

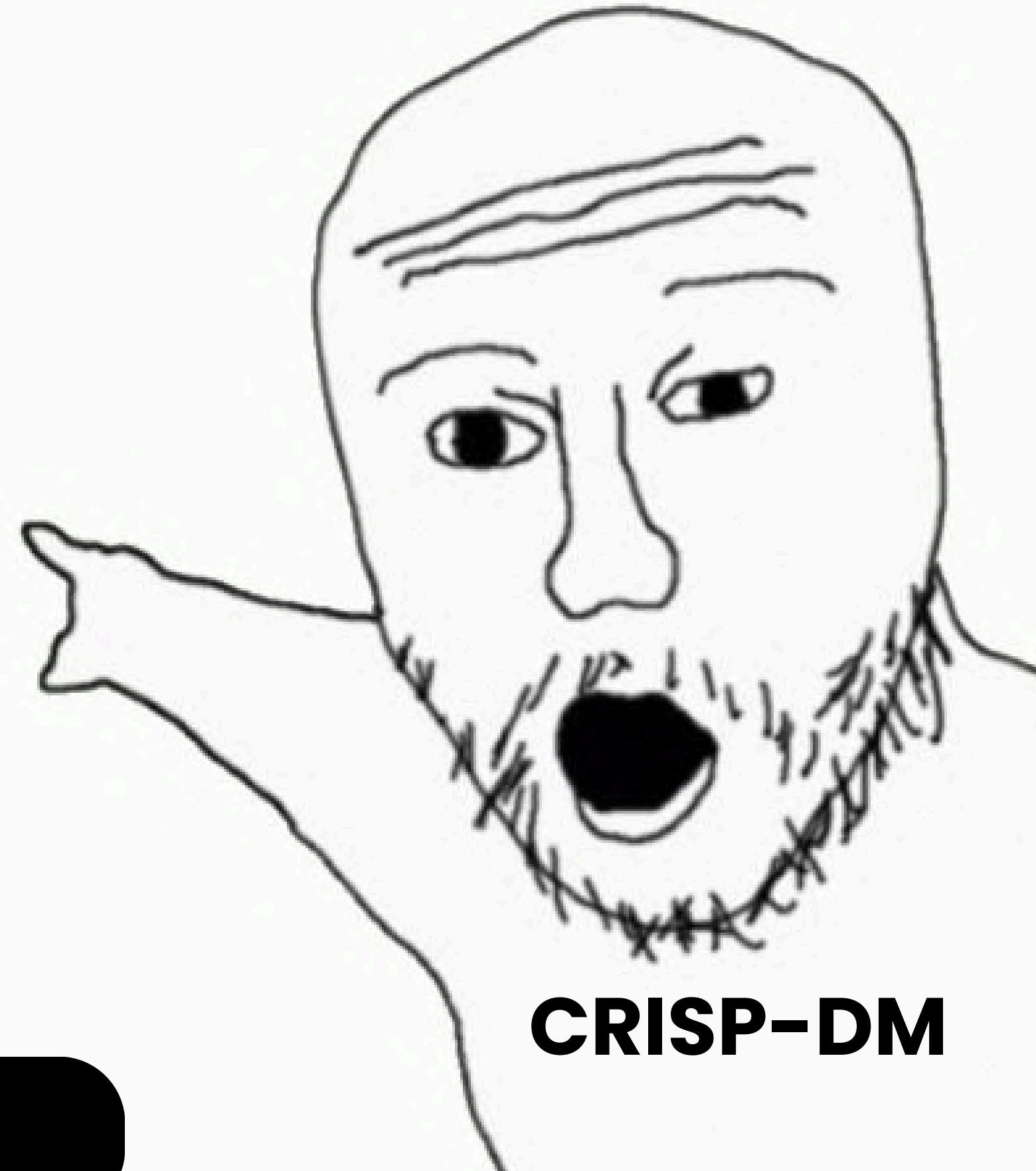
