Mental Health Data Analysis in R: A Comprehensive Guide to Insights in the Tech Industry

DATA 230 Final Project

Dr. Mehdi Ghayoumi

Eliza Ochoa 12/7/24 White Paper

ABSTRACT

This white paper presents an in-depth analysis of mental health trends within the tech industry using R programming. The study involves data cleaning, exploratory data analysis (EDA), and predictive modeling to uncover actionable insights. Findings emphasize the role of factors like age, gender, and family history in shaping treatment-seeking behaviors, providing foundation for data-driven policy recommendations.

Table of Contents

1. Introduction 1
2. Methodology 1
   2.1 Data Cleaning..1
   2.2 Data Cleaning..1
   2.3 Data Transformation..1
3. Exploratory Data Analysis (EDA) 2
   3.1 Descriptive Statistics 2
   3.2 Visualizations 2
4. Modeling and Analysis 5
   4.1 Logistic Regression..5
   4.2 Practical Implications 5
5. Discussions and Recommendations 6
6. Conclusions..6
7. References 6

Figure 1. Proportion of Treatment by Age Group..2
Figure 2. Age Distribution (Bar Chart) 3
Figure 3. Proportion of Treatment by Gender 3
Figure 4. Box Plot of Age by Age Group 4
Figure 5. Age Distribution (Histogram)..4
Figure 6. Scatter Plot of Age Group vs. Age 5

1. Introduction

The tech industry is renowned for its high-pressure environments, long work hours, and intense workloads. These factors contribute to widespread mental health challenges among employees. Despite growing awareness, many workers still face barriers to seeking treatment, stemming from stigma and insufficient workplace support.

This study uses a dataset from Kaggle's "Mental Health in Tech Survey" to analyze treatment-seeking behaviors in the tech industry. The study's objectives include identifying key predictors of mental health treatment and providing data-driven recommendations for organizations.

2. Methodology

2.1 Data Cleaning

The dataset, sourced from Kaggle's "Mental Health in Tech Survey," includes various demographic and workplace factors, such as age, gender, treatment-seeking behaviors, and employer-provided mental health benefits. It consists of 1,251 responses and 27 variables.

2.2 Data Cleaning

The dataset underwent rigorous cleaning to manage missing values, duplicates, and inconsistencies. For instance, gender values like 'M' and 'Male' were standardized. The following R code demonstrates the cleaning process:

```
# Load the dataset
survey_data <- read.csv('survey.csv')

# Remove missing values
cleaned_data <- na.omit(survey_data)

# Standardized gender variable
cleaned_data$Gender <- tolower(cleaned_data$Gender)
```

2.3 Data Transformation

To enhance the dataset's utility, age was categorized into groups, and binary treatment values were recoded for analysis. The R code snippet below demonstrates these transformations:

```
# Create age groups

cleaned_data$AgeGroup <- cut(cleaned_data$Age, breaks = c(0, 18, 35, 50, 65, Inf),

                Labels = c('0-18', '19-35', '36-50', '51-65', '65+'))
```

Recode treatment variable

cleaned_data$Treatment <- ifelse(cleaned_data$Treatment == 'Yes,' 1, 0)

3. Exploratory Data Analysis (EDA)

3.1 Descriptive Statistics

Descriptive statistics summarized key variables. For numerical variables like age, mean, median, and variance were calculated. Categorical variables such as gender and treatment were represented using frequency tables.

3.2 Visualizations

The following charts highlight key trends in the dataset:

