NORTH-WEST UNIVERSITY
YUNIBESITI YA BOKONE-BOPHIRIMA
NOORDWES-UNIVERSITEIT

Module Description: Artificial Intelligence **Module Code:**CMPG 313

Practical Assessment: Clustering fitness tracker data.

Introduction

In this assignment, students will apply clustering techniques to analyze simulated fitness tracker data. The goal is to identify distinct user groups based on their activity levels (steps per day) and sleep patterns (hours slept per night). This exercise aims to enhance students' understanding of clustering algorithms, particularly KMeans, and their application in exploratory data analysis.

Learning Objectives

- Become familiar with using Google Colab/Jupyter Notebook for data manipulation.
- Gain practical experience with libraries like NumPy and scikit-learn.
- Explore data visualization with Matplotlib.
- Use these libraries for contextually relevant information.

Steps

1. **Data Generation:**
   Simulate data for three user clusters representing different activity and sleep patterns. Define parameters for each cluster, including:
   - Number of users for each cluster (Use a random interval of 100 between 100 and 1000)
   - Ranges for steps per day and hours slept per night using the following table:

Cluster	Steps per Day (Min)	Steps per Day (Max)	Hours Slept per Night (Min)	Hours Slept per Night (Max)
Cluster 1 (Active)	8000	12000	7	9
Cluster 2 (Moderately active)	5000	8000	6	7.5
Cluster 3 (Least active)	2000	5000	5	6.5

2. **Data Visualization:**

   Plot the generated data on a scatter plot. Label the x-axis with "Steps per Day" and the y-axis with "Hours Slept per Night". Title the plot as "Fitness Tracker Data Visualization".

3. **KMeans Clustering:**

   Implement the KMeans clustering algorithm using the scikit-learn library. Fit the KMeans model to the generated data. Retrieve the cluster labels assigned by the KMeans algorithm.

4. **Cluster Visualization:**

   Visualize the clustered data on a scatter plot. Use different colours to represent each cluster. Label the x-axis with "Steps per Day" and the y-axis with "Hours Slept per Night". Title the plot as "KMeans Clustering of Fitness Tracker Data".

5. **Cluster Analysis:**

   - Analyse the characteristics of each cluster based on average sleep values.
   - Print the average steps and sleep values for each cluster.

Report:
Finally, write a report that will be submitted as a .pdf containing screenshots of the written code and the output. Add personal identifiers (your name and student number) to each part of your code. Ensure the code is well-documented and follows best coding style and readability practices. Include any additional insights or observations gained from the clustering analysis. Please include full copies of the code (not image screenshots) and remember that any suspicion of plagiarism will be strictly punished.