 

RLAB PROJECT

**Branch:** MCA(AIML) **Section/Group:**1(B)

# Semester: I Date of Performance:21/10/24

**Subject Name:** STATISTICAL TECHNIQUES USING R LAB

**Subject Code:** 24CAP-614

* Performed by Rachitesh Sharma and Harsimran Singh (24MCI10257,24MCI10048)

# Aim/Overview of the practical:

Choose a dataset from a repository like Kaggle or UCI Machine Learning Repository and perform exploratory data analysis using R. Explore the distribution of variables, identify outliers, and visualize relationships between variables using plots like histograms, scatter plots, and boxplots.

# Task to be done:

### **Task 1: Load the Dataset**

### Step 1.1: Load the titanic Dataset

Step 1.2: View the First Few Rows

Step 1.3: Check the Structure of the Data

**Task 2: Summarize the Dataset**

Step 2.1: Generate Summary Statistics

Step 2.2: Generate Structure (STR) Statistics

# Task 3: Visualizing the Distribution of Variables

# Step 3.1: Create Histograms

# 

# Task 4: Identifying Outliers

# Step 4.1: Create Boxplot

# 

# Task 5: Analyzing Relationships Between Variables

# Step 5.1: Scatter Plots

# 

# Steps/Commands involved to perform project:

# Loading the dataset titanic and performing EDA

# 

# 

# 

# 

# 

# 

# 

# 

# 

# 

# OUTPUTS-:

# 

# 

# 

# 

# 

# 

# 

# 

# Summary:

# 

# Performed exploratory data analysis (EDA) on the titanic dataset, or alternatively, downloaded a dataset from an online source if you prefer not to use the pre-installed dataset.

# Summarizing the dataset.

# Visualizing the distribution of numerical variables using histograms and density plots.

# Identifying outliers with boxplots.

# Exploring relationships between variables using scatter plots.