PTSDM PPI DUNIA

START

DIGIHUB DATA SCIENCE

ADVANCED

AGENDA HARI INI

> 1. Pengenalan Data Science

> 2. Data Processing

> 3. Data Manipulation

> 4. Function

> 5. Error Handling

> 6. Sesi Tanya Jawab

PENGENALAN DATA SCIENCE

PERHATIKAN GAMBAR !

APA ITU DATA SCIENTIST ?

> Definisi Data Scientist

Data Scientist (Ilmuwan Data) terdiri dari 2 kata, yaitu Data dan Scientist. Data adalah ahl - hal yang abstrak (angka, simbol, gambar, suara) yang terkumpul lalu diolah oleh Scientist (Ilmuwan) untuk menghasilkan informasi yang bermanfaat dalam pengambilan keputusan. Jadi, seorang Data Scientist adalah orang yang mengumpulkan, mengolah, dan memberikan informasi berdasarkan data

APA JOB DESC DATA SCIENTIST ?

> Deskripsi pekerjaan seorang Data Scientist

1 MEMAHAMI DATA

data dapat dipahami dengan memperhatikan apakah datanya lengkap, apakah datanya tersedia dan cukup atau tidak, dan mampu memanipulasi data agar siap digunakan.

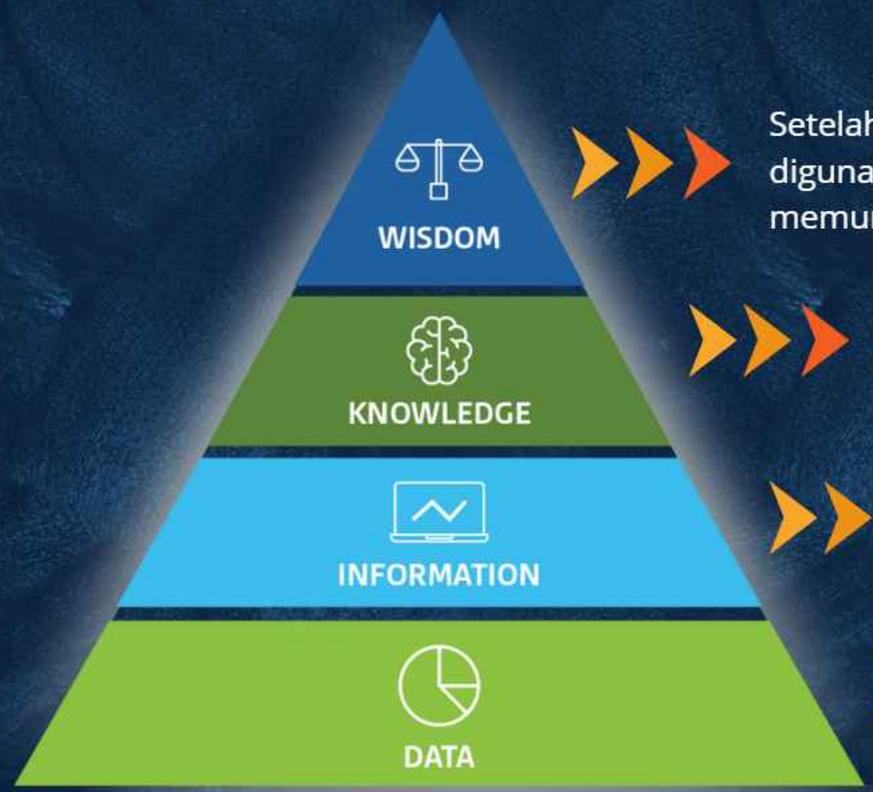
2 EXPLAINABLE INSIGHT

Setelah data siap digunakan, maka akan diubah menjadi suatu informasi / insight yang berguna untuk tim produksi mengeksekusi keputusan berdasar data.

3 MELAKUKAN MODELLING

Lalu melakukan prediksi dengan menyusun sebuah model algoritma yang berguna untuk memprediksi informasi masa depan dari data yang ada.

