

# FUNDAMENTAL DATA VISUALIZATION

*With Python*



# WHAT'S IN WHUP?



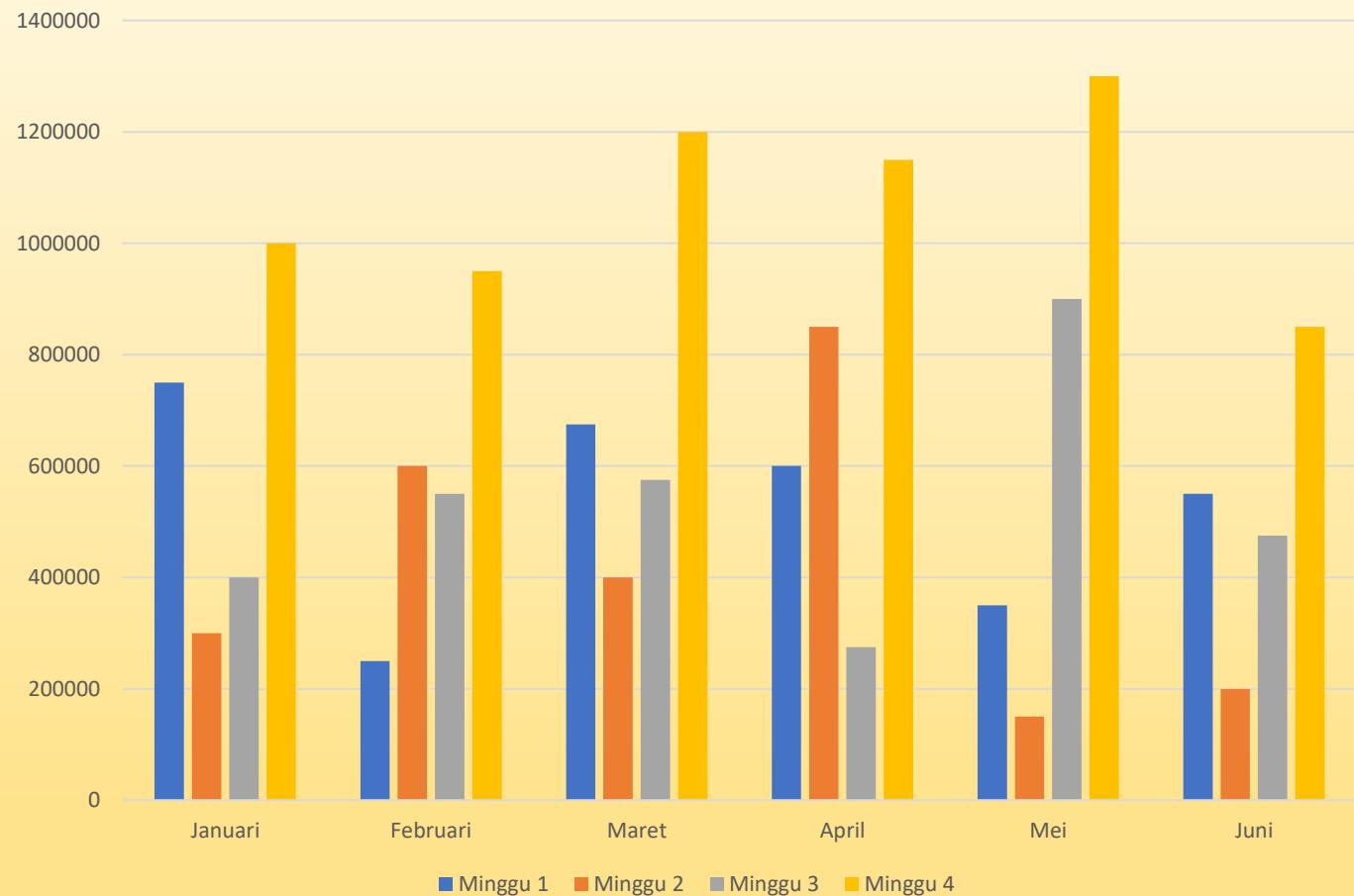
Representasi grafis dari informasi dan data dengan menggunakan elemen visual seperti bagan, grafik dan peta.

Tujuannya untuk melihat dan memahami tren, outliers dan pola dalam data yang digunakan dalam proses analisa dan pengambilan keputusan ;

## Pengeluaran Bulanan

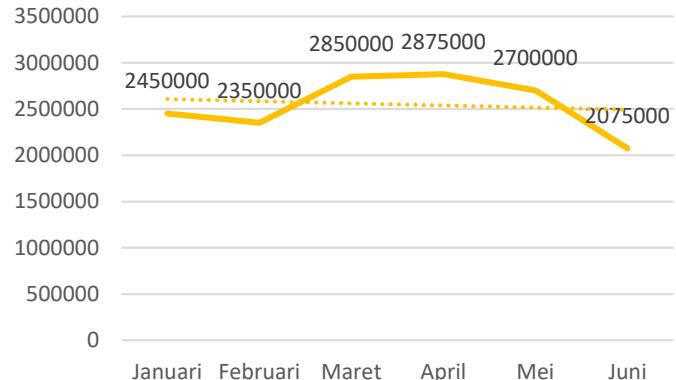
	Minggu 1	Minggu 2	Minggu 3	Minggu 4
Januari	750000	300000	400000	1000000
Februari	250000	600000	550000	950000
Maret	675000	400000	575000	1200000
April	600000	850000	275000	1150000
Mei	350000	150000	900000	1300000
Juni	550000	200000	475000	850000

## Pengeluaran Bulanan



# REPORT PENGELUARAN PRIBADI

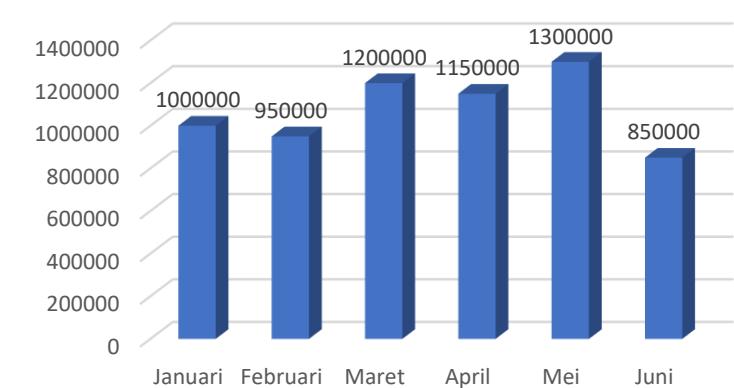
Sum Per Month



Average Per Month



Max Per Month

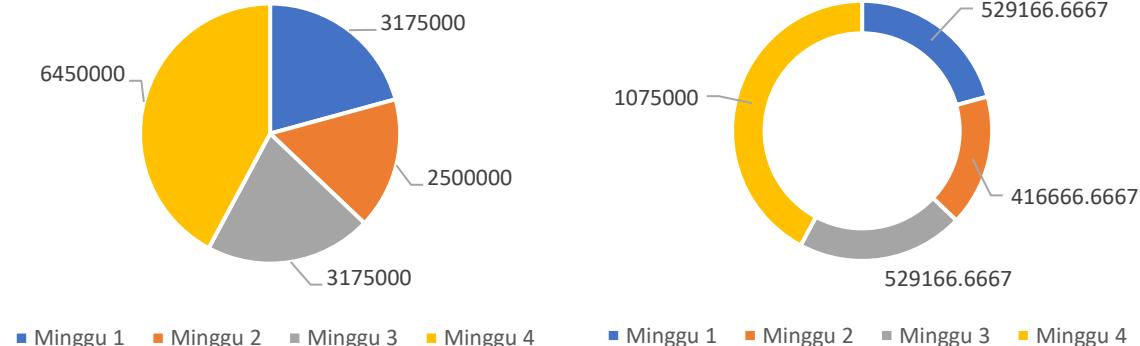


**MAX  
OUTCOME**

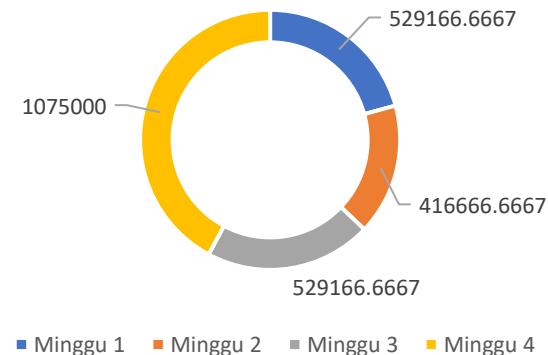
1

week

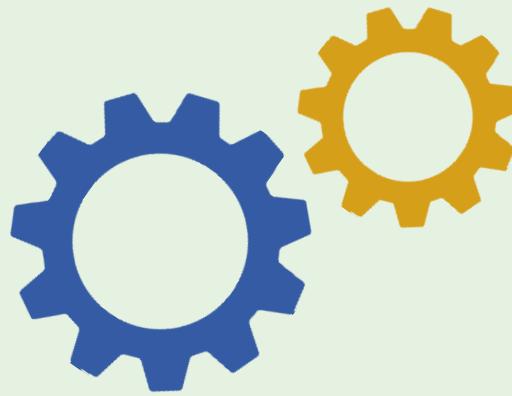
Sum Per Week



Average Per Week



- Pengeluaran bulanan bergerak fluktuatif
- Pengeluaran paling banyak terjadi di bulan April sebesar Rp. 2.875.000
- Rata – rata pengeluaran terbesar terjadi saat minggu 1
- Berdasarkan tradeline kemungkinan bulan depan pengeluaran bulanan menurun



# WHY DATAVIZ?

- 
1. LEBIH MUDAH MENGANALISA GRAFIK DARIPADA DATA NUMERIK / ANGKA
  2. KEEP OUR EYES TO THE MESSAGES OR INSIGHT
  3. PROSES ANALISA JAUH LEBIH CEPAT
  4. MEMPERMUDAH KOMUNIKASI ANTAR USER



# MEDIA DATAVIZ

- **TABEL**  
susunan data dalam bentuk baris dan kolom
- **DIAGRAM**  
representasi simbolis informasi dalam bentuk geometri dua dimensi sesuai teknik visualisasi



# TABEL



Pendapatan Toko Sederhana pada bulan Januari tahun 2021 tercatat Rp. 10.500.000, sedangkan pada bulan Februari di tahun yang sama pendapatan menurun 20% sehingga menjadi Rp. 8.400.000. Kabar baiknya di bulan Maret 2021 tercatat peningkatan pendapatan hingga 2 kali pendapatan dari bulan sebelumnya yakni sebesar Rp. 16.800.000 dan stabil saat bulan berikutnya yakni April 2021. Namun di bulan Mei 2021, toko mengalami kerugian sebesar Rp. 5.000.000 akibat diterapkannya lockdown lokal di wilayah tersebut.

