

Brand Personality

Aditya Agral Serhansyah

December 6, 2023

Contents

1. Background	1
1.1 Analysis Objective	2
2. Theory	2
2.1 Hieararchical Clustering	2
2.2 Fuzzy C-Means Clustering	2
2.3 Cluster Validation: Dunn Coefficient	2
3. Read Data	3
3.1 Preprocessing Data	3
4. Brand Clustering using Hierarchical Clustering	5
4.1 Agglomerative	5
4.2 Divisive Hierarchical Clustering	7
4.3 Choose Cluster	7
5. Brand Clustering using Fuzzy C Means Clustering	12
5.1 Cluster Exploration	13
6. Cluster Validation	16
7. Conclusion	17

1. Background

Brand personality merupakan salah satu faktor yang membuat seseorang membeli suatu produk. Brand satu dengan brand lainnya memiliki karakteristik yang berbeda-beda seperti manusia, misalnya: elegan, sederhana, ceria, ramah dan lain-lain. Konsumen ternyata cenderung membeli produk yang memiliki brand personality yang cocok dengan mereka sehingga inilah yang membuat brand personality menjadi salah satu faktor yang mendorong minat beli konsumen. Brand personality juga merupakan tools bagi perusahaan untuk membedakan produk mereka dengan produk perusahaan lain. Dengan mempertimbangkan aspek personality, penting bagi perusahaan untuk menentukan strategi pemasaran seperti market segmentation dan brand positioning. Dengan menentukan strategi pemasaran, perusahaan dapat lebih fokus melayani segmen pasar dan mampu unggul dalam persaingan bisnis yang semakin ketat.

Dikarenakan terdapat sangat banyak brand pada saat ini, besar kemungkinan terdapat brand-brand yang memiliki brand personality yang serupa. Sehingga perusahaan perlu membuat target untuk membangun brand personality yang efektif. Untuk mengetahui persaingan dalam pasar, perusahaan harus mengetahui brand lain mana saja yang kira-kira memiliki brand personality yang serupa. Untuk itu, akan dilakukan clustering untuk menemukan kelompok-kelompok brand yang memiliki personality serupa. Clustering adalah metode machine learning jenis unsupervised yang bertujuan untuk mengelompokkan berdasarkan karakteristiknya.

1.1 Analysis Objective

Tujuan penelitian ini adalah untuk mengetahui kelompok segmentasi brand berdasarkan personality nya, sehingga perusahaan dapat mengetahui persaingan pasar.

```
library("readxl")
library("dplyr")
library("ggplot2")
library("ggdist")
library(cluster)      # clustering algorithms
library(factoextra)   # clustering visualization
library(dendextend)   # for comparing two dendrograms
library(tidyr)        # pivot longer
library(tidytext)
library(fpc)          # cluster validation
```

Pada data awal, terdapat 16 kolom dengan 1 kolom berisi nama-nama brand ("Brand") dan 15 kolom lainnya merupakan atribut-atribut personality.

2. Theory

2.1 Hieararchical Clustering

Pengelompokkan hierarki (hierarchical clustering) adalah metode analisis cluster dengan cara membangun sebuah hierarki kelompok. Strategi untuk mengelompokkan hierarki pada umumnya dibagi menjadi dua jenis yaitu agglomerative (pemusatan) dan divisive (penyebaran). Metode agglomerative (pemusatan) biasanya dipergunakan pada bidang ekonomi dan sosial masyarakat.

Hieararchical clustering dibagi menjadi dua, yaitu Agglomerative clustering dan Divisive hierarchical clustering.

1. **Agglomerative clustering** dikenal sebagai AGNES (Agglomerative Nesting). Pengelompokan agglomerative bekerja dengan pendekatan bottom-up. Artinya, setiap observasi pada data awalnya dianggap sebagai cluster elemen tunggal (leaf). Pada setiap langkah algoritma, dua cluster yang paling mirip digabungkan menjadi cluster baru (node) yang lebih besar. Prosedur ini diulang sampai semua titik menjadi anggota hanya satu cluster besar (root). Hasilnya adalah pohon (tree) yang dapat diplot sebagai dendrogram.
2. **Divisive hierarchical clustering** dikenal sebagai DIANA (penyebaran) dan bekerja secara top-down. Algoritmanya adalah urutan terbalik dari AGNES. Itu dimulai dengan root, dimana semua objek termasuk dalam satu cluster. Pada setiap langkah iterasi, cluster yang paling heterogen dibagi menjadi dua. Proses ini diulang sampai semua objek berada di cluster masing-masing.

Perhatikan bahwa pengelompokan aglomeratif bagus untuk mengidentifikasi kelompok kecil. Pengelompokan hirarkis divisif bagus dalam mengidentifikasi kelompok besar.

2.2 Fuzzy C-Means Clustering

Fuzzy C-Means adalah metode clustering yang pertama kali diperkenalkan oleh Jim Bezdek pada tahun 1981. Fuzzy C-Means mengelompokkan setiap observasi data ke suatu cluster berdasarkan derajat keanggotaannya.

2.3 Cluster Validation: Dunn Coefficient

Dunn index menilai seberapa rapat anggota dalam suatu cluster dengan variansi yang kecil dan seberapa terpisah cluster yang satu dengan yang lainnya. Cluster yang baik adalah yang memiliki nilai dunn index besar.

3. Read Data

Data yang digunakan pada analisis ini didapat dari paper milik Jack Hoare & Tim Bock (2019). Data tersebut merupakan sebuah tabel kontingensi untuk 28 brand dan 15 atribut personality. Data dikumpulkan menggunakan metode nonprobability sampling dari 3173 orang dewasa di Australia pada tahun 2005. Responden diminta untuk menentukan tiga atribut personality untuk setiap brand yang ditunjukkan. Untuk setiap atribut personality, responden ditunjukkan 10 brand.

3.1 Preprocessing Data

Pada data awal, terdapat 16 kolom dengan 1 kolom berisi nama-nama brand ("Brand") dan 15 kolom lainnya merupakan atribut-atribut personality.

```
data_brand <- read_excel("brand_dataset.xlsx")
data_brand <- as.data.frame(data_brand)
```

Nilai kolom brand dijadikan sebagai index row untuk memudahkan visualisasi. selanjutnya kolom "Brand" dibuang. Berikut adalah 6 baris pertama dari data:

```
# meng-assign nilai dari kolom brand menjadi rownames
rownames(data_brand) <- data_brand$Brand

# membuang kolom yang tidak digunakan
data_brand <- data_brand %>%
  select(-Brand)
head(data_brand)
```

```
##           charming cheerful daring down_to_earth honest imaginative
## AmericanExpress           20          9          15           10          14           17
## Applecomputers           25          27          23           14          27           54
## Avoncosmetics            33          33           7           26          25           20
## CalvinKleinPerfume        59          25          50           7          13           31
## Campbell'sSoup            22          30           8           62          50           11
## ColgateToothpaste         20          45          13           51          56           17
##           intelligent outdoorsy reliable spirited successful tough
## AmericanExpress           44           14           26           14           63           20
## Applecomputers           71           4           32           34           43           13
## Avoncosmetics            10           6           23           7           23           6
## CalvinKleinPerfume        16           11           21           38           44           5
## Campbell'sSoup            11           16           53           14           31           12
## ColgateToothpaste         32           12           82           19           56           23
##           up_to_date upper_class wholesome
## AmericanExpress           30           69           8
## Applecomputers           53           33           24
## Avoncosmetics            20           6           26
## CalvinKleinPerfume        30           81           10
## Campbell'sSoup            19           11           80
## ColgateToothpaste         37           9           52
```

