CLEANING AND ORGANIZING STUDENT ATTENDANCE RECORDS IN R

A PROJECT REPORT

Submitted by

PRAVINA M(2303811724322084)

in partial fulfillment of requirements for the award of the course

AGI1252 - FUNDAMENTALS OF DATA SCIENCE USING R

in

ARTIFICIAL INTELLIGENCE AND DATA SCIENCE

K. RAMAKRISHNAN COLLEGE OF TECHNOLOGY

(An Autonomous Institution, affiliated to Anna University Chennai and Approved by AICTE, New Delhi)

SAMAYAPURAM – 621 112

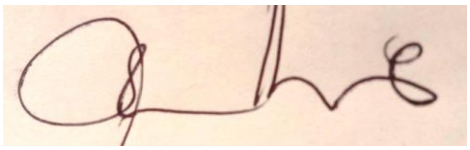
JUNE- 2025

K. RAMAKRISHNAN COLLEGE OF TECHNOLOGY (AUTONOMOUS)

SAMAYAPURAM – 621 112

BONAFIDE CERTIFICATE

Certified that this project report on **" CLEANING AND ORGANIZING STUDENT ATTENDANCE RECORDS IN R"** is the bonafide work of **PRAVINA M (2303811724322084)** who carried out the project work during the academic year 2024 - 2025 under my supervision.



SIGNATURE

Dr.T. AVUDAIAPPAN, M.E.,Ph.D.,

HEAD OF THE DEPARTMENT

PROFESSOR

Department of Artificial Intelligence

K.Ramakrishnan College of Technology (Autonomous)

Samayapuram–621112.

SIGNATURE

Ms.S.Murugavalli., M.E.,(Ph.D).,

SUPERVISOR

ASSISTANT PROFESSOR

Department of Artificial Intelligence

K.Ramakrishnan College of Technology (Autonomous)

Samayapuram–621112.

Submitted for the viva-voce examination held on ……**02.06.2025**.

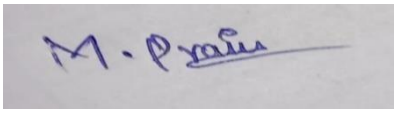
INTERNAL EXAMINER **EXTERNAL EXAMINER**

DECLARATION

I declare that the project report on **"CLEANING AND ORGANIZING STUDENT ATTENDANCE RECORDS IN R"** is the result of original work done by us and best of our knowledge, similar work has not been submitted to **"ANNA UNIVERSITY CHENNAI"** for the requirement of Degree of **BACHELOR OF TECHNOLOGY**. This project report is submitted on the partial fulfilment of the requirement of the completion of the course **AGI1252 - FUNDAMENTALS OF DATA SCIENCE USING R**

.

Signature



PRAVINA M

Place: Samayapuram

Date:30.05.2025

ACKNOWLEDGEMENT

It is with great pride that I express our gratitude and in-debt to our institution "**K.Ramakrishnan College of Technology (Autonomous)**", for providing us with the opportunity to do this project.

I glad to credit honourable chairman **Dr. K. RAMAKRISHNAN**, **B.E.,** for having provided for the facilities during the course of our study in college.

I would like to express our sincere thanks to our beloved Executive Director **Dr. S. KUPPUSAMY, MBA, Ph.D.,** for forwarding to our project and offering adequate duration in completing our project.

I would like to thank **Dr. N. VASUDEVAN, M.Tech., Ph.D.,** Principal, who gave opportunity to frame the project the full satisfaction.

I whole heartily thanks to **Dr. T. AVUDAIAPPAN, M.E.,Ph.D.,** Head of the department, **ARTIFICIAL INTELLIGENCE** for providing his encourage pursuing this project.

I express our deep expression and sincere gratitude to our project supervisor **Ms.S.Murugavalli., M.E.,(Ph.D).,** Department of **ARTIFICIAL INTELLIGENCE,** for her incalculable suggestions, creativity, assistance and patience which motivated us to carry out this project.

I render our sincere thanks to Course Coordinator and other staff members for providing valuable information during the course.

I wish to express our special thanks to the officials and Lab Technicians of our departments who rendered their help during the period of the work progress.

