Excel for Data Analytic

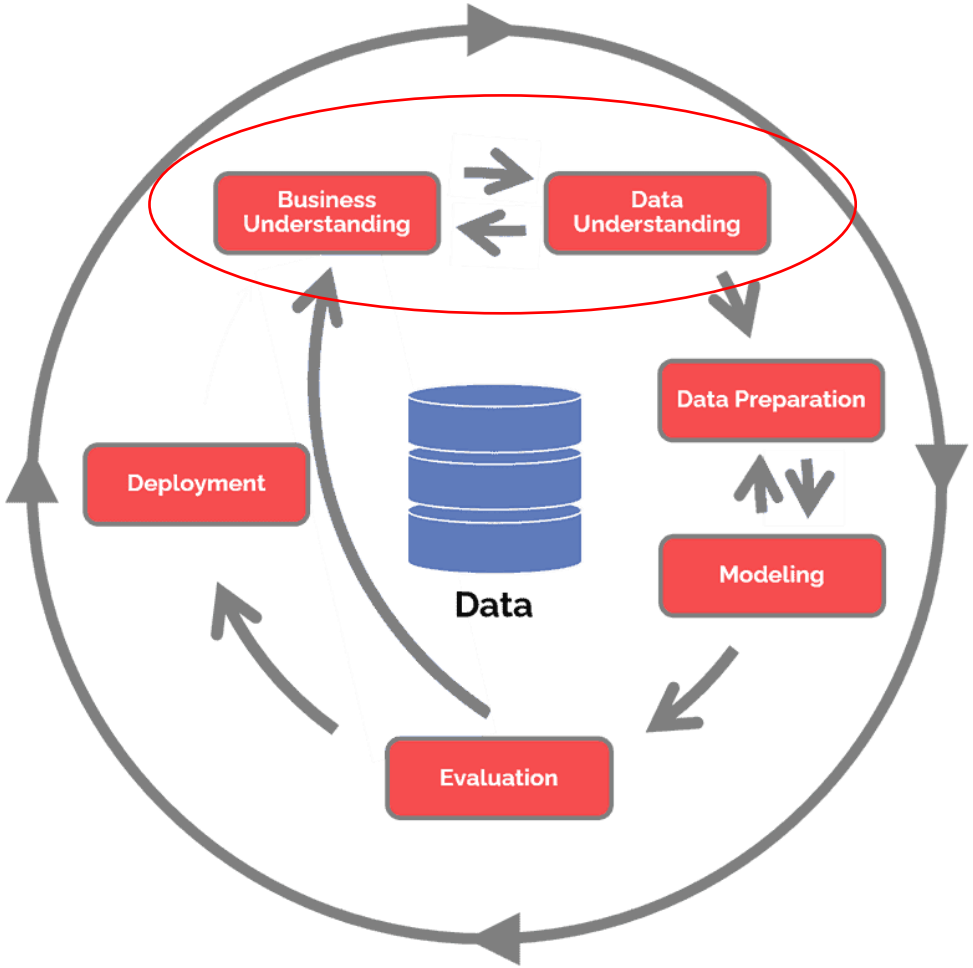
Kemenkeu Corporate University

WARNING
Pastikan anda sudah makan sebelum mengikuti materi ini!!!

Data Mining Lifecycle

- Business Understanding
- Data Understanding
- Data Preparation
- Modeling
- Evaluation
- Deployment

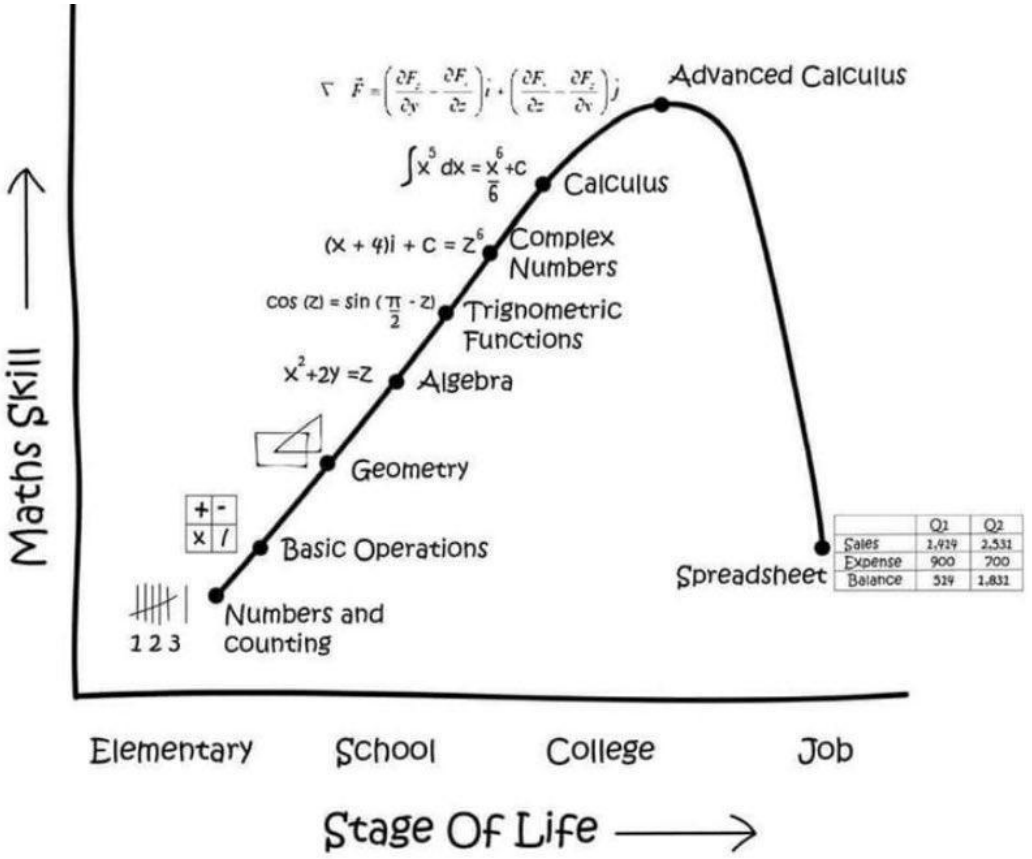
Notice the iteration!