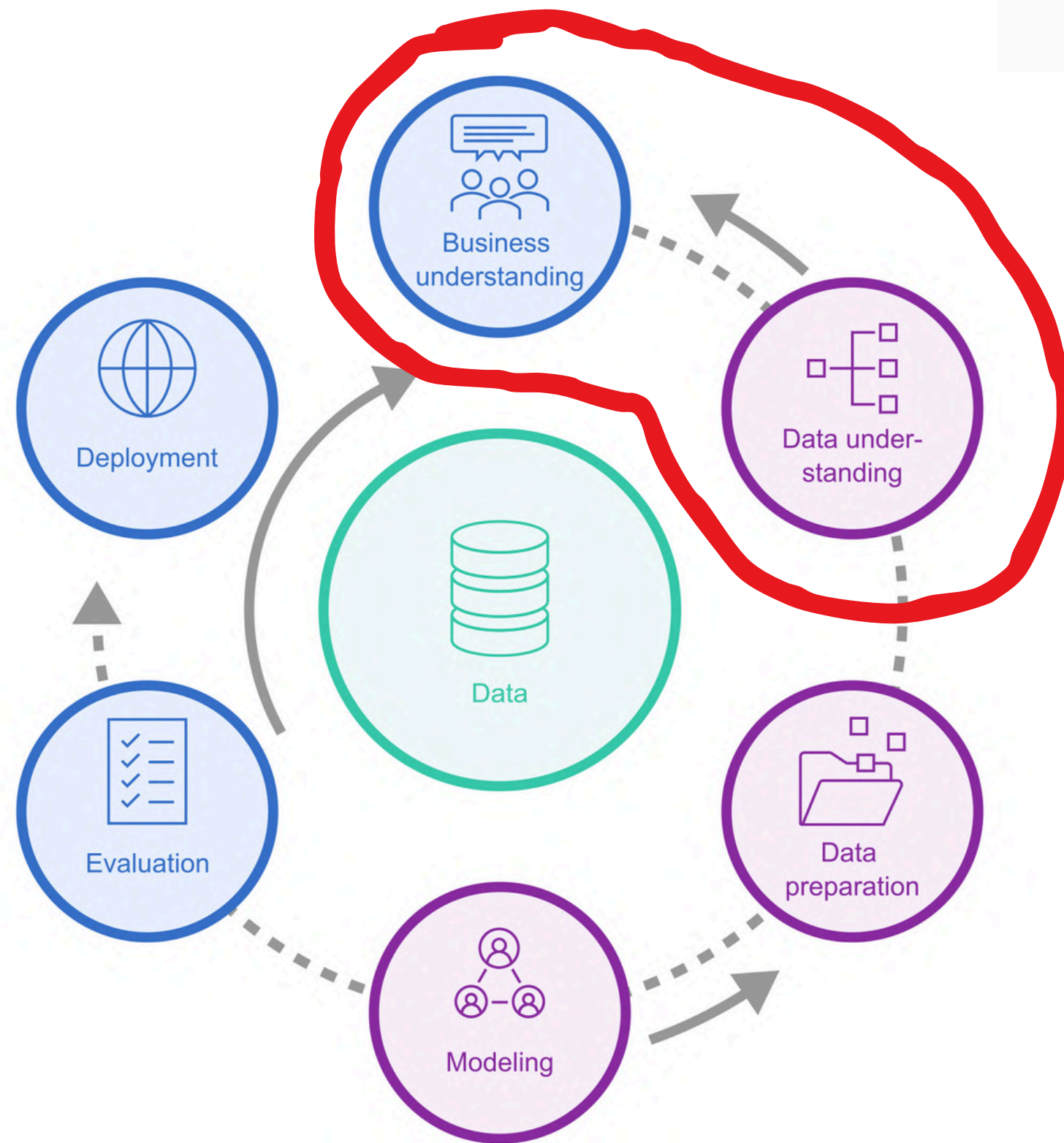


