

--	--	--	--	--	--	--	--	--	--

**Sixth Semester B.E. Degree Examination, June/July 2024**  
**Data Science and Visualization**

Time: 3 hrs.

Max. Marks: 100

**Note:** Answer any FIVE full questions, choosing ONE full question from each module.

**Module-1**

- 1 a. Define the Data Science. What are the skill sets required for Data Scientists? (08 Marks)  
 b. Explain the following concepts : (12 Marks)  
     i) Datafication                      ii) Statistical modeling.

**OR**

- 2 a. Discuss the "Fitting Model" in Data Science. (08 Marks)  
 b. Explain the following concepts with suitable examples : (12 Marks)  
     i) Population                      ii) Samples.

**Module-2**

- 3 a. Briefly explain the Exploratory Data Analysis (EDA) process with an example. (08 Marks)  
 b. Explain the working of the K – Nearest Neighbour (K – NN) algorithm, with a suitable example. (12 Marks)

**OR**

- 4 a. Briefly explain the Data Science Process with a neat diagram. (08 Marks)  
 b. Explain the working of the K – mean algorithm with a suitable example. (12 Marks)

**Module-3**

- 5 a. What is Feature Extraction? Discuss the Principal Component Analysis (PCA) with example. (08 Marks)  
 b. What is a Decision tree and how does the decision tree play a role in feature selection? Explain the implementation with suitable example. (12 Marks)

**OR**

- 6 a. What is the Feature selection? Discuss the need for feature selection of classification of feature selection algorithms. (08 Marks)  
 b. What is Random Forest? Analyze the working of the random forest algorithm, with an example. (12 Marks)

**Module-4**

- 7 a. What is Data Visualization and the importance of Data Visualization? What makes a good visualization? (08 Marks)  
 b. Explain the following concepts : (12 Marks)  
     i) Scatter Plot                      ii) Histogram                      iii) Box Plot.

**OR**

- 8 a. Define “Data Wrangling” and Identify the steps in the flow of the data wrangling process with a neat diagram to measure employee engagement. (08 Marks)
- b. Explain the following concepts : (12 Marks)
- i) Choropleth Map      ii) Line chart      iii) Bubble chart.

### Module-5

- 9 a. Apply a pie chart for water usage. To understand the reason behind it, generate a visual representation of water usage using a pie chart in matplotlib library.

Usage	Clothes washer	Leak	Other	Toilet	Shower	Faucet
Percentage (%)	17	12	8	24	20	19

- (10 Marks)
- b. What is the Visualization layout in matplotlib? Explain to create subplots and grid space using matplotlib with suitable examples. (10 Marks)

OR

- 10 a. Explain the Vertical and Horizontal bar chart with parameters. Create a vertical bar chart with suitable example using matplotlib library. (10 Marks)
- b. What is the “Legend” function and Basic text direction? How to implement legend and basic text in the “Stacked bar chart”? (10 Marks)

\*\*\*\*\*