# Third Year B. Tech (EL & CE)

**Semester: VI Subject:** Data Science for Engineering

**Name: Shreerang Mhatre Class: TY**

**Roll No: 52 Batch: A2**

# Experiment No: 07

**Name of the Experiment**: **Clustering using Python**

**Performed on: 25/04/2024**

**Submitted on: 25/04/2024**

**Problem Statement:**

**Aim:**

Write a python program to perform Clustering: We have the data for workout as below.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Date | Distance\_km | Duration\_min | Delta\_last\_workout | Day\_category |
| 10/17/17 | 4.3 | 21.58 | 1 | 0 |
| 11/04/17 | 1.9 | 9.25 | 18 | 1 |
| 11/18/17 | 1.9 | 9.0 | 14 | 1 |
| 11/23/17 | 1.9 | 8.93 | 5 | 0 |
| 11/28/17 | 2.3 | 11.94 | 5 | 0 |
| 11/29/17 | 2.8 | 14.05 | 1 | 0 |

To keep track of your performance you need to identify similar workout sessions. Clustering can help you group the data into distinct groups, guaranteeing that the data points in each group are similar to each other. Perform following steps:

1. Load the Data
2. Data Exploratory Analysis: Pair Plot and Distance versus workout duration, distance versus duration with the number of days, correlation (Scatter plot) to get idea about correlation between different features.
3. Select K-means clustering for model and get the clusters.
4. Evaluate the performance of the model.

