INSTITUTE

Vision:

- To serve the society by offering top-notch technical education on par with global standards.

Mission:

- Be a center of excellence for technical education in emerging technologies by exceeding the needs of industry and society.
- Be an institute with world class research facilities.
- Be an institute nurturing talent and enhancing competency of students to transform them as all – round personalities respecting moral and ethical values.

DEPARTMENT

Vision:

- To excel in education, innovation, and research in Artificial Intelligence and Data Science to fulfil industrial demands and societal expectations.

Mission

- To educate future engineers with solid fundamentals, continually improving teaching methods using modern tools.
- To collaborate with industry and offer top-notch facilities in a conducive learning environment.
- To foster skilled engineers and ethical innovation in AI and Data Science for global recognition and impactful research.
- To tackle the societal challenge of producing capable professionals by instilling employability skills and human values.

PROGRAM EDUCATIONAL OBJECTIVES (PEO)

- **PEO1:** Compete on a global scale for a professional career in Artificial Intelligence and Data Science.
- **PEO2:** Provide industry-specific solutions for the society with effective communication and ethics.
- **PEO3** Enhance their professional skills through research and lifelong learning initiatives.

PROGRAM SPECIFIC OUTCOMES (PSOs)

- **PSO1:** Capable of finding the important factors in large datasets, simplify the data, and improve predictive model accuracy.
- **PSO2:** Capable of analyzing and providing a solution to a given real-world problem by designing an effective program.

PROGRAM OUTCOMES (POs)

Engineering students will be able to:

1. **Engineering knowledge:** Apply knowledge of mathematics, natural science, computing, engineering fundamentals, and an engineering specialization to develop solutions to complex engineering problems.

2. **Problem analysis:** Identify, formulate, review research literature and analyze complex engineering problems reaching substantiated conclusions with consideration for sustainable development.

3. **Design/development of solutions:** Design creative solutions for complex engineering problems and design/develop systems/components/processes to meet identified needs with consideration for the public health and safety, whole-life cost, net zero carbon, culture, society and environment as required.

4. **Conduct investigations of complex problems:** Conduct investigations of complex engineering problems using research-based knowledge including design of experiments, modelling, analysis & interpretation of data to provide valid conclusions.

5. **Engineering Tool Usage:** Create, select and apply appropriate techniques, resources and modern engineering & IT tools, including prediction and modelling recognizing their limitations to solve complex engineering problems.

6. **The Engineer and The World:** Analyze and evaluate societal and environmental aspects while solving complex engineering problems for its impact on sustainability with reference to economy, health, safety, legal framework, culture and environment.

7. **Ethics:** Apply ethical principles and commit to professional ethics, human values, diversity and inclusion; adhere to national & international laws.

8. **Individual and Collaborative Team work:** Function effectively as an individual, and as a member or leader in diverse/multi-disciplinary teams.

9. **Communication:** Communicate effectively and inclusively within the engineering community and society at large, such as being able to comprehend and write effective reports and design documentation, make effective presentations considering cultural, language, and learning differences.

10. **Project management and finance:** Apply knowledge and understanding of engineering management principles and economic decision-making and apply these to one's own work, as a member and leader in a team, and to manage projects and in multidisciplinary environments.

11. **Life-long learning:** Recognize the need for, and have the preparation and ability for i) independent and life-long learning ii) adaptability to new and emerging technologies and iii) critical thinking in the broadest context of technological change.

ABSTRACT

This project presents a web-based Student Attendance Management System developed using R and the Shiny framework, aimed at streamlining and automating the process of attendance management in educational institutions. The application supports the complete data pipeline—from collection to visualization—enabling users to upload raw CSV files and transform them into clean, standardized, and analyzable datasets. This reduces the manual workload associated with tracking attendance and enhances data accuracy. The system is built with a modular architecture comprising five key components: data collection, data cleaning, data standardization, analysis and reporting, and visualization. By leveraging robust R libraries such as dplyr, ggplot2, plotly, and DT, the application ensures real-time data processing and interactive presentation of attendance records. It empowers teachers and administrators to derive insights through summary tables and dynamic graphs, enabling data-driven decision-making. This solution is particularly beneficial for institutions seeking a lightweight yet powerful tool to improve operational efficiency and academic oversight through digital transformation.

