

CHANDIGARH UNIVERSITY

Discover. Learn. Empower.

MINI PROJECT

Statistical Techniques using R

TITLE: Exploratory Data Analysis On Social Media Addiction

MCA (DATA SCIENCE)

Submitted by:

Chaitanya Sharma – 25MCD10056

Submitted to:

Amarjeet Arora Sir

Acknowledgement

I would like to express my sincere gratitude to all those who have contributed to the successful completion of this project. Their constant guidance, encouragement, and valuable support have played an instrumental role throughout this journey.

First and foremost, I extend my heartfelt appreciation to **Amarjeet Arora**, whose exceptional teaching and mentorship have been the foundation of my learning. Her dedication, patience, and ability to simplify complex concepts in R, and data visualization enabled me to acquire the knowledge and confidence required to create this project. Her guidance not only helped me technically but also inspired me to think critically, explore data with curiosity, and approach problems methodically.

I would also like to thank **the University of Computing** for providing the academic environment and resources necessary to carry out this work. The support and collaboration of my peers further enhanced my understanding and motivation throughout this project.

In conclusion, this project would not have been possible without the valuable guidance and inspiration of **Amarjeet Arora**. Her mentorship has been a cornerstone of my progress, and I remain deeply grateful for her invaluable contribution to my academic and professional development.

1. Abstract

This project, *Social Media Addiction Analysis*, aims to explore and visualize patterns in students' social media usage using statistical and graphical methods in R. By examining parameters such as time spent online, academic performance, and psychological indicators, the study seeks to identify trends that may point toward addictive behavior.

The data underwent extensive preprocessing and exploratory data analysis (EDA) to uncover correlations, distributions, and outliers. Visualizations such as histograms, boxplots, and scatter plots were employed to highlight behavioral patterns and relationships between variables.

The project ultimately demonstrates how data analytics and visualization can be used to gain meaningful insights into social media habits among students, offering a foundation for further research into digital well-being and academic impact.

2. Introduction

In the digital era, social media has become deeply intertwined with students' daily routines. While it facilitates communication and learning, excessive use can lead to addictive behavior and negatively affect academic and mental well-being.

This project focuses on exploring patterns of social media addiction among students using an interactive dashboard built in R Shiny. The goal was to visualize and analyze relationships among behavioral, academic, and psychological factors, and to uncover patterns in how social media usage correlates with other lifestyle attributes such as sleep hours, academic performance, and mental health.

3. Objectives and Scope

The primary objective of this project is to explore and analyze students' social media usage patterns using R's data analysis and visualization capabilities. The project demonstrates how statistical tools and interactive dashboards can be used to understand behavioral trends and correlations related to social media addiction among students.

The scope of this work includes developing an interactive Shiny application that enables users to visualize data dynamically through histograms, correlation heatmaps, and summary statistics. It focuses on identifying key numerical variables that influence social media addiction levels and presenting them in an accessible, visually appealing format. The project highlights how R, when combined with libraries like ggplot2, dplyr, and plotly, can serve as a powerful environment for educational and psychological data exploration.

4. Tools and Technologies

This project was developed using R and the Shiny framework, along with several key packages for visualization and data manipulation.

Software & Libraries:

- R (v4.5) — Core programming environment.
- Shiny — Used for building the interactive web-based dashboard.
- ggplot2 — For creating elegant and layered visualizations.
- plotly — To make the plots interactive (zoom, hover, pan).
- dplyr — For efficient data manipulation and filtering.
- readr — For importing and handling CSV data.
- corrplot — For visualizing correlation matrices.
- psych — For descriptive statistics and psychological variable exploration.

5. Dataset Overview

The dataset contains 13 attributes describing demographic, behavioral, and psychological traits of students. Below is a brief summary of each column:

Attributes:

Student_ID – Unique identifier for each student.

Age – Age of the student.

Gender – Gender of the student.

Academic_Level – Level of study (e.g., Undergraduate, Postgraduate).

Country – Country of the student.

Avg_Daily_Usage_Hours – Average daily time spent on social media.

Most_Used_Platform – The social media platform most frequently used.

Affects_Academic_Performance – Indicates if academic performance is affected.

Sleep_Hours_Per_Night – Average sleep duration per night.

Mental_Health_Score – Self-assessed mental well-being score.

Relationship_Status – Relationship status of the student.

Conflicts_Over_Social_Media – Reports of conflicts arising due to social media use.

Addicted_Score – Quantified score representing addiction level.

6. Exploratory Data Analysis (EDA)

6.1 Histograms

Users can dynamically select any numeric variable (e.g., Age, Avg_Daily_Usage_Hours, Addicted_Score) from a dropdown menu. The histogram then updates interactively to visualize its distribution.

Observations:

- Most students reported moderate to high daily usage hours.
- A negative skew was observed in Sleep_Hours_Per_Night, indicating shorter sleep among heavy users.
- Addicted_Score showed a right-skewed distribution, suggesting that a significant portion of students fall into higher addiction ranges.

6.2 Correlation Heatmap

The second module visualizes the correlation matrix among all numeric variables using the corrplot library.

