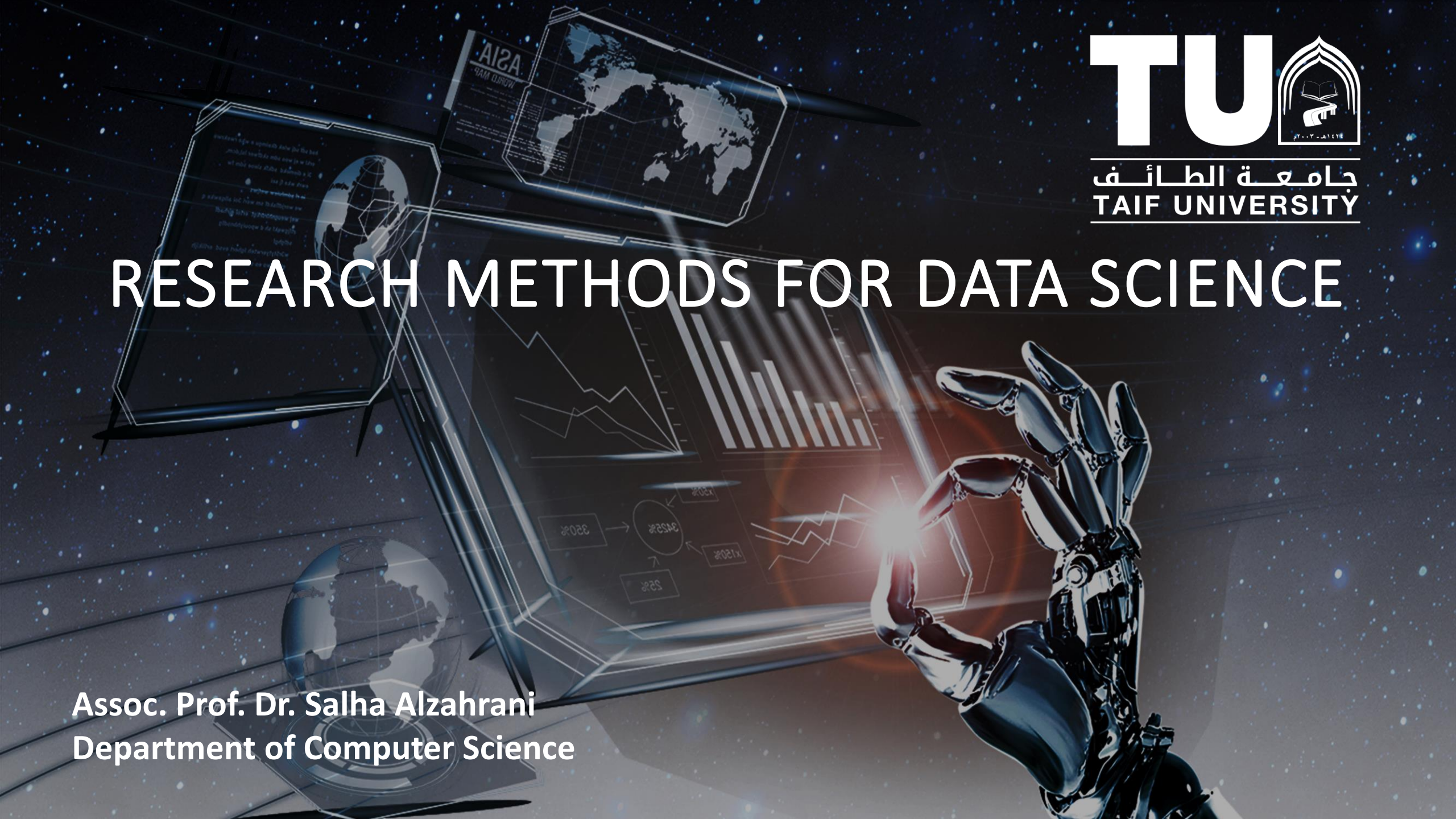


جامعة الطائف
TAIF UNIVERSITY

RESEARCH METHODS FOR DATA SCIENCE

Assoc. Prof. Dr. Salha Alzahrani
Department of Computer Science

Step 2 : Conceptualising a research design

My Notes

~Research Design

~Research Process in Data Science

What is a research design?