Dapat diperhatikan juga bahwa setiap variabel dari data brand di atas sudah memiliki tipe data yang benar untuk dilakukan clustering, yaitu bertipe data numerik. Selanjutnya akan dilakukan pengecekan apakah terdapat missing values dalam data.

```
anyNA(data_brand)
```

```
## [1] FALSE
```

Karena dalam melakukan clustering akan dilakukan perhitungan jarak, maka skala data harus sama. Berikut akan dicek skala data pada data brand:

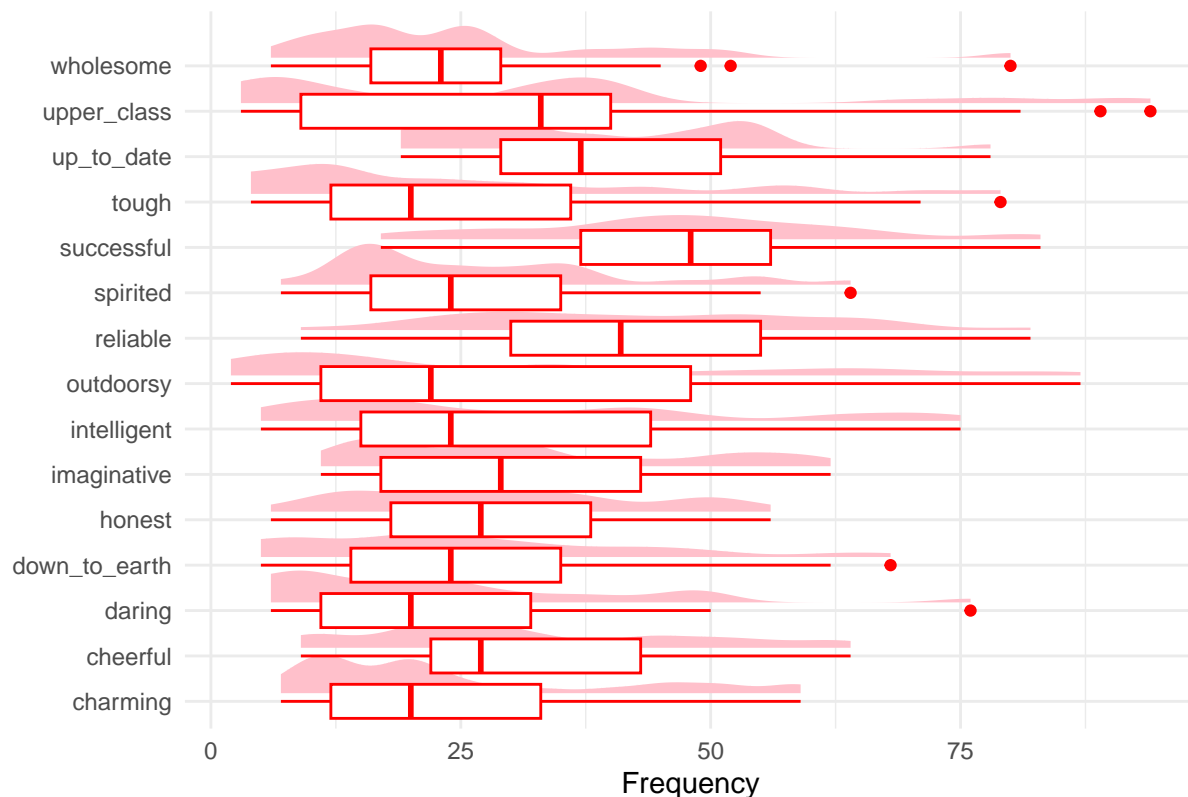
```
summary(data_brand)
```

```
##      charming      cheerful      daring      down_to_earth
## Min.   : 7.00    Min.   : 9.00    Min.   : 6.00    Min.   : 5.00
## 1st Qu.:12.00    1st Qu.:22.00    1st Qu.:11.00    1st Qu.:14.00
## Median :20.00    Median :27.00    Median :20.00    Median :24.00
## Mean   :24.72    Mean   :31.48    Mean   :24.24    Mean   :26.52
## 3rd Qu.:33.00    3rd Qu.:43.00    3rd Qu.:32.00    3rd Qu.:35.00
## Max.   :59.00    Max.   :64.00    Max.   :76.00    Max.   :68.00
##      honest      imaginative      intelligent      outdoorsy      reliable
## Min.   : 6.00    Min.   :11.00    Min.   : 5.00    Min.   : 2.00    Min.   : 9.0
## 1st Qu.:18.00    1st Qu.:17.00    1st Qu.:15.00    1st Qu.:11.00    1st Qu.:30.0
## Median :27.00    Median :29.00    Median :24.00    Median :22.00    Median :41.0
## Mean   :29.14    Mean   :31.66    Mean   :30.93    Mean   :30.21    Mean   :42.9
## 3rd Qu.:38.00    3rd Qu.:43.00    3rd Qu.:44.00    3rd Qu.:48.00    3rd Qu.:55.0
## Max.   :56.00    Max.   :62.00    Max.   :75.00    Max.   :87.00    Max.   :82.0
##      spirited      successful      tough      up_to_date      upper_class
## Min.   : 7      Min.   :17.00    Min.   : 4.00    Min.   :19.00    Min.   : 3.00
## 1st Qu.:16      1st Qu.:37.00    1st Qu.:12.00    1st Qu.:29.00    1st Qu.: 9.00
## Median :24      Median :48.00    Median :20.00    Median :37.00    Median :33.00
## Mean   :27      Mean   :47.38    Mean   :26.38    Mean   :38.48    Mean   :32.48
## 3rd Qu.:35      3rd Qu.:56.00    3rd Qu.:36.00    3rd Qu.:51.00    3rd Qu.:40.00
## Max.   :64      Max.   :83.00    Max.   :79.00    Max.   :78.00    Max.   :94.00
##      wholesome
## Min.   : 6.00
## 1st Qu.:16.00
## Median :23.00
## Mean   :25.66
## 3rd Qu.:29.00
## Max.   :80.00
```

Berikut adalah visualisasi dari distribusi data brand untuk melihat lebih lanjut skalanya.

```
data_brand %>%
  pivot_longer(col = names(data_brand), names_to = "personality", values_to = "values") %>%
  ggplot(mapping = aes(x = values, y = personality)) +
  ggdist::stat_halfeye(adjust = 0.5, width = 0, justification = -.2, fill = "pink") +
  geom_boxplot(.width = 12, colour = "red") +
  theme_minimal() +
  labs(title = "Boxplot for Personality in Data Brand",
       y = "",
       x = "Frequency")
```

Boxplot for Personality in Data Brand



Karena terdapat beberapa outlier, diputuskan untuk melakukan scaling agar skala data brand pada tiap variabel sama. Akan dilakukan scaling dengan fungsi `scale()` yang artinya mengubah skala data ke distribusi normal atau disebut juga dengan z-score standarization.