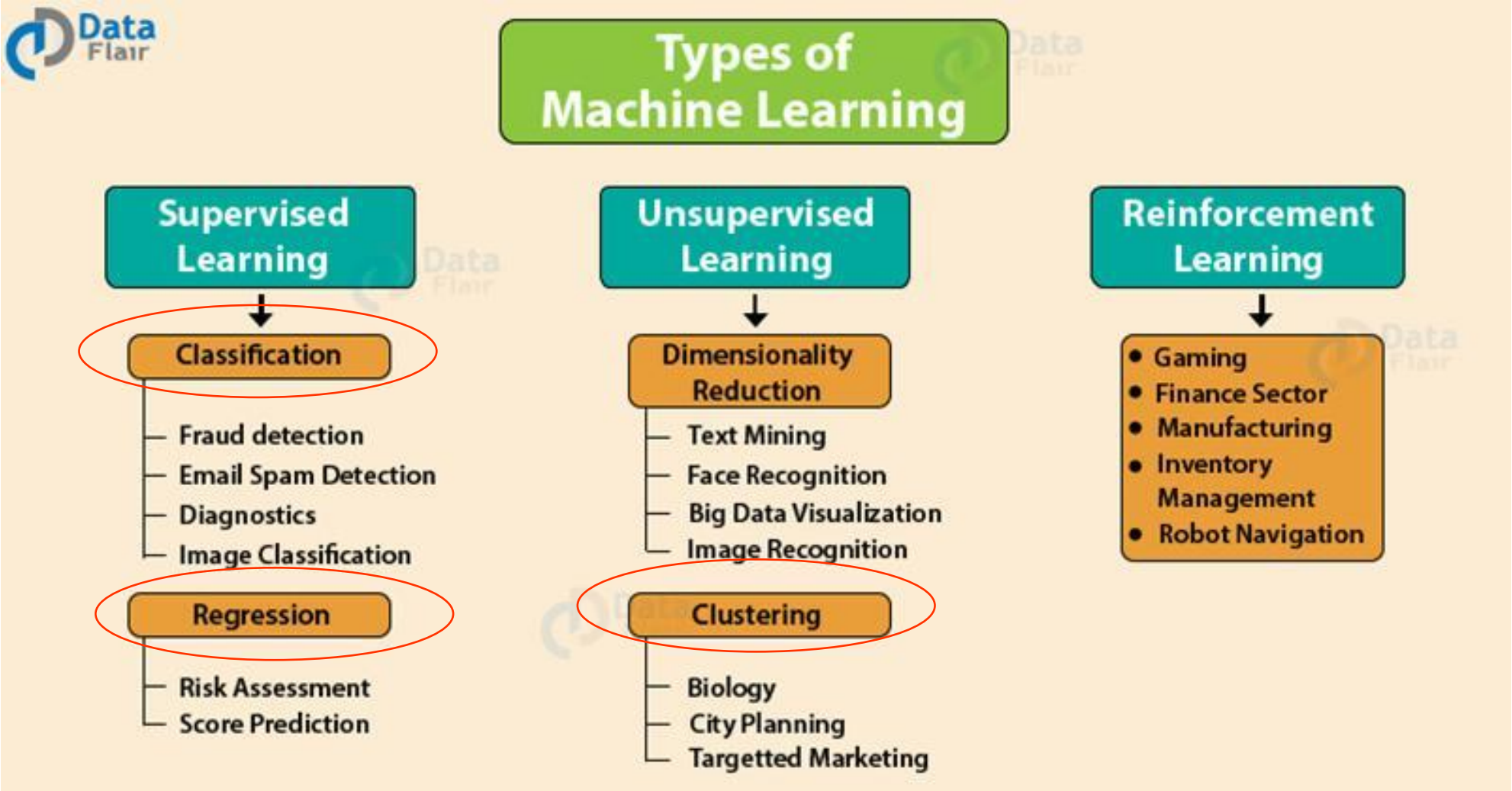
THE FAMOUS CRISP-DM

Why Excel?

- Software yang paling sering digunakan di kantor: Word, PowerPoints, dan Excel… dan Whatsapp
- Excel bisa memecahkan hampir semua permasalahan yang dihadapi di kantor.
- Excel selain fungsi-fungsi dasar, juga bisa digunakan sebagai pengolahan data analytics sederhana.
- Apapun yang dipelajari pada akhirya ke Excel juga….



https://twitter.com/paul__mangione/status/1508998049321955337?s=20&t=ZoB57yJE5G9IvT-gE3Rkxw


Data Flair
Types of Machine Learning
Supervised Learning
Classification
Fraud detection
Email Spam Detection
Diagnostics
Image Classification
Regression
Risk Assessment
Score Prediction
Unsupervised Learning
Dimensionality Reduction
Text Mining
Face Recognition
Big Data Visualization
Image Recognition
Clustering
Biology
City Planning
Targetted Marketing
Reinforcement Learning
Gaming
Finance Sector
Manufacturing
Inventory Management
Robot Navigation

DataExploration and Preparation

Data Understanding

date	nilai_export	usd_to_idr	inflasi	oil_price	neraca_perdagangan
Jan-20	13636,4	13643,18066	2,68%	51,56	-0,64
Feb-20	14042,1	14311,5	2,98%	44,76	2,51
Mar-20	14031,3	16346,96191	2,94%	20,48	0,72
Apr-20	12159,8	15234,63672	2,67%	18,84	-0,37
Mei-20	10452,6	14639,59961	2,19%	35,49	2,02
Jun-20	12006,8	14470,2002	1,96%	39,27	1,25
Jul-20	13689,9	14486,66895	1,54%	40,27	3,24
Agu-20	13055,3	14618	1,32%	42,61	2,35
Sep-20		14910	1,42%	40,22	2,39
Okt-20	14363,4	14585,02148	1,44%	35,79	3,58
Nov-20	15258,4	14061,64453	1,59%	44,55	2,595
Des-20	16539,6	13832,55957	1,68%	48,52	2,101
Jan-21	15293,7	14023	1,55%		1,964
Feb-21	15256,2	14332,40039	1,38%	61,5	1,991
Mar-21	18354,4	14605	1,37%	59,16	1,567
Apr-21	18490,7		1,42%	63,58	2,286
Mei-21	16932,9	14278	1,68%	66,32	2,698
Jun-21	18542,4	14629,09961	1,33%	73,47	1,324
Jul-21	19385,8	14419,7002	1,52%	73,95	2,599
Agu-21	21427,1	14362	1,59%	68,5	4,782
Sep-21	20605,6		1,60%	75,03	4,371
Okt-21	22029,7	14228,5	1,66%	83,57	5,736
Nov-21	22844,4	14377	1,75%	66,18	3,516
Des-21	22359,5	14285,2002	1,87%		1,008
Jan-22	19173,7	14381,59961	2,18%	88,15	0,933
Feb-22	20472,9	14334,2002	2,06%	95,72	3,834

1. Buka file Rodaku neraca perdagangan.xlsx.
2. Langkah pertama dalam pengolahan data adalah memahami data, salah satunya dengan mengetahui deskripsi data, contoh adalah sebagai berikut:

Kolom	Definisi	Bentuk Data
date	Tanggal disajikan per bulan	Datetime - MMM-YY
nilai_export	Nilai Export dalam jutaan USD	float, numerik
usd_to_idr	Nilai 1 USD dikonversi menjadi rupiah, kurs BI per akhir bulan	float, numerik
inflasi	Nilai inflasi Indeks Harga Konsumen yoy akumulasi berdasarkan data BPS	float, numerik, percent
oil_price	Harga rerata bulanan Indonesian Crude Oil Price (ICP) dalam USD	float, numerik
neraca_perdagangan	Surplus neraca perdagangan dalam miliar USD	float, numerik

