

# Dolanan Data #3 Preprocessing Data: Excel vs Python

Webinar Series by :

 **DATA SCIENCE**  
ITTELKOM SURABAYA

 **IT Telkom Surabaya**  
Solution for The Nation

 **Yayasan Pendidikan Telkom**

PRESENT

 **Dolanan Data**  
Preprocessing Data:  
Excel vs Python

SABTU, 28 AGUSTUS 2021  
08.30 WIB

LIVE ON  
ZOOM MEETING

REGRISTRASI NOW  
[bit.ly/DolananData3](https://bit.ly/DolananData3)

Narasumber:  
**CAK MAW**  
Dosen Program Studi S1 Sains Data  
Institut Teknologi Telkom Surabaya

**FREE**  
\* Benefit Pengalaman dan ilmu bermanfaat

 [ds.itelkomsby](https://www.instagram.com/ds.itelkomsby)  [ds@ittelkom-sby.ac.id](mailto:ds@ittelkom-sby.ac.id)  <https://ds.ittelkom-sby.ac.id>

 **DATA SCIENCE**  
ITTELKOM SURABAYA

 **IT Telkom Surabaya**  
Solution for The Nation

 **Yayasan Pendidikan Telkom**

PRESENT

 **Dolanan Data**  
Pengolahan Data Sederhana  
Dengan Python

SABTU, 21 AGUSTUS 2021  
08.30 WIB

LIVE ON  
ZOOM MEETING

REGRISTRASI NOW  
[bit.ly/DolananData2](https://bit.ly/DolananData2)

Narasumber:  
**CAK SENO**  
Kepala Program Studi S1 Sains Data  
Institut Teknologi Telkom Surabaya

**FREE**  
\* Benefit Pengalaman dan ilmu bermanfaat

 [ds.itelkomsby](https://www.instagram.com/ds.itelkomsby)  [ds@ittelkom-sby.ac.id](mailto:ds@ittelkom-sby.ac.id)  <https://ds.ittelkom-sby.ac.id>

**FREE**  
\* Benefit Pengalaman dan ilmu bermanfaat

Narasumber:  
**CAK OBY**  
Dosen Program Studi S1 Sains Data  
Institut Teknologi Telkom Surabaya

 [ds.itelkomsby](https://www.instagram.com/ds.itelkomsby)  [ds@ittelkom-sby.ac.id](mailto:ds@ittelkom-sby.ac.id)  <https://ds.ittelkom-sby.ac.id>

# Meet Me

## Mawanda Almuhayar

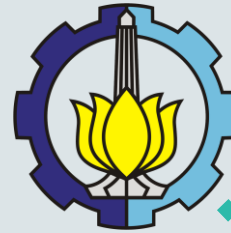
 Prodi Data Science, ITTelkom SBY

 Surabaya

 mawandaal@ittelkom-sby.ac.id

 mawandaal

 wanderational



S1 Statistika (Ekonometrika)  
S2 Statistika (Komputasi Stat.)



S2 Statistika (Image)  
(Double Degree)



