Hierarchical Clustering Algorithms

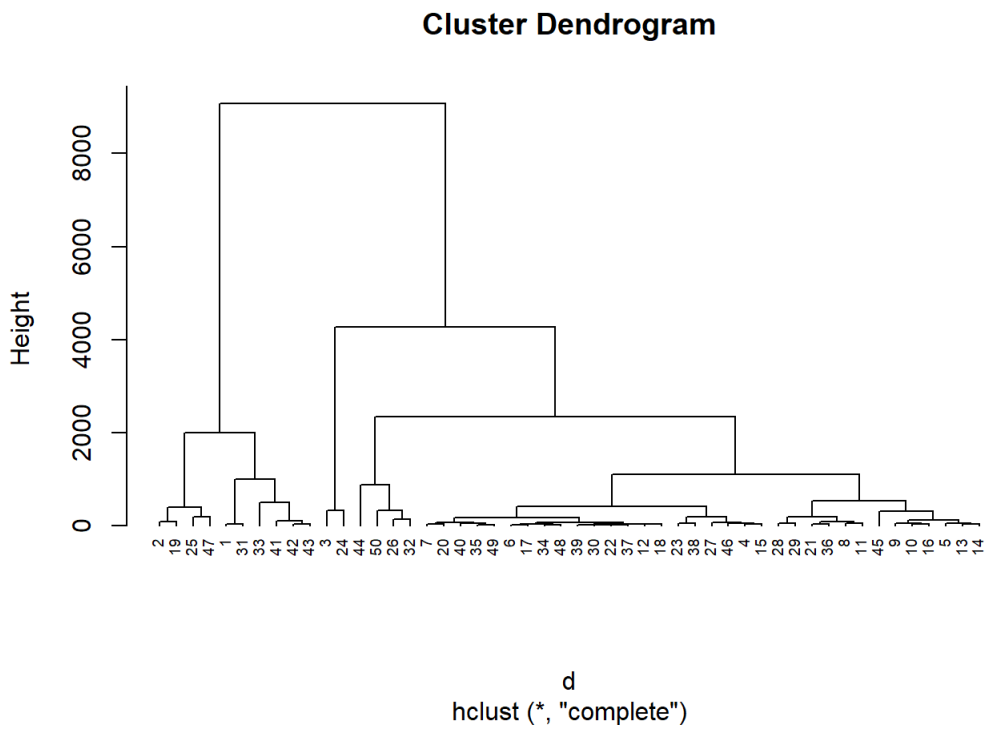
Customer Purchasing Behavior via Clustering Analysis

Clustering analysis uses mathematical models to discover groups of customers with similar purchasing behaviors on the smallest variations among customers within each group.

