

Nama Anggota :

- Ahmad Shobirin 3311901047
- Nuzela Rahma Fitra 3311901058
- Indah Fauziyyah Utami 3311901060

TUGAS BESAR DATA MINING

(Manajemen Data Menggunakan RStudio Pada Data Chocolate Bar Rating)

Data manajemen merupakan bagian penting dalam setiap proses analisa data. Proses import dan ekspor data pada berbagai format penting untuk dipelajari. Selain itu, proses perapihan data sebelum analisa menjadi bagian yang harus ada pada awal proses analisa. Proses-proses tersebut akan kami ulas pada laporan ini.

Konteks

Cokelat adalah salah satu permen paling populer di dunia. Setiap tahun, penduduk Amerika Serikat secara kolektif makan lebih dari 2,8 miliar pound. Namun, tidak semua batang cokelat dibuat sama! Kumpulan data ini berisi peringkat ahli lebih dari 1.700 batang coklat, bersama dengan informasi tentang asal daerah, persentase kakao, varietas biji coklat yang digunakan, dan di mana biji itu ditanam.

Sistem Pemeringkatan Kakao:

- 5 = Elite (Melampaui melampaui batas biasa)
- 4 = Premium (Pengembangan rasa, karakter, dan gaya superior)
- 3 = Memuaskan (3,0) hingga terpuji (3,75) (dibuat dengan baik dengan kualitas khusus)
- 2 = Mengecewakan (Lumayan tetapi mengandung setidaknya satu kekurangan yang signifikan)
- 1 = Tidak menyenangkan (kebanyakan tidak enak)

Setiap coklat dievaluasi dari kombinasi kualitas objektif dan interpretasi subjektif. Peringkat di sini hanya mewakili pengalaman dengan satu batang dari satu kelompok. Nomor batch, vintages dan tanggal tinjauan disertakan dalam database jika diketahui.

Basis data secara sempit difokuskan pada cokelat hitam biasa dengan tujuan menghargai rasa kakao saat dibuat menjadi cokelat. Peringkat tersebut tidak mencerminkan manfaat kesehatan, misi sosial, atau status organik.

Flavour adalah komponen terpenting dari peringkat Flavours of Cacao. Keragaman, keseimbangan, intensitas, dan kemurnian rasa semuanya dipertimbangkan. Hal ini dimungkinkan untuk satu nada coklat lurus ke depan untuk menilai setinggi profil rasa kompleks yang berubah secara keseluruhan. Genetika, terroir, teknik pasca panen, pemrosesan dan penyimpanan semuanya dapat didiskusikan ketika mempertimbangkan komponen rasa.

Tekstur memiliki pengaruh yang besar pada keseluruhan pengalaman dan juga mungkin masalah terkait tekstur mempengaruhi rasa. Ini adalah cara yang baik untuk mengevaluasi visi pembuat, perhatian terhadap detail dan tingkat kemahiran.

Aftermelt adalah pengalaman setelah cokelat meleleh. Cokelat dengan kualitas lebih tinggi akan bertahan lama dan nikmat. Karena peleburan adalah kesan terakhir yang Anda peroleh dari cokelat, itu sama pentingnya dalam peringkat keseluruhan.

Nama Anggota :

- Ahmad Shobirin 3311901047
- Nuzela Rahma Fitra 3311901058
- Indah Fauziyyah Utami 3311901060

Opini Keseluruhan benar-benar di mana peringkat mencerminkan opini subjektif. Idealnya ini adalah evaluasi saya apakah komponen di atas bekerja sama atau tidak dan pendapat tentang pengembangan rasa, karakter, dan gaya. Di sinilah setiap cokelat biasanya dapat diringkas oleh kesan paling menonjol yang akan Anda ingat tentang setiap cokelat.

Peringkat ini disusun oleh Brady Brelinski, Anggota Pendiri Manhattan Chocolate Society.

Eksplorasi

- Tahun kapan rating cocoa tertinggi terjadi ?
- Berapa nilai cocoa rating terendah ?
- Apa nama bar yang memiliki rating tertinggi ?
- Company apa saja yang memiliki lokasi di Netherlands?