Dosen Prodi  
Data Science





# Outline

## Introduction

- Data Science Hierarchy of Needs
- Data Science Life Cycle
- Python vs Excel

## Coding

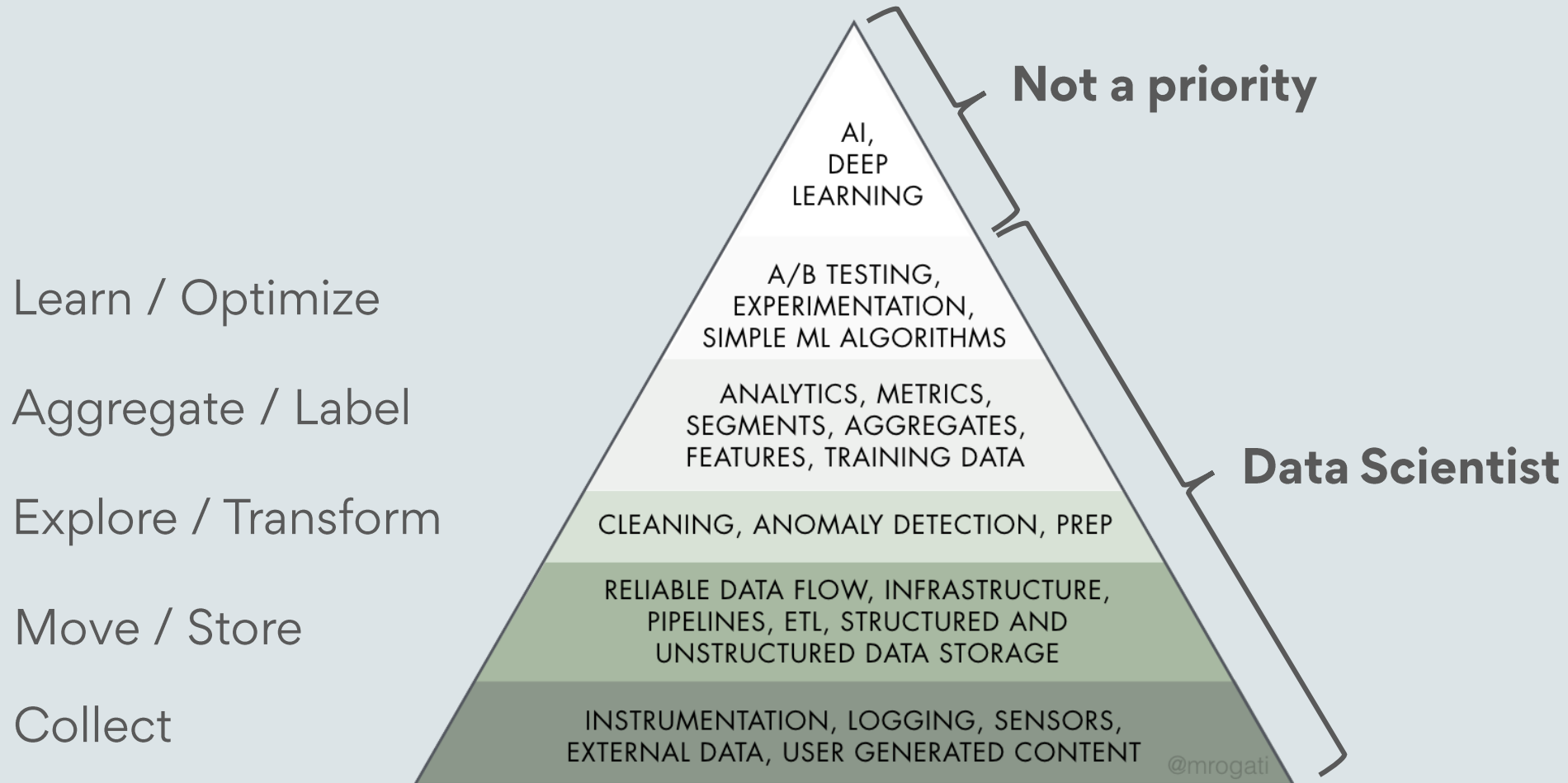


