

Kuliah 3 STA1381 – Pengantar Sains Data

Septian Rahardiantoro



## Outline

- Manajemen data
- Visualisasi data

**Manajemen Data** 

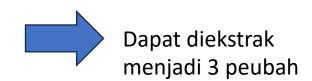
# Manajemen Data

 Pada data web scraping yang sudah diperoleh, akan dilakukan proses manajemen data

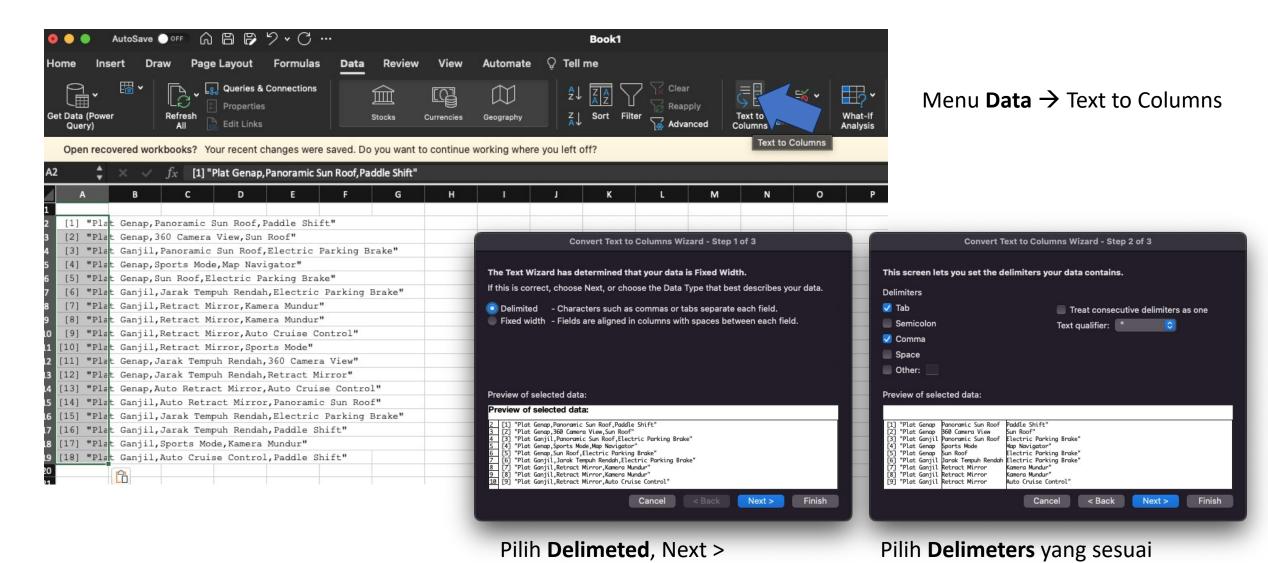
- Manajemen data dimaksudkan untuk:
  - Menyesuaikan format peubah (numerik, teks)
  - Menyusun peubah-peubah tertentu

#### 1. Perhatikan data hasil web scraping Data Mobil Bekas

```
[1] "Plat Genap, Panoramic Sun Roof, Paddle Shift"
 [2] "Plat Genap, 360 Camera View, Sun Roof"
    "Plat Ganjil, Panoramic Sun Roof, Electric Parking Brake"
 [4] "Plat Genap, Sports Mode, Map Navigator"
    "Plat Genap, Sun Roof, Electric Parking Brake"
    "Plat Ganjil, Jarak Tempuh Rendah, Electric Parking Brake"
 [7] "Plat Ganjil, Retract Mirror, Kamera Mundur"
    "Plat Ganjil, Retract Mirror, Kamera Mundur"
    "Plat Ganjil, Retract Mirror, Auto Cruise Control"
[10] "Plat Ganjil, Retract Mirror, Sports Mode"
[11] "Plat Genap, Jarak Tempuh Rendah, 360 Camera View"
    "Plat Genap, Jarak Tempuh Rendah, Retract Mirror"
[13] "Plat Genap, Auto Retract Mirror, Auto Cruise Control"
[14] "Plat Ganjil, Auto Retract Mirror, Panoramic Sun Roof"
    "Plat Ganjil, Jarak Tempuh Rendah, Electric Parking Brake"
[16] "Plat Ganjil, Jarak Tempuh Rendah, Paddle Shift"
[17] "Plat Ganjil, Sports Mode, Kamera Mundur"
[18] "Plat Ganjil, Auto Cruise Control, Paddle Shift"
```