Data Cleansing > Imputation

Buka dataset Neraca Perdagangan Impute.xlsx

Value Imputation

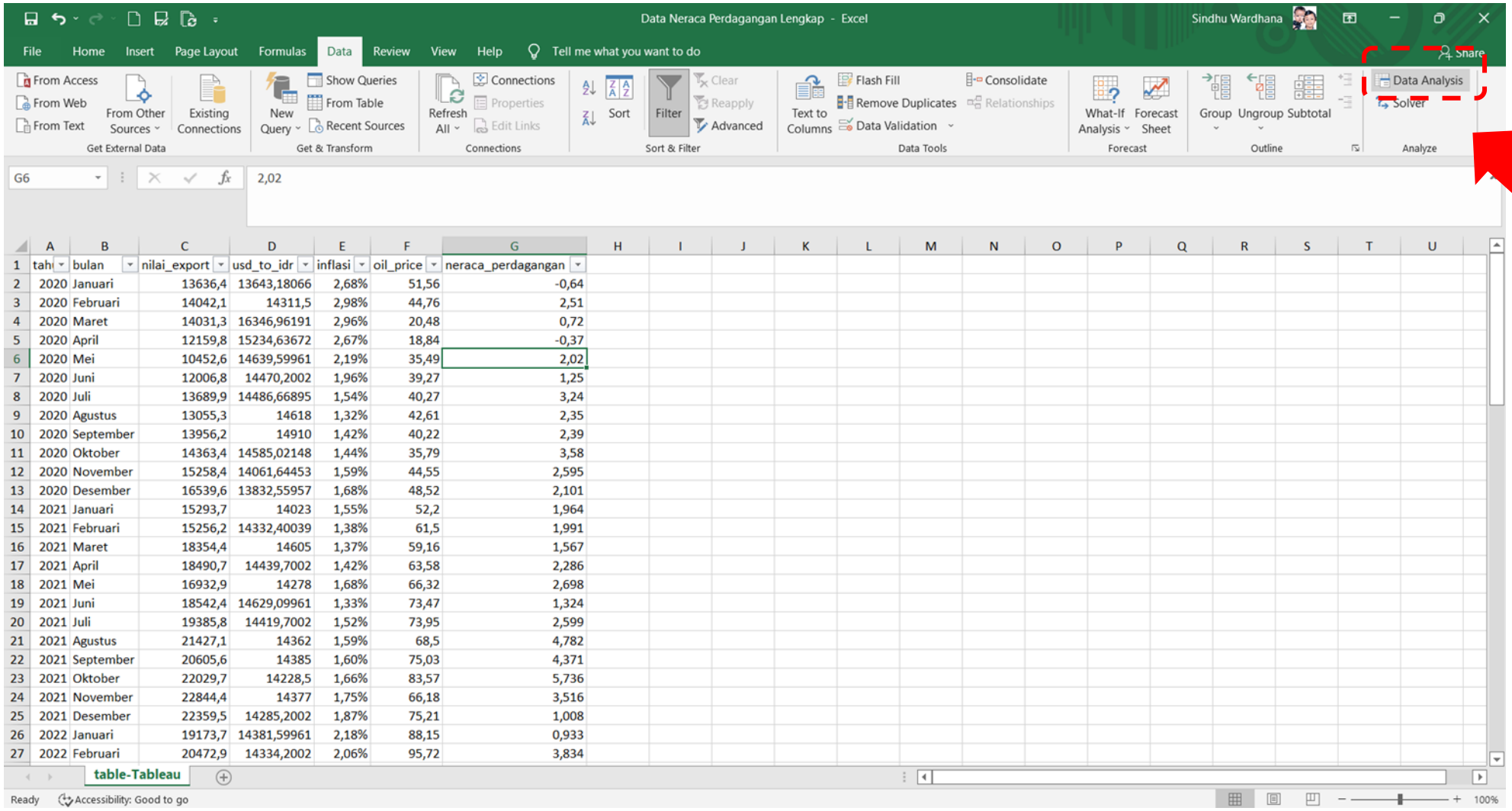
- mode / most frequent (categorical)
- mean / median (numerical)
- Random / defined value
- Model based imputation

Time Series Imputation

- Last Observation Carried Forward (LOCF)
- Next Observation Carried Backward (NOCB)
- Rolling Statistics
- Interpolation

nilai_export	usd_to_idr	inflasi	oil_price
13636,4	13643,18066	2,68%	51,56
14042,1	14311,5	2,98%	44,76
14031,3	16346,96191	2,94%	20,48
12159,8	15234,63672	2,67%	18,84
10452,6	14639,59961	2,19%	35,49
12006,8	14470,2002	1,96%	39,27
13689,9	14486,66895	1,54%	40,27
13055,3	14618	1,32%	42,61
12884,275	14910	1,42%	40,22
14363,4	14585,02148	1,44%	35,79
15258,4	14061,64453	1,59%	44,55
16539,6	13832,55957	1,68%	48,52
15293,7	14023	1,55%	55,476
15256,2	14332,40039	1,38%	61,5
18354,4	14605	1,37%	59,16
18490,7	14411,5498	1,42%	63,58
16932,9	14278	1,68%	66,32
18542,4	14629,09961	1,33%	73,47
19385,8	14419,7002	1,52%	73,95
21427,1	14362	1,59%	68,5
20605,6	14256,5	1,60%	75,03
22029,7	14228,5	1,66%	83,57
22844,4	14377	1,75%	66,18
22359,5	14285,2002	1,87%	78,368
19173,7	14381,59961	2,18%	88,15
20472,9	14334,2002	2,06%	95,72

Mengaktifkan Data Analysis Toolpak (1)



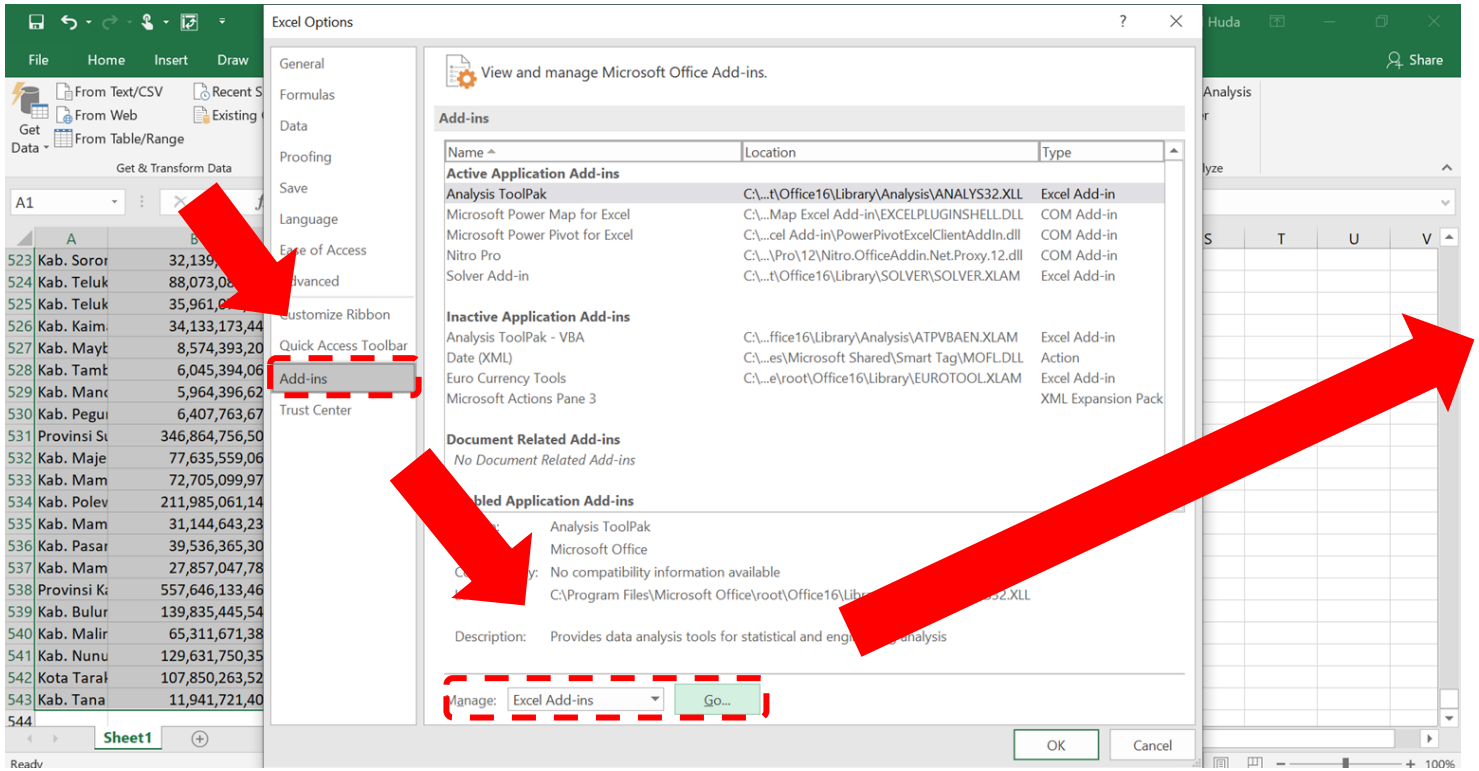
	A	B	C	D	E	F	G
1	tah	bulan	nilai_export	usd_to_idr	inflasi	oil_price	neraca_perdagangan
2	2020	Januari	13636,4	13643,18066	2,68%	51,56	-0,64
3	2020	Februari	14042,1	14311,5	2,98%	44,76	2,51
4	2020	Maret	14031,3	16346,96191	2,96%	20,48	0,72
5	2020	April	12159,8	15234,63672	2,67%	18,84	-0,37
6	2020	Mei	10452,6	14639,59961	2,19%	35,49	2,02
7	2020	Juni	12006,8	14470,2002	1,96%	39,27	1,25
8	2020	Juli	13689,9	14486,66895	1,54%	40,27	3,24
9	2020	Agustus	13055,3	14618	1,32%	42,61	2,35
10	2020	September	13956,2	14910	1,42%	40,22	2,39
11	2020	Oktober	14363,4	14585,02148	1,44%	35,79	3,58
12	2020	November	15258,4	14061,64453	1,59%	44,55	2,595
13	2020	Desember	16539,6	13832,55957	1,68%	48,52	2,101
14	2021	Januari	15293,7	14023	1,55%	52,2	1,964
15	2021	Februari	15256,2	14332,40039	1,38%	61,5	1,991
16	2021	Maret	18354,4	14605	1,37%	59,16	1,567
17	2021	April	18490,7	14439,7002	1,42%	63,58	2,286
18	2021	Mei	16932,9	14278	1,68%	66,32	2,698
19	2021	Juni	18542,4	14629,09961	1,33%	73,47	1,324
20	2021	Juli	19385,8	14419,7002	1,52%	73,95	2,599
21	2021	Agustus	21427,1	14362	1,59%	68,5	4,782
22	2021	September	20605,6	14385	1,60%	75,03	4,371
23	2021	Oktober	22029,7	14228,5	1,66%	83,57	5,736
24	2021	November	22844,4	14377	1,75%	66,18	3,516
25	2021	Desember	22359,5	14285,2002	1,87%	75,21	1,008
26	2022	Januari	19173,7	14381,59961	2,18%	88,15	0,933
27	2022	Februari	20472,9	14334,2002	2,06%	95,72	3,834

