SIMATS ENGINEERING

Saveetha Institute of Medical and Technical Sciences
Chennai- 602105

Engineer to Excel

Student Name: Papudesi Mohitha **Reg. No.:** 192324025

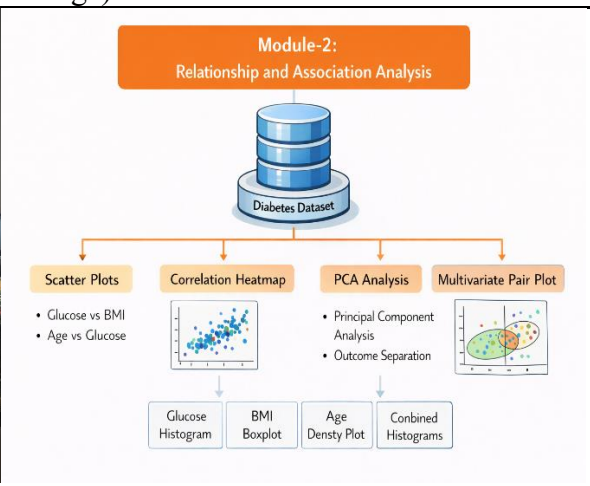
Course Code: DSA0613 **Slot:** A

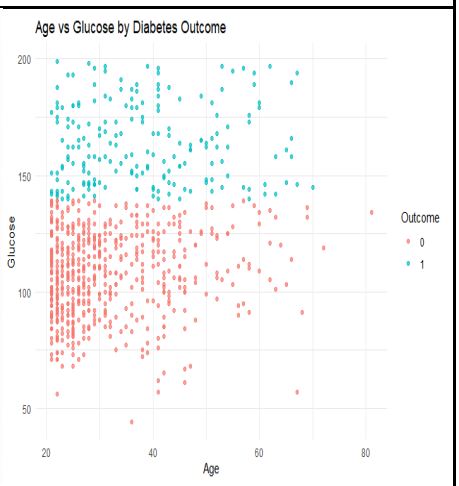
Course Name: Data Handling and Visualization for Data Analytics

Course Faculty: Dr. T. Kumaragurubaran and Dr. Senthilvadivu S

Project Title: Exploratory Data Analysis and Visualization of Diabetes Prediction Dataset.

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

This project focuses on performing Exploratory Data Analysis (EDA) and visualization on a diabetes dataset to understand key factors influencing diabetes and their relationships using the R programming language. The dataset was initially loaded and inspected to study its structure, feature types, and statistical characteristics. Basic preprocessing was performed by identifying medically invalid zero values in features such as Glucose, Blood Pressure, Insulin, and BMI, which were treated as missing values and handled using median imputation to improve data quality.

As part of data enrichment, a rule-based outcome variable was created based on glucose levels to classify individuals as diabetic or non-diabetic, and the updated dataset was saved as a new CSV file for consistent analysis. The major contribution of this project lies in Module-2, where advanced visualization techniques such as scatter plots, correlation heatmaps, principal component analysis (PCA), and multivariate plots were used to analyze relationships and associations among features and the outcome variable. These visualizations helped identify glucose and BMI as influential factors associated with diabetes. Overall, the project provides meaningful visual insights into diabetes-related patterns and serves as a foundation for future predictive modeling. The analysis emphasizes visualization-based understanding rather than clinical diagnosis, making it suitable for exploratory and academic purposes. This approach helps in clearly understanding data trends and feature interactions that influence diabetes outcomes.

Student Signature **Guide Signature**