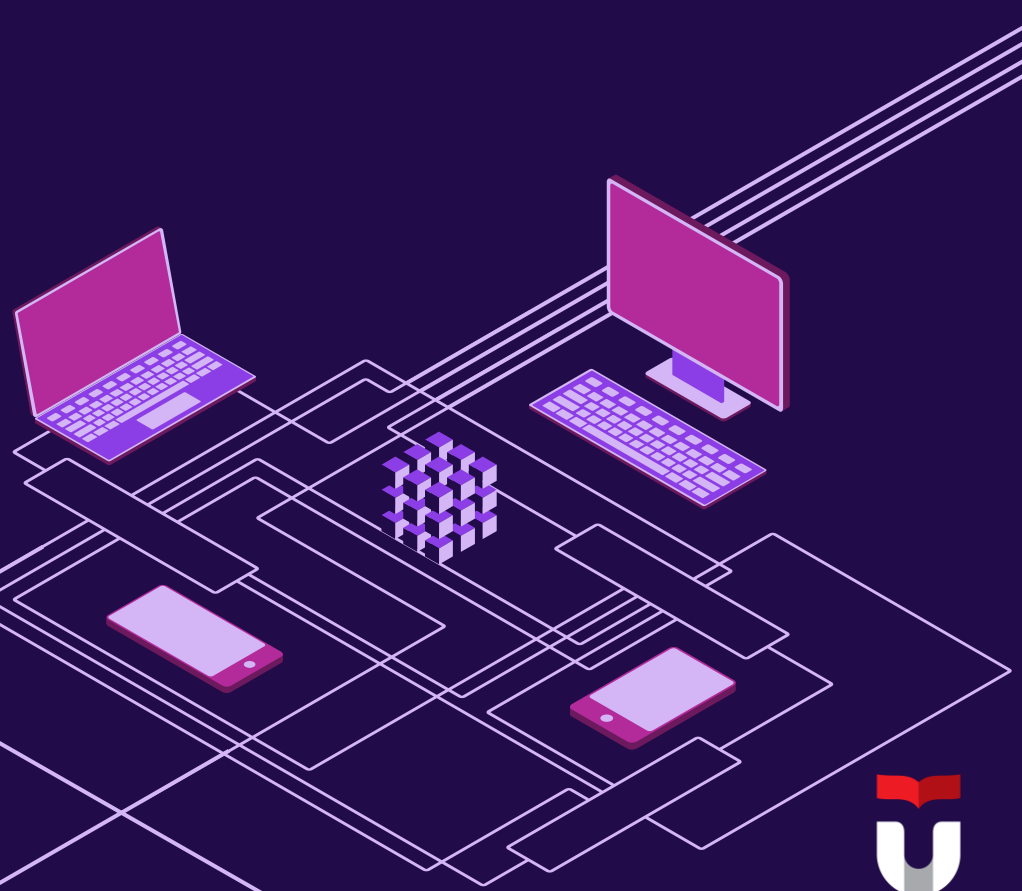
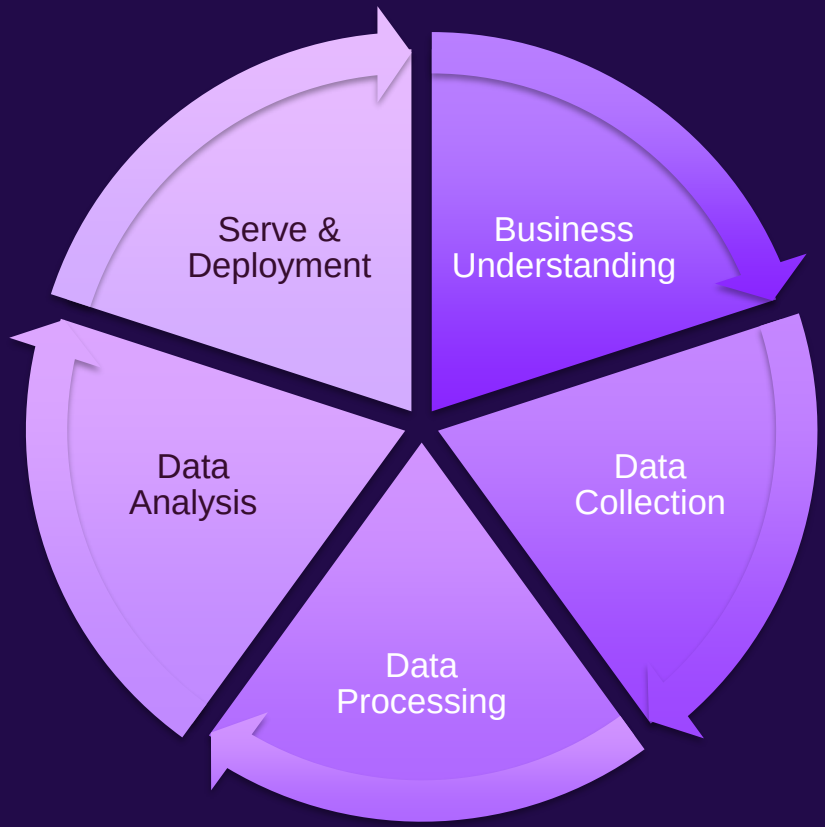
HOW

