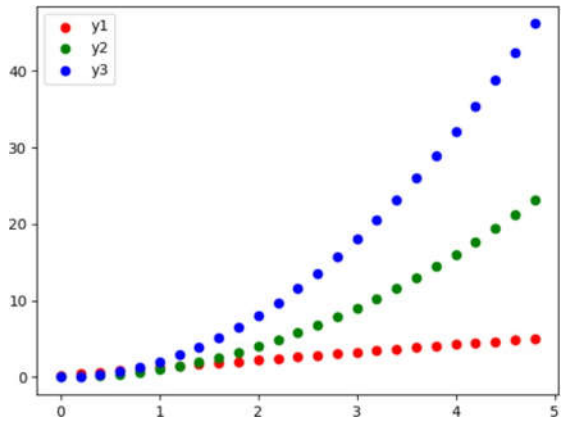
Question 1 - Basic Plotting/Visualization

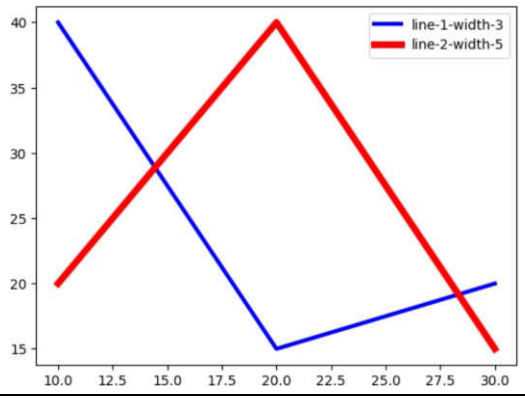
(a) Write a program with Matplotlib to visualize scatter plot below.

Input Data	Output

import numpy as np
 	
# Values of coordinate x 	
x = np.arange(0, 5, 0.2)
 	
# Values of coordinate y1 (linear line) 	
linearFunction = np.vectorize(lambda x : x + 0.25)
y1 = linearFunction(x)
 	
# Values of coordinate y2 (parabolic 1) 	
parabolicFunction1 = np.vectorize(lambda x : x ** 2)
y2 = parabolicFunction1(x)
 	
# Values of coordinate y3 (parabolic 2) 	
parabolicFunction2 = np.vectorize(lambda x : 2 * (x ** 2))
y3 = parabolicFunction2(x)
	

(b) Write a program with matplotlib to visualize line plot below.

Input Data	Output

import numpy as np
 	
# Coordinate (x,y) for line 1 (linewidth : 3) 	
x1 = [10, 20, 30]
y1 = [40, 15, 20]
 	
# Coordinate (x,y) for line 2 (linewidth : 5) 	
x2 = [10, 20, 30]
y2 = [20, 40, 15]
	

Question 2 – Visualization of Senior High School's Dataset in Province West Java

This assignment will use dataset from education ministry in Province West Java Indonesia referring to senior high school data, taken from : https://opendata.jabarprov.go.id/id/dataset?organization=dinas-pendidikan&q=murid&suggestion=on. From that link, we only choose three datasets to be

worked on for data visualization, including amount of schools, students, and teachers for each regencies/districts in West Java province. Here are the things you have to do :

(a) Read all of the CSV files from following three datasets and display the top 5 rows each.
- Number of senior high school unit in West Java : **number_of_senior_high_school_in_west_java_province.csv**
- Number of senior high school's student in West Java : **number_of_student_in_senior_high_school_in_west_java_province.csv**
- Number of senior high school's teacher in West Java : **number_of_teacher_in_senior_high_school_in_west_java_province.csv**

(b) Please do the following operations on the three datasets above :
- Join the three datasets by **merge**'s pandas operation based on three same columns as pivot : **nama_provinsi**, **kode_kabupaten_kota**, **tahun_ajaran**. **Hint** : please use code snippet below and fill in the blank part.

```
merge_result_1_df = pd.merge(……, ……, on=[……], suffixes=(‘’,
                        ‘_remove’), how=”inner”)
final_high_school_df = pd.merge(……, ……, on=[……], suffixes=(‘’,
                        ‘_remove’), how=”inner”)
```

- Display the 5 top rows of final dataframe.
- In process of merging three datasets, the three columns used as pivot before are duplicated in final dataframe (column name with suffix "_remove"). Remove them from final dataframe and display again 5 top rows after operation done. **Hint** : you can use directly code snippet below or any other suitable method.

```
final_high_school_df.drop([i for i in final_high_school_df.columns if
                        'remove' in i], axis=1, inplace=True)
```

- Display the data types of every column in final dataframe.
- Check if there are some missing values in final dataframe. If any of the columns contains missing value, fill in missing values with zero (for integer) and/or most frequent values (for categorical/non integer).

(c) Visualize 10 regency/district with **lowest number of teacher** in year "2020/2021" with any diagram/chart you think is appropriate

(d) Visualize 10 regency/district with **lowest number of teachers and students ratio** in year "2020/2021" with any diagram/chart you think is appropriate

(e) Visualize 10 regencies/districts with **highest percentage of number unit of schools** in year "2020/2021" in **pie chart** or any diagram/chart you think is appropriate

(f) Visualize 10 regency/district with **highest number of teachers and schools unit ratio** in year "2020/2021" with any diagram/chart you think is appropriate

(g) Visualize 10 regency/district with **highest number of students and schools unit ratio** in year "2020/2021" with any diagram/chart you think is appropriate

Question 3 - Plotting with Seaborn

You are strictly only allowed to use seaborn library for the visualization on this number (you are allowed subplots using plt / matplotlib.pyplot, but the chart have to use seaborn)

a. Read the Pokemon.csv data and put it to a dataframe
b. Preprocess the data if you think it's necessary. If not, explain your answer
    a. *Special note for Pokemon Types: each pokemon can either have 1, or 2 types, you can choose whatever technique you want to preprocess this
c. Create a pie chart representing the number of legendaries vs non-legendary pokemons.
d. Visualize the number of pokemon and their types (type 1 & type 2) in a bar plot
    a. 1 bar plot for type 1
    b. 1 bar plot for type 2
e. Create a line chart which plots the average stat (HP, Attack, Defense, Special Attack, Special Defense, and Speed) of pokemons across it's generations.

Question 4 - Hotel Booking

There is Jakarta hotel booking data (hotel_booking.csv provided). Use data only arrival_date_year in 2016. Then do the following processing.

a. Read the CSV file and get the average number of bookings each month and visualize it in Barchart
b. Visualize the number of adult guests into diagram / chart that you think is appropriate
c. Visualize the number of canceled bookings (is_cancelled, 0 = not cancelled, 1 = canceled) each week using the appropriate chart