```r
# Load required packages
install.packages(c("dplyr", "ggplot2", "cluster", "factoextra"))
library(dplyr)
library(ggplot2)
library(cluster)
library(factoextra)
# Set seed for reproducibility
 set.seed(123)
#import Data
data1<-survey_for_use_of_social_media_and_its_impacts
#data clean
New_data<-survey_for_use_of_social_media_and_its_impacts[,-
c(1,3,6,12,13,14,15,17,18,19,20,24,27)]
New_data
#convert categorical variable to factors
New_data$`2. which age group belongs to you?` <- as.factor(New_data$`2. which age group
belongs to you?` )
New_data$`3. what is your designation?` <- as.factor(New_data$`3. what is your
designation?`)
New_data$`4. what is your region?`<- as.factor(New_data$`4. what is your region?`)
#New_data$`10. Time spent daily on Social media networks:` <- as.factor(New_data$`10. Time
spent daily on Social media networks:`)
#New_data$`14. Are your parents aware about your social media activities?` <-
as.factor(New_data$`14. Are your parents aware about your social media activities?`)
#New_data$`19. Do you aware about privacy policy of Social Media?` <-
as.factor(New_data$`19. Do you aware about privacy policy of Social Media?`)
#New_data$`20. Do you  think  social media has negative impact?` <-
as.factor(New_data$`20. Do you  think  social media has negative impact?`)
#New_data$`22. Level of impact of social media on well - being` <- as.factor(New_data$`22.
Level of impact of social media on well - being`)
#New_data$`24. Do you think social media affect on mental health?` <-
as.factor(New_data$`24. Do you think social media affect on mental health?`)
# Create dummy variables for check box questions
New_data <- New_data %>%
 mutate(Instagram = ifelse(grepl("Instagram",New_data$`6. Which social media network/s do
you use?`), 1,0),
        Facebook = ifelse(grepl("Facebook",New_data$`6. Which social media network/s do
you use?`), 1, 0),
        Snapchat = ifelse(grepl("Snapchat",New_data$`6. Which social media network/s do
you use?`), 1, 0),
        Whatsapp = ifelse(grepl("Whatsapp",New_data$`6. Which social media network/s do
you use?`), 1,0),
        Linkedin = ifelse(grepl("Linkedin",New_data$`6. Which social media network/s do
you use?`), 1,0),
        Twitter = ifelse(grepl("Twitter",New_data$`6. Which social media network/s do you
use?`), 1,0),
        Google = ifelse(grepl("Google+",New_data$`6. Which social media network/s do you
use?`), 1, 0),
        Greeting = ifelse(grepl("Greeting Messages",New_data$`7. What is  the purpose of
using social media networks?`), 1, 0),
        Jokes = ifelse(grepl("Sharing Jokes",New_data$`7. What is  the purpose of using
social media networks?`), 1,0),
        Sharing = ifelse(grepl("Sharing Photos & Videos",New_data$`7. What is  the purpose
of using social media networks?`), 1,0),
        Bussiness = ifelse(grepl("Bussiness activities",New_data$`7. What is  the purpose
of using social media networks?`),1,0),
        Update_trends = ifelse(grepl("to keep update with news/trends",New_data$`7. What
is  the purpose of using social media networks?`),1,0),
        Information = ifelse(grepl("Sharing useful information",New_data$`7. What is  the
purpose of using social media networks?`),1,0),
        Shoping = ifelse(grepl("online shoping",New_data$`7. What is  the purpose of using
social media networks?`),1,0),
        Awareness = ifelse(grepl("Social awareness",New_data$`7. What is  the purpose of
using social media networks?`),1,0),
```

```r
        Professinal_network = ifelse(grepl("Build Professional network",New_data$`15.
Benefits of using Social Media Networks`),1,0),
        Connect = ifelse(grepl("Staying connect with friends",New_data$`15. Benefits of
using Social Media Networks`),1,0),
        Educate_yourself = ifelse(grepl("Educate yourself",New_data$`15. Benefits of using
Social Media Networks`),1,0),
        Usefull_information = ifelse(grepl("Useful Information",New_data$`15. Benefits of
using Social Media Networks`),1,0),
        New_skill = ifelse(grepl("Learn new skill",New_data$`15. Benefits of using Social
Media Networks`),1,0),
        Fun_entetainment = ifelse(grepl("Entertainment and fun",New_data$`15. Benefits of
using Social Media Networks`),1,0),
        Job_opportunities = ifelse(grepl("In quest of job opportunities",New_data$`15.