# Introduction

# Data Science Hierarchy of Needs

Hirarki Kebutuhan Ilmu Sains Data

Startup Companies

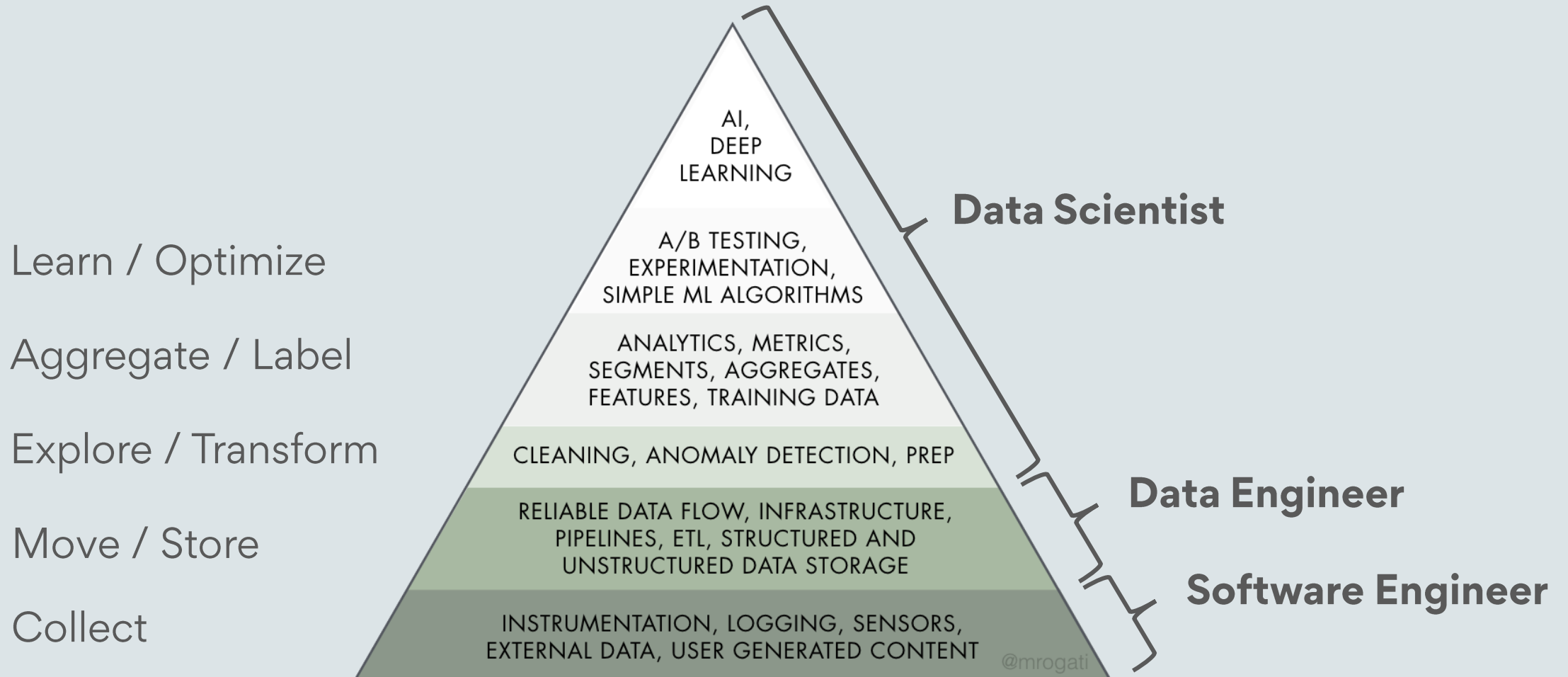




# Data Science Hierarchy of Needs (2)

Hirarki Kebutuhan Ilmu Sains Data (2)