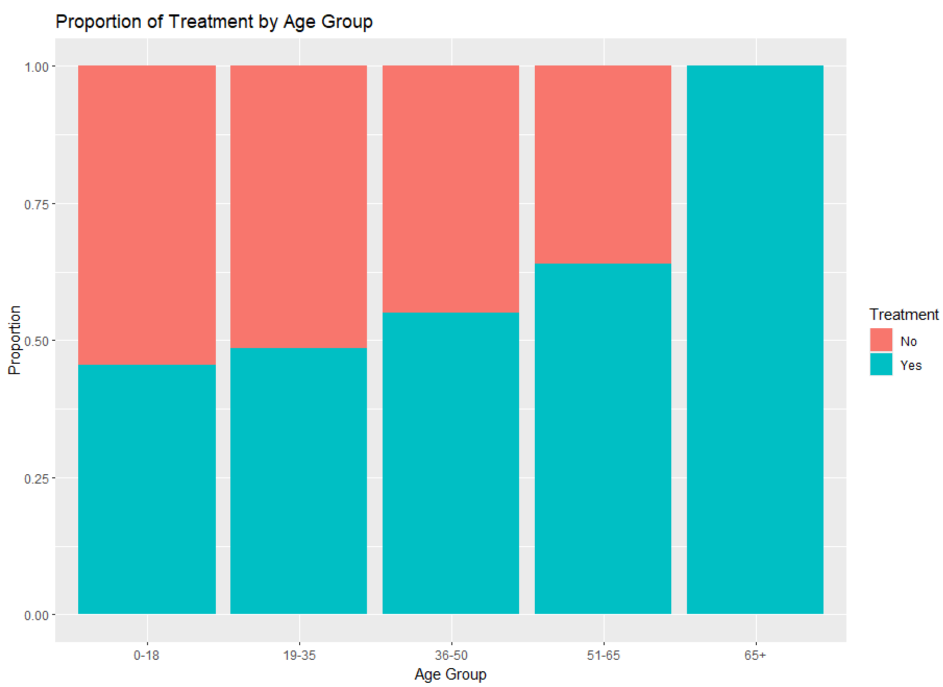

Figure 1. Proportion of Treatment by Age Group

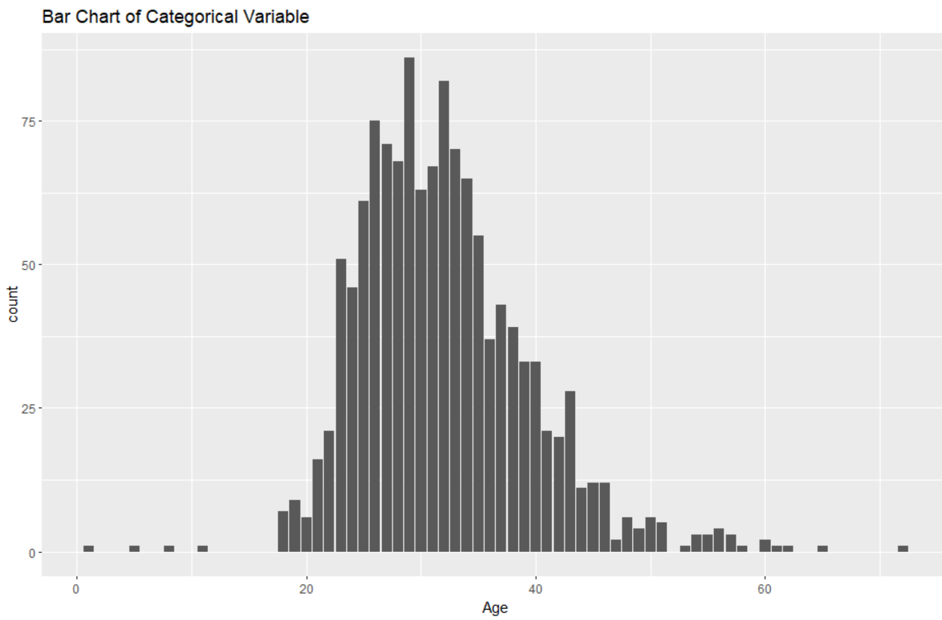

Figure 2. Age Distribution (Bar Chart)