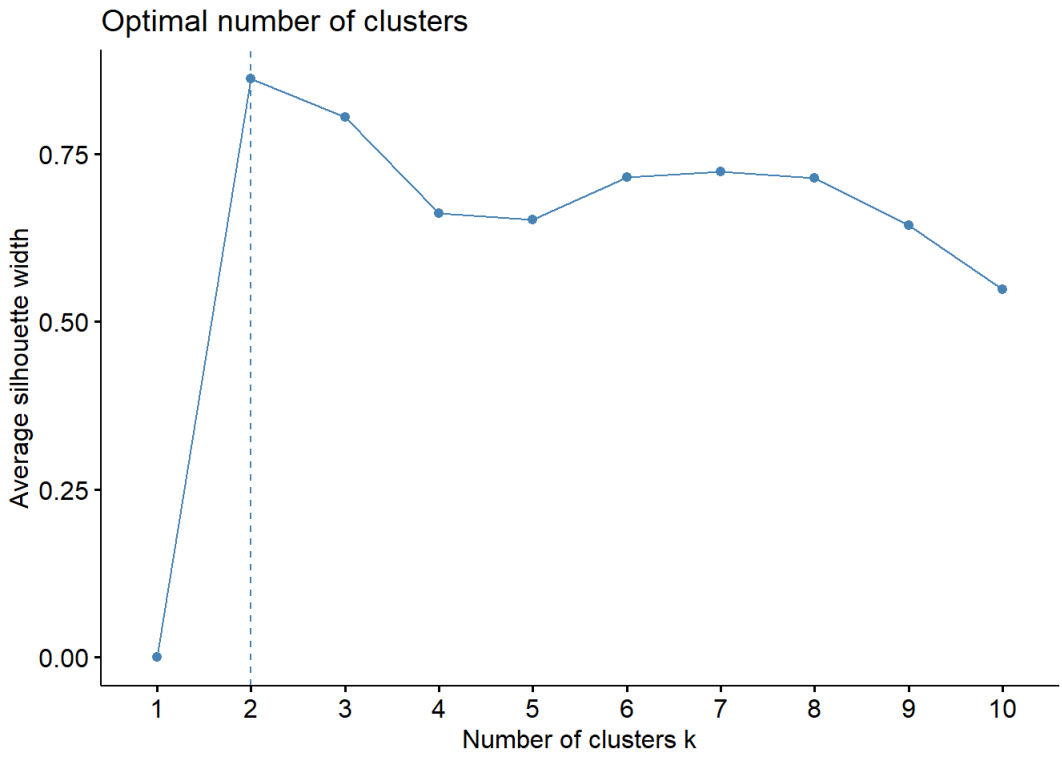
Clustering Analysis

The goal of the current hierarchical clustering analysis in marketing is to accurately segment customers in order to achieve more effective customer marketing via personalization. The clusters from the results would assist in better customer modeling and predictive analytics.

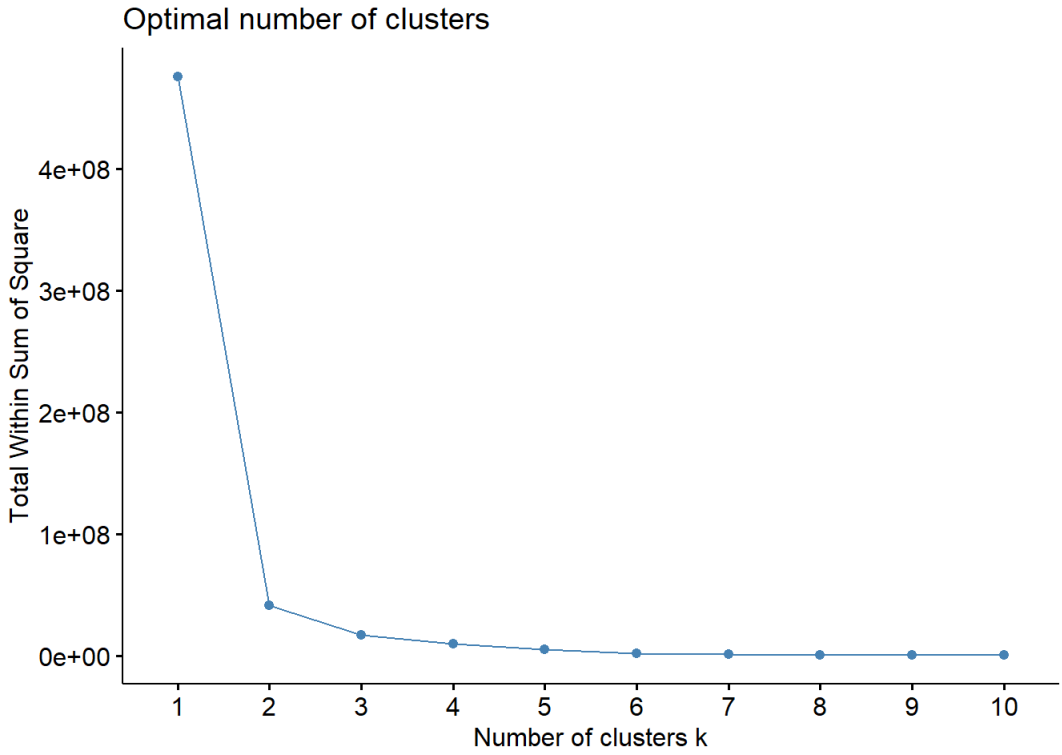
```
## required packages
library(cluster) # clustering analysis algorithms
library(factoextra) # clusters visualization
library(dendextend) # comparing dendrograms
library(dplyr) # data wrangling
library(knitr) # html table for rmarkdown
library(kableExtra) # html table for rmarkdown
## load data
df <- read.csv(file = "Purchasing Behavior.csv", header = TRUE)
## calculate matrix distance
d <- dist(df, method = "euclidean")
## Hierarchical clustering using complete linkage
hca1 <- hclust(d, method = "complete")
## plot dendrogram
plot(hca1, cex = 0.6, hang = -1)
```



