

# Pengantar Data Analisis

Memahami dasar-dasar analisis data dan penerapannya untuk pengambilan keputusan berbasis data.

Jehan Afwazi Ahmad, S.Kom, M.Kom.

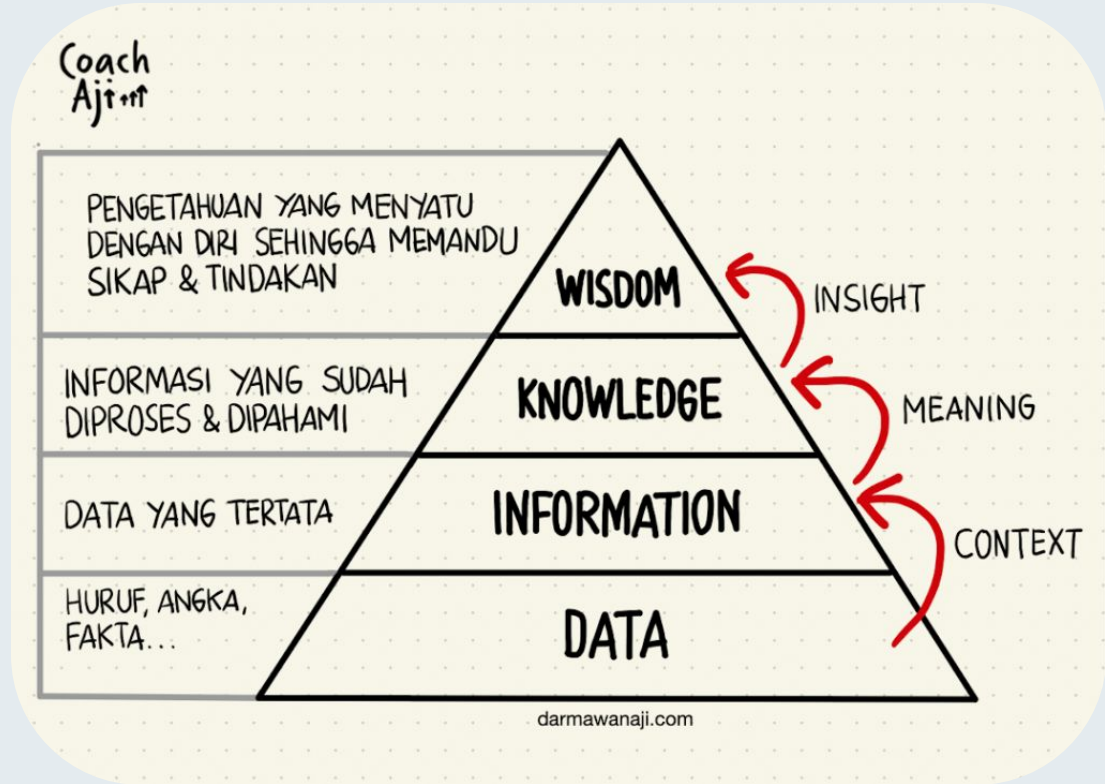
# Pendahuluan

Kita hidup di era ledakan data digital dari berbagai sumber seperti transaksi, media sosial, dan sensor IoT.

Data mentah tidak selalu bermanfaat tanpa proses pengolahan.

Analisis data adalah proses mengubah data menjadi informasi, wawasan, dan keputusan yang bermanfaat.

