**Homework #3**

**Your name:**

**Your student ID:**

**Please submit this word docx file and the .ipynb format of your code. .py is not accepted. Make sure you run your code; points will be deducted if you did not run the code.**

**(Hint: please use the class code as your reference to start with)**

Instructions:

1] Please answer the following questions and submit this document as well.

2] Please download the data called *data\_Q1.csv* for Problem 1 *data\_Q2.csv* for question 2 and *Homework\_3\_2024.ipynb*

Problems:

**Q1-1. Do an initial exploratory of the data (using the describe method) and answer the following questions:**

a. What is the mean of the annual income? [5 points]

b. What is the max value of the Age? [5 points]

c. What is the range of Spending Score? [5 points]

**Q1-2. There are three dimensions (Age, Annual Income, Spending Score) that could be used in K-Means Clustering, select two of them and show all possible cases (in total 3).**

a. Using plt.scatter to plot the data, based on the plot decide the appropriate K value by yourself. [10 points]

b. Perform a K-Means Clustering on data using "Kmeans". [10 points]

c. Print out the results by plotting the data colored by their labels and plot the cluster centers as determined by the K-Means estimator. [10 points]

**Q2. The data data\_Q2 can be clustered into different clusters. Try different K values (MAX 7) and find out which K value is the best for this dataset.** [30 points]