A research design is the road map that you decide to follow during your research journey to find answers to your research questions as validly, objectively, accurately and economically as possible. It is a procedural-cum- operational plan that details what and how different methods and procedures to be applied during the research process. According to Kerlinger (1986: 279):

> A research design is a plan, structure and strategy of investigation so conceived as to obtain answers to research questions or problems. The plan is the complete scheme or programme of the research. It includes an outline of what the investigator will do from writing the hypotheses and their operational implications to the final analysis of data.

My Notes

What is a research design?

According to Thyer (1993: 94):

A traditional research design is a blueprint or detailed plan for how a research study is to be completed – operationalizing variables so they can be measured, selecting a sample of interest to study, collecting data to be used as a basis for testing hypotheses, and analysing the results.

My Notes

The functions of research design

A research design is a plan through which you decide for yourself and communicate to others your decisions regarding what study design you propose to use, how you will collect information from your respondents, how you will select your respondents, how the information you will collect is to be analysed and how you will communicate your findings. In addition, you detail your rationale and justification for each decision that shapes your answers to the 'how' of the research journey. In presenting your rationale and justification you need to support them critically from the literature reviewed. You also need to assure yourself and others that the path you have proposed will yield valid and reliable results.

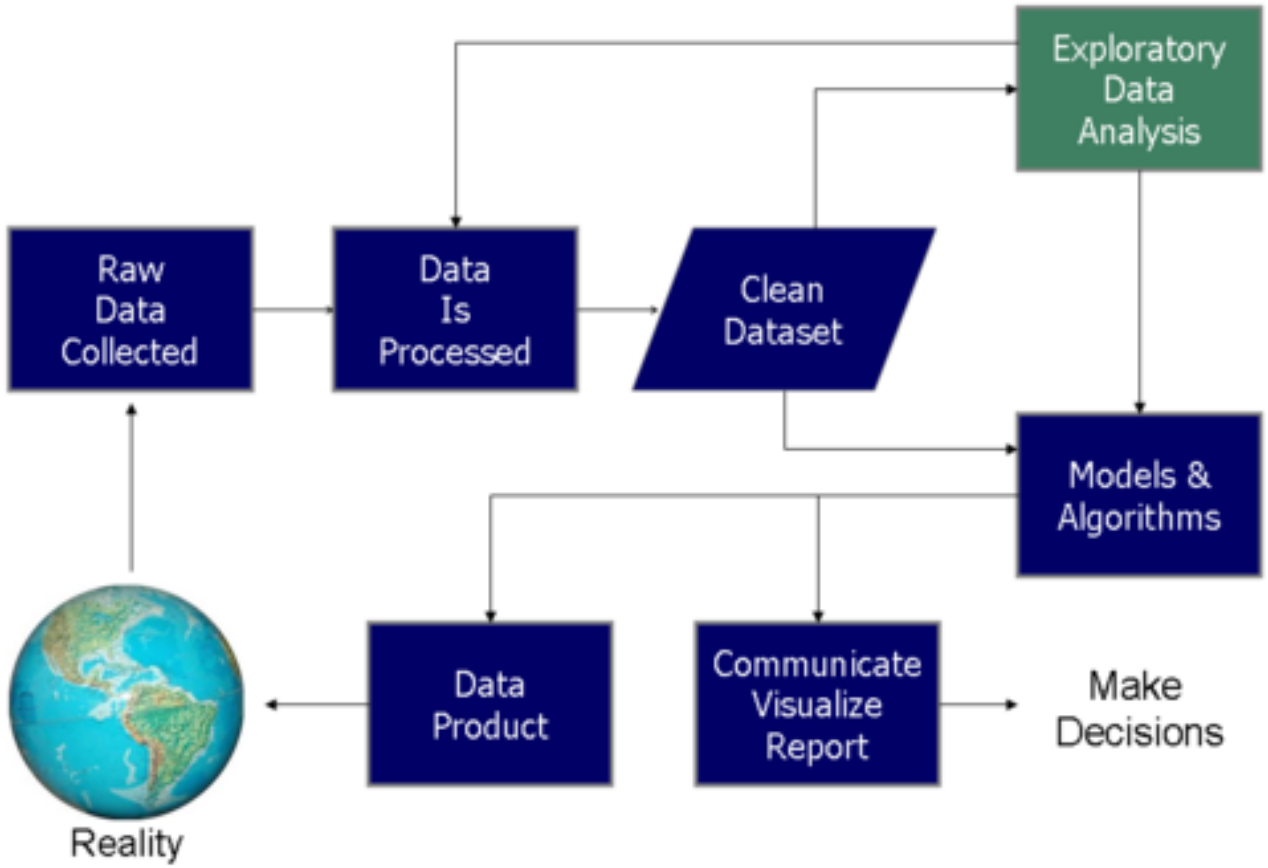
My Notes

Research designs

- Name the study design per se – that is, 'cross-sectional', 'before-and-after', 'comparative', 'control experiment' or 'random control' (Chapter 8 describes some of the commonly used study designs).
- Provide detailed information about the following aspects of the study (details of these are covered in the subsequent chapters of the book):
    - Who will constitute the study population?
    - How will the study population be identified?
    - Will a sample or the whole population be selected?
    - If a sample is selected, how will it be contacted?
    - How will consent be sought?
    - What method of data collection will be used and why?
    - In the case of a questionnaire, where will the responses be returned?
    - How should respondents contact you if they have queries?
    - In the case of interviews, where will they be conducted?
    - How will ethical issues be taken care of?

My Notes

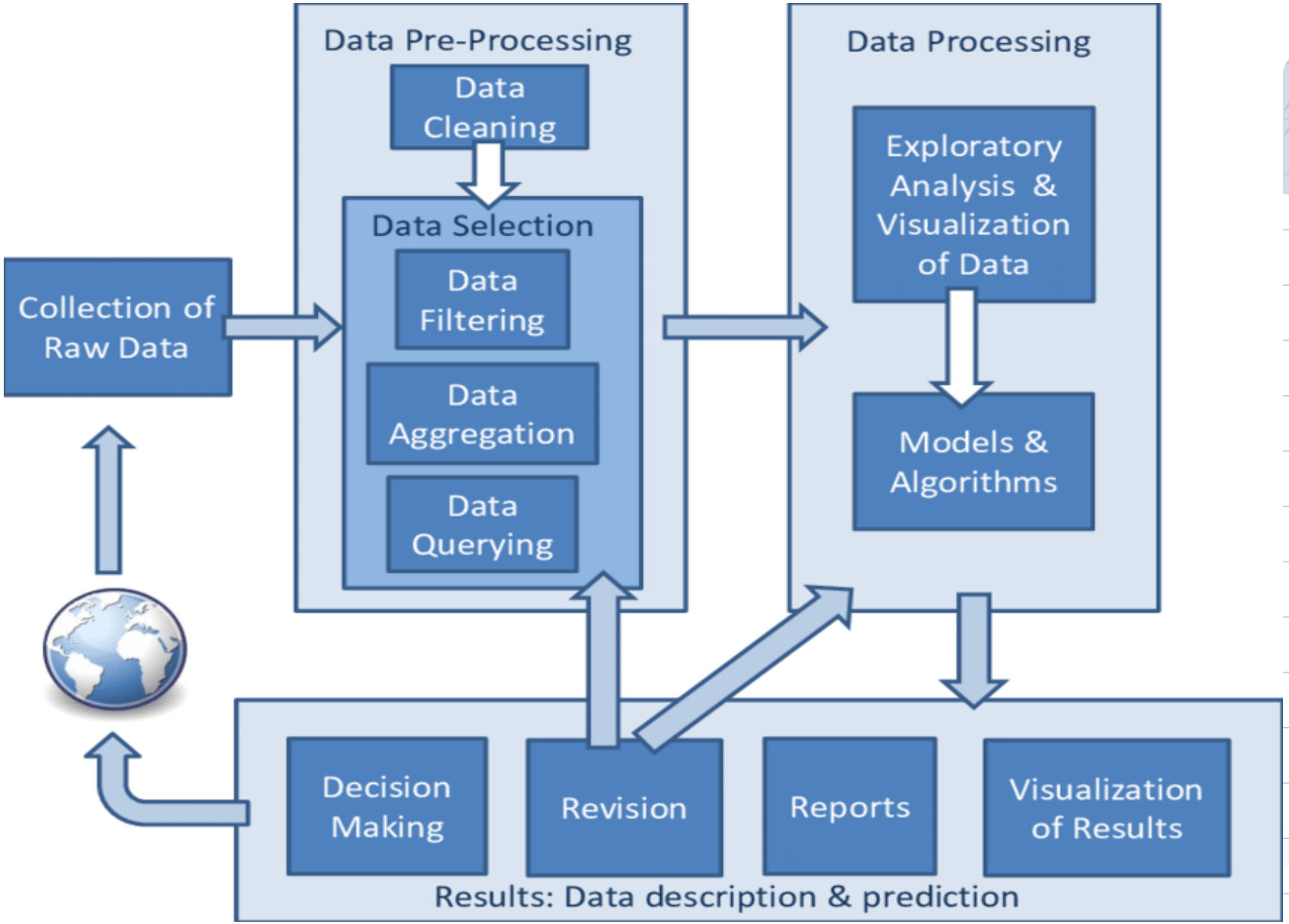
What is a research design in data science?



