

LAPORAN FINAL PROJECT 2

PENERAPAN METODE K-MEANS KLUSTERING
BERDASARKAN NEGARA UNTUK MENGETAHUI
STRATEGI PENJUALAN YANG DITERAPKAN DI
TIAP NEGARA

Humaira Nur Pradani	(05211640000011)
Nevada Veterino	(05211640000096)
Rizki Ahmad Fauzi	(05211640000111)

DAFTAR ISI

DAFTAR ISI.....	2
BAB I : PENDAHULUAN.....	3
1.1. LATAR BELAKANG.....	3
1.2. RUMUSAN MASALAH.....	3
BAB II : TINJAUAN PUSTAKA.....	4
2.1. DATA MINING	4
2.2. DATA CLUSTERING	4
2.3. ALGORITMA K-MEANS	5
BAB III : METODOLOGI	7
BAB IV : IMPLEMENTASI.....	9
4.1. DATA PRE-PROCESSING	9
4.2. K-MEANS CLUSTERING.....	12
BAB V : HASIL DAN PEMBAHASAN.....	15
5.1. PRE-PROCESSING DATA	15
5.2. CLUSTERING PERTAMA.....	17
5.3. CLUSTERING KEDUA.....	20
5.4. CLUSTER GABUNGAN.....	23
5.5. PENENTUAN STRATEGI	23
5.6. VISUALISASI	26
6.1. KESIMPULAN	28
BAB VI : KESIMPULAN DAN SARAN	28
6.2. SARAN	30
REFERENSI	31

1.1. LATAR BELAKANG

Saat ini, perkembangan teknologi semakin pesat. Perkembangan tersebut dapat dilihat dengan majunya teknologi yang diciptakan oleh orang-orang yang mahir pada bidang masing-masing. Teknologi yang diciptakan diharapkan mampu membuat mudah hidup manusia dalam melakukan kegiatan sehari-hari. Dari teknologi yang telah diciptakan terdapat teknologi yang telah digunakan sehari-hari namun banyak orang belum mengetahui bahwa teknologi tersebut bermanfaat bagi manusia. Teknologi tersebut yaitu *Online Retail*.

Online Retail merupakan transaksi yang didapatkan dari transaksi jual-beli yang dilakukan oleh manusia secara online atau melalui internet. Seiring berjalannya waktu, *Online Retail* makin berkembang dengan cepat dan semakin banyak bermunculan bermacam-macam jenis *Online Retail*. Dengan berkembangnya *Online Retail* akan mendapatkan pelanggan yang banyak jika sebuah *Online Retail* sudah berkembang pesat. Dengan bertambahnya jumlah pelanggan dari sebuah *Online Retail* maka diperlukan suatu strategi penjualan agar dapat mempertahankan jumlah pelanggan yang ada dan menambah jumlah pelanggan.

Dalam merencanakan strategi dalam penjualan, dibutuhkan pengolahan data yang diambil dari seluruh transaksi yang terjadi pada *Online Retail* tersebut. Diperlukan analisis clustering data untuk mengelompokkan data dengan karakteristik yang sama. Dalam hal ini akan mencari tahu pola hubungan antara data jumlah customer, total penjualan, dan jumlah barang yang terjual. Dengan begitu dapat diketahui negara mana saja yang merupakan satu kluster dan dapat menentukan strategi penjualannya.

1.2. RUMUSAN MASALAH

Adapun rumusan masalah pada penelitian kali ini adalah sebagai berikut.

- Berapa jumlah cluster yang terbentuk pada hasil k-means clustering yang dilakukan?
- Apa saja karakteristik masing-masing cluster yang terbentuk?
- Bagaimana strategi penjualan yang perlu diterapkan pada *Online Retail*?
- Apakah terdapat persamaan pada penerapan strategi antar negara?

BAB II : TINJAUAN PUSTAKA

Sebelum melakukan penelitian lebih lanjut, perlu dilakukan studi literatur terlebih dahulu guna memperjelas ilmu yang akan digunakan pada penelitian kali ini. Berikut hasil dari studi literatur yang dilakukan.

2.1. DATA MINING

Menurut Gartner Group, Data mining adalah suatu proses menemukan hubungan yang berarti, pola, dan kecenderungan dengan memeriksa dalam sekumpulan besar data yang tersimpan dalam penyimpanan dengan menggunakan teknik pengenalan pola seperti teknik statistik dan matematika [1].

Menurut Iarose [2], data mining dibagi menjadi beberapa kelompok berdasarkan tugas yang dapat dilakukan, yaitu :

1. *Deskripsi* : Terkadang peneliti dan analisis secara sederhana ingin mencoba mencari cara untuk menggambarkan pola dan kecenderungan yang terdapat dalam data.
2. *Estimasi* : Model dibangun menggunakan record lengkap yang menyediakan nilai dari variabel target sebagai nilai prediksi. Selanjutnya, pada peninjauan berikutnya estimasi nilai dari variabel target dibuat berdasarkan nilai variabel prediksi.
3. *Prediksi* : Prediksi hampir sama dengan klasifikasi dan estimasi, kecuali bahwa dalam prediksi nilai dari hasil akan ada dimasa mendatang.
4. *Klasifikasi* : Dalam klasifikasi, terdapat variabel kategori. Sebagai contoh, Penggolongan pendapatan dapat dipisahkan dalam tiga kategori, yaitu pendapatan tinggi, pendapatan sedang, dan pendapatan rendah.
5. *Pengklusteran* : Pengklusteran merupakan pengelompokan record, pengamatan, atau memperhatikan dan membentuk kelas objek-objek yang memiliki kemiripan. Cluster adalah kumpulan record yang memiliki kemiripan suatu dengan yang lainnya dan memiliki ketidakmiripan dengan record dengan cluster lain.
6. *Asosiasi* : Tugas asosiasi dalam data mining adalah menemukan atribut yang muncul dalam suatu waktu. Dalam dunia bisnis lebih umum disebut analisis keranjang belanja.

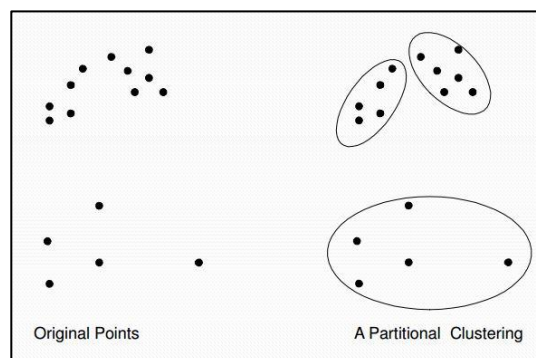
2.2. DATA CLUSTERING

Clustering adalah metode penganalisaan data, yang sering dimasukkan sebagai salah satu metode Data Mining, yang tujuannya adalah untuk mengelompokkan data dengan karakteristik yang sama ke suatu 'wilayah' yang sama dan data dengan karakteristik yang berbeda ke 'wilayah' yang lain. Ada beberapa pendekatan yang digunakan dalam mengembangkan metode clustering. Dua

pendekatan utama adalah clustering dengan pendekatan partisi dan clustering dengan pendekatan hirarki. Clustering dengan pendekatan partisi atau sering disebut dengan partition-based clustering mengelompokkan data dengan memilah-milah data yang dianalisa ke dalam clustercluster yang ada. Clustering dengan pendekatan hirarki atau sering disebut dengan hierarchical clustering mengelompokkan data dengan membuat suatu hirarki berupa dendogram dimana data yang mirip akan ditempatkan pada hirarki yang berdekatan dan yang tidak pada hirarki yang berjauhan. Di samping kedua pendekatan tersebut, ada juga clustering dengan pendekatan automatic mapping (Self-Organising Map/SOM).

2.3. ALGORITMA K-MEANS

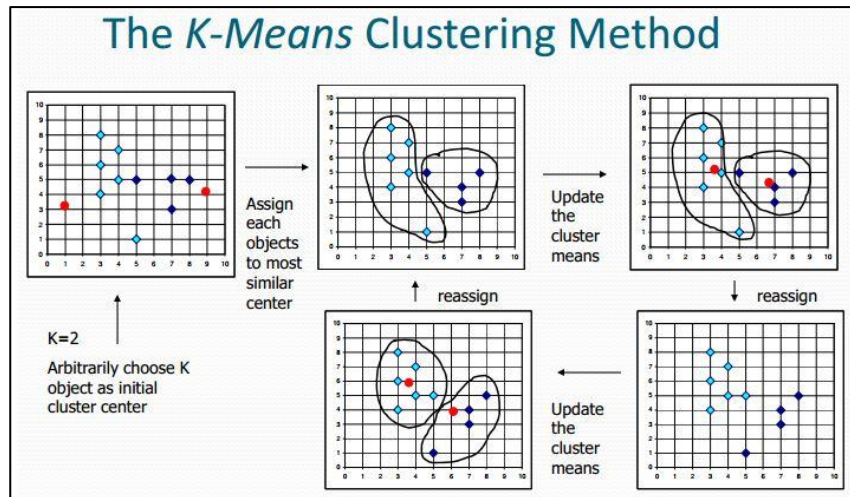
K-Means Clustering adalah suatu metode penganalisaan data atau metode Data Mining yang melakukan proses pemodelan tanpa supervisi (unsupervised) dan merupakan salah satu metode yang melakukan pengelompokan data dengan sistem partisi. Terdapat dua jenis data clustering yang sering dipergunakan dalam proses pengelompokan data yaitu Hierarchical dan Non-Hierarchical, dan K-Means merupakan salah satu metode data clustering non-hierarchical atau Partitional Clustering.



Gambar 1 Partitional Clustering

Metode K-Means Clustering berusaha mengelompokkan data yang ada ke dalam beberapa kelompok, dimana data dalam satu kelompok mempunyai karakteristik yang sama satu sama lainnya dan mempunyai karakteristik yang berbeda dengan data yang ada di dalam kelompok yang lain.

Dengan kata lain, metode K-Means Clustering bertujuan untuk meminimalisasikan objective function yang diset dalam proses clustering dengan cara meminimalkan variasi antar data yang ada di dalam suatu cluster dan memaksimalkan variasi dengan data yang ada di cluster lainnya.



