|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Roll No |  |  |  |  |  |  |  |  |  |  |  |  |

****

**Presidency University**

**Bengaluru**

**Ph. D Course Work End Term Examinations – JAN-FEB 2025**

**Semester**:

**Course Code**: CSE881

**Course Name**: Mining the Social Web Data-Facebook, Twitter, LinkedIn, Instagram

**School:** SOCSE

**Date**: 04-02-2025

**Time**: 9.30 AM TO 12.30 PM

**Max Marks**: 100

**Weightage**: 50%

**Instructions:**

1. *Read the all questions carefully and answer accordingly.*
2. *Do not write any matter on the question paper other than roll number.*

**PART A**

**Answer all the Questions. Each question carries 10 marks. (6Qx 10M= 60M)**

|  |  |
| --- | --- |
| **1.** | **Explain the significance of data structure in data analysis. How does structured data simplify the analysis process compared to unstructured data?** |
| **2.** | **Discuss the role of exploratory data analysis (EDA) in identifying patterns and relationships in structured data. How does it help in decision-making?** |
| **3.** | **Describe three key methods for collecting unstructured data from online platforms. Provide examples of tools and techniques used for this purpose.** |
| **4.** | **What is the importance of data visualization in data exploration and presentation? Compare the use of bar charts, scatter plots, and histograms in summarizing data.** |
| **5.** | **Explain how Python can be used for preparing tools and packages for data analysis. Why is package management important in Python programming?** |
| **6.** | **Discuss the common challenges faced during statistical analysis of large datasets and how python can help overcome these challenges.** |

**PART B**

**Answer all the Questions. Each question carries 20 marks. (2Qx 20M= 40M)**

|  |  |
| --- | --- |
| **7.** | **Design a workflow for collecting data from YouTube using Python. Include the steps for data cleaning, storage, and analysis. Mention key packages required for this process.** |
| **8.** | **Describe the process of performing a statistical analysis using python on a real-world dataset (e.g., sales data or survey results). Explain how to use python to generate descriptive statistics, visualizations, and hypothesis testing. Provide relevant code examples.** |