Bulan	Pendapatan
Januari	10500000
Februari	8400000
Maret	16800000
April	16800000
Mei	11800000

# PENULISAN TABEL



- **PENULISAN JUDUL**

Pastikan judul mencangkup isi dari tabel

- **SIMPLE**

Desain tabel sesederhana mungkin

- **PENJELASAN SIMBOL**

Jika terdapat symbol, beri catatan kaki

- **PENEKANAN**

Memfokuskan perhatian pembaca pada pokok data

- **SUMBER TABLE**

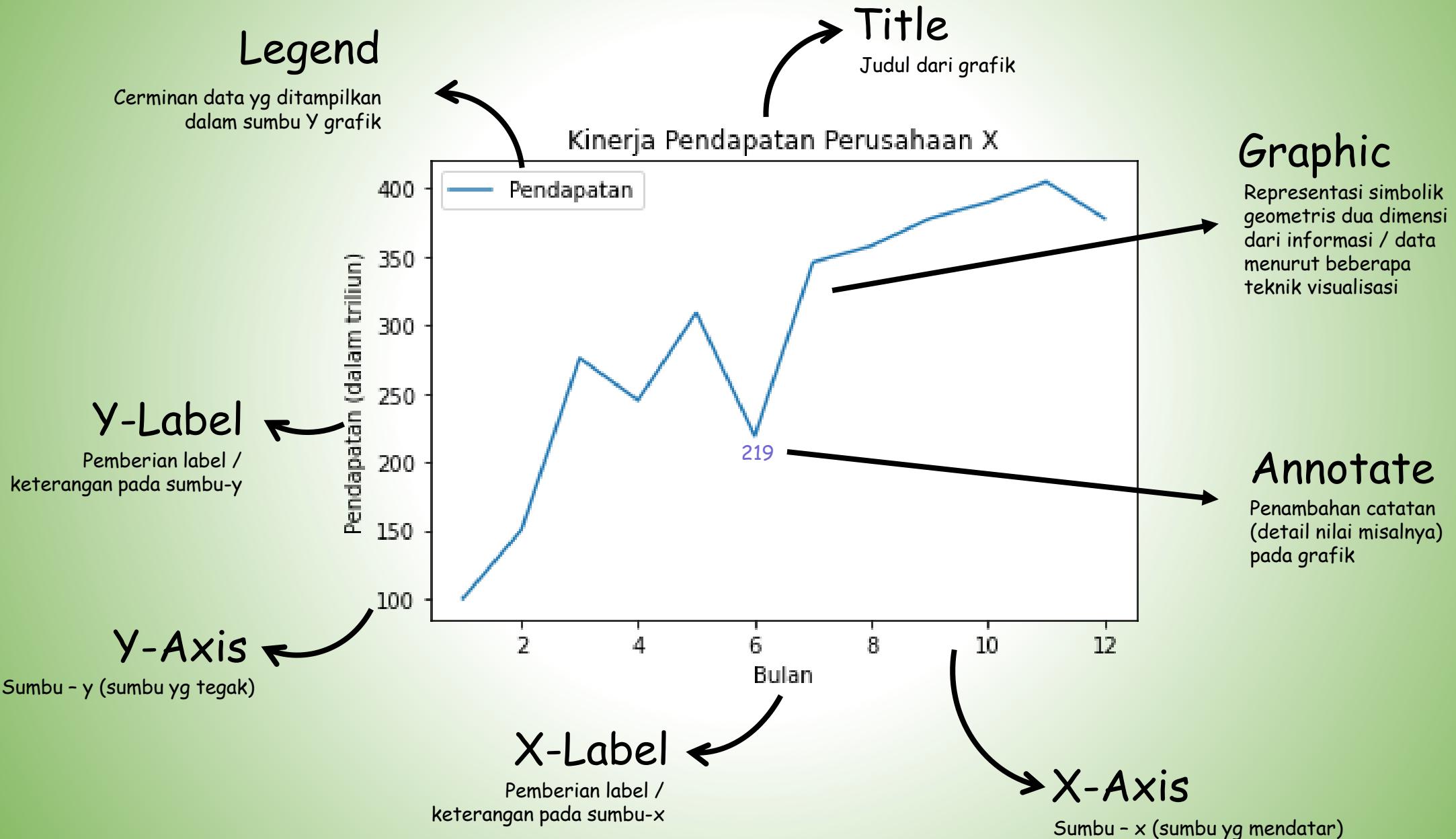
Sertakan sumber table jika diperoleh dari luar

# Diagram

Sebuah representasi data yang digambarkan dalam bentuk grafik.



# ANATOMI DIAGRAM



# DIAGRAM GARIS

Menggambarkan bagaimana variable berubah terhadap variable lain (semisal : waktu)

```
import matplotlib.pyplot as plt

x = [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12]
y1 = [100, 150, 276, 245, 309, 219, 346, 358, 378, 390, 405, 378]
plt.plot(x, y1, label="Pendapatan")
plt.plot()

plt.xlabel("Bulan")
plt.ylabel("Pendapatan (dalam triliun)")
plt.title("Kinerja Pendapatan Perusahaan X")
plt.legend()
plt.show()
```



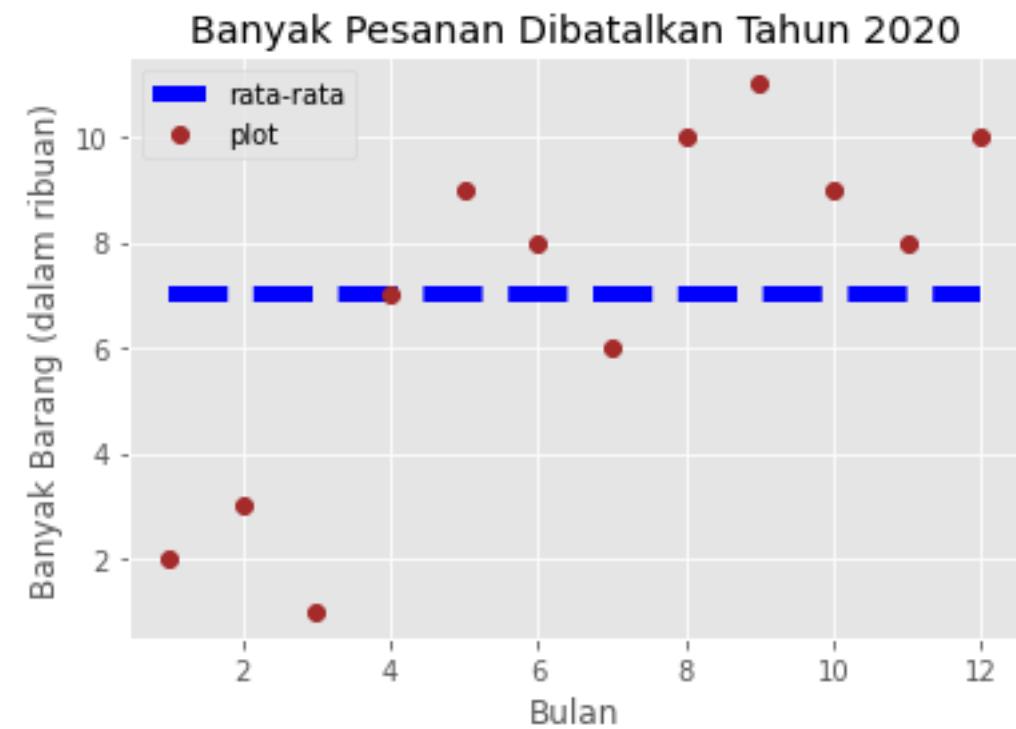
# DIAGRAM TITIK

Digunakan untuk menggambarkan hubungan antara dua variabel

```
import matplotlib.pyplot as plt
import numpy as np

x1 = [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12]
y1 = [2, 3, 1, 7, 9, 8, 6, 10, 11, 9, 8, 10]

fig, ax = plt.subplots()
y_avg = [np.mean(y1)] * len(y1)
ax.plot(x1, y_avg, color='blue', lw=6, ls='--', label="rata-rata")
ax.plot(x1, y1, 'o', color = 'brown', label='plot')
plt.xlabel("Bulan")
plt.ylabel("Banyak Barang (dalam ribuan)")
plt.title("Banyak Pesanan Dibatalkan Tahun 2020")
plt.legend(loc=0)
plt.show()
```



# DIAGRAM AREA

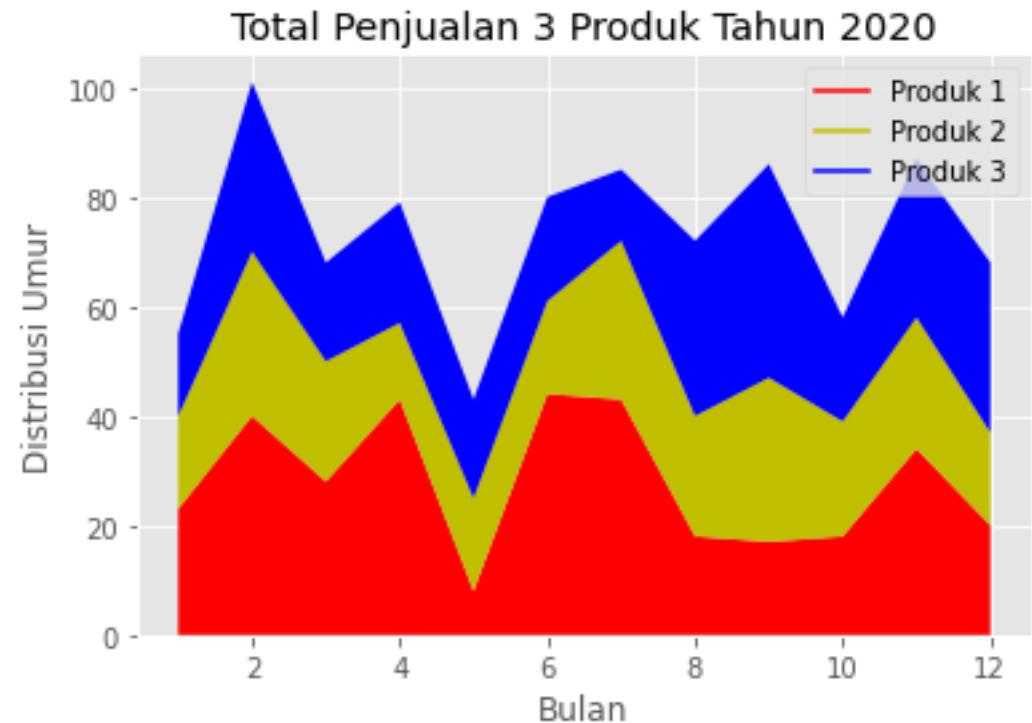
Digunakan untuk melihat trend data dari waktu ke waktu. Daerah berbayang di bawah garis yang diplot pada grafik untuk menunjukkan volume data yang diwakili oleh garis.

```
import matplotlib.pyplot as plt

idxes = [ 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12]
produk1 = [23, 40, 28, 43, 8, 44, 43, 18, 17, 18, 34, 20]
produk2 = [17, 30, 22, 14, 17, 17, 29, 22, 30, 21, 24, 17]
produk3 = [15, 31, 18, 22, 18, 19, 13, 32, 39, 19, 29, 31]

# Adding legend for stack plots is tricky.
plt.plot([], [], color='r', label = 'Produk 1')
plt.plot([], [], color='y', label = 'Produk 2')
plt.plot([], [], color='b', label = 'Produk 3')

plt.stackplot(idxes, produk1, produk2, produk3, colors= ['r', 'y', 'b'])
plt.title('Total Penjualan 3 Produk Tahun 2020')
plt.xlabel("Bulan")
plt.ylabel("Distribusi Umur")
plt.legend()
plt.show()
```

