MySkill

#RintisKarirImpian

Intro to Data Analytics

Achmad Rozie
Data Analyst at Flip

Outline

- Data Analytics
- Steps in Data Analysis

What is Analysis

Analisis merupakan upaya yang dilakukan untuk *mengamati sesuatu secara* ***mendalam dan mendetail*** melalui proses penguraian berbagai komponen pembentuknya atau juga penyusunan kompenen tersebut untuk dipelajari atau diselidiki lebih lanjut - *brainly*

#RintisKarirImpian

MySkill

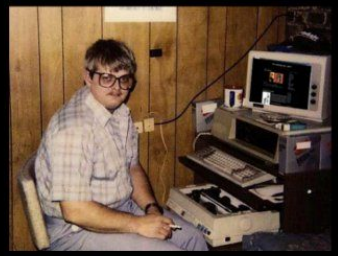
#RintisKarirImpian

MySkill

Data Analytics

Data Analytics
(?)

Data Analyst

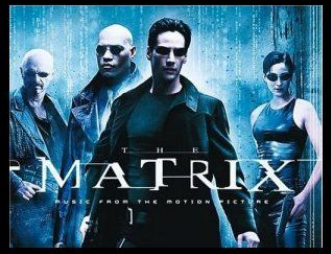


What my friends think I do

What my Mom thinks I do

What my boss thinks I do

What my customers think I do



What I think I do

What I really do

Simple definition

Process of examining raw data to extract **useful information**

DATA ANALYTICS

PEOPLE WHO DON'T KNOW

PEOPLE WHO KNOW

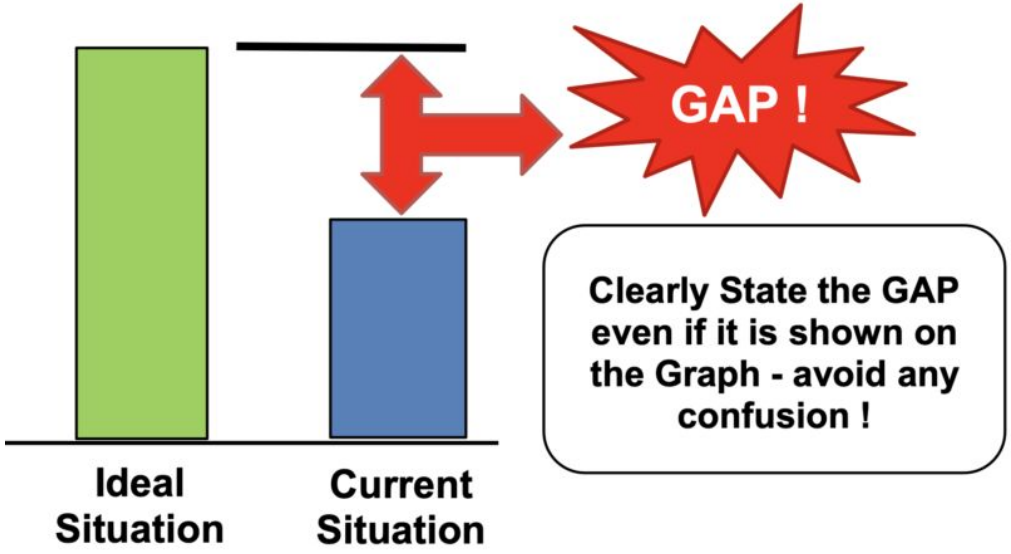
Steps in Data Analysis

- Define problems
- *Confirm the analysis point first! with your business user ofc.*
- Find the data
- Preprocess the data
- Do the analysis
- Interpret the result