# Piramida Pengetahuan

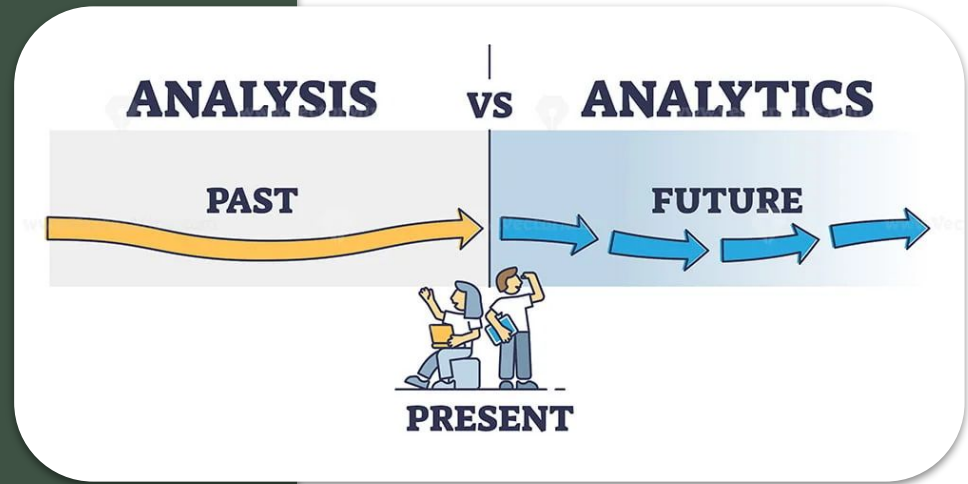



# Perbedaan Data Analysis dan Data Analytics

Data Analysis fokus pada proses pengolahan data seperti cleaning, transformasi, dan pemodelan.

Data Analytics mencakup keseluruhan proses analitik termasuk prediksi dan visualisasi.

Keduanya mendukung pengambilan keputusan berbasis data.



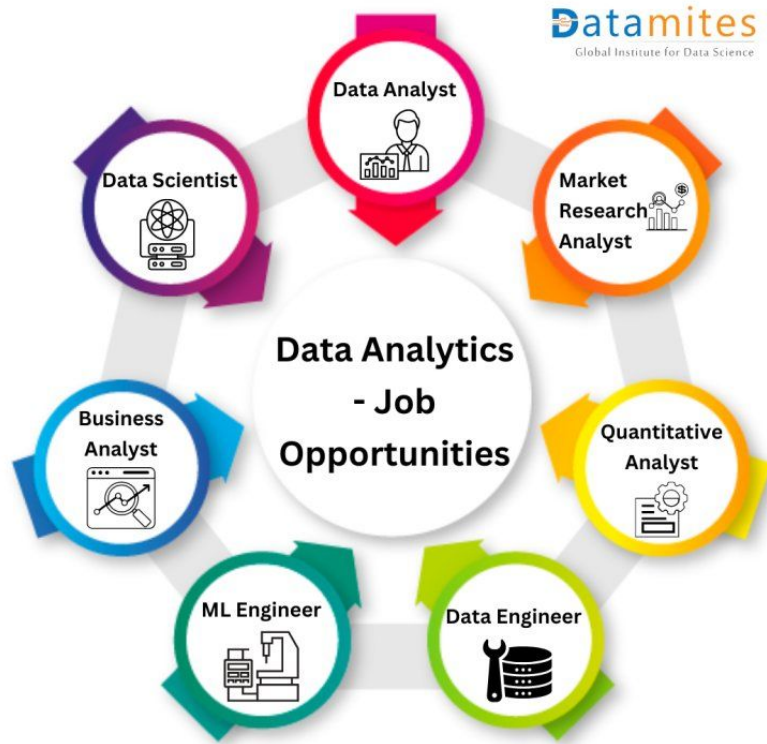


**Memahami apa yang terjadi  
dan kenapa terjadi.**

**Menentukan apa yang bisa  
dilakukan selanjutnya.**

**Contoh:**  
**Video Streamer: rekomendasi**  
**tontonan,**  
**Ecommerce: diskon otomatis.**

## **Kenapa Analisis Data Penting?**



# Peran Data Analyst

- Mengolah dan membersihkan data mentah
- Mengorganisasi dan menyajikan data agar mudah dipahami
- Menggunakan analisis statistik untuk menarik insight
- Mendukung pengambilan keputusan berdasarkan data

Contoh peran: analis keuangan, pemasaran, media, market, dll.

# Tahapan Proses Analisis Data (1)

## Langkah 1: Kumpulkan Data

- ❖ Ambil data dari berbagai sumber seperti file, API, atau database.
- ❖ Pastikan data relevan dan lengkap untuk analisis selanjutnya.



## Langkah 2: Bersihkan Data

- ❖ Tangani data kosong, format tanggal, dan duplikat.
- ❖ Data yang bersih penting untuk hasil analisis yang akurat.

