**Homework #3**

**Your name:** Michael Savino

**Your student ID:** 10453739

**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:**

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

**The mean of annual income is 60.56.**

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

**The max value of Age is 70.**

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

**The range of Spending Score is 98.**

**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).**

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

**For Age vs Annual Income, there is no distinct cluster so I left it to do it as 1 cluster.**

**For Age vs Spending Score, an argument can be made there there is two clusters so I decided to let it be 2.**

**For Annual Income vs Spending Score, there is a clear distinction between the 5 clusters, so the cluster was equal to 5.**

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]

A graph with blue dots

Description automatically generatedA chart with many colored dots

Description automatically generatedA graph with blue dots

Description automatically generatedA chart of different colored circles

Description automatically generatedA diagram of a number of blue dots

Description automatically generatedA diagram of a number of colored dots

Description automatically generated

**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]

A graph of blue dots

Description automatically generated

A graph showing a number of dots

Description automatically generated

A chart with yellow and purple dots

Description automatically generated

A group of colored dots

Description automatically generated

A group of colored dots

Description automatically generated

By looking at the different amounts of clusters (Tested with 1, 2, 4, and 6 clusters) we can determine that 4 clusters is the best for this dataset as there is clearly 4 distinct clusters shown.