DIKW HIERARCHY

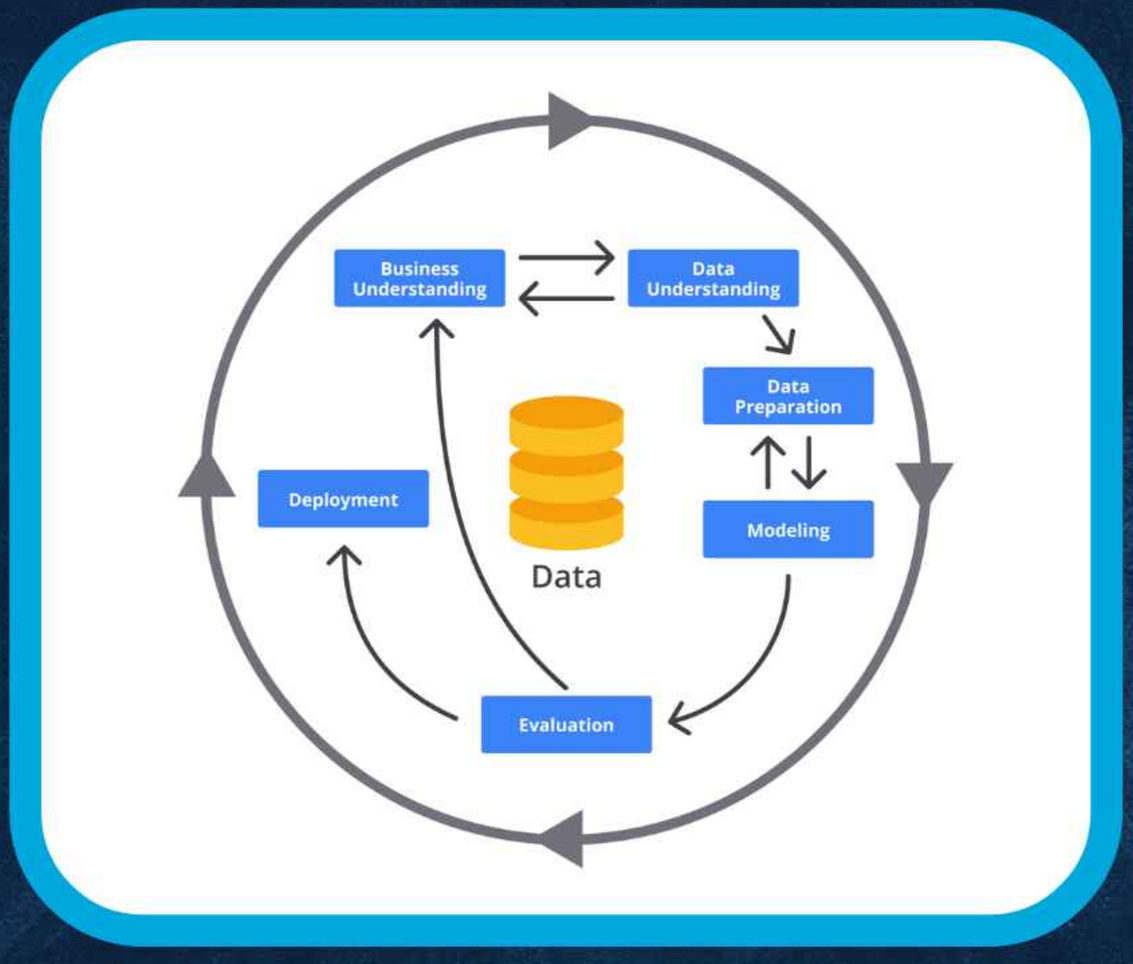


Setelah tahu knowledge / pemahaman yang dipelajari, maka digunakan untuk memprediksi masa depan sehingga memunculkan kebijaksanaan / wisdom.

Informasi yang disusun lalu digabungkan sehingga memiliki makna dan manfaat.

Data yang tersusun dari data bstrak menjadi data yang bisa dapat dibaca disebut informasi.

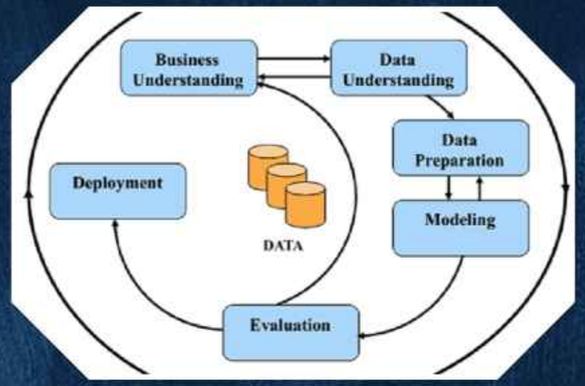
Data yang abstrak, tidak dapat dibaca, dan tidak ada konteks dari data. Dapat berupa pernyataan, simbol, angka, gambar, dll



CRISP - DM

CRoss Industry Standard Process for Data Mining

CRISP - DM



> Merupakan suatu Guideline / Kerangka berpikir untuk mengerjakan proyek secara terukur.