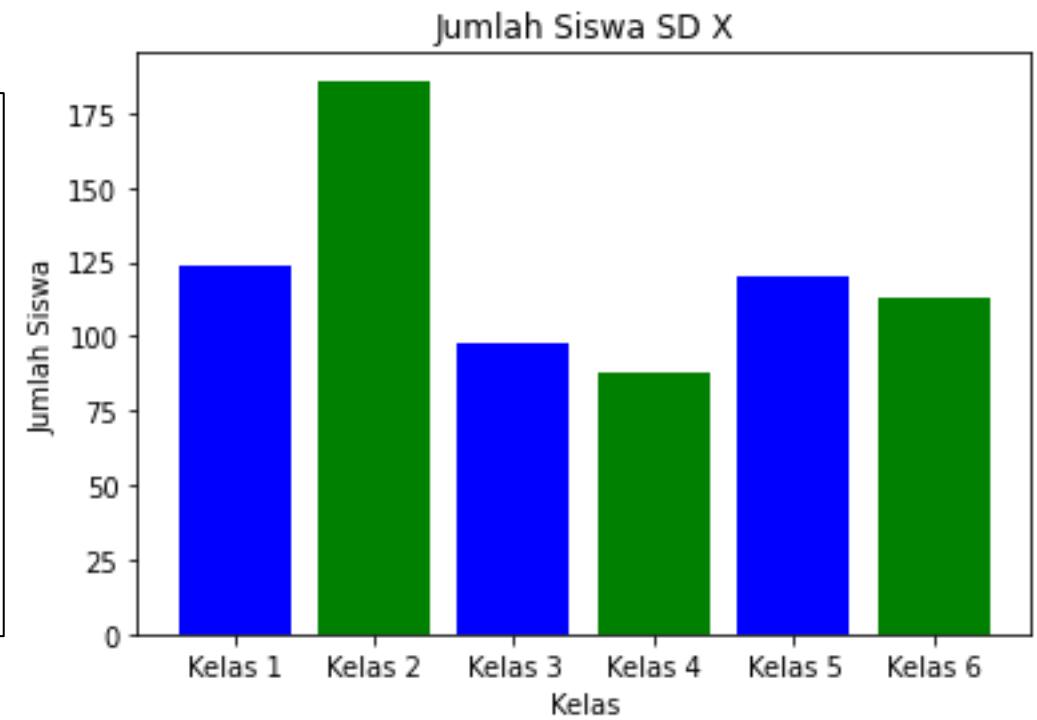


# D I A G R A M   B A T A N G

Paling cocok untuk komparasi data dengan satu atau lebih kategori

```
import matplotlib.pyplot as plt
import numpy as np

jumlah = [124, 186, 98, 88, 120, 113]
kelas = ('Kelas 1','Kelas 2','Kelas 3','Kelas 4','Kelas 5','Kelas 6')
x_pos = np.arange(len(kelas))
plt.bar(x_pos, jumlah)
plt.bar(kelas, jumlah, color = ['blue','green'])
plt.xlabel("Kelas")
plt.ylabel("Jumlah Siswa")
plt.title("Jumlah Siswa SD X")
plt.xticks(x_pos, kelas)
plt.show()
```



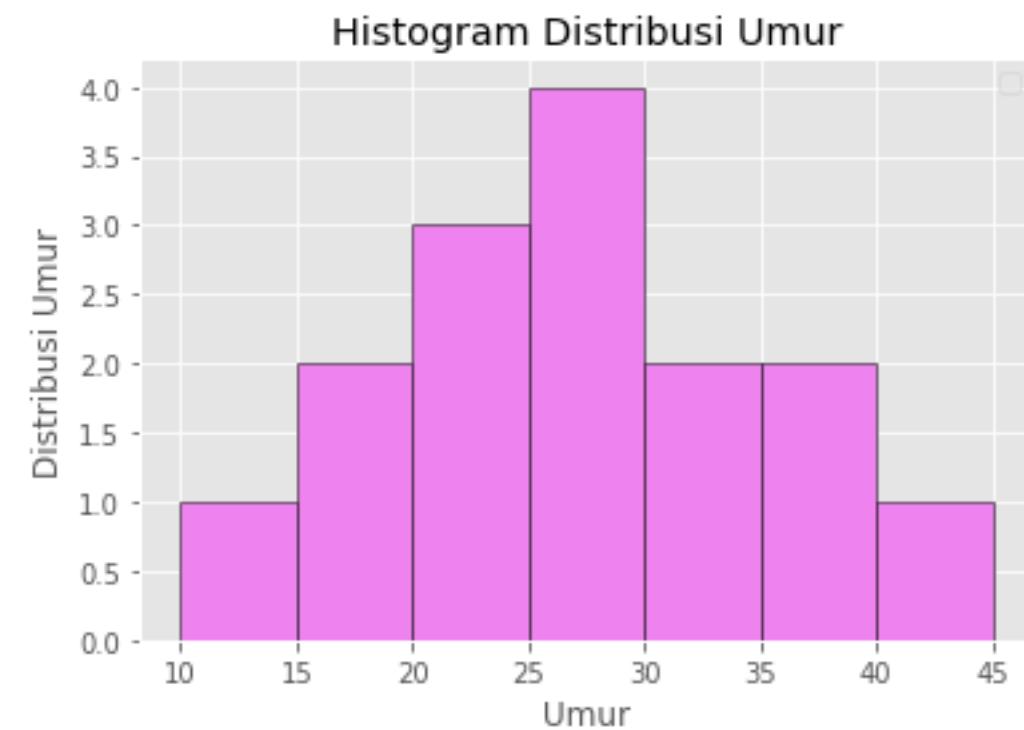
# HISTOGRAM

Menampilkan distribusi frekuensi pada variable kuantitatif

```
import matplotlib.pyplot as plt

ages = [12, 15, 19, 26, 28, 29, 21, 20, 24, 26, 32, 39, 49, 31, 42, 36]
bins = [10, 15, 20, 25, 30, 35, 40, 45]

plt.hist(ages, bins=bins, color = 'violet', edgecolor='black')
plt.xlabel("Umur")
plt.ylabel("Distribusi Umur")
plt.title("Histogram Distribusi Umur")
plt.legend()
plt.show()
```



# DIAGRAM LINGKARAN

Menggambarkan komposisi antarbagian pada suatu kesatuan utuh.

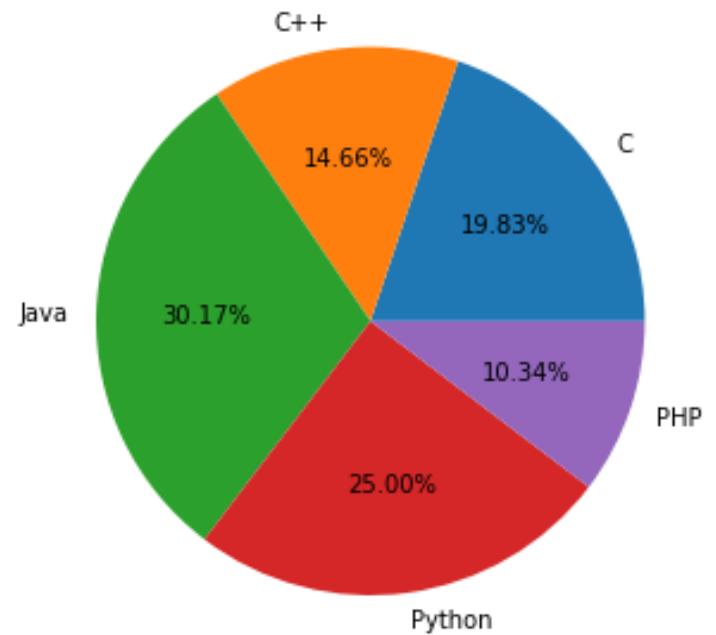
```
from matplotlib import pyplot as plt
import numpy as np

fig = plt.figure()
ax = fig.add_axes([0,0,1,1])
ax.axis('equal')

langs = ['C', 'C++', 'Java', 'Python', 'PHP']
students = [23,17,35,29,12]

ax.pie(students, labels = langs, autopct='%1.2f%%')
plt.title("Minat Bahasa Pemrograman Mahasiswa TI")
plt.show()
```

Minat Bahasa Pemrograman Mahasiswa TI



# DIAGRAM RADAR

Menampilkan data multivariabel yang digambarkan dalam koordinat polar pada sumbu berbeda-beda dengan satuan/unit yang sama

```
import plotly.express as px
import pandas as pd
df = pd.DataFrame(dict(
    r=[46, 52, 23, 28, 32],
    minat=['Fotografi', 'Koding', 'Musik', 'Olahraga', 'Sinematografi']))
fig = px.line_polar(df, r='r', minat='theta', line_close=True)
fig.update_layout(title="Jumlah Peminat Rencana Pelatihan Minat-
Bakat Siswa SMA X", title_x=0.5)
fig.show()
```

Jumlah Peminat Rencana Pelatihan Minat-Bakat Siswa SMA X

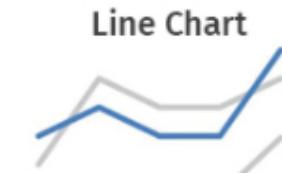


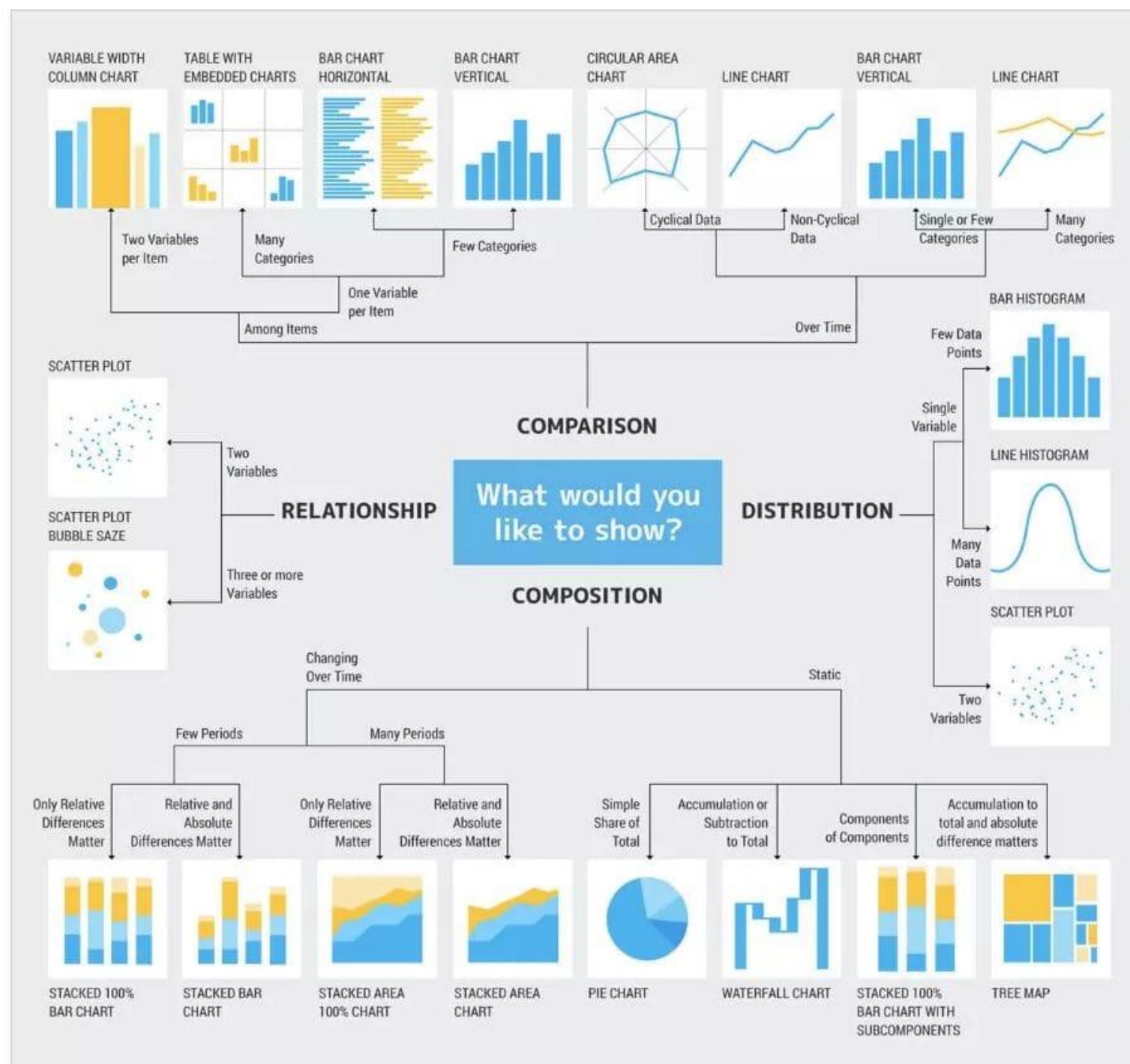
## Pilah-pilih grafik

Membaca grafik bukanlah satu pekerjaan mudah bagi sebagian orang, terutama yang tidak terbiasa melihat data. Melihat dua garis saling bersinggungan, misalnya, bisa menimbulkan arti berbeda bagi orang yang tidak memahami data.