ABSTRACT WITH POs AND PSOs MAPPING

CO 5 : BUILD DATA SCIENCE USING R PROGRAMMING FOR SOLVING REAL-TIME PROBLEMS.

ABSTRACT	POs MAPPED	PSOs MAPPED
This project introduces a web-based Student Attendance Management System using R and Shiny. It automates CSV-based attendance data processing through modular components like data collection, cleaning, analysis, and visualization. Leveraging R libraries such as dplyr, ggplot2, plotly, and DT, the system offers a user-friendly interface to help educators efficiently manage and analyze attendance records.	**PO1 -3** **PO2 -3** **PO3 -3** **PO4 -3** **PO5 -3** **PO6 -3** **PO7 -3** **PO8 -3** **PO9 -3** **PO10 -3** **PO11-3**	**PSO1 -3** **PSO2 -3**

Note: 1- Low, 2-Medium, 3- High

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

1.1 INTRODUCTION

The Student Attendance Management System using R Shiny is developed to address the inefficiencies and limitations of traditional attendance tracking systems. Manual methods are often time-consuming, error-prone, and lack the ability to visualize trends or perform real-time analysis. This system offers an automated solution that simplifies data collection, cleaning, and reporting by leveraging powerful R packages and the Shiny web framework. It empowers educators to analyze attendance patterns efficiently, ensure data accuracy, and make data-driven decisions.

1.2 OBJECTIVE

The primary objective of this project is to design and develop an automated system that simplifies student attendance management using R Shiny. This system is intended to streamline the process of collecting, organizing, analyzing, and visualizing attendance data in a user-friendly, interactive manner. By leveraging modern data handling and visualization libraries, this tool ensures accurate tracking of attendance records and provides meaningful insights to educators. It seeks to reduce manual effort, avoid errors, and facilitate decision-making by delivering comprehensive summaries and visual representations of student participation across time and classes.

1.3 DATA SCIENCE RELATED CONCEPTS

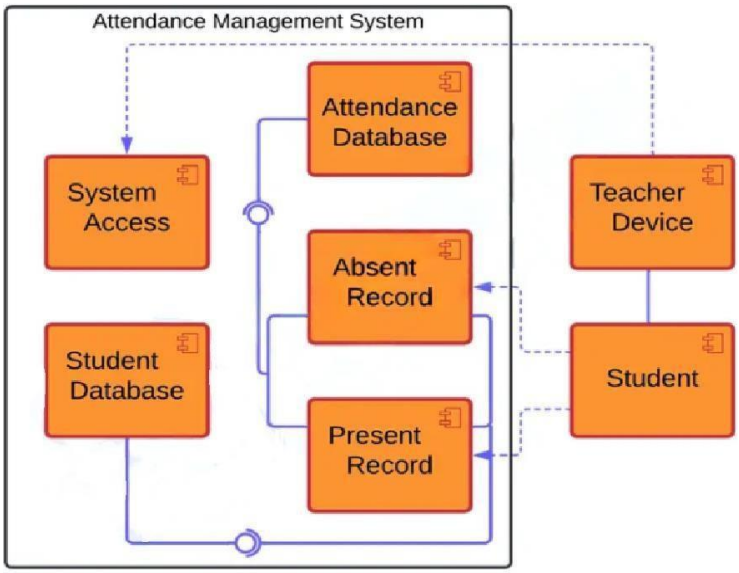
In this project, several core data science concepts were implemented to manage and analyze student attendance effectively. The system begins with **data collection**, where attendance data is imported from CSV files, simulating real-world data acquisition. This is followed by **data cleaning**, where missing values are handled, duplicates are removed, and inconsistent entries are corrected to ensure high data quality. The **data standardization** process transforms and formats fields like dates and attendance status for consistency. Using **exploratory data analysis (EDA)** techniques, the project summarizes trends by class and date through grouping and aggregation. These insights are visually communicated using **data visualization**, where ggplot2 and plotly generate interactive charts. The application also employs **reactive programming** through Shiny to dynamically update outputs based on user actions, enhancing interactivity. Collectively, these data science practices enable accurate tracking, meaningful analysis, and intuitive presentation of attendance data within the system.