Enterprise  
Data  
Management  
Laboratory

**Grow With EDM**

# **Business & Data Understanding**

**GWE #3**



**CRISP-DM**

**GWE #3**





Enterprise  
Data  
Management  
Laboratory

# Apa Itu Business Understanding?

**GWE #3**

# Apa itu Business Understanding?

Business Understanding adalah tahap pertama dalam metodologi CRISP-DM (Cross-Industry Standard Process for Data Mining). Tujuan utamanya adalah untuk memahami tujuan bisnis dari proyek yang akan dijalankan dan bagaimana data dapat membantu mencapai tujuan tersebut.

stonks



# Langkah-langkah dalam Business Understanding

## Menentukan Tujuan Bisnis

- Apa yang ingin dicapai oleh organisasi?
- Misalnya: Meningkatkan penjualan, mengurangi churn pelanggan, atau memprediksi permintaan produk.

## Mengonversi Tujuan Bisnis ke Tujuan Analitik

- Bagaimana data dan analitik dapat membantu mencapai tujuan tersebut?
- Misalnya: Memprediksi pelanggan yang akan berhenti berlangganan atau mengklasifikasikan produk berdasarkan kategori.

## Menentukan Success Criteria

- Apa metrik yang digunakan untuk menilai apakah proyek berhasil?
- Misalnya: Tingkat akurasi model, ROI, atau penurunan tingkat churn.



# Contoh Penerapan Business Understanding

## Kasus: Meningkatkan Retensi Pelanggan pada E-commerce

1. Tujuan Bisnis: Mengurangi tingkat churn pelanggan dan meningkatkan retensi pelanggan.
2. Tujuan Analitik: Mengembangkan model prediksi yang dapat mengidentifikasi pelanggan yang berisiko berhenti berlangganan.
3. Kriteria Sukses: Model yang dihasilkan harus dapat memprediksi dengan akurasi minimal 85% untuk pelanggan yang berisiko churn.





# Apa Itu Data Understanding?

# Apa itu Data Understanding?

Data Understanding adalah tahap kedua dalam CRISP-DM yang fokus pada pengumpulan, eksplorasi, dan pemahaman lebih lanjut tentang data yang ada. Tujuan utama adalah untuk mengeksplorasi dataset yang sudah tersedia, memverifikasi kualitas data, dan menentukan apakah data tersebut relevan dengan tujuan analitik.





# Langkah-langkah dalam Data Understanding

1

## Mengumpulkan Data:

- Proses pertama adalah memperoleh dataset yang akan dianalisis. Biasanya, data datang dari berbagai sumber seperti database, file CSV, atau API.
- Contoh: Mengunduh data transaksi pelanggan dari sistem e-commerce / kaggle



# Memahami Struktur & Karakteristik Data

2

| Tahapan  | Library | Syntax Relevan   |
|--|---------|--|
| Data understanding:<br>Loading data                      | pandas  | <ul style="list-style-type: none"><li>• <b>pd.read_csv()</b>: membaca data dari csv</li><li>• <b>pd.read_excel()</b>: membaca data dari excel</li><li>• <b>pd.read_sql_query()</b>: membaca data dari sql</li></ul>  |
| Data understanding:<br>menampilkan<br>karakteristik data | pandas  | <ul style="list-style-type: none"><li>• <b>head()</b>: Menampilkan beberapa baris pertama data.</li><li>• <b>tail()</b>: Menampilkan beberapa baris terakhir data.</li><li>• <b>head(n)</b>: Menampilkan n baris pertama. Defaultnya menampilkan 5 baris.</li><li>• <b>tail(n)</b>: Menampilkan n baris terakhir. Defaultnya menampilkan 5 baris.</li><li>• <b>describe()</b>: Menyediakan statistik deskriptif seperti mean, std, min, max, dll.</li><li>• <b>info()</b>: Menampilkan informasi tentang DataFrame, termasuk tipe data dan jumlah nilai kosong.</li><li>• <b>value_counts()</b>: Menghitung frekuensi nilai unik dalam suatu kolom.</li><li>• <b>value_counts(normalize=False, sort=True, ascending=False, dropna=True)</b>: Menghitung frekuensi unik dari suatu kolom. normalize=True mengembalikan proporsi, sort=True mengurutkan berdasarkan frekuensi.</li><li>• <b>corr()</b>: Menghitung korelasi antar variabel numerik.</li><li>• <b>shape</b>: Menampilkan dimensi dari DataFrame (baris dan kolom).</li><li>• <b>dtypes</b>: Menampilkan tipe data dari masing-masing kolom.</li><li>• <b>unique()</b>: Menampilkan nilai unik dari suatu kolom.</li><li>• <b>nunique()</b>: Menghitung jumlah nilai unik dari suatu kolom.</li><li>• <b>nunique(dropna=True)</b>: Menghitung jumlah nilai unik. Jika dropna=False, NaN juga dihitung.</li><li>• <b>df['column_name'].mean()</b>: Menampilkan rata-rata</li><li>• <b>df['column_name'].median()</b>: Menampilkan median</li><li>• <b>df['column_name'].mode()</b>: Menampilkan modus</li></ul> |

Contohnya di handson ye



3

# Apa itu Exploratory Data Analysis (EDA) ?



Enterprise  
Data  
Management  
Laboratory

# Apa itu EDA?

- Langkah awal dalam analisis data
- Tujuan: Memahami data, menemukan pola, dan melihat hubungan antar fitur
- Membantu mengidentifikasi outlier atau data tidak wajar
- Dilakukan sebelum analisis statistik lebih lanjut atau pemodelan