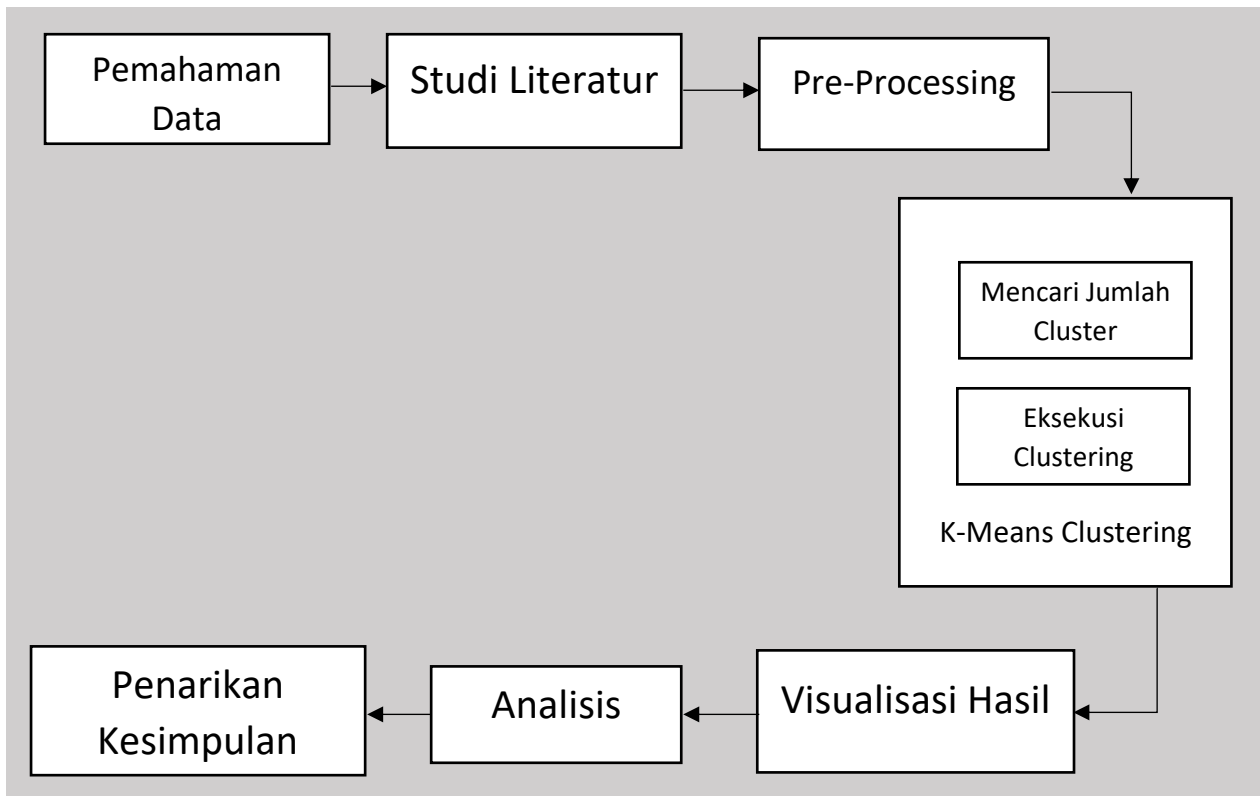
Gambar 2 Ilustrasi Proses K-Means Clustering

Data clustering menggunakan metode K-Means Clustering ini secara umum dilakukan dengan algoritma dasar sebagai berikut:

1. Tentukan jumlah cluster
2. Alokasikan data ke dalam cluster secara random
3. Hitung centroid/rata-rata dari data yang ada di masing-masing cluster
4. Alokasikan masing-masing data ke centroid/rata-rata terdekat
5. Kembali ke Step 3, apabila masih ada data yang berpindah cluster atau apabila perubahan nilai centroid, ada yang di atas nilai threshold yang ditentukan atau apabila perubahan nilai pada objective function yang digunakan di atas nilai threshold yang ditentukan

Beberapa distance space dapat diimplementasikan untuk menghitung jarak (distance) antara data dan centroid termasuk di antaranya Manhattan/City Block Distance, Euclidean Distance dan Minkowski Distance. Tetapi secara umum distance space yang sering digunakan adalah Manhattan dan Euclidean. Euclidean sering digunakan karena penghitungan jarak dalam distance space ini merupakan jarak terpendek yang bisa didapatkan antara dua titik yang diperhitungkan, sedangkan Manhattan sering digunakan karena kemampuannya dalam mendeteksi keadaan khusus seperti keberadaan outliers dengan lebih baik[3].

Pada studi kasus yang diambil yaitu Penerapan metode k-means untuk clustering berdasarkan negara dengan mempertimbangkan jumlah customer, total penjualan, dan jumlah barang yang terjual untuk mengetahui strategi penjualan yang dapat diterapkan pada tiap clusternya. Seperti judul dari studi kasus, metodologi yang digunakan dalam tugas ini yaitu menggunakan metodologi k-means.



Grafik 1 Metodologi

1. Pemahaman Data

Dalam pemahaman data ini dilakukan identifikasi mengenai data apa saja yang akan digunakan untuk diolah sehingga mendapatkan hasil yang sesuai dengan tujuan dilakukan pengolahan data. Mulai dari penentuan kolom, baris, dan sebagainya.

2. Studi Literatur

Pada tahap ini dilakukan studi literatur mengenai data yang telah didapat dengan mencari referensi yang relevan dengan rumusan masalah. Dengan begitu akan terkoleksi referensi yang relevan dengan perumusan masalah. Dengan studi literatur dapat memperkuat permasalahan serta dapat digunakan sebagai dasar teori dalam melakukan studi.

3. Pre-processing data

Pada tahap ini dilakukan pembersihan data *Online Retail* dari data null, data yang negatif. Kemudian data-data yang null dan data negative tersebut diubah menjadi factor. Selanjutnya dalam tahap PreProcessing, dilakukan penghapusan data-data yang sekiranya tidak dibutuhkan dalam pengolahan data. Dalam menghapus data-data yang tidak diperlukan, digunakan fungsi untuk mendeteksi kata kunci yang akan dihapus. Kemudian dilakukan penambahan sebuah *derived data* bernama *TotalPrice* untuk menyatakan jumlah total transaksi yang dilakukan. Lalu dilakukan agregasi data pada data *Online Retail* yaitu pada negara, jumlah pelanggan, dan jumlah quantity.

4. K-Means Clustering

Pada K-Means Clustering dilakukan pencarian jumlah kluster yang dibutuhkan dalam pengolahan data. Pada K-Means juga dilakukan pengelompokan terhadap data yang memiliki karakteristik yang sama satu sama lain dan mengelompokkan ke kluster lain jika terdapat perbedaan karakteristik pada data. Dengan kata lain, metode K-Means Clustering bertujuan untuk meminimalisasikan objective function yang diset dalam proses clustering dengan cara meminimalkan variasi antar data yang ada di dalam suatu cluster dan memaksimalkan variasi dengan data yang ada di cluster lainnya.

5. Visualisasi Hasil

Setelah melakukan K-Means Clustering maka dilakukan visualisasi hasil yang akan menggambarkan hasil yang didapat setelah dilakukannya clustering dengan algoritma K-Means.

6. Analisis

Dari hasil yang telah divisualisasikan, maka dilakukan analisa terhadap hasil visualisasi yang bertujuan untuk mencari jawaban yang diperlukan dari hasil yang telah divisualisasikan terhadap permasalahan yang ada. Dalam kasus ini maka akan dilakukan analisa mengenai strategi penjualan yang cocok diterapkan pada hasil yang telah dibuat.

7. Penarikan Kesimpulan

Setelah melakukan analisis pada hasil, maka dilakukan penarikan kesimpulan dari hasil analisis tersebut. Kesimpulan ini akan menjawab rumusan masalah yang dimana akan diketahui bagaimana strategi penjualan yang akan dilakukan pada tiap kluster negara.

BAB IV : IMPLEMENTASI

Pada bab ini dijelaskan bagaimana kode-kode digunakan selama pengerjaan untuk melakukan analisis clustering.

4.1. DATA PRE-PROCESSING

Sebelum melakukan menerapkan algoritma pada data Online Retail, dilakukan tahap *pre-processing*. Pada tahap ini dilakukan proses pembersihan pada data. Pada proses *pre-processing* dilakukan untuk memastikan data yang akan diolah merupakan data yang baik. Pada *pre-processing* ini melakukan pembersihan jika terdapat data yang tidak lengkap, membuang duplikasi data, dan memperbaiki kesalahan jika terdapat kesalahan pada data.

Untuk menghapus data null, hilang (missing) kita perlu menggunakan fungsi bernama `!is.na`. Sedangkan untuk menghilangkan data yang negatif, gunakan simbol `">=0"`.

MENGHAPUS DATA NULL, HILANG DAN NEGATIF

```
EcommDataClean <- Ecommdata[which(Ecommdata$Quantity>=0 &
!is.na(Ecommdata$Quantity) & Ecommdata$UnitPrice >=0 &
!is.na(Ecommdata$UnitPrice) & !is.na(Ecommdata$CustomerID)),]
```

Selanjutnya variabel-variabel tersebut diubah menjadi faktor dengan fungsi `factor()`. Variabel-variabel yang diubah adalah InvoiceNo, StockCode, InvoiceDate, CustomerID, Country.

CONVERT VARIABEL MENJADI FAKTOR

```
EcommDataClean$InvoiceNo<- factor(EcommDataClean$InvoiceNo)
EcommDataClean$StockCode<- factor(EcommDataClean$StockCode)
EcommDataClean$CustomerID<- factor(EcommDataClean$CustomerID)
EcommDataClean$Country<- factor(EcommDataClean$Country)
EcommDataClean$InvoiceDate<- as.Date(EcommDataClean$InvoiceDate,
'%m/%d/%Y %H:%M')
```

Selanjutnya adalah menghapus data-data yang tidak dibutuhkan dengan menggunakan bantuan library `stringr` pada fungsi `str_detect` untuk mendeteksi adanya kata kunci dari sebuah tuple yang harus dihapus. Kata kunci yang digunakan adalah Away, Charge, Fee, Fault, Sales, Adjust, Counted, Incorrect, Wrong, Crushed, Lost, Damage, Found, Thrown, Smashed, \\\?, Broken, Barcode, Returned, Mailout, Delivery, Mix up, Mouldy, Put Aside, Error, Destroyed, Rusty, Manual, Amazon, Postage, Pads, dan Bank. Baris yang mengandung kata-kata tersebut dianggap tidak dibutuhkan dan tidak akan diikuti untuk diolah selanjutnya. Selain itu, ada juga penghilangan data Carriage

MENGHAPUS DATA YANG TIDAK DIBUTUHKAN

```
library(stringr)
isUndesirable = function(x) {
  str_detect(toupper(x), "AWAY") | str_detect(toupper(x), "CHARGES") |
  str_detect(toupper(x), "FEE") | str_detect(toupper(x), "FAULT")
  str_detect(toupper(x), "SALES") | str_detect(toupper(x), "ADJUST") |
```

```

str_detect(toupper(x), "COUNTED") |
str_detect(toupper(x), "INCORRECT") |
str_detect(toupper(x), "WRONG") | str_detect(toupper(x), "LOST") |
str_detect(toupper(x), "CRUSHED") | str_detect(toupper(x), "DAMAGE") |
str_detect(toupper(x), "FOUND") | str_detect(toupper(x), "THROWN") |
str_detect(toupper(x), "SMASHED") |
str_detect(toupper(x), "\\?") |
str_detect(toupper(x), "BROKEN") | str_detect(toupper(x), "BARCODE") |
str_detect(toupper(x), "RETURNED") |
str_detect(toupper(x), "MAILOUT") | str_detect(toupper(x), "DELIVERY") |
str_detect(toupper(x), "MIX UP") | str_detect(toupper(x), "MOULDY") |
str_detect(toupper(x), "PUT ASIDE") | str_detect(toupper(x), "ERROR") |
str_detect(toupper(x), "DESTROYED") | str_detect(toupper(x), "RUSTY") |
str_detect(toupper(x), "MANUAL") | str_detect(toupper(x), "AMAZON") |
str_detect(toupper(x), "POSTAGE") |
str_detect(toupper(x), "PADS") | str_detect(x, "Bank")
}

EcommDataClean =
EcommDataClean[which(!isUndesirable(as.character(EcommDataClean$Description))),]

EcommDataClean =
EcommDataClean[which(!startsWith(as.character(EcommDataClean$StockCode),
"C")),]

```

Selanjutnya, menambahkan sebuah *derived data* bernama TotalPrice untuk menyatakan jumlah total transaksi yang dilakukan. TotalPrice didapatkan dari perkalian antara kolom UnitPrice dan Quantity.