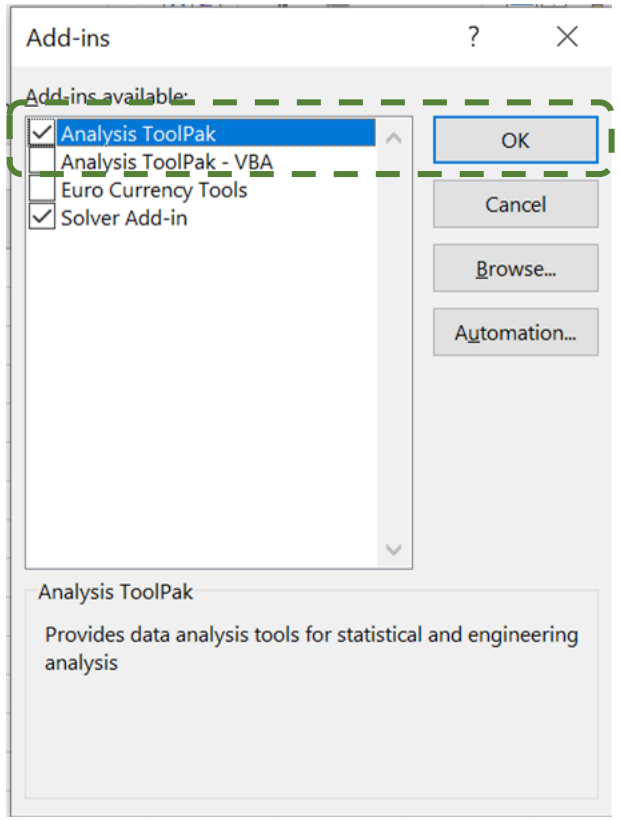
Excel Memiliki fungsi bawaan yang tidak langsung aktif yaitu **Data Analysis Toolpack,** yang dapat kita gunakan dalam kegiatan data analytics.

Mengaktifkan Data Analysis Toolpak (2)

Cara Memunculkan Data Analysis

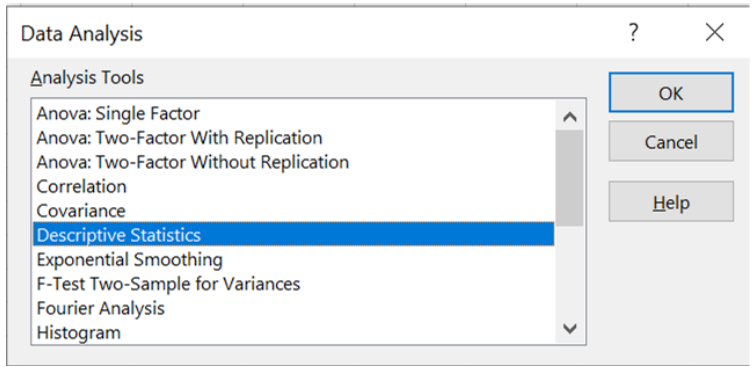
- Klik Tab File Options sehingga muncul jendela berikut
- Klik Add-ins Excel Add-ins Go
- Check Analysis Toolpak OK



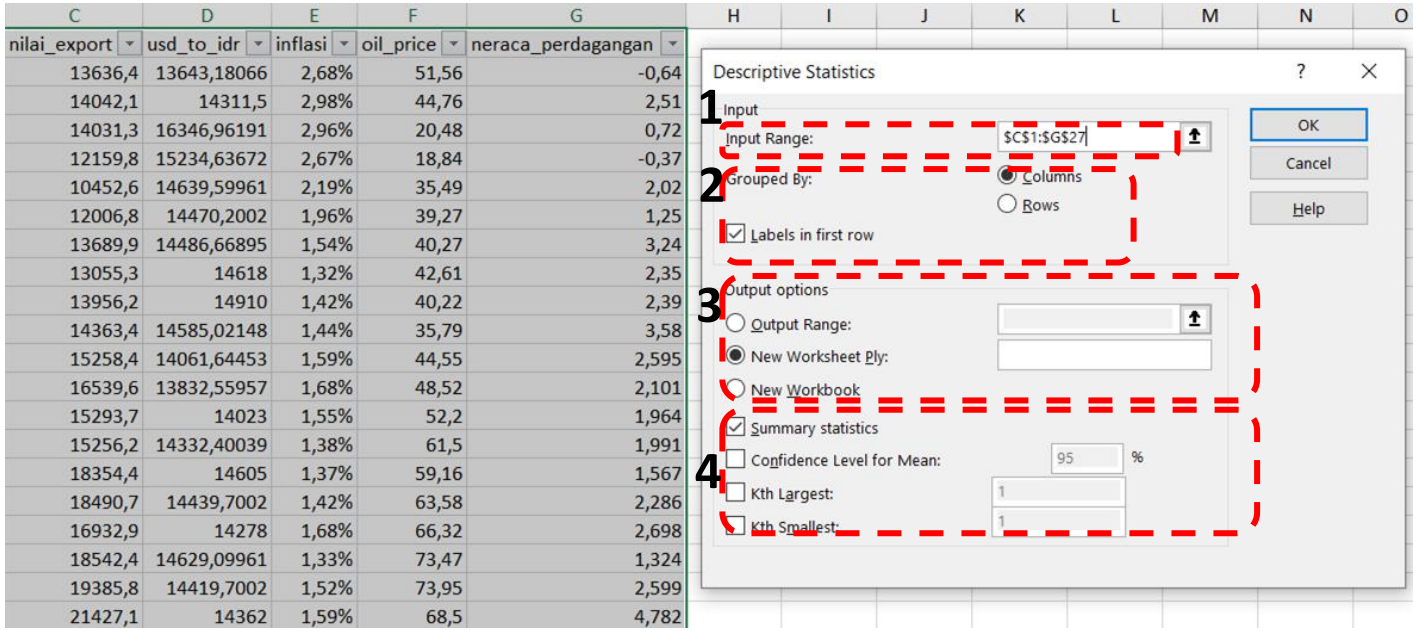


Descriptive Analysis (1)