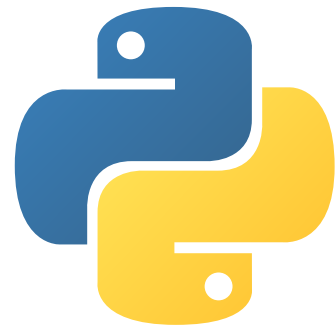




# Kenapa ~~bandung~~ EDA?

- ✓ Memahami struktur dan distribusi data
- ✓ Menemukan pola tersembunyi & hubungan antar variabel
- ✓ Mendeteksi outlier atau anomali
- ✓ Menentukan teknik analisis atau pemodelan yang sesuai

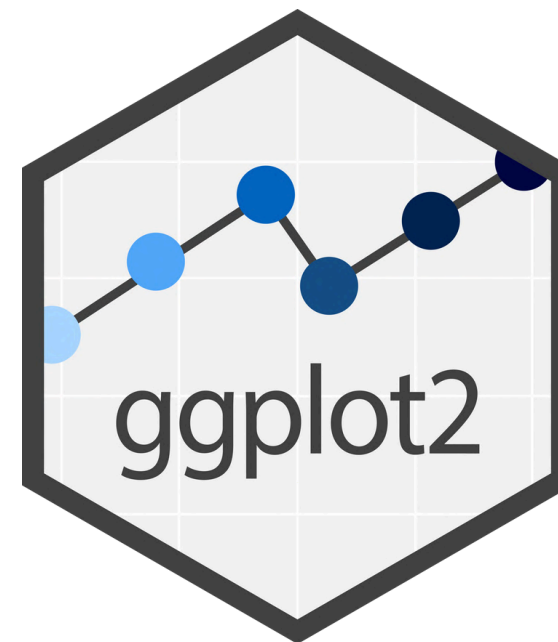
# Viz Tools yang biasa dipake



*matplotlib*

 plotly

 seaborn



 + a b l e a u

 Power BI

 Looker

 Excel



# Types of Exploratory Data Analysis

Univariate  
Analysis

Bivariate  
Analysis

Multivariate  
Analysis

# Univariate Analysis

## Contoh

### Box Plot

Analisis data yang hanya melibatkan **satu variabel**.

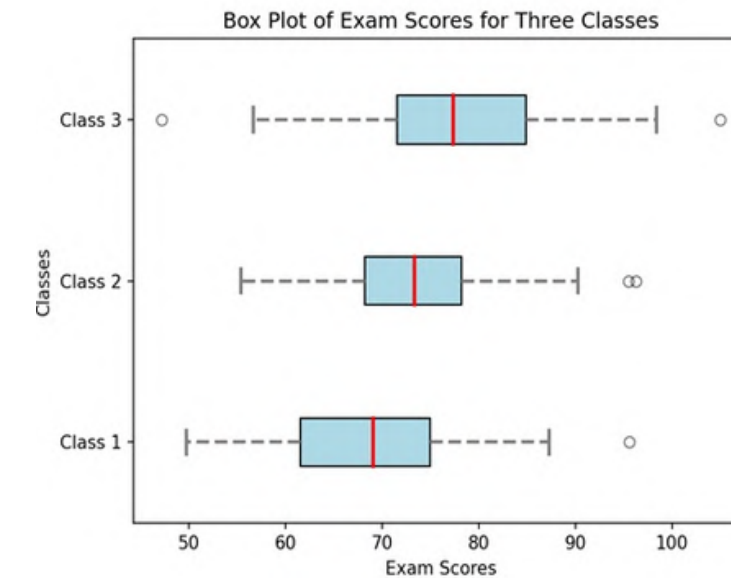
📌 Metode Umum:

✅ Statistik Deskriptif:

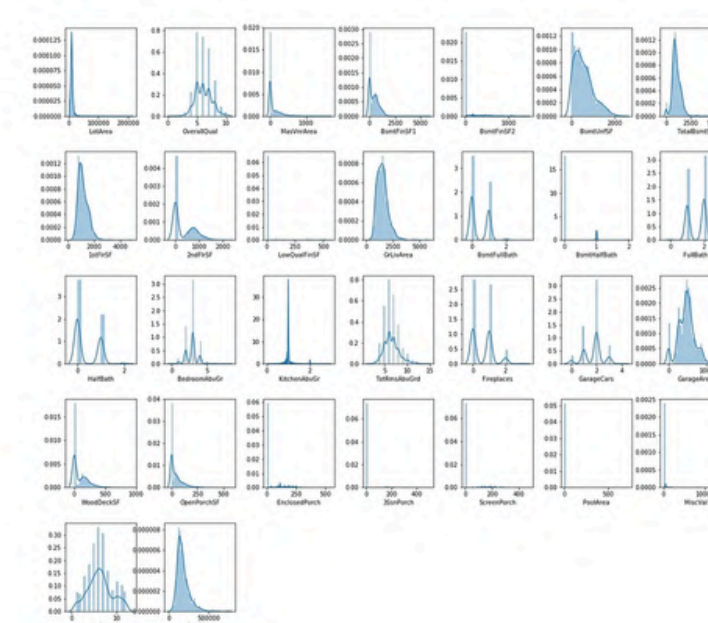
Mean, Median, Mode, Range, Standard Deviation

✅ Visualisasi:

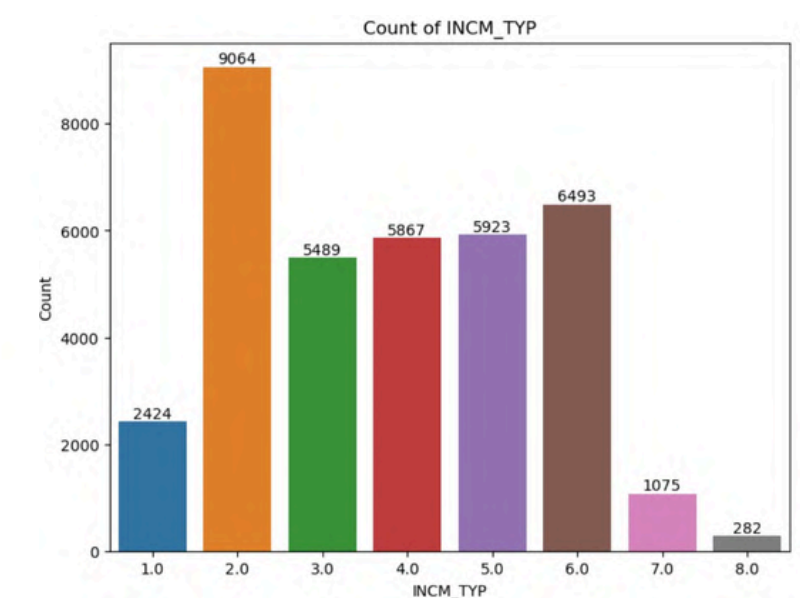
- Histogram 📊 → Distribusi frekuensi variabel
- Box Plot 📦 → Menampilkan outlier dan distribusi data
- Violin Plot 🎻 → Kombinasi histogram dan box plot



