SIMATS ENGINEERING

Saveetha Institute of Medical and Technical Sciences
Chennai- 602105

Engineer to Excel

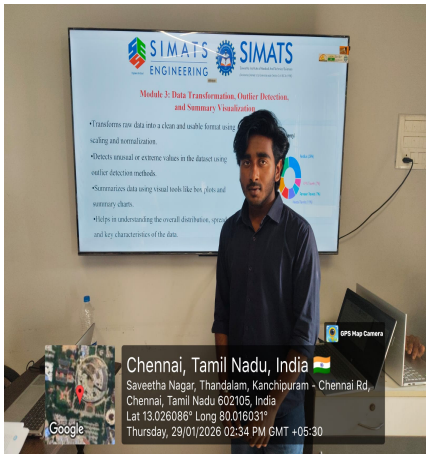
Student Name: G.Vinay kumar Reddy **Reg. No.:192424245**

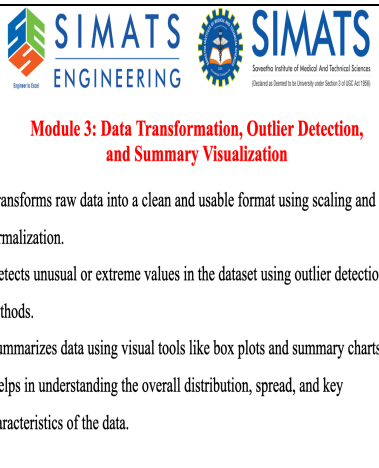
Course Code: DSA-0613 **Slot:** A

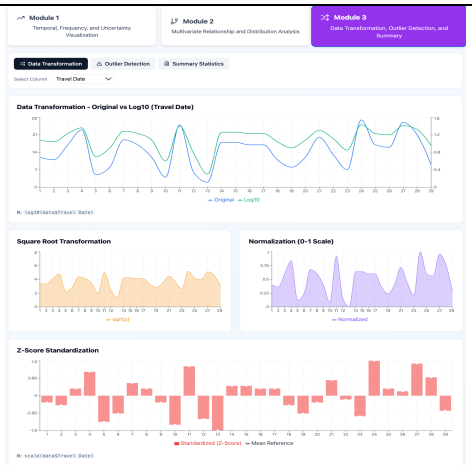
Course Name: Data Handling and Visualization for Data Analytics
Course Faculty: Dr. Kumaragurubaran T, Senthilvadivu Subramanian

Project Title: Exploratory Data Analysis and Visualization of Bus Fare Data Using R Techniques

Module Photographs: (3 photographs –Module Photo, Individual student contribution module work in the project and presentation image)







Project Description: (here you write what you did in this project (contribution) including Model Description

Module 3: Data Transformation, Outlier Detection, and Summary Visualization focuses on preparing and refining Indian bus fare data to improve the accuracy and clarity of exploratory data analysis. Real-world bus fare datasets often contain missing values, extreme fares, inconsistent formats, and noise due to operational and regional variations. This module applies data transformation techniques such as normalization, scaling, log transformation, and categorical encoding to make fare data suitable for meaningful analysis. Using R visualization tools like **ggplot2**, outlier detection is performed through box plots, density plots, and scatter plots to identify unusually high or low fares that may indicate pricing errors, policy exceptions, or special service routes. Summary visualizations, including bar charts, line plots, and aggregated plots, are used to present key insights such as average fare by bus type, route category, and region. These visual summaries help simplify complex datasets and highlight overall trends in the Indian bus fare system. By combining transformation, outlier analysis, and concise visual summaries, this module supports reliable interpretation, better fare comparison, and informed decision-making in public transport planning.

Student Signature **Guide Signature**