CHAPTER 2
PROJECT METHODOLOGY

2.1 PROPOSED WORK

This project proposes the development of a modular, web-based student attendance management application using R and the Shiny framework. The goal is to create an interactive and automated solution for uploading, processing, analyzing, and visualizing student attendance data. The system is designed to be both user-friendly and highly functional, allowing teachers and administrators to upload CSV files, clean and standardize the input, perform group-wise and time-wise analysis, and generate insightful reports. It is built with five major modules—data collection, data cleaning, data standardization, analysis and reporting, and visualization—each contributing to a streamlined and effective data processing pipeline. These modules work together to ensure data integrity, analytical accuracy, and visual clarity, ultimately supporting better academic management through informed decision- making.

2.1 BLOCK DIAGRAM



CHAPTER 3
MODULE DESCRIPTION

3.1 DATA COLLECTION MODULE:

Function: Handles user input of CSV files. **Implementation:** Uses fileInput() in the Shiny UI to allow users to select and upload a CSV file. The server reads the file using read.csv() and validates it by checking for mandatory columns like Student_ID, Name, Class, Section, Date, and Attendance. If the format is incorrect, the app stops with an error message. This ensures data integrity right at the entry point.

3.2 DATA CLEANING MODULE:

Data Cleaning Module plays a crucial role in preparing the raw attendance data for reliable analysis. It is responsible for ensuring that the input data is free from inconsistencies, inaccuracies, and redundancies. Upon uploading the CSV file, this module identifies and fills missing values—typically by marking them as "Absent"—to prevent incomplete records. Duplicate rows are removed to avoid double counting in analysis, preserving data accuracy. Furthermore, it standardizes various non-uniform attendance entries such as "p", "pres", and "abs" into unified labels like "Present" and "Absent". These steps, implemented using R functions like tolower() and !duplicated(), are essential for maintaining a clean and uniform dataset, setting a solid foundation for standardization, analysis, and visualization in subsequent modules.

.

3.3 DATA STANDARDIZATION MODULE:

Converts all date strings to Date type using as.Date() to enable proper sorting and filtering by time.Normalizes text fields such as attendance status to ensure consistent casing and terminology.These steps ensure that the entire dataset is coherent and compatible with visualization and analysis tools used later in the pipeline.

3.4 DATA ANALYSIS AND REPORTING MODULE:

Uses dplyr::group_by() and summarise() to calculate totals per class, section, or date. Displays summaries in structured data tables with renderDT() from the DT package. Users can sort, filter, and explore the data interactively. This module supports identifying patterns such as frequently absent students, low-attendance classes, or peak absence dates. It acts as the backbone for strategic decision-making.

3.5 VISUALIZATION MODULE:

Uses ggplot2 to create visual representations like bar plots showing class-wise and daily attendance. Integrates with plotly using ggplotly() to make the graphs interactive—users can zoom, hover for details, and dynamically explore data. This module helps users instantly recognize anomalies, detect behavioral trends, and interpret attendance patterns with ease.

CHAPTER 4

CONCLUSION

The Student Attendance Management System achieves its goal of simplifying and enhancing the attendance tracking process through automation and interactivity. By leveraging R Shiny's reactive capabilities, the system provides a smooth user experience for uploading, cleaning, and analyzing student attendance records. The interface is intuitive and allows educators to generate insights without needing advanced technical skills. The modular structure makes the system scalable, maintainable, and extensible. Overall, the system serves as an effective educational analytics tool that supports both administrative operations and student performance evaluation.

FUTURE SCOPE:

- **Database Integration**: Replace CSV input with a SQL database to enable persistent storage, real-time access, and multi-user functionality.
- **Authentication System**: Add role-based logins (e.g., for teachers, students, admins) to secure and personalize the system.
- **Export Options**: Enable export of summaries and charts to PDF, Excel, or PNG for reporting and archival.
- **Real-time Data Sync**: Integrate with cloud-based tools like Google Sheets, APIs from biometric devices, or learning management systems (LMS) for automated updates.
- **Enhanced Dashboard**: Add filters by date range, student performance, and integrate notifications or email alerts for irregular attendance behaviour.

