

**K. J. Somaiya Institute of Technology, Sion, Mumbai-22**  
(Autonomous College Affiliated to University of Mumbai)

April – May 2023	
(B.Tech Program: Electronics and Telecommunication Engineering Scheme I/II: II)	
Examination: LY	Semester: VIII
Course Code: EXDLC8032 <sup>1</sup>	Course Name: Fundamentals of Data Science
Date of Exam: 18/05/2023	Duration: 2.5 Hours <span style="float: right;">Max. Marks: 60</span>

Instructions:				
(1) All questions are compulsory.				
(2) Draw neat diagrams wherever applicable.				
(3) Assume suitable data, if necessary.				
		Max. Marks	CO	BT level
Q 1	Solve any six questions out of eight:	12		
i)	List 6 features of python programming language	2	CO1	U
ii)	List strategies involved in Data Dimensionality reductions	2	CO2	U
iii)	Explain Group by functions in Pandas. Illustrate with suitable examples	2	CO3	U
iv)	Explain steps involved in KDD.	2	CO4	U
v)	List Tasks involved in Data Mining	2	CO4	U
vi)	Define data Warehouse	2	CO5	U
vii)	List various schemes for warehouse designing	2	CO5	U
viii)	Explain Autoregressive	2	CO6	U
Q2.	Solve any four questions out of six.	16		
i)	How to classify data mining system? Discuss	4	CO1	U
ii)	Describe about Data discretization?	4	CO2	U
iii)	What is the drawback of k-means algorithm? How can we modify the algorithm to diminish that problem?	4	CO3	U
iv)	Write partitioning around methods.	4	CO4	U
v)	Define Clustering? Explain about Types of Data in Cluster Analysis?	4	CO5	U

**K. J. Somaiya Institute of Technology, Sion, Mumbai-22**  
(Autonomous College Affiliated to University of Mumbai)

April - May 2023

(B.Tech Program: Electronics and Telecommunication Engineering Scheme I/II: II

Examination: LY

Semester: VIII

Course Code: 1UEXDLC8032

Course Name: Fundamentals of Data Science

Date of Exam: 18/05/2023

Duration: 2.5 Hours

Max. Marks: 60

vi)	Why is data integration required in a data warehouse, more so than in an operational Application?	4	CO6	U										
Q3.	Solve any two questions out of three.	16												
i)	Suppose that the data for analysis includes the attribute salary. We have the following values of salary (in Thousands of INR), shown in increasing order: 30, 36, 47, 50, 52, 52, 56, 60, 63, 70, 70, 110. i. What are the mean, median, mode and midrange of the data? ii. Find first quartile, second quartile and third quartile of the data. iii. Show the boxplot of the data	8	CO1	U										
ii)	Explain decision tree induction algorithm for classifying data tuples and with suitable example	8	CO4	U										
iii)	Use the k-means algorithm and Euclidean distance to cluster the following 8 examples into 3 clusters: A1=(4,10), A2=(2,10), A3=(6,8), A4=(10,16), A5=(5,7), A6=(4,6), A7=(4.6), A8=(2,5).	8	CO5	U										
Q4.	Solve any two questions out of three.	16												
i)	Consider the following data set given in table below and solve using apriori algorithm for two iterations with minimum support equals to 2. <table border="1" style="margin-left: auto; margin-right: auto;"> <thead> <tr> <th>Transaction Id</th> <th>Items</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>1,3,4,5</td> </tr> <tr> <td>2</td> <td>2,3,4</td> </tr> <tr> <td>3</td> <td>1,2,3,5</td> </tr> <tr> <td>4</td> <td>2,5</td> </tr> </tbody> </table>	Transaction Id	Items	1	1,3,4,5	2	2,3,4	3	1,2,3,5	4	2,5	8	CO2	U
Transaction Id	Items													
1	1,3,4,5													
2	2,3,4													
3	1,2,3,5													
4	2,5													
ii)	What are the components of Data Mining (DM) and explain the various operation of Data mining and DM Techniques?	8	CO4	U										
iii)	Explain in Brief Autoregressive Integrated Moving average (ARIM)	8	CO6	U										