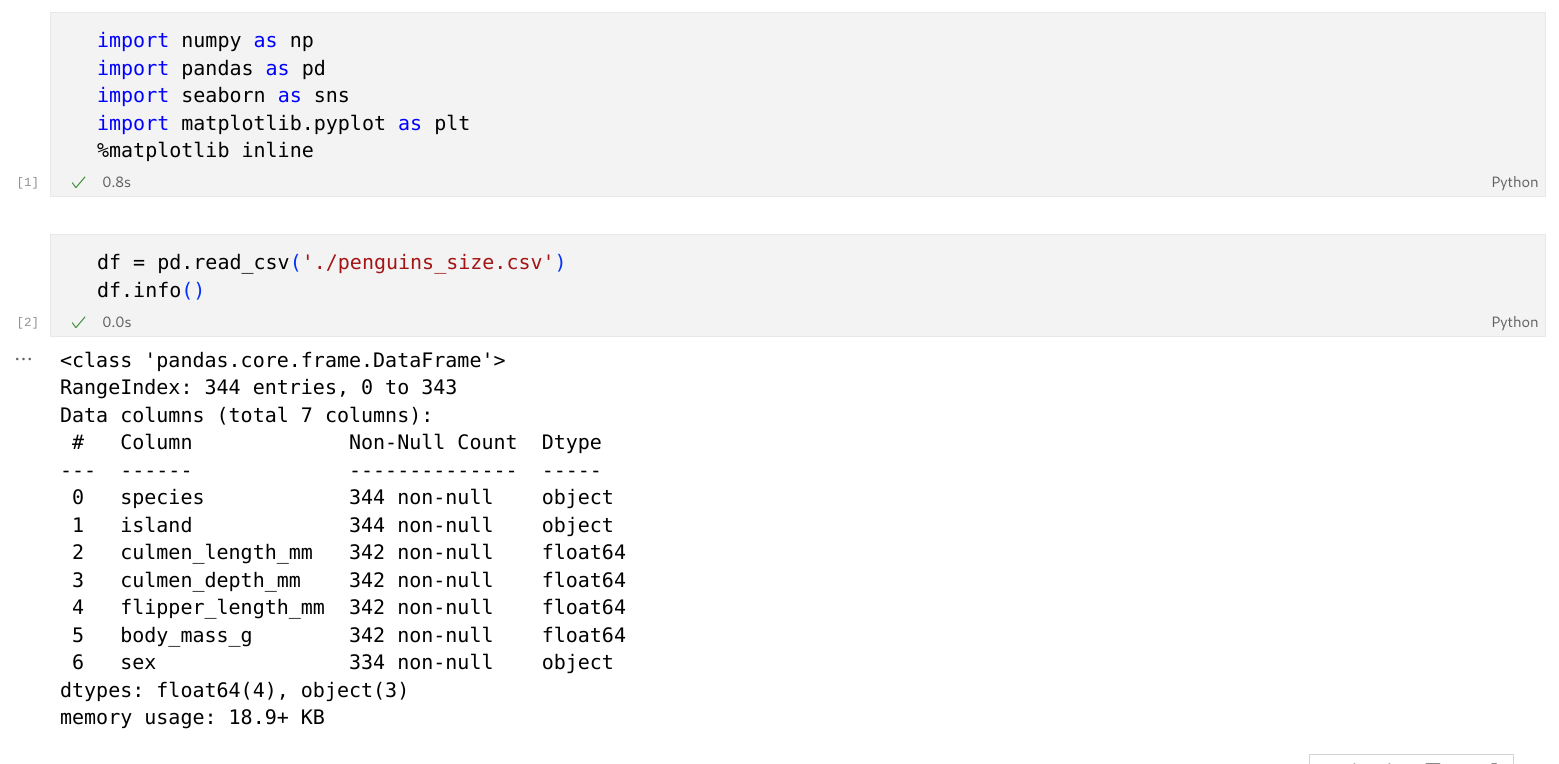
**Notebook:**

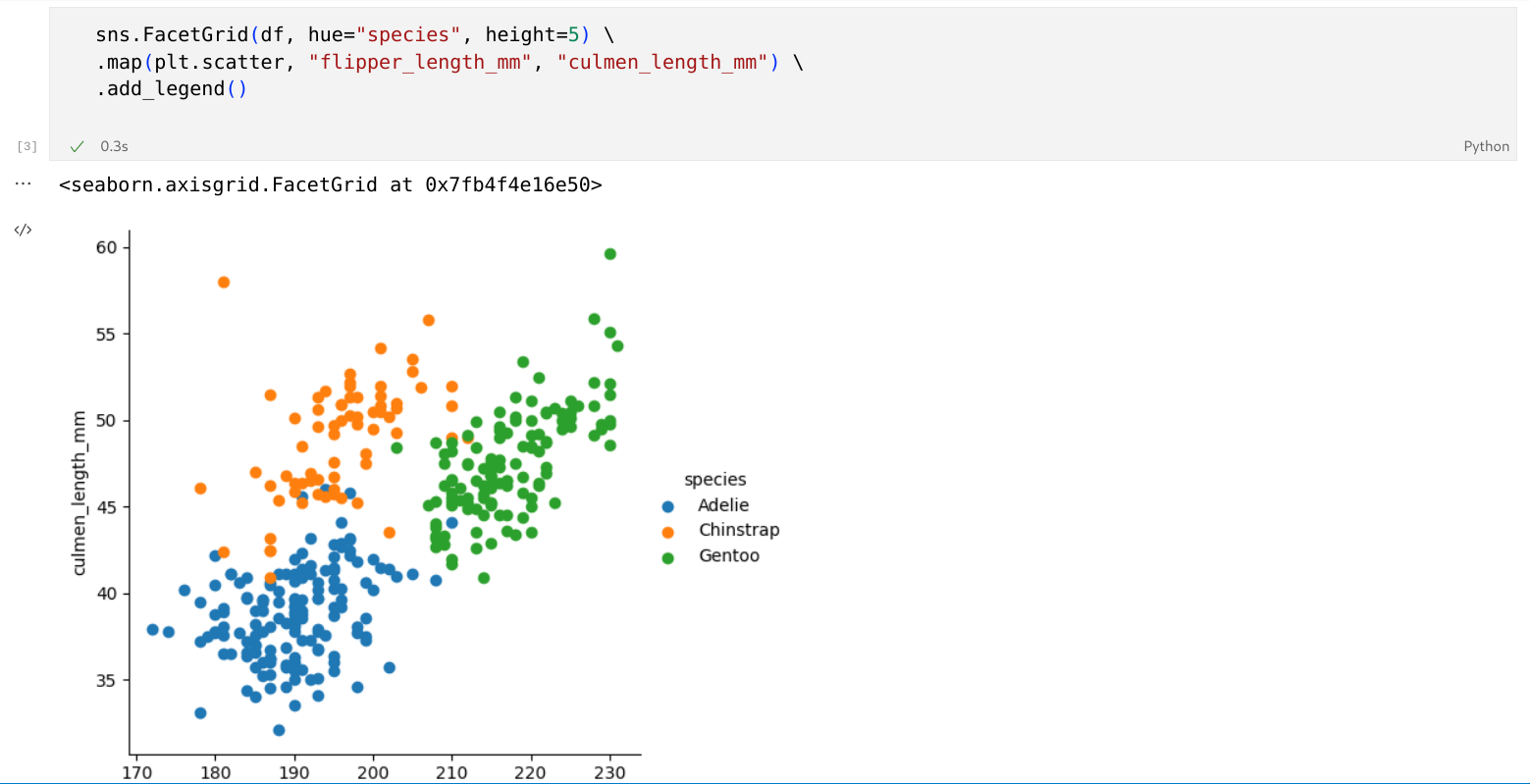
**1)**

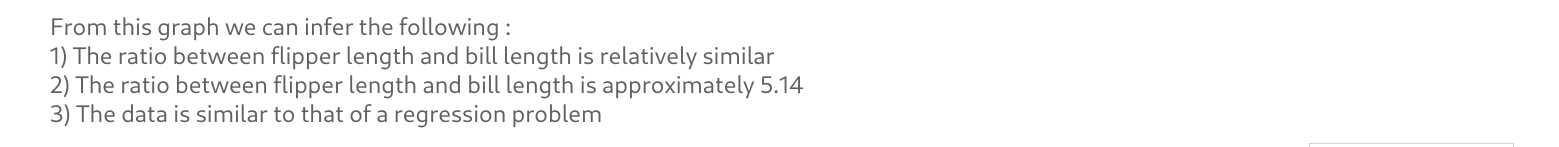
 

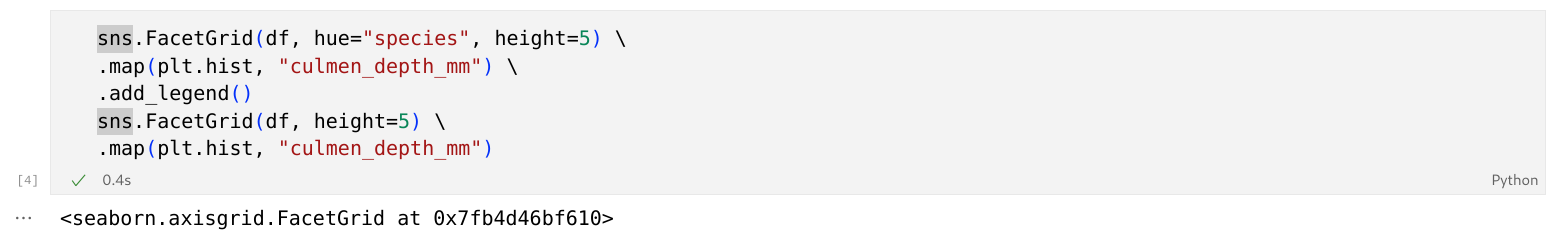


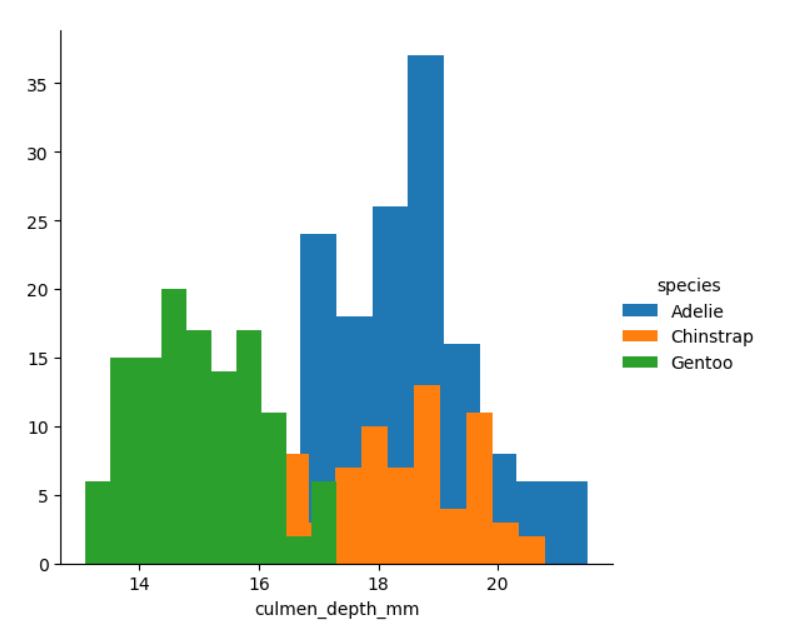
**2)**

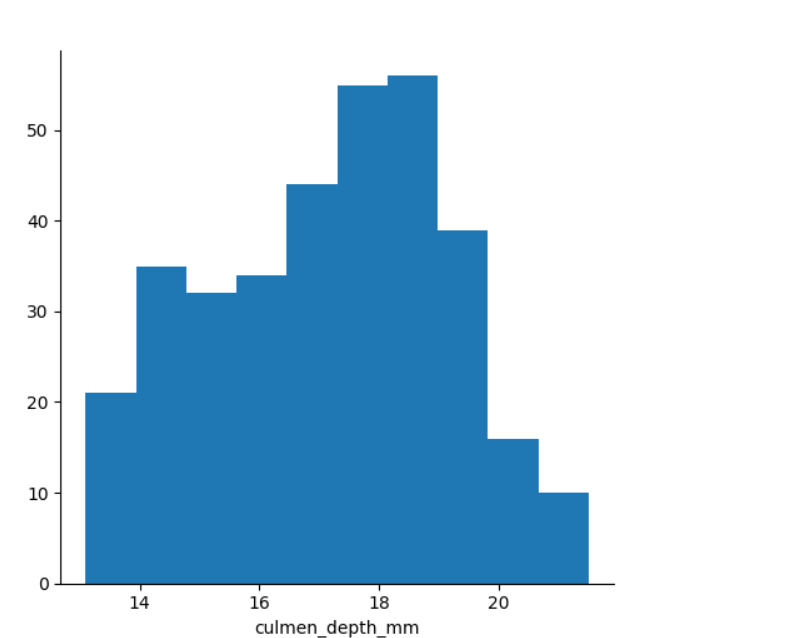
****





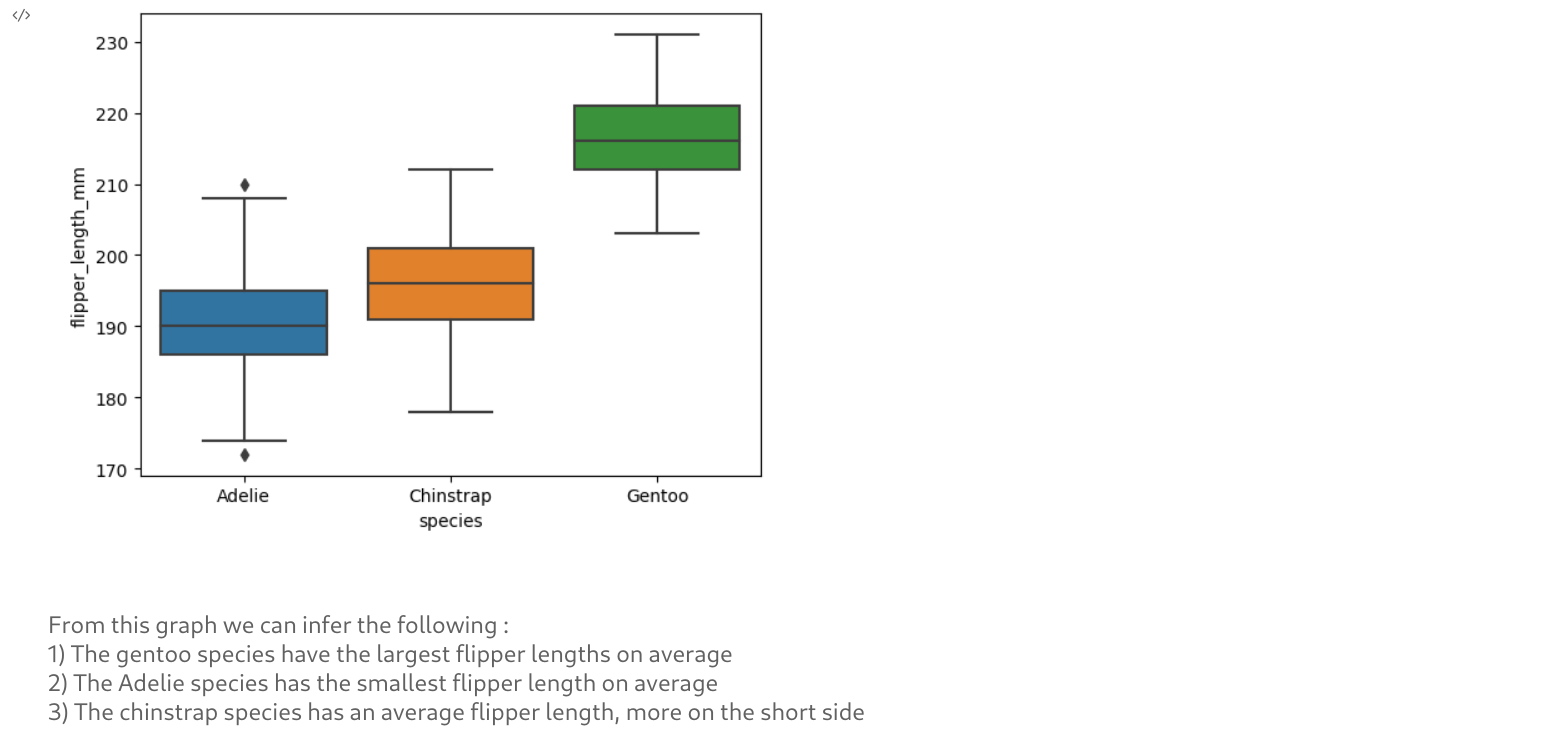




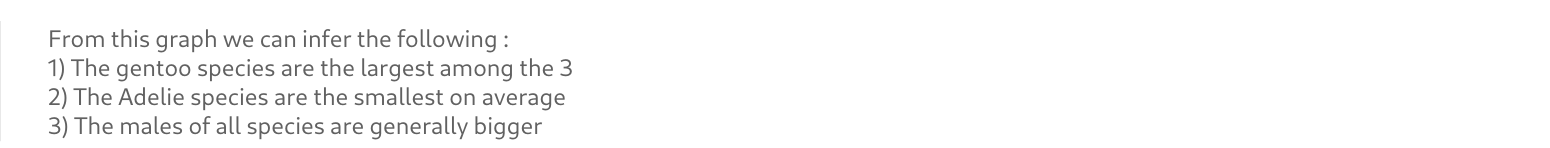
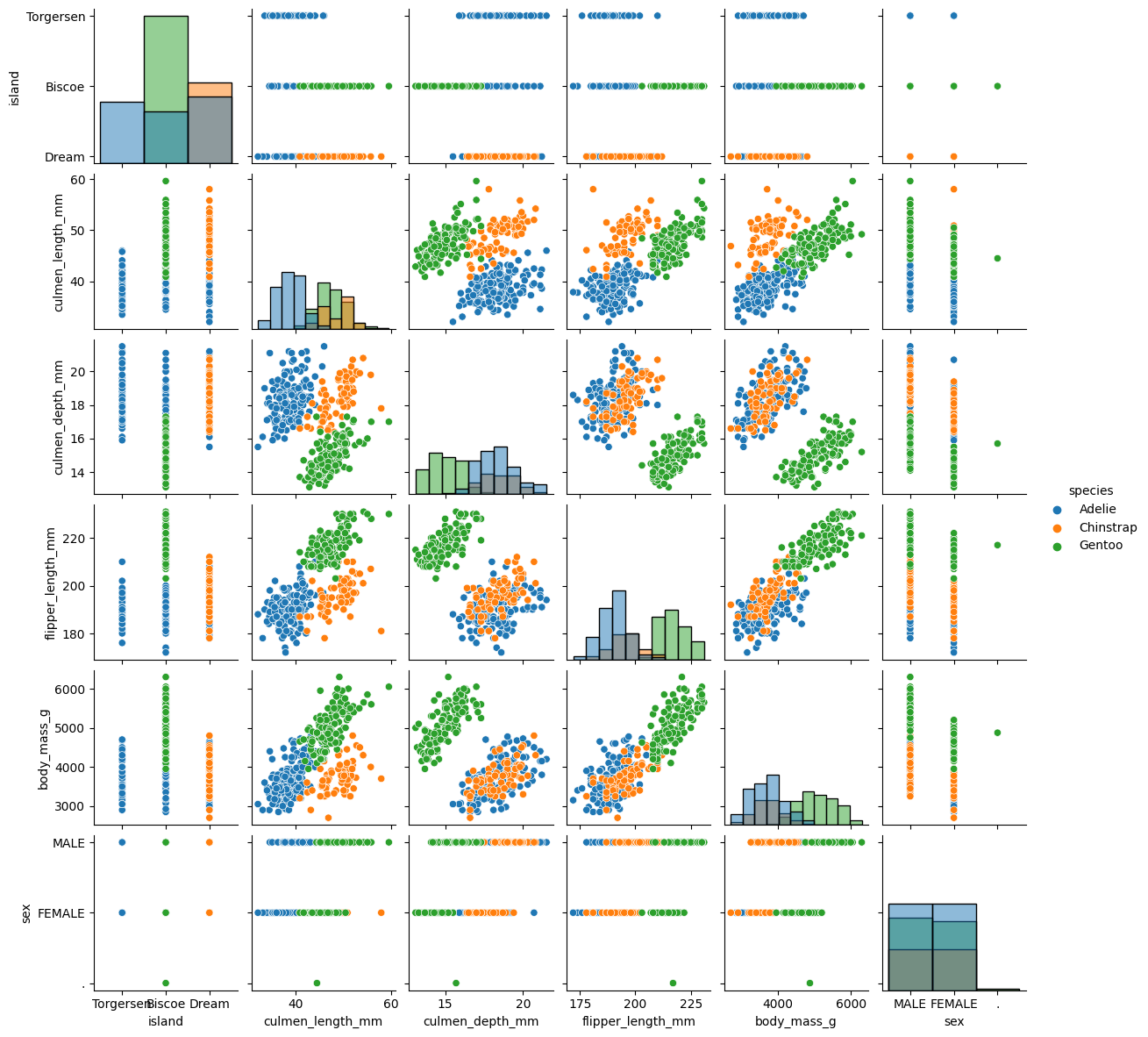






****



****