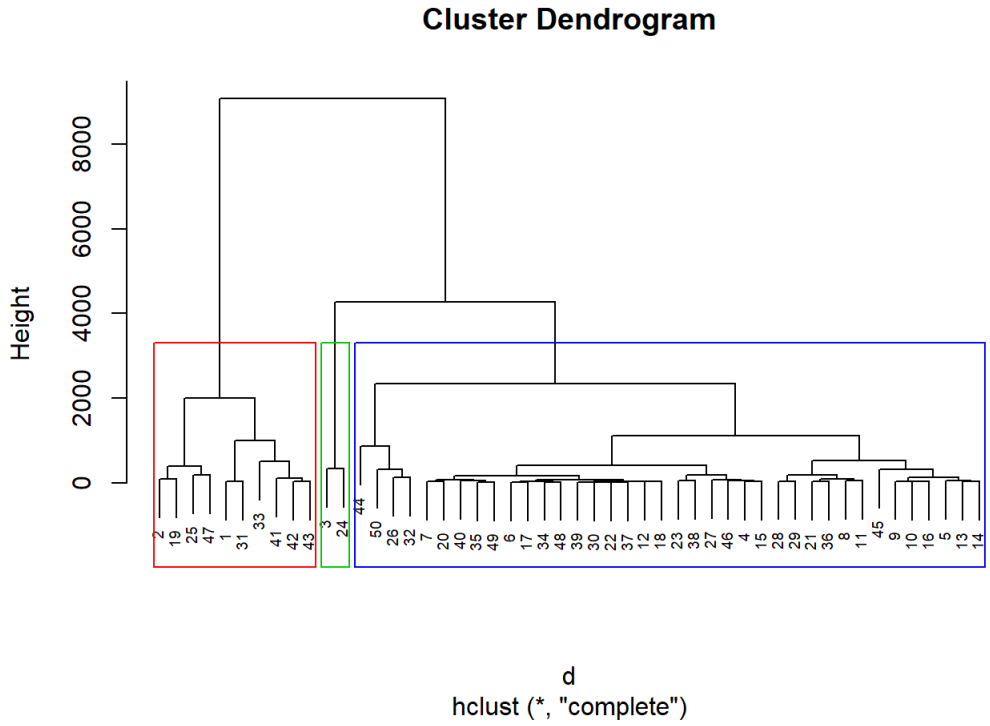

```
## determine clusters
fviz_nbclust(df, FUN = hcut, method = "silhouette") # average silhouette method
```