https://www.kdnuggets.com/2016/03/data-science-process.html

My Notes

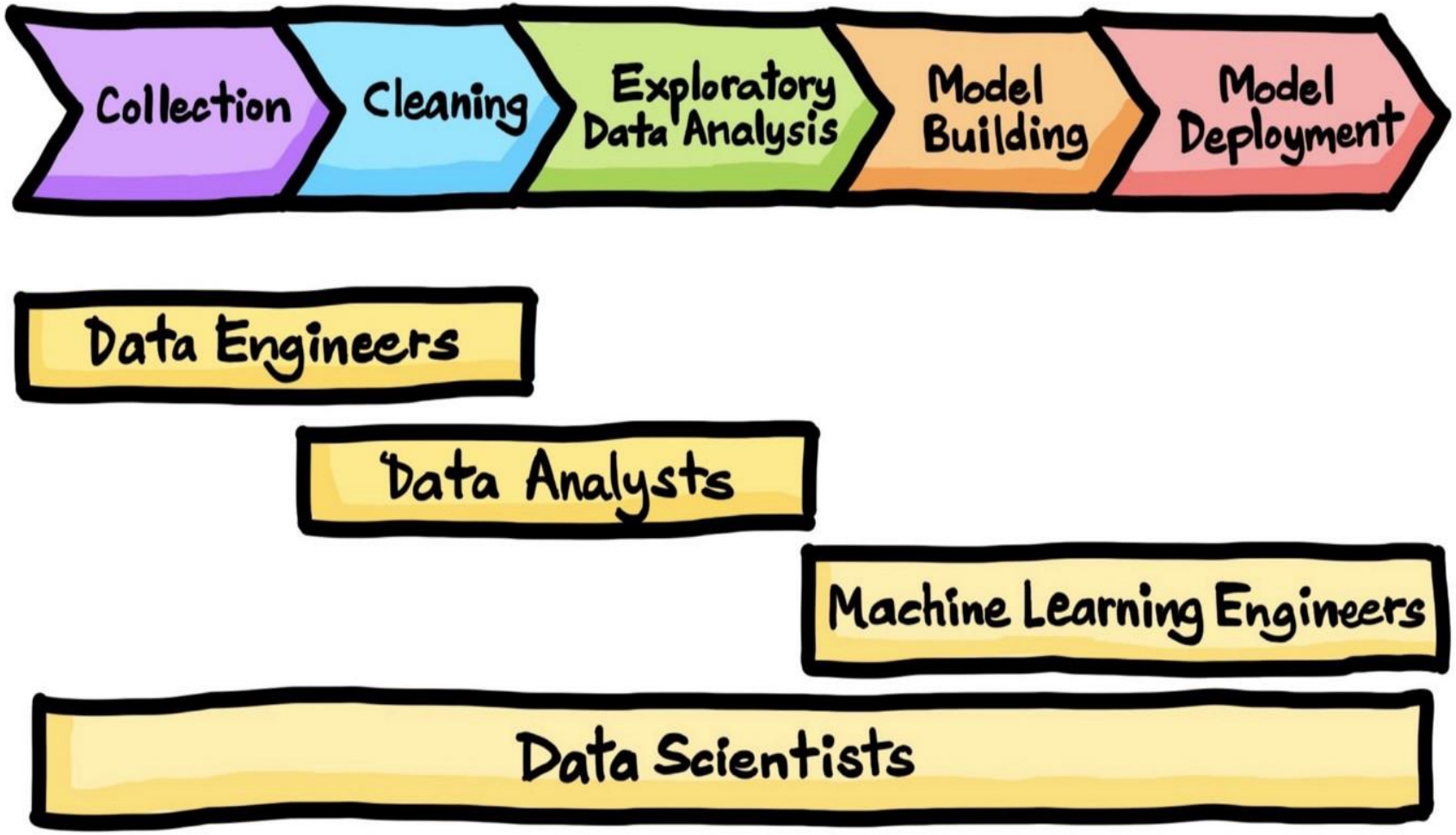
~Is it a sequential process?



My Notes

~Is it a sequential process?

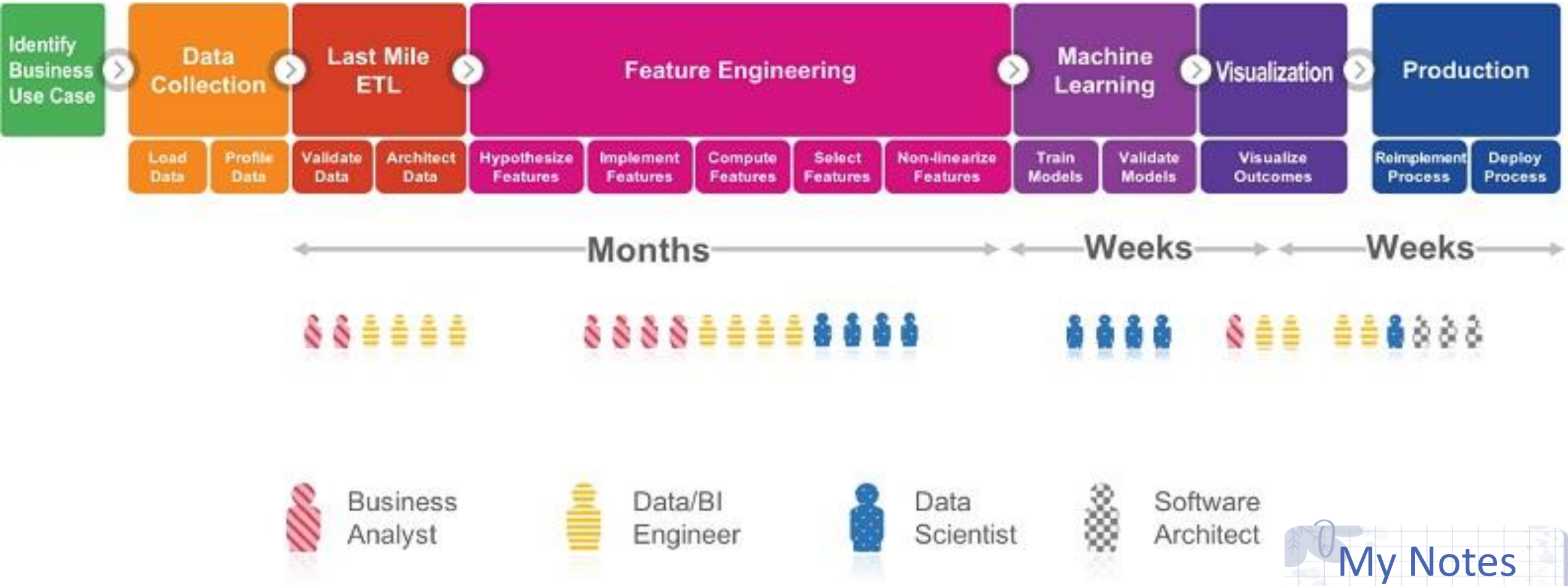
Molina-Solana, Miguel & Ros, María & Ruiz, M. & Gómez-Romero, Juan & Martin-Bautista, Maria. (2017). Data science for building energy management: A review. Renewable and Sustainable Energy Reviews. 70. 598-609. 10.1016/j.rser.2016.11.132.