## Tahapan Proses Analisis Data (2)



### Langkah 3: Eksplorasi Data (EDA)

- ❖ Temukan pola, ringkasan statistik, dan distribusi data.
- ❖ EDA membantu memahami karakteristik data sebelum analisis lanjutan

### Langkah 4 & 5: Analisis Lanjutan dan Insight

- ❖ Lakukan visualisasi, segmentasi, dan analisis statistik.
- ❖ Buat narasi dan rekomendasi berdasarkan insight untuk pengambilan keputusan



# Level-level Analitik

**Descriptive:** Apa yang terjadi?

→ Contoh: Ringkasan penjualan harian/mingguan

**Diagnostic:** Kenapa itu terjadi?

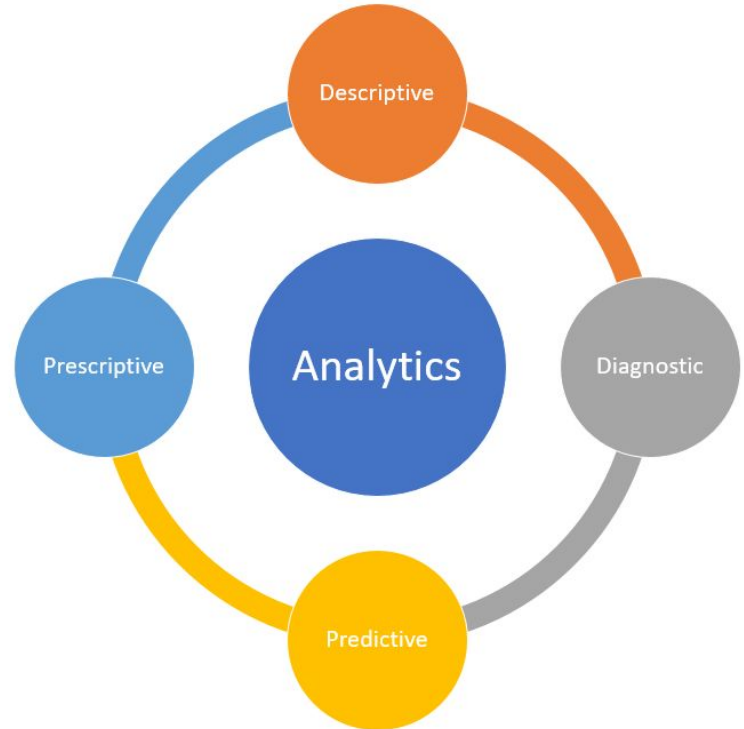
→ Contoh: Mengapa penjualan turun di bulan tertentu?