MENAMBAH KOLOM TOTAL PRICE
<pre>EcommDataClean\$TotalPrice <- EcommDataClean\$UnitPrice * EcommDataClean\$Quantity</pre>

Untuk atribut pertama yang dibutuhkan untuk analisa, yaitu total penjualan tiap negara, maka dilakukan agregasi pada data penjualan berdasarkan negara dengan menggunakan fungsi aggregate() diikuti dengan 'by=list' negara.

MELAKUKAN AGREGAT TERHADAP NEGARA
<pre>AggCountry <- aggregate(EcommDataClean\$TotalPrice, by=list(EcommDataClean\$Country), FUN=sum) names(AggCountry) <- c("Country", "Total Sales")</pre>

Sedangkan untuk mengetahui total dari pelanggan pada tiap negaranya, yaitu atribut yang kedua, gunakan library plyr dan dplyr. Dimana, kita pertama-tama mengambil kolom ke-7(Country) dan ke-8(CustomerID) pada data Online Retail kemudian fungsi ddply dijalankan untuk menghitung frekuensi untuk masing-masing transaksi yang dilakukan pelanggan pada tiap negara.

MELAKUKAN AGREGAT JUMLAH PELANGGAN
<pre>library(plyr) library(dplyr)</pre>

```
CountryCustomer <- select(EcommDataClean, c(7,8))
count_CustID <- ddply(CountryCustomer, .(CountryCustomer$Country,
CountryCustomer$CustomerID), nrow)
names(count_CustID) <- c("Country", "CustomerID", "Frequency")

CountryCustomer <- select(count_CustID, c(1))
count_CustID <- ddply(CountryCustomer, .(CountryCustomer$Country), nrow)
names(count_CustID) <- c("Country", "Number of Unique Customer")
CountryCustomer <- count_CustID
```

Setelah kedua dataframe tersebut terbentuk, yaitu CountryCustomer dan AggCountry, selanjutnya keduanya digabungkan dengan fungsi merge() sebagai berikut.

PENGGABUNGAN CountryCustomer & AggCountry
AggCountry <- merge(AggCountry, CountryCustomer, by.x = "Country", by.y = "Country")

Untuk atribut ketiga yang dibutuhkan, yaitu total kuantitas barang yang terjual di tiap negara, gunakan kembali fungsi aggregate() sebagai berikut. Agregat dilakukan dengan melibatkan kolom Quantity dengan kolom Country. Names() digunakan untuk memberikan nama kolom pada data frame yang telah dibuat.

MELAKUKAN AGREGAT JUMLAH QUANTITY
AggQuantity <- aggregate(EcommDataClean\$Quantity, by=list(EcommDataClean\$Country), FUN=sum) names(AggQuantity) <- c("Country", "Total Quantity")

Selanjutnya, setelah ketiga atribut berhasil terbentuk, ketiganya dilakukan penggabungan kembali dengan fungsi merge() sebagai berikut. Kedua atribut sebelumnya telah digabungkan pada tabel AggCountry, sehingga merge dilakukan pada tabel AggCountry dan AggQuantity.

MENGGABUNGAN AggCountry & AggQuantity
AggCountry <- merge(AggCountry, AggQuantity, by.x = "Country", by.y = "Country")

Untuk melihat hasil tabel yang terbentuk, gunakan fungsi View().

MELIHAT DATAFRAME AKHIR YANG TERBENTUK
View(AggCountry)

Agar data siap untuk dicluster, kolom negara perlu dihilangkan sehingga hasil data hanya berupa matrix angka. Sehingga, nama negara diganti menjadi nama tiap barisnya dengan menggunakan fungsi row.names() sebagai berikut. Syntax yang dijalankan adalah sebagai berikut.

MENGANTI BARIS & KOLOM DATA NEGARA
row.names(AggCountry) <- AggCountry\$Country AggCountry <- AggCountry[,c(2:4)]

Data kemudian dilakukan normalisasi agar persebaran data lebih seimbang dan tidak berat pada data-data tertentu saja. Normalisasi yang digunakan adalah normalisasi zscore, cara melakukan normalisasi tersebut yaitu dengan menggunakan fungsi `scale()`.

NORMALISASI DATA
<code>ZAggCountry <- as.data.frame(scale(AggCountry[1:3]))</code>

Setelah melakukan semua tahap diatas, maka data akan siap dilakukan clustering.

4.2. K-MEANS CLUSTERING

Hal yang pertama kali dilakukan adalah mencari nilai optimal k dengan menggunakan elbow method, yaitu dengan syntax sebagai berikut.

MENCARI NILAI OPTIMAL K PERTAMA
<pre>set.seed(123) k.max <- 15 wss <- sapply(1:k.max, function(k){kmeans(ZAggCountry, k, nstart=50, iter.max = 15) \$tot.withinss}) wss plot(1:k.max, wss, type="b", pch = 19, frame = FALSE, xlab="Number of clusters K", ylab="Total within-clusters sum of squares")</pre>

Selanjutnya dilakukan clustering dengan syntax `kmeans()`, namun sebelumnya dengan menggunakan syntax `set.seed` sebagai berikut.

K-MEANS CLUSTERING PERTAMA
<pre>set.seed(20) CountryCluster <- kmeans(ZAggCountry, 2, nstart = 20)</pre>

Untuk memanggil anggota cluster, panggil kolom cluster pada hasil `kmeans` yang telah terbentuk dengan syntax `$cluster`.

MENAMPILKAN ANGGOTA CLUSTER PERTAMA
<code>CountryCluster\$cluster</code>

Karakteristik tiap cluster dapat digambarkan pada plot dengan menggunakan syntax `plot()` dengan bantuan beberapa fungsi rekursi (`for`) dan `text()` untuk memunculkan label sebagai berikut.

MENAMPILKAN KARAKTERISTIK TIAP CLUSTER PERTAMA
<pre>library(ggplot2) plot(c(0), xaxt = 'n', ylab = "", type = "l",</pre>

```

ylim = c(min(CountryCluster$centers), max(CountryCluster$centers)),
xlim = c(0, 3))

axis(1, at = c(1:3), labels = names(ZAggCountry))

for (i in c(1:2))
  lines(CountryCluster$centers[i,], lty = i, lwd = 2, col = ifelse(i %in%
c(1,3,5),
                                "black", "blue"))

text(x = 0.5, y = CountryCluster$centers[, 1], labels = paste("Cluster",
c(1:2)))

```

Langkah selanjutnya adalah membuat dataframe baru untuk dengan tidak melibatkan negara United Kingdom. Penjelasan lebih lanjut terhadap penghilangan

MENGHILANGKAN NEGARA UNITED KINGDOM

```

AggCountryNoUK <- AggCountry[-c(35),]
View(AggCountryNoUK)
row.names(AggCountryNoUK) <- AggCountryNoUK$Country
AggCountryNoUK <- AggCountryNoUK[,c(2:4)]

```

Lakukan kembali normalisasi data terhadap dataframe baru yang terbentuk tanpa negara United Kingdom.

MELAKUKAN NORMALISASI DATA TANPA NEGARA UK

```

ZAggCountryNoUK <- as.data.frame( scale(AggCountryNoUK[1:3] ))
View(ZAggCountryNoUK)

```

Selanjutnya mencari kembali nilai optimal jumlah cluster.

MENCARI NILAI ORTIMAL K KEDUA

```

set.seed(123)
k.max <- 15
wss <- sapply(1:k.max,
              function(k){kmeans(ZAggCountryNoUK, k, nstart=50,iter.max =
15 )$tot.withinss})
wss
plot(1:k.max, wss,
     type="b", pch = 19, frame = FALSE,
     xlab="Number of clusters K",
     ylab="Total within-clusters sum of squares")

```

Jalankan k-means clustering yang kedua sebagai berikut.

K-MEANS CLUSTERING KEDUA

```

set.seed(20)
CountryCluster <- kmeans(ZAggCountryNoUK, 4, nstart = 20)

```

Untuk menampilkan anggota cluster, gunakan syntax berikut.

MENAMPILKAN ANGGOTA CLUSTER KEDUA
CountryCluster\$cluster

Sama seperti sebelumnya, untuk menampilkan karakteristik tiap cluster, digunakan syntax seperti berikut.

MENAMPILKAN KARAKTERISTIK TIAP CLUSTER KEDUA
<pre>plot(c(0), xaxt = 'n', ylab = "", type = "l", ylim = c(min(CountryCluster\$centers), max(CountryCluster\$centers)), xlim = c(0, 3)) axis(1, at = c(1:3), labels = names(ZAggCountryNoUK)) for (i in c(1:4)) lines(CountryCluster\$centers[i,], lty = i, lwd = 2, col = ifelse(i %in% c(1,3,5), "black", "blue")) text(x = 0.5, y = CountryCluster\$centers[, 1], labels = paste("Cluster", c(1:4)))</pre>

Untuk membuat visualisasi data, digunakan visualisasi 3D karena data yang digunakan berjumlah 3 atribut. Syntax yang digunakan adalah sebagai berikut.