#### Ekstraksi peubah → dengan Ms. Excel



(dalam hal ini "Comma"), Next >

- [1] "Plat Genap, Panoramic Sun Roof, Paddle Shift"
- [2] "Plat Genap, 360 Camera View, Sun Roof"
- [3] "Plat Ganjil, Panoramic Sun Roof, Electric Parking Brake"
- [4] "Plat Genap, Sports Mode, Map Navigator"
- [5] "Plat Genap, Sun Roof, Electric Parking Brake"
- [6] "Plat Ganjil, Jarak Tempuh Rendah, Electric Parking Brake"
- [7] "Plat Ganjil, Retract Mirror, Kamera Mundur"
- [8] "Plat Ganjil, Retract Mirror, Kamera Mundur"
- [9] "Plat Ganjil, Retract Mirror, Auto Cruise Control"
- [10] "Plat Ganjil, Retract Mirror, Sports Mode"
- [11] "Plat Genap, Jarak Tempuh Rendah, 360 Camera View"
- [12] "Plat Genap, Jarak Tempuh Rendah, Retract Mirror"
- [13] "Plat Genap, Auto Retract Mirror, Auto Cruise Control"
- [14] "Plat Ganjil, Auto Retract Mirror, Panoramic Sun Roof"
- [15] "Plat Ganjil, Jarak Tempuh Rendah, Electric Parking Brake"
- [16] "Plat Ganjil, Jarak Tempuh Rendah, Paddle Shift"
- [17] "Plat Ganjil, Sports Mode, Kamera Mundur"
- [18] "Plat Ganjil, Auto Cruise Control, Paddle Shift"



Plat Genap	Panoramic Sun Roof	Paddle Shift
Plat Genap	360 Camera View	Sun Roof
Plat Ganjil	Panoramic Sun Roof	Electric Parking Brake
Plat Genap	Sports Mode	Map Navigator
Plat Genap	Sun Roof	Electric Parking Brake
Plat Ganjil	Jarak Tempuh Rendah	Electric Parking Brake
Plat Ganjil	Retract Mirror	Kamera Mundur
Plat Ganjil	Retract Mirror	Kamera Mundur
Plat Ganjil	Retract Mirror	Auto Cruise Control
Plat Ganjil	Retract Mirror	Sports Mode
Plat Genap	Jarak Tempuh Rendah	360 Camera View
Plat Genap	Jarak Tempuh Rendah	Retract Mirror
Plat Genap	Auto Retract Mirror	Auto Cruise Control
Plat Ganjil	Auto Retract Mirror	Panoramic Sun Roof
Plat Ganjil	Jarak Tempuh Rendah	Electric Parking Brake
Plat Ganjil	Jarak Tempuh Rendah	Paddle Shift
Plat Ganjil	Sports Mode	Kamera Mundur
Plat Ganjil	Auto Cruise Control	Paddle Shift

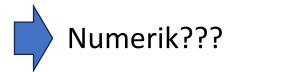
#### Bagaimana caranya?