Oleh karena itu, penting sekali memperkenalkan visualisasi data dengan sesederhana mungkin agar alur membaca dan mengambil kesimpulan atas elemen visual yang ditampilkan dengan lebih mudah.

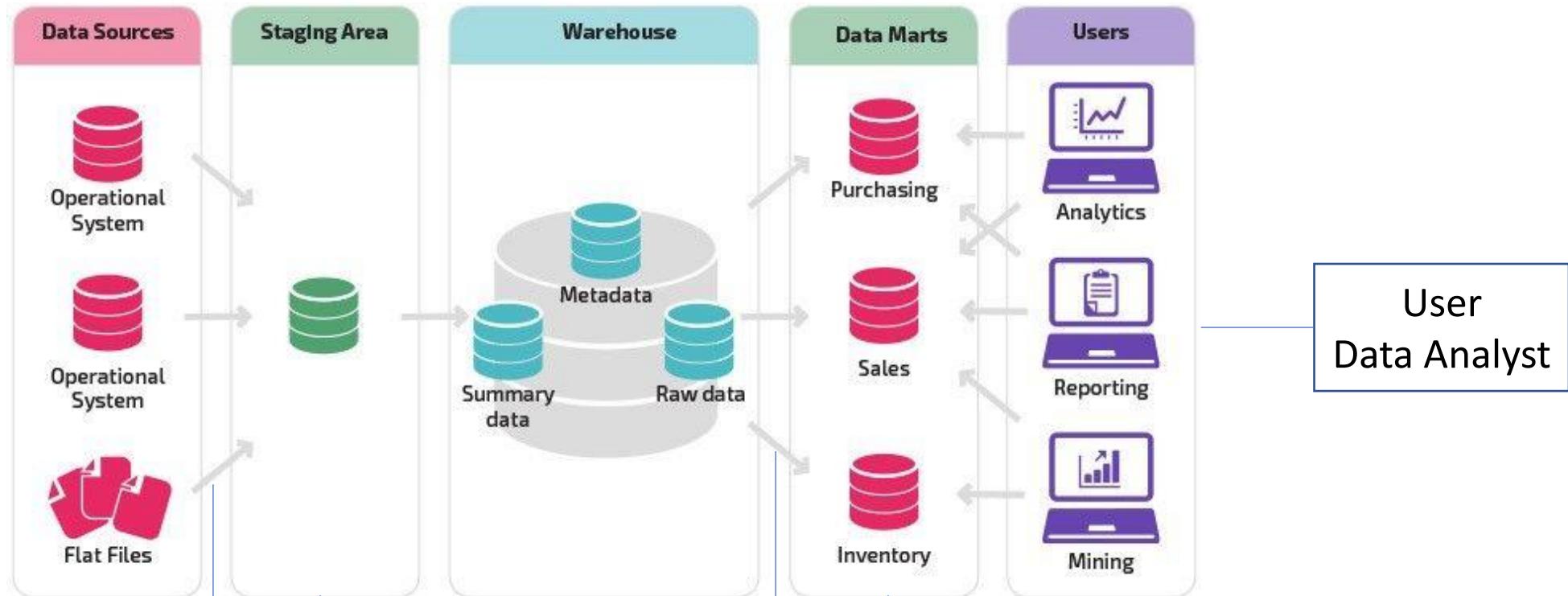
Ada berbagai macam cara menampilkan data, entah dalam bentuk grafik, angka, atau teks. Cara menampilkannya tergantung pada pesan yang ingin disampaikan kepada audiens. Berikut penjelasan singkat tentang sejumlah pilihan visualisasi data.

Jenis Grafik	Bentuk	Keterangan
Teks dan angka		Data tidak harus ditampilkan dalam bentuk grafik. Dapat menggunakan teks dan angka saja, dengan catatan hanya 1-2 data yang ingin ditampilkan. Berikan penekanan atau warna berbeda pada angka atau teks yang ingin ditonjolkan agar perhatian pembaca terarah pada bagian tersebut.
Diagram garis		Line chart paling tepat digunakan untuk menunjukkan tren dari waktu ke waktu. Sumbu X biasanya mewakili periode waktu, sumbu Y menggambarkan nilai/kuantitas. Contohnya jumlah penjualan dari pekan ke pekan selama satu tahun. Grafik ini dapat memuat banyak titik data yang dapat diatur saling berdekatan sesuai kerapatan periode waktu. Karena visualisasi yang sederhana, bisa menggunakan banyak garis.



# DATA VIZ PADA BISNIS





User, Dev, Data Analyst & Scientist



Data Engineer

**Apache** **hadoop** **Apache** **cloudera**  
**Scala** **Spark** **Apache** **HUE**  
**python**

Data Engineer (Modeller)

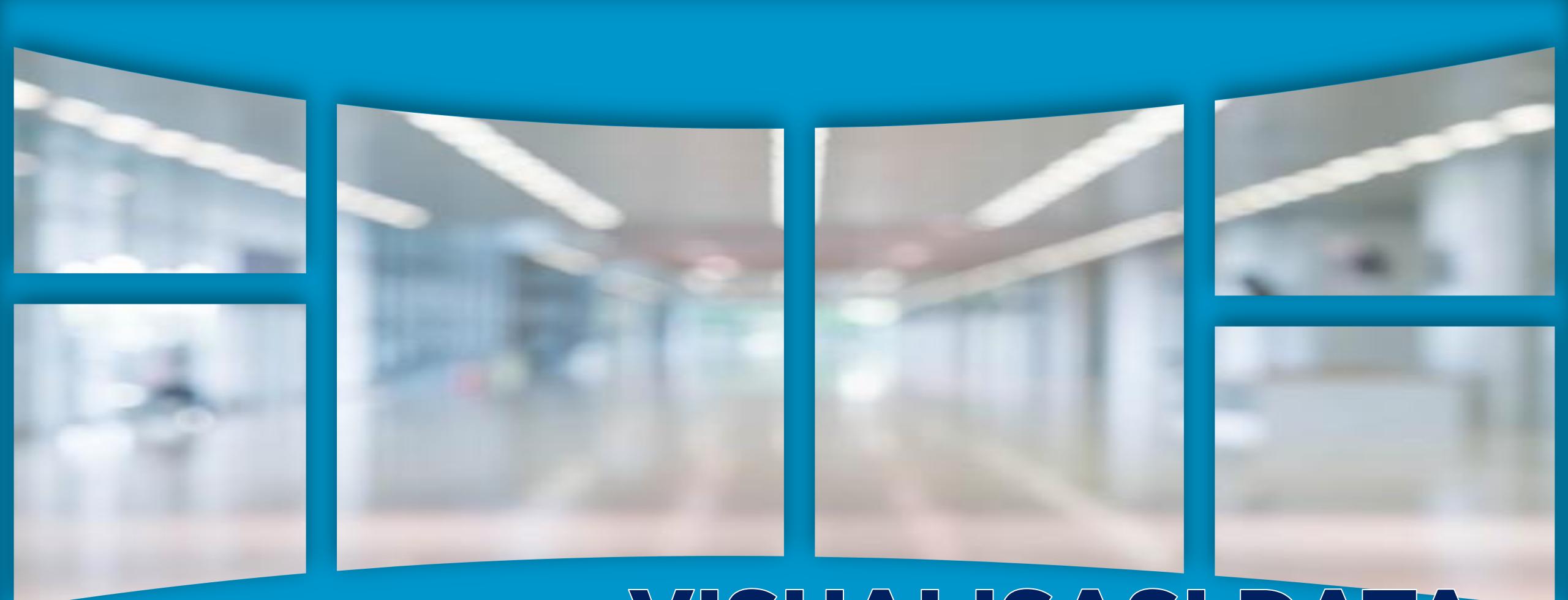
Data Engineer Data Scientist

Data Scientist Bus. Intelligence



**MicroStrategy**





# **VISUALISASI DATA**

# SCORECARD

- Sebuah visualisasi yang focus pada sebuah jenis data dan biasanya terdiri dari 1 bentuk visualisasi saja
- Contoh : jumlah pendapatan, kepuasan pelanggan, dll
- Menggambarkan salah satu Key Performance Indicators lembaga / perusahaan
- Memantau kemajuan progres



type: Web 

Date: Last 6 months 

+ NEW

Total clicks

223

Total impressions

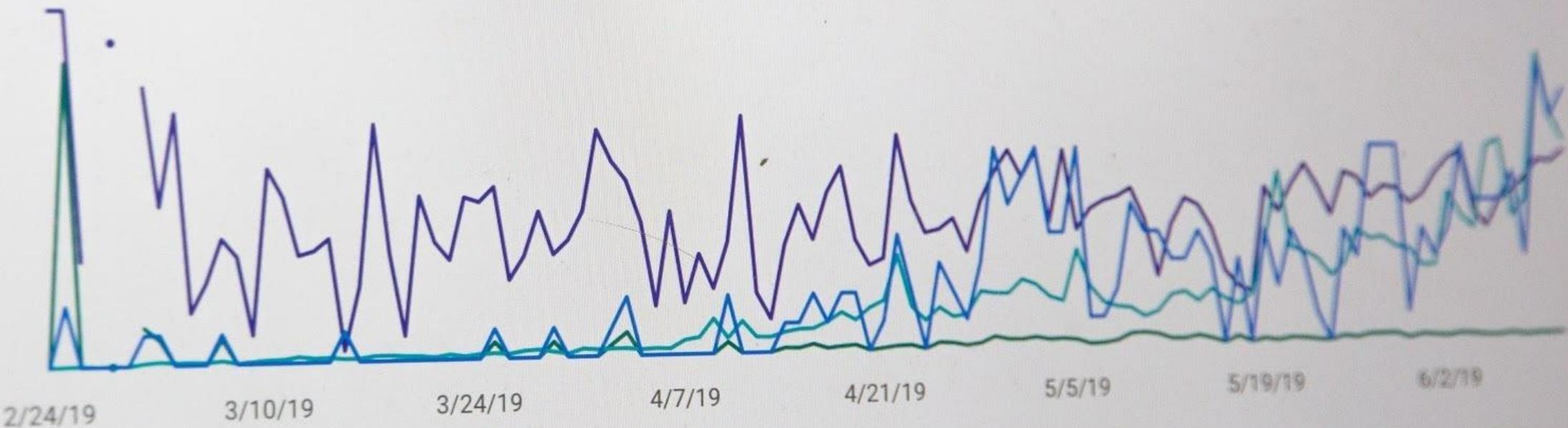
17.6K

Average CTR

1.3%

Average position

25.2



# DASHBOARD



- Kumpulan dari berbagai macam visualisasi yg menggabungkan dan merangkum informasi atau data
- Contoh : laporan penjualan bulanan, kinerja karyawan, dll
- Visualisasi dan analisa yang kompleks
- Mendapatkan detail laporan yang lebih kompleks dan beragam