Untuk mengetahui gambaran mengenai ukuran pemusatan dan sebaran data



1. **Buka dataset Neraca Perdagangan.xlsx**
2. **Click Data Analysis**
3. **Pilih Descriptive Statistic**
4. **Klik "OK"**



1. **Pilihan Input Range dari data yang mau dianalisis**
2. **Grouped By:**
   a. **Columns ⍰ data dikelompokkan berdasarkan kolom seperti pada contoh**
   b. **Rows ⍰ data dikelompokkan berdasarkan baris**

   Check Labels in First Row jika data mempunyai nama kolom seperti pada contoh
3. **Output dari analisis dapat dipilih di Range tertentu, Worksheet baru, atau Workbook baru**
4. **Check Summary statistics untuk mendapatkan hasil dari Descriptive Statistics**

Descriptive Analysis (2)

untuk mengetahui gambaran mengenai ukuran pemusatan dan sebaran data

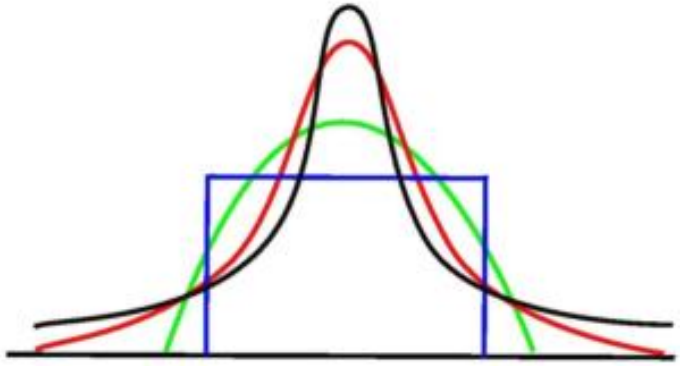
Hasil Descriptive Statistic:

oil_price	
Mean	56,34230769
Standard Error	3,915612745
Median	55,68
Mode	#N/A
Standard Deviation	19,9657858
Sample Variance	398,6326025
Kurtosis	-0,595901013
Skewness	0,027533458
Range	76,88
Minimum	18,84
Maximum	95,72
Sum	1464,9
Count	26

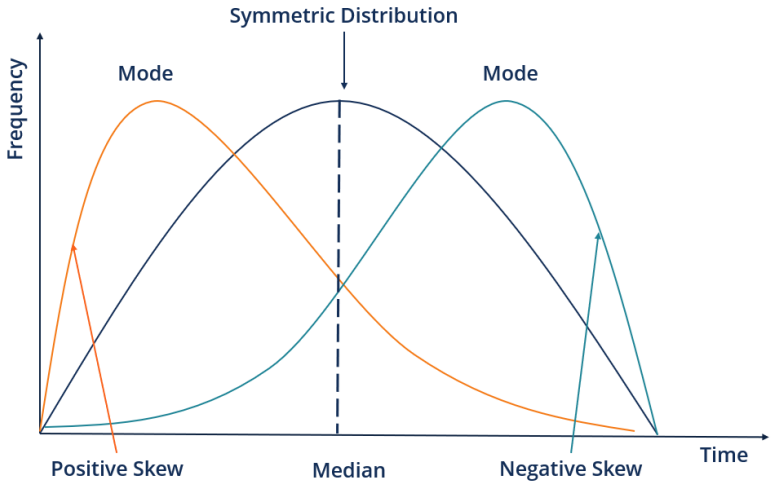
Glossary:

Parameter	Deskripsi	Fungsi Excel
Mean	Mean adalah rata-rata dari suatu dataset.	AVERAGE
Standar Error	Perkiraan kesalahan dalam sampel Mean.	STDEV.S/SQRT(COUNT)
Median	Nilai tengah dari suatu dataset.	MEDIAN
Mode/Modus	Nilai yang paling sering muncul/umum dalam suatu dataset	MODE
Standar Deviation	Ukuran dari variabilitas data dari mean	STDEV.S
Sample Variance	Kuadrat dari tandar deviasi	VAR
Kurtosis	Mengukur perbedaan ekor distribusi dari distribusi normal	KURT
Skewness	Mengukur ketidaksimetrisan distribusi dari distribusi normal	SKEW
Range	Selisih antara nilai maksimum dan minimum dari dataset	MAX-MIN
Minimum	Nilai terkecil dari dataset	MIN
Maximum	Nilai terbesar dari dataset	MAX
Sum	Jumlah dari semua nilai dari dataset	SUM
Count	Jumlah row/tuple/nilai dari dataset	COUNT

