



SIMATS ENGINEERING

Saveetha Institute of Medical and Technical Sciences
Chennai- 602105



Student Name: S.Tushar Krishna Sai

Reg. No.: 192424406

Course Code: DSA-0613

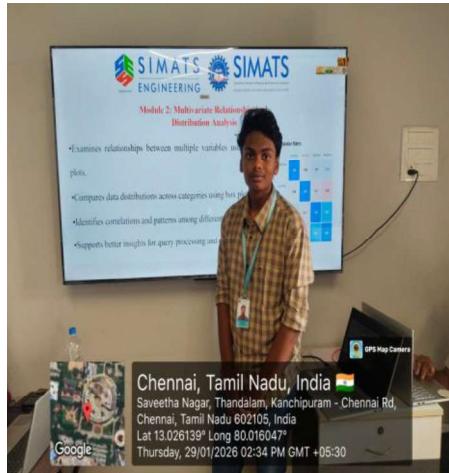
Slot: A

Course Name: Data handling & visualization for data analytics

Course Faculty: DR. Kumaragurubaran T

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)



Module 2: Multivariate Relationship And Distribution Analysis

- Examines relationships between multiple variables in data using visual methods.
 - Compares data distributions across categories using box plots.
 - Identifies correlations and patterns among different variables.
 - Supports better insights for query processing and transportation planning.
- Exploratory Data Analysis (EDA) helps in understanding patterns, trends, and structure of Indian bus fare data using visual methods.
- Scatter plots are used to examine relationships between fare, travel distance, and journey duration, helping identify positive or negative trends.
 - Box plots compare fare distributions across different categories and highlight median values, spread, and outliers.
 - A correlation matrix identifies the strength and direction of relationships among numerical variables influencing bus fare.
 - These visual insights support efficient query processing, better pricing decisions, and improved transportation planning.

Correlation Matrix

	Travel Dist...	Fare Price...	Total Seat...	Duration ...
Travel Dist...	1.00	-0.06	0.10	0.11
Fare Price...	-0.06	1.00	-0.11	-0.20
Total Seat...	0.10	-0.11	1.00	0.54
Duration ...	0.11	-0.20	0.54	1.00

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

Module 2: Multivariate Relationship and Distribution Analysis aims to understand how multiple factors interact to influence bus fare structures in the Indian bus transportation system. Bus fares depend on a combination of variables such as travel distance, bus category (ordinary, express, AC, deluxe), route type, passenger demand, and time of travel. This module applies advanced exploratory data analysis techniques in R to study these complex relationships in an effective visual manner. Using visualization libraries like **ggplot2**, multivariate relationships are explored through scatter plots, faceted plots, grouped box plots, violin plots, and correlation heatmaps. These plots help reveal how fares change across different service types and regions, and how distance and fare are related under varying conditions. Distribution analysis further highlights fare variability, skewness, and outliers, allowing meaningful comparison across states and bus services. By visualizing multiple dimensions simultaneously, this module uncovers hidden patterns and pricing inconsistencies that are not easily detected through numerical analysis alone. Overall, it provides valuable insights for fare regulation, transport planning, and policy formulation, contributing to a more efficient and transparent public bus fare system in India.

Student Signature

Guide Signature