VISUALISASI DATA 3D
<pre>library(rgl) library(car) #2 cluster(dengan menggunakan k-means versi 2 cluster) scatter3d(x = ZAggCountry\$`Total Sales`, y = ZAggCountry\$`Number of Unique Customer`, z = ZAggCountry\$`Total Quantity`, groups = as.factor(CountryCluster\$cluster), surface=FALSE) rglwidget() #5 cluster(dengan menggunakan k-means versi 5 cluster) scatter3d(x = ZAggCountry\$`Total Sales`, y = ZAggCountry\$`Number of Unique Customer`, z = ZAggCountry\$`Total Quantity`, groups = as.factor(CountryCluster\$cluster), surface=FALSE) rglwidget() #4 cluster tanpa UK(dengan menggunakan k-means versi 4 cluster tanpa UK) scatter3d(x = ZAggCountryNoUK\$`Total Sales`, y = ZAggCountryNoUK\$`Number of Unique Customer`, z = ZAggCountryNoUK\$`Total Quantity`, groups = as.factor(CountryCluster\$cluster), surface=FALSE) rglwidget()</pre>

BAB V : HASIL DAN PEMBAHASAN

Pada bab ini, dijelaskan bagaimana hasil dan analisa dari hasil yang didapatkan setelah menjalankan kode-kode pada BAB IV : Pengerjaan.

5.1. PRE-PROCESSING DATA

Pada tahap pertama pre-processing data, data dibersihkan dari data-data yang tidak diinginkan seperti nilai NA (null), negatif, dan data item yang tidak digunakan seperti Bank Charges, Manual, Postage, Carriage, dll.

DATA CLEANSING										
Sebelum						Sesudah				
InvoiceNo	StockCode	Description	Quantity	InvoiceDate		InvoiceNo	StockCode	Description	Quantity	InvoiceDate
1 541431	23166	MEDIUM CERAMIC TOP STORAGE JAR	74215	1/18/11 10:01		1 541431	23166	MEDIUM CERAMIC TOP STORAGE JAR	74215	0011-01-18
2 C541433	23166	MEDIUM CERAMIC TOP STORAGE JAR	-74215	1/18/11 10:17		3 537626	85116	BLACK CANDELABRA T-LIGHT HOLDER	12	0010-12-07
3 537626	85116	BLACK CANDELABRA T-LIGHT HOLDER	12	12/7/10 14:57		4 537626	22375	AIRLINE BAG VINTAGE JET SET BROWN	4	0010-12-07
4 537626	22375	AIRLINE BAG VINTAGE JET SET BROWN	4	12/7/10 14:57		5 537626	71477	COLOUR GLASS. STAR T-LIGHT HOLDER	12	0010-12-07
5 537626	71477	COLOUR GLASS. STAR T-LIGHT HOLDER	12	12/7/10 14:57		6 537626	22492	MINI PAINT SET VINTAGE	36	0010-12-07
6 537626	22492	MINI PAINT SET VINTAGE	36	12/7/10 14:57		7 537626	22771	CLEAR DRAWER KNOB ACRYLIC EDWARDIAN	12	0010-12-07
7 537626	22771	CLEAR DRAWER KNOB ACRYLIC EDWARDIAN	12	12/7/10 14:57		8 537626	22772	PINK DRAWER KNOB ACRYLIC EDWARDIAN	12	0010-12-07
8 537626	22772	PINK DRAWER KNOB ACRYLIC EDWARDIAN	12	12/7/10 14:57		9 537626	22773	GREEN DRAWER KNOB ACRYLIC EDWARDIAN	12	0010-12-07
9 537626	22773	GREEN DRAWER KNOB ACRYLIC EDWARDIAN	12	12/7/10 14:57		10 537626	22774	RED DRAWER KNOB ACRYLIC EDWARDIAN	12	0010-12-07
10 537626	22774	RED DRAWER KNOB ACRYLIC EDWARDIAN	12	12/7/10 14:57		11 537626	22775	PURPLE DRAWERKNOB ACRYLIC EDWARDIAN	12	0010-12-07
11 537626	22775	PURPLE DRAWERKNOB ACRYLIC EDWARDIAN	12	12/7/10 14:57		12 537626	22805	BLUE DRAWER KNOB ACRYLIC EDWARDIAN	12	0010-12-07
12 537626	22805	BLUE DRAWER KNOB ACRYLIC EDWARDIAN	12	12/7/10 14:57		13 537626	22725	ALARM CLOCK BAKELIKE CHOCOLATE	4	0010-12-07

Pada data *Online Retail* jumlah baris yang dimiliki yaitu 541.909 baris, yang pada seluruh baris tersebut terdapat data-data yang sekiranya tidak dibutuhkan untuk pengolahan data. Setelah dilakukan Data Cleansing jumlah baris yang dimiliki sekarang yaitu 393.568 baris.

Pada analisis k-means kali ini, atribut yang digunakan adalah :

- **Number of Unique Customer** : Jumlah pelanggan unik pada tiap negara
- **Total Sales** : jumlah penjualan yang terjadi pada tiap negara
- **Total Quantity** : jumlah barang yang terjual pada tiap negara

AGREGAT DATA 1						
CountryCustomer				AggCountry		
	Country	Number of Unique Customer		Country	Total Sales	
1	Australia	9		1	Australia	136028.56
2	Austria	11		2	Austria	8722.88
3	Bahrain	2		3	Bahrain	548.40
4	Belgium	25		4	Belgium	36462.94
5	Brazil	1		5	Brazil	1143.60
6	Canada	3		6	Canada	2534.40
7	Channel Islands	9		7	Channel Islands	19939.44
8	Cyprus	8		8	Cyprus	13239.84
9	Czech Republic	1		9	Czech Republic	786.74
10	Denmark	9		10	Denmark	18117.74
11	EIRE	3		11	EIRE	254333.37
				12	European Community	1139.45

Gabungan (CountryCustomer + AggCountry)				
	Country	Total Sales	Number of Unique Customer	
1	Australia	136028.56	9	
2	Austria	8722.88	11	
3	Bahrain	548.40	2	
4	Belgium	36462.94	25	
5	Brazil	1143.60	1	
6	Canada	2534.40	3	
7	Channel Islands	19939.44	9	
8	Cyprus	13239.84	8	
9	Czech Republic	786.74	1	
10	Denmark	18117.74	9	
11	EIRE	254333.37	3	
Showing 1 to 11 of 37 entries				

Untuk membentuk tabel dengan ketiga atribut yang dibutuhkan tersebut, perlu dilakukan agregat pada beberapa data awal. Pada agregat data 1, tabel CountryCustomer adalah hasil dari agregat yang dilakukan untuk mencari jumlah pelanggan unik tiap-tiap negaranya. Sedangkan pada tabel AggCountry awal, agregat dilakukan terhadap total penjualan yang dilakukan tiap negara. Lalu kemudian kedua tabel tersebut digabungkan menjadi tabel AggCountry. Pada agregat 2, tabel AggCountry yang tadi telah terbentuk digabungkan kembali dengan tabel AggCountry dari hasil agregat terhadap jumlah item barang yang terjual di tiap negara. Hasil akhir dari agregat-agregat yang telah dilakukan adalah sebuah tabel dengan kolom negara, total sales, number of unique customer, dan total quantity.

AGREGAT DATA 2

AggCountry				AggQuantity			
	Country	Total Sales	Number of Unique Customer		Country	Total Quantity	
1	Australia	136028.56	9	1	Australia	83587	
2	Austria	8722.88	11	2	Austria	4840	
3	Bahrain	548.40	2	3	Bahrain	260	
4	Belgium	36462.94	25	4	Belgium	22810	
5	Brazil	1143.60	1	5	Brazil	356	
6	Canada	2534.40	3	6	Canada	2234	
7	Channel Islands	19939.44	9	7	Channel Islands	9471	
8	Cyprus	13239.84	8	8	Cyprus	6356	
9	Czech Republic	786.74	1	9	Czech Republic	670	
10	Denmark	18117.74	9	10	Denmark	8145	
11	EIRE	254333.37	3	11	EIRE	139136	
				12	European Community	486	
Showing 1 to 11 of 37 entries				Showing 1 to 12 of 37 entries			

Gabungan (AggCountry + AggQuantity)				
	Country	Total Sales	Number of Unique Customer	Total Quantity
1	Australia	138521.31	9	84209
2	Austria	10198.68	11	4881
3	Bahrain	548.40	2	260
4	Belgium	41196.34	25	23237
5	Brazil	1143.60	1	356
6	Canada	3666.38	4	2763
7	Channel Islands	20450.44	9	9491
8	Cyprus	13590.38	8	6361
9	Czech Republic	826.74	1	671
10	Denmark	18955.34	9	8235
11	EIRE	265545.90	3	140525
12	European Community	1300.25	1	499
13	Finland	22546.08	12	10704
14	France	209024.05	87	111472
15	Germany	228867.14	94	119263
16	Greece	4760.52	4	1557
17	Iceland	4310.00	1	2458

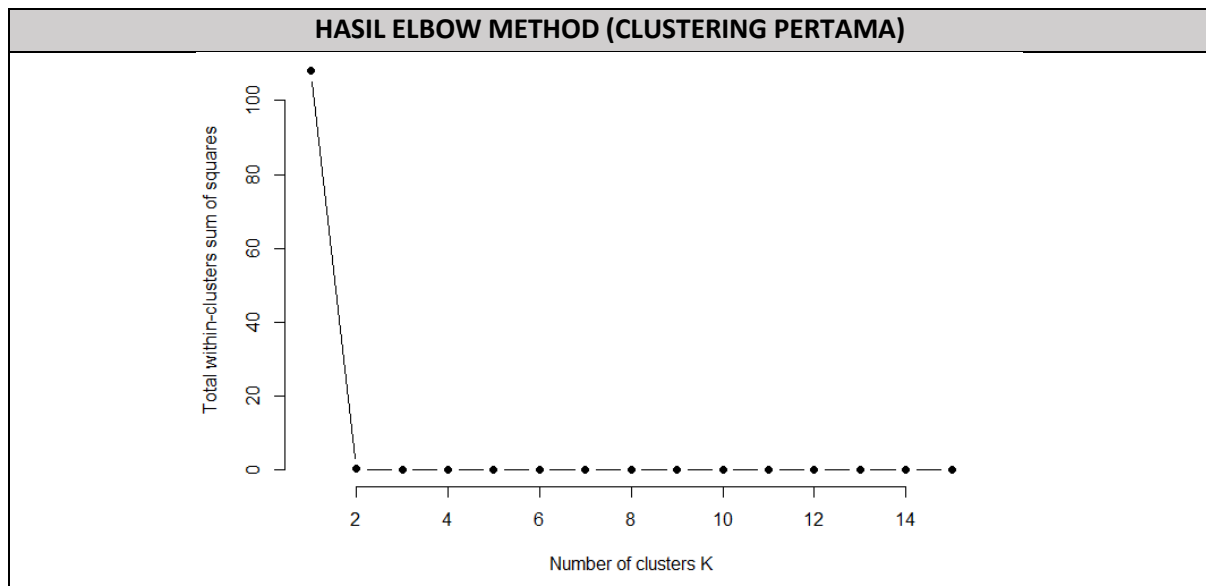
Selanjutnya kolom Country diubah menjadi nama baris (row names) dari masing-masing baris, sehingga kolom Country dihilangkan. Hal tersebut dimaksudkan agar data dapat selanjutnya diolah dengan menggunakan algoritma K-Means Cluster.