Hasil dari scaling data disimpan dalam objek `brand_scale`.

```
brand_scale <- scale(data_brand)
```

4. Brand Clustering using Hierarchical Clustering

4.1 Agglomerative

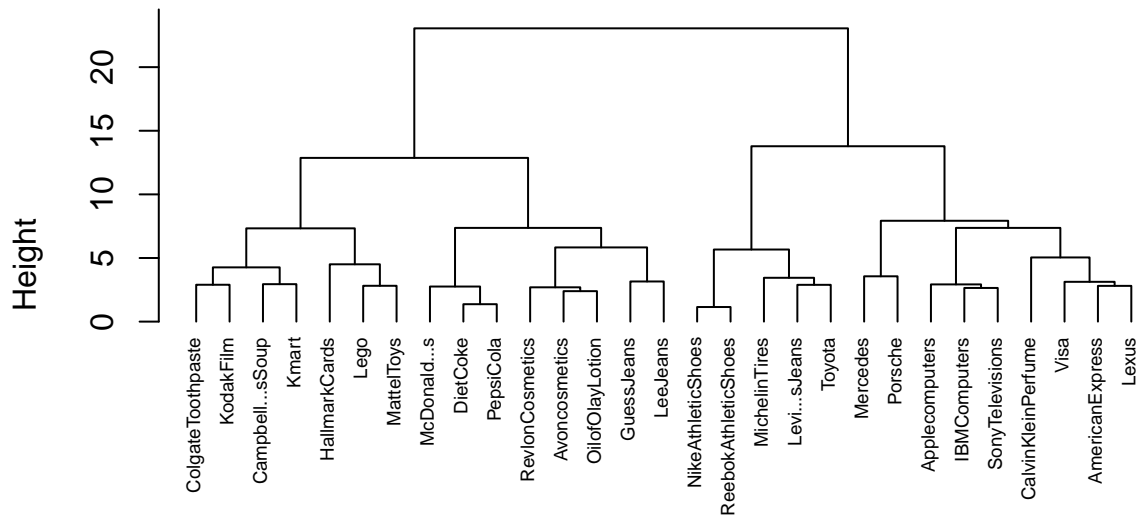
Kita dapat melakukan Agglomerative Hierarchical Clustering dengan `hclust` dari package `stat (base)`. Pertama kita menghitung nilai dissimilarity dengan `dist` yang pada case ini dipilih ukuran jarak euclidean dan kemudian memasukkan nilai ini ke dalam `hclust`. Selanjutnya, menentukan metode aglomerasi yang akan digunakan ("complete", "average", "single", "ward.D").

```
# Dissimilarity matrix
d <- dist(brand_scale, method = "euclidean")

# Hierarchical clustering using Complete Linkage
hc_agg1 <- hclust(d, method = "ward.D" )

# Plot the obtained dendrogram
plot(hc_agg1, cex = 0.6, hang = -1)
```

Cluster Dendrogram



d
hclust (*, "ward.D")

Sementara itu, kita juga dapat melakukan dengan Hierarchical Clustering agglomerative menggunakan fungsi `agnes()` dari package `cluster`. Secara default, ukuran jarak yang digunakan pada fungsi `agnes()` adalah euclidean. Salah satu kelebihanannya adalah bisa didapatkan koefisien agglomerative, yang mengukur jumlah struktur pengelompokan yang ditemukan (nilai yang mendekati 1 menunjukkan struktur pengelompokan yang kuat).

Menggunakan koefisien ini, memungkinkan kita untuk menemukan metode aglomerasi tertentu yang dapat mengidentifikasi struktur pengelompokan yang lebih kuat.

```
# methods to assess
m <- c("average", "single", "complete", "ward")
names(m) <- c("average", "single", "complete", "ward")

# function to compute coefficient
ac <- function(x) {
  agnes(brand_scale, method = x)$ac
}

library(purrr)
map_dbl(m, ac)
```

```
## average single complete ward
## 0.5265406 0.3076316 0.7201097 0.8099598
```

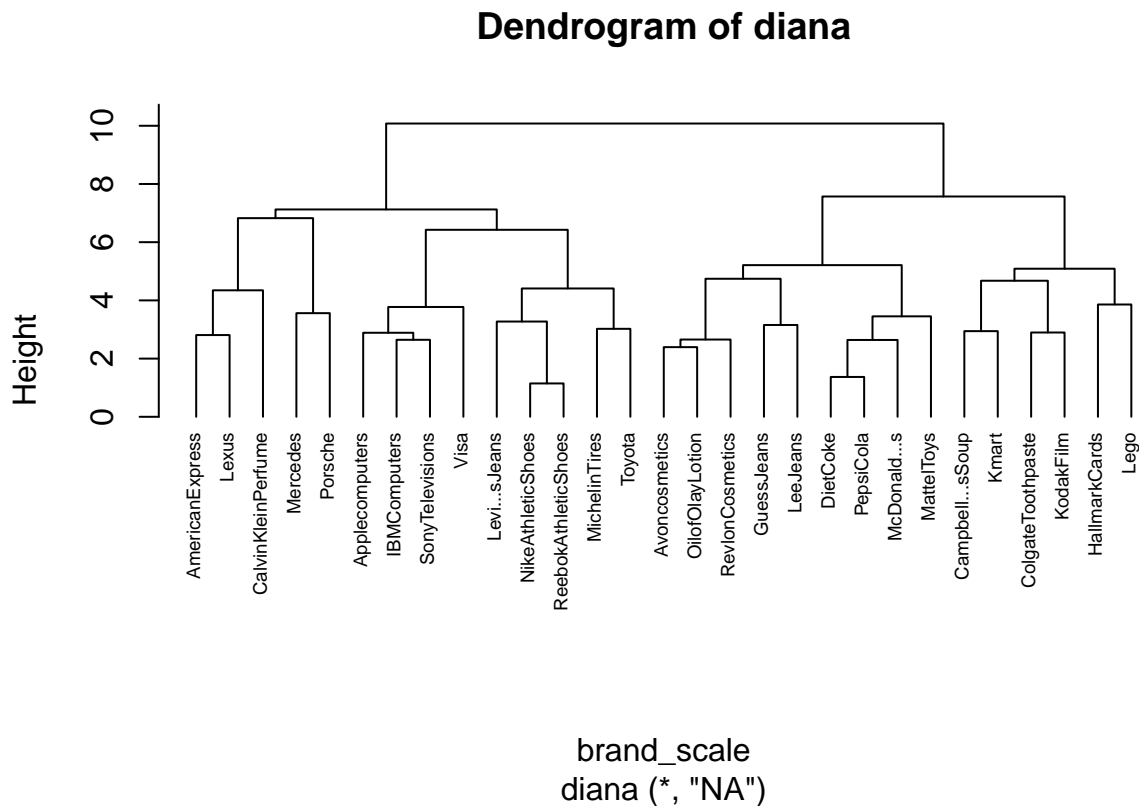
Dapat dilihat bahwa metode Ward mengidentifikasi struktur pengelompokan terkuat dari empat metode yang ada.

4.2 Divisive Hierarchical Clustering

Divisive Hierarchical Clustering dapat dilakukan dengan memanfaatkan fungsi `diana()` yang disediakan oleh package `cluster`. Sama seperti fungsi `agnes()`, secara default fungsi `diana()` menggunakan euclidean sebagai ukuran jaraknya.

```
# compute divisive hierarchical clustering
hc_div <- diana(brand_scale)

# plot dendrogram
pltree(hc_div, cex = 0.6, hang = -1, main = "Dendrogram of diana")
```



Sementara itu berikut adalah koefisien divide yang juga mengukur jumlah struktur pengelompokan yang ditemukan:

```
# Divide coefficient; amount of clustering structure found
hc_div$dc
```