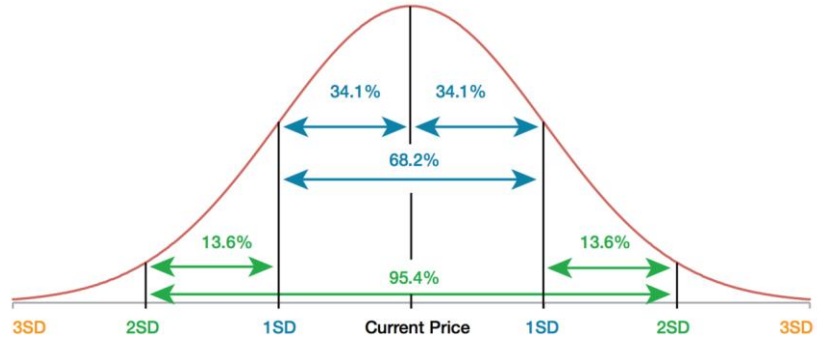
Descriptive Analysis (3) – Penjelasan



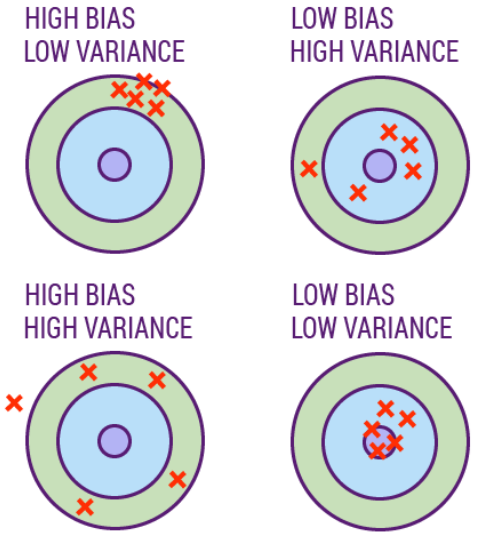
kurtosis



skewness



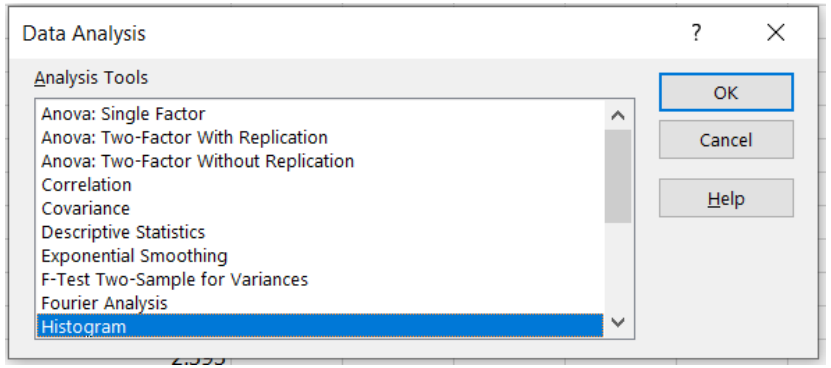
standard deviation



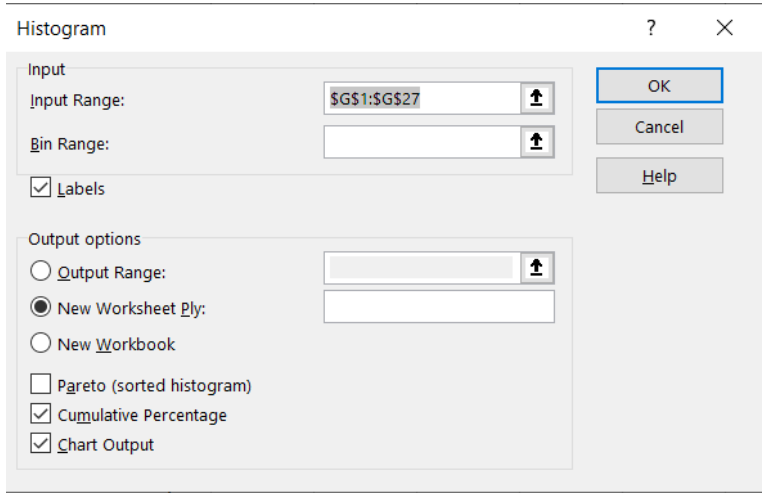
bias & variance

Descriptive Analysis (4) – Distribusi Data

Visualisasi Data untuk mendapatkan gambaran distribusi data, misal: **Oil Prices**



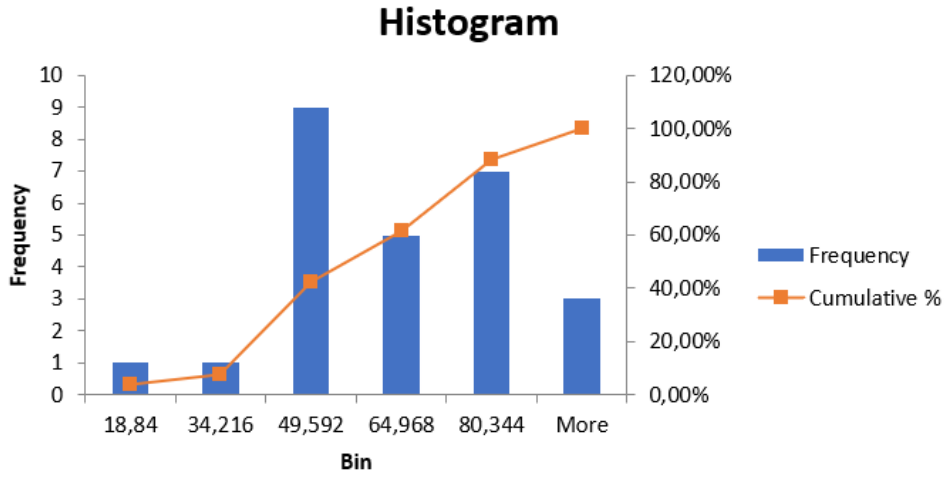
1. Pilih Histogram di Data Analysis



2. Pilih Dataset missal kolom G, centang Cumulative dan Chart Output

Bin	Frequency	Cumulative %
18,84	1	3,85%
34,216	1	7,69%
49,592	9	42,31%
64,968	5	61,54%
80,344	7	88,46%
More	3	100,00%

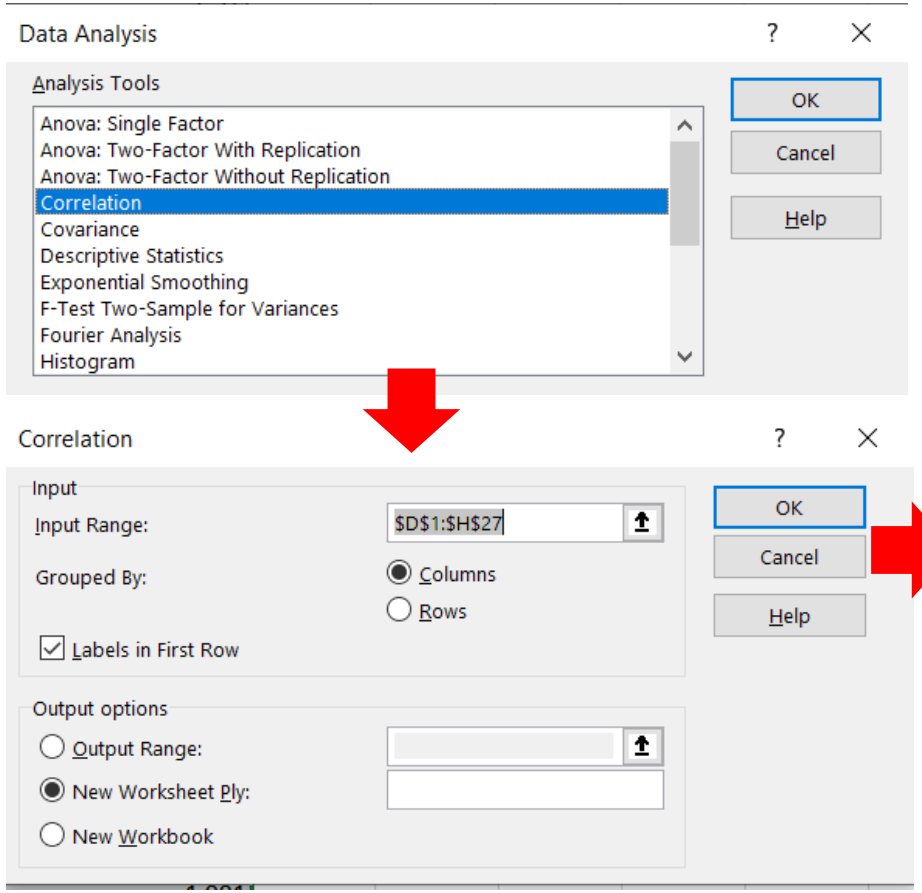
3. Akan secara otomatis mengelompokkan ke Bin

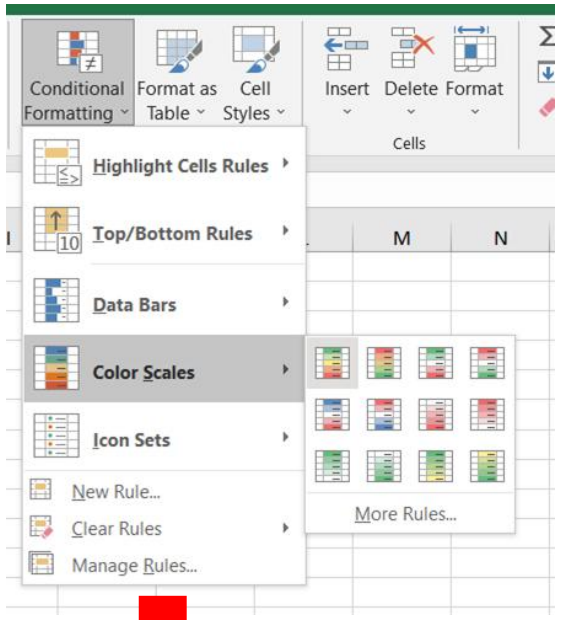


4. Akan secara otomatis membuat distribusi dataset

Descriptive Analysis (5) - Korelasi

Korelasi antar variable makro dan neraca perdagangan





1. Pilihan Correlation di Data Analysis.
2. Pilih kolom yang akan dicari korelasinya, korelasi ini hanya berlaku untuk angka.
3. Jangan lupa centang “Label in The First Row”.
4. Setelah jadi lakukan “conditional formatting> color scales” agar bisa secara intuitif melihat korelasi dengan lebih mudah.
5. Semakin mendekati 1 berarti korelasi positif semakin kuat, semakin mendekati -1 berarti korelasi negative semakin kuat.
6. Ini akan memudahkan kita dalam melakukan pencarian fitur mana yang akan kita gunakan sebagai fitur.

	nilai_export	*usd_to_idr*	*inflasi*	*oil_price*	*neraca_perdagangan*
nilai_export	1				
usd_to_idr	-0,25672504	1			
inflasi	-0,311056539	0,316073134	1		
oil_price	0,839498984	-0,47267549	-0,325959534	1	
neraca_perdagangan	0,490820055	-0,211617555	-0,458476591	0,415238722	1

Kriteria Data untuk Visualisasi Data

Tabel 2 dimensi / flat

Terdiri dari baris dan kolom tanpa merge cell

Bukan data agregasi

Berupa data mentah

Mana yang memenuhi kriteria?