5.2. CLUSTERING PERTAMA

Sebelum melakukan clustering, perlu adanya normalisasi data. Normalisasi data sendiri dimaksudkan agar data tidak timpang pada data-data tertentu. Hal ini dilakukan pada studi kasus kali ini karena dari tabel sebelum normalisasi dapat dilihat bahwa total sales memiliki angka hingga ratusan ribu, sedangkan number of unique customer hanya mencapai ribuan, dan total quantity mencapai puluhan ribu. Normalisasi yang digunakan pada studi kasus kali ini adalah normalisasi Z-score. Dari normalisasi yang dilakukan, terbentuklah tabel yang memiliki data yang relatif seimbang, yang dapat dilihat pada tabel dibawah.

NORMALISASI (CLUSTERING PERTAMA)				
Sebelum				Sesudah
	Total Sales	Number of Unique Customer	Total Quantity	
Australia	138521.31	9	84209	Australia -0.08551244 -0.16872317 -0.0798531764
Austria	10198.68	11	4881	Austria -0.19274828 -0.16561268 -0.1933014485
Bahrain	548.40	2	260	Bahrain -0.20081277 -0.17960989 -0.1999100163
Belgium	41196.34	25	23237	Belgium -0.16684435 -0.14383923 -0.1670502323
Brazil	1143.60	1	356	Brazil -0.20031537 -0.18116513 -0.1997727252
Canada	3666.38	4	2763	Canada -0.19820715 -0.17649940 -0.1963304350
Channel Islands	20450.44	9	9491	Channel Islands -0.18418116 -0.16872317 -0.1867086120
Cyprus	13590.38	8	6361	Cyprus -0.18991393 -0.17027841 -0.1911848763
Czech Republic	826.74	1	671	Czech Republic -0.20058017 -0.18116513 -0.1993222385
Denmark	18955.34	9	8235	Denmark -0.18543057 -0.16872317 -0.1885048382
EIRE	265545.90	3	140525	EIRE 0.02063866 -0.17805464 0.0006852577
European Community	1300.25	1	499	European Community -0.20018447 -0.18116513 -0.1995682185
Finland	22546.08	12	10704	Finland -0.18242989 -0.16405743 -0.1849738809
France	209024.05	87	111472	France -0.02659516 -0.04741398 -0.0408639135
Germany	228867.14	94	119263	Germany -0.01001281 -0.03652726 -0.0297218768
Greece	4760.52	4	1557	Greece -0.19729281 -0.17649940 -0.1980551554
Iceland	4310.00	1	2458	Iceland -0.19766930 -0.18116513 -0.1967666205

Sebelum menjalankan algoritma k-means clustering, perlu diketahui lebih dulu jumlah cluster yang hendak dibuat. Untuk menentukan jumlah cluster tersebut, dapat dilakukan metode yang dinamakan Elbow Method. Dari hasil grafik yang terbentuk, dapat dilihat nilai mana yang menjadi titik acuan dimana grafik sudah tidak lagi turun secara signifikan. Pada hasil elbow method di clustering pertama ini, jumlah cluster optimal adalah 2. Karena, setelah angka 2 sudah tidak terjadi lagi penurunan yang signifikan.

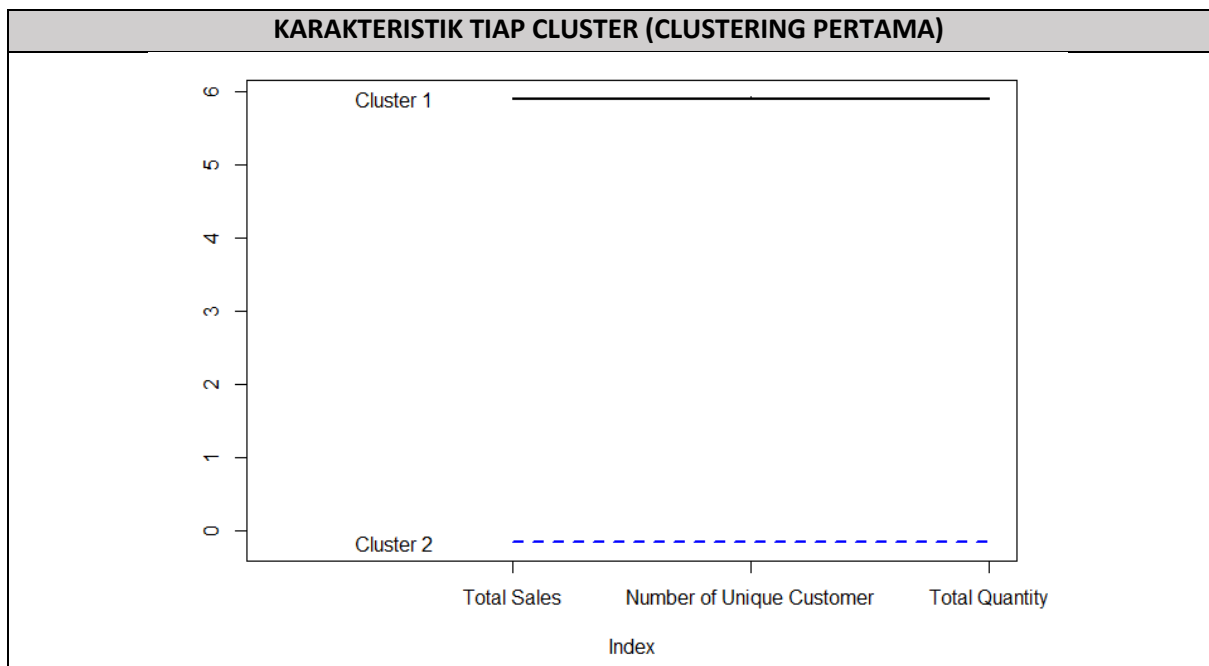


Setelah menjalankan algoritma k-means clustering, ternyata ditemukan bahwa Cluster pertama hanya diisi oleh United Kingdom dan Cluster kedua diisi oleh negara-negara lainnya.

DAFTAR ANGGOTA (CLUSTERING PERTAMA)

CountryCluster\$cluster					
Australia	2	Austria	2	Bahrain	2
Channel Islands	2	Cyprus	2	Belgium	2
Finland	2	France	2	Denmark	2
Italy	2	Germany	2	EIRE	2
Norway	2	Japan	2	Greece	2
Spain	2	Lebanon	2	Iceland	2
USA	2	Lithuania	2	Malta	2
		Poland	2	Saudi Arabia	2
		Portugal	2	Switzerland	2
		RSA	2	United Arab Emirates	2
		Sweden	2	United Kingdom	1
		Switzerland	2	Unspecified	2

Untuk karakteristik dari kedua cluster yang terbentuk tersebut, dapat dilihat pada gambar grafik dibawah. Pada gambar grafik tersebut, informasi yang didapatkan yaitu bahwa cluster pertama memiliki karakteristik total sales, number of unique customer, dan total quantity yang tinggi sedangkan cluster kedua memiliki karakteristik total sales, number of unique customer, dan total quantity yang rendah.



Dari hasil identifikasi anggota cluster dan karakteristiknya, berikut adalah ringkasan informasi cluster yang terbentuk disajikan dalam bentuk tabel.

No Cluster	Anggota Cluster	Karakteristik		
		Total Penjualan	Jumlah Pelanggan	Jumlah barang yang terjual
1	United Kingdom	Sangat Tinggi	Sangat Tinggi	Sangat Tinggi
2	Australia, Austria, Bahrain, Belgium, Brazil, Canada, Channel Islands, Cyprus, Czech Republic, Denmark, EIRE, European Community, Finland, France, Germany, Greece, Iceland, Israel, Italy, Japan, Lebanon, Lithuania, Malta, Netherlands, Norway, Poland, Portugal, RSA, Saudi Arabia, Singapore, Spain, Sweden, Switzerland, United Arab Emirates, Unspecified USA	Rendah	Rendah	Rendah

5.3. CLUSTERING KEDUA

Pada hasil clustering pertama didapati bahwa terdapat dua cluster yang dihasilkan, namun jika dilihat dari karakteristik dan nilai yang yang dihasilkan, pada kluster satu hanya diisi dengan United Kingdom yang jarak nilainya sangatlah jauh dengan cluster satunya yang berisi seluruh negara kecuali United Kingdom. Dilihat dari sini maka cluster yang dihasilkan kurang valid, sehingga dilakukan clustering tanpa menggunakan United Kingdom.

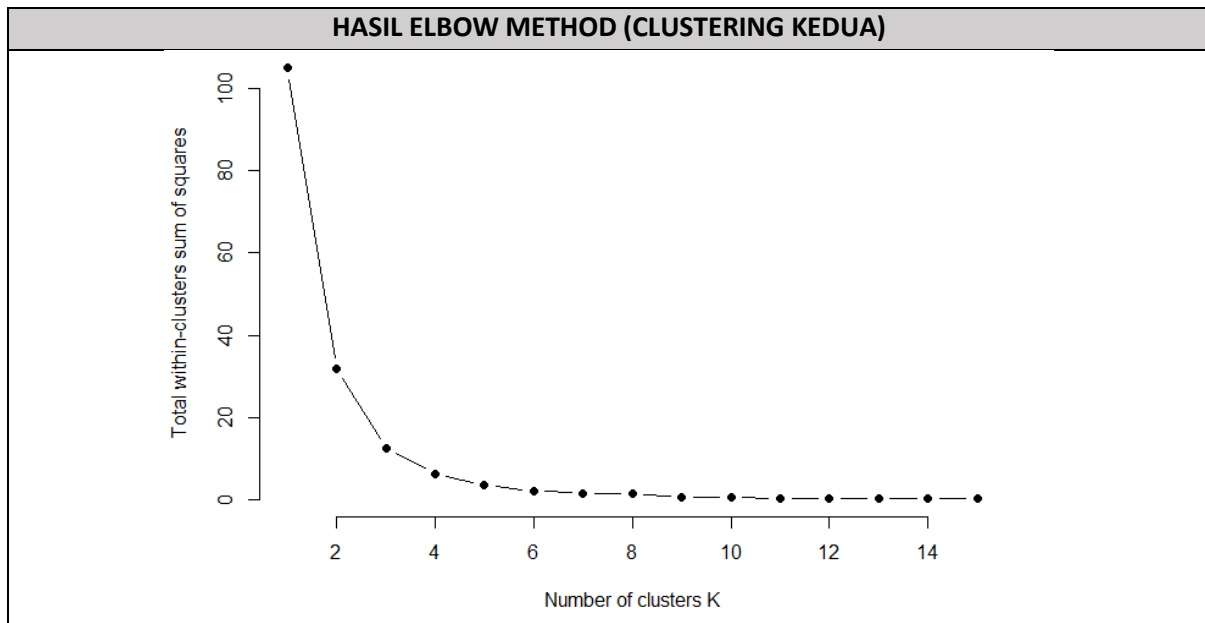
SAMPEL DATA TANPA UNITED KINGDOM				
	Country	Total Sales	Number of Unique Customer	Total Quantity
37	USA	3580.39	4	2458
36	Unspecified	2667.07	4	1789
34	United Arab Emirates	1902.28	2	982
33	Switzerland	56443.95	21	30083
32	Sweden	38378.33	8	36083
31	Spain	61577.11	30	27951
30	Singapore	21279.29	1	5241
29	Saudi Arabia	145.92	1	80
28	RSA	1002.31	1	352
27	Portugal	33439.89	19	16122
26	Poland	7334.65	6	3684
25	Norway	36165.44	10	19338