https://towardsdatascience.com/the-data-science-process-a19eb7ebc41b

My Notes

~Who did what?

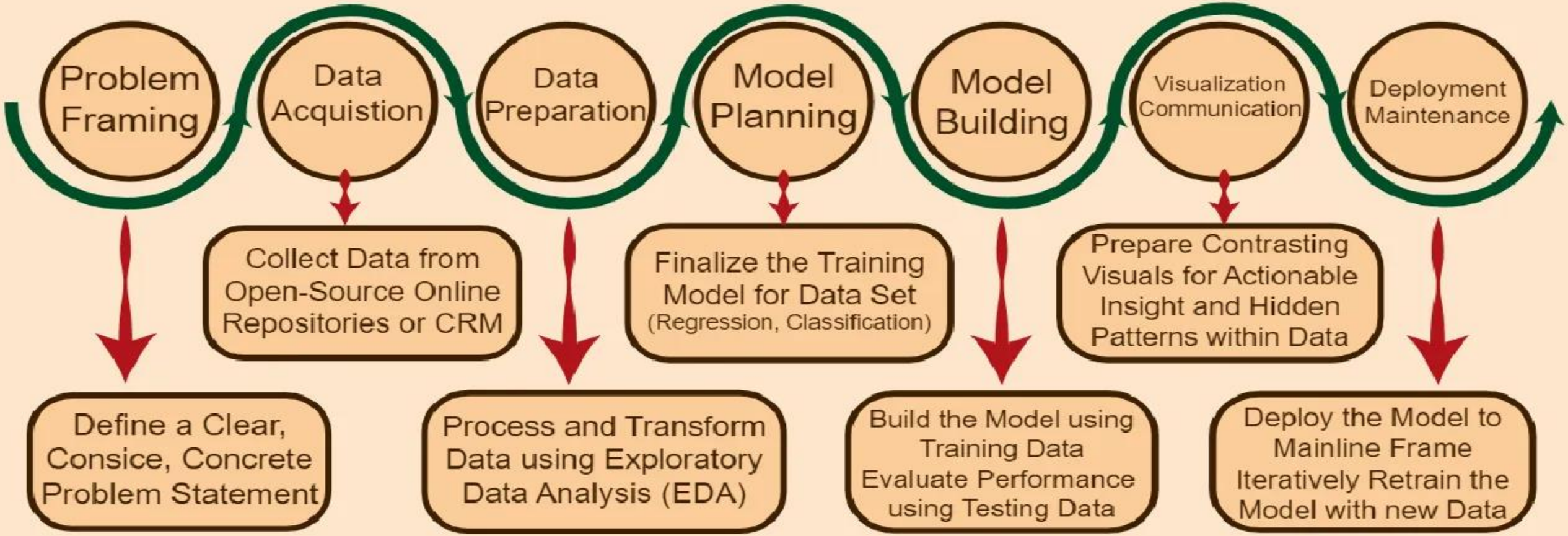


[How and Why Your Enterprise Should Democratize Data Science | Transforming Data with Intelligence (tdwi.org)]