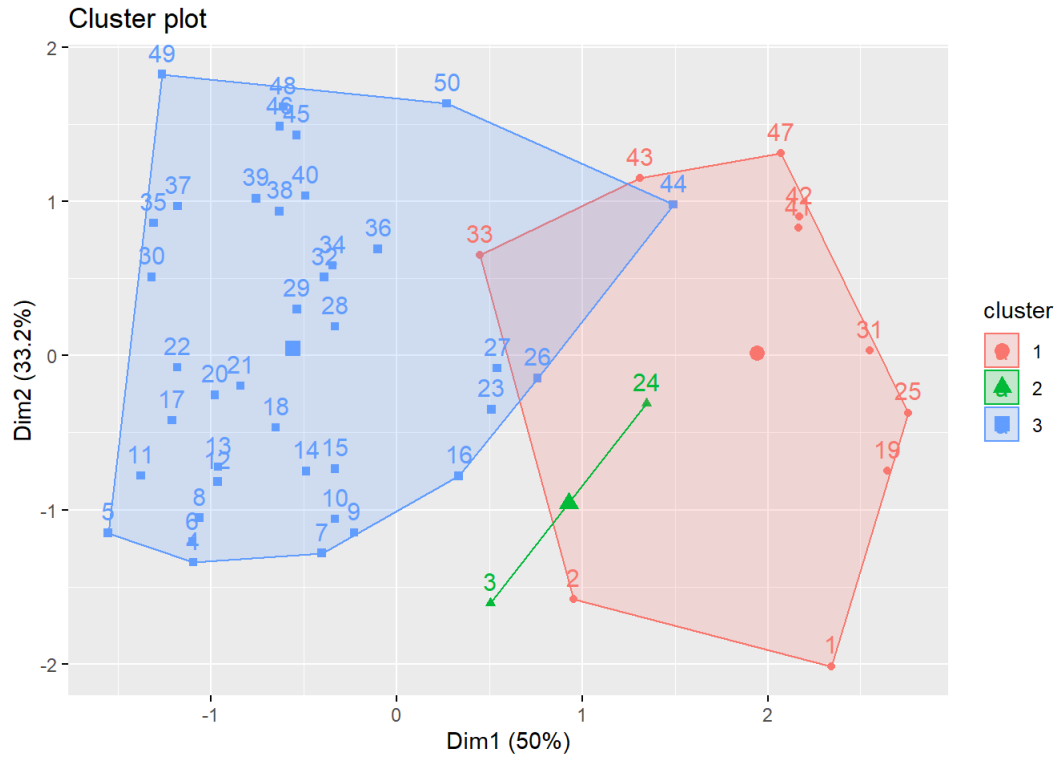

```
fviz_nbclust(df, FUN = hcut, method = "wss") # Elbow Method
```




```
## plot clusters in dendrogram
plot(hca1, cex = 0.6)
rect.hclust(hca1, k = 3, border = 2:5)
```




```
## plot clusters in a scatter plot
df_cluster <- cutree(hca1, k = 3)
fviz_cluster(list(data = df, cluster = df_cluster))
```




```
## create data file with clusters
df1 <-
  df %>%
  mutate(cluster = df_cluster)
head(df1)
```

```
##   ID Purchase Spend cluster
## 1  1      101  7078       1
## 2  2       29  9007       1
## 3  3       56  3955       2
## 4  4       26   223       3
## 5  5        2   856       3
## 6  6       27    10       3
```

Sample Clustering Analysis Results

The following table shows the results of the culstering analysis performed on the customer purchasing behavior data. The analysis resulted in discovering three clusters of customers.

```
## create table with mean values for different clusters
df2 <-
  df1 %>%
  group_by(cluster) %>%
  summarise_at(vars(-ID), funs(mean(., na.rm = TRUE)))
## generate table
df2 %>%
  kable() %>%
  kable_styling(bootstrap_options = "striped", full_width = F) %>%
  row_spec(1, bold = T, background = "lightyellow") %>%
  row_spec(2, bold = T, background = "lightblue") %>%
  row_spec(3, bold = T, background = "lightgrey")
```

cluster	Purchase	Spend
1	67.10000	8078.0000
2	67.50000	4118.5000
3	36.97368	533.7368

Cluster 1: customer with a lot of purchasing behavior + spent a lot of money in the period of analysis. Cluster 2: customer with a lot of purchasing behavior + spent only half of money compared to cluster 1. Cluster 3: customer with little purchasing behavior + spent small amount of money.

Cluster 1: motivated shopper + high spender. Cluster 2: motivated shopper + mid spender. Cluster 3: unmotivated shopper + low spender.

By understanding more about customers' purchasing behaviors, the marketing team will have a more clear picture about various customers. In that case, they will be able to make better interactions to customers based on their types and preferences.