Selanjutnya sama seperti pada clustering pertama, kolom Country diubah menjadi nama baris (row names) dari masing-masing baris, sehingga kolom Country ditiadakan. Hal tersebut dimaksudkan agar data dapat selanjutnya diolah dengan menggunakan algoritma K-Means Cluster.

Sebelum melakukan clustering, perlu dilakukan normalisasi data. Normalisasi data sendiri dimaksudkan agar data tidak timpang pada data-data tertentu. Normalisasi yang digunakan pada studi kasus kali ini adalah normalisasi Z-score. Dari normalisasi yang dilakukan, terbentuklah tabel yang memiliki data yang relative seimbang, yang dapat dilihat pada tabel dibawah.

NORMALISASI (CLUSTERING KEDUA)							
Sebelum				Sesudah			
	Total Sales	Number of Unique Customer	Total Quantity		Total Sales	Number of Unique Customer	Total Quantity
Netherlands	285446.34	9	200937	Australia	1.20661788	-0.137289261	1.26269867
EIRE	265545.90	3	140525	Austria	-0.44069895	-0.040379194	-0.43881832
Germany	228867.14	94	119263	Bahrain	-0.56458255	-0.476474494	-0.53793478
France	209024.05	87	111472	Belgium	-0.04277252	0.637991272	-0.04509800
Australia	138521.31	9	84209	Brazil	-0.55694179	-0.524929528	-0.53587566
Spain	61577.11	30	27951	Canada	-0.52455609	-0.379564428	-0.48424759
Switzerland	56443.95	21	30083	Channel Islands	-0.30909398	-0.137289261	-0.33993781
Belgium	41196.34	25	23237	Cyprus	-0.39715866	-0.185744294	-0.40707360
Sweden	38378.33	8	36083	Czech Republic	-0.56100941	-0.524929528	-0.52911918
Japan	37416.37	8	26016	Denmark	-0.32828703	-0.137289261	-0.36687792
Norway	36165.44	10	19338	EIRE	2.83727138	-0.428019461	2.47062816
Portugal	33439.89	19	16122	European Community	-0.55493082	-0.524929528	-0.53280843
Finland	22546.08	12	10704	Finland	-0.28219161	0.008075839	-0.31392001
Singapore	21279.29	1	5241	France	2.11168311	3.642203337	1.84746644
Channel Islands	20450.44	9	9491	Germany	2.36641492	3.981388570	2.01457665
Denmark	18955.34	9	8235	Greece	-0.51051028	-0.379564428	-0.51011525
Italy	17483.24	14	8112	Iceland	-0.51629374	-0.524929528	-0.49078958

Setelah itu dilakukan metode Elbow Method yang bertujuan untuk mengetahui jumlah cluster yang akan dibuat. Grafik yang dihasilkan dari Elbow Method ini terdapat titik acuan dimana jika titik acuan tersebut sudah tidak turun secara signifikan, maka pada nilai titik itu yang menjadi acuan jumlah cluster yang terbentuk. Dan pada hasil Elbow Method clustering kedua ini ditemukan jumlah cluster yang optimla yaitu sejumlah 4 cluster.

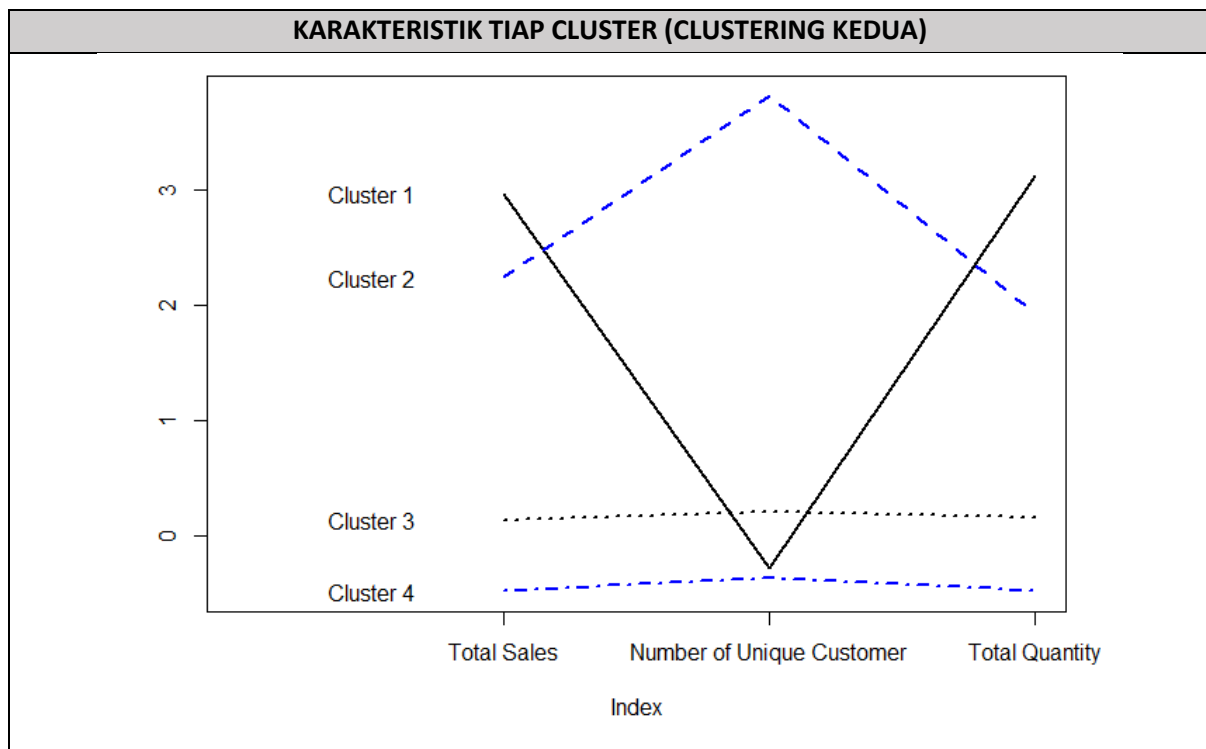


Setelah menjalankan algoritma k-means clustering, ternyata ditemukan bahwa Cluster pertama diisi oleh negara Netherlands dan EIRE, lalu pada cluster kedua diisi oleh negara France dan Germany, pada cluster ketiga diisi oleh negara Australia, Norway, Spain, Japan, Sweden, Portugal, Switzerland dan Belgium. Pada kluster keempat diisi oleh negara Channel Island, Finland, Italy, Austria, Cyprus, Poland, Bahrain, Czech Republic, Lebanon, Denmark, Greece, Lithuania, RSA, United Arab Emirates, Brazil, Iceland, Malta, Saudi Arabia, Unspecified, Canada, European Community, Israel, Singapore dan USA.

DAFTAR ANGGOTA (CLUSTERING KEDUA)					
> CountryCluster\$cluster					
Australia	Austria	Bahrain	Belgium	Brazil	Canada
3	4	4	3	4	4
Channel Islands	Cyprus	Czech Republic	Denmark	EIRE	European Community
4	4	4	4	1	4
Finland	France	Germany	Greece	Iceland	Israel
4	2	2	4	4	4
Italy	Japan	Lebanon	Lithuania	Malta	Netherlands
4	3	4	4	4	1
Norway	Poland	Portugal	RSA	Saudi Arabia	Singapore
3	4	3	4	4	4
Spain	Sweden	Switzerland	United Arab Emirates	Unspecified	USA
3	3	3	4	4	4

Berikut ini merupakan karakteristik dari tiap-tiap cluster yang terbentuk. Pada cluster 1 memiliki kondisi karakteristik Total Sales dan Total Quantity yang paling tinggi, namun dari segi Number of Unique Customer berada pada tingkat yang agak rendah. Lalu pada cluster 2 kondisi Total Sales dan

Total Quantity yang agak tinggi, namun pada Number of Unique Customer memiliki nilai yang paling tinggi. Lalu karakteristik pada cluster 3 memiliki Total Sales dan Total Quantity yang agak rendah dan Number of Unique Customer sedang. Lalu pada karakteristik cluster 4 memiliki Total Sales, Number of Unique Customer dan Total Quantity yang terendah.



Dari hasil identifikasi anggota cluster dan karakteristiknya, berikut adalah ringkasan informasi cluster yang terbentuk disajikan dalam bentuk tabel.

No Cluster	Anggota Cluster	Karakteristik		
		Total Penjualan	Jumlah Pelanggan	Jumlah barang yang terjual
1	Netherlands, EIRE	Paling Tinggi	Agak Rendah	Paling Tinggi
2	France, Germany	Agak Tinggi	Paling Tinggi	Agak Tinggi
3	Australia, Belgium, Japan, Norway, Portugal, Spain, Sweden, Switzerland	Agak Rendah	Sedang	Agak Rendah
4	Austria, Bahrain, Brazil, Canada, Channel Islands, Cyprus, Czech Republic, Denmark, European	Paling rendah	Paling rendah	Paling rendah

	Community, Finland, , Greece, Iceland, Israel, Italy, Lebanon, Lithuania, Malta, Poland, USA, RSA, Arab Emirates, Unspecified, Saudi Arabia, Singapore			
--	---	--	--	--

5.4. CLUSTER GABUNGAN

Dengan menggunakan hasil dua kali cluster tersebut, didapatkan lima cluster dengan karakteristik sebagai berikut.