Define Problems

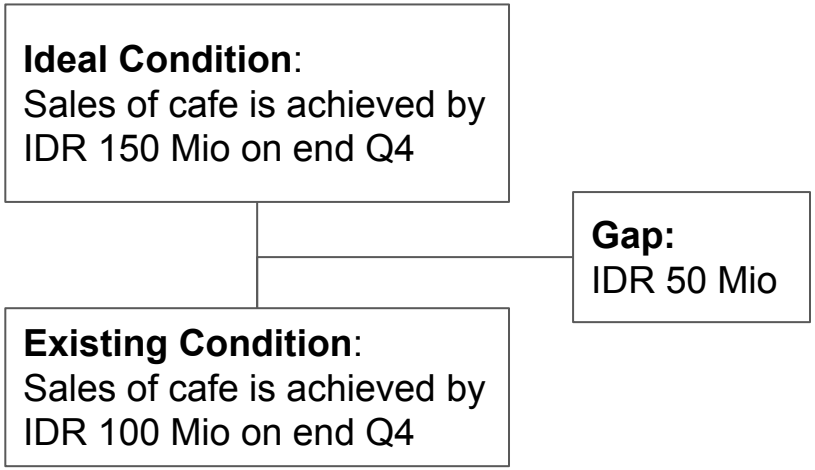
What is problem?

Problem = Gap


GAP !
Clearly State the GAP even if it is shown on the Graph - avoid any confusion !
Ideal Situation
Current Situation

Example

- You are the manager of a cafe. The sales on Q4 2022 is decreased if compared with with Q3 2022. The sales is only achieved IDR 100 Mio while the target is 150 Mio.



Remember!

- Good gap / problem is able to be **quantified**.
- If you can’t **measure** it, you can’t improve it!
- How if the problem is subjective?
  - try to look the **benchmark**,
  - or set your own **threshold**!
- Redefine the problem with your own stakeholder!

Define Problems

- Related to “why do you do this analysis”?
- Refine and realign the problems with all stakeholders
- Cutting as much **assumptions** as possible
- Stating the **hypotheses**
- Setting the **priority and urgency**

What are problem statements of these cases?

- You are the director of a hospital. Lately you feel that the hospital is more crowded than usual and the rooms are unusually full, but the number of patients seems steady from the record. What problem will you need to address to your analysts?

- You are the manager of a cafe. The sales of coffee is decreased 20% if compared with last month. What problem will you address to your analysts?

What metrics can you propose to these cases?

- How to rank best students in the class?
- How to categorize consumers loyalty?
- How to measure our delivery speed?
- How to know the users satisfaction?

Example

You are the director of a hospital. Lately you feel that the hospital is **more crowded than usual** and the rooms are unusually full, but the number of patients seems steady from the record. What problem will you need to address to your analysts?

Ideal Condition:
Average people in a room is 3 people / patients

Gap:
There is an excess of 2 people in the room

Existing Condition:
Average people in a room is achieved more than 5 people / patients

Confirm your analysis point first!

Why?

In order to **minimize** the rework.

Remember, most of analysis is never ending process, we need to minimize and limit it.