### Histogram



### Count Plot (1 Feature)





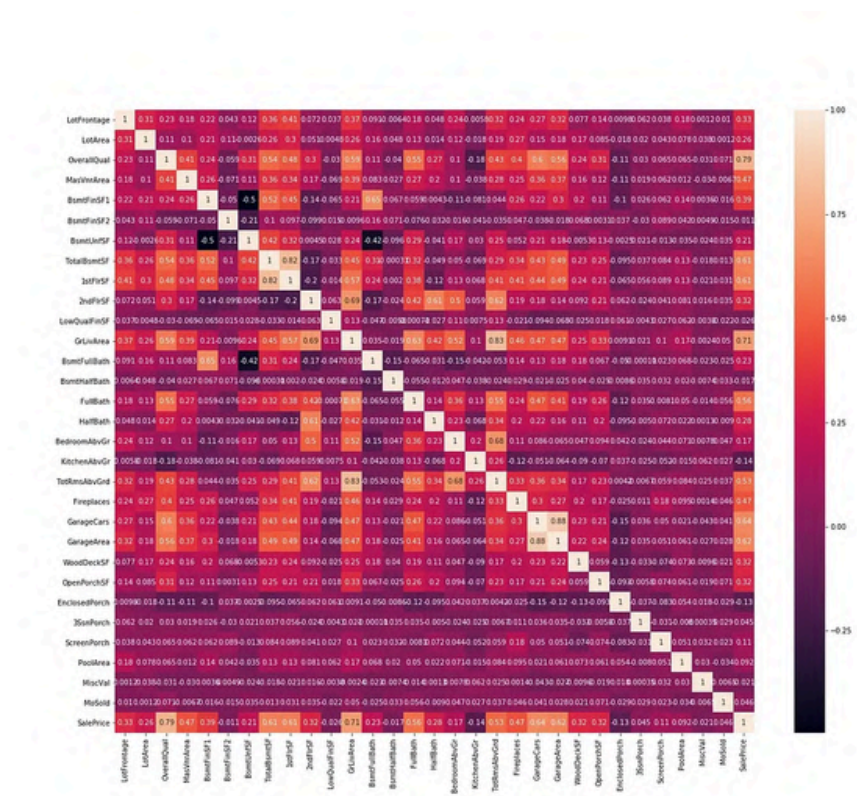
# Bivariate Analysis

Membandingkan **dua variabel** untuk melihat hubungan atau pola antar keduanya.

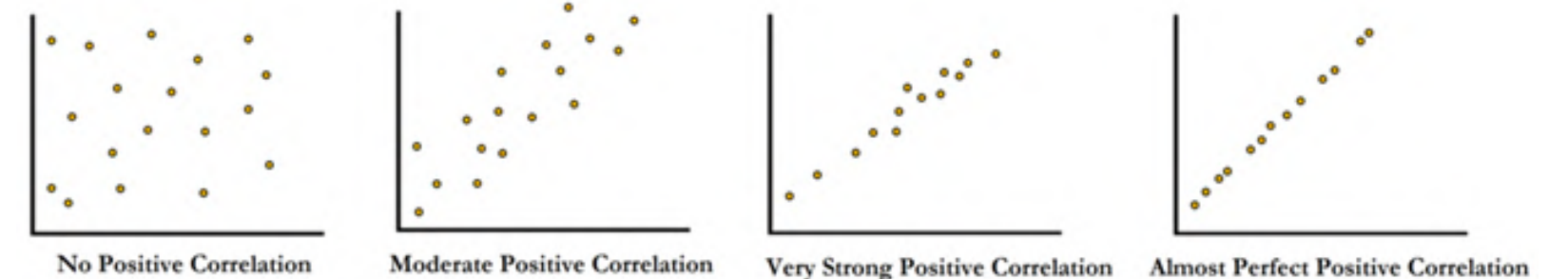
## Contoh

📌 Metode Umum:

- ✅ Scatter Plot ● → Melihat pola hubungan antara dua variabel
- ✅ Regression Plot 📈 → Menganalisis tren dalam data
- ✅ Correlation Coefficient (Pearson/Spearman) → Menentukan kekuatan hubungan antar variabel



Corelation Matrix & Scatter Plot



# Multivariate Analysis

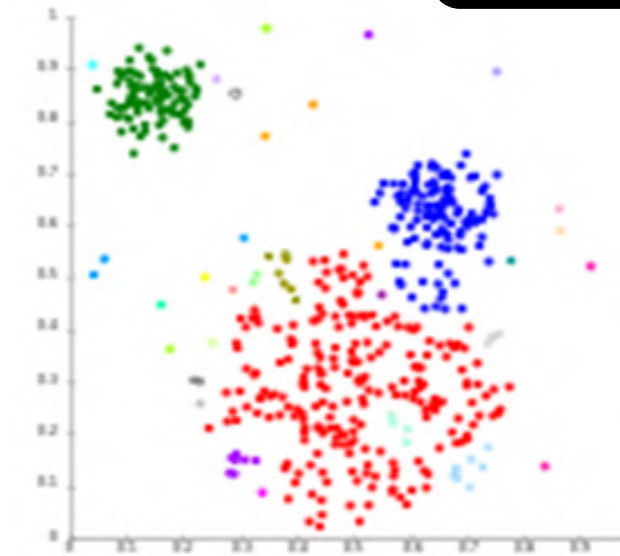
## Contoh

Melibatkan **lebih dari dua variabel** untuk melihat hubungan kompleks dalam dataset.