No Cluster	Anggota Cluster	Karakteristik		
		Total Penjualan	Jumlah Pelanggan	Jumlah barang yang terjual
1	United Kingdom	Terlampau Tinggi	Terlampau Tinggi	Terlampau Tinggi
2	Netherlands, EIRE	Tinggi	Agak Rendah	Tinggi
3	France, Germany	Agak Tinggi	Tinggi	Agak Tinggi
4	Australia, Belgium, Japan, Norway, Portugal, Spain, Sweden, Switzerland	Agak Rendah	Sedang	Agak Rendah
5	Austria, Bahrain, Brazil, Canada, Channel Islands, Cyprus, Czech Republic, Denmark, European Community, Finland, Greece, Iceland, Israel, Italy, Lebanon, Lithuania, Malta, Poland, USA, RSA, Arab Emirates, Unspecified, Saudi Arabia, Singapore	Rendah	Rendah	Rendah

5.5. PENENTUAN STRATEGI

Setelah mendapatkan hasil clustering tersebut, selanjutnya dirumuskanlah strategi-strategi perusahaan yang sekiranya bisa diterapkan sesuai dengan karakteristik negaranya.

Untuk Cluster pertama yang berisi negara United Kingdom dengan memiliki karakteristik total penjualan, jumlah pelanggan dan jumlah barang yang terjual sangat tinggi, strategi yang mungkin diterapkan adalah tetap mempertahankan marketshare di negara tersebut sehingga dapat menjaga tingkat loyalitas pelanggan agar tetap tinggi. Hal tersebut dapat dilakukan dengan cara mempertahankan cara penjualan yang berada pada United Kingdom dan mengadakan suatu promo pada waktu tertentu agar customer tertarik untuk bertransaksi pada *Online Retail* tersebut.

Cluster kedua berisi negara Netherlands dan EIRE dimana kedua negara tersebut memiliki karakteristik yaitu memiliki total penjualan dan jumlah barang yang terjual cukup tinggi namun jumlah pelanggan yang agak rendah. Dari karakteristik tersebut, dapat dilihat bahwa sebenarnya kedua negara tersebut memiliki potensi untuk dapat menghasilkan profit yang tinggi bagi perusahaan. Sehingga, untuk pengoptimalan dapat dilakukan pengiklanan mengenai *Online Retail* sehingga jika terdapat orang yang belum mengerti tentang *Online Retail* tersebut dapat mengerti dengan *Online Retail* tersebut dan ingin menggunakan *Online Retail* tersebut. Dan juga dengan mengadakan fitur-fitur yang membuat orang penasaran sehingga, orang-orang ingin menggunakan *Online Retail* tersebut.

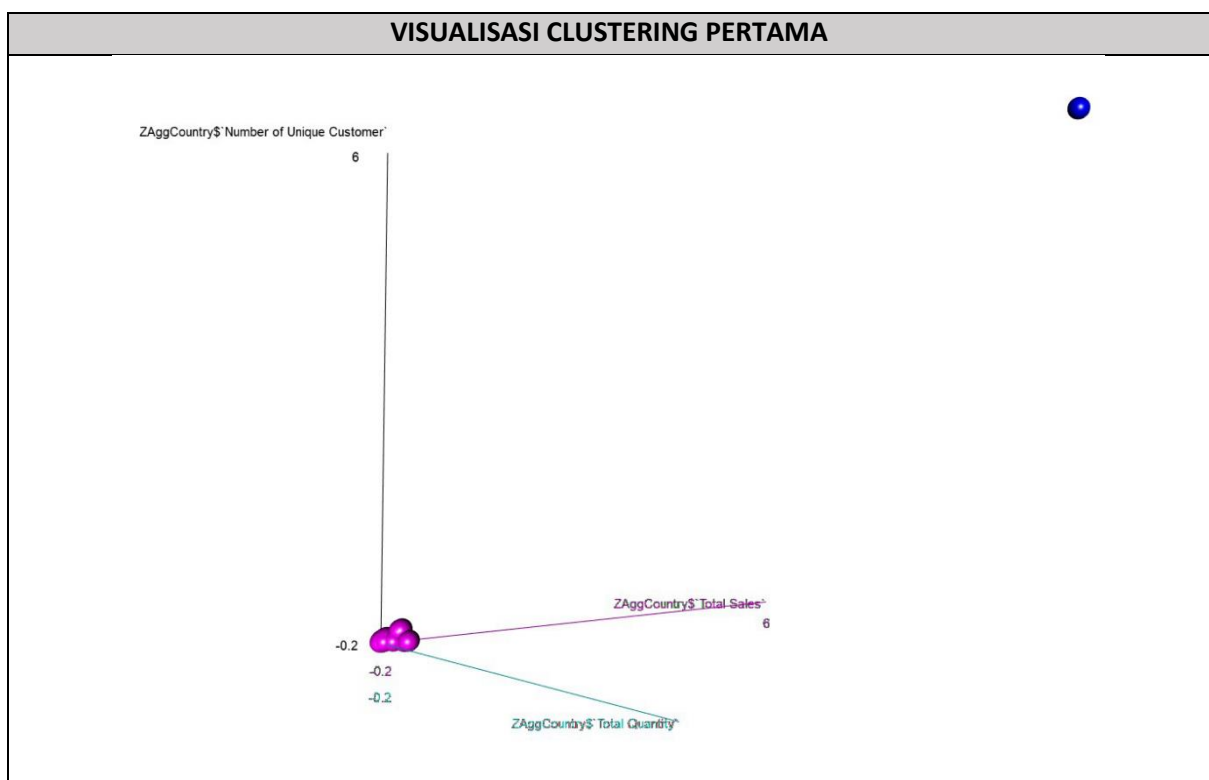
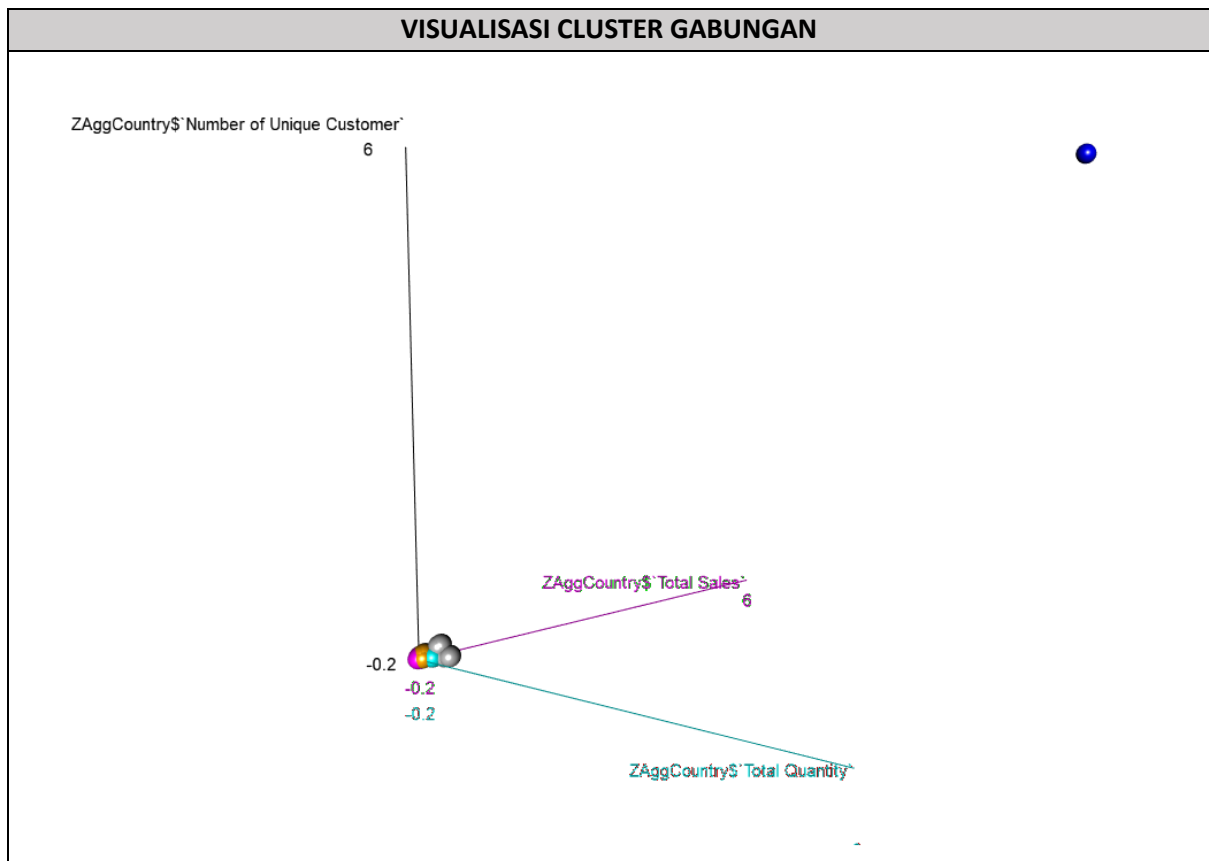
Cluster ketiga berisi negara France dan Germany dimana kedua negara tersebut memiliki karakteristik yaitu memiliki total penjualan dan jumlah barang yang terjual agak tinggi serta jumlah pelanggan yang cukup tinggi. Dari karakteristik tersebut, dapat dilihat bahwa sebenarnya kedua negara tersebut memiliki potensi untuk dapat menghasilkan profit yang tinggi bagi perusahaan. Sehingga, untuk pengoptimalan dapat melakukan strategi mengiklankan produk-produk yang dijual pada *Online Retail* dan juga membandingkan dengan harga pesaing. Berikan harga yang dibawah harga pesaing maka pelanggan akan membeli barang di *Online Retail* ini.

