PRESIDENCY UNIVERSITY
BENGALURU

Department of Research & Development
Mid - Term Examinations - SEPTEMBER 2024

Odd Semester: Ph.D. Course Work		**Date**: 28 /09/2024	
Course Code: CSE879		**Time**: 2:00pm – 3:30pm	
Course Name: Data Analysis and Visualization		**Max Marks**: 50	
Department: CSE		**Weightage**: 25%	

Instructions:

(i) Read all questions carefully and answer accordingly.

(ii) Do not write anything on the question paper other than roll number.

Part A

Answer ALL the Questions. Each question carries 4 marks.		4Qx5M=20M
1	Analyze the effectiveness of using pie charts versus bar charts to visualize categorical data. Provide examples of when each chart type may be misleading and suggest alternative visualizations that would better represent the data.	**5 Marks**
2	Examine the relationship between data outliers and the type of visualizations used to represent data distributions. How would you identify outliers in a dataset, and which visualization techniques can help highlight or mitigate the impact of outliers?	**5 Marks**
3	Explain the principles of data cleaning to the given dataset by identifying and handling missing values. Explain the methods you would use to address these issues and why.	**5 Marks**
4	Examine the relationships between multiple variables in a dataset using a scatter plot matrix. Interpret the insights that this visualization provides, and suggest any correlations or patterns observed.	**5 Marks**

Part B

Answer ALL Questions. Each question carries 15 marks.		2QX15M=30M
5	Using a real-world dataset of your choice (e.g., sales, healthcare, finance), the explain following tasks: a) Clean the dataset by handling missing values, duplicates, and outliers. b) Apply appropriate data transformations (e.g., normalization, encoding) based on the types of variables in the dataset.	**15 Marks**

	c) Create at least three different types of visualizations (e.g., bar chart, heatmap, scatter plot) to uncover insights from the data. d) Explain the rationale behind choosing each visualization technique and how it helps in understanding the dataset.	
6	Suggest a suitable tool (e.g., Power BI, Tableau, Python libraries), explain the procedure to design a dashboard that displays the following: i. Key performance indicators (KPIs) from a dataset of your choice. ii. A multi-dimensional analysis using charts like histograms, box plots, and line graphs to track trends, anomalies, and distributions. iii. Interactive elements (e.g., filters, drill-down options) to allow users to explore different segments of the data. iv. Describe the process you followed to design the dashboard and how each visualization aids in decision-making.	**15 Marks**