JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT

TEST-2 EXAMINATION- 2022

B. Tech (CSE/IT) Semester VI

COURSE CODE: 19BIWCI637   **MAX. MARKS:** 25

COURSE NAME: Statistics and Exploratory Data Analytics

COURSE CREDITS: 2   **MAX. TIME:** 1 Hour 30 Minutes

Note: All questions are compulsory.

1. Write a Python program to import any two attributes (say A and B) from any given CSV file (say sample.csv). Next, draw a subplot with three rows and three columns as shown in Fig.1 below. Use three different transformation methods (square, log, and Box-Cox) for each row (as shown in Fig 1). In the first column, display the distribution plot before transformation. In the second column, display the distribution plot after the transformation. Finally, in the third column, display the QQ plot after the transformation. (CO-2) **[6 marks]**

Transformation Type	Distribution Plot Before Transformation	Distribution Plot After Transformation	QQ Plot After Transformation
Square Transformation			
Log Transformation			
Box-Cox Transformation			

Fig. 1. Scheme of Plotting

2. A dataset of students' marks in three subjects is given as below. Based on this dataset, write a Python program to perform and display the following operations involved in Principal Component Analysis (PCA). (CO-3) **[5 marks]**

Student	Math	English	Arts
1	90	60	90
2	90	90	30
3	60	60	60
4	60	60	90
5	30	30	30

- Covariance matrix
- Eigen values
- Eigen vectors
- Reduced dimensional matrix (transformation of the original matrix to a reduced dimensional matrix)

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) have been 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) **[6 marks]**

Object	Attribute 1 (X): weight index	Attribute 2 (Y): pH
Medicine A	1	1
Medicine B	2	1
Medicine C	4	3
Medicine D	5	4

5. Differentiate between the following terms. (CO-3) **[2×2=4 marks]**
   - PCA and Singular Value Decomposition
   - Box-Cox Transformation and Yeo-Johnson Transformation