

**Materi II**

# **Business & Data Understanding**



**Reza R Pratama**

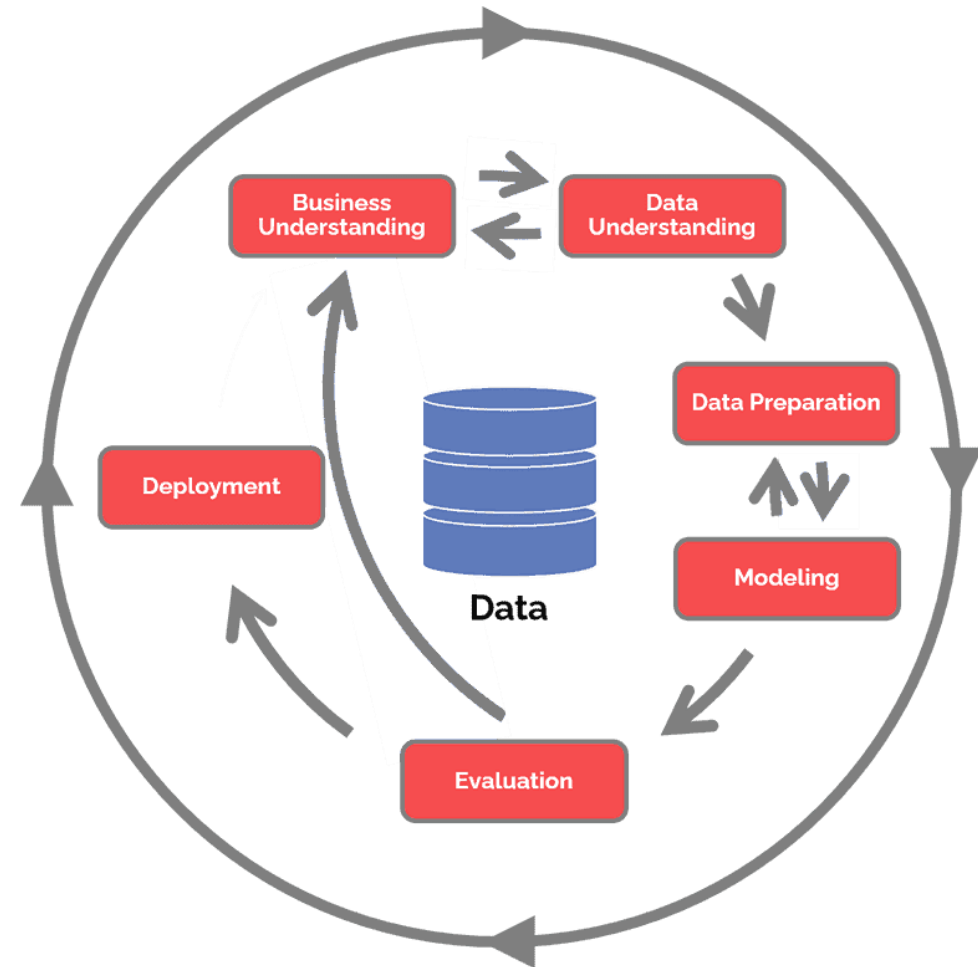
# Introduction

A person is working at a desk with multiple laptops. They are holding a white coffee cup in their right hand and a blue folder in their left hand. The background is slightly blurred, showing more laptops and papers on the desk. A large blue semi-circle is overlaid on the left side of the image, containing the word "Introduction" in white text.

# CRISP-DM Data Mining Lifecycle

- Business Understanding
- Data Understanding
- Data Preparation
- Modeling
- Evaluation
- Deployment