DO WE DO DATA ANALYTICS?

ADITYA FIRMAN IHSAN



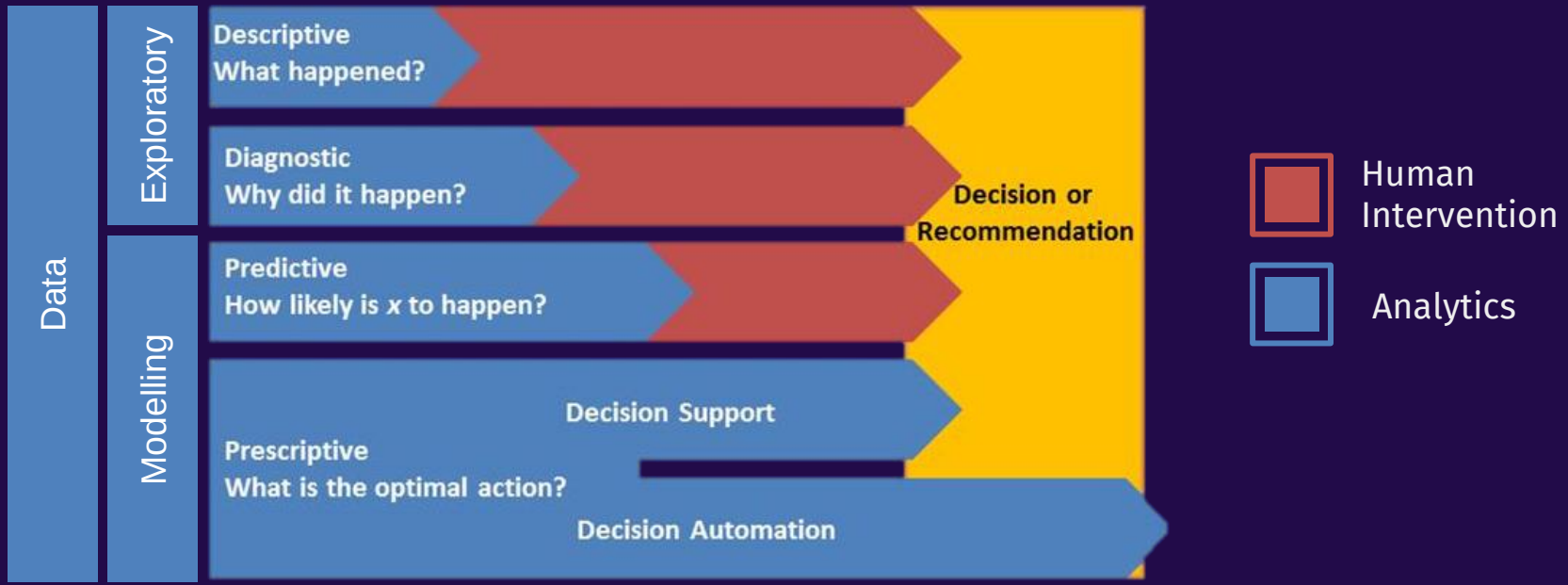
Telkom University



GENERAL STEP OF DA

It is a cyclic process

4 TYPES OF DATA ANALYSIS



DATA ANALYTICS TASKS

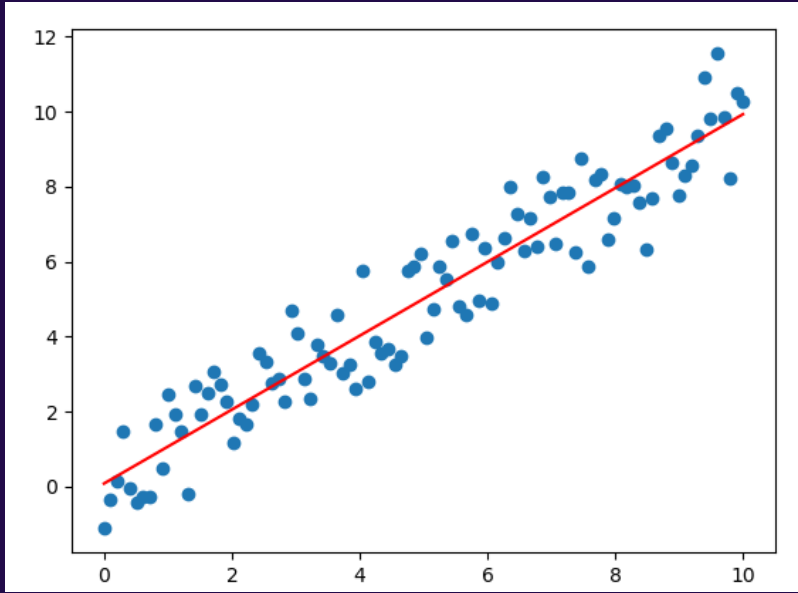
Exploratory

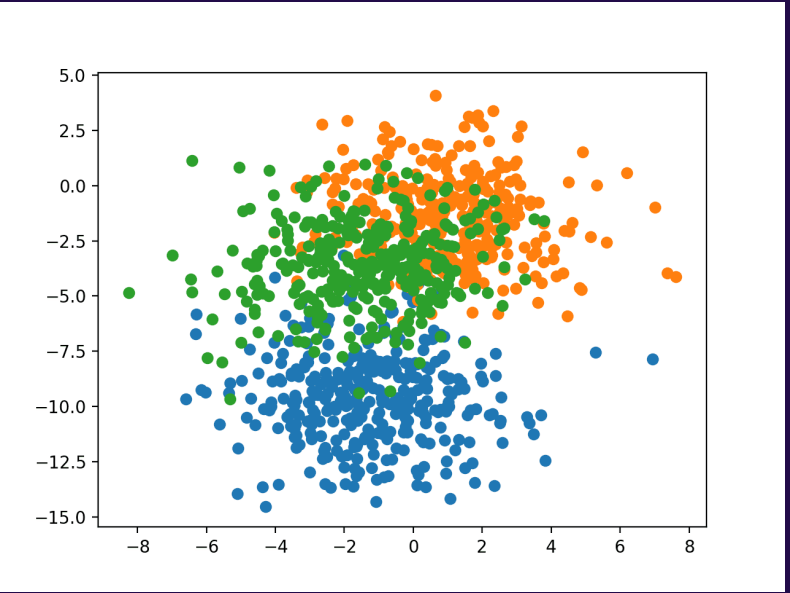
- Sorting
- Grouping
- Visualization
- Basic Statistics
- Feature Engineering

