

# capstone\_model\_3\_1 (kmeans)

2022-12-17

```
library(dplyr)
library(ggplot2)
library(stringr)
library(gridExtra)
library(tidyverse)
library(cluster)
library(factoextra)
```

```
#import dataset
```

```
radiomics <- read.csv("radiomics_completedata.csv")
```

```
str(radiomics)
glimpse(radiomics)
```

## initial dimension

```
dim(radiomics)
```

```
#check for missing values
```

```
is.na(radiomics)
sum(is.na(radiomics))
na.omit(radiomics)
```

## Scale data

```
scale(radiomics)
head(radiomics)
newdf1 = subset(radiomics, select = c(-Institution))
newdf1
```

```
#Determining Optimal Number of Clusters
```

```
set.seed(123)
```

## Compute k-means clustering with $k = 2$

```
final <- kmeans(newdf1, 2, nstart = 25)  
print(final)
```

```
#final data
```

```
fviz_cluster(final, data = newdf1)
```