Analysis Framework Component

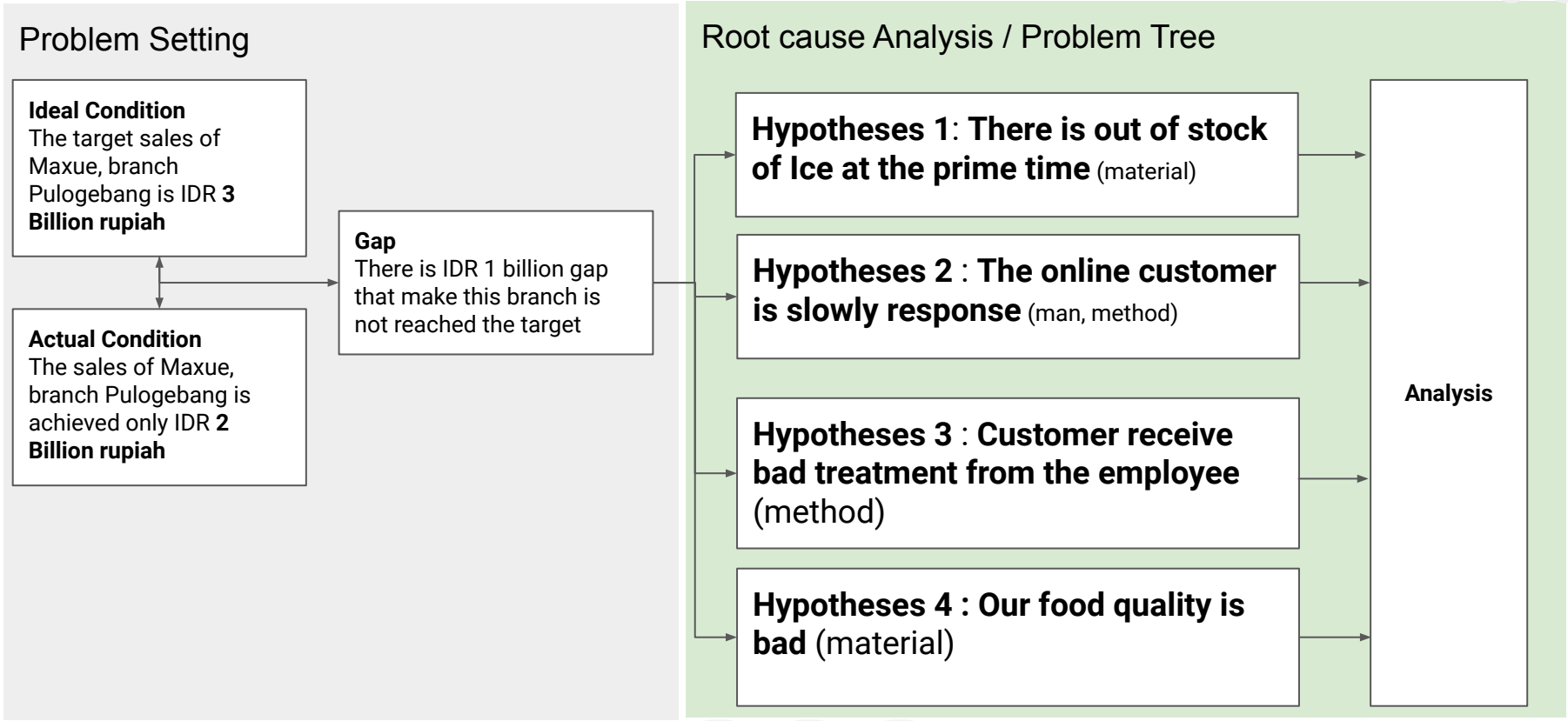
Analysis Framework that usually I used is contained of 8 steps problems and problem tree. I usually divide to 3 things. **Problem Setting, Problem Breakdown, Root Cause analysis**

1. Problem Setting

2. Problem Breakdown (Optional)

3. Root cause Analysis / Problem Tree

Example of Analysis Framework (1) - *easier to be used*



Example of Analysis Framework (2) - *more targeted*

Problem Setting

Ideal Condition
The target sales of Maxue, branch Pulogebang is IDR **3 Billion rupiah**

Actual Condition
The sales of Maxue, branch Pulogebang is achieved only IDR **2 Billion rupiah**

Gap
There is IDR 1 billion gap that make this branch is not reached the target

Food
The food is IDR 200 Mio below the target

Beverage
The food is IDR 800 Mio below the target

Juice & Latte
This cluster is achieved 150 mio below the target

Ice Cream
This cluster is achieved IDR 550 mio below the target

Root cause Analysis / Problem Tree

Hypotheses 1: **There is out of stock of Ice at the prime time** (material)

- Stock fulfillment
- Out of stock ratio
- Supplier availability

Hypotheses 2 : **The online customer is slowly response** (man, method)

Hypotheses 3 : **Customer receive bad treatment from the employee** (method)

Hypotheses 4 : Our food quality of ice cream is bad (material)

Breakdown *(similar with the above)*

MySkill

How? we can use / **optimize** through analysis framework!

After you got the problem (gap):

Define your analysis point. It can be:

a. Hypotheses,
b. Point that may be affected
c. Factor / event that may cause the problem

Discuss with business user

Confirm is it enough or is there anything else that not inserted on your analysis framework

Also confirm the constraints

It can be the:

a. the data should we used
b. The time period
c. the condition

Additional tips for “confirm your analysis point first!”