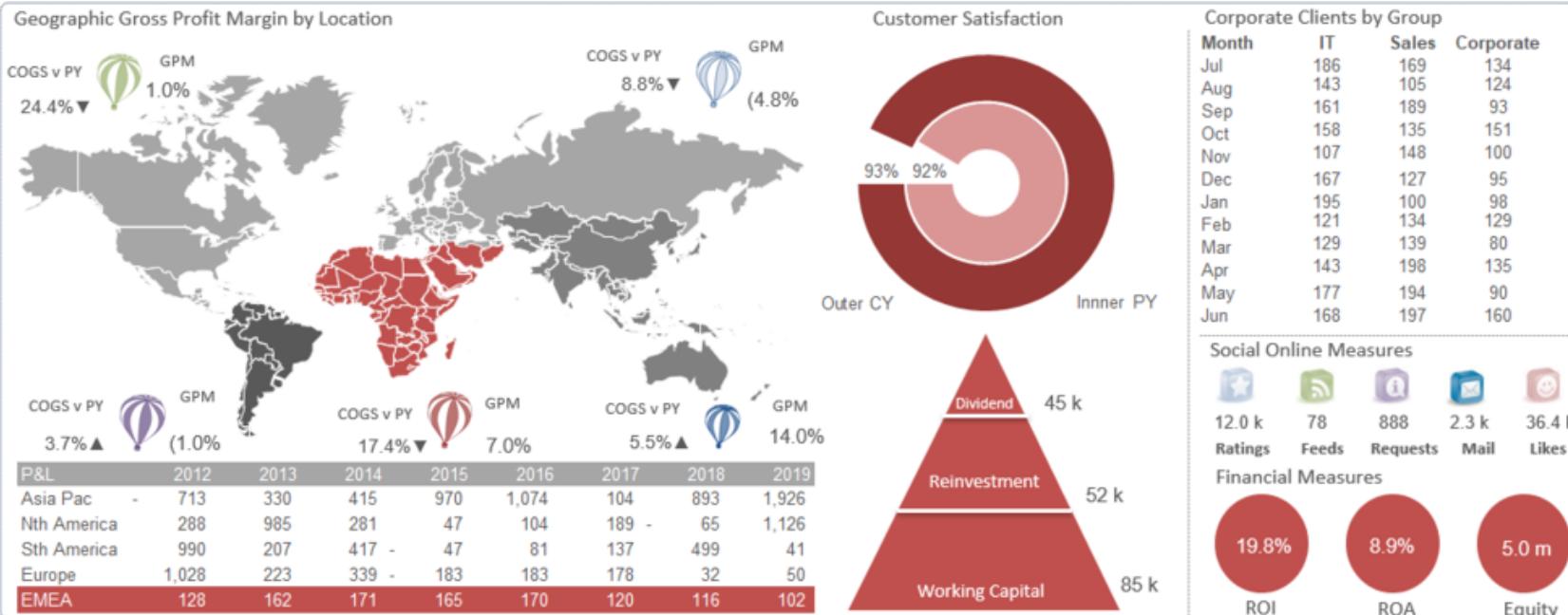
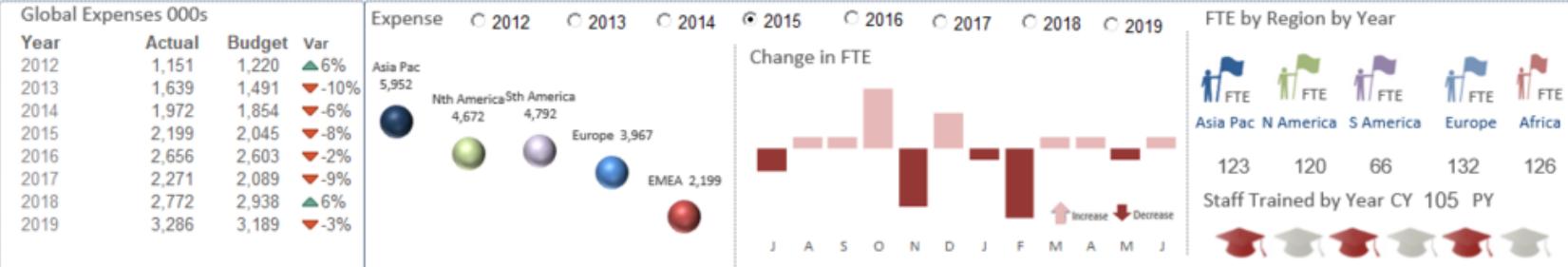
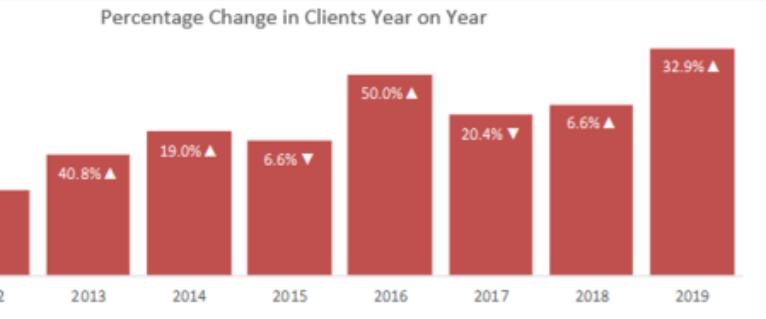


# Global Sales Dashboard

Summary of EMEA

TheSmallman.com

Year	Global Sales 000s EMEA		Current Yr v PY by Region (m)					
	Actual	Budget	Var	Asia Pac	Nth America	Sth America	Europe	EMEA
2012	1,279	1,215	+5.2%	6,922				
2013	1,801	1,696	+6.2%	4,719				
2014	2,143	2,042	+5.0%		16			
2015	2,364	2,116	+11.7%	4,744				
2016	2,825	2,605	+8.5%		19			
2017	2,391	2,709	(-11.7%)	3,784				
2018	2,888	3,092	(-6.6%)					
2019	3,388	3,383	0.1%	2,364				



# REPORT

- Penyampaian berita, keterangan, pemberitahuan ataupun pertanggung-jawaban sebagai bentuk pelaksanaan komunikasi dari satu pihak ke pihak yang lain
- Visualisasi data untuk membuat summary dari apa yang terjadi di perusahaan dalam waktu tertentu sehingga dapat digunakan untuk memahami hal yang sedang terjadi dari suatu perusahaan dengan secepat mungkin





—  
—  
—

— 10 —

1995-1996 Annual Report

1996-1997  
Yearbook

1995-1996  
Yearbook

1990-1991 學年上學期

**DETtaglio consumi storici**  
Mese: 12.988 kWh

Consumo annuo: 12.988 kWh  
Grafico consumi in kWh



Sintesi periodo dal 01/05/2017 al 31/05/2017  
PER I CONSUMI

Periodo di lettura dal 01/05/2017 al 31/05/2017  
Letture e consumi

ENERGIA REATIVA		POTENZA	
Consumo Effettivo (kWh)	Impegnata (kW)	Consumo Effettivo (kWh)	Potenza Impegnata (kW)
0	0	0	10,2
0	0	0	10,2
0	0	0	10,2

# ANALYTIC REPORT



- Laporan yang berfokus pada analisis untuk menentukan keputusan
- Memberikan pemahaman yang mudah dipahami dengan membaca sekilas saja

ANZEIGE

## BANTILEON

Investmentfonds  
Investmentfonds.com

ZEITUNGSSCHLUSS DIESER AUSGABE: 21.45 UHR

EURO in DOLLAR  
→ 1.3276



## FINANZIELLER

WERT PAPER

ONLINE-FINANZTOOL: [www.finanztool.de](http://www.finanztool.de)  
Unternehmen und NEC punkten mit NG Partner  
chiba und NG Partner