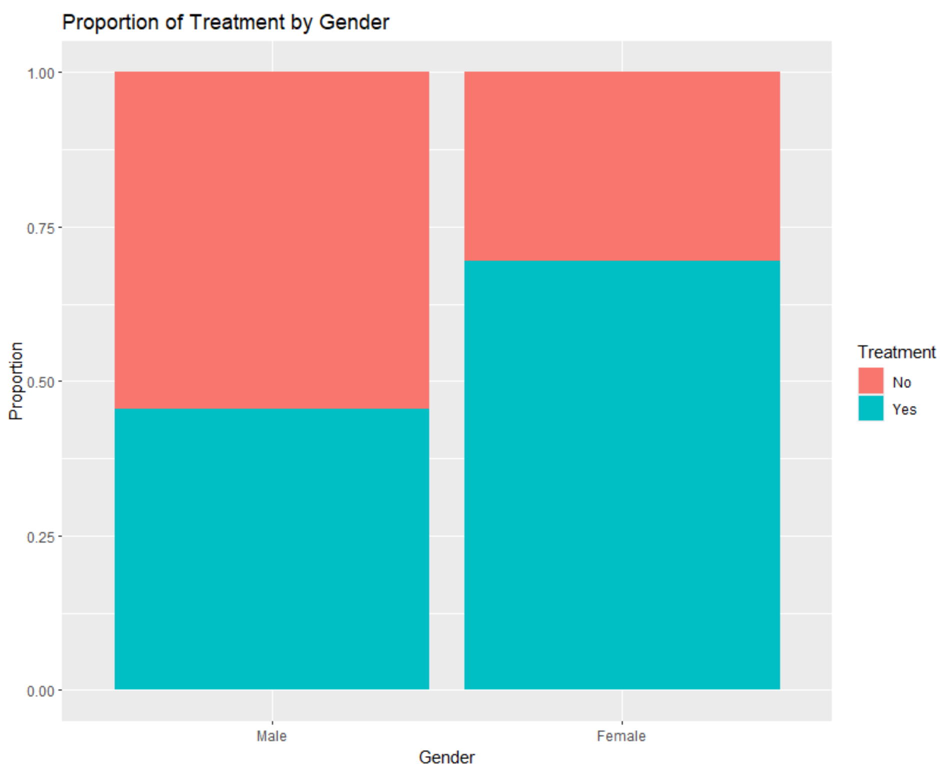

Figure 3. Proportion of Treatment by Gender

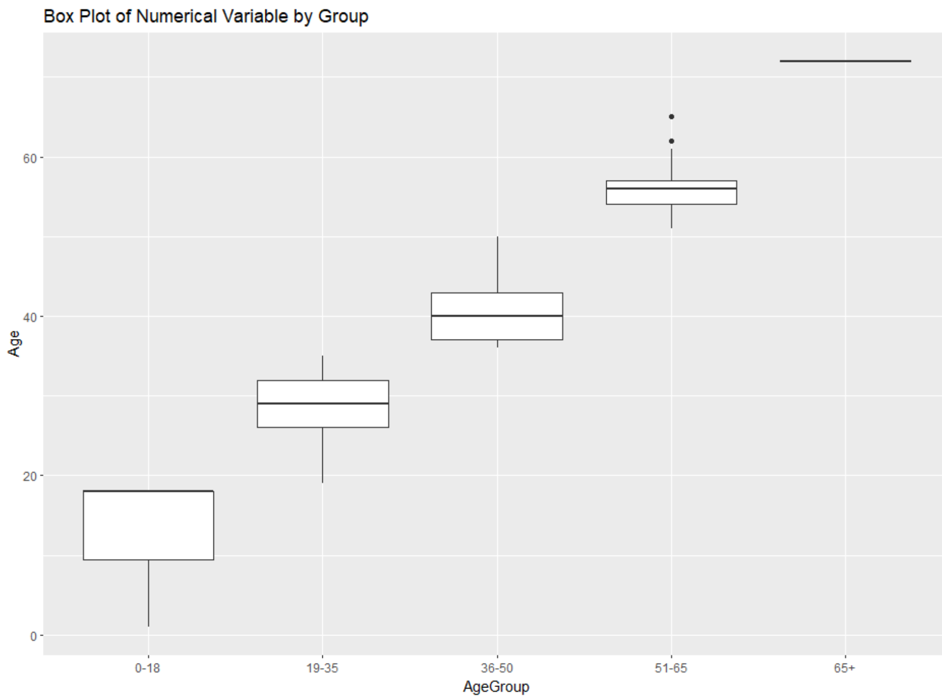

Figure 4. Box Plot of Age by Age Group

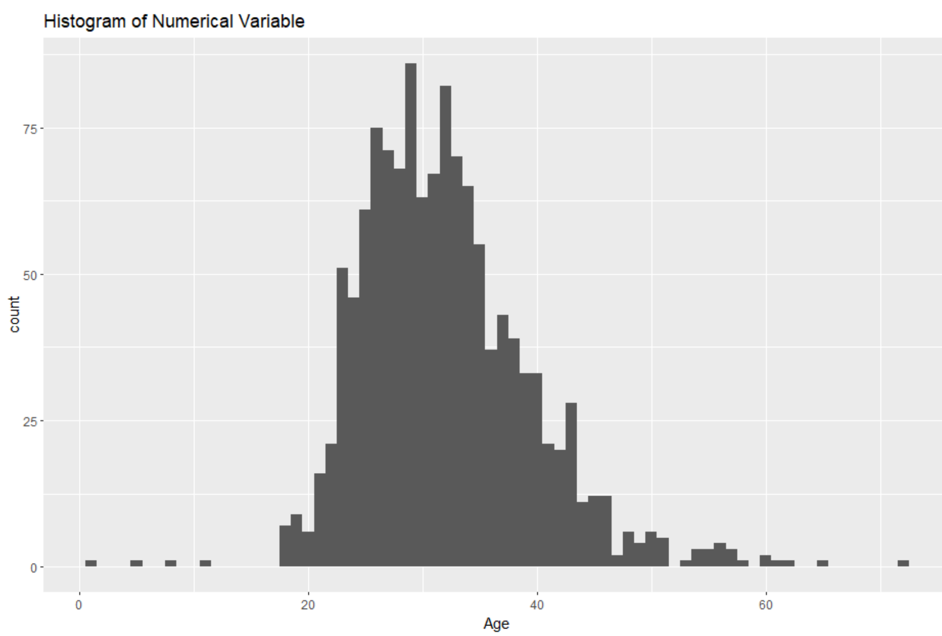

Figure 5. Age Distribution (Histogram)