1. **At the beginning, usually i write it on a piece of paper / tab first!**
   → help **to organize** what we want to look for

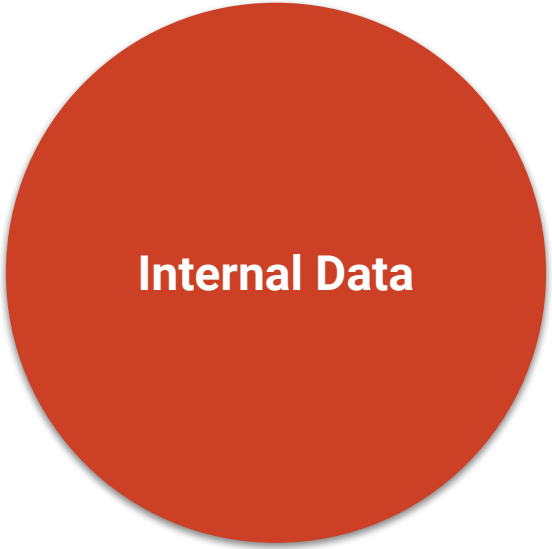
1. **Don’t make it too complex!**
   → because the analysis will come and come continuously, we must finish it quickly (of course still within our work pace)

Continue your previous problem definition (5 mins)
At least 3 hypotheses

- You are the director of a hospital. Lately you feel that the hospital is more crowded than usual and the rooms are unusually full, but the number of patients seems steady from the record. What problem will you need to address to your analysts?

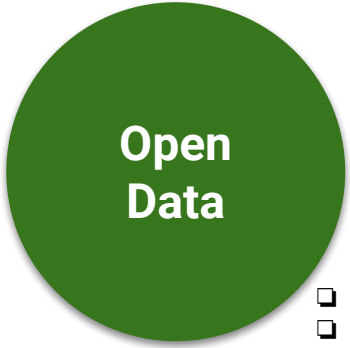
- You are the manager of a cafe. The sales of coffee is decreased 20% if compared with last month. What problem will you address to your analysts?

Find the Data



- ❏ App tracking
- ❏ Business logs
- ❏ Customers data

- ❏ Google analytics
- ❏ Competitors data
- ❏ Vendors data



- ❏ Public holidays
- ❏ Postal codes
- ❏ Landmark locations

This is what we do, daily …

Raw sources

Analytics databases

Analytics tools
(Big Query, Google Sheets, Python, Redash, Looker, or even your note)

Common issues

- ❏ Are the data available?
- ❏ Are the data ready to use?
- ❏ Are the data reliable?
- ❏ Are the data restricted?
- ❏ Are the data condition is right? *(underrated but important)*
- ❏ etc

Preprocess the Data

MY FAMILY'S REACTION TO DIRTY DATA

#RintisKarirImpian

MySkill

Data preprocessing

- Remove **duplicates**
- Handle **anomalies** and dirty data
- Take action on **missing** data
- **Standardize** the format and types

