Week 1 : Data Analysts meet?

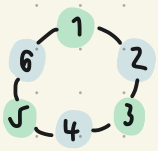
The Six Data Analysis Phase

1

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
4. Identifying themes → grouping them into broader themes (In a user study, examples of themes) (ตัวอย่างของธีม)
5. Discovering connections
6. Finding Patterns

* SMART Questions *

Avoid asking questions that:
- ✕ close-ended Questions (answer with yes/no)
- ✕ Vague & Lacks content
- ✕ Leading Questions ??? → Bias, Unfair to same groups :(

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

19/12/2022

Week 2: Data, Report, Dashboard

Data

Small Data	Big Data
- specific metrics, short defined time	- large, less-specific, cover longer time period
- usually organized and analyzed in spreadsheets.	- 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

Report vs. Dashboard

Report

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

Cons
- Lack of continual maintenance
- no visual appealing
- Static

Dashboard

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)

Metric

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

Return on Investment (ROI)

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

Week 3: Spreadsheets

1. Organize (sort & filter) e.g. Pivot table
2. Calculate data using operators)

→ helps doing calculations faster
e.g. sum, Average, min, mean, max

- Data Ranges D2:D4 (lock both col and row)
- Cell References (C2 E3 D8)
  - Absolute Referencing ($) such as A10 (lock cell)
- operators (+, -, ×, ÷)
- Formulas is a set of instructions to perform calculation. e.g. =10
- Errors
  - #DIV/0! → divide value in a cell by 0 / 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

Spreadsheets & Data Life Cycle

1. Plan — headings, color scheme, ensure consistency
2. Capture — connecting data sources to spreadsheets (e.g. online survey)
3. Manage — storing, organizing, filtering, and updating information
4. Analyze — aggregate data, create reports, and pivot tables
5. Archive — store not often use data (historical data)
6. Destroy — never need it again

Structured thinking

recognize current situation, organize info., and identify options

เข้าใจภาพรวม / แก้ไขปัญหา

Scope of Work (SOW) ขอบเขตงาน
* ไม่ตายตัว, แต่ละงาน different กัน
- Deliverables หน้าที่ๆได้รับมอบหมาย (e.g. Estimate budget for an event)
- Timeline ตารางเวลา (e.g. Invite all attendees by June 1)
- Milestones ขั้นตอน (e.g. Confirm list of employees who will attend)
- Reports รายงาน, paper (e.g. Employee feedback after the training)

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," to put information into context
  - "Data has little value if it is not paired with context."
  - Turn Raw data → Meaningful info.

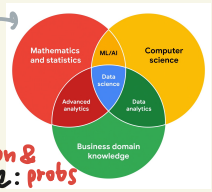
Week 4: Communication

→ 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.
- Data Science team

Company ↔ CFT ↔ Customer

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



→ Clear Communication (to avoid confusion & probs)
- Before communicate:
  - → who your audience is?
  - → what they already know?
  - → what they need to know?
  - → How you can communicate that effectively to them?