*Notice the iteration!*



THE FAMOUS CRISP-DM

Business Understanding	Data Understanding	Data Preparation	Modeling	Evaluation	Deployment
<p><b>Determine Business Objectives</b> Background Business Objectives Business Success Criteria</p> <p><b>Situation Assessment</b> Inventory of Resources Requirements, Assumptions, and Constraints Risks and Contingencies Terminology Costs and Benefits</p> <p><b>Determine Data Mining Goal</b> Data Mining Goals Data Mining Success Criteria</p> <p><b>Produce Project Plan</b> Project Plan Initial Assessment of Tools and Techniques</p>	<p><b>Collect Initial Data</b> Initial Data Collection Report</p> <p><b>Describe Data</b> Data Description Report</p> <p><b>Explore Data</b> Data Exploration Report</p> <p><b>Verify Data Quality</b> Data Quality Report</p>	<p><i>Data Set</i> <i>Data Set Description</i></p> <p><b>Select Data</b> <i>Rationale for Inclusion / Exclusion</i></p> <p><b>Clean Data</b> <i>Data Cleaning Report</i></p> <p><b>Construct Data</b> <i>Derived Attributes</i> <i>Generated Records</i></p> <p><b>Integrate Data</b> <i>Merged Data</i></p> <p><b>Format Data</b> <i>Reformatted Data</i></p>	<p><b>Select Modeling Technique</b> <i>Modeling Technique</i> <i>Modeling Assumptions</i></p> <p><b>Generate Test Design</b> <i>Test Design</i></p> <p><b>Build Model</b> <i>Parameter Settings</i> <i>Models</i> <i>Model Description</i></p> <p><b>Assess Model</b> <i>Model Assessment</i> <i>Revised Parameter Settings</i></p>	<p><b>Evaluate Results</b> <i>Assessment of Data Mining Results w.r.t. Business Success Criteria</i> <i>Approved Models</i></p> <p><b>Review Process</b> <i>Review of Process</i></p> <p><b>Determine Next Steps</b> <i>List of Possible Actions</i> <i>Decision</i></p>	<p><b>Plan Deployment</b> <i>Deployment Plan</i></p> <p><b>Plan Monitoring and Maintenance</b> <i>Monitoring and Maintenance Plan</i></p> <p><b>Produce Final Report</b> <i>Final Report</i> <i>Final Presentation</i></p> <p><b>Review Project</b> <i>Experience</i> <i>Documentation</i></p>



The image features a background of a person in a business suit sitting at a desk. The person's left hand is on a laptop keyboard, and their right hand is holding a white coffee cup with a black lid. On the desk, there are several documents with charts and graphs, and a black pen. A large green semi-circle is on the left side of the image, and a blue semi-circle is on the right side, overlapping the green one. The text "Business Understanding" is written in white on the green semi-circle.

# Business Understanding



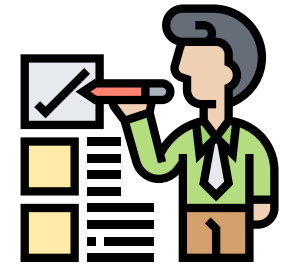
**Menentukan  
tujuan bisnis**



**Analisis  
Situasi saat ini**



**Menentukan  
tujuan proyek dan matriks  
kesuksesan**



**Membuat rencana  
proyek**



**Business Unit & Tim Data  
Analytics**

# 5W 1H in Data Analytics

## Who:

Who are the final users?  
Who is going to approve these solutions?  
Who do we report to?

## What:

What is/are the problems and/or pain points?  
What resources are available to us?  
What known data issues exist?  
What does success look like?  
What is it worth to solve the problem?

## When:

When is/are the project(s) due?  
When are progress reports taken place?  
When can results be verified?

## Where:

Where can we retrieve \_\_\_\_\_ data?  
Where can we store \_\_\_\_\_ data?  
Where is the process currently?

## Why:

Why is this process currently set up this way?  
Why is this information  
important/proprietary/sensitive?  
Why was this equipment/software/cloud  
service mentioned??



# Tools for Business Understanding: Project Charter

## PROBLEM STATEMENT

- Target to lower WACC by new business model (peer to peer gold)
- Need to find potential lender within 16++ million of Pegadaian existing customers
- Need to forecast the risk in gold price volatility
- Need to forecast the growth in lender
- Need to define optimum capital structure

## GOAL STATEMENT

- Descriptive analytics to make targeted marketing into particular customer segment
- Predictive analysis to predict the risk in gold volatility and growth of lender
- Prescriptive analysis to determine the portfolio for optimum capital structure.