Load Data Chocolate Bar Ratings

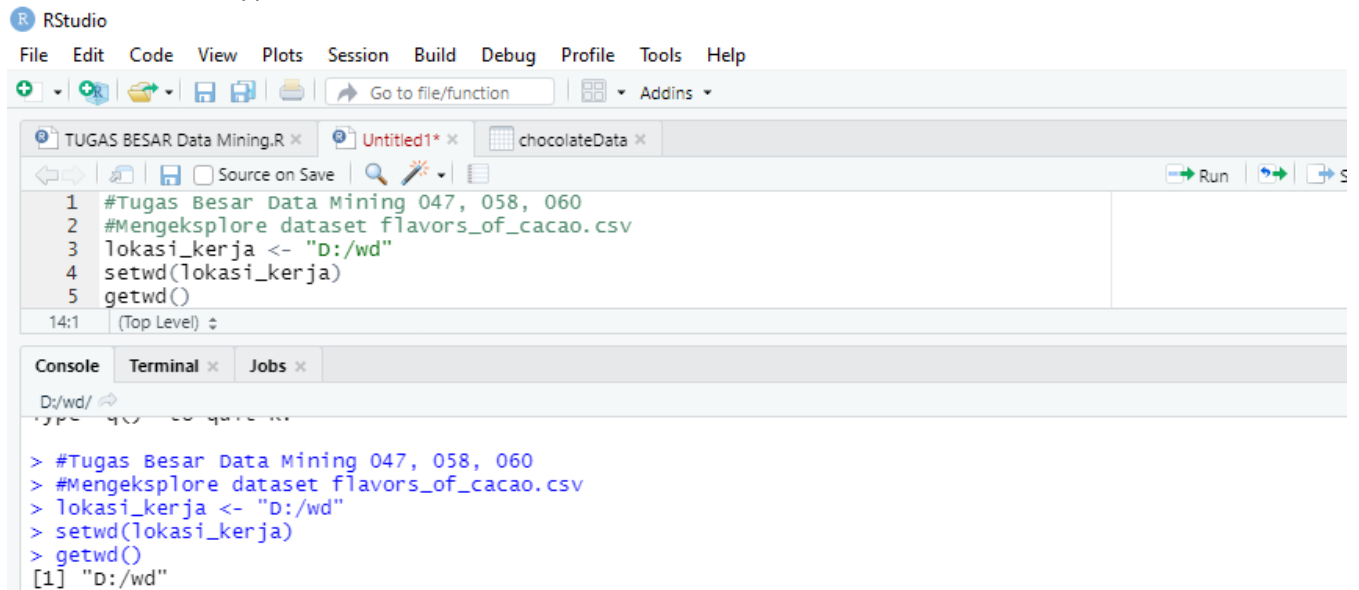
Download dan simpan dalam folder D:/wd file .csv yang akan digunakan sebagai dataset dalam pengerjaan tugas ini.

Company	Specific Bean Origin	Bar Name	REF	ReviewDate	CocoaPercent	CompanyLocation
A. Morin	Agua Grande		1876	2016	0,63	France
A. Morin	Kpime		1676	2015	0,7	France
A. Morin	Atsane		1676	2015	0,7	France
A. Morin	Akata		1680	2015	0,7	France
A. Morin	Quilla		1704	2015	0,7	France
A. Morin	Carenero		1315	2014	0,7	France
A. Morin	Cuba		1315	2014	0,7	France
A. Morin	Sur del Lago		1315	2014	0,7	France
A. Morin	Puerto Cabello		1319	2014	0,7	France
A. Morin	Pablino		1319	2014	0,7	France
A. Morin	Panama		1011	2013	0,7	France
A. Morin	Madagascar		1011	2013	0,7	France
A. Morin	Brazil		1011	2013	0,7	France
A. Morin	Equateur		1011	2013	0,7	France
A. Morin	Colombie		1015	2013	0,7	France
A. Morin	Birmanie		1015	2013	0,7	France
A. Morin	Papua New Guinea		1015	2013	0,7	France
A. Morin	Chuao		1015	2013	0,7	France
A. Morin	Piura		1019	2013	0,7	France
A. Morin	Chanchamayo Province		1019	2013	0,7	France

Menyiapkan lembar kerja dan lokasi file

Nama Anggota :

- Ahmad Shobirin 3311901047
- Nuzela Rahma Fitra 3311901058
- Indah Fauziyyah Utami 3311901060