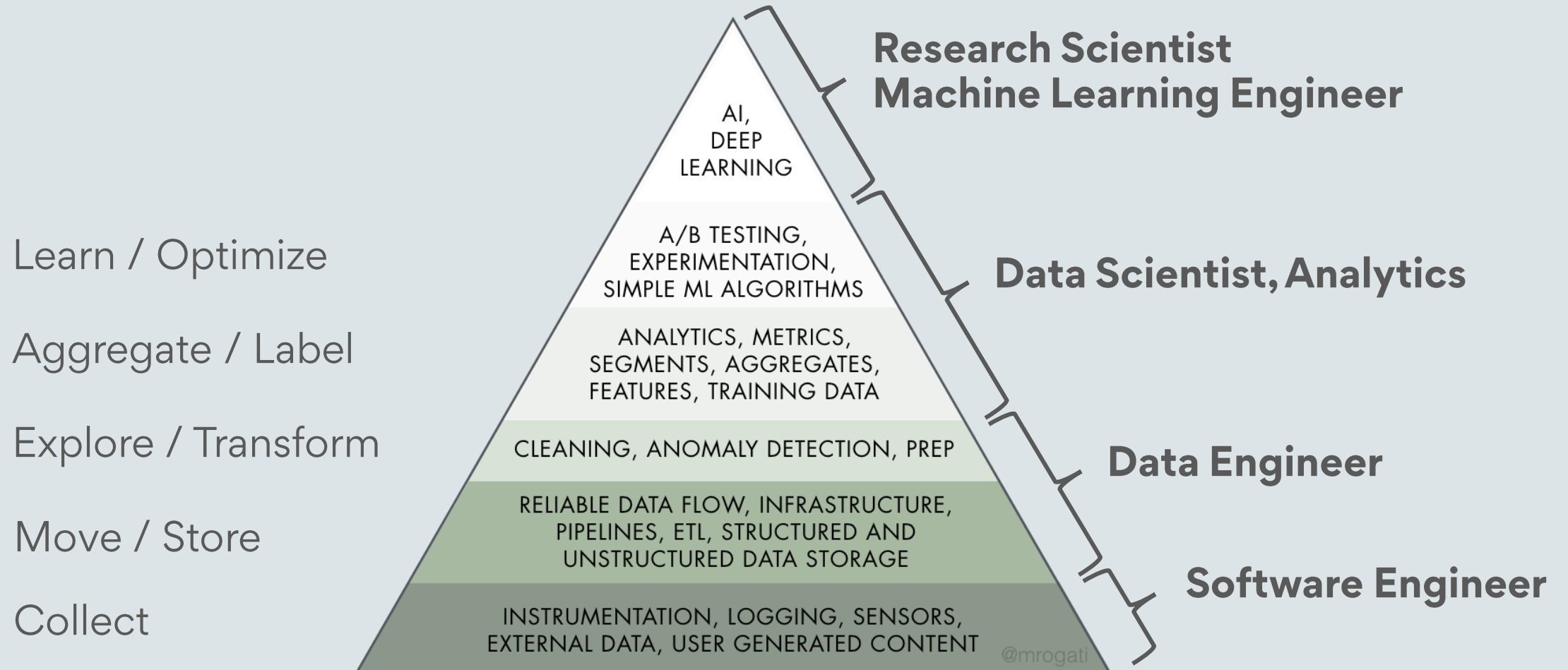
Medium-sized Companies



# Data Science Hierarchy of Needs (3)

Hirarki Kebutuhan Ilmu Sains Data (3)