BUSINESS UNDERSTANDING

Memahami apa yang menjadi kebutuhan bisnis, apa yang ingin dicapai? dengan memahami masalah bisnis, akan mempermudah kita untuk mencari solusinya.

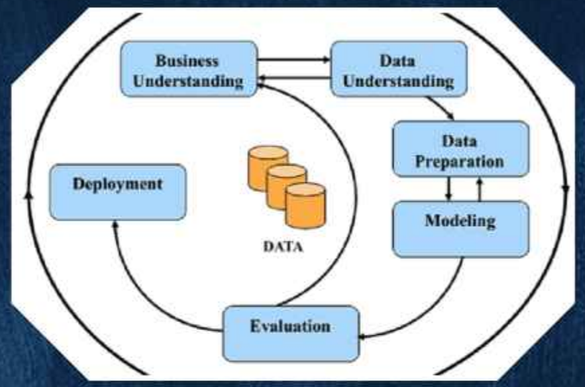
DATA UNDERSTANDING

Setelah memahami masalah, kita akan mulai mencari data apa saja yang dibutuhkan dari suatu pertanyaan bisnis. Jika datanya tidak lengkap, maka balik ke business understanding.

DATA PREPARATION

Tahap setelah data yang dibuthkan ada. Data Preparation adalah proses menyiapkan data untuk bisa divisualisasikan atau dibuat prediksi sesuai dengan kebutuhan.

CRISP - DM



> Merupakan suatu Guideline / Kerangka berpikir untuk mengerjakan proyek secara terukur.

MODELLING

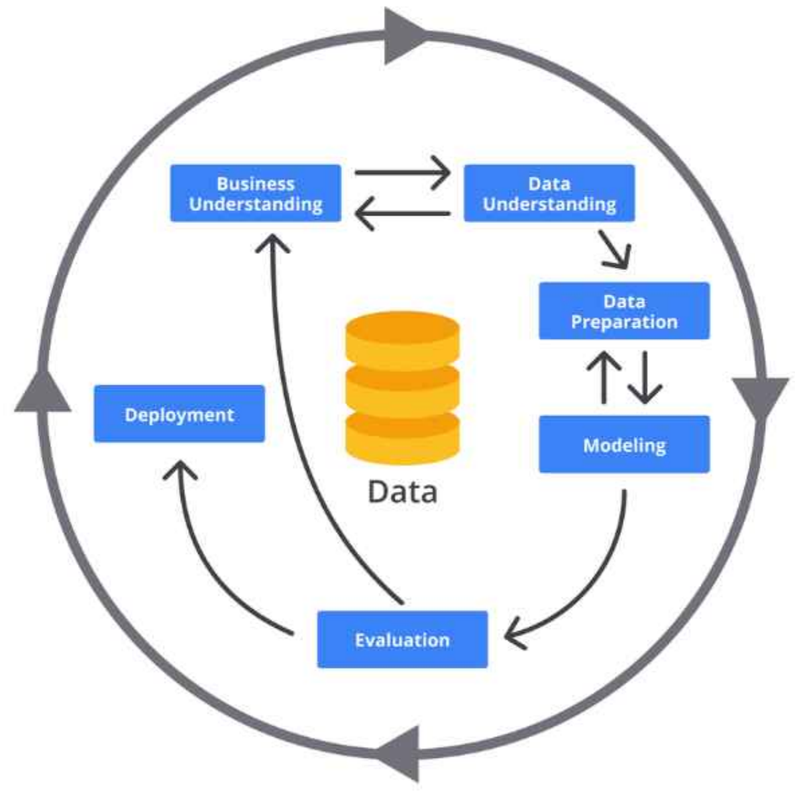
Modelling adalah proses untuk menciptakan / memilih suatu model prediksi. Ketika model yang dibuat membutuhkan data yang lebih banyak lagi, maka balik ke data preparation.

