# CSCI 2410 Introduction to Data Analytics Using Python Homework Assignment #4

**HW Programming #4: - Data Analytics with K-Means Clustering**

Tasks: Experiment with the **K-means** clustering technique on ‘iris.xlsx’ dataset. Assignment Instructions:

1. Run the **K-Means** clustering technique on the sheet 1 of ‘iris.xlsx’ dataset from your local directory
   1. **[30%]** Run with k = 2, k= 3, k = 4, k = 5, and k = 6, respectively

Done. Next page

* 1. **[10%]** Get the **silhouette coefficients** for each run of the k values

Done. Silhouette coefficients are [0.6808, 0.5526, 0.4978, 0.4885, 0.3682]

* 1. **[30%]** Plot the clustered data in each of the attribute pairs (Total 6 plots) for each run of k

Done. Next page

* 1. **[30%]** Calculate the optimal **k** number by using the **elbow method**

Done

Python libraries needed: numpy, pandas, sklearn.cluster-KMeans, klearn.metrics- silhouette\_score, matplotlib.pyplot, scipy.spatial.distance-cdist

A computer screen shot of a program code

Description automatically generated

A graph with a blue line

Description automatically generatedA screen shot of a computer

Description automatically generatedA screenshot of a computer program

Description automatically generated

A screenshot of a computer

Description automatically generated

**A chart with blue dots and red dots

Description automatically generated**A chart with blue and orange dots

Description automatically generatedA chart with blue dots and green and orange dots

Description automatically generated**A chart with blue dots and red x

Description automatically generatedA chart with blue dots and orange dots

Description automatically generated**A diagram of a number of dots

Description automatically generated

**A chart with green and orange dots

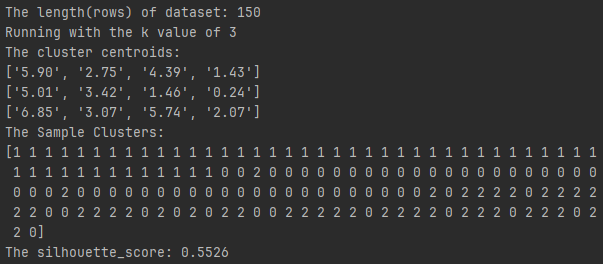
Description automatically generatedA chart with different colored dots

Description automatically generatedA chart with different colored dots and numbers

Description automatically generatedA chart with different colored dots

Description automatically generatedA chart with different colored dots

Description automatically generatedA diagram of different colored dots

Description automatically generated**

**A chart with different colored dots

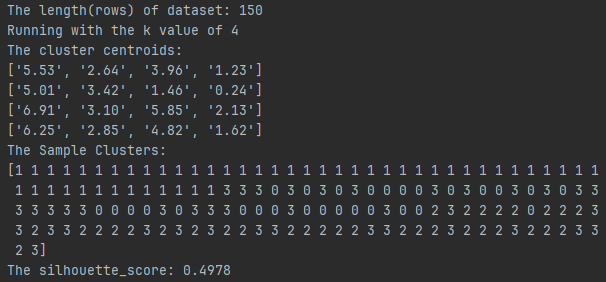
Description automatically generatedA chart with different colored dots and numbers

Description automatically generatedA chart with colored dots and numbers

Description automatically generatedA diagram of a number of dots

Description automatically generatedA chart with different colored dots

Description automatically generatedA diagram of a number of dots

Description automatically generated****A chart with different colored dots

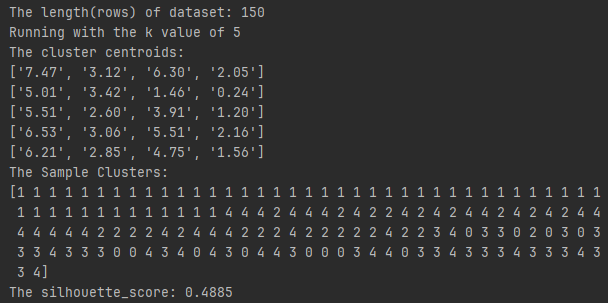
Description automatically generatedA chart with different colored dots

Description automatically generatedA chart with different colored dots

Description automatically generatedA diagram of a number of dots

Description automatically generated with medium confidenceA chart with different colored dots

Description automatically generatedA diagram of a number of dots

Description automatically generated with medium confidence****A chart with different colored dots

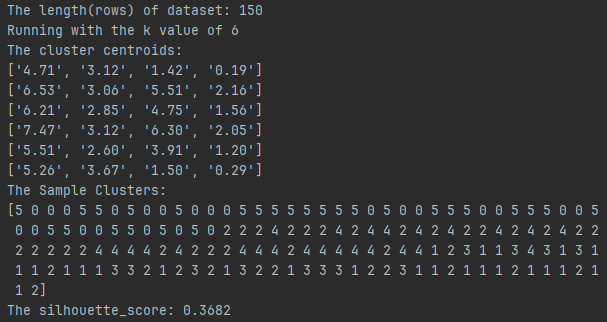
Description automatically generatedA chart with different colored dots and numbers

Description automatically generatedA diagram of a clustering data

Description automatically generated with medium confidenceA chart with different colored dots and numbers

Description automatically generated with medium confidenceA chart with different colored triangles

Description automatically generatedA chart with different colored dots and numbers

Description automatically generated**

# Requirements for the Submission of Programming/Homework Assignments

1. Well-documented program list (the .py files)

**20% of total points** if no .py file submitted.

Done

1. Three annotated program test and run examples (screenshots) that **show different and representative test cases** with **input, output, and the parameter settings of the program runs clearly marked/annotated**. You can do the annotations by
   1. Pasting the screenshots into a WORD document,

Done

* 1. Editing on the WORD document pages for the required marks and annotations,

Done.

Testing and running examples, as well as annotations, were provided inside the screenshots.

* 1. Converting the document to pdf for submission (it is ok to submit the WORD file directly without converting to pdf).

Done

**20% of total points** will be taken off if run examples are not representative.

**20% of total points** will be taken off if run examples are not clearly marked/annotated.

1. A discussion page
2. Hardware and software used by your program,

I completed this assignment using my personal computer with PyCharm Professional Version: 2023.2.1.

1. Features of your program, e.g., data structures, algorithms, programming styles, etc.

The program incorporates various data analysis and visualization features. It employs the K-means clustering algorithm to analyze the Iris dataset and includes an elbow method for determining the optimal number of clusters. The program effectively loads, analyzes, and visualizes data.

1. Problems you encountered during your work, and

None

1. Assigned discussion problems, if there is any.

No assigned discussion problems

1. Fill in the following table and submit it along with your above submissions.

|  |  |  |  |
| --- | --- | --- | --- |
| Total (approximate) time spent on the assignment | 14 hours | Total (approximate) time for the correction part | 2 hours |

|  |  |
| --- | --- |
| Problems and difficulties encountered | None |
| Reflections (good and bad) on the assignment | Good: A snippet of lines of code was provided  Bad: None |
| Any comments and suggestions | None |

**20% of total points** will be taken off if no discussion page is submitted.