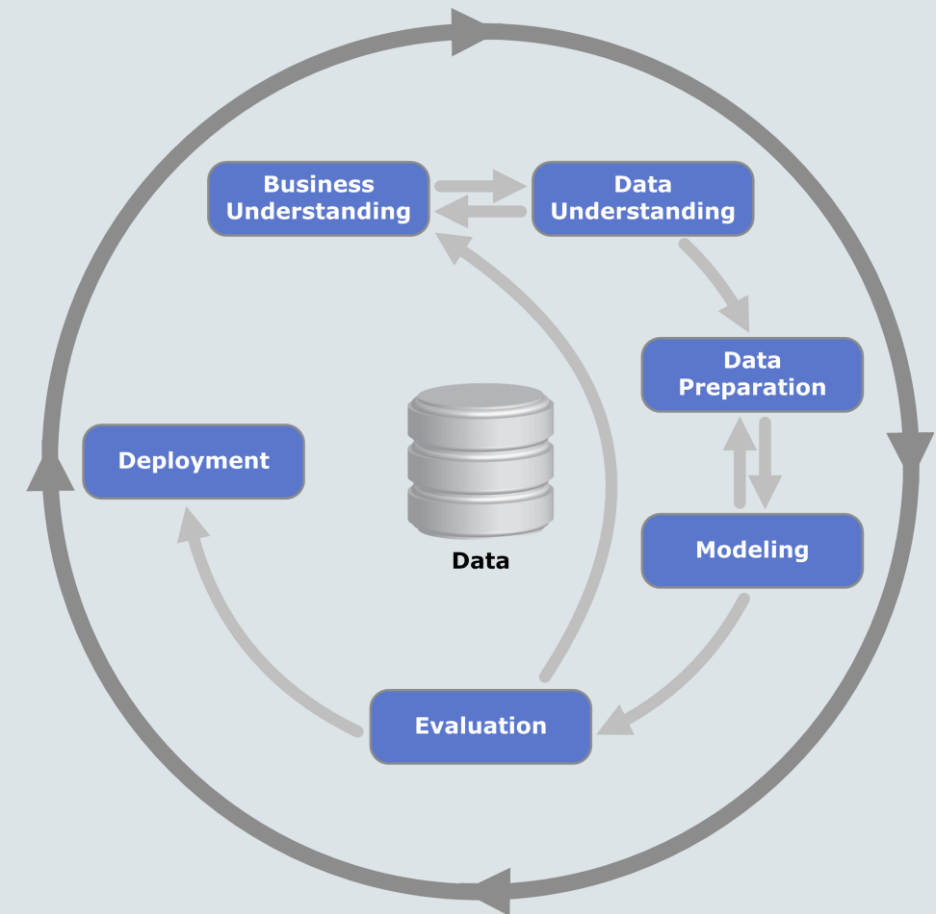
Large Companies



# Data Science Life Cycle/Process

Proses/Siklus Hidup Ilmu Sains Data

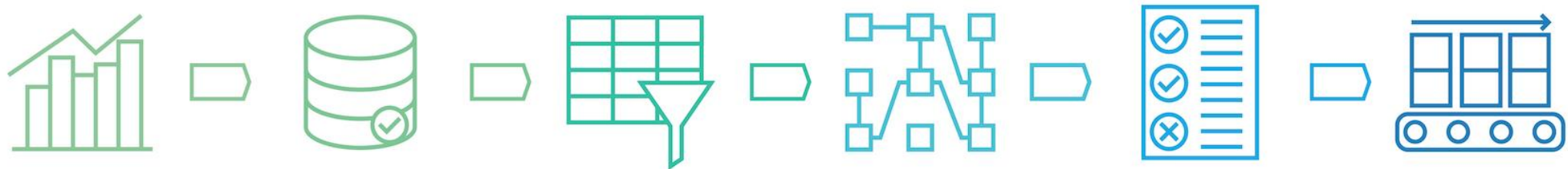
- **Business/Problem Understanding**  
What does the business/problem need?
- **Data Understanding**  
What data do we have / need? Is it clean?
- **Data preparation**  
How do we organize the data for modeling?
- **Modeling**  
What modeling techniques should we apply?
- **Evaluation**  
Which model best meets the business objectives?
- **Deployment**  
How do stakeholders access the results?





# Data Science Life Cycle/Process (2)

Proses/Siklus Hidup Ilmu Sains Data (2)