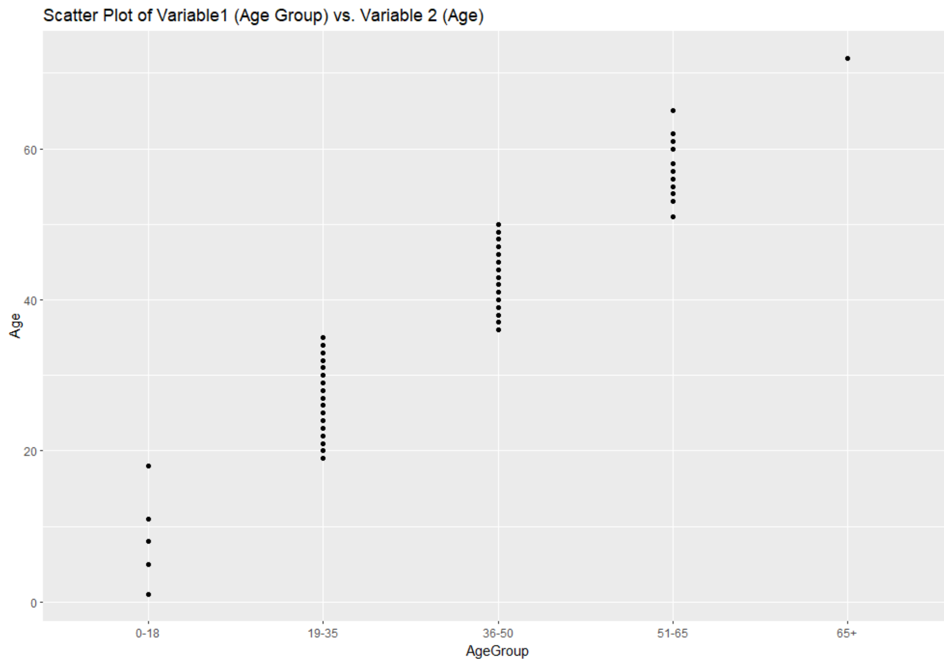

Figure 6. Scatter Plot of Age Group vs. Age

4. Modeling and Analysis

4.1 Logistic Regression

A logistic regression model was used to identify predictors of treatment-seeking behavior. Significant predictors included:
-Age Group: Younger individuals were more likely to seek treatment.
-Gender: Females exhibited slightly higher treatment-seeking behavior.
-Employer-Provided Benefits: Availability of benefits strongly correlated with seeking treatment.

Below is the R code for fitting the model:

```
# Fit logistic regression model
model <- glm(Treatment ~ AgeGroup + Gender + Benefits, data = cleaned_data, family = binomial)
summary(model)
```

4.2 Practical Implications

The findings underscore the need for workplace policies promoting mental health awareness and accessibility. Organizations should prioritize tailored mental health programs and benefits to encourage employees to seek treatment.

5. Discussions and Recommendations

This study highlights key barriers to mental health treatment in the tech industry and proposes actionable strategies:

Workplace Interventions:

Employers should offer comprehensive mental health benefits.

Targeted Campaign:

Age specific and gender-sensitive initiatives can reduce stigma.

Continuous Monitoring:

Regular surveys can help organizations adapt their policies to evolving needs.

6. Conclusions

Mental health remains a critical concern in the tech industry. This analysis identifies predictors of treatment-seeking behavior and provides actionable recommendations for organizations. Implementing these strategies can foster a healthier, more productive workforce.

7. References

1. Open Sourcing Mental Illness, L. and Kim, E. (2016). Mental Health in Tech Survey, Kaggle.
2. Bhandari, P. (2023). Missing data: Types, explanation, & imputation, Scribbr.
3. Hadley Wickham (2023). R for Data Science. 2nd Edition.
4. Murthy, V. (2023, April 27). Work in America survey: Psychological health and well-being. American Psychological Association (APA).
5. Velasco, A. A., Cruz, I. S. S., et al. (2020). Barriers and interventions targeting mental health problems. BMC Psychiatry.