Benefits of using Social Media Networks`),1,0),
        cyber_theft = ifelse(grepl("Cyber theft",New_data$`21. Negative impact of social
media`),1,0),
        Health = ifelse(grepl("Health issue",New_data$`21. Negative impact of social
media`),1,0),
        Time_waste = ifelse(grepl("Time waste",New_data$`21. Negative impact of social
media`),1,0),
        Text_abuse = ifelse(grepl("Wrong Text Abuse",New_data$`21. Negative impact of
social media`),1,0),
        Culture_issue = ifelse(grepl("Culture issue",New_data$`21. Negative impact of
social media`),1,0),
        Happiness = ifelse(grepl("Happiness",New_data$`25. what emotions do you experience
when using social media?`),1,0),
        Motivation = ifelse(grepl("Motivation",New_data$`25. what emotions do you
experience when using social media?`),1,0),
        Inspiration = ifelse(grepl("inspiration",New_data$`25. what emotions do you
experience when using social media?`),1,0),
        Sense_belonging = ifelse(grepl("sense of belonging",New_data$`25. what emotions do
you experience when using social media?`),1,0),
        Rejection = ifelse(grepl("rejection",New_data$`25. what emotions do you experience
when using social media?`),1,0),
        self_esteem = ifelse(grepl("Boost self esteem",New_data$`25. what emotions do you
experience when using social media?`),1,0),
        Lower_esteem = ifelse(grepl("Lower of esteem",New_data$`25. what emotions do you
experience when using social media?`),1,0),
        left_alone = ifelse(grepl("No fear of left alone",New_data$`25. what emotions do
you experience when using social media?`),1,0)
        )
#convert hour per day to numeric data
New_data$`20. Do you  think  social media has negative impact?`<-
as.numeric(recode(New_data$`20. Do you  think  social media has negative impact?`,
                                                                       "Yes"
= "1",
                                                                       "No" =
"2"))
New_data$`22. Level of impact of social media on well - being`<-
as.numeric(recode(New_data$`22. Level of impact of social media on well - being`,
                                                                       "High
level" = "1",
                                                                       "Medium
level" = "2",
                                                                       "Low
level" = "3"))
New_data$`24. Do you think social media affect on mental health?`<-
as.numeric(recode(New_data$`24. Do you think social media affect on mental health?`,

"Yes" = "1",
                                                                          "No"
= "2",

"Maybe" = "3"))
New_data$`9. How long you have been active on social media networks?`<-
```

```r
as.numeric(recode(New_data$`9. How long you have been active on social media networks?`,

"Less than 6 month" = "1",

"1 year" = "2",

"2 year" = "3",

"More than 2 year" = "4"))
# Prepare data for clustering (select relevant columns)
Clustering_data1 <- New_data %>%
  select(Instagram,Facebook,Snapchat,Whatsapp,Linkedin,Twitter,Google,Greeting,Jokes,

Sharing,Bussiness,Update_trends,Information,Shoping,Awareness,Professinal_network,Connect,
        Educate_yourself,Usefull_information,New_skill,Fun_entetainment,
        Job_opportunities,cyber_theft,Health,Time_waste,Culture_issue,Text_abuse,

Happiness,Motivation,Inspiration,,Sense_belonging,Rejection,self_esteem,Lower_esteem,
        left_alone)
na.omit(Clustering_data1 )




# Scale the data
clustering_data_scaled <- scale(Clustering_data1)
clustering_data_scaled
distinct_points <- nrow(unique(clustering_data_scaled))
print(distinct_points)
# Check for NA values
na_check <- sum(is.na(clustering_data_scaled))
print(paste("Number of NA values:", na_check))

# Check for NaN values
nan_check <- sum(is.nan(clustering_data_scaled))
print(paste("Number of NaN values:", nan_check))

# Check for Inf values
#inf_check <- sum(is.infinite(clustering_data_scaled))
#print(paste("Number of Inf values:", inf_check))

clustering_data_clean <- clustering_data_scaled[complete.cases(clustering_data_scaled), ]
clustering_data_scaled[is.na(clustering_data_scaled)] <- mean(clustering_data_scaled,
na.rm = TRUE)
clustering_data_scaled[is.nan(clustering_data_scaled)] <- mean(clustering_data_scaled,
na.rm = TRUE)
#clustering_data_scaled[is.infinite(clustering_data_scaled)] <-
mean(clustering_data_scaled[!is.infinite(clustering_data_scaled)])
#kmeans_result <- kmeans(clustering_data_clean, centers = 5, nstart = 80)
distinct_points <- nrow(unique(clustering_data_clean))
print(distinct_points)
df<-as.data.frame(clustering_data_clean)




# Perform K-means clustering
set.seed(123)
kmeans_result <- kmeans(df, centers = 5, nstart = 80)  # Change the number of clusters as
needed
# Step 1: Identify constant columns
constant_columns <- sapply(df, function(x) length(unique(x)) == 1)
print(df[, constant_columns, drop = FALSE])

# Step 2: Remove constant columns
clustering_data_clean <- df[, !constant_columns, drop = FALSE]
```

```r
# Step 3: Run k-means clustering
kmeans_result <- kmeans(clustering_data_clean, centers = 5, nstart = 25)

# Step 4: Visualize clusters
library(factoextra)
fviz_cluster(kmeans_result, data = clustering_data_clean)
# Perform PCA
# Perform PCA
pca_result <- prcomp(clustering_data_clean, scale. = TRUE)

# Perform PCA
pca_result <- prcomp(clustering_data_clean, scale. = TRUE)

# Visualize PCA results
fviz_pca_ind(pca_result,
             col.ind = as.factor(kmeans_result$cluster), # Color by clusters
             palette = "jco",
             addEllipses = TRUE,
             ellipse.level = 0.95)




#visualize dendrogram
fviz_dend(hcluster,show_labels = FALSE ,rect = TRUE)
fviz_cluster(hcluster,ellipse.type = "convex")


# Determine the optimal number of clusters using the Elbow method
fviz_nbclust(clustering_data_scaled, kmeans, method = "wss")
```