EVALUATION

Evaluation terjadi ketika tim bisnis ingin melakukan settinggan pada model. Misal, Tim bisnis ingin modelnya memprediksi hanya dalam waktu 1 detik, bukan 5 menit. Maka balik ke Busniness Undertsnding

DEPLOYMENT

Deployment berbentuk dashboard / aplikasi yang bisa diakses oleh Business User atau Stakeholders untuk mengakses model hasil Evalution.



CRISP - DM

Tidak berhenti dampai di Deployement. Seorang Data Scientist harus bersiap siaga ketika user meminta untuk update model ketika ada tambahan fitur baru atau tambahan permintaan dari user dalam model. Hal ini untuk keberlangsungan model yang bermanfaat pada bisnis

> Guido Van Rossum

GET TO KNOW

> Python (Since 1991)

Python adalah salah satu bahasa pemrograman yang mudah dipahami oleh manusia (**High - Level Programming Language**) sehingga banyak digunakan oleh Data Scientist untuk mengolah data dan membuat prediksi berdasarkan hasil modelling.

Python memiliki banyak built-in atau packages yang lengkap untuk mengolah data dan komunitas yang besar secara Internasional. Maka dari itu, ketika mengalami error atau kesulitan untuk running kode, banyak sumber yang menyediakan solusi, salah satunya adalah **Stack Overflow**.

Apa maksud dari Python yang merupakan High - Level Programming Language ?

HIGH LEVEL LANGUAGE

High Level vs Low Level Languages

High Level Languages

Python

C++

JavaScript

Java

Low Level Languages

Assembly Language

```
1 lea di, buffer
2 mov dx, size
3 putc 13d
```

C++ "HelloWorld"

```
#include <iostream>
int main() {
    cout<<"Hello World";
    return 0;
}
```

JAVA "Hello World"

```
class HelloWorld
{
 public static void main(String[] args) {
   system.out.println("Hello World");
  }
}
```

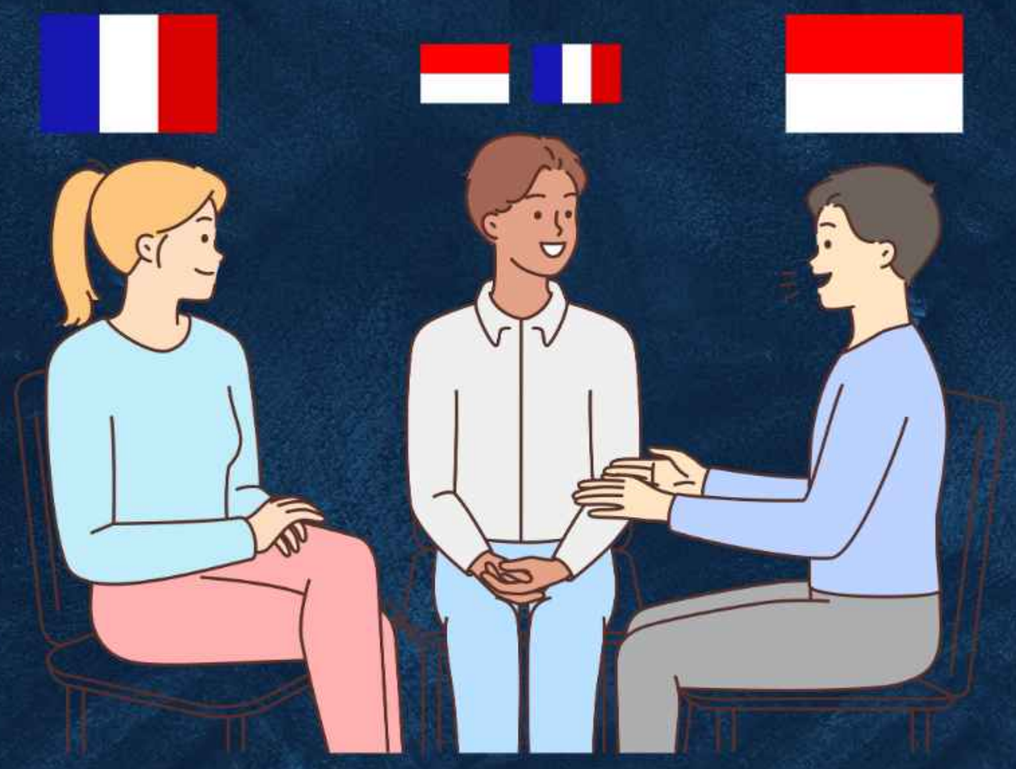
Python "Hello World"

```
print("Hello World")
```

High - Level programming language adalah bahasa pemrograman yang mudah dimengerti dan ditulis oleh manusia karena susunan sintaksnya yang mirip dengan bahasa alami (bahasa inggris), sedangkan Low - Level programming language adalah bahasa pemrograman yang langsung menggunakan bahasa mesin.