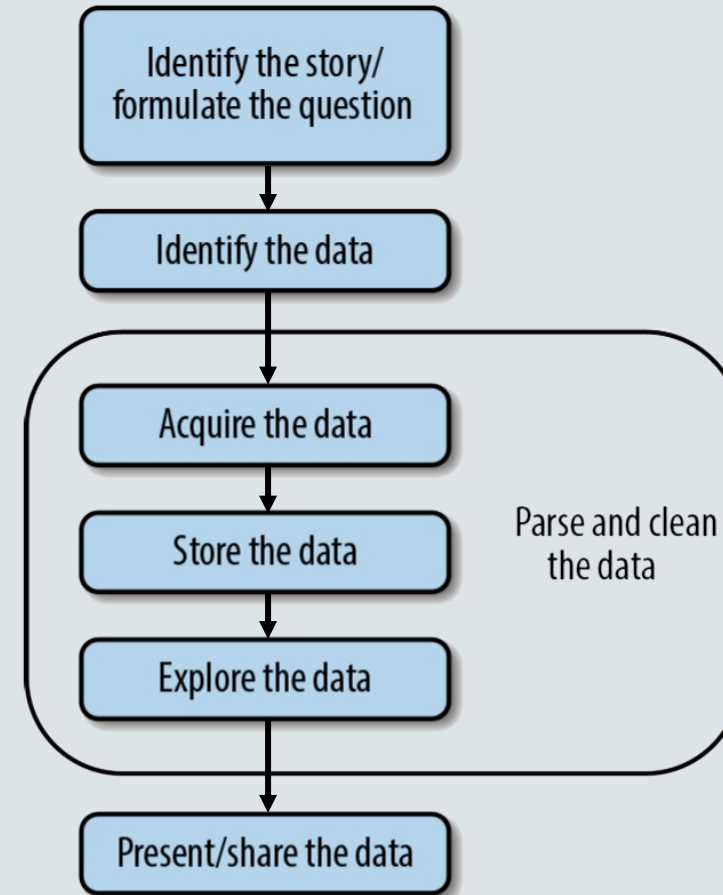


Business/ Problem Understanding	Data Understanding	Data Preparation	Modeling	Evaluation	Deployment
....	....	....	....	....	....
....	....	....	....	....	....
....	....	....	....	....	....
....	....	....	....	....	....
....	....	....	....	....	....

# Data Handling Process

Proses Penanganan Data

- Data Preparation
- Data Preprocessing
- Data Wrangling
- Data Munging
- Data Cleaning
- Data Exploration
- Exploratory Data Analysis