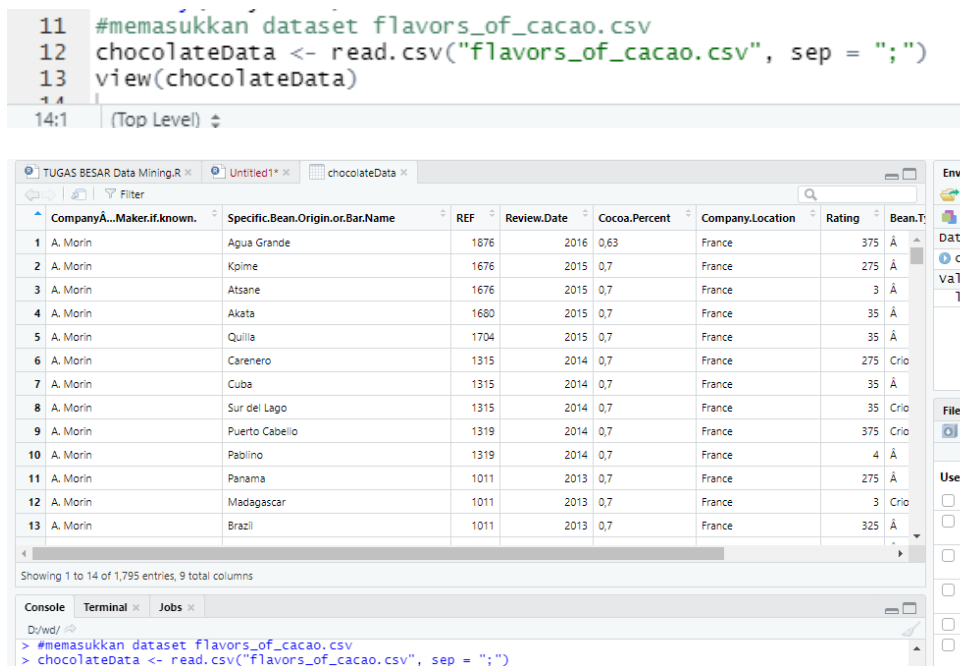


```
#Tugas Besar Data Mining 047, 058, 060
#Mengeksplor dataset flavors_of_cacao.csv
lokasi_kerja <- "D:/wd"
setwd(lokasi_kerja)
getwd()
```

Console

```
D:/wd/
> #Tugas Besar Data Mining 047, 058, 060
> #Mengeksplor dataset flavors_of_cacao.csv
> lokasi_kerja <- "D:/wd"
> setwd(lokasi_kerja)
> getwd()
[1] "D:/wd"
```

Menampilkan dataset flavors_of_cacao.csv



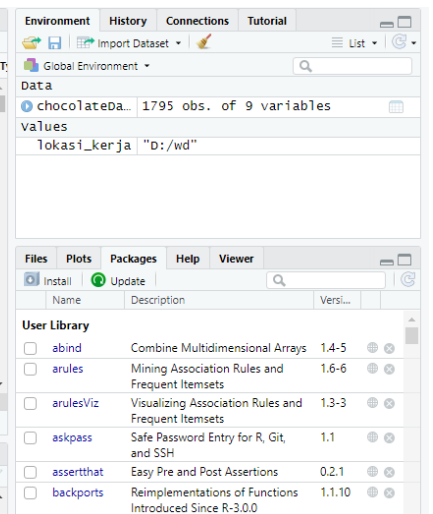
```
#memasukkan dataset flavors_of_cacao.csv
chocolateData <- read.csv("flavors_of_cacao.csv", sep = ";")
view(chocolateData)
```

	Company&...Maker.if.known.	Specific.Bean.Origin.or.Bar.Name	REF	Review.Date	Cocoa.Percent	Company.Location	Rating	Bean.T
1	A. Morin	Agua Grande	1876	2016	0,63	France	375	Å
2	A. Morin	Kpime	1676	2015	0,7	France	275	Å
3	A. Morin	Atsane	1676	2015	0,7	France	3	Å
4	A. Morin	Akata	1680	2015	0,7	France	35	Å
5	A. Morin	Quilla	1704	2015	0,7	France	35	Å
6	A. Morin	Carenero	1315	2014	0,7	France	275	Crio
7	A. Morin	Cuba	1315	2014	0,7	France	35	Å
8	A. Morin	Sur del Lago	1315	2014	0,7	France	35	Crio
9	A. Morin	Puerto Cabello	1319	2014	0,7	France	375	Crio
10	A. Morin	Pablino	1319	2014	0,7	France	4	Å
11	A. Morin	Panama	1011	2013	0,7	France	275	Å
12	A. Morin	Madagascar	1011	2013	0,7	France	3	Crio
13	A. Morin	Brazil	1011	2013	0,7	France	325	Å