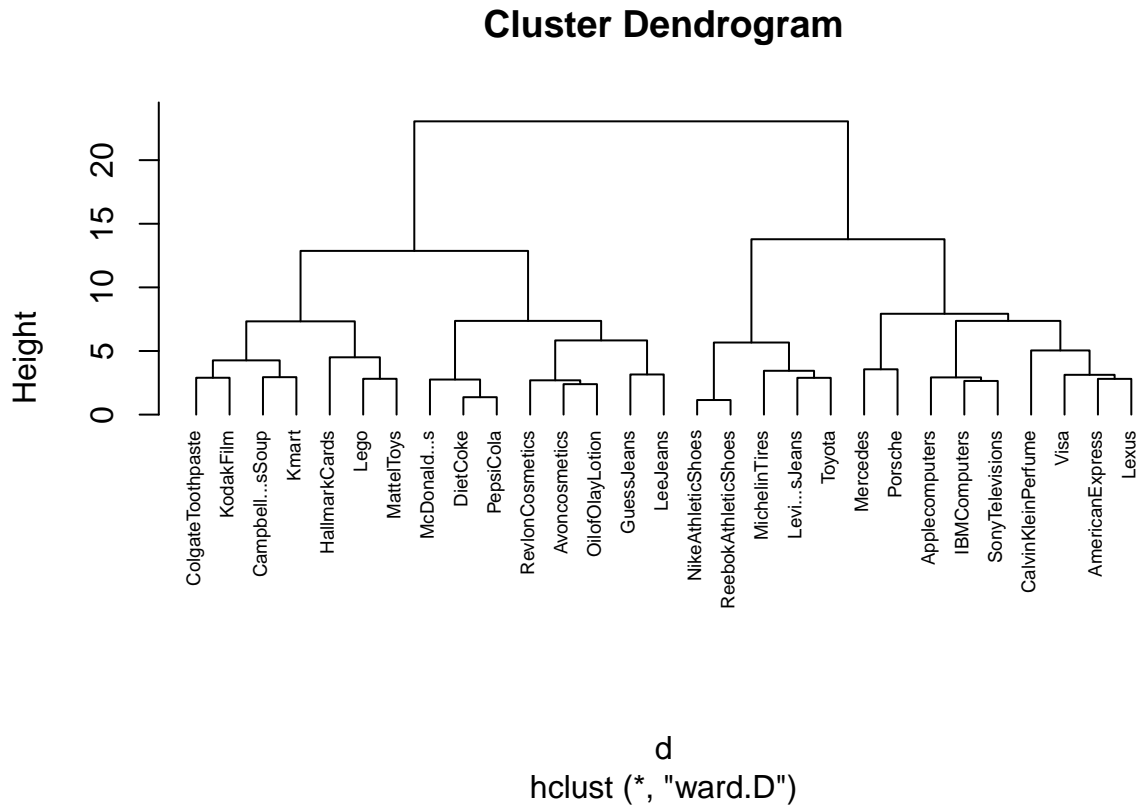
```
## [1] 0.7172142
```

Karena nilai koefisien milik agglomerative hierarchical clustering menggunakan metode ward lebih besar dari koefisien divide, yaitu sebesar 0.8099598. Maka analisis hierarchial clustering selanjutnya akan menggunakan agglomerative hierarchical clustering menggunakan metode ward.

4.3 Choose Cluster

Dendrogram yang dihasilkan Agglomerative Hierarchical Clustering `hc_agg1`.

```
plot(hc_agg1, cex = 0.6, hang = -1)
```



Interpretasi Dendrogram: Dalam dendrogram yang ditampilkan di atas, setiap leaf merupakan dengan satu observasi data. Saat kita naik ke atas pohon, obserdata data yang mirip satu sama lain digabungkan menjadi cabang. Ketinggian pengelompokan, yang diberikan pada sumbu vertikal, menunjukkan dissimilarity antara dua observasi data. Semakin tinggi ketinggian pengelompokan, semakin tidak mirip pengamatannya begitupun sebaliknya. Contohnya adalah tinggi cabang Nike dan Rebook yang sangat pendek, menandakan kedekatan sifat mereka.

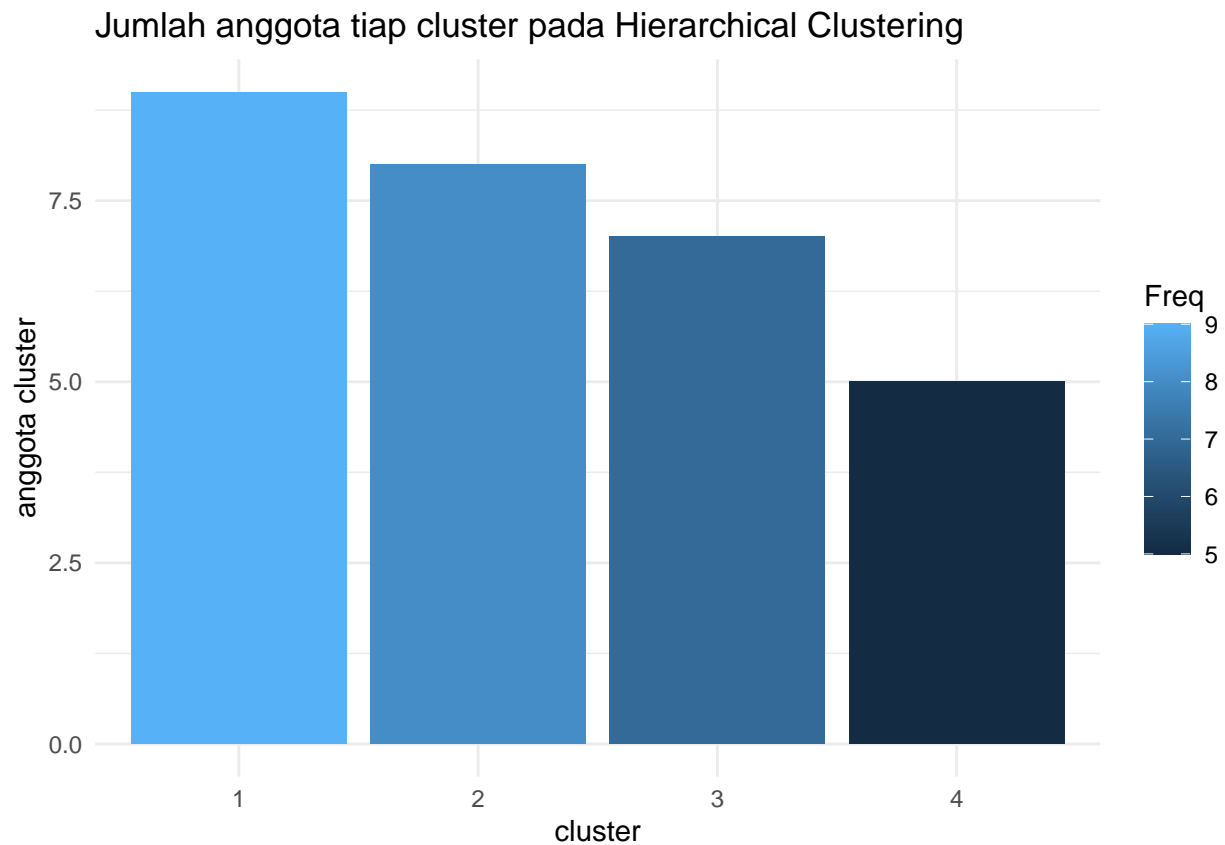
Untuk mengidentifikasi sub-kelompok (yaitu cluster), potong dendrogram pada ketinggian tertentu seperti yang dijelaskan pada bagian berikutnya. Ketinggian potongan ke dendrogram mengontrol jumlah cluster yang diperoleh.