## Project: Project CS (Crowd Sourcing)

### PROJECT TEAM

- Project Sponsor
- Business PM
- Technical PM
- Functional and Tech
- Expert User

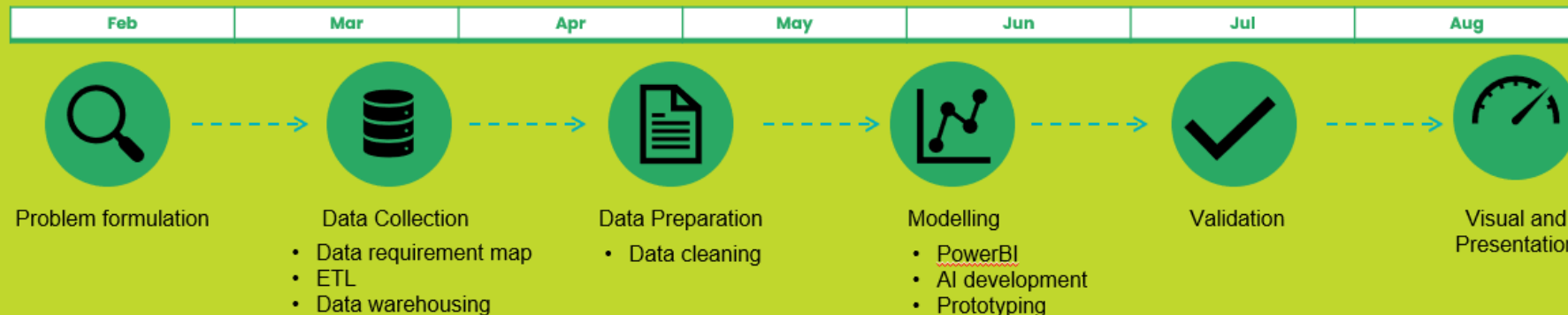
### PROJECT STAKEHOLDERS

- Marketing & Sales
- PDS
- IT
- Appraiser expert

### CRITICAL SUCCESS FACTORS

- Descriptive analysis before May 2022
- Predictive analysis before June 2022
- Prescriptive analysis before July 2022