Findings:

- Strong positive correlation between Avg_Daily_Usage_Hours and Addicted_Score.
- Negative correlation between Sleep_Hours_Per_Night and Addicted_Score.
- Mental_Health_Score showed moderate negative correlation with Addicted_Score.

7. Implementation

```
# install.packages(c("ggplot2", "dplyr", "readr", "psych", "plotly", "shiny", "corrplot"))

library(shiny)
library(dplyr)
library(plotly)
library(tidyverse)
library(psych)
library(readr)
library(corrplot)

data <- read_csv("Students Social Media Addiction.csv")
head(data)
str(data)
summary(data)
summary(data)
colSums(is.na(data))
```

```

# Select only numeric columns
num_data <- data %>%
  dplyr::select(where(is.numeric)) %>%
  dplyr::select(-Student_ID)

# Define UI
ui <- fluidPage(
  titlePanel("📊 Social Media Addiction - EDA Dashboard"),
  sidebarLayout(
    sidebarPanel(
      h4("Histogram Explorer"),
      selectInput("num_var", "Choose a numeric variable:", choices = names(num_data)),
      br(),
      h4("Correlation Matrix"),
      helpText("Displays relationships among all numeric variables")),
    mainPanel(
      tabsetPanel(
        tabPanel("Histogram", plotlyOutput("histPlot")),
        tabPanel("Correlation Heatmap", plotOutput("corrPlot"))))))
)

# Define server logic
server <- function(input, output) {
  # Histogram (interactive)
  output$histPlot <- renderPlotly({
    p <- ggplot(num_data, aes_string(x = input$num_var)) + geom_histogram(binwidth = 1, fill =
      "#69b3a2", color = "white") + theme_minimal(base_size = 14) +
      labs(
        title = paste("Distribution of", input$num_var),
        x = input$num_var,
        y = "Count")
    ggplotly(p)}))
}

```

```

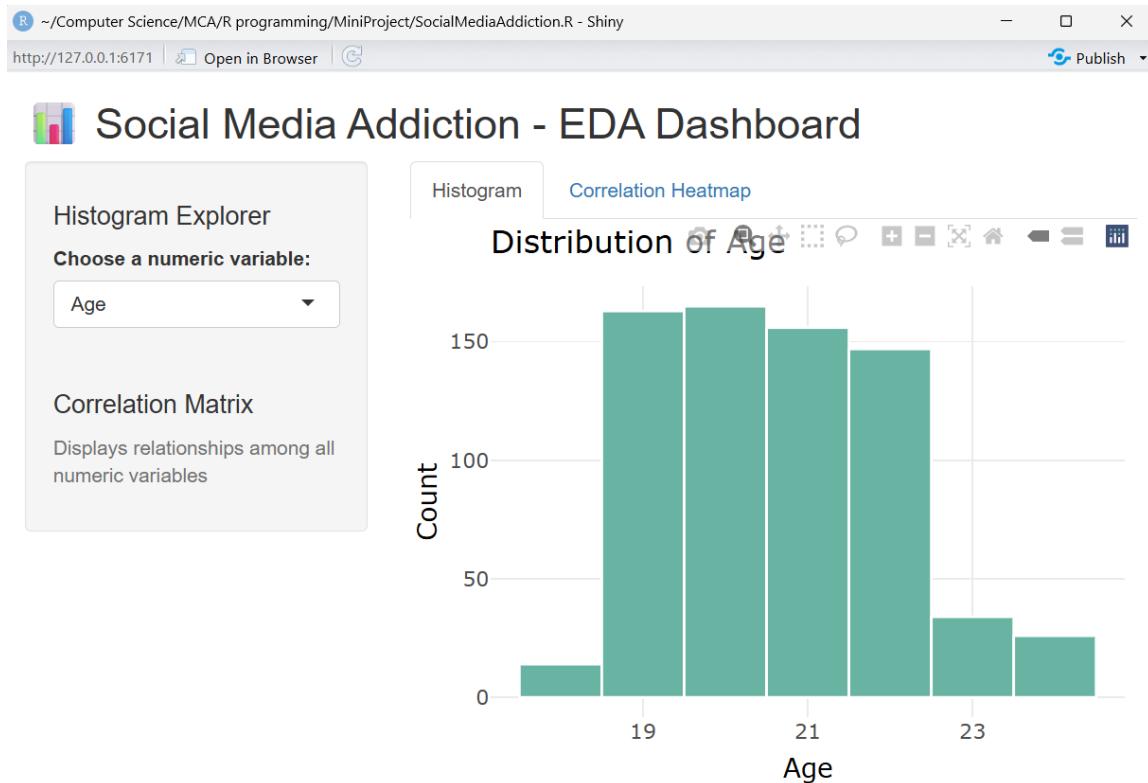
# Correlation Heatmap

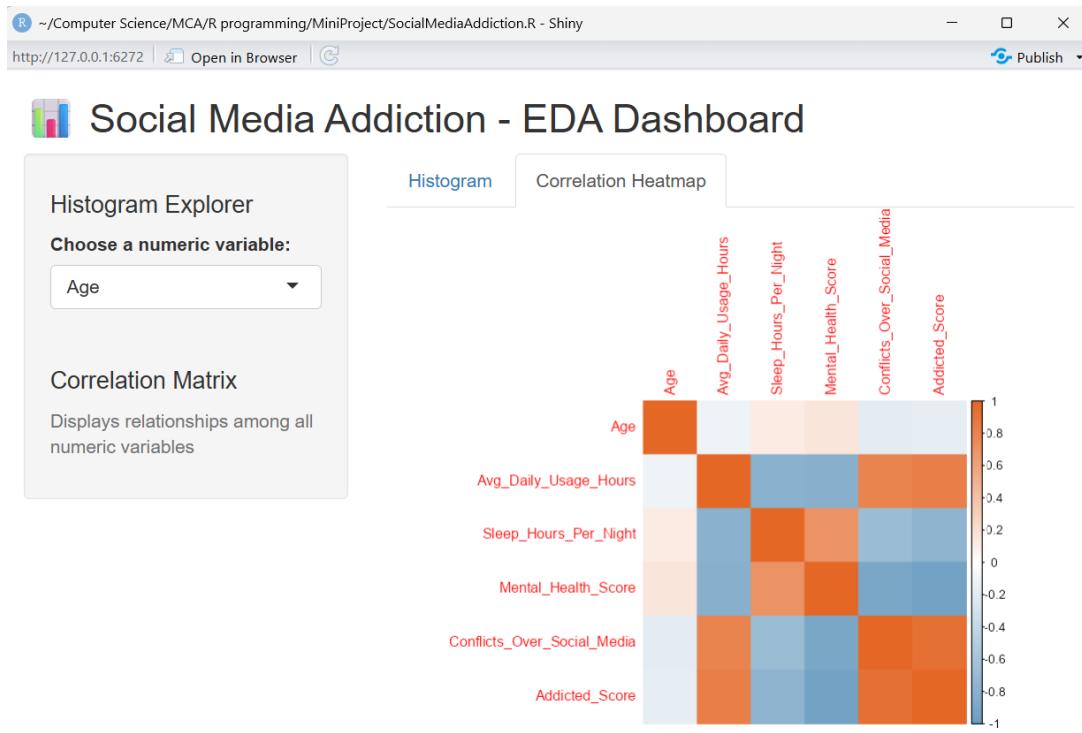
output$corrPlot <- renderPlot({
  cor_matrix <- cor(num_data, use = "complete.obs")
  corrplot(cor_matrix, method = "color", tl.cex = 0.9, number.cex = 0.7,
    col = colorRampPalette(c("#6D9EC1", "white", "#E46726"))(200))
})

# Run the application

shinyApp(ui = ui, server = server)