APPENDICES

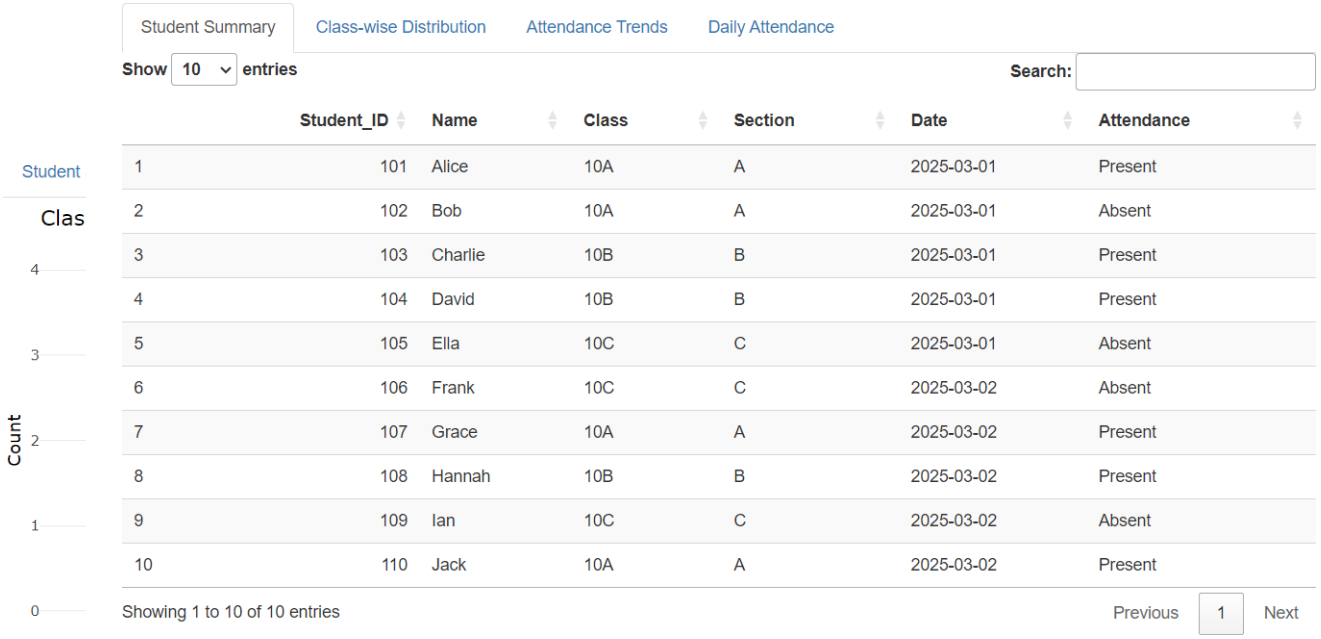
APPENDIX A – SOURCE CODE

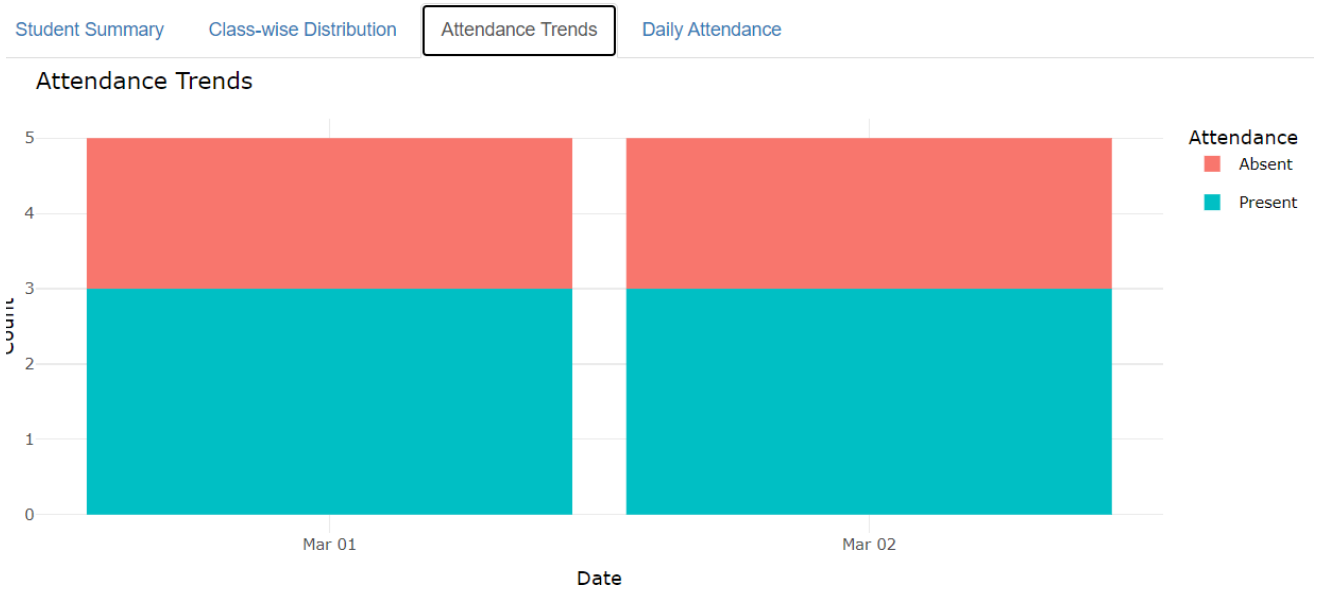
```
library(shiny)
library(dplyr)
library(ggplot2)
library(plotly)
library(DT)

# Data Collection Module – Imports attendance data from CSV files
ui <- fluidPage(
titlePanel("Interactive Student Records Management"),
sidebarLayout(
sidebarPanel(
fileInput("file", "Upload CSV File", accept = ".csv"),
helpText("Ensure the CSV file has 'Student_ID',
 'Name', 'Class', 'Section', 'Date', and 'Attendance' columns."),
actionButton("process", "Process Data")
),
mainPanel(
tabsetPanel(
tabPanel("Student Summary", DTOutput("data_summary")),
tabPanel("Class-wise Distribution", plotlyOutput("class_distribution")),
tabPanel("Attendance Trends", plotlyOutput("attendance_trend")),
tabPanel("Daily Attendance", plotlyOutput("daily_bar"))
)
)
)
)

server <- function(input, output) {

# Data Cleaning Module – Handles missing values, removes
 duplicates, and corrects inconsistencies
student_data <- reactive({
req(input$file)
df <- read.csv(input$file$datapath, stringsAsFactors = FALSE)

if (!all(c("Student_ID", "Name", "Class", "Section", "Date",
```

```
"Attendance") %in% colnames(df))) {
stop("Error: Missing required columns.")
}