## PROJECT MILESTONES





# Tools for Business Understanding: As Is – To Be

## As is condition (Before)

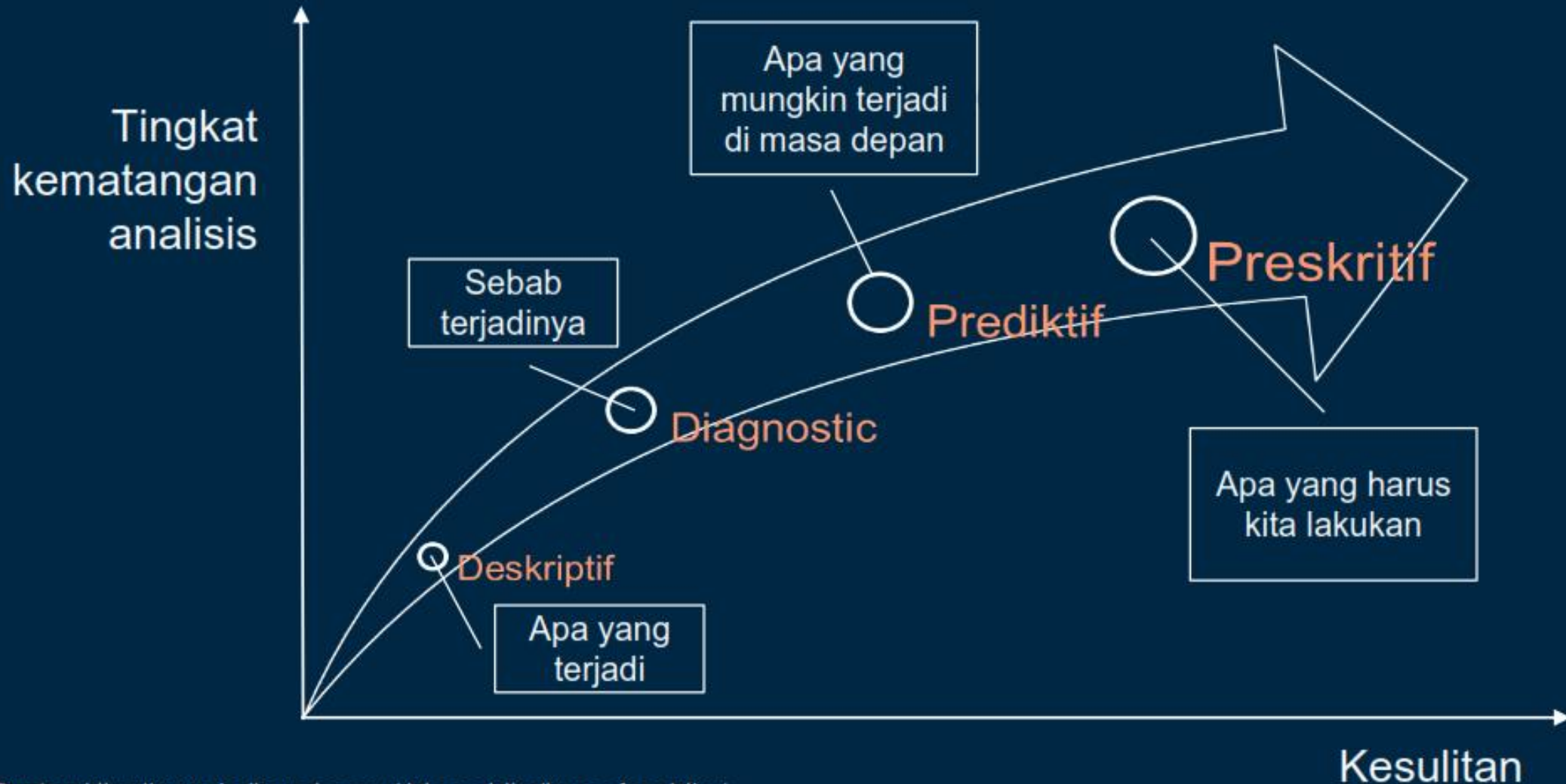
- Pegadaian Funding depends on Bank, Securities and Equity only
- Option investment in Gold Product
- Physical Auction
- Customers should present to Pegadaian to bring the collateral



## To be condition: (After Idea implementation)

- Pegadaian Funding for Pawn Product also from Gold Investor
- Alternative of product investment to get certain benefit through fix return
- Online auction through Community
- Alternative picked up collateral by Agent or through partnership with logistic services