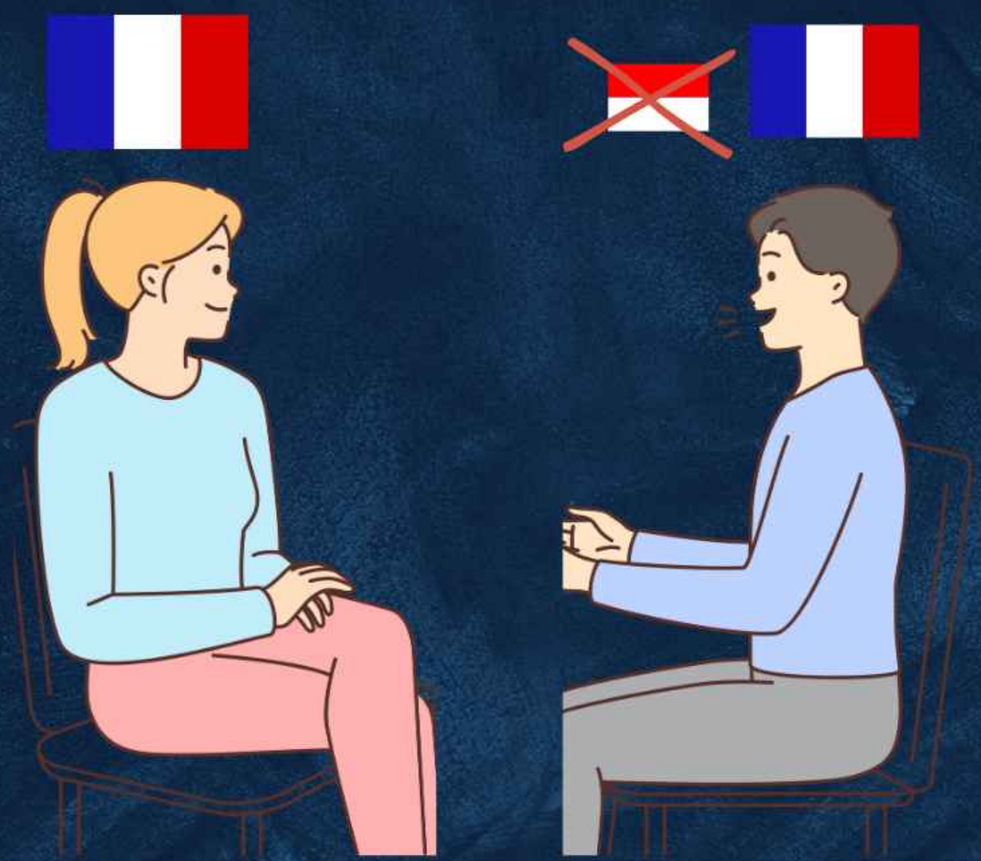
HIGH LEVEL

> Analogi High - Level Programming



Misal, Budi yang hanya menguasai Bahasa Indonesia ingin menyuruh Maeva yang hanya bisa berbahasa Prancis untuk **menelepon seseorang di telepon genggamnya**. Bahasa high - level adalah translatornya. Translator ini akan memberikan isntruksi ke Maeva bagaimana caranya menelepon seseorang.

Budi tidak harus tahu bagaimana cara menelepon seseorang di telepon genggam Maeve karena Translator yang mengurus detailnya. Budi hanya memberikan perintah umum.



LOW LEVEL

> Analogi Low - Level Programming

Dalam kasus yang sama, Budi **harus menguasai bahasa Prancis** untuk menyuruh Maeve menelepon seseorang dengan **memberikan instruksi detail** tanpa bantuan penerjemah, inilah yang disebut Low - Level Programming Language.

Instruksi dalam bahasa Prancis harus sesuai dengan struktur bahasa (grammar) yang benar, kosakata yang tepat, dan kemampuan berbicara yang jelas. Jika salah sedikit, Maeve akan mengira untuk melakukan tindakan lain selain menelepon seseorang.