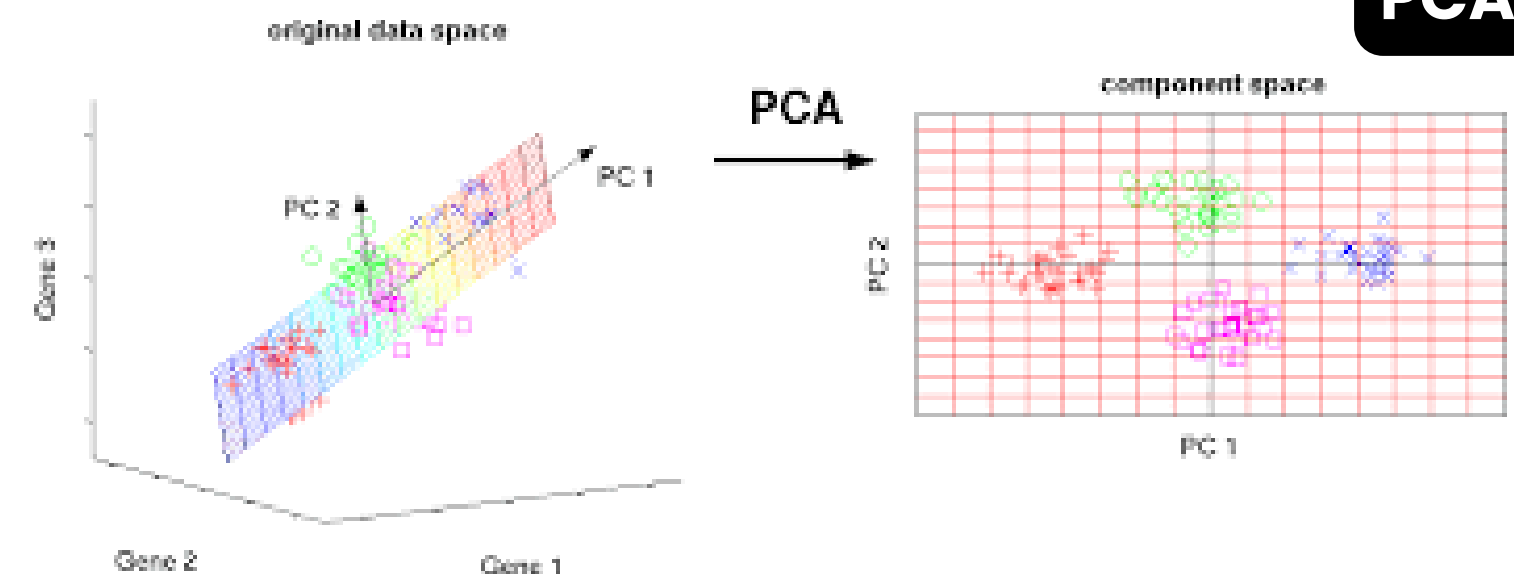
📌 Metode Umum:

- ✓ Pairplot (Scatter Plot Matrix) 📊 → Visualisasi banyak hubungan dalam satu gambar
- ✓ Principal Component Analysis (PCA) 🔍 → Mengurangi dimensi data tanpa kehilangan informasi penting
- ✓ Cluster Analysis 🔗 → Mengelompokkan data berdasarkan kemiripan

