AIM: To perform partition based clustering and understand about clustering

SOFTWARE REQUIRED: RStudio

R CODE:

```
data <- read.csv("iris.csv", row.names = 1)
df <- scale(data)
set.seed(112)
fit <- kmeans(df,3)
fit$size
fit$withinss
fit$tot.withinss
Kmax <- 15
WCSS <- rep(NA, Kmax)
nClust <- list()
for (i in 1:Kmax) {
  fit <- kmeans(df, i)
  WCSS[i] <- fit$tot.withinss
  nClust[[i]] <- fit$size
}
plot(1:Kmax, WCSS, type = "b", pch=19)
library(factoextra)
fviz_nbclust(df, kmeans, method = "wss")
library(cluster)
fit <- pam(df, 3, metric = "manhattan")
fviz_nbclust(df, pam, method = "silhouette")
```

OUTPUT:

```
Console   Terminal ×   Background Jobs ×
R  R 4.3.2 · /cloud/project/

> data <- read.csv("iris.csv", row.names = 1)
> df <- scale(data)
> set.seed(112)
> fit <- kmeans(df,3)
> fit$size
[1] 47 53 50
> fit$withinss
[1] 47.45019 44.08754 47.35062
> fit$tot.withinss
[1] 138.8884
> Kmax <- 15
> WCSS <- rep(NA, Kmax)
> nClust <- list()
> for (i in 1:Kmax) {
+    fit <- kmeans(df, i)
+    WCSS[i] <- fit$tot.withinss
+    nClust[[i]] <- fit$size
+ }
> plot(1:Kmax, WCSS, type = "b", pch=19)
> library(factoextra)
> fviz_nbclust(df, kmeans, method = "wss")
> library(cluster)
> fit <- pam(df, 3, metric = "manhattan")
> fviz_nbclust(df, pam, method = "silhouette")
```
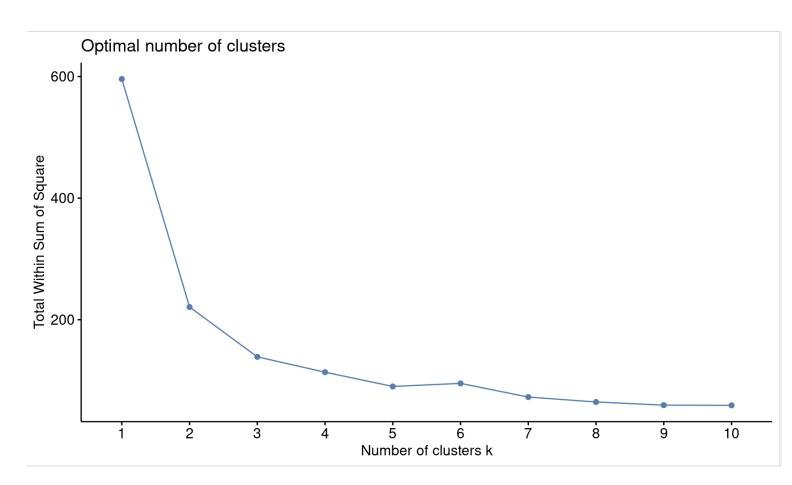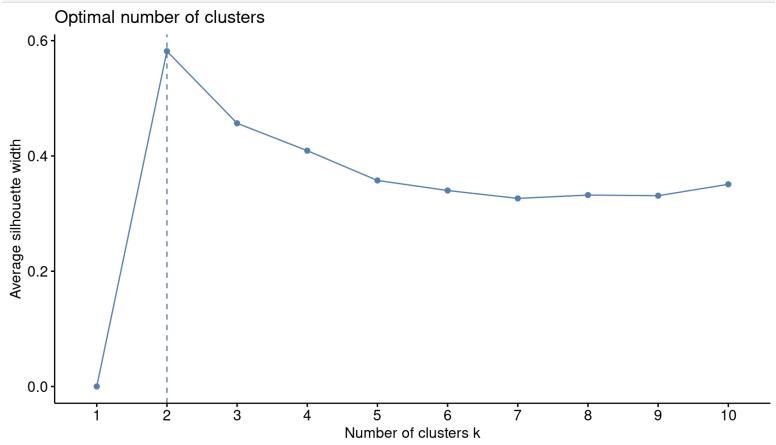
BCSE352E–Essentials of Data Analytics – Lab [Winter Semester 2023–24]
Name of the Student: **Sujay Ghosh**
Register Number: **21BLC1607**

Optimal number of clusters

Optimal number of clusters

BCSE352E–Essentials of Data Analytics – Lab [Winter Semester 2023–24]
Name of the Student: **Sujay Ghosh**
Register Number: **21BLC1607**