Instructor:
- Mr. Nazeef Ul Haq

Registration No._____

Name: _____

Guide Lines/Instructions:
You may talk with your fellow CS261-ers about the problems. However:
- Try the problems on your own *before* collaborating.
- Write up your answers yourself, in your own words. You should never share your typed-up solutions with your collaborators.
- If you collaborated, list the names of the students you collaborated with at the beginning of each problem.

Today's Task:
- Scrap the data from the web pages.
- Learn to manipulate with matrices
- Data plot using graphs

Part 1: Data Plot using Graphs
Download the data https://www.kaggle.com/datasets/tanuprabhu/population-by-country-2020?resource=download

```
!pip install matplotlib #if already not installed
import matplotlib.pyplot as plt
import pandas as pd


df = pd.read_csv('population_by_country_2020.csv' )



print(df.dtypes)
list1 = df['Country (or dependency)'].values.tolist()
list2 = df['Population (2020)'].values.tolist()

plt.bar(list1, list2,width = 1, color = ['red', 'green'])

plt.show()
```

Download the following data
https://www.kaggle.com/datasets/nurudeenabdulsalaam/fitbit-fitness-tracker-data
1. Draw the line chart for the total number of steps on daily basis.
2. Draw the bar chart for the daily distance covered.
3. Draw the scatter chart for the total time in the bed.
4. Draw the Pie chart for the hourly steps on the 12th April 2016.

Part2: Differentiating the Disease Category

You are given data of some individuals who have one of four diseases. The data is collected using the various symptoms of the patient. The names of four diseases are as follow:

- ALLERGY
- COVID
- COLD
- FLU

The symptoms upon which data is collected are given below:

Name	Possible Values	Description
COUGH	0 , 1	
MUSCLE_ACHES	0 , 1	
TIREDNESS	0 , 1	
SORE_THROAT	0 , 1	
RUNNY_NOSE	0 , 1	Mucus dripping or "running" out of your nose
STUFFY_NOSE	0 , 1	A stuffy or congested nose occurs when the tissues lining it become swollen.
FEVER	0 , 1	
NAUSEA	0 , 1	Stomach discomfort and the sensation of wanting to vomit.
VOMITING	0 , 1	
DIARRHEA	0 , 1	Diarrhea is loose or watery stool

SHORTNESS_OF_BREATH	0 , 1	
DIFFICULTY_BREATHING	0 , 1	
LOSS_OF_TASTE	0 , 1	
LOSS_OF_SMELL	0 , 1	
ITCHY_NOSE	0 , 1	
ITCHY_EYES	0 , 1	
ITCHY_MOUTH	0 , 1	
ITCHY_INNER_EAR	0 , 1	
SNEEZING	0 , 1	
PINK_EYE	0 , 1	

- In short, data contains 26 columns: 25 symptoms and last columns is named as label. Data is available on the following link(https://bit.ly/3DbA7Ee):

You are required to perform the following tasks.
1. Load data using pandas

2. Draw scattered graph between each symptom and the label. Discuss how each symptom is dependent on the label.
3. All symptoms are labeled with the categorization from one of the four classes/labels. Now let's take some data (test data) provided here(https://bit.ly/3D7FQv4), and compare the each entry with the provided data earlier. You need to calculate the distance of each entry in test data with main data file according to following Euclidian formula.

$$d(x,y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$$

4. After calculation of distance, assign the label to test data entry with the same label as in main data.
5. Analyze the time complexity of your code.
6. Now split the main data in two parts: Part1 contain the data with no change, Part 2: contains the data without label.
7. Now run your algorithm on Part2 while taking Part1 data as reference. How much labels you assign correctly? Give the answer in percentage. Can you increase the percentage of ratio of correctly classified instances.
8. Provide the recursive algorithm for above task, and provide complexity analysis.
9. Improve your algorithm and design an efficient algorithm. (you can change the mechanism of data storage) Compare it with previous algorithm.

Part 3: Learn to Manipulate Matrices

1. Write a function to print a 2D matrix in the following format
   `printMatrix(A, starting_index, rows, columns)` – starting_index is tuple
   `(2,4), A is matrix`
   3 4 5
   2 5 7
2. Add A and B. Write a function with following prototype(do not use built in function)
   `MatAdd(A,B)`
3. Now instead of addition of two full matrices, we need to add partition of two matrices.
   `MatAddPartial(A, B, start, size)`
   Where start is the vector representing position (x,y). size is the integer. Lets start is (2,3) and size is 2 then you need to add submatrices of A and B.
4. Write a function `MatMul(A,B)` with iterative function.
5. Write a recursive function `MatMulRecursive(A,B)` with partitions.
6. Implement Strassen Method as function `MatMulStrassen(A,B)`

Part 4: Scrap the data from web pages

```
!pip install selenium #if already not installed
!pip install beautifulsoup4 #if already not installed

from selenium import webdriver
from bs4 import BeautifulSoup
import pandas as pd #install chrom webdriver
#webdriver can be downloaded from
#https://sites.google.com/chromium.org/driver/downloads/version-
selection?authuser=0
driver = webdriver.Chrome(executable_path='G:\Program
Files\Anaconda3\chromdriver\chromedriver.exe')

products=[] #List to store name of the product
prices=[] #List to store price of the product
ratings=[] #List to store rating of the product
driver.get("https://www.flipkart.com/gaming-
laptops-
store?otracker=nmenu_sub_Electronics_0_Gaming%20Laptops&otracker=nmenu_sub_E
lectr onics_0_Gaming%20Laptops")

content = driver.page_source
soup = BeautifulSoup(content)
# print(soup)
for a in soup.findAll('div',attrs={'class':'_37K3-p'}):
    print (a)
    name=a.find('a', attrs={'class':'s1Q9rs'})
    price=a.find('div',attrs={'class':'_30jeq3'})
    rating=a.find('div',attrs={'class':'_3LWZlK'})
    products.append(name.text)
    prices.append(price.text)
    ratings.append(rating.text)

df = pd.DataFrame({'Product Name':products,'Price':prices,'Rating':ratings})
df.to_csv('products.csv', index=False, encoding='utf-8')
```

Problem:

Scrap all the courses from eduko and save in the csv file with the following columns in this order.
CourseCode, Title, Description, CLO1, CLO2, ClO3, CLO4, TextBook1, TextBook2, Instructor, Semester

What to Submit:

1. Only .py files are allowed.
2. You are required to submit the following files.
   a. Part1.py
   b. Part2.py
   c. Part3.py

    d. Part4.py

    e. Lab4.docx with all the descriptive answers typed. (you may take snapshot of handwritten work to paste in the file, but text should be typed)

3. Functions names, input and output should be exactly same.
4. Zip all files, and submit on eduko