### 2. Perhatikan data hasil web scraping Data Handphone Android

web-scraper-order	web-scraper-start-url	nama_hp	harga_hp	
1692063197-1	https://www.blibli.com/cari/handphone%20android	Xiaomi Redmi 9C Smartphone [4GB/ 64GB]	Rp1.589.000	
1692063197-2	https://www.blibli.com/cari/handphone%20android	Asus Zenfone 3 ZE520KL Smartphone [32GB/ 4GB]	Rp1.299.000	
1692063197-3	https://www.blibli.com/cari/handphone%20android	Xiaomi Poco M3 Pro 5G Smartphone [4GB/64GB]	Rp2.250.000	
1692063197-4	https://www.blibli.com/cari/handphone%20android	Samsung Galaxy S21 FE 8/256GB RAM 8 ROM 256 Handphone Hp Smartphone Android Garansi Resmi	Rp7.099.000	
1692063197-5	https://www.blibli.com/cari/handphone%20android	Vivo Y15S 3/64 GB RAM 3 ROM 64GB Handphone Hp Smartphone Android Original Murah Garansi Resmi	Rp1.480.000	
1692063197-6	https://www.blibli.com/cari/handphone%20android	OPPO A31 Handphone Android - Black [4/128]	Rp2.547.902	
1692063197-7	https://www.blibli.com/cari/handphone%20android	Advan NasaPlus 2GB 16GB Smartphone Android10 Garansi Resmi	Rp845.000	
1692063197-8	https://www.blibli.com/cari/handphone%20android	OPPO A57 4/64 GB Smartphone Android Original Garansi Resmi	Mulai	Rp2.399.00
1692063197-9	https://www.blibli.com/cari/handphone%20android	Advan NasaPlus 2GB 16GB Smartphone Android10 Garansi Resmi Deep Blue	Rp845.000	
1692063197-10	https://www.blibli.com/cari/handphone%20android		Rp2.100.000	
1692063197-11	https://www.blibli.com/cari/handphone%20android	Samsung Galaxy A04e 3/32 GB RAM 3GB 32GB Smartphone Android	Rp1.137.000	
1692063197-12	https://www.blibli.com/cari/handphone%20android		Rp1.475.000	
1692063197-13	https://www.blibli.com/cari/handphone%20android	·	Rp1.258.000	
1692063197-14	https://www.blibli.com/cari/handphone%20android	·	Rp700.000	
1692063197-15		XIAOMI REDMI 10A RAM 3GB - 64GB GARANSI RESMI	Mulai	Rp1.149.00
1692063197-16	https://www.blibli.com/cari/handphone%20android		Rp7.899.000	
1692063197-17	https://www.blibli.com/cari/handphone%20android		Mulai	Rp2.500.00
1692063197-18	https://www.blibli.com/cari/handphone%20android		Rp2.750.000	Кр2.300.00
1692063197-19	https://www.blibli.com/cari/handphone%20android		Rp799.000	
1692063197-20	https://www.blibli.com/cari/handphone%20android	·	Rp8.518.000	
1692063197-21	https://www.blibli.com/cari/handphone%20android		Rp680.000	
1692063197-21	https://www.blibli.com/cari/handphone%20android	HP Oppo A53S murah terbaru 2022   OPPO A53 S 5G RAM 8 + 256 GB   Original New Baru Smartphone Android 6.52in HP OPPO Murah garansi 1 tahun	Rp2.459.000	
1692063197-23	https://www.blibli.com/cari/handphone%20android		Rp7.149.000	
1692063197-24	https://www.blibli.com/cari/handphone%20android	• • • • • • • • • • • • • • • • • • • •	Rp1.500.000	
1692063197-25	https://www.blibli.com/cari/handphone%20android		Rp2.399.000	
1692063197-26				
	https://www.blibli.com/cari/handphone%20android		Rp6.199.000	
1692063197-27	https://www.blibli.com/cari/handphone%20android		Rp3.231.000	
1692063197-28	https://www.blibli.com/cari/handphone%20android		Rp3.699.000	
1692063197-29	https://www.blibli.com/cari/handphone%20android		Rp3.849.000	
1692063197-30	https://www.blibli.com/cari/handphone%20android		Rp5.599.000	-
1692063197-31	https://www.blibli.com/cari/handphone%20android		Rp5.999.000	
1692063197-32	https://www.blibli.com/cari/handphone%20android		Rp1.199.000	
1692063197-33	https://www.blibli.com/cari/handphone%20android		Rp1.625.000	
1692063197-34	https://www.blibli.com/cari/handphone%20android		Rp1.799.000	
1692063197-35	https://www.blibli.com/cari/handphone%20android		Mulai	Rp1.799.00
1692063197-36	https://www.blibli.com/cari/handphone%20android	TIONAL NI DECLEA MONIONI NONDESTA NOLINAN	Rp2.269.000	
1692063197-37	https://www.blibli.com/cari/handphone%20android		Rp2.487.500	
1692063197-38	https://www.blibli.com/cari/handphone%20android	Ç , ·	Mulai	Rp3.799.00
1692063197-39	https://www.blibli.com/cari/handphone%20android	,	Rp1.110.000	
1692063197-40	https://www.blibli.com/cari/handphone%20android	Vivo Y16 Smartphone [4GB/128GB]	Rp1.899.000	
1692063197-41	https://www.blibli.com/cari/handphone%20android	REALME C31 RAM 3/32 GB GARANSI RESMI	Rp1.470.000	
1692063197-42	https://www.blibli.com/cari/handphone%20android	Redmi 9A 2/32GB	Rp1.500.500	
1692063197-43	https://www.blibli.com/cari/handphone%20android	Infinix Note 7 Lite Smartphone [4GB- 64GB]	R	
1692063197-44	https://www.blibli.com/cari/handphone%20android	Xiaomi redmi 9T RAM 4/64 GB garansi resmi	🖪 Belum ni	umeri
1692063197-45	https://www.blibli.com/cari/handphone%20android	Samsung M23 5G 6/128 garansi resmi	R	

