COURSE CODE (CREDITS): 19BIWCI637 (2)   **MAX. MARKS:** 25

COURSE NAME: STATISTICS AND EXPLORATORY DATA ANALYTICS

COURSE INSTRUCTORS: Dr. AMOL VASDEVA   **MAX. TIME:** 1 Hour 30 Minutes

Note:
(a) All questions are compulsory.
(b) Marks are indicated against each question in square brackets.
(c) The candidate is allowed to make Suitable numeric assumptions wherever required for solving Problems
(d) Any kind of calculator is not allowed in the examination

1. You are given a matrix M. Write a Python program to decompose the given matrix into three components U, S, and V using Singular Value Decomposition (SVD) method. Next, use U, S, and V to convert into original matrix M. **(CO-3)**   **[4 marks]**

2. The ratings of various online books are given in a data set "bookrating.csv". The fields include user_id, book_id, and rating (1 to 5). Write a Python program to perform the following:
   a. Display the active users, i.e. the user who reviewed more than ten books.
   b. Display the poplar books, i.e., the books that were reviewed by more than ten users.
   c. Keep and display only poplar books and active users. **(CO-3)**   **[6 marks]**

3. A hierarchical cluster is given as ((((D, F), E), C), (A, B)). The steps involved in the formation of hierarchical cluster are given as below. Draw a Dendogram for this hierarchy based on the following information. **(CO-4)**   **[4 marks]**
   • In the beginning, there are six clusters: A, B, C, D, E, and F.
   • The clusters D and F have been merged into a cluster (D, F) at a distance 0.50
   • The clusters A and B have been merged into a cluster (A, B) at a distance 0.71
   • The cluster E and (D, F) have been merged into a cluster ((D, F), E) at a distance 1.00
   • The clusters ((D, F), E) and cluster C have been merged into a cluster (((D, F), E), C) at a distance 1.41
   • The clusters (((D, F), E), C) and cluster (A, B) are merged into a cluster ((((D, F), E), C), (A, B)) at a distance 2.50

4. Based on the following dataset, use K-mean clustering algorithm to allocate each of the six objects into two groups (i.e. K=2). **(CO-4)**   **[5 marks]**

Object	Weight Index	pH Value
Medicine A	1	1
Medicine B	2	1
Medicine C	4	3
Medicine D	5	4

5. Differentiate between the following terms. **(CO-3)**   **[2×3=6 marks]**
   a. Box-Cox Transformation and Yeo-Johnson Transformation
   b. Hierarchical Clustering and K-Mean Clustering
   c. Data Standardization and Data Normalization