GOOGLE COLAB

> Cloud Storage

Google Colab merupakan salah satu jenis Notebook yang memiliki keunikan tersendiri, yaitu terintegrasi dengan produk Google, instan dan GRATIS. Karena Google Colab terintegrasi dengan produk google, maka jenis penyimpanannya adalah penyimpanan awan (**Cloud Storage**).

Google Colab tersedia secara online sehingga mempermudah Data Scientist untuk bekerja sama dengan rekan tim untuk mengembangkan kode dalam mengolah data. Google Colab tidak memperhitungkan spesifikasi, ram, maupun memori dari laptop / komputer lokal karena Google Colab hanya mempertimbangan spesifikasi dari Cloud Storage saja.

DETAIL MATERI LENGKAP

> Silahkan mengakses playlist Youtube Agung Krishna untuk Beginner Level berikut

SCAN ME!

DATA PROCESSING

VARIABLE

Variabel dan Data Types wajib diketahui untuk memahami sebuah kode, tidak hanya pada bahasa pemrograman Python, tetapi semua bahasa pemrograman memiliki variable dan tipe data. Variabel adalah simbol berupa huruf atau angka untuk mereferensikan / assign suatu object.

#Assign Object ke Variable

```
X = 1990
y = "Hello World"
z = 3.5
```

#Print Variable

```
Print(X)
```

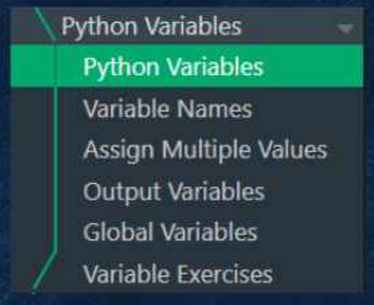
#Penamaan tidak bisa didahului dengan angka

```
1X = 1990 ✗
X1 = 1990 ✓
```

#Penamaan tidak bisa dengan spasi

```
X y = 1990 ✗
X, y = 1990, "Hello World" ✓
```

MORE INFO



w3schools.com

DATA TYPES

Variabel dan Data Types wajib diketahui untuk memahami sebuah kode, tidak hanya pada bahasa pemrograman Python, tetapi semua bahasa pemrograman memiliki variable dan tipe data. Variable bisa berisi tipe data berbeda, dan tipe data tersebut memiliki fungsi yang berbeda - beda.

String

```
nama = “John Doe”
angka = “52.5”
```

Integer

```
lima_puluh_dua = 52
satu = 1
```

Float

```
satu_koma_lima = 1.5
```

Boolean

```
Betul = True
komparasi = satu == 1
```

MORE INFO

Text Type:	str
Numeric Types:	int, float, complex
Sequence Types:	list, tuple, range
Mapping Type:	dict
Set Types:	set, frozenset
Boolean Type:	bool
Binary Types:	bytes, bytearray, memoryview
None Type:	NoneType

w3schools.com

DATA TYPES

- List : [ ], Mutable, Nested
- Tuple : ( ), Immutable, Nested
- Set : { }, tidak menyimpan nilai duplicate, seperti set di matematika
- Dictionary : { }, pasangan key dan values