df$Date <- as.Date(df$Date, format="%d/%m/%Y")
df$Attendance <- tolower(df$Attendance)
df$Attendance[df$Attendance %in% c("p", "pres", "present")] <- "Present"
df$Attendance[df$Attendance %in% c("a", "abs", "absent")] <- "Absent"
df[is.na(df)] <- "Absent"
df <- df[!duplicated(df), ]
df
})

# Data Standardization Module – Formats dates, normalizes
attendance status, and standardizes student names
output$data_summary <- renderDT({
req(input$process)
df <- student_data()
datatable(df)
})

# Data Analysis & Reporting Module – Summarizes attendance trends
output$class_distribution <- renderPlotly({
req(input$process)
df <- student_data()
summary_table <- df %>% group_by(Class) %>% summarise(Count = n())
plot <- ggplot(summary_table, aes(x = Class, y = Count, fill = Class)) +
geom_bar(stat = "identity") +
labs(title = "Class-wise Student Distribution", x = "Class", y = "Count") +
theme_minimal()
ggplotly(plot)
})

# Visualization Module – Uses ggplot2 to create graphs for attendance trends
output$attendance_trend <- renderPlotly({
req(input$process)
df <- student_data()
df$Attendance <- as.factor(df$Attendance)
plot <- ggplot(df, aes(x=Date, fill=Attendance)) +
geom_bar(position="stack") +
labs(title="Attendance Trends", x="Date", y="Count") +
theme_minimal()
ggplotly(plot)
```