Modelling

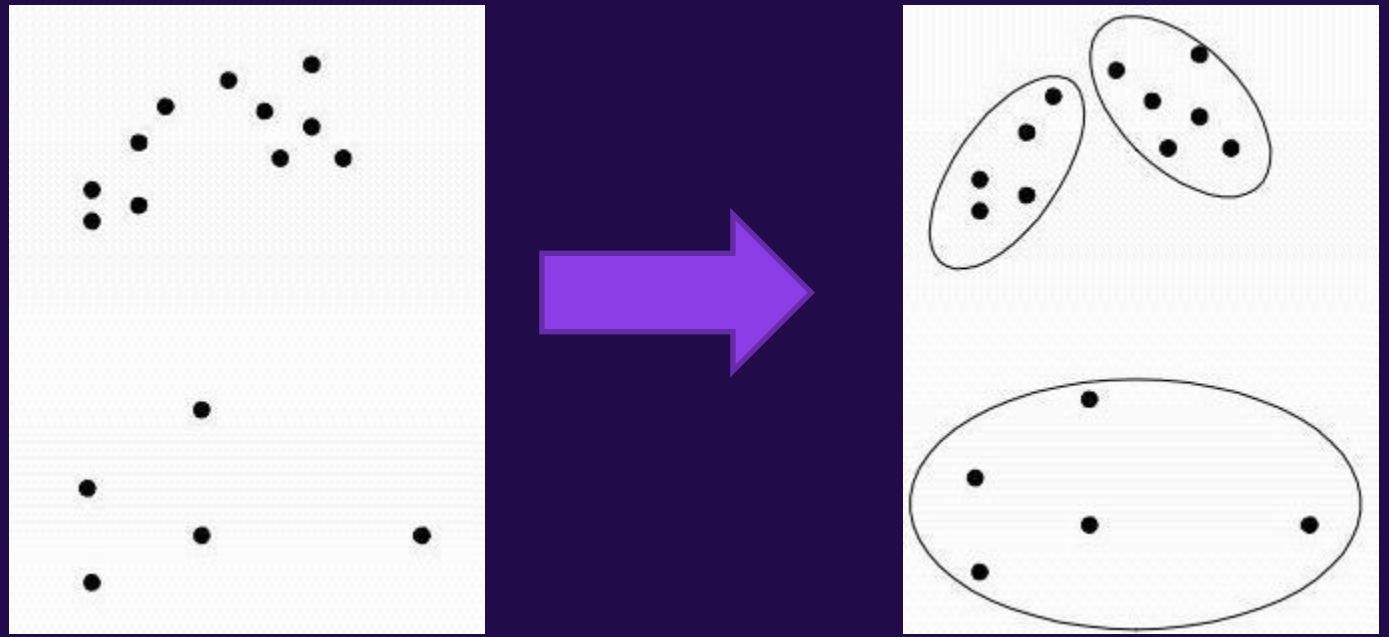
- Regression
- Classification
- Clustering
- Forecasting
- Anomaly Detection

