## SOMAIYA
### VIDYAVIHAR UNIVERSITY

| Semester: January 2023 – May 2023 |||||
|---|---|---|---|---|
| **Examination: In-Semester Examination** |||||
| Programme code: 54 <br> Programme: B.Tech. Computer Engineering <br> (Honours in Data Science and analytics ) || Class: SY | Semester: IV <br> (SVU 2020) ||
| Name of the Constituent College: <br> K. J. Somaiya College of Engineering || Name of the department: <br> COMP ENGG |||
| Course Code:    116h54C401 | Name of the Course: Applied Data Science ||||

| Question No. | | Max. Marks | CO Mapped | BT Level |
|---|---|---|---|---|
| Q1 | Attempt any TWO (2) of the following <br> A. Illustrate with suitable example how data science can be used to add value to eLearning platform(ex. Coursera) <br> B. Explain Velocity, veracity and volume with example. <br> C. Is data science is expensive process? Justify your answers. | 05 mark Each | CO1 <br> CO1 <br> CO1 | Apply <br> U <br> U |
| Q2 | Attempt any TWO (2) of the following <br> a) What is sample and population? List types of hypothesis. Explain each with example. <br> b) Eleven football players and their score for 30 matches are: 12, 13, 20, 19, 14, 15, 21, 30, 22, 09, 06.Give the five-number summary and draw the boxplot. <br> c) A company packages almonds in 300gm packets using the machine. A sample of 16 packets is taken from the production line at random time intervals and their contents are weighted. The mean weight of 16 packets is found as 299.5. Can we conclude that machine is working properly using statistical hypothesis testing? In the hypothesis testing clearly show all the 5 steps in the calculation, considering the significance level of 5%.   (Assume value of  test statistic=2.131) | 05 mark Each | CO1 <br><br> CO2 <br><br> CO3 | R <br><br> Apply <br><br> Evaluate |
| Q3 | Attempt any TWO (2) of the following <br> a) What are different plots? Explain significance of each. <br> b) Explain the quality metric for dataset selection. <br> c) Use min-max normalization method to normalize the following group of data by setting min=0 and max=1: 150, 200, 100, 500,800 | 05 mark Each | CO3 <br> CO3 <br> CO3 | R <br> U <br> Apply |