**50% of  
all process!**



# Excel: Advantage and Disadvantage

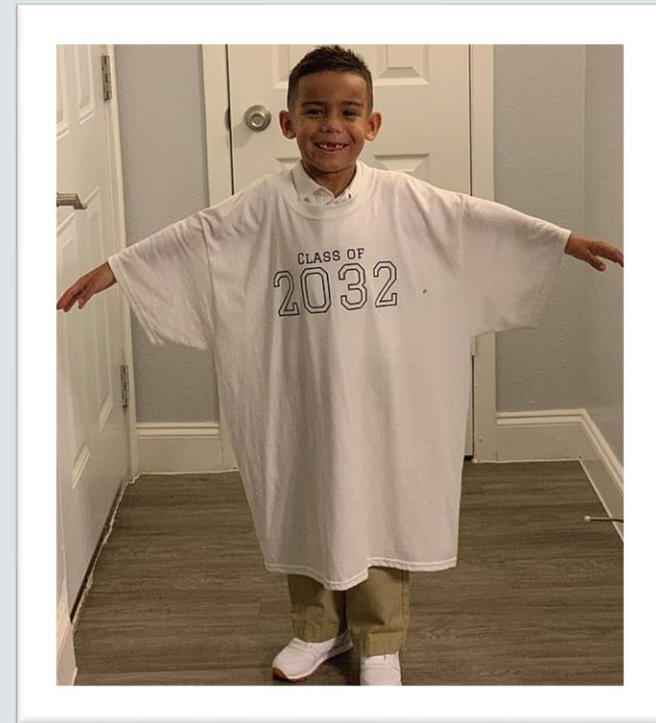
Kelebihan dan Kekurangan

## Pros:

- Widely known
- Easy to use because of GUI
- Built-In functions

## Cons:

- Not free
- Many unused features
- Not integrated to programming language





# Python: Advantage and Disadvantage

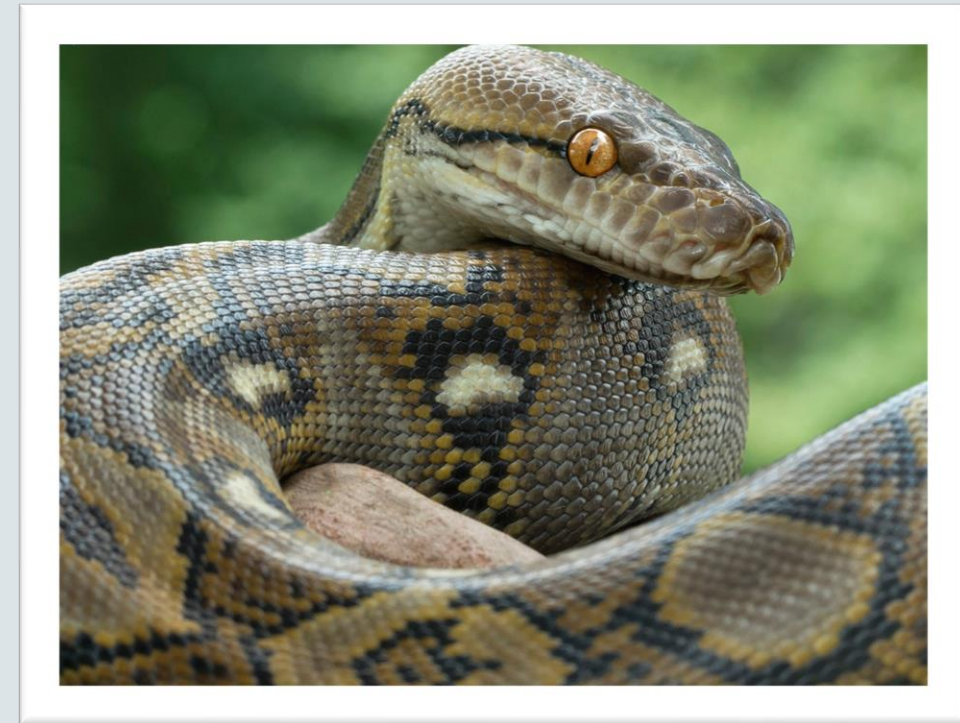
Kelebihan dan Kekurangan

## Pros:

- Free and open source
- High-level programming language
- Extensive support libraries

## Cons:

- You have to learn to code
- Speed limitations
- ...







# Pandas: Advantage and Disadvantage

Kelebihan dan Kekurangan

## Pros:

- Efficient handling of huge data
- Seamless data analysis workflow
- Easy collaboration with other tools

## Cons:

- You have to learn to code in Python
- Speed limitations
- ...





# Case Study

Studi Kasus

- **Clean the data**

Handle missing value

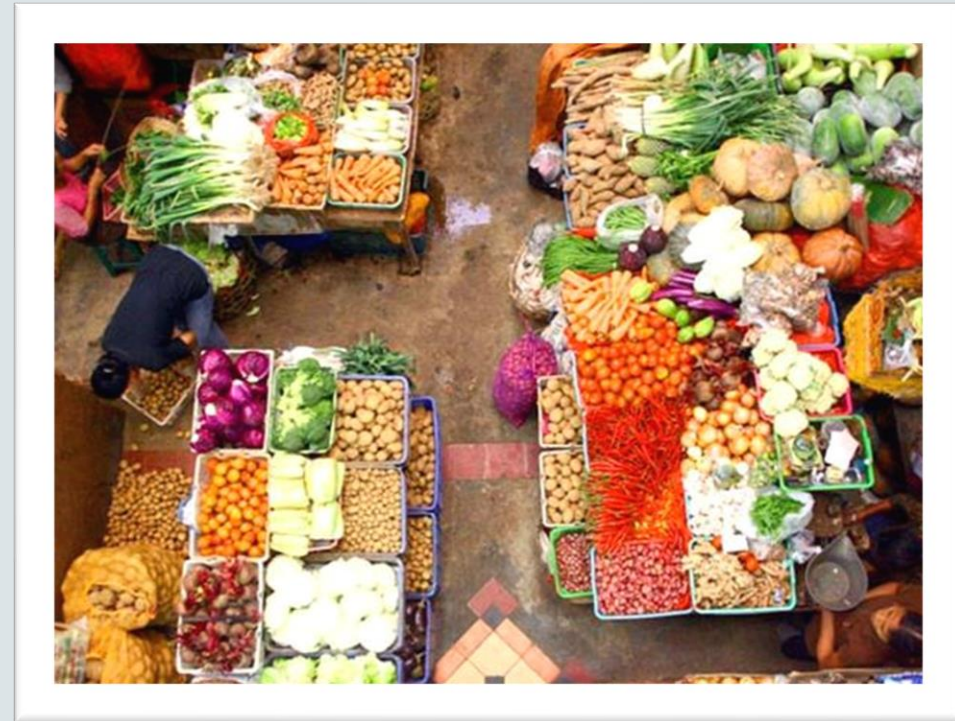
Check data type

Transform the dataframe

- **Visualize the data with the plot**

Link to the data and notebook:

<https://github.com/mawandaal/dolanan-data>



**Let's Code**