Cluster keempat berisi negara Australia, Belgium, Japan, Norway, Portugal, Spain, Sweden, dan Switzerland dimana negara-negara pada cluster tersebut memiliki karakteristik yaitu memiliki total penjualan dan jumlah barang yang terjual agak tinggi dan memiliki jumlah pelanggan yang cukup tinggi. Dari karakteristik yang ada pada cluster ini perlu melakukan strategi penjualan yaitu dengan merekomendasikan produk-produk yang sekiranya dekat dengan pelanggan atau barang-barang yang sekiranya sejenis dan saling berkaitan dengan barang yang sering dibeli oleh pelanggan. Dengan memberi fitur rekomendasi barang pada tiap akun pelanggan supaya jumlah barang yang terjual juga dapat bertambah. Dan juga dengan mengiklankan produk pada sosial media sehingga pelanggan yang

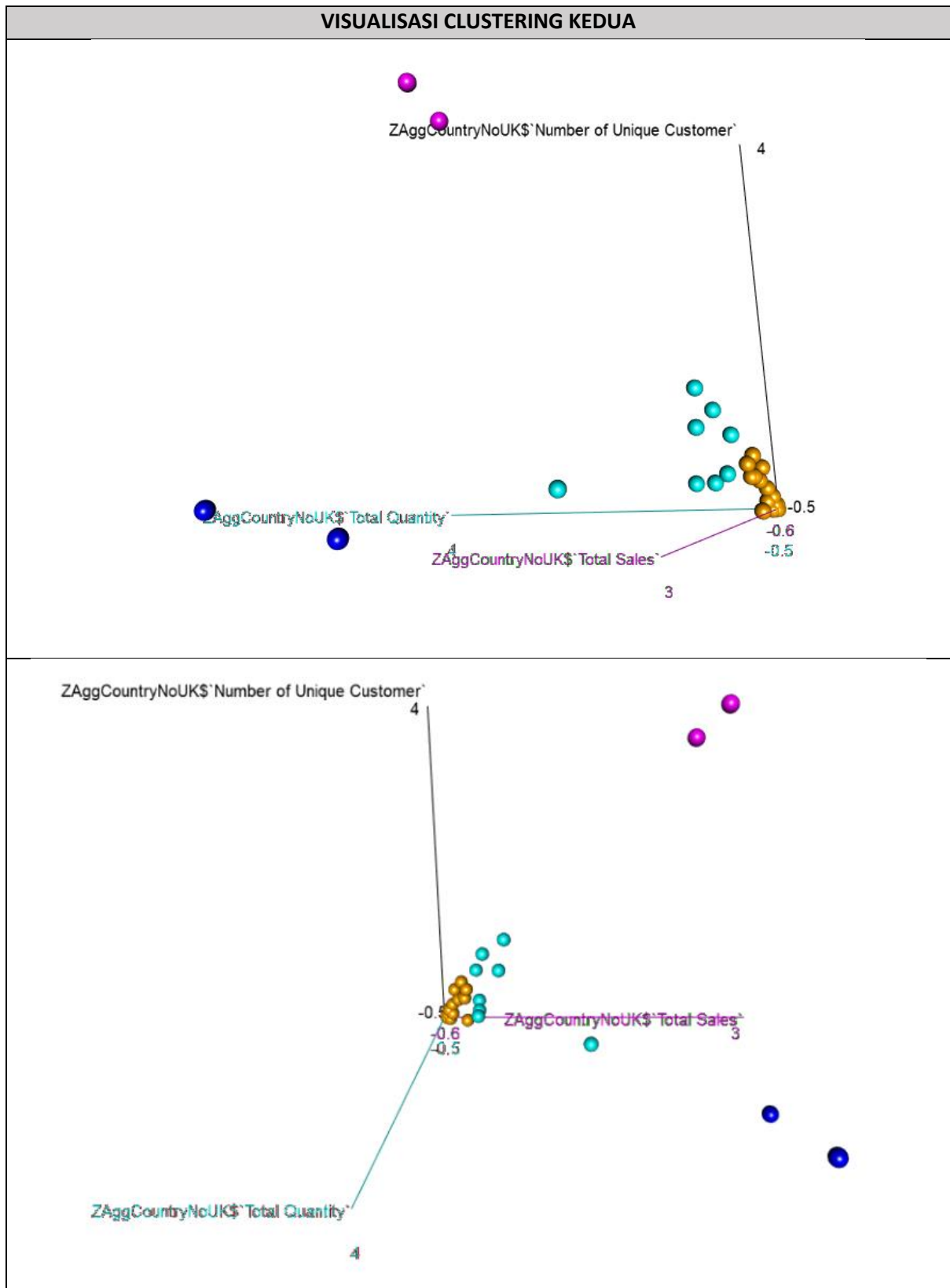
membuka sosial media dapat tertarik untuk belanja pada *Online Retail*. Dan juga dengan menjadi sponsor suatu kegiatan yang sekiranya kegiatan tersebut banyak dihadiri orang sehingga orang-orang dapat melihat *Online Retail* ini.

Kemudian pada cluster lima berisi negara Austria, Bahrain, Brazil, Canada, Channel Islands, Cyprus, Czech Republic, Denmark, European Community, Finland, Greece, Iceland, Israel, Italy, Lebanon, Lithuania, Malta, Poland, USA, RSA, Arab Emirates, Unspecified, Saudi Arabia, dan Singapore dimana negara-negara pada cluster tersebut memiliki karakteristik yang sama yaitu memiliki total penjualan, jumlah barang yang terjual, dan jumlah pelanggan yang rendah. Dari karakteristik tersebut diperlukan strategi penjualan yang sangat matang agar *Online Retail* tersebut tidak tersisih dari persaingan. Dengan mengiklankan *Online Retail* tersebut secara massal maka dapat menarik perhatian customer sehingga pelanggan dapat melihat *Online Retail* tersebut dan menggunakannya. Lalu dengan memberikan promo-promo yang tidak kalah dengan pesaing sehingga dengan harga yang murah dapat menarik pelanggan dan pelanggan dapat membeli barang dengan jumlah yang banyak. Kemudian dengan menggunakan Brand Ambassador tokoh-tokoh yang ada pada negara cluster lima. Dengan memanfaatkan Brand Ambassador maka penggemar dari Brand Ambassador dapat terpengaruh dan akan menggunakan *Online Retail* tersebut.

5.6. VISUALISASI



VISUALISASI CLUSTERING KEDUA



BAB VI : KESIMPULAN DAN SARAN

6.1. KESIMPULAN

Dari hasil analisa dengan k-means cluster yang dilakukan pada data Online Retail, dapat disimpulkan bahwa cluster yang terbentuk berjumlah lima dengan rincian anggota tiap cluster dan karakteristiknya sebagai berikut.

No Cluster	Anggota Cluster	Karakteristik		
		Total Penjualan	Jumlah Pelanggan	Jumlah barang yang terjual
1	United Kingdom	Terlampau Tinggi	Terlampau Tinggi	Terlampau Tinggi
2	Netherlands, EIRE	Tinggi	Agak Rendah	Tinggi
3	France, Germany	Agak Tinggi	Tinggi	Agak Tinggi
4	Australia, Belgium, Japan, Norway, Portugal, Spain, Sweden, Switzerland	Agak Rendah	Sedang	Agak Rendah
5	Austria, Bahrain, Brazil, Canada, Channel Islands, Cyprus, Czech Republic, Denmark, European Community, Finland, , Greece, Iceland, Israel, Italy, Lebanon, Lithuania, Malta, Poland, USA, RSA, Arab Emirates, Unspecified, Saudi Arabia, Singapore	Rendah	Rendah	Rendah

Selanjutnya, dari cluster yang terbentuk tersebut dianalisis strategi penjualan tiap clusternya, sehingga strategi yang dirumuskan untuk tiap clusternya adalah sebagai berikut.

No Cluster	Anggota Cluster	Strategi
1	United Kingdom	- Mempertahankan cara penjualan yang berada pada United Kingdom

		<ul style="list-style-type: none"> - Mengadakan suatu promo pada waktu tertentu agar customer tetap tertarik untuk bertransaksi pada <i>Online Retail</i> tersebut.
2	Netherlands, EIRE	<ul style="list-style-type: none"> - Pengiklanan mengenai <i>Online Retail</i> sehingga jika terdapat orang yang belum mengerti tentang <i>Online Retail</i> tersebut dapat mengerti dengan <i>Online Retail</i> tersebut dan ingin menggunakan <i>Online Retail</i> tersebut. - Mengadakan fitur-fitur yang membuat orang penasaran sehingga, orang-orang ingin menggunakan <i>Online Retail</i> tersebut.
3	France, Germany	<ul style="list-style-type: none"> - Mengiklankan produk-produk yang dijual pada <i>Online Retail</i> - Berikan harga yang dibawah harga pesaing maka pelanggan akan membeli barang di <i>Online Retail</i> ini. - Merekomendasikan barang kepada pelanggan yang sekiranya barang tersebut berkaitan dengan barang yang pernah dibeli pelanggan
4	Australia, Belgium, Japan, Norway, Portugal, Spain, Sweden, Switzerland	<ul style="list-style-type: none"> - Merekomendasikan produk-produk yang sekiranya dekat dengan pelanggan atau barang-barang yang sekiranya sejenis dan saling berkaitan dengan barang yang sering dibeli oleh pelanggan. Dengan memberi fitur rekomendasi barang pada tiap akun pelanggan supaya jumlah barang yang terjual juga dapat bertambah. - Mengiklankan produk pada sosial media sehingga pelanggan yang membuka sosial media dapat tertarik untuk belanja pada <i>Online Retail</i>. - Menjadi sponsor suatu kegiatan yang sekiranya kegiatan tersebut banyak dihadiri orang sehingga orang-orang dapat melihat <i>Online Retail</i> ini.
5	Austria, Bahrain, Brazil, Canada, Channel Islands, Cyprus, Czech Republic,	<ul style="list-style-type: none"> - Dengan mengiklankan <i>Online Retail</i> tersebut secara massal maka dapat menarik perhatian customer

	Denmark, European Community, Finland, Greece, Iceland, Israel, Italy, Lebanon, Lithuania, Malta, Poland, USA, RSA, Arab Emirates, Unspecified, Saudi Arabia, Singapore	<p>sehingga pelanggan dapat melirik <i>Online Retail</i> tersebut dan menggunakannya.</p> <ul style="list-style-type: none"> - Memberikan promo-promo yang tidak kalah dengan pesaing sehingga dengan harga yang murah dapat menarik pelanggan dan pelanggan dapat membeli barang dengan jumlah yang banyak. - Menggunakan Brand Ambassador tokoh-tokoh yang ada pada negara cluster lima. Dengan memanfaatkan Brand Ambassador maka penggemar dari Brand Ambassador dapat terpengaruh dan akan menggunakan <i>Online Retail</i> tersebut.
--	--	--

6.2. SARAN

Dari penelitian ini, landasan pembentukan strategi untuk masing-masing cluster negara yang terbentuk masih belum dilihat secara mendalam dari segi bisnis. Strategi yang ditawarkan hanya gambaran besarnya saja, sehingga perlu adanya kajian lebih lanjut mengenai hal tersebut.

REFERENSI

- [1] Daniel T. Larose, *Discovering Knowledge In Data*. United States Of America: John Wiley & Sons, Inc., 2005.
- [2] Firdausi Nabila and Wessiani Naning Aranti, "Analisis Financial Distress dengan Pendekatan Data Mining pada Industri Manufaktur Go-Public di Indonesia," Institut Teknologi Sepuluh Nopember (ITS), Surabaya, 2013
- [3] Informatikalogi, 12 November 2016. [Online]. Available: <https://informatikalogi.com/algorithm-k-means-clustering/>. [Accessed 23 May 2019]