No	Pelatihan	Lokasi Pelatihan	Batch	ID Pelatihan	Kelas	Mulai Pelatihan	Selesai Pelatihan	Rekap Pendaftar Awal	Rekap Peserta Diterima	Jenis Kelamin		Rekap Peserta Onboarding	Jenis Kelamin		Rekap Peserta Completion	Jenis Kelamin	
										L	P		L	P		L	P
1	The Science of Well Being	Online Target: 100	1	567	A	1 Januari 2023	7 Januari 2023	100	34	20	14	30	16	14	28	14	14
					B				34	17	17	31	15	16	30	15	15
					C				32	20	12	32	20	12	30	19	11
		Timika Target: 25	2	876	A	4 Maret	10 Maret 2023	26	24	13	11	24	13	11	18	10	8
2	Successful Negotiation	Batam Target: 60	I	213	A	1 Februari 2023	5 Februari 2023	55	28	15	13	25	13	12	23	11	12
					B				27	16	11	27	16	11	26	16	10
		Offline (Kab. Mojokerto) Target: 25	II	789	A	03 Juli 2023	06 Juli 2023	25	25	10	15	23	9	14	23	9	14

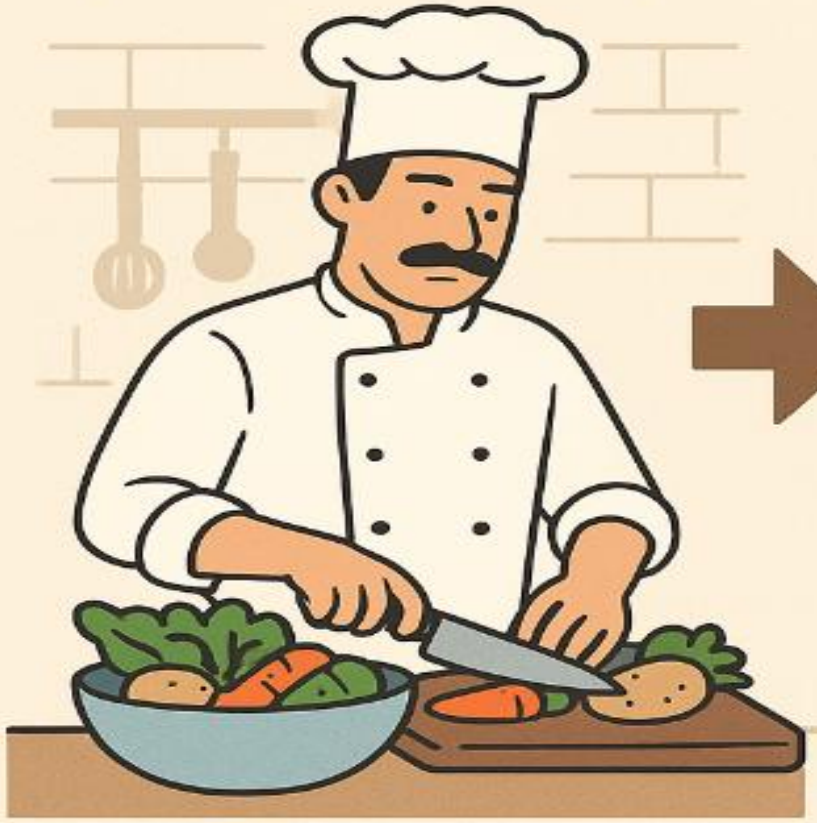
VS

Nama	Jenis Kelamin	Batch	ID Pelatihan	Nama Pelatihan	Lokasi Pelatihan	Status Diterima	Status Onboarding	Status Completion
Andi	Pria	1	567	The Science of Well Being	Online	1	1	1
Ani	Wanita	1	568	Successful Negotiation	Bekasi	1	1	1
Budi	Pria	1	568	Successful Negotiation	Bekasi	1	0	0
Beta	Wanita	1	567	The Science of Well Being	Online	0	0	0
Cici	Wanita	1	580	The Science of Well Being	Online	1	1	1
Ciko	Pria	1	568	Successful Negotiation	Bekasi	1	1	1
Didi	Pria	1	580	The Science of Well Being	Online	1	0	0
Dini	Wanita	1	568	Successful Negotiation	Bekasi	0	0	0
Eri	Pria	1	568	Successful Negotiation	Bekasi	1	1	1
Edi	Pria	1	580	The Science of Well Being	Online	1	1	0

Extract
Transform
Load

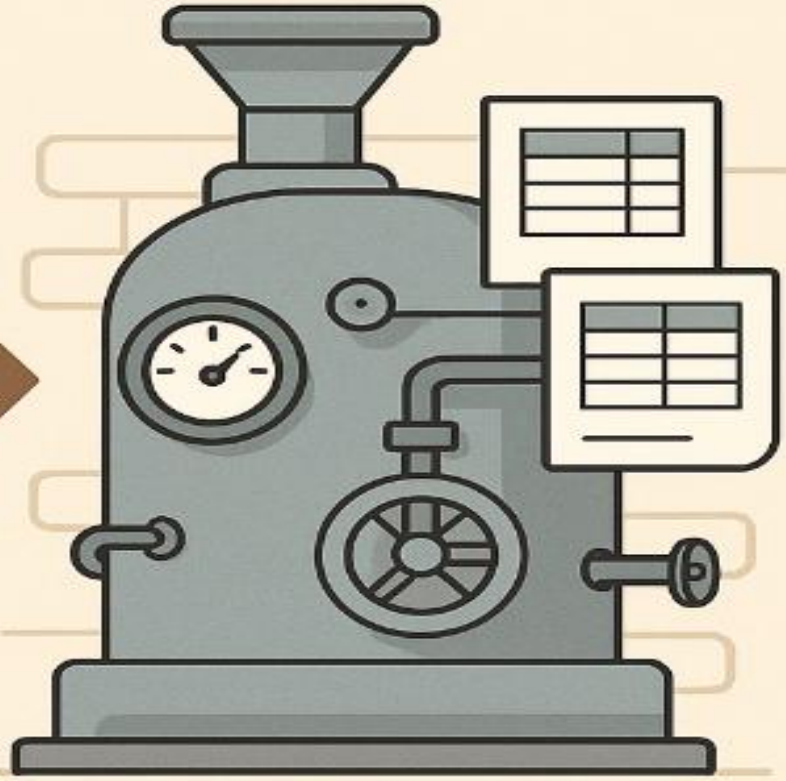
Kemenkeu
Corporate University

DAPUR
(RAW MATERIAL)