**Predictive:** Apa yang mungkin terjadi?

→ Contoh: Prediksi penjualan menggunakan machine learning

**Prescriptive:** Apa yang sebaiknya dilakukan?

→ Contoh: Menentukan strategi diskon otomatis berdasarkan perilaku konsumen



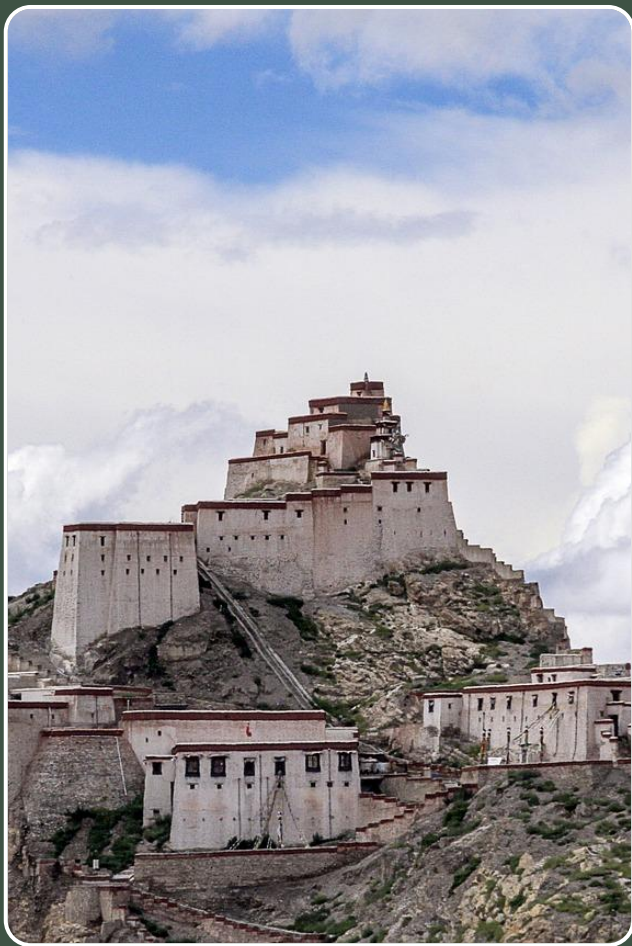
# Pengenalan Pandas



- Pandas adalah library Python yang berfungsi untuk manipulasi dan analisis data tabular.
- Mengapa Pandas?
  - Mudah digunakan & powerful.
  - Memiliki banyak fungsi siap pakai untuk filtering, grouping, merge, transformasi, dll.
- Dua struktur utama:
  - Series: 1D array (seperti 1 kolom Excel)
  - DataFrame: 2D table (seperti lembar kerja Excel)
- Pandas mampu membaca/menulis:
  - File CSV, Excel, Feather, JSON, SQL

Contoh

```
import pandas as pd
df = pd.read_csv("data.csv")
df.head()
```



# Basic Usage Pandas

```
import pandas as pd

# Membaca file
df = pd.read_csv("data.csv")

# Menjelajah isi
df.head(), df.info(), df.describe()

# Seleksi dan filter
df[df["Gender"] == "Female"]

# Agregasi
df.groupby("Country")["Amount"].sum()

# Penanganan missing value
df.dropna(), df.fillna(0)
```

# Sebelum Mulai Praktik

1. Menguasai dasar Python seperti struktur data dan logika
2. Pahami data tabular
3. Install library yang dibutuhkan
4. Siapkan editor kerja seperti Google Colab, Jupyter, atau VSCode



- **Python** (bahasa pemrograman utama)
- **Pandas** (manipulasi data)
- **Matplotlib / Seaborn** (visualisasi data)
- **Jupyter / Google Colab / VSCode** (notebook interaktif)
- **Excel & Feather dataset**
- **Google Drive** (penyimpanan cloud)
- Terminal untuk instalasi library (`pip install`)

**Tools yang  
Digunakan**

# Implementasi Tahapan Analisis

## # ✅ 1. KUMPULKAN DATA

```
# Membaca dataset dari file Excel yang berisi transaksi retail
import pandas as pd
df = pd.read_excel("Online Retail.xlsx")
```

## # ✅ 3. EDA (EXPLORATORY DATA ANALYSIS)

```
# Ringkasan statistik dari data numerik
df.describe()

# 10 produk terlaris berdasarkan quantity
top_products = df.groupby("Description")["Quantity"].sum()
                .sort_values(ascending=False).head(10)

# Negara dengan transaksi terbanyak
top_countries = df["Country"].value_counts().head(5)
```

## ✅ 5. Insight dan Narasi

- Produk paling laris adalah home decor.
- Transaksi tertinggi dari **United Kingdom**.
- Penjualan meningkat menjelang akhir tahun.

*"Penjualan tertinggi berasal dari item dekorasi rumah. Q4 mengalami peningkatan penjualan signifikan yang kemungkinan besar dipengaruhi oleh musim liburan."*

## # ✅ 2. BERSIHKAN & PERSIAPKAN DATA

```
# Melihat struktur awal dataset
df.info()

# Hapus baris yang memiliki nilai kosong pada kolom penting
df = df.dropna(subset=["CustomerID", "Description"])

# Ubah kolom 'InvoiceDate' menjadi format tanggal
df["InvoiceDate"] = pd.to_datetime(df["InvoiceDate"])

# Tambahkan kolom 'Total' untuk nilai transaksi
df["Total"] = df["Quantity"] * df["UnitPrice"]
```

## # ✅ 4. VISUALISASI DATA

```
import matplotlib.pyplot as plt

top_products.plot(kind="bar", figsize=(10, 5), title="Top 10 Produk Terlaris")
plt.ylabel("Jumlah Terjual")
plt.tight_layout()
plt.show()
```