Fungsi `cutree()` digunakan dan mengembalikan vektor yang berisi nomor cluster dari setiap observasi. Jika kita memotong dendrogram pada ketinggian 10 maka akan didapatkan 4 cluster.

```
# Cut tree into 4 groups
cluster <- cutree(hc_agg1, k = 4)
```

Berikut akan ditampilkan berapa banyak observasi data yang menjadi anggota tiap cluster:

```
table(cluster) %>%
  as.data.frame() %>%
  ggplot(mapping = aes(x = cluster, y = Freq, fill = Freq)) +
  geom_col() +
  theme_minimal() +
  labs(title = "Jumlah anggota tiap cluster pada Hierarchical Clustering",
       y = "anggota cluster")
```

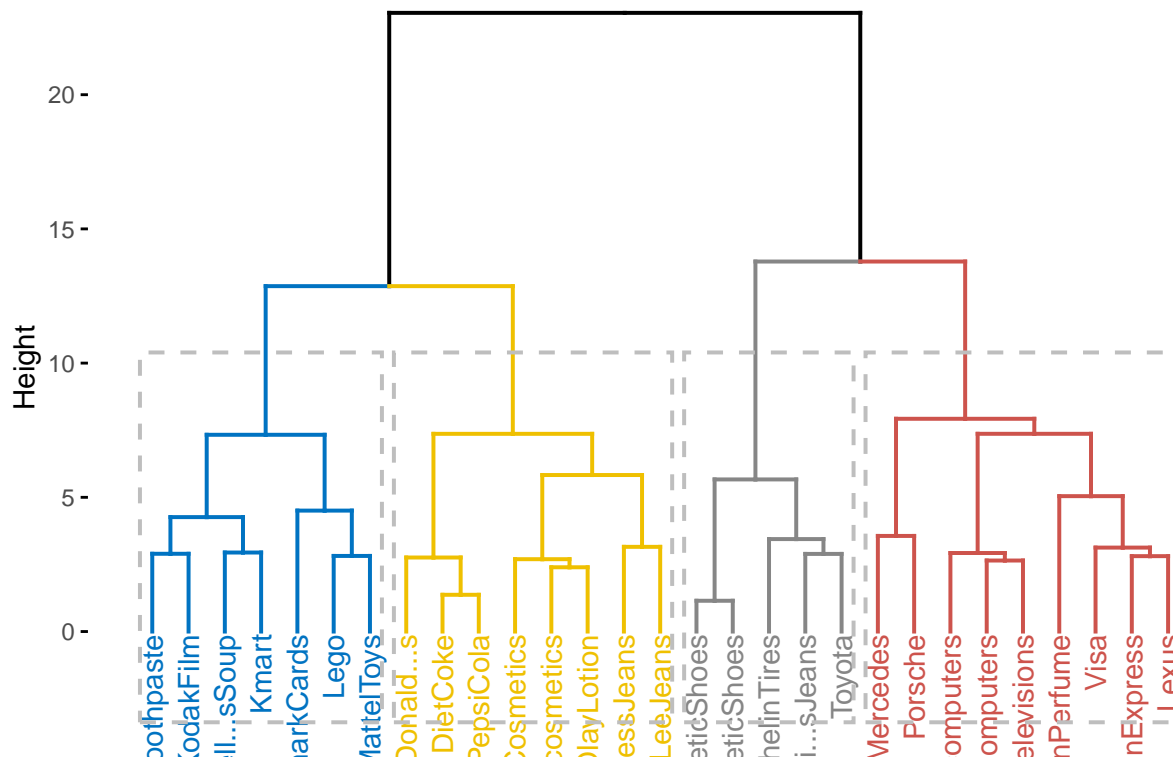



Didapatkan bahwa anggota tiap cluster yang dihasilkan ternyata tidak sama. Dengan cluster 1 beranggotakan 9 brand, cluster 2 beranggotakan 8 brand, cluster 3 beranggotakan 7 brand dan cluster 4 beranggotakan 5 brand.

Berikut adalah visualisasi anggota tiap cluster pada dendrogram yang dibedakan oleh warnanya:

```
fviz_dend(hc_agg1, k = 4, k_colors = "jco", rect = T,  
          main = "Ward Linkage Cluster")
```

Ward Linkage Cluster



Langkah selanjutnya adalah memasukkan hasil anggota clustering ke data brand dalam variabel “cluster.

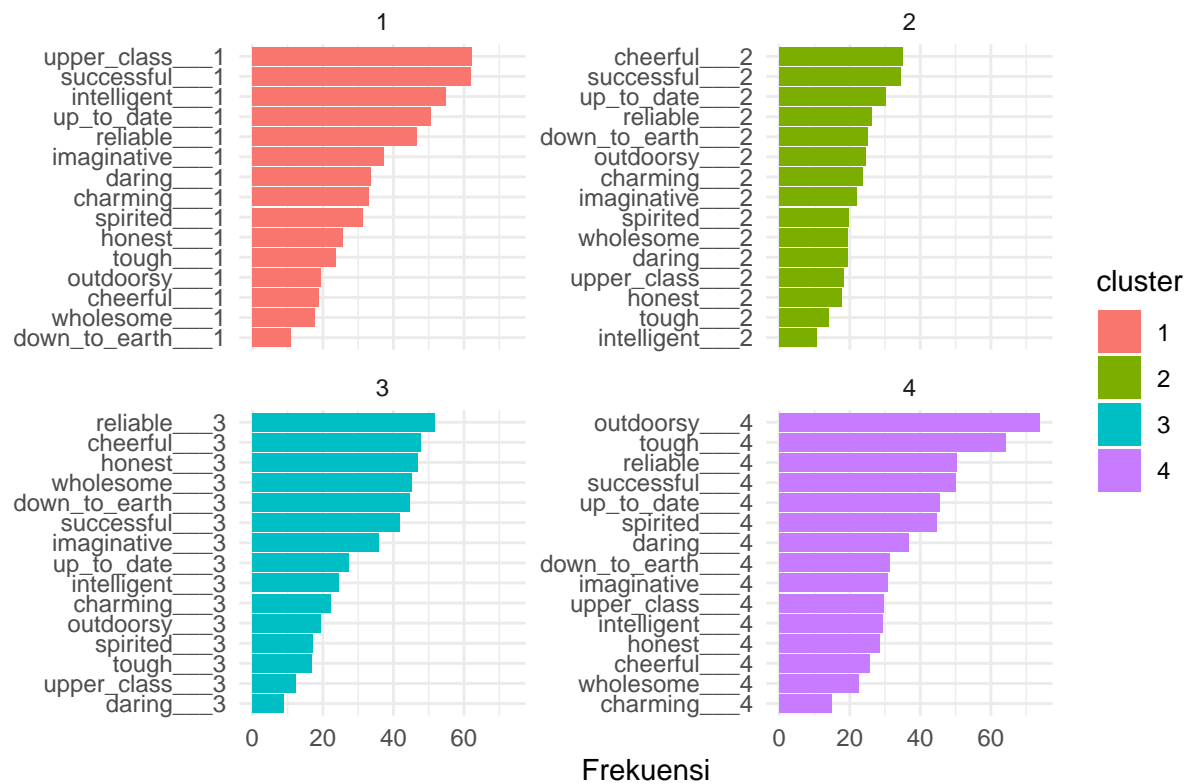
```
data_cluster1 <- data_brand %>%
  mutate(cluster = cluster)

## profiling cluster
brand_mean <-
data_cluster1 %>%
  group_by(cluster) %>%
  summarise_all(mean) %>%
  mutate_if(is.numeric, function(x) {round(x, 2)})
```

Untuk mempermudah dalam melakukan profiling cluster dan mengetahui karakteristik dari masing-masing cluster, berikut adalah visualisasi tingkat personality pada masing-masing cluster.