REGRESSION & CLASSIFICATION

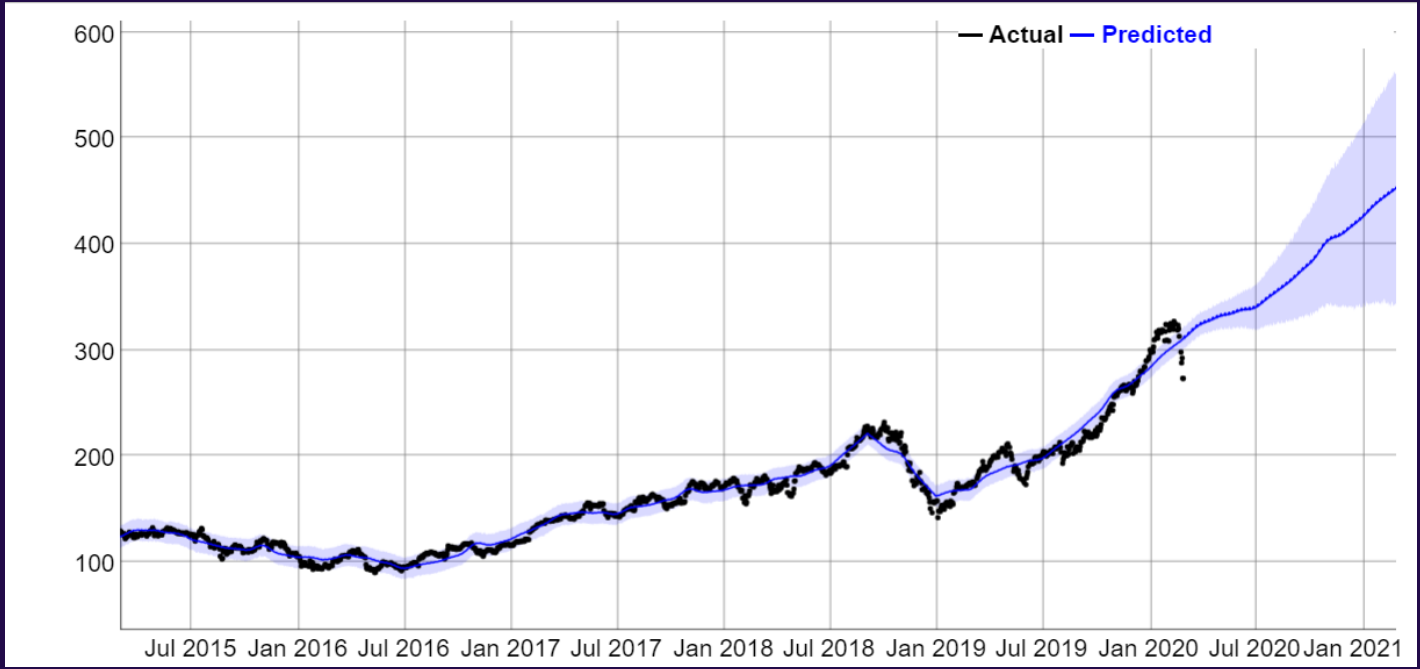




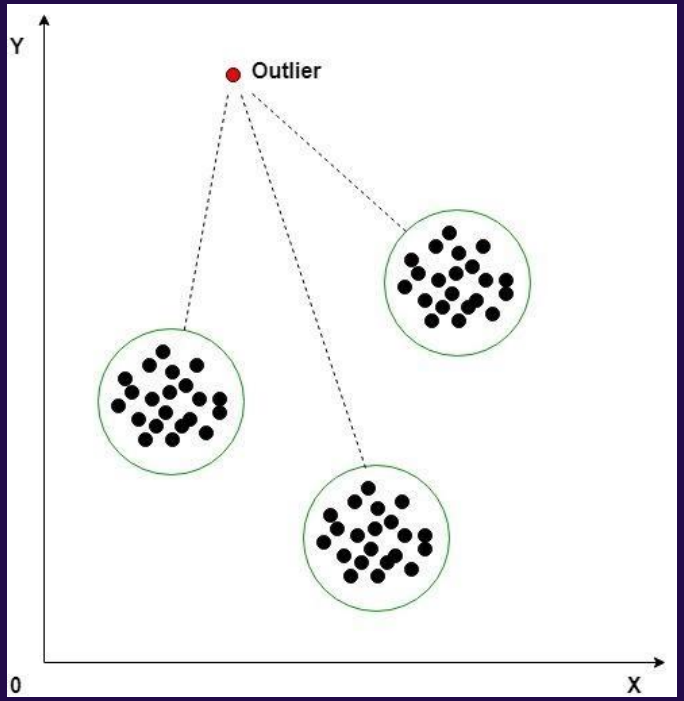
CLUSTERING



FORECASTING

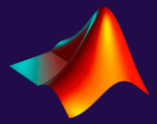


ANOMALY DETECTION



Google Sheets

orange



MATLAB®

Open for Innovation ®
KNIME

CODE-BASED

Keyboard-focused:
Type this and type that

Flexible and light

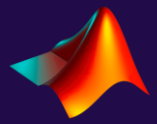
GUI-BASED

Mouse-focused:
Click here and click there

Intuitive and easy

Google Sheets

orange



MATLAB®

Open for Innovation ®
KNIME

CODE-BASED

Keyboard-focused:
Type this and type that

Flexible and light

GUI-BASED

Mouse-focused:
Click here and click there

Intuitive and easy

EXCEL

One of the most used data manager in most traditional business process

WHY EXCEL?

Graphical UI

Visually Intuitive

User-Friendly

Just learn few clicks

Readily available

All Desktop should have it

Simplicity

Less effort to understand

WHY NOT EXCEL?

Data Volume

Only powerful in relatively small dataset

Security Risk

Vulnerable in digital attack

Syntax Problem

Hard to see and understand any bug in formulas

PYTHON

MOST POPULAR
PROGRAMMING LANGUAGE
SINCE 2016

WHY PYTHON?

01

BUILD FOR READABILITY

Easy to be read is Python's purpose of life!