#### Penyesuaian peubah harga untuk menjadi peubah numerik

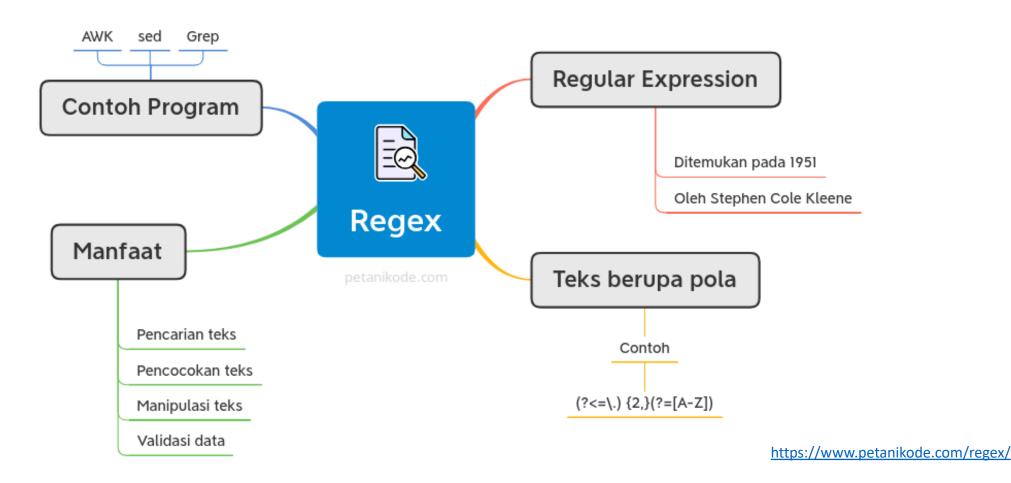
- · · <b>/</b> · · · · ·	
harga_hp	
Rp1.589.000	
Rp1.299.000	
Rp2.250.000	
Rp7.099.000	
Rp1.480.000	
Rp2.547.902	
Rp845.000	
Mulai	Rp2.399.00
Rp845.000	
Rp2.100.000	
Rp1.137.000	
Rp1.475.000	
Rp1.258.000	
Rp700.000	
Mulai	Rp1.149.00
Rp7.899.000	
Mulai	Rp2.500.00
Rp2.750.000	
Rp799.000	
Rp8.518.000	
Rp680.000	
Rp2.459.000	
Rp7.149.000	
Rp1.500.000	
Rp2.399.000	
Rp6.199.000	
Rp3.231.000	
Rp3.699.000	
Rp3.849.000	
Rp5.599.000	
Rp5.999.000	
Rp1.199.000	
Rp1.625.000	
Rp1.799.000	
Mulai	Rp1.799.00
Rp2.269.000	
Rp2.487.500	
Mulai	Rp3.799.00
Rp1.110.000	
Rp1.899.000	
Rp1.470.000	
Rp1.500.500	
Rp2.100.000	
Rp2.475.000	
Rp3.309.800	

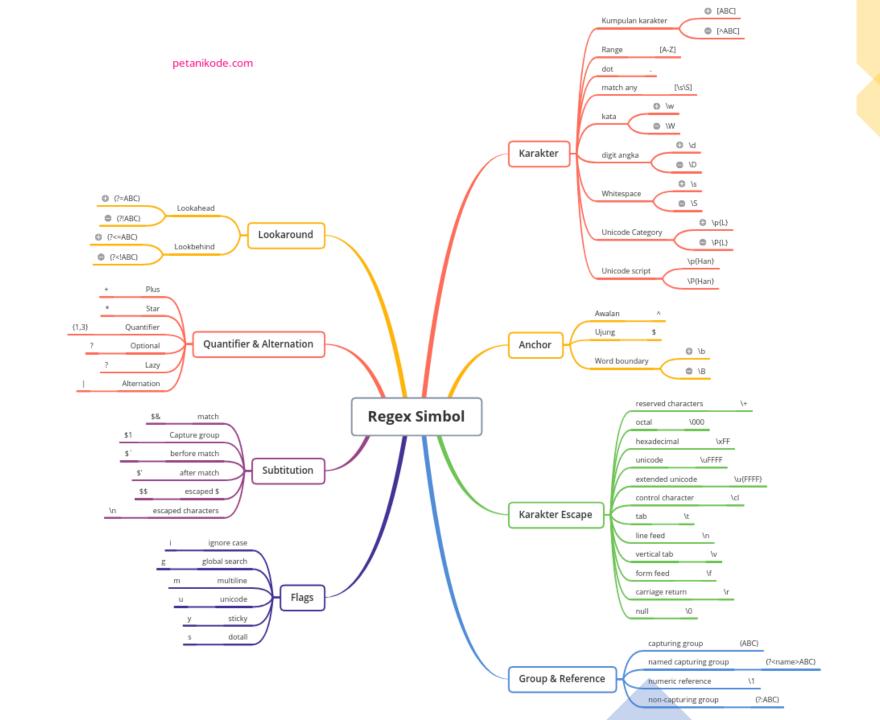


Gunakan Regular Expression (RegEx)

## Regular Expression (RegEx)

→ Susunan string yang mendefinisikan pola pencarian sehingga dapat membantu kita untuk melakukan matching (pencocokan), locate (pencarian), dan manipulasi teks.





