K-Means Cluster Analysis on Weekly Product Sales

Code

Hide

```
# Importing the dataset
dataset = read.csv('Sales_Transactions_Dataset_Weekly.csv')
dataset = dataset[1:53]
dim(dataset)
```

```
[1] 811  53
```

Hide

```
print("Number of products = ")
```

```
[1] "Number of products = "
```

Hide

```
print(nrow(dataset)-1)
```

```
[1] 810
```

Hide

```
print("Number of weeks of data available for the year:")
```

```
[1] "Number of weeks of data available for the year:"
```

Hide

```
print(ncol(dataset)-1)
```

```
[1] 52
```

Hide

```
str(dataset)
```

```
'data.frame':   811 obs. of  53 variables:
 $ Product_Code: Factor w/ 811 levels "P1","P10","P100",..: 1 112 223 332 443 554 663 770 801 2
...
 $ W0          : int  11 7 7 12 8 3 4 8 14 22 ...
 $ W1          : int  12 6 11 8 5 3 8 6 9 19 ...
 $ W2          : int  10 3 8 13 13 2 3 10 10 19 ...
 $ W3          : int  8 2 9 5 11 7 7 9 7 29 ...
 $ W4          : int  13 7 10 9 6 6 8 6 11 20 ...
 $ W5          : int  12 1 8 6 7 3 7 8 15 16 ...
 $ W6          : int  14 6 7 9 9 8 2 7 12 26 ...
 $ W7          : int  21 3 13 13 14 6 3 5 7 20 ...
 $ W8          : int  6 3 12 13 9 6 10 10 13 24 ...
 $ W9          : int  14 3 6 11 9 3 3 10 12 20 ...
 $ W10         : int  11 2 14 8 11 1 5 8 15 31 ...
 $ W11         : int  14 2 9 4 18 1 2 8 15 22 ...
 $ W12         : int  16 6 4 5 8 5 3 15 16 23 ...
 $ W13         : int  9 2 7 4 4 4 4 9 10 19 ...
 $ W14         : int  9 0 12 15 13 3 5 5 9 15 ...
 $ W15         : int  9 6 8 7 8 5 3 11 9 19 ...
 $ W16         : int  14 2 7 11 10 3 7 10 13 22 ...
 $ W17         : int  9 7 11 9 15 5 10 7 8 23 ...
 $ W18         : int  3 7 10 15 6 10 0 13 10 20 ...
 $ W19         : int  12 9 7 4 13 8 3 9 18 33 ...
 $ W20         : int  5 4 7 6 11 4 7 12 18 16 ...
 $ W21         : int  11 7 13 7 6 9 5 11 17 23 ...
 $ W22         : int  7 2 11 11 10 7 1 5 10 23 ...
 $ W23         : int  12 4 8 7 9 5 5 11 16 16 ...
 $ W24         : int  5 5 10 9 8 4 7 11 14 25 ...
 $ W25         : int  9 3 8 6 12 2 5 12 10 27 ...
 $ W26         : int  7 5 14 10 8 1 2 3 4 12 ...
 $ W27         : int  10 8 5 10 9 3 4 10 7 15 ...
 $ W28         : int  5 5 3 2 13 2 3 12 7 15 ...
 $ W29         : int  11 5 13 6 3 4 1 9 10 11 ...
 $ W30         : int  7 3 11 7 5 0 3 9 3 14 ...
 $ W31         : int  10 1 9 2 3 3 2 10 13 29 ...
 $ W32         : int  12 3 7 5 5 2 2 8 9 23 ...
 $ W33         : int  6 2 8 12 5 11 4 9 7 12 ...
 $ W34         : int  5 3 7 5 9 2 2 8 9 16 ...
 $ W35         : int  14 10 9 19 7 1 6 9 8 9 ...
 $ W36         : int  10 5 6 8 4 4 4 15 7 23 ...
 $ W37         : int  9 2 12 6 8 4 5 6 9 22 ...
 $ W38         : int  12 7 12 8 8 3 1 7 15 15 ...
 $ W39         : int  17 3 9 8 5 2 3 8 8 18 ...
 $ W40         : int  7 2 3 12 5 5 5 3 9 13 ...
 $ W41         : int  11 5 5 6 8 4 8 9 8 17 ...
 $ W42         : int  4 2 6 9 7 4 2 10 11 14 ...
 $ W43         : int  7 4 14 10 11 2 3 14 5 17 ...
 $ W44         : int  8 5 5 3 7 4 3 4 13 11 ...
 $ W45         : int  10 1 5 4 12 3 6 8 3 24 ...
 $ W46         : int  12 1 7 6 6 6 2 8 7 13 ...
 $ W47         : int  3 4 8 8 6 5 6 6 7 16 ...
 $ W48         : int  7 5 14 14 5 3 2 7 10 18 ...
 $ W49         : int  6 1 8 8 11 3 4 4 12 23 ...
```

```
 $ W50         : int  5 6 8 7 8 10 2 9 7 18 ...
 $ W51         : int  10 0 7 8 9 6 1 9 13 20 ...
```