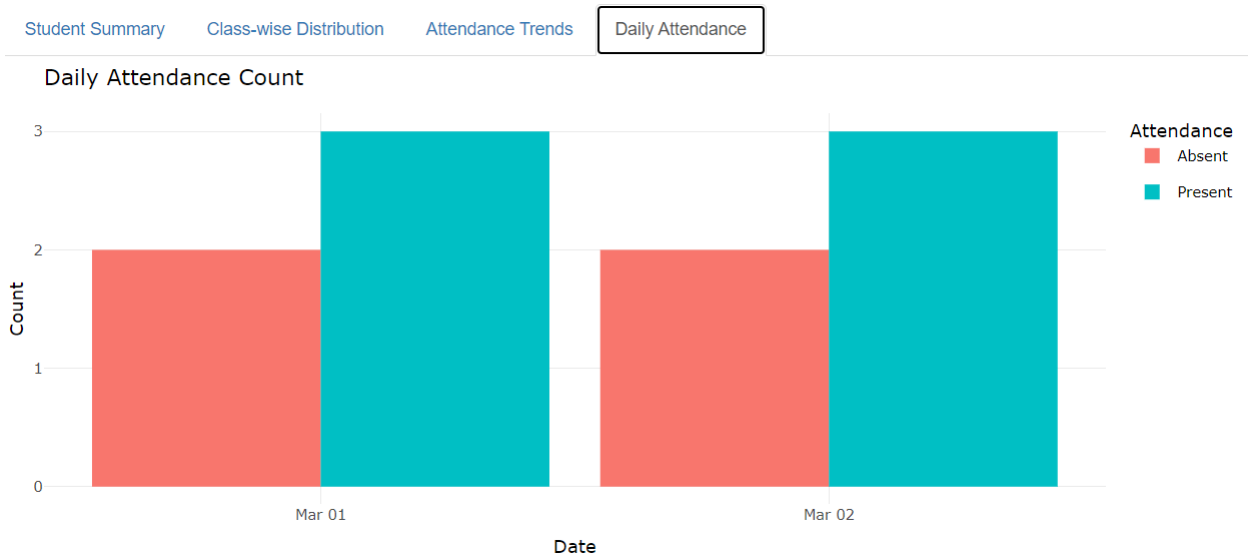
```
})

output$daily_bar <- renderPlotly({
req(input$process)
df <- student_data()
df$Attendance <- as.factor(df$Attendance)
plot <- ggplot(df, aes(x=Date, fill=Attendance)) +
geom_bar(position="dodge") +
labs(title="Daily Attendance Count", x="Date", y="Count") +
theme_minimal()
                ggplotly(plot)
                })
                }

                # Run the Shiny app
                shinyApp(ui = ui, server = server)
```

APPENDIX B – SCREENSHOTS




Student Summary
Class-wise Distribution
Attendance Trends
Daily Attendance
Attendance Trends
Count
5
4
3
2
1
0
Mar 01
Mar 02
Date
Attendance
Absent
Present

s


Student Summary
Class-wise Distribution
Attendance Trends
Daily Attendance
Daily Attendance Count
Count
3
2
1
0
Mar 01
Mar 02
Date
Attendance
Absent
Present

REFERENCES:

1. Chang, W., Cheng, J., Allaire, J., Xie, Y., & McPherson, J. (2022). Shiny: Web Application Framework for R. R package version 1.7.4. https://CRAN.R-project.org/package=shiny

2. Grolemund, G., & Wickham, H. (2017). R for Data Science. O'Reilly Media. https://r4ds.had.co.nzs

3. R Core Team. (2023). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. shttps://www.R-project.org/

4. Sievert, C. (2020). Interactive Web-Based Data Visualization with R, plotly, and shiny. CRC Press. https://plotly-r.com/

5. Wickham, H., Francois, R., Henry, L., & Muller, K. (2023). dplyr: A Grammar of Data Manipulation. R package version 1.1.2. https://CRAN.R-project.org/package=dplyr