RAJALAKSHMI ENGINEERING COLLEGE

DEPARTMENT OF COMPUTER SCIENCE AND DESIGN

CD19P10 – FOUNDATIONS OF DATA SCIENCE

ASSESSMENT TYPE: PROJECT BASED

Internal Marks Split

Exercises Completion	Project Review 1	Project Review 2	Report	Total Marks
10	40	40	10	100

Review 1

24th and 25th October 2024- Offline

Review 2

4th and 6th Nov 2024 – Online

Review 1 - Data Understanding and Cleaning (50%)

Review Objectives:

1. Evaluate the understanding of the dataset and project scope.
2. Assess the data collection strategy, data structure, and completeness.
3. Review the data cleaning techniques applied, including handling missing values and outliers.

Review 1 Presentation slides should include:

Abstract

Understanding the Dataset

Data Cleaning Process

Exploratory Data Analysis

Data Integration and Transformation

Review 1 Rubrics:

Criteria	Max Marks	Description	Marks Obtained

Understanding the Dataset	10	Clarity in explaining the dataset attributes, types of data, and the relationship between features.
Data Cleaning Process	10	Appropriateness and completeness of techniques like dealing with missing values, duplicate removal, handling outliers, and feature standardization.
Exploratory Data Analysis (EDA)	15	Insights from initial data exploration, summary statistics, and visualization methods such as histograms, box plots, or correlation matrices.
Data Integration and Transformation	05	Use of proper data integration and transformation techniques like normalization or aggregation to make the data ready for further analysis.
Project Presentation	10	Quality and clarity of the report and presentation for the current stage.
Total	50	

Review 2 - Insight Generation and Visualization (50%)

Review Objectives:

1. Evaluate the quality of insights derived from the data.
2. Assess the use of visualization techniques to communicate insights.
3. Review model evaluation and residuals if applicable.

Review 2 ppt includes:

1. Model Building Output and explanation
2. Visualization Techniques
3. Model Evaluation

Review 2 Rubrics:

Criteria	Max Marks	Description	Marks Obtained
Model Building	10	Depth and relevance of the insights generated from the cleaned data using statistical or machine learning techniques.	

Visualization Techniques	20	Application of appropriate visualization techniques (e.g., heatmaps, pivot tables, distribution plots) to communicate findings effectively.
Model Evaluation (If Applicable)	10	Quality of regression or other model evaluations through residual plots, distribution plots, and other metrics.
Presentation and Reporting	10	Final presentation and report, including the logical flow of analysis, professionalism, and clarity in communication.
Total	50	