\*EuroSTOXX 50

Dax  
Stoxx\*  
TecDax  
Mittel  
DAX  
Nasdaq

4146,98  
289115  
513,81  
10977,89  
10511,74  
2108,84

28988,95  
514,10  
10832,75  
10522,23  
210197

0,27  
0,06  
1,03  
0,10  
0,10

101,15  
0,01  
0,01  
0,01  
0,01

101,80  
0,01  
0,01  
0,01  
0,01

101,57  
0,01  
0,01  
0,01  
0,01

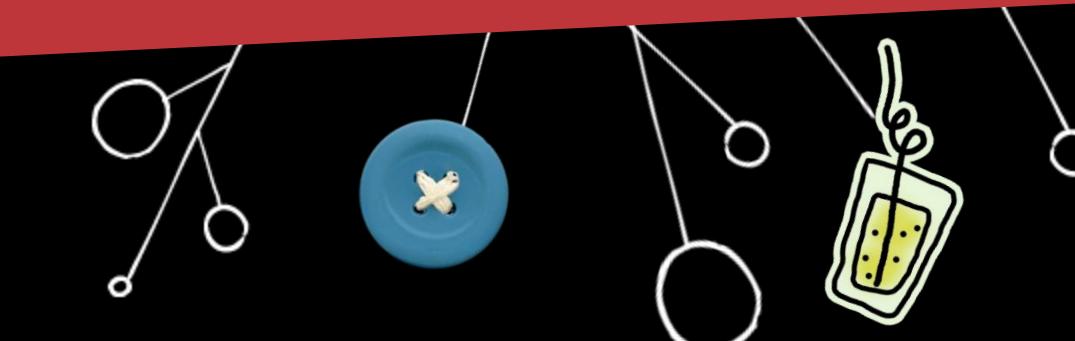
101,80  
0,01  
0,01  
0,01  
0,01

101,80  
0,01  
0,01  
0,01  
0,01

101,80  
0,01  
0,01  
0,01  
0,01

101,80  
0,01  
0,01  
0,01  
0,01

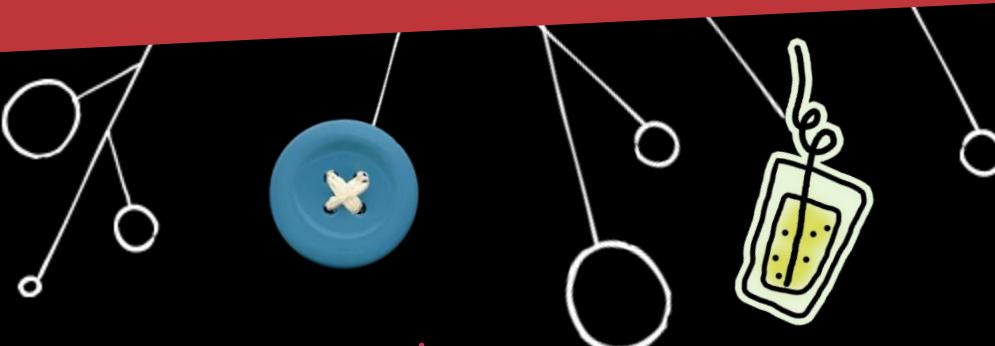
# KONTEKS DATA



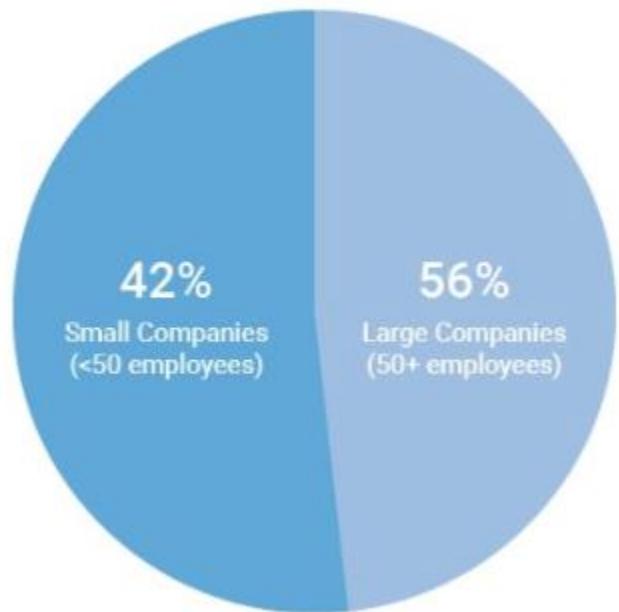
```
<!DOCTYPE html>
<html lang="id">
<head>
    <title> URGENSI VISUALISASI DATA </title>
</head>
<body>
    <h1> 1. KEPADA SIAPA DATAINI AKAN DISAMPAIKAN? </h1>
    <h2> 2. DATA APA YANG INGIN SAYA SAMPAIKAN? </h2>
    <ol>
        <li> a. Data yang ingin disampaikan harus relevan dengan audien </li>
        <li> b. Sasarannya agar audien dapat memproses fakta dan data </li>
        <li> c. Insightful dan Kontekstual </li>
    </ol>
    <h3> 3. BAGAIMANA CARA MENYAMPAIKAN DATA TERSEBUT? </h3>
    <h4> 4. IMPACT VISUALISASI DATA, BUKAN HANYA TAMPIL CANTIK NAMUN PESAN YANG INGIN  
DISAMPAIKAN HARUS TERCAPAI </h4>
</body>
</html>
```



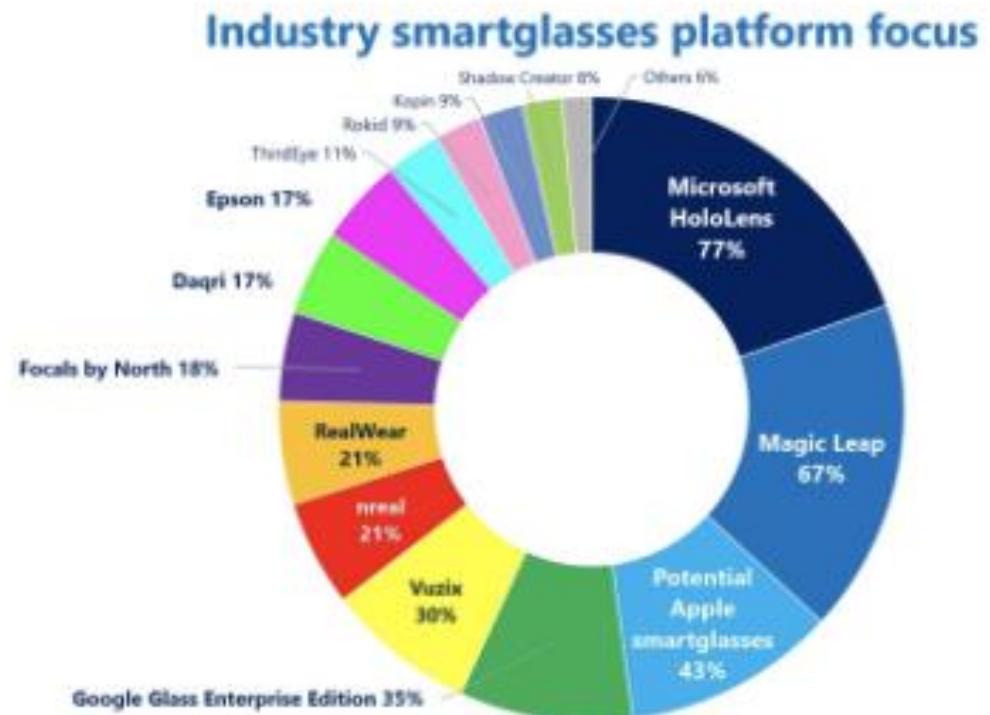
# KESALAHAN UMUM VISUALISASI DATA



## Nilai Persentase Tidak Sesuai



## Terlalu Banyak Irisan



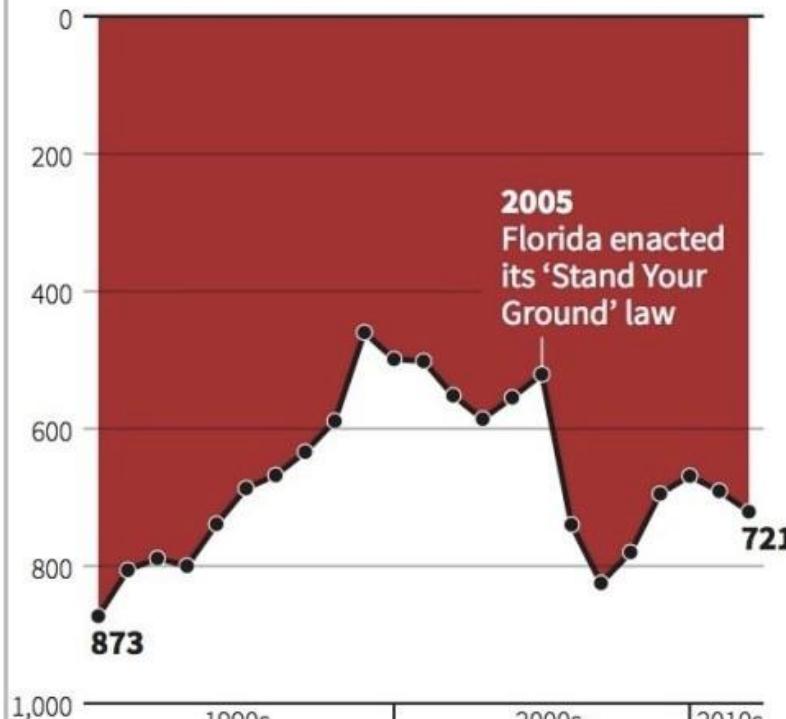
Digi-Capital

Note: Percentages not cumulative. "Others" include OnyaGlass, EverySight, Glasses, iM888, Viatick, Lensbox, Leftrar, Lumus, Optinvent, Realmax, Snap Spectacles, Tilt Five. © 2019 Digi-Capital. All rights reserved. No publication, extraction, modification, reproduction or compilation without written permission from Digi-Capital.