02

Wide Community

A lot of users that can help each other!

03

Rich Libraries

Anything you need is there

04

Multi-function

Cover up almost everything (From game, app, hack, analysis, crawling, robotic, etc)

05

Free!

Just need coffee and a few snacks

06

Glue Language

Connect to many systems

Zen of Python

Beautiful is better than ugly.
Explicit is better than implicit.
Simple is better than complex.
Complex is better than complicated.
Flat is better than nested.
Sparse is better than dense.
Readability counts.
Special cases aren't special enough to break the rules.
Although practicality beats purity.
Errors should never pass silently.
Unless explicitly silenced.

In the face of ambiguity, refuse the temptation to guess.
There should be one-- and preferably only one --obvious way to do it.
Although that way may not be obvious at first unless you're Dutch.
Now is better than never.
Although never is often better than *right* now.
If the implementation is hard to explain, it's a bad idea.
If the implementation is easy to explain, it may be a good idea.
Namespaces are one honking great idea -- let's do more of those!

WHY PYTHON FOR DA?

01

SCALABILITY

More data? More complex? Easy!

02

Automation

Updating data, renaming, formatting, checking, etc, in automatically

03

Data-Agnostic

Give Python any data file, and it will work.

04

Reproducibility

Anyone can do the same things with the same code

05

Supporting Tools

Colaborate with Git and work with Notebook

06

Advanced Capabilities

From statistical inference to cool visualization

WHY NOT PYTHON?

.... Well, nothing!