## Cluster Analysis



## PCA





4

# Apa itu Data Quality Assessment?



# Apa itu Data Quality Assessment?

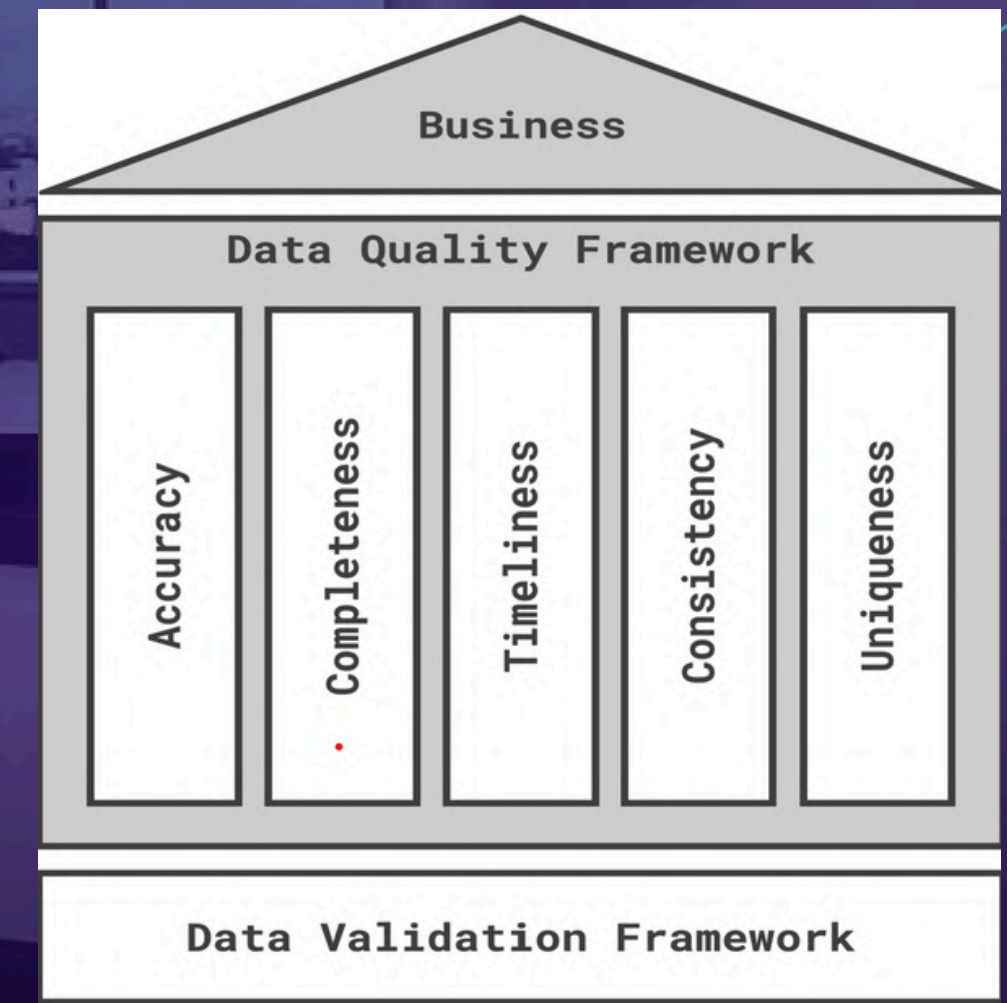
Proses untuk mengevaluasi kualitas data agar dapat digunakan dengan baik. Proses ini memastikan bahwa data yang digunakan akurat, lengkap, konsisten, relevan, dan dapat diandalkan.





# DQA mencakup beberapa aspek

- ✓ Accuracy (Akurasi) – Apakah data sesuai dengan fakta?
- ✓ Completeness (Kelengkapan) – Apakah ada data yang hilang?
- ✓ Consistency (Konsistensi) – Apakah data selaras di berbagai sumber atau kolom?
- ✓ Uniqueness (Keunikan) – Apakah data tidak memiliki duplikasi yang tidak perlu?





Enterprise  
Data  
Management  
Laboratory

# Let's try



# Q & A



Enterprise  
Data  
Management  
Laboratory

# Quiz



## PENUGASAN MODUL 3

### **Kasus: Analisis Faktor yang Mempengaruhi Kesehatan Mental Mahasiswa dan Pekerja Profesional**

1. Melakukan Exploratory Data Analysis (EDA) pada dataset.
2. Memahami distribusi data terkait kesehatan mental individu.
3. Menganalisis hubungan antar variabel dalam dataset.
4. Mengidentifikasi pola yang berpotensi mempengaruhi kesehatan mental mahasiswa dan pekerja profesional.
5. Penugasan bisa di akses di link berikut :





# ATURAN PENUGASAN MODUL 3

1. Buat Repository Kelompok di Github dengan format nama : **[KODE MENTOR] – KELOMPOK X – GWE G6**  
contoh : **ALVR – KELOMPOK 1 – GWE G6**
2. Ganti Repository Visibilitynya jadi **Private**
3. Invite **alvaroocls**, **ahmfzui**, dan **mentor masing masing** sebagai collaborator
4. Push hasil pengerjaan ke repository kelompok sebelum **Jumat 20 Februari 23:59**



# Username Github Mentor

| Kode Mentor | Username Github |
|-------------|-----------------|
| ZIZI        | ahmfzui         |
| ALVR        | alvaroocls      |
| JHON        | AriAnanta       |
| VLSN        | vilsonnn        |
| MALE        | MatthewSitorus  |
| NDOW        | CuPid168        |
| MAAZ        | aliimron26      |
| EVIE        | dyahanggunhnp   |
| KING        | itsqal          |



Enterprise  
Data  
Management  
Laboratory

**BARCODE**

# ABSENSI

**EDM Laboratory**

**GWE #3**



# Presensi GWE #3

<https://tel-u.ac.id/presensigwe>



**GWE #3**

# Terima Kasih



**See U!**

**GWE #4**



@lifeatedmlab



@186mcgwc