**UNIVERSITY OF PETROLEUM & ENERGY STUDIES, DEHRADUN**

| Program | MCA | Semester | I |
|---|---|---|---|
| Course | **Introduction to Data Science** | **Course Code** | **CSDS 7001** |

| SN | Questions | COs | Marks |
|---|---|---|---|
| 1 | Define data science and explain its significance in various sectors like business, industry, and government. | CO1 | 10 |
| 2 | Explain the following terms: (*i*) types of data and datasets, (*ii*) data quality, and (*iii*) data issues, and (*iv*) data models. | CO1 | 10 |
| 3 | Define the data wrangling. Write a short note on data cleaning, data aggregation, and data sampling. | CO1 | 10 |
| 4 | The frequency distribution of weight in grams of mangoes of a given variety is given below. Weight in grams:   410-419   420-429   430-439   440-449   450-459   460-469   470-479 Number of mangoes:   14          20          42          54          45          18          7 Calculate the arithmetic mean, and the median. | CO2 | 10 |
| 5 | Explain the discretization, binarization, normalization, and data smoothening for numerical data in data science. | CO2 | 10 |