Do the analysis


Kowalski, analysis

SQL based

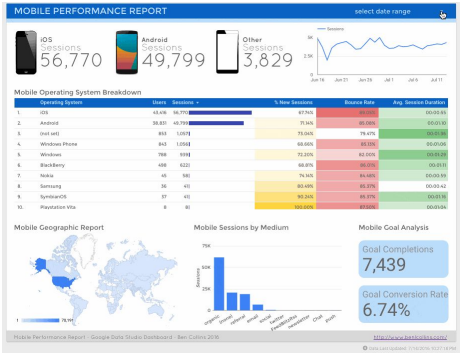
Spreadsheet based

Coding based

Google Forms

Google Sheets

Looker



#RintisKarirImpian

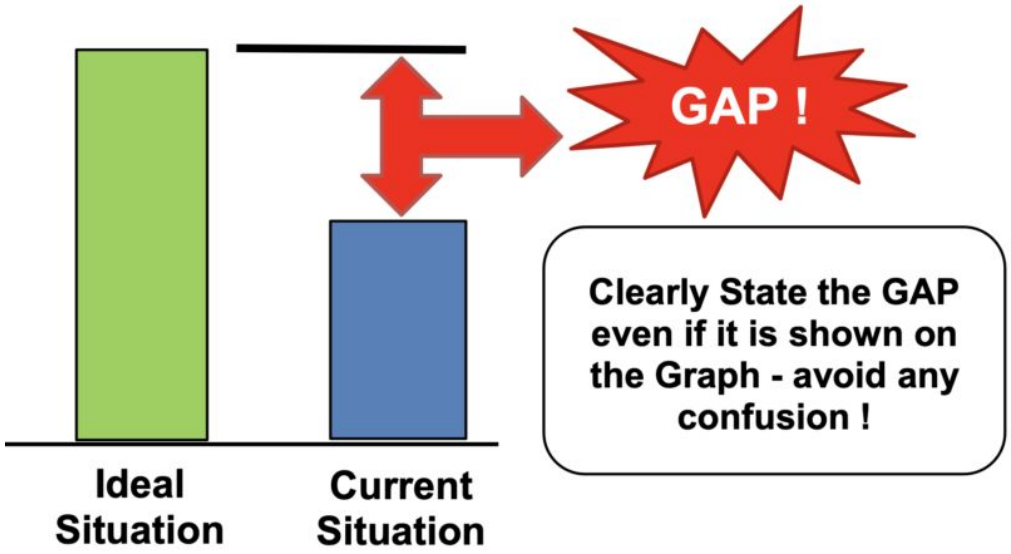
MySkill

Common statistics in data analysis

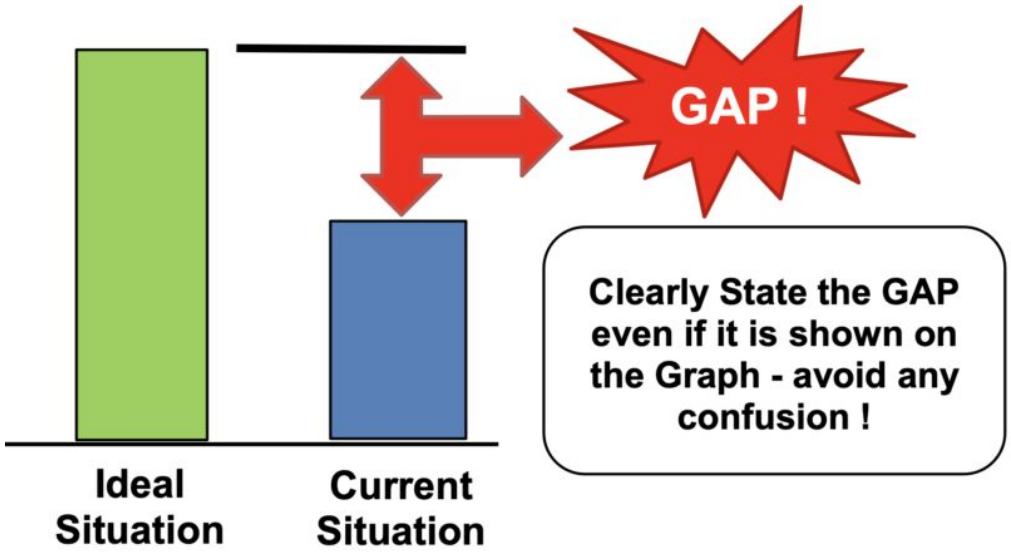
- ❏ Count
- ❏ Count distinct
- ❏ Sum
- ❏ Min
- ❏ Max
- ❏ Average
- ❏ Median
- ❏ Percentile
- ❏ etc