#### Beberapa simbol yang sering digunakan:

```
    → untuk grouping;
    → simbol untuk logika or (atau alternatif);
    → untuk mencari karekter dari huruf a sampai z;
    → untuk mencari karakter dari huruf A sampai Z;
    → Untuk mencari angka 0 sampai 9;
    → untuk mencari kata;
    → untuk mencari angka;
```

Dapat diterapkan di R

#### Regular Expression (RegEx) di R

- R mendukung regular expression
- Dapat pula menggunakan package tambahan, seperti stringr (recommended), stringi, dll
- Beberapa fungsi di antaranya:

Fungsi	stringr	R Base
Mengembalikan semua element string yang cocok	str_subset()	grep()
Memeriksa keberadaan pola dalam string	str_detect()	grepl()
Ekstraksi pola yang cocok dari suatu string	<pre>str_extract(), str_extract_all()</pre>	<pre>regexpr(), regmatches(), grep()</pre>
Menemukan posisi dari pola dalam suatu string	str_locate()	regexpr()
Menemukan semua posisi dari pola dalam suatu string	str_locate_all()	gregexpr()
Mengganti pola yang cocok dari suatu string	str_replace()	sub()
Mengganti semua yang cocok dg pola dari suatu string	str_replace_all()	gsub()

https://stringr.tidyverse.org/articles/from-base.html

#### Ilustrasi di R

```
> str_extract_all(perusahaan, '[0-9]+')
> str subset(perusahaan, 'a')
[1] "Roti Bangsawan" "RB@email.id"
                                                    [[1]]
[3] "roti_bang29@gmail.com"
                                                    character(0)
> str_detect(perusahaan, '[0-9]')
[1] FALSE TRUE FALSE TRUE TRUE
                                                    [[2]]
                                                    [1] "08128712"
> str_extract_all(perusahaan, '[0-9]')
[[1]]
character (0)
                                                    [[3]]
                                                    character(0)
[[2]]
[1] "0" "8" "1" "2" "8" "7" "1" "2"
                                                    [[4]]
                                                    [1] "29"
[[3]]
character (0)
                                                    [[5]]
                                                    [1] "3823"
[[4]]
[1] "2" "9"
                                                    > str_extract(perusahaan, '[a-z]+')
                                                    [1] "oti" NA "email" "roti" NA
[[5]]
[1] "3" "8" "2" "3"
```

```
> str extract all(perusahaan, '[a-z]+')
[[1]]
[1] "oti" "angsawan"
[[2]]
character(0)
[[3]]
[1] "email" "id"
[[4]]
[1] "roti" "bang" "gmail" "com"
[[5]]
character(0)
> str extract all(perusahaan, '[A-Za-z]+')
[[1]]
[1] "Roti" "Bangsawan"
[[2]]
character (0)
[[3]]
[1] "RB" "email" "id"
[[4]]
[1] "roti" "bang" "gmail" "com"
[[5]]
[1] "F"
        "KSB"
```

#### Kembali ke data

#### Penyesuaian peubah harga untuk menjadi peubah numerik

,	
harga_hp	
Rp1.589.000	
Rp1.299.000	
Rp2.250.000	
Rp7.099.000	
Rp1.480.000	
Rp2.547.902	
Rp845.000	
Mulai	Rp2.399.00
Rp845.000	
Rp2.100.000	
Rp1.137.000	
Rp1.475.000	
Rp1.258.000	
Rp700.000	
Mulai	Rp1.149.00
Rp7.899.000	
Mulai	Rp2.500.00
Rp2.750.000	
Rp799.000	
Rp8.518.000	
Rp680.000	
Rp2.459.000	
Rp7.149.000	
Rp1.500.000	
Rp2.399.000	
Rp6.199.000	
Rp3.231.000	
Rp3.699.000	
Rp3.849.000	
Rp5.599.000	
Rp5.999.000	
Rp1.199.000	
Rp1.625.000	
Rp1.799.000	
Mulai	Rp1.799.00
Rp2.269.000	
Rp2.487.500	D=2.700.00
Mulai	Rp3.799.00
Rp1.110.000	
Rp1.899.000	
Rp1.470.000	
Rp1.500.500	
Rp2.100.000	
Rp2.475.000	
Rp3.309.800	