Eh, except

01

Hard to start

Steep learning curve

Libraries in Python

Analytics

- Statsmodels
- Pandas
- Dask
- NumPy

Visualization

- Matplotlib
- Seaborn
- Plotly

Machine Learning

- Tensorflow
- PyTorch
- Sckit-Learn

BUT, I CAN'T PROGRAMMING

STEEP LEARNING CURVE

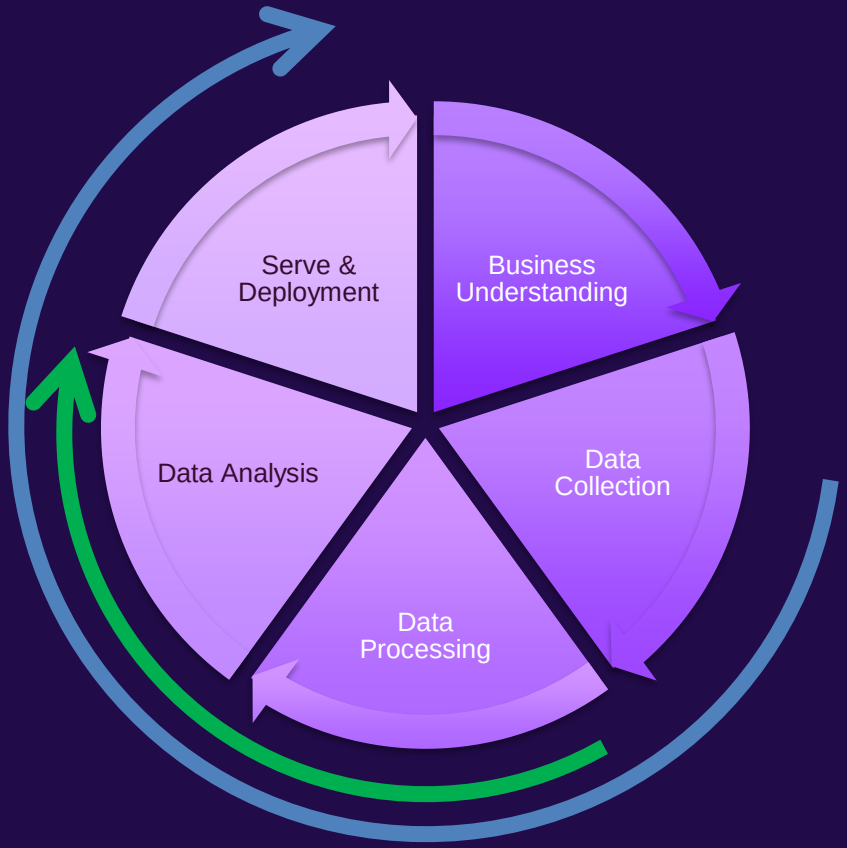
Python is difficult to learn initially

especially for those who never do programming

And yeah, we need to set up environment, install libraries, etc

IN THIS ERA, PROGRAMMING WILL EMPOWERED MOST JOBS

GENERAL STEP OF DA



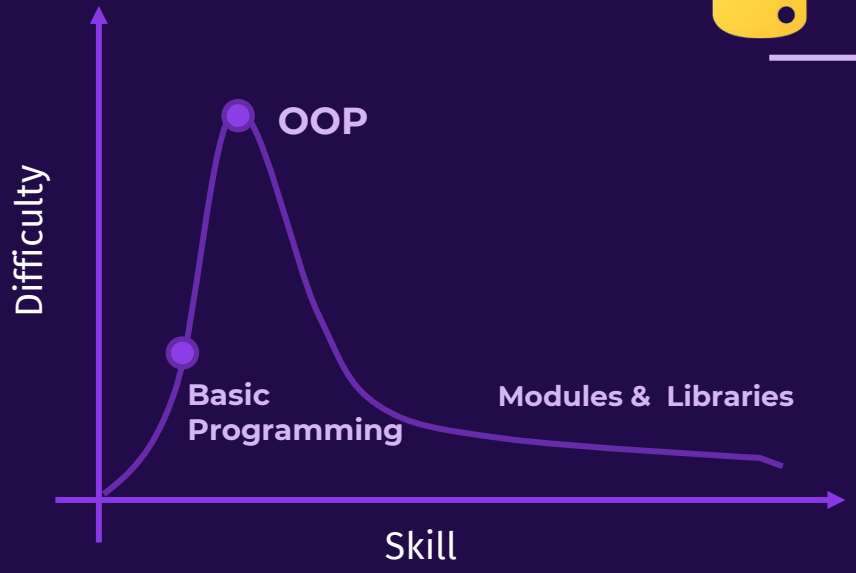
DATA ANALYTICS TASKS

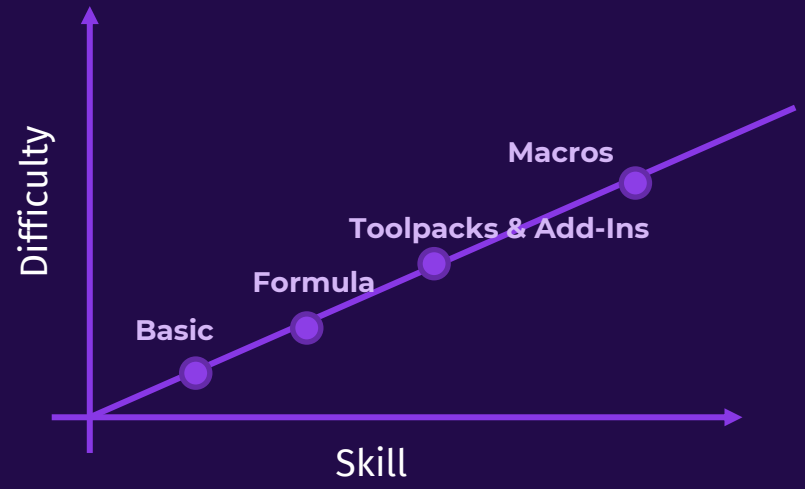
Exploratory

- Sorting
- Grouping
- Visualization
- Basic Statistics
- Feature Engineering

Modelling

- Regression
- Classification
- Clustering
- Forecasting
- Optimization


Difficulty
OOP
Basic
Programming
Modules & Libraries
Skill


Difficulty
Macros
Toolpacks & Add-Ins
Formula
Basic
Skill

OPPORTUNITY COST

	Excel	Python
Simplicity	+	–
Capability	–	+

Keep using excel, no learning effort, but only able to do the same things

Move to python, learn all the tools, advance the analysis

LET'S SEE HOW IT'S DONE!

THANKS!

Do you have any questions?

adityaihsan@telkomuniversity.ac.id
+62 8574 1852 615
phoenixfin.github.io

Aditya Firman Ihsan

CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon**, and infographics & images by **Freepik**.

Please keep this slide for attribution.