## Tidak Mengikuti Standar Penulisan

### Gun deaths in Florida

Number of murders committed using firearms

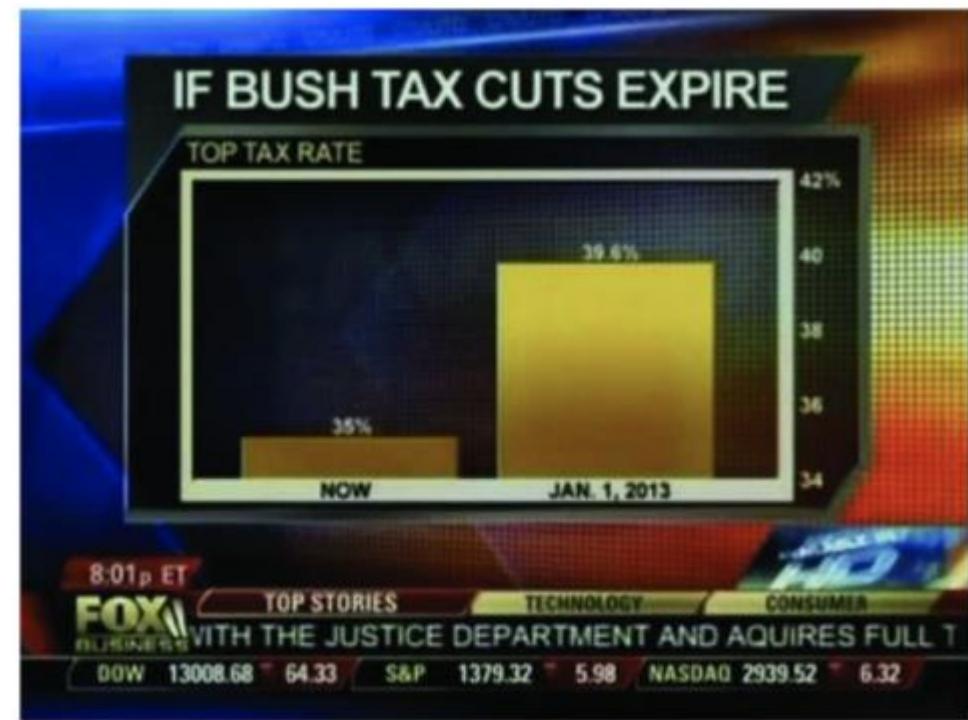


Source: Florida Department of Law Enforcement

© Chan 16/02/2014

REUTERS

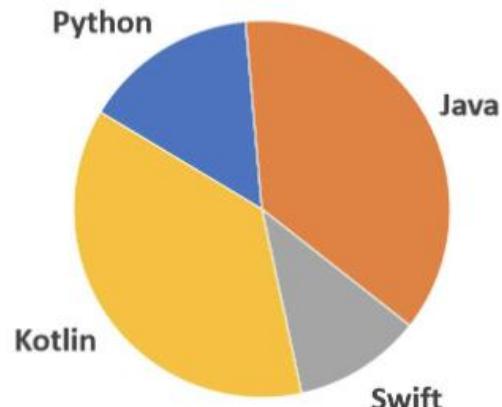
## Sumbu yang Terpotong



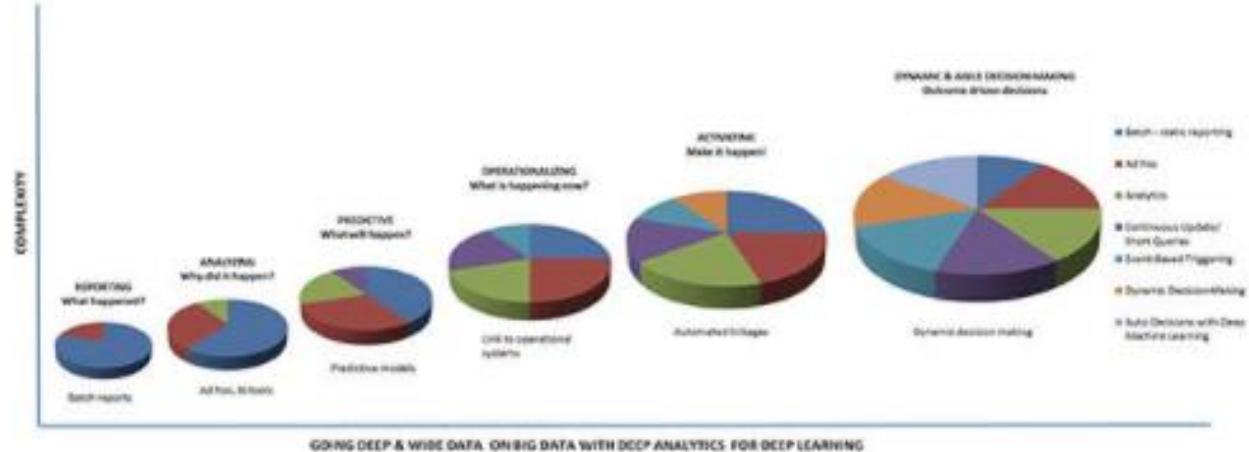
# Penggunaan 3D yang Kurang Sesuai



Penggunaan Bahasa Pemrograman dalam Bekerja

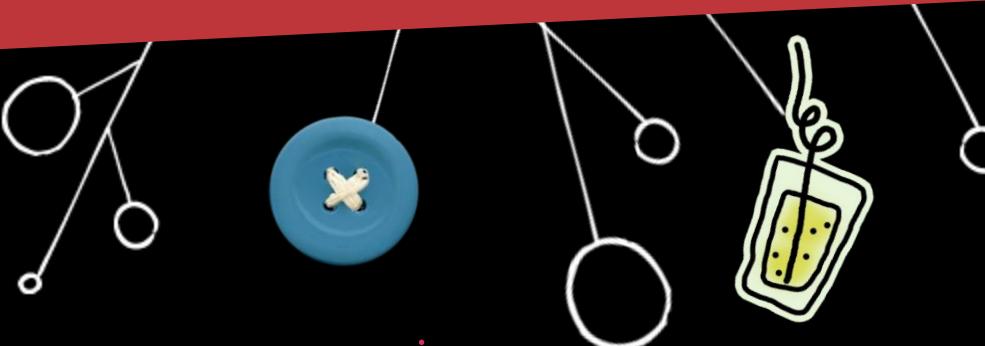


## Susah Dibandingkan





# TOOLS VISUALISASI DATA





<https://powerbi.microsoft.com/en-us/>



<https://www.microstrategy.com/en/education>



<https://www.tableau.com/products/public>





# VISUALISASI DATA YANG EFEKTIF

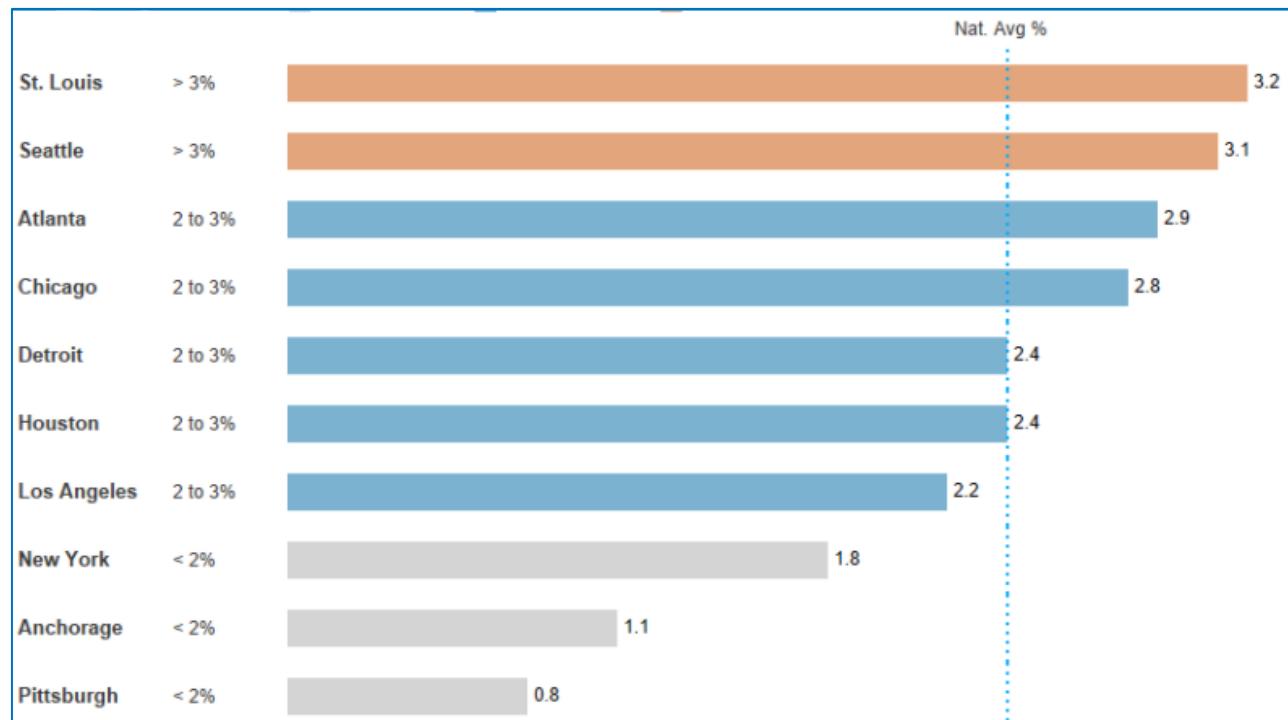
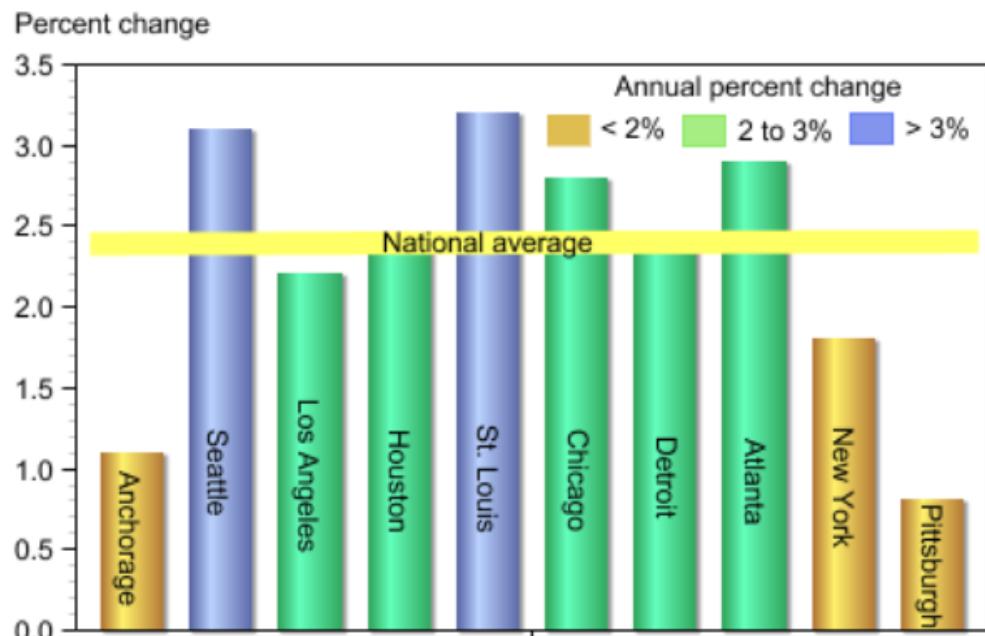


# KETENTUAN

- Gunakan table, teks atau diagram sesuai kebutuhan dan tonjolkan dalam presepsi visual dan audiens
- Meminimalkan cognitive load (ex : clutter merupakan elemen visual yang tidak menambah pemahaman)
- Maksimalkan Preattentive attributes (komponen atau atribut visual yang langsung tertangkap perhatian kita secara tanpa sadar)
- Perhatikan Affordances, Accessibility dan Aesthetics

# Cognitive Load

In 2014, food price inflation was higher in Seattle, St. Louis, Chicago, and Atlanta than in other metropolitan areas



## Preattentive Attributes

75639506847

36586630375

76860372658

60284658910

75**6**3950**6**847

3658**66**30375

76860372**6**58

60284**6**58910

## 10 masalah desain teratas

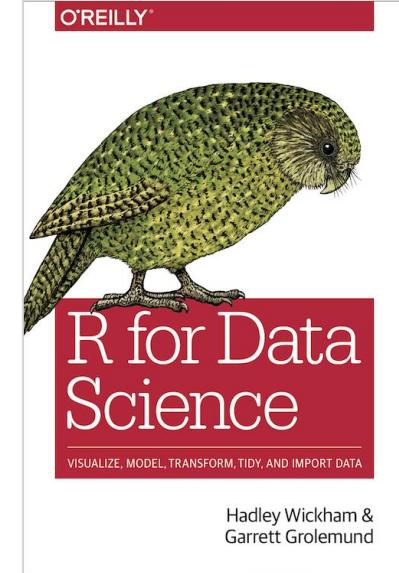
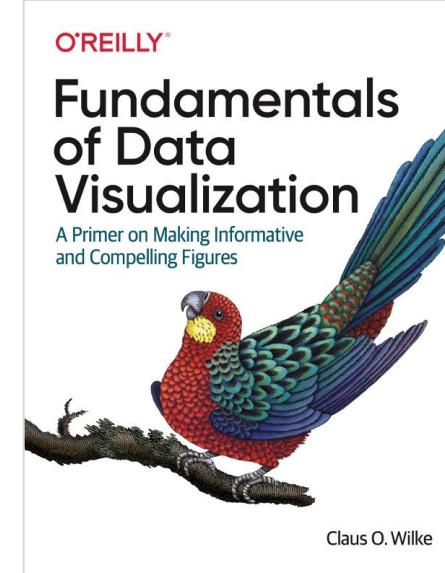
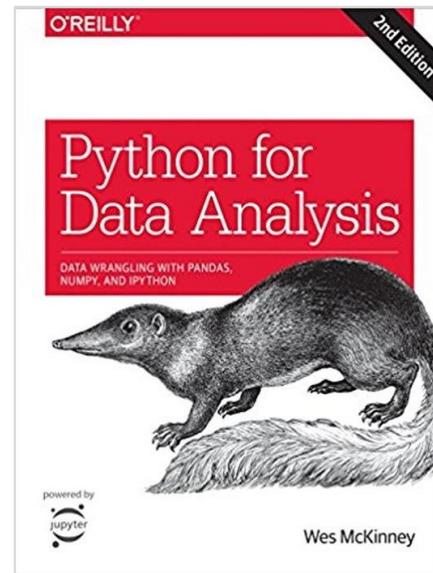
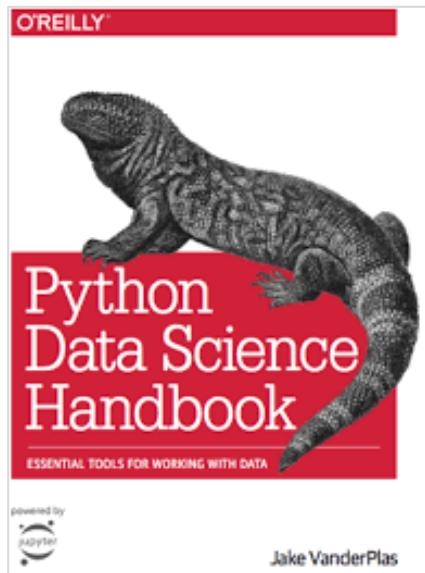


**7 dari 10 masalah desain teratas terjadi sebanyak 10x untuk setiap 1.000 mobil.** Diskusi: Apakah ini merupakan standar kejadian yang dapat diterima atau tidak?