#### Ekstraksi peubah

nama hp	
Xiaomi Redmi 9C Smartphone [4GB/(64GB])	
Asus Zenfone 3 ZE520KL-Smartphone [32GB/ 4GB]	→ Memori Internal
Xiaomi Poco M3 Pro 5G Smartphone [4GB/64GB]	
Samsung Galaxy S21 FE 8/256GB RAM 8 ROM 256 Handphone Hp Smartphone Android Garansi Resmi	→ RAM
Vivo Y15S 3/64 GB RAM 3 ROM 64GB Handphone Hp Smartphone Android Original Murah Garansi Resmi	
OPPO A31 Handphone Android - Black [4/128]	Nama merek HP
Advan NasaPlus 2GB 16GB Smartphone Android10 Garansi Resmi	
OPPO A57 4/64 GB Smartphone Android Original Garansi Resmi	
Advan NasaPlus 2GB 16GB Smartphone Android10 Garansi Resmi Deep Blue	
Hp xiomi not 9	
Samsung Galaxy A04e 3/32 GB RAM 3GB 32GB Smartphone Android	
Infinix Hot 30i 8/128 RAM 8GB ROM 128GB 8 128 GB HP Android	
POCO C40 3/32 GB - Resmi	
Handphone OPPO A37	
XIAOMI REDMI 10A RAM 3GB - 64GB GARANSI RESMI	
Xiaomi MI 11T Pro 12/256 GB RAM 12 ROM 256GB Handphone Smartphone Android Original Garansi Resmi	
Samsung Galaxy A13 4/128 6/128 4GB 6GB 128GB RAM 4 6 ROM 128 GB HP Smartphone Android	
handphone samsung A12 6/128 resmi sein	

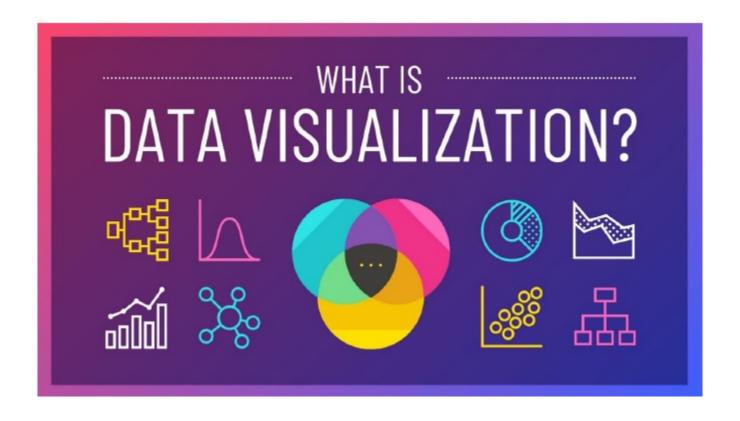
Bagaimana Caranya ???

#### Ekstraksi peubah Nama merek HP

```
##ekstraksi peubah
#ekstraksi nama merek HP
nama.hp <- data.hp$nama hp</pre>
#identifikasi list nama HP
list.nama.hp <- c("Redmi", "Poco", "Xiaomi", "Asus", "Samsung",</pre>
                    "Vivo", "OPPO", "Advan", "Infinix", "Huawei",
                    "Realme", "Oppo", "SAMSUNG", "Sharp", "POCO",
                    "RENO", "VIVO", "Evercross", "xiomi", "REDMI",
                    "samsung", "ALDO", "Tecno", "REALME", "XIAOMI",
                    "Wiko", "Evercoss", "Meizu", "Haier", "Lava", "Itel",
                    "infinix", "Alcatel", "MI", "ITEL", "Aldo", "Soyes",
                    "EVERCOSS", "Galaxy", "Mito", "TECNO", "oppo", "MITO",
                    "Maxtron", "opppo", "vivo", "VIvo", "SHARP", "Honor",
                    "ADVAN", "Nokia", "INFINIX", "redmi")
m1.cek <- str detect(nama.hp,paste(list.nama.hp, collapse = "|"))</pre>
sum(m1.cek)
nama.hp[!m1.cek]
#ekstrak nama HP
m2 <- str extract all(nama.hp,paste(list.nama.hp, collapse = "|"))</pre>
m3 <- NULL
for(i in 1:800){
  a <- paste(m2[i],collapse = " ")</pre>
  m3 < -c(m3,a)
                                                              Bagaiamana
                                                              Ekstraksi peubah
m4 <- str remove all(m3,"[c(\",)]")</pre>
                                                              RAM dan Memori Internal?
summary(as.factor(m4))
#namun masih perlu dicleaning!
```

## **Visualisasi Data**





penyajian informasi dalam bentuk gambar atau grafik yang diperoleh dari data melalui proses analitika



- 90% of the information transmitted to the brain is visual
- Humans process images 60,000 times faster than text
- 70% of our sensory receptors are in our eyes
- 65% of people are visual learners

In short, visual data is easier to remember than words.



