Assignment 1

In this assignment, you will be required to analyze the distribution of life expectancy and design several health improvement strategies by using AI methods. You will need to use Python but notably: Please do NOT use external libraries (e.g., Numby) except for loading data. All the Python codes should have appropriate comments to help us better understand your code logic.

You can find and download the data in Life Expectancy Data.csv or via https://www.kaggle.com/datasets/kumarajarshi/life-expectancy-who

Task 1:

By using the original data, please calculate the mean (μ) and standard deviation (σ) of life expectancy of "Developing Countries" and "Developed Countries". Note that each country has numerous years of data, and you will need to first calculate the mean life expectancy of the target country. Please fill the results in the following table:

	mean (μ)	standard deviation (σ)
Developing Countries		
Developed Countries		

Please submit your code (named ***calculate_mean_std.py***). Please explain why standard deviation can be important to characterize the life expectancy statistics:

_____

_____

_____

Task 2:

By using the numbers you calculated from the last task, please calculate the label of each country with the following logic:

For the target country, compared with the other countries from the same category (either "Developing Countries" or "Developed Countries"), if the target country's life expectancy is smaller than μ - σ, label this country as "Short"; if the target country's life expectancy is larger than μ + σ, label this country as "Long"; otherwise, label this country as "Normal". Please print the results in a ***true_label.txt*** file.

Task 3:

There are 20 predicting variables. All predicting variables can be divided into several broad categories: Immunization related factors, Mortality factors, Economical factors and Social factors. Please design a machine learning model (you don't need to implement it), to predict the country labels that you generated from Task 2.

Overall Model Description:

_____
_____
_____

Please answer the following questions:

- What features will you use the predict the label? Why they can be important

_____
_____
_____

- As each country has multiple years of data, how to preprocess your data to generate the features that you need?

_____
_____
_____

- Can you use dynamic information (e.g., GDP is increasing/decreasing rates) to enhance the model that you just created?

_____
_____
_____

Task 4:

In **sample.csv** file, there are some results generated by a machine learning, please use the labels (Groundtruth) you generate from Task 2 to calculate the **precision, recall and F1 score**s. Please save your code as ***Evaluate_result.py***

To start this Assignment, please make sure your laptop has the Python environment. We provide a zip file consisting of "Life Expectancy Data.csv", "Sample.csv" and "Assignment 1.ipynb" files. In this, two .csv files are original data and result data, separately. "Assignment 1.ipynb" is a Jupyter Notebook file, in this file, we give the function frame (you only need to fill in the blank with your own solution). Using this file requires you to have Jupyter Notebook on your laptop, you can install this following this link (https://docs.jupyter.org/en/latest/install/notebook-classic.html). You can either choose to create a python file to do the above tasks or use this .ipynb file.

Submission: Your code files and a PDF file (containing your solution for tasks and the results to report).