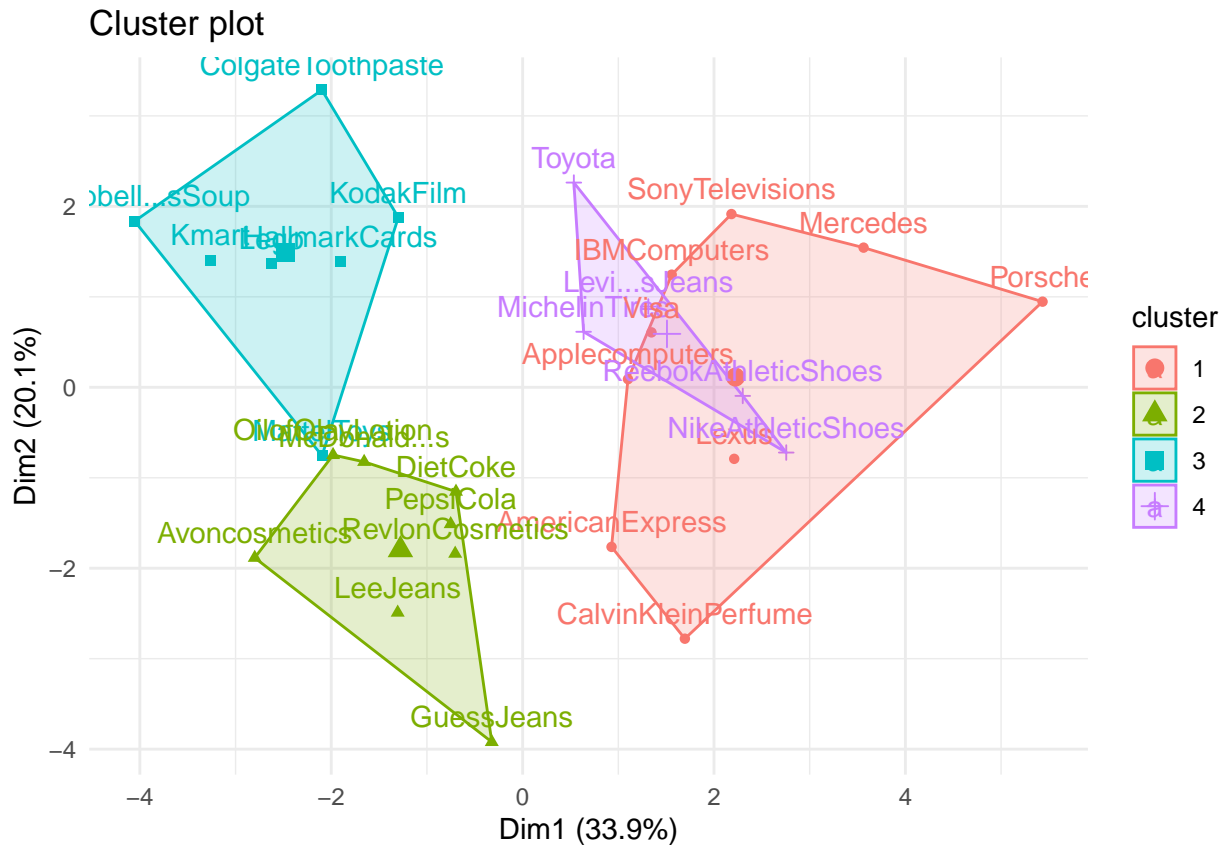
```
brand_mean %>%
  pivot_longer(-cluster) %>%
  mutate(cluster = as.factor(cluster)) %>%
  ggplot(aes(x = value, y = reorder(name, value), fill = cluster)) +
  geom_col(mapping = aes(x = value, y = tidytext::reorder_within(name, value, cluster))) +
  facet_wrap(~cluster, scales = "free_y") +
  theme_minimal() +
  labs(title = "Personality each Cluster",
       y = "",
       x = "Frekuensi")
```

Personality each Cluster



Berikut adalah visualisasi hasil clustering pada plot dua dimensi untuk melihat brand apa saja dalam satu cluster.

```
fviz_cluster(list(data = brand_scale, cluster = cluster)) +  
  theme_minimal()
```



Didapatkan bahwa:

- Personality yang paling tinggi di cluster 1 di banding cluster lainnya adalah successful, upper class, intelligent, dan up to date. Cluster ini berisi brand Sony Televisions, Mercedes, IBM Computers, Porsche, Apple, American Express, Calvin Klein Perfume.
- Tidak ada personality yang paling tinggi di cluster 2. Cluster ini berisi brand Olay, Revlon, Avoncos, Pepsi Cola, McDonalds, Guess Jeans.
- Personality yang dominan di cluster 3 adalah reliable, cheerful, honest dan wholesome. Cluster ini berisi brand Colgate Toothpaste, Campbessl's Soup, Kodak Film, Kmark.
- Personality yang dominan di cluster 4 adalah outdoorsy, tough, daring dan spirited. Cluster ini berisi brand Toyota, Levis's Jeans, Mechelin, Reebok dan Nike

5. Brand Clustering using Fuzzy C Means Clustering

Untuk melakukan Fuzzy C Means Clustering dapat digunakan fungsi `fanny()` dari package `cluster`, dengan parameter sebagai berikut:

- `x`: data untuk clustering
- `k`: banyak cluster yang akan dibentuk
- `metric`: metric yang digunakan untuk mengukur dissimilarity antar observasi

```
# clustering using funny()
fuz_clust <- fanny(x = brand_scale, k = 4, metric = "euclidean")
```

```
## Warning in fanny(x = brand_scale, k = 4, metric = "euclidean"): the memberships
## are all very close to 1/k. Maybe decrease 'memb.exp' ?
```

5.1 Cluster Exploration

```
# melihat derajat keanggotaan masing-masing cluster
fuz_clust$membership %>% head()
```

```
##           [,1] [,2] [,3] [,4]
## AmericanExpress 0.25 0.25 0.25 0.25
## Applecomputers  0.25 0.25 0.25 0.25
## Avoncosmetics    0.25 0.25 0.25 0.25
## CalvinKleinPerfume 0.25 0.25 0.25 0.25
## Campbell'sSoup   0.25 0.25 0.25 0.25
## ColgateToothpaste 0.25 0.25 0.25 0.25
```

Dari 6 data pertama, diketahui derajat keanggotaan setiap obesrvasi memiliki nilai yang sama yaitu sebesar 0.25. Sehingga dapat diartikan setiap observasi data memiliki peluang atau kesempatan yang sama untuk masuk ke dalam seluruh cluster yang terbentuk dan membuat hasil clustering bias. Hal ini dapat disebabkan oleh:

- Metric yang digunakan yaitu euclidian belum mampu membedakan setiap karakteristik dari masing-masing cluster yang akan dibentuk.
- Banyak k yang digunakan belum optimal

Sehingga kita akan mencoba mengganti metric dissimilarity terlebih dahulu menjadi square root dari euclidean distance atau SqEuclidean.

```
fuz_clust2 <- fanny(x = brand_scale, k = 4, metric = "SqEuclidean")
```