Interpret the Result

Problem = Gap. **Know the gap first!**

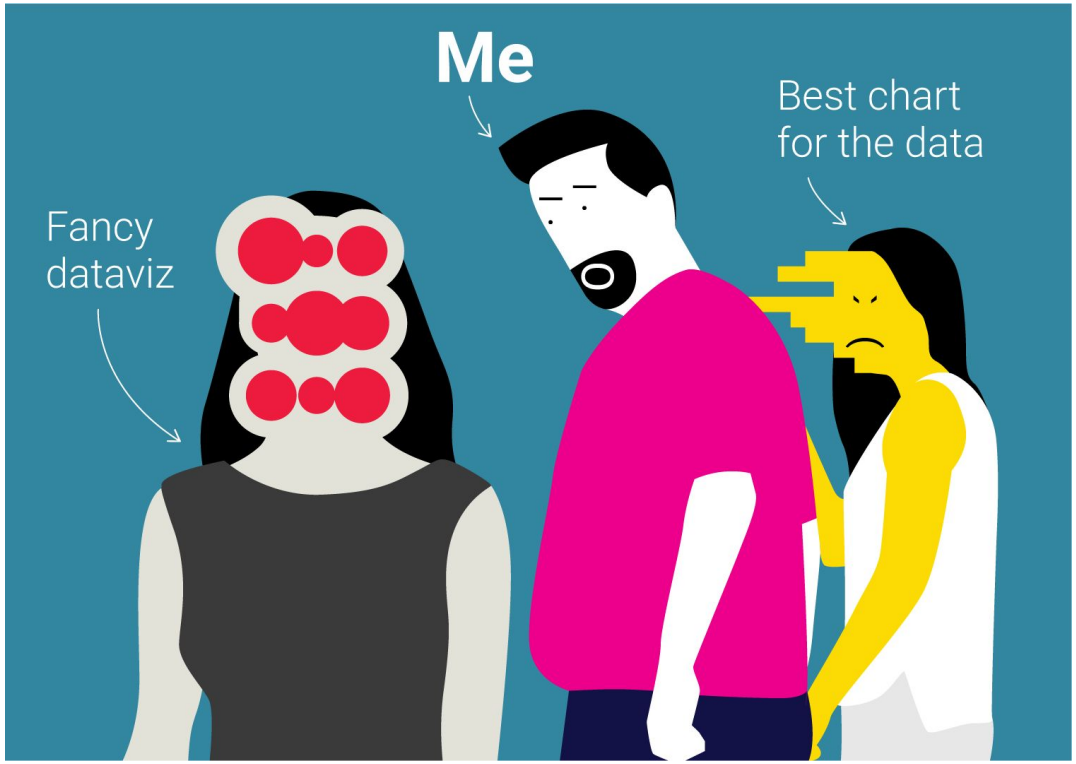


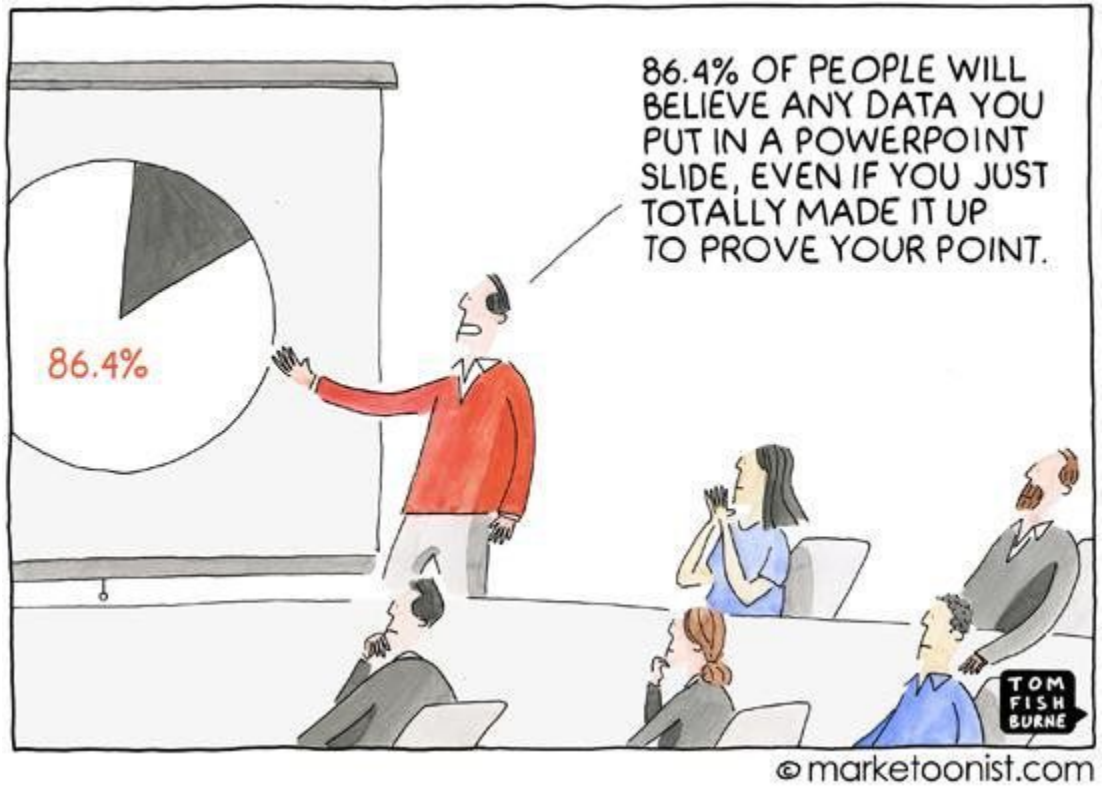
Problem = Gap. **Know the gap first!**

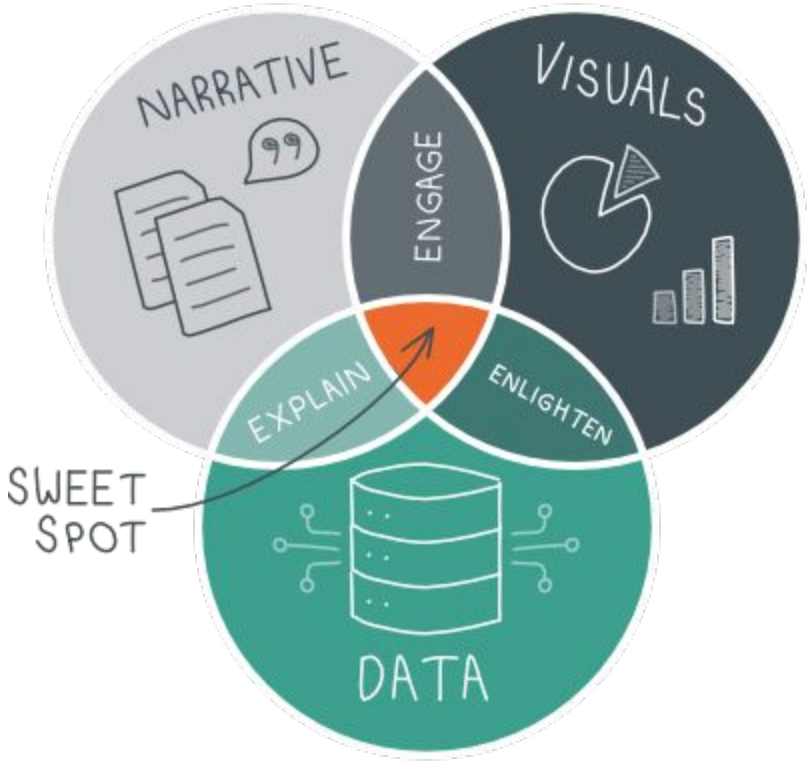


#RintisKarirImpian

MySkill


Me
Best chart
for the data
Fancy
dataviz


86.4% OF PEOPLE WILL BELIEVE ANY DATA YOU PUT IN A POWERPOINT SLIDE, EVEN IF YOU JUST TOTALLY MADE IT UP TO PROVE YOUR POINT.
86.4%
TOM FISH BURNE
©marketoonist.com


NARRATIVE
VISUALS
ENGAGE
EXPLAIN
ENLIGHTEN
SWEET SPOT
DATA

end.