Showing 1 to 14 of 1,795 entries, 9 total columns

Console

```
D:/wd/
> #memasukkan dataset flavors_of_cacao.csv
> chocolateData <- read.csv("flavors_of_cacao.csv", sep = ";")
```



Environment

Global Environment

Data

chocolateData... 1795 obs. of 9 variables

Values

lokasi_kerja "D:/wd"

Files Plots Packages Help Viewer

Install Update

User Library

Name	Description	Vers...
abind	Combine Multidimensional Arrays	1.4-5
arules	Mining Association Rules and Frequent Itemsets	1.6-6
arulesViz	Visualizing Association Rules and Frequent Itemsets	1.3-3
askpass	Safe Password Entry for R, Git, and SSH	1.1
assertthat	Easy Pre and Post Assertions	0.2.1
backports	Reimplementations of Functions Introduced Since R-3.0.0	1.1.10

Nama Anggota :

- Ahmad Shobirin 3311901047
- Nuzela Rahma Fitra 3311901058
- Indah Fauziyyah Utami 3311901060

```
14 head(chocolateData)
15 # some of our column names have spaces in them. This line changes the column names to
16 # versions without spaces, which let's us talk about the columns by their names.
17 names(chocolateData) <- make.names(names(chocolateData), unique=TRUE)
18 #Now we want to make sure that it all read in correctly, and get an idea of what's in our data file.
19 # the head() merupakan fungsi untuk menampilkan line/baris pertama dari file
20 head(chocolateData)
21 |
```

21:1 (Top Level) R Script

Console Terminal Jobs

D:/wd/

```
> # some of our column names have spaces in them. This line changes the column names to
> # versions without spaces, which let's us talk about the columns by their names.
> names(chocolateData) <- make.names(names(chocolateData), unique=TRUE)
> #Now we want to make sure that it all read in correctly, and get an idea of what's in our data file.
> # the head() merupakan fungsi untuk menampilkan line/baris pertama dari file
> head(chocolateData)
```

	CompanyA...Maker.if.known.	Specific.Bean.origin.or.Bar.Name	REF	Review.Date	Cocoa.Percent	Company.Location
1	A. Morin	Agua Grande	1876	2016	0,63	France
2	A. Morin	Kpime	1676	2015	0,7	France
3	A. Morin	Atsane	1676	2015	0,7	France
4	A. Morin	Akata	1680	2015	0,7	France
5	A. Morin	Quilla	1704	2015	0,7	France
6	A. Morin	Carenero	1315	2014	0,7	France

Rating.Bean.Type.Broad.Bean.Origin

Penamaan Objek

Berikut aturan penamaan objek di R.

- Menggunakan kombinasi alfabet (a-z, A-Z), angka (0-9), titik atau *underscore*.
- Diawali alfabet, titik atau *underscore*. Tidak boleh diawali dengan angka.
- Tidak mengandung spasi, tab atau karakter khusus seperti !, @, # dan lainnya.
- Sebaiknya tidak menggunakan beberapa penamaan atau nilai yang sudah digunakan oleh R (function dan keyword lainnya). Misalnya c, q, TRUE, FALSE, df, dt, rnorm, runif, rf, exp, dan lain-lain. Untuk mengetahui nama-nama yang sudah digunakan oleh R Anda dapat mengetikkan perintah ?reserved di console RStudio Anda.