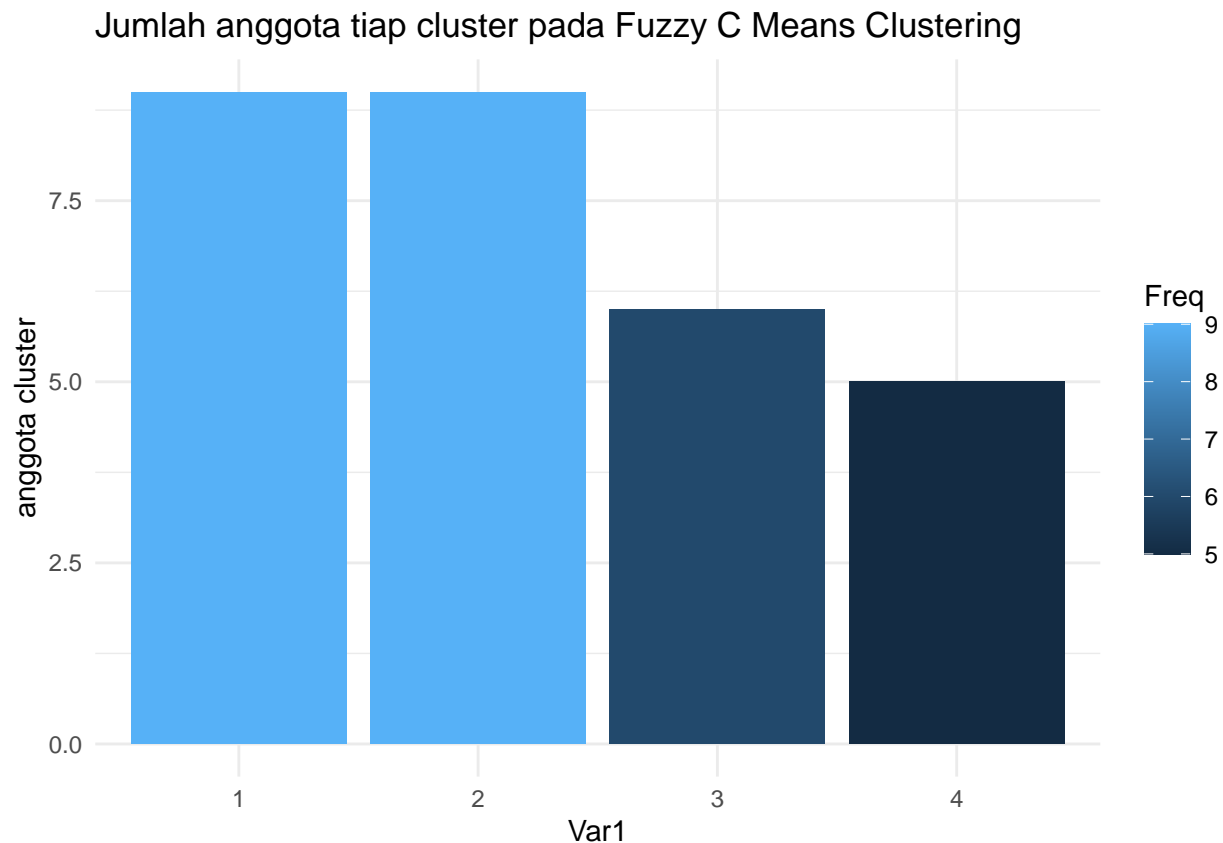
```
#cek membership
fuz_clust2$membership %>% head()
```

```
##           [,1]      [,2]      [,3]      [,4]
## AmericanExpress 0.3950083 0.2893952 0.1449438 0.17065268
## Applecomputers  0.4612579 0.2166803 0.1484940 0.17356779
## Avoncosmetics    0.1283338 0.5113433 0.2620218 0.09830104
## CalvinKleinPerfume 0.3188046 0.3090762 0.1517916 0.22032765
## Campbell'sSoup   0.1424282 0.2162236 0.5055515 0.13579674
## ColgateToothpaste 0.1499203 0.1630462 0.5554645 0.13156895
```

Sekarang derajat keanggotaan setiap observasi lebih terlihat jelas berpeluang besar masuk ke dalam cluster yang mana. Misalnya pada baris pertama, American Express cenderung akan masuk cluster 1 karena di memiliki derajat keanggotaan terbesar disana.

Berikut akan ditampilkan berapa banyak observasi data yang menjadi anggota tiap cluster:

```
table(fuz_clust2$clustering) %>%
  as.data.frame() %>%
  ggplot(mapping = aes(x = Var1, y = Freq, fill = Freq)) +
  geom_col() +
  theme_minimal() +
  labs(title = "Jumlah anggota tiap cluster pada Fuzzy C Means Clustering",
       y = "anggota cluster")
```



Didapatkan bahwa cluster 1 beranggotakan 9 brand, cluster 2 beranggotakan 9 brand, cluster 3 beranggotakan 6 brand dan cluster 4 beranggotakan 5 brand.

Langkah selanjutnya adalah memasukkan hasil anggota clustering ke data brand dalam variabel "cluster.

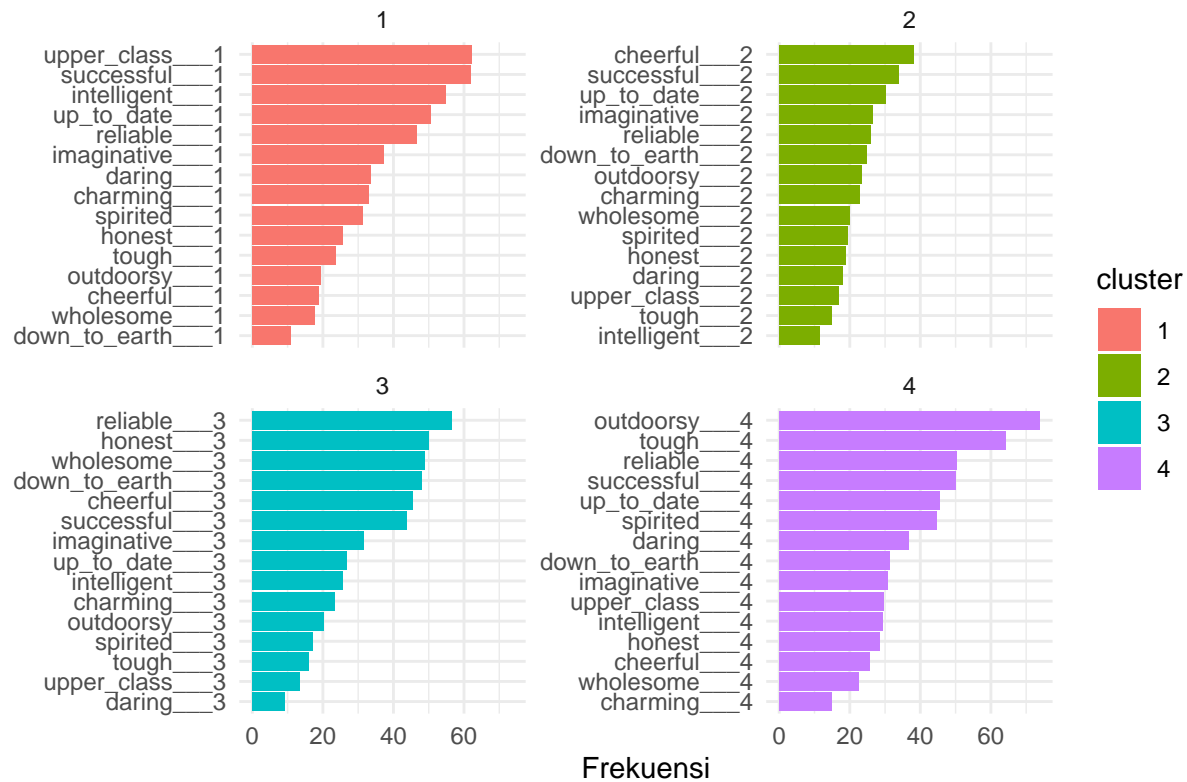
```
data_cluster2 <- data_brand %>%
  mutate(cluster = fuz_clust2$clustering)

# profiling cluster
brand_mean <-
data_cluster2 %>%
  group_by(cluster) %>%
  summarise_all(mean) %>%
  mutate_if(is.numeric, function(x) {round(x, 2)})
```

Untuk mempermudah dalam melakukan profiling cluster dan mengetahui karakteristik dari masing-masing cluster, berikut adalah visualisasi tingkat personality pada masing-masing cluster.