## 10 masalah desain teratas



# REFERENSI BUKU

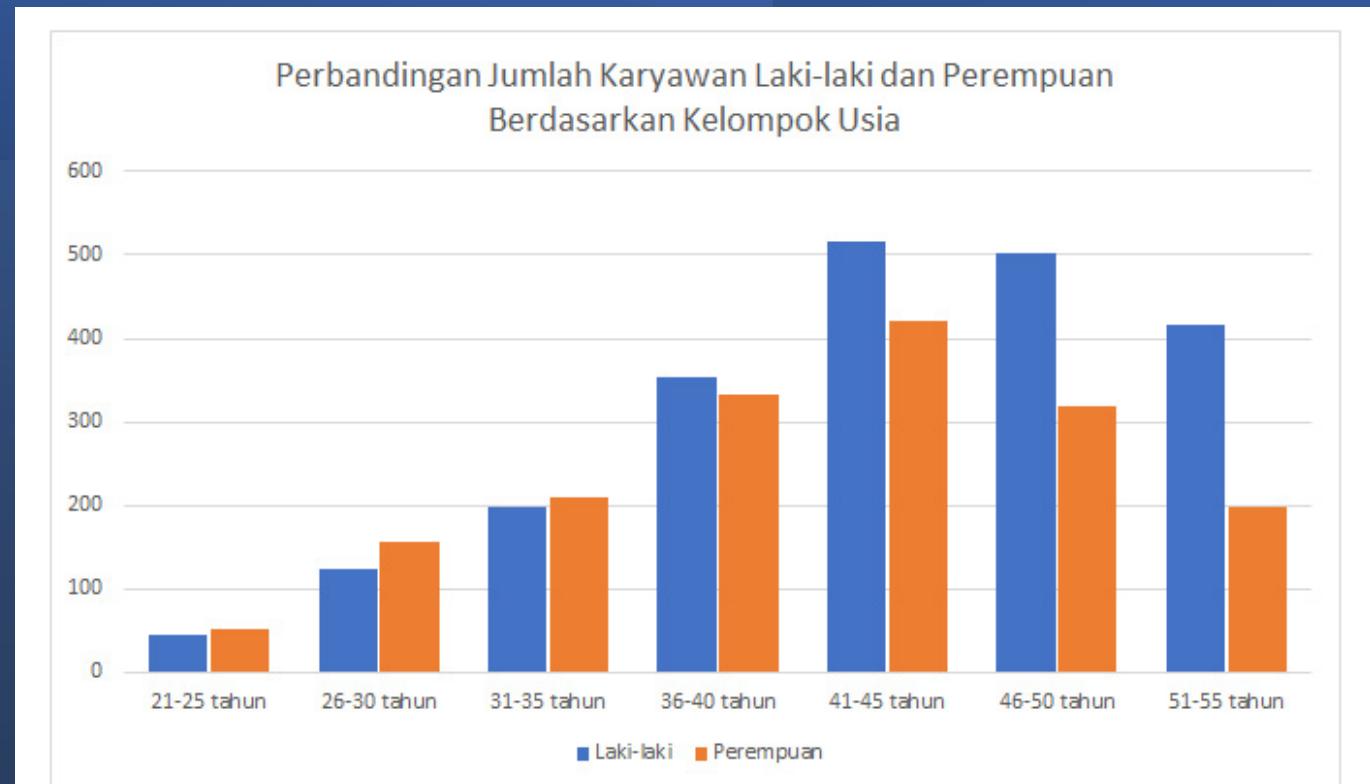


# FUN QUIZ



Berikut adalah data karyawan di perusahaan A.

Kelompok Usia	Laki-laki	Perempuan
21-25 tahun	45	53
26-30 tahun	124	156
31-35 tahun	198	208
36-40 tahun	353	332
41-45 tahun	516	420
46-50 tahun	502	319
51-55 tahun	417	199



Gambar 1: Grafik perbandingan jumlah karyawan laki-laki dan perempuan berdasarkan rentang usia.

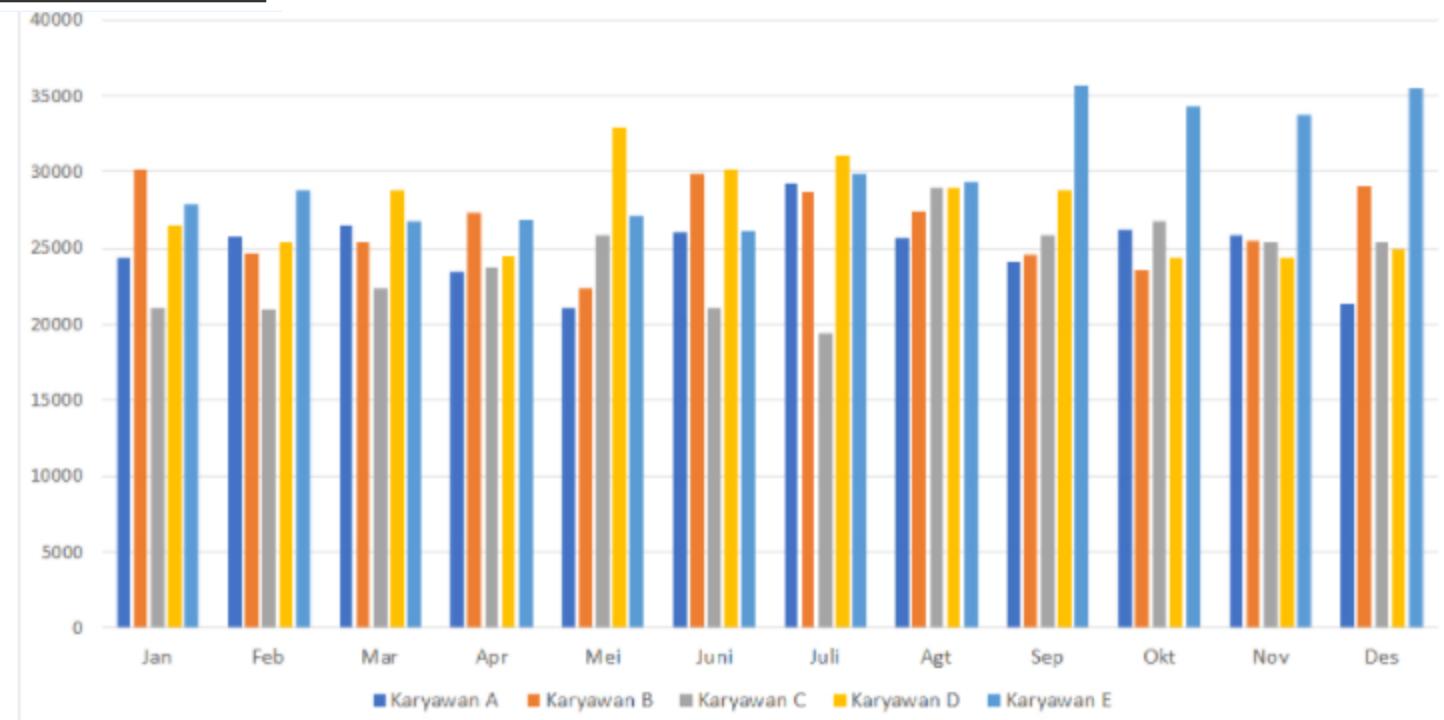
## Perbandingan Jumlah Karyawan Laki-laki dan Perempuan Berdasarkan Kelompok Usia



Gambar: Hasil akhir dari perbaikan grafik untuk Praktek Visualisasi Data 1.

Bulan	Karyawan A	Karyawan B	Karyawan C	Karyawan D	Karyawan E
Januari	24345	30128	21039	26472	27894
Februari	25739	24678	20983	25382	28763
Maret	26489	25402	22321	28757	26748
April	23455	27320	23759	24443	26839
Mei	21093	22390	25873	32901	27110
Juni	25987	29871	21023	30211	26112
Juli	29223	28654	19420	31118	29878
Agustus	25667	27431	28999	28998	29320
September	24091	24536	25831	28752	35733
Oktober	26201	23547	26758	24379	34289
November	25890	25439	25385	24390	33741
Desember	21345	29087	25380	24879	35478

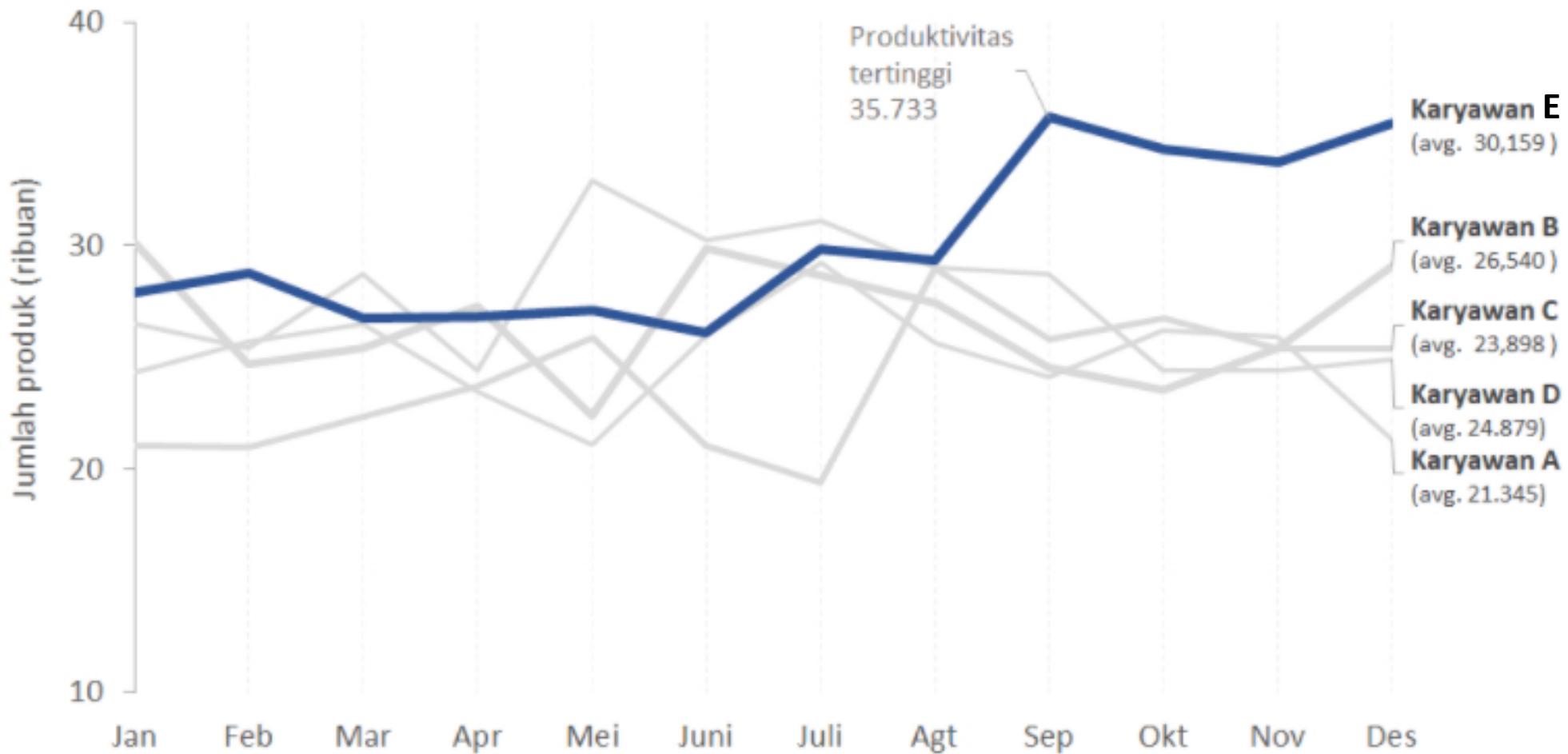
Grafik Jumlah Produksi Per Karyawan



Gambar 1: Grafik Jumlah Produksi Per Karyawan.

# Best Employee 2019

Lima karyawan dengan produktivitas tertinggi



Gambar 2: Hasil akhir dari perbaikan grafik.



Terima Kasih