Pertama kita mulai manajemen data dengan mencari tahun rating tertinggi

```
jawab1=chocolateData[(which.max(chocolateData$Rating)),4]
```

disini jawab1 sebagai objeknya.

```
> #tahun kapan rating cocoa tertinggi ?
> jawab1=chocolateData[(which.max(chocolateData$Rating)),4]
> print(jawab1)
[1] 2016
> |
```

Maka didapat lah tahun 2016

Fungsi which dapat juga menampilkan nilai terendah pada data tertentu :

```
> #Berapa nilai rating cocoa terendah ?
> jawab2=chocolateData[which.min(chocolateData$Rating)]
Error in chocolateData[which.min(chocolateData$Rating)] :
could not find function "chocolateData"
> jawab2=which.min(chocolateData$Rating)
> print(jawab2)
[1] 327
> |
```

Nama Anggota :

- Ahmad Shobirin 3311901047
- Nuzela Rahma Fitra 3311901058
- Indah Fauziyyah Utami 3311901060

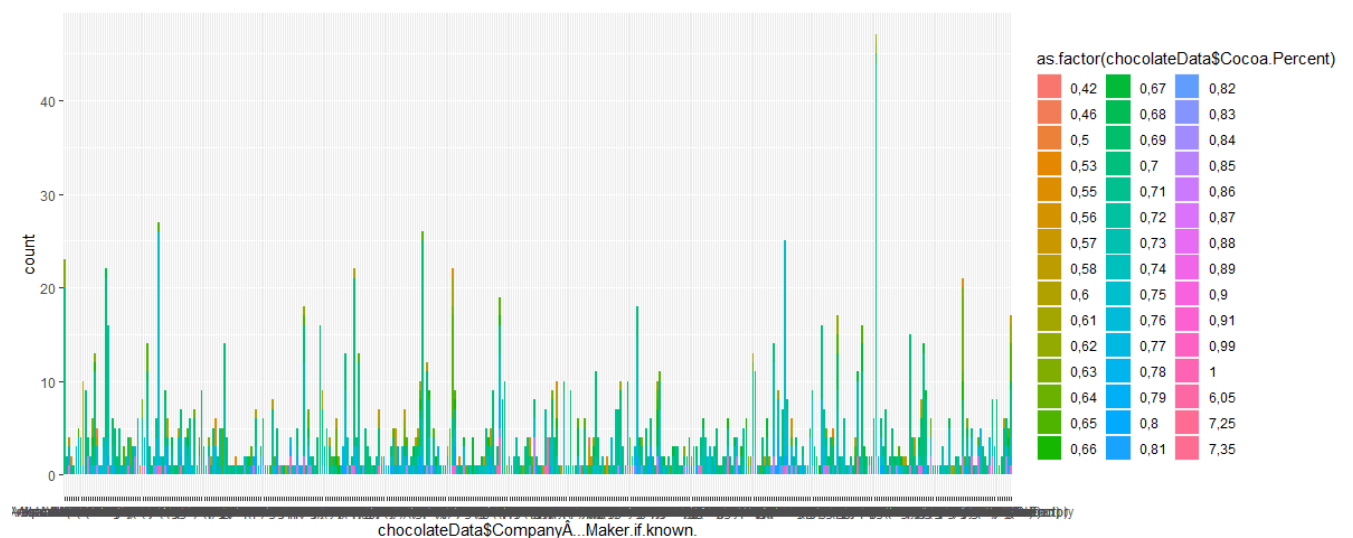
```
> #Apa nama bar yang memiliki rating tertinggi ?
> jawab4=chocolateData[which.max(chocolateData$Rating),2]
> print(jawab4)
[1] "Agua Grande"
>
```

Mengetahui perusahaan mana saja yang berlokasi di netherlands

```
> #Company apa saja yang memiliki lokasi di netherlands?
> jawab5=chocolateData[(which(chocolateData$Company.Location=="Netherlands")),2]
> print(jawab5)
[1] "winak Coop, Napo"      "La Dalia, Matagalpa"
[3] "Tien Giang"           "Makwale village, Kyela"
>
```

Selanjutnya Anda dapat install juga package {ggplot2}. Package ini sangat berguna untuk membuat visualisasi data di R.

```
> #plot data
> library(ggplot2)
> ggplot(data=chocolateData)+
+   geom_bar(mapping = aes(chocolateData$Company...Maker.if.known.,
+                           fill=as.factor(chocolateData$Cocoa.Percent)))
```



Dari data diatas terlihat warna – warna data yang tertera yang merupakan frekuensi data dari persentase cocoa.

Referensi

<https://www.kaggle.com/rtatman/chocolate-bar-ratings>

<https://bookdown.org/aepstk/intro/basicr.html#naming>

Nama Anggota :

- Ahmad Shobirin 3311901047
- Nuzela Rahma Fitra 3311901058
- Indah Fauziyyah Utami 3311901060