Hide

```
#Checking for missing data
d3=dataset
for(i in 1:ncol(d3))
   {
    print(colnames(d3[i]))
    print(sum(is.na(d3[i])))
   }
```

```
[1] "Product_Code"
[1] 0
[1] "W0"
[1] 0
[1] "W1"
[1] 0
[1] "W2"
[1] 0
[1] "W3"
[1] 0
[1] "W4"
[1] 0
[1] "W5"
[1] 0
[1] "W6"
[1] 0
[1] "W7"
[1] 0
[1] "W8"
[1] 0
[1] "W9"
[1] 0
[1] "W10"
[1] 0
[1] "W11"
[1] 0
[1] "W12"
[1] 0
[1] "W13"
[1] 0
[1] "W14"
[1] 0
[1] "W15"
[1] 0
[1] "W16"
[1] 0
[1] "W17"
[1] 0
[1] "W18"
[1] 0
[1] "W19"
[1] 0
[1] "W20"
[1] 0
[1] "W21"
[1] 0
[1] "W22"
[1] 0
[1] "W23"
[1] 0
[1] "W24"
[1] 0
[1] "W25"
```

```
[1] 0
[1] "W26"
[1] 0
[1] "W27"
[1] 0
[1] "W28"
[1] 0
[1] "W29"
[1] 0
[1] "W30"
[1] 0
[1] "W31"
[1] 0
[1] "W32"
[1] 0
[1] "W33"
[1] 0
[1] "W34"
[1] 0
[1] "W35"
[1] 0
[1] "W36"
[1] 0
[1] "W37"
[1] 0
[1] "W38"
[1] 0
[1] "W39"
[1] 0
[1] "W40"
[1] 0
[1] "W41"
[1] 0
[1] "W42"
[1] 0
[1] "W43"
[1] 0
[1] "W44"
[1] 0
[1] "W45"
[1] 0
[1] "W46"
[1] 0
[1] "W47"
[1] 0
[1] "W48"
[1] 0
[1] "W49"
[1] 0
[1] "W50"
[1] 0
[1] "W51"
[1] 0
```

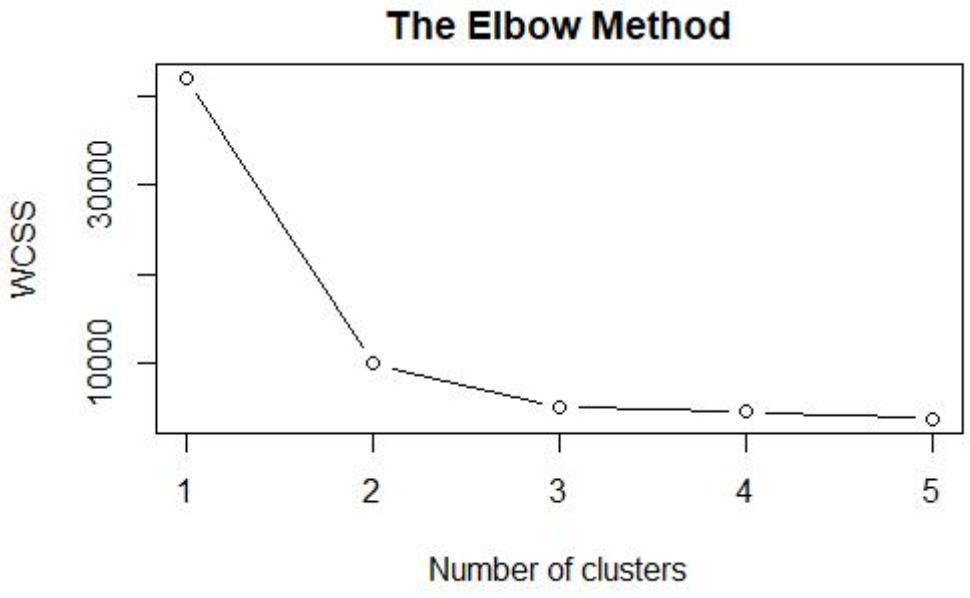
Hide

```
# There is no missing data
# Encoding the target feature as factor
dataset$Product_Code = factor(dataset$Product_Code)
# Training Set
training_set = dataset
# Feature Scaling
training_set = scale(dataset[-1])
str(training_set)
```

```
 num [1:811, 1:52] 0.1738 -0.1577 -0.1577 0.2567 -0.0748 ...
 - attr(*, "dimnames")=List of 2
  ..$ : NULL
  ..$ : chr [1:52] "W0" "W1" "W2" "W3" ...
 - attr(*, "scaled:center")= Named num [1:52] 8.9 9.13 9.39 9.72 9.57 ...
  ..- attr(*, "names")= chr [1:52] "W0" "W1" "W2" "W3" ...
 - attr(*, "scaled:scale")= Named num [1:52] 12.1 12.6 13 13.6 13.1 ...
  ..- attr(*, "names")= chr [1:52] "W0" "W1" "W2" "W3" ...
```