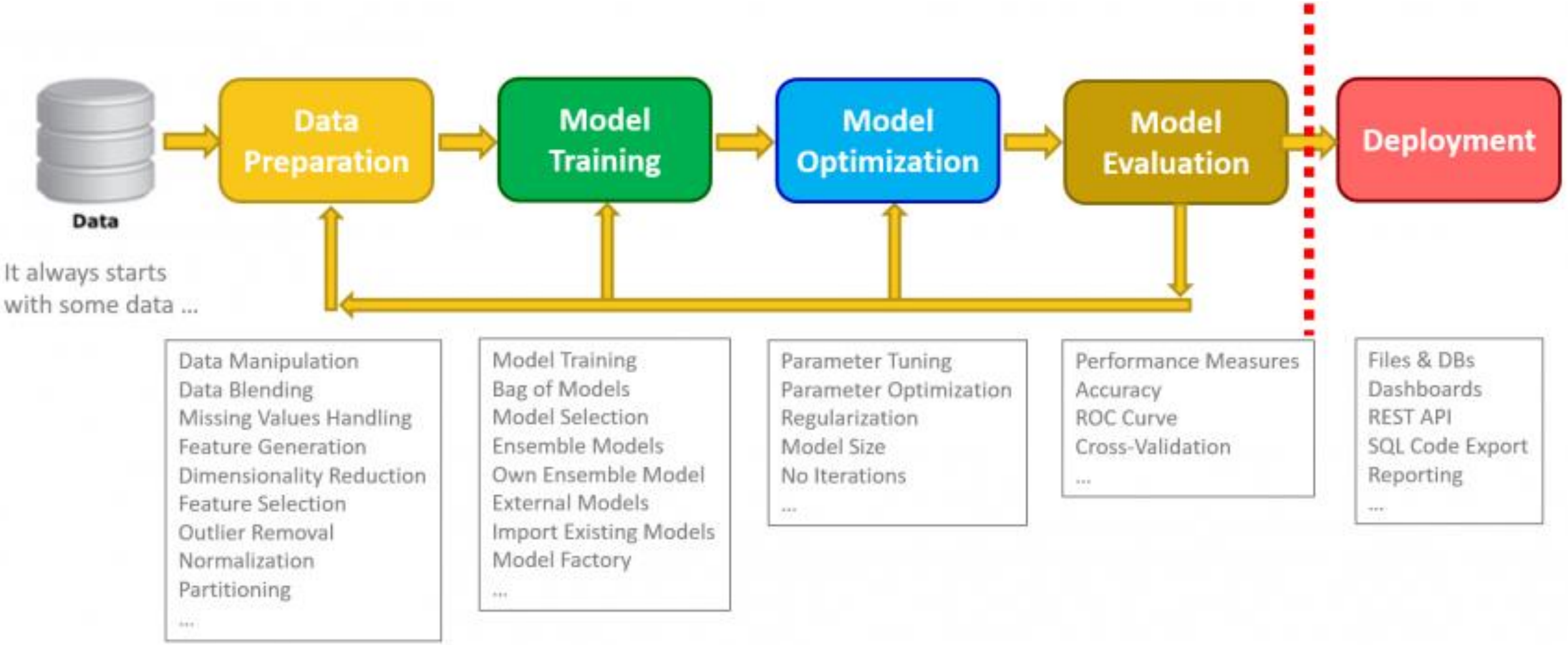
My Notes

~Who did what? And for how long?

DATA SCIENCE PROCESS