Peranan Visualisasi dalam Proses Analisis Data selama proses analisis data

setelah proses analisis data

#### Peranan Visualisasi dalam Proses Analisis Data





- memahami karakteristik perilaku data yang dimiliki (eksplorasi)
- mengarahkan pendalaman-pendalaman analisis data (elaborasi)
- membangkitkan ide baru analisis data

#### Peranan Visualisasi dalam Proses Analisis Data

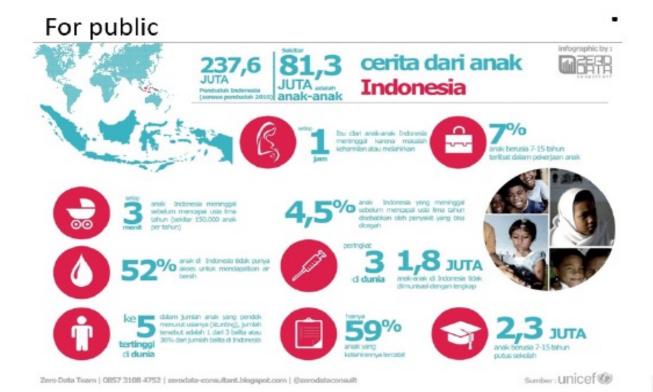




- Meringkas informasi pada data besar sehingga mudah melihat pola pada data dan memahami cerita di belakangnya
- Memudahkan menyampaikan informasi sehingga lebih cepat menyusun aksi
- Membantu pembuat keputusan memahami apa yang sedang atau telah terjadi
- Mengarahkan audiens untuk fokus pada hal yang dituju untuk diperhatikan
- Menyibak informasi yang tidak terlihat sebelumnya sehingga memudahkan menuliskan narasi dan laporan

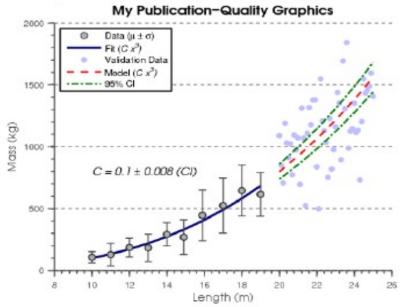
## **Target audiens**





For administrators

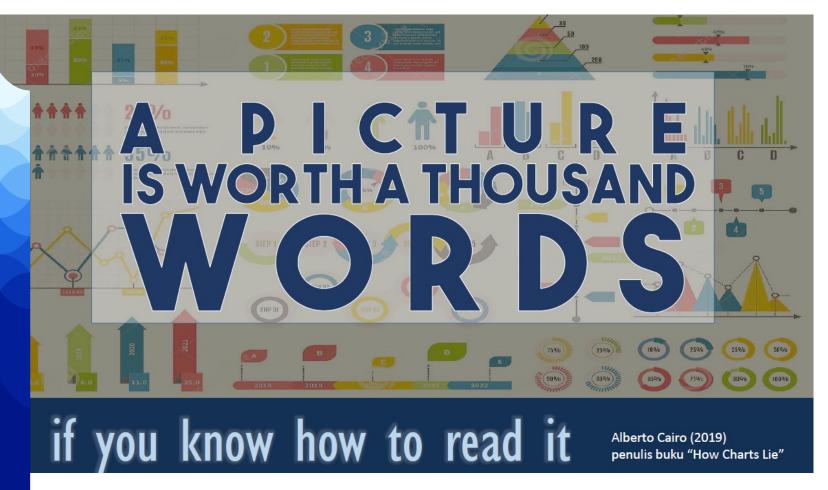
#### For professional colleagues





## Skala dalam grafik (visualisasi)

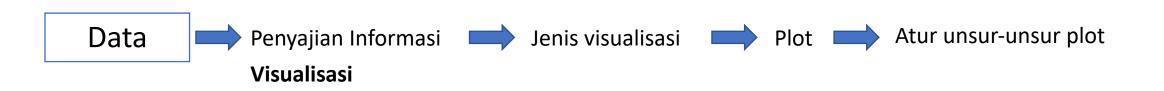




- Skala menerjemahkan nilai menjadi isyarat visual
- Pertanyaan utamanya adalah bagaimana jarak dalam grafik data diterjemahkan menjadi perbedaan kuantitas yang berarti
- Tersesat dengan skala itu mudah → berpotensi mendistorsi informasi