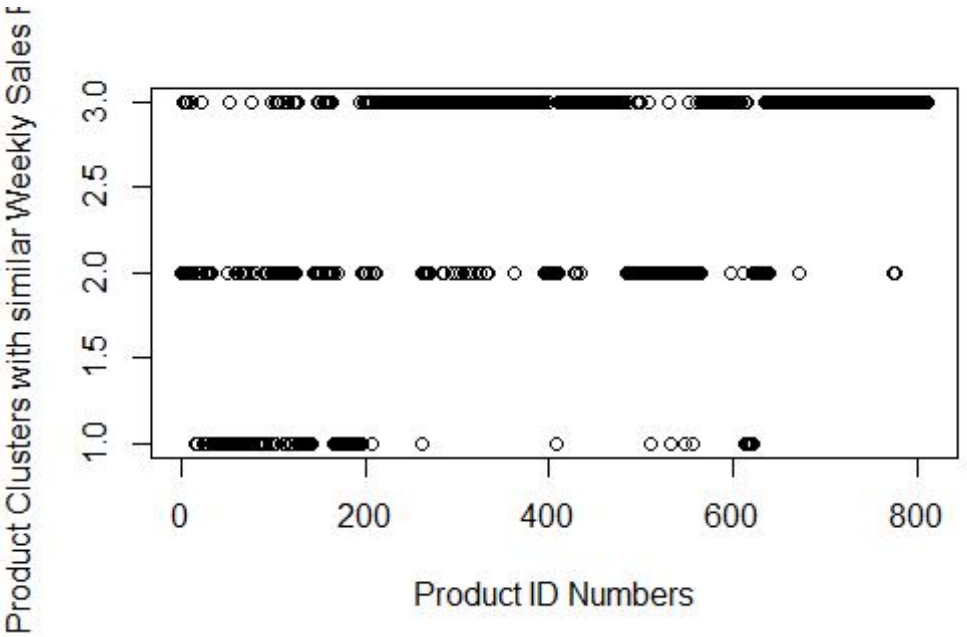
Hide

```
# Number of Clusters
k = 5
# Using the elbow method to find the optimal number of clusters
set.seed(123)
wcss = vector()
for (i in 1:k) wcss[i] = sum(kmeans(training_set, i, algorithm = "Hartigan-Wong")$withinss)
plot(1:k,
     wcss,
     type = 'b',
     main = paste('The Elbow Method'),
     xlab = 'Number of clusters',
     ylab = 'WCSS')
```



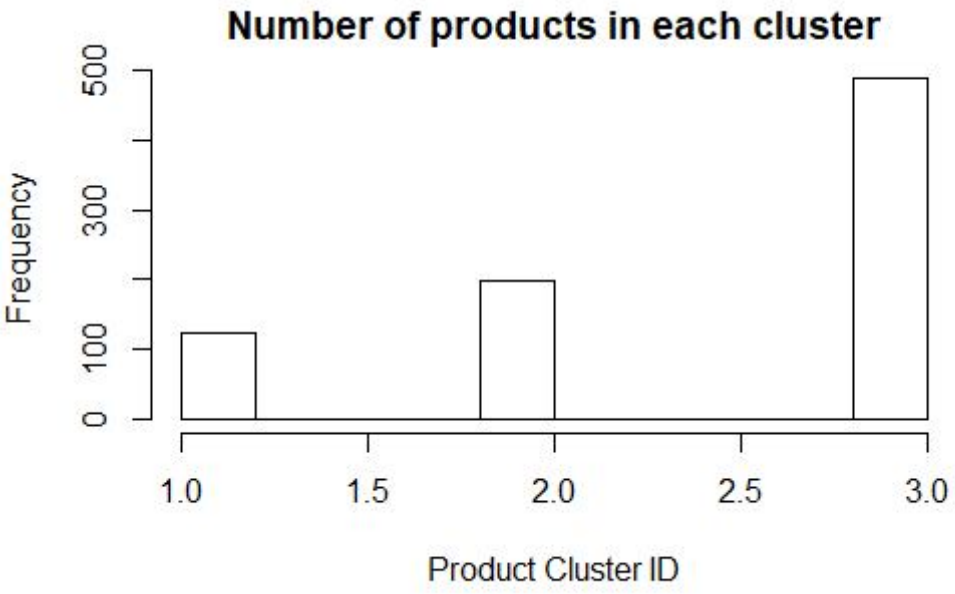
Hide

```
# Optimal Clusters Found
Optimal_Clusters = 3
# Fitting K-Means to the dataset
set.seed(456)
kmeans = kmeans(x = training_set, centers = Optimal_Clusters)
y_kmeans = kmeans$cluster
plot(y_kmeans,
     ylab = "Product Clusters with similar Weekly Sales Pattern",
     xlab = "Product ID Numbers")
```



Hide

```
hist(y_kmeans, main  = "Number of products in each cluster", xlab = "Product Cluster ID")
```



Hide

```
# Visualising the clusters
library(cluster)
clusplot(training_set,
         y_kmeans,
         lines = 0,
         shade = TRUE,
         color = TRUE,
         labels = 2,
         plotchar = FALSE,
         span = TRUE,
         main = paste('Clusters of Products'),
         xlab = 'Qty of Sales',
         ylab = 'Sales Location')
```