```





8. Results and Discussion

Key Insights:

- Students spending more than 4–5 hours daily on social media exhibit higher addiction scores.
- Reduced sleep hours are associated with increased social media usage.
- Mental health scores decline as addiction scores rise.
- Academic performance appears indirectly affected, correlating with behavioral fatigue and reduced focus.

9. Conclusion and Future Work

This project successfully demonstrates how data visualization and interactive analytics can uncover behavioral patterns within psychological datasets. The R Shiny dashboard provides a powerful, intuitive interface for exploring correlations and distributions.

Key takeaways:

- Excessive social media use correlates strongly with higher addiction scores and poorer sleep.
- Mental health and academic outcomes appear negatively influenced by overuse.
- The dashboard offers educators and researchers an adaptable platform for future exploration.

Future improvements could include integrating machine learning models, adding categorical analysis, and deploying the app online.

9.1. GitHub Link: <https://github.com/Chai5K5/EDA-using-R-on-Social-Media-Addiction-Data>

10. Bibliography:

- Lin, L. Y., Sidani, J. E., Shensa, A., Radovic, A., Miller, E., Colditz, J. B., Hoffman, B. L., Giles, L. M., & Primack, B. A. (2016). Association between social media use and depression among U.S. young adults. **Depression and Anxiety**, 33(4), 323–331. <https://doi.org/10.1002/da.22466>
- Kuss, D. J., & Griffiths, M. D. (2017). *Social networking sites and addiction: Ten lessons learned. **International Journal of Environmental Research and Public Health**, 14(3), 311. <https://doi.org/10.3390/ijerph14030311>
- Wickham, H. (2016). **ggplot2: Elegant Graphics for Data Analysis** Springer-Verlag.
- Chang, W., Cheng, J., Allaire, J. J., Sievert, C., Schloerke, B., Xie, Y., & Allen, J. (2024). **Shiny: Web Application Framework for R**. R package version 1.9.0. <https://shiny.rstudio.com/>
- Revelle, W. (2024). **psych**: Procedures for Psychological, Psychometric, and Personality Research. R package version 2.4.0. Northwestern University, Evanston, Illinois.
- Friendly, M. (2002). **Corrplot: Visualization of Correlation Matrices**. The R Journal..