Power Query

MESIN

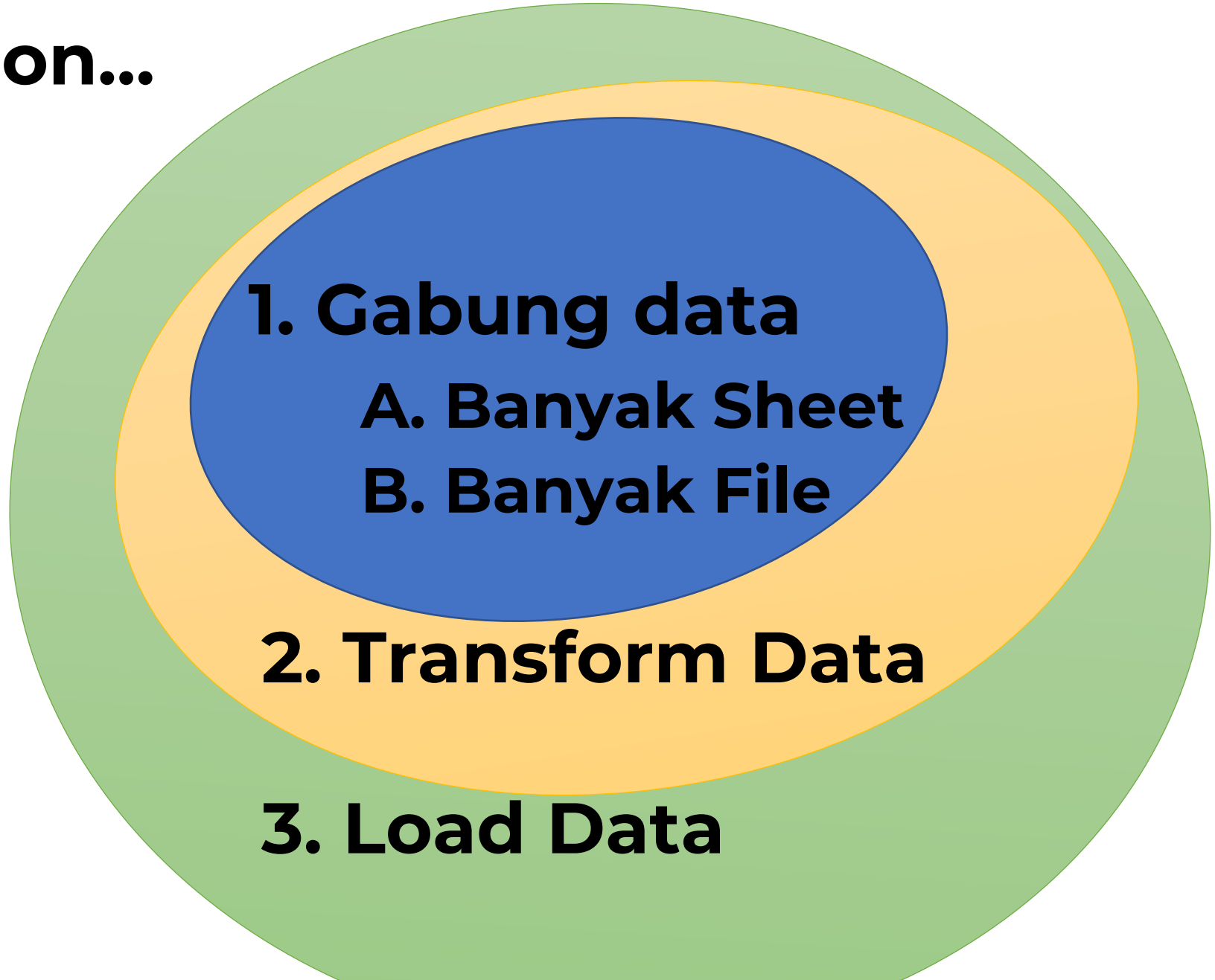


Power Pivot

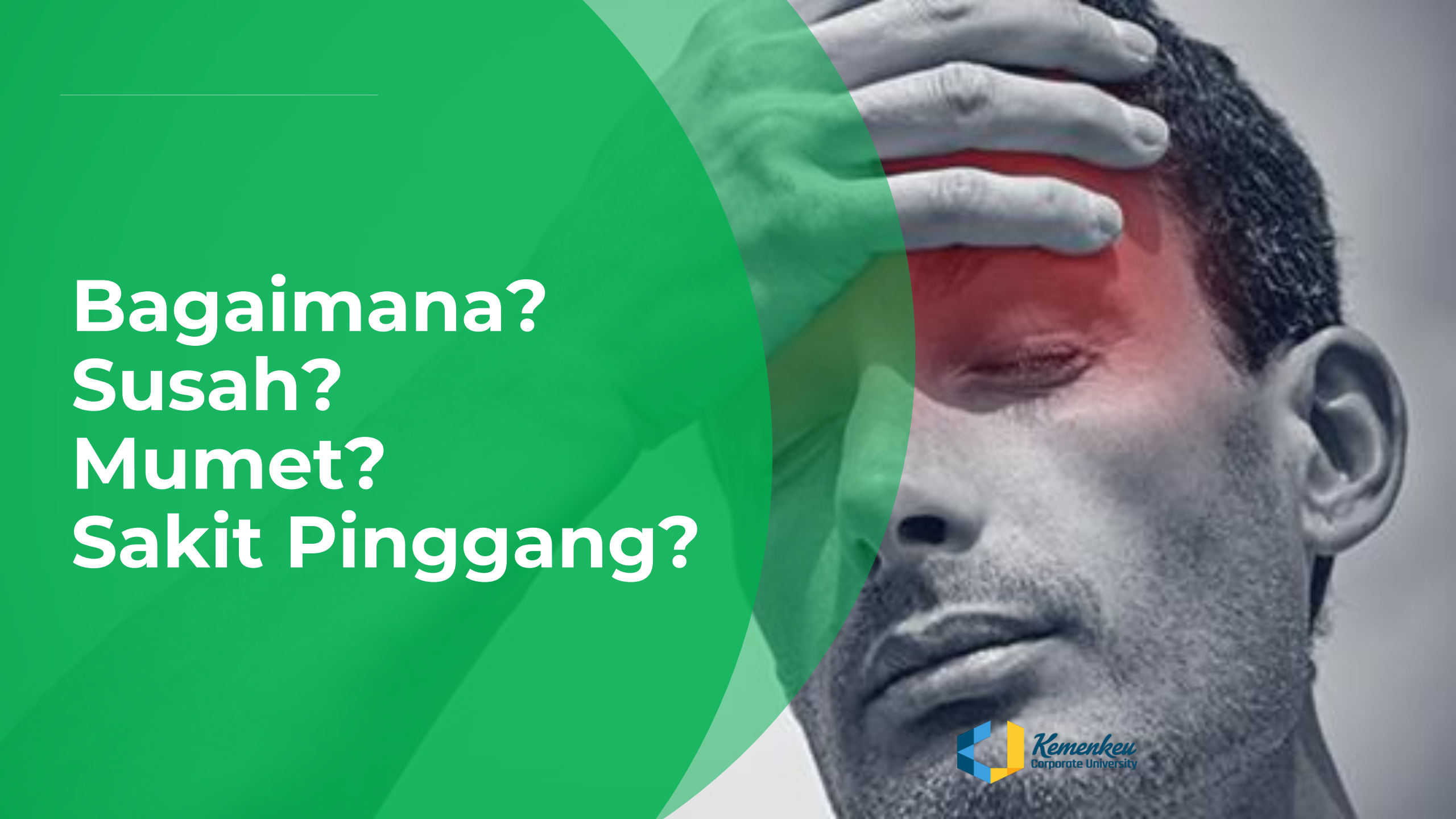
ETALASE
& RUANG MAKAN

Power BI
(or Excel PivotTable)

Hands on...


1. Gabung data
A. Banyak Sheet
B. Banyak File
2. Transform Data
3. Load Data

Bagaimana?
Susah?
Mumet?
Sakit Pinggang?



Aman Boss

Kemenkeu
Corporate University

Terima Kasih.

Kemenkeu Bicara DATA

Kemenkeu Corporate University