List

```
data = ['John', 2]
data = [2.5, True]
```

Tuple

```
data = ('John', 2)
data = (2.5, True)
```

Set

```
data = {'John', 2}
data = {1,2,1,2}
```

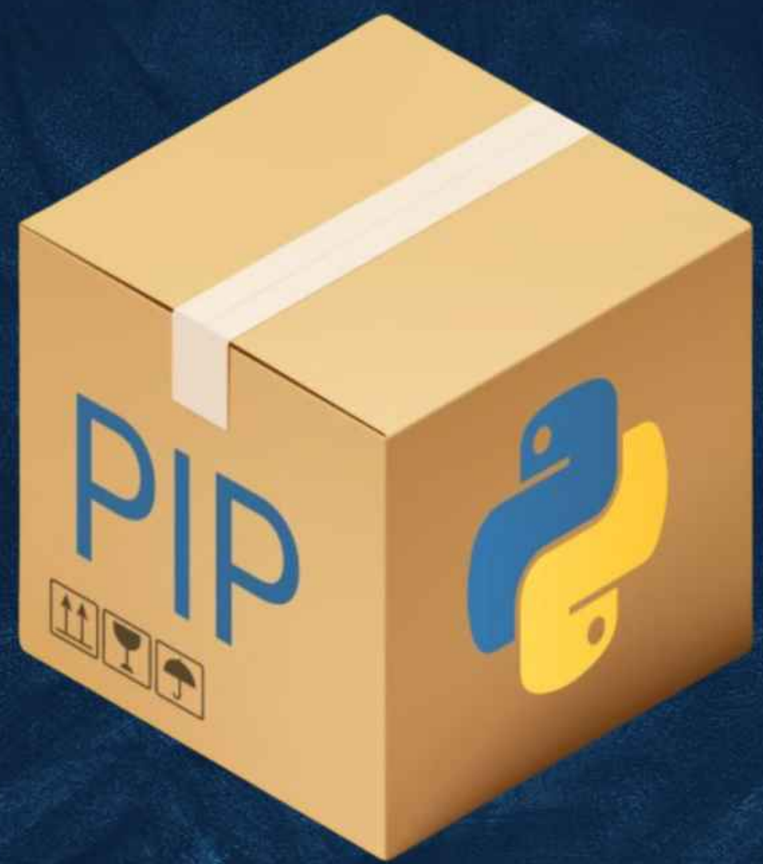
Dictionary

```
data = {'nama':'John',
        'height':180,
        'Suhu':35.4}
```

Text Type:	str
Numeric Types:	int, float, complex
Sequence Types:	list, tuple, range
Mapping Type:	dict
Set Types:	set, frozenset
Boolean Type:	bool
Binary Types:	bytes, bytearray, memoryview
None Type:	NoneType

w3schools.com

YUK PRAKTEK !



PACKAGE

> Package / Library / Tools

Package adalah kumpulan function function yang telah dibuat oleh ilmuwan luar untuk mengerjakan tugas / kode yang sangat panjang menjadi lebih singkat. Dalam satu baris kode, bisa memunculkan banyak output sekaligus.

PANDAS

Pandas berasal dari istilah “Panel Data” yaitu kumpulan data multidimensional yang sering digunakan untuk memuat, membersihkan, menganalisis, dan manipulasi data. Salah satu struktur Pandas adalah DataFrame. DataFrame adalah kumpulan data 2 dimensi dalam bentuk frame (baris dan kolom) seperti tabel yang bisa dianalisis.

METODE

> Metode Data Processing yang akan kita pelajari

- **import data** : Mengimpor data ke dalam Python, biasanya menggunakan **pandas** dengan format seperti CSV, Excel, atau database SQL.
- **shape** : Menampilkan jumlah baris dan kolom dalam DataFrame.
- **info** : Menampilkan ringkasan informasi tentang DataFrame, seperti jumlah non-null, tipe data, dan memori yang digunakan.
- **describe** : Memberikan statistik ringkasan untuk kolom numerik dalam DataFrame.
- **head** : Menampilkan beberapa baris pertama dari DataFrame.

YUK PRAKTEK !

METODE

> Metode Data Processing yang akan kita pelajari

- **tail :** Menampilkan beberapa baris terakhir dari DataFrame.
- **sum, mean, min, max :** Melakukan operasi agregasi seperti penjumlahan, rata-rata, nilai minimum, dan maksimum pada DataFrame.
- **Mengubah tipe data :** Mengubah tipe data suatu kolom, misalnya dari string ke integer atau sebaliknya.
- **Filtering Column :** Memilih kolom tertentu dari DataFrame.
- **Add column :** Menambahkan kolom baru ke dalam DataFrame.

YUK PRAKTEK !

METODE

> Metode Data Processing yang akan kita pelajari

- **Dropping Columns :** Menghapus kolom tertentu dalam DataFrame.
- **Duplicated Columns :** Mendeteksi dan menghapus data yang duplikat dalam DataFrame.
- **Mengurutkan DataFrame :** Mengurutkan DataFrame berdasarkan satu atau beberapa kolom.
- **Checking Null Value :** Mengecek nilai yang kosong (missing values) dalam DataFrame.
- **Export DataFrame :** Mengekspor DataFrame ke dalam berbagai format seperti CSV atau Excel.

YUK PRAKTEK !

BERSAMBUNG BESOK ...

SESI TANYA JAWAB

SESI FOTO BERSAMA

End

THANK YOU

We are excited about the future of Data Sceince and OISAA's role in shaping it. Let's work together to build solutions that drive innovation and positive change.

Welcome to our course and have a great day !

@ppidunia @ptsdm.ppidunia