#### Jenis visualisasi

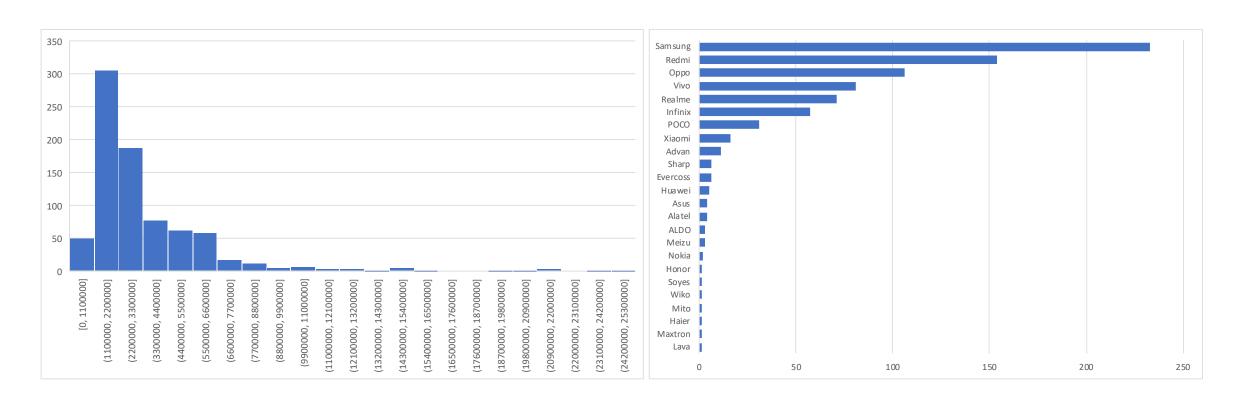
- 1. Visualisasi Besaran dan Perbandingannya
- 2. Visualisasi Sebaran Data dan Perbandingannya
- 3. Visualisasi Proporsi
- 4. Visualisasi Hubungan Antar Variabel
- Visualisasi Data Deret Waktu dan Trend
- 6. Visualisasi Geospatial
- 7. Visualisasi Aliran dan Jaringan



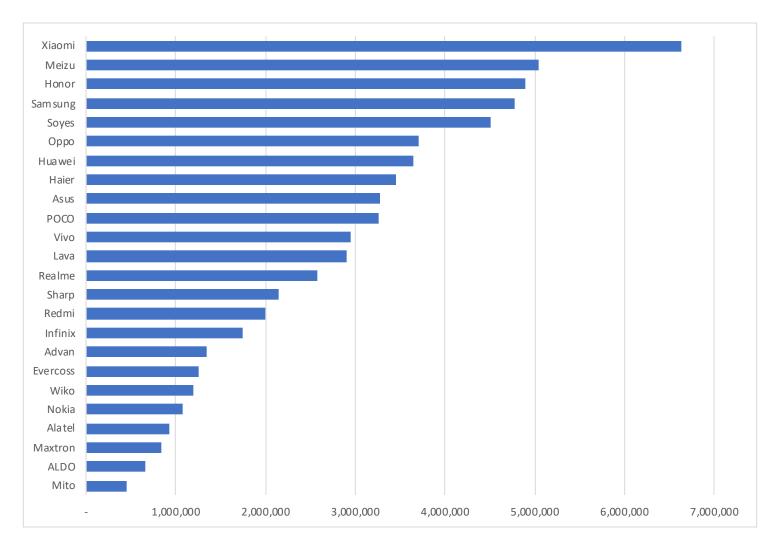
# Perhatikan kembali data hasil web scraping Data Handphone Android setelah dilakukan proses regular expression

Visualisasi peubah harga

Visualisasi peubah nama HP



#### Harga berdasarkan Nama HP



# Projek

Berdasarkan data scraping yang diperoleh untuk topik yang ditentukan:

Topik	P1	P2
Topik 1: Harga rumah dan spesifikasinya	K4	K2
Topik 2: Harga leptop dan spesifikasinya	K2	К3
Topik 3: Harga mobil bekas dan spesifikasinya	K1	K1
Topik 4: Harga handphone dan spesifikasinya	К3	K4

Lakukanlah manajemen dan visulaisasi data yang berguna!

# Terima kasih ©