

Week 1: Data Analysts meet?

## The Six Data Analysis Phase



### 1) Ask

- to define a problem to be solved.
- to help focusing on the actual problem and avoid any distractions.
- to make sure you fully understand the stakeholder's expectations.



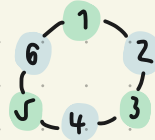
### 2) Prepare

- what metrics to measure?
- locate data in your database.
- create security measures to protect that data (policy agreement)



### 3) Process

- to clean up your data (get rid of possible errors, inaccuracies, and inconsistencies)
- to use spreadsheet functions to find incorrectly entered data.
- to use SQL functions to check for extra spaces.



### 4) Analyze

- to perform calculations.
- to combine data from multiple sources.
- to create table with your results.



### 5) Share

- to make more informed decisions via Graphs or Dashboards.
- to share results with stakeholders.



### 6) Act

- to act on your data.
- to provide your stakeholders with recommendations based on your findings.



## Six Problem Types

Data Analysts typically work with these:

- 1) Making Predictions → carrot → vegetable
- 2) Categorizing things → assigning items to categories
- 3) Spotting sth. unusual → eye icon
- 4) Identifying themes → grouping them into broader themes (In a User Study, examples of themes)
- 5) Discovering connections → butterfly icon
- 6) Finding Patterns

### \* SMART Questions \*

Avoid asking questions that:

- ✗ Close-ended Questions (answer with yes/no)
- ✗ Vague & Lacks content
- ✗ Leading Questions ???

Bias, Unfair  
to some groups in



Asking questions that's SMART

- ✓ S (Specific)
- ✓ M (Measurable)
- ✓ A (Action-Oriented) → encourage change
- ✓ R (Relevant)
- ✓ T (Time-bound) → specify time to be studied.

SMART Question



## Week 2: Data, Report, Dashboard

### Data

#### Small Data

- specific metrics, short defined time
- usually organized and analyzed in spreadsheets.

#### Big Data

- large, less-specific, cover longer time period
- usually organized and analyzed in Databases
- \* Needs to be broken into smaller pieces in order to be organized & analyzed effectively

- 3V
- Volume: Amount of data
  - Variety: Different kinds of data
  - Velocity: How fast data can be processed
  - Veracity: Quality and Reliability of the data

Some data analysts consider a fourth V.

### Metric

a single, quantifiable type of data that is used for measurement.

## Report vs. Dashboard

### Pros

- represents high-level historical data
- easy to design
- pre-cleaned & sorted data

### Cons

- Lack of continual maintenance
- no visual appealing
- Static

### Pros

- Dynamic, Interactive
- suitable when sharing information across many peoples promptly

### Cons

- Labor-intensive design (a lot of effort)
- can lead to misunderstanding (if it's not well-designed)

### Return on Investment (ROI)

cost of an investment to the net profit over a period of time.

## Week 3: Spreadsheets

helps doing calculations faster using operators

e.g. Sum, Average, Min, Max, Data Ranges D2:D4

Cell References (C2 E3 D8)

Absolute Referencing (\$) such as \$A\$10

operators (+, -, \*, /)

Formulas is a set of instructions to perform calculation.

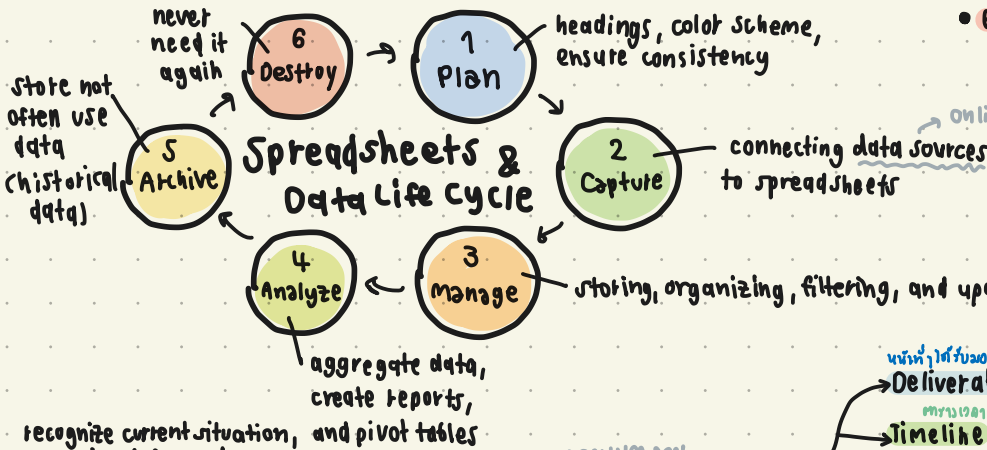
Errors

#DIV/0! → divide value in a cell by a empty cell

#ERROR! (parsing error) → not correctly place the order

#VALUE! #NAME #REF! → accidentally deleted row/col

Functions e.g. SUM(), is a preset command to perform a specific process



## Structured thinking

### Problem Domain

Specific area of analysis that encompasses every activity affecting / affected by the problem.

Need to understand this first! \*\* before discovering the story. (whole image)

Ask clarifying questions, Define what to accomplish, Specify project boundaries

Ask "who, what, when, where, why, how" to put information into context

"Data has little value if it is not paired with context."

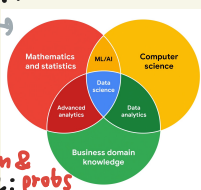
turn raw data into meaningful info.

## Week 4: Communication

### Communication

→ stakeholder → hold stakes on what you are doing, invest time & resources to a project.

- Executive team → provides strategic and operational leadership to the company. → Primary stakeholders
- Customer-facing team → who interacts with customers. Company ↔ CFT ↔ customer
- Data science team



### Clear Communication

to avoid confusion & Before communicate: probs

- who your audience is?
- what they already know?
- what they need to know?
- How you can communicate that effectively to them?

Linkage between team, Organize overall proj & team "Project manager"