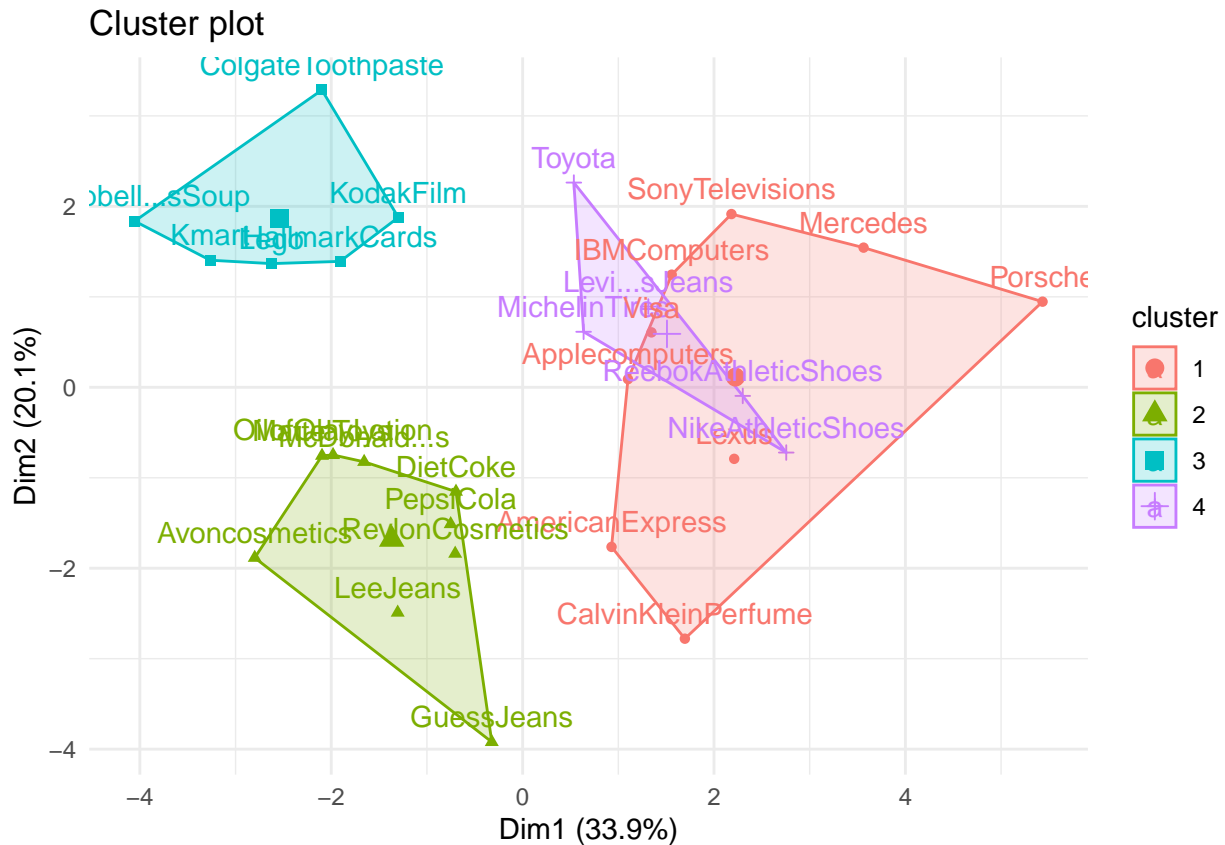
```
brand_mean %>%
  pivot_longer(-cluster) %>%
  mutate(cluster = as.factor(cluster)) %>%
  ggplot(aes(x = value, y = tidytext::reorder_within(name, value, cluster), fill = cluster)) +
  geom_col() +
  facet_wrap(~cluster, scales = "free_y") +
  theme_minimal() +
  labs(title = "Personality each Cluster",
       y = "",
       x = "Frekuensi")
```

Personality each Cluster



Berikut adalah visualisasi hasil clustering pada plot dua dimensi untuk melihat brand apa saja dalam satu cluster. N

```
fviz_cluster(list(data = brand_scale, clustering = fuz_clust2$clustering)) +  
  theme_minimal()
```



Didapatkan bahwa:

- Personality yang paling tinggi di cluster 1 di banding cluster lainnya adalah successful, upper class, intelligent, dan up to date. Cluster ini berisi brand Sony Televisions, Mercedes, IBM Computers, Porsche, Apple, American Express, Calvin Klein Perfume, Lexus, Visa.
- Tidak ada personality yang paling tinggi di cluster 2. Cluster ini berisi brand Olay, Revlon, Pepsi Cola, McDonalds, Guess Jeans, LeeJeans, Avoncosmetics, DietCoke, MattelToys.
- Personality yang dominan di cluster 3 adalah reliable, imaginative, cheerful, honest, down to earth dan wholesome. Cluster ini berisi brand Colgate Toothpaste, Campbessl's Soup, Kodak Film, Kmark, Lego, HallmarkCards.
- Personality yang dominan di cluster 4 adalah outdoorsy, tough, daring dan spirited. Cluster ini berisi brand Toyota, Levis's Jeans, Michelin Tires, Reebok dan Nike

6. Cluster Validation

Untuk menentukan keakuratan hasil clustering dapat digunakan Dunn Index. Dunn index menilai seberapa rapat anggota dalam suatu cluster dengan variansi yang kecil dan seberapa terpisah cluster yang satu dengan yang lainnya.

```
# Dunn index
stats_fuz <- cluster.stats(sqrt(dist(brand_scale)), clustering = fuz_clust2$clustering)
stats_agg <- cluster.stats(dist(brand_scale),clustering = cluster)

cbind(Dunn_agg = stats_agg$dunn, Dunn_fuz = stats_fuz$dunn)
```

```
##      Dunn_agg  Dunn_fuz
```



```
## [1,] 0.4662699 0.6423549
```

Didapatkan bahwa Dunn Index dari hasil clustering Fuzzy C Means lebih tinggi dibanding hierarchical clustering, dengan nilai 0.6423549.

7. Conclusion

Untuk mengetahui kelompok segmentasi brand berdasarkan personality nya, kita dapat menggunakan metode clustering. Dari analisis yang telah dilakukan, didapatkan bahwa berdasarkan Dunn Index, jenis clustering yang lebih baik dalam membuat kelompok segmentasi brand berdasarkan personality nya adalah Fuzzy C Means. Clustering Fuzzy C Means membuat 4 cluster dengan rincian sebagai berikut: cluster 1 adalah brand Sony Televisions, Mercedes, IBM Computers, Porsche, Apple, American Express, Calvin Klein Perfume, Lexus, Visa; cluster 2 adalah brand brand Olay, Revlon, Pepsi Cola, McDonalds, Guess Jeans, LeeJeans, Avoncosmetics, DietCoke, MattelToys; cluster 3 adalah brand Colgate Toothpaste, Campbessl's Soup, Kodak Film, Kmark, Lego, HallmarkCards; cluster 4 adalah Toyota, Levis's Jeans, Michelin Tires, Reebok dan Nike.

Dengan mengetahui kelompok-kelompok brand yang memiliki personality serupa, perusahaan dapat lebih aware dengan persaingan yang ketat serta dapat menyusun strategi pemasaran yang tepat.