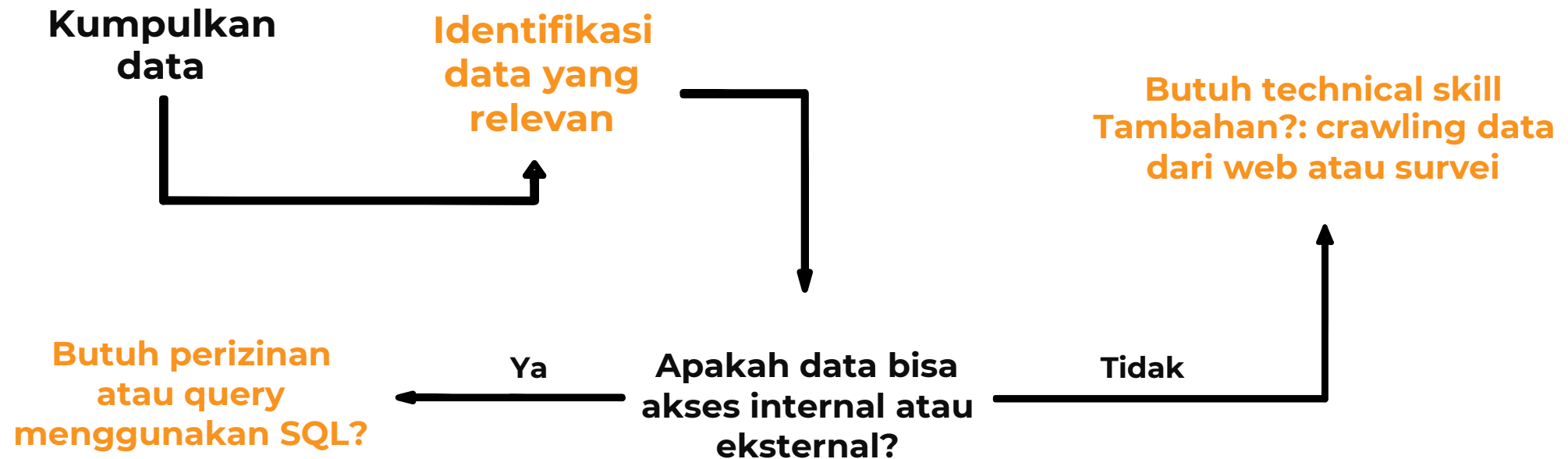
# Analytics approach



# Data Understanding

A person is working at a desk with multiple laptops. They are holding a white coffee cup in their right hand and a blue pen in their left hand. The desk is cluttered with papers and documents. An orange circular overlay is on the left side of the image, containing the text 'Data Understanding'.

# Data Understanding



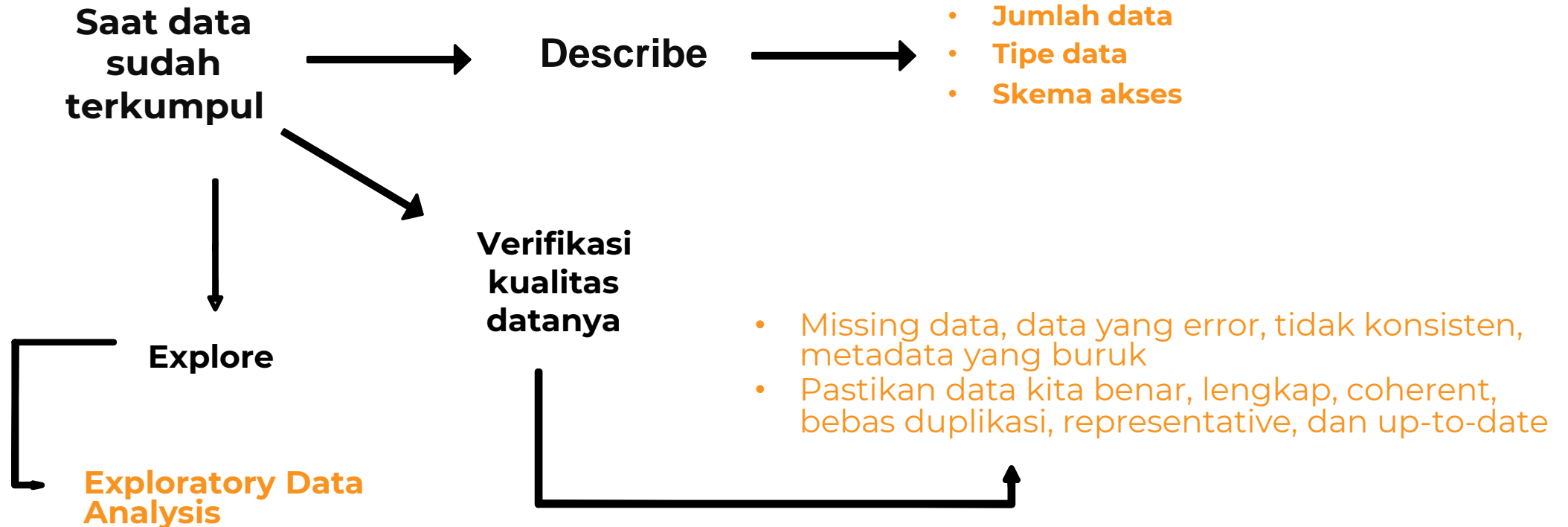
# Data Collection Plan



Type of Data	What data will be collected?	How?	What will this tell us?
House Listing Data	Location, price, house specification, location information	Manual scraping	To predict house price based on specification and location
Customer Credit Application Data	Since we have no access to credit data so we will use open dataset	Kaggle	For credit risk rating



# Data Understanding



# Data Understanding with python<sup>TM</sup>

