1. Basics

   Create two arrays: x, y and z where:

   - x contains 12 evenly spaced numbers between 1 and 20
   - y contains the values of x multiplied by 2 ie. 2*x
   - z contains the squares of elements of x ie x**2

   a. Plot y vs x, set the label of x-axis as 'x', the label of y-axis as 'y' and the title of the plot as 'y = 2x'. Increase the x axis range to (0, 25) and y axis range to (0, 450)

   b. Create a new plot and plot y vs x and z vs x in the same plot. Differentiate between the two functions using legend.

   c. Create a figure containing two subplots. Set the size of the figure to be 12X6. In the first subplot, plot y vs x and in the second subplot, plot z vs y. Set the title and axis labels accordingly.

   d. Finally, redo all the above plots using **seaborn** which is a higher level plotting library. It creates more aesthetically pleasing charts than matplotlib.

2. Applications

   a. Create 3 arrays:

   ```
   x1 = [5,7,8,7,2,17,2,9,4,11,12,9,6]

   x2 = [99,86,87,88,111,86,103,87,94,78,77,85,86]

   y = [0, 1, 1, 0, 0, 1, 0, 0, 1, 0, 0, 1, 1]
   ```

   Plot the scatterplot of x1 and x2. Color the points in the scatterplot according to their values in y: blue for 0 and red for 1. For example, color the point (5, 99) blue as the corresponding y value is 0 and the point (7, 86) red as the corresponding y value is 1.

   b. Download the dataset (PIMA Diabetes Dataset)

   - Plot a bar plot of the 'outcome' variable. Is the dataset balanced or unbalanced? Try to explore the consequences of having such a dataset.

   - Create a figure containing 8 subplots arranged in (2, 4) grid. Plot the histograms of the 8 features (all the columns except the outcome) present in the dataset. Identify the features having normal distribution and skewed distributions.

- Create a figure containing 8 subplots arranged in (4, 2) grid. Plot the boxplot for each of the features in each of the subplots. Outliers are a set of observations that vary significantly from the other observations. Which feature do you think has the most outliers? And Why?

- A correlation matrix represents how much the features are correlated to each other. You can compute the correlation matrix of a dataframe using the dataframe.corr() method. Use matplotlib.pyplot.matshow to plot the correlation matrix in matplotlib. Also, you can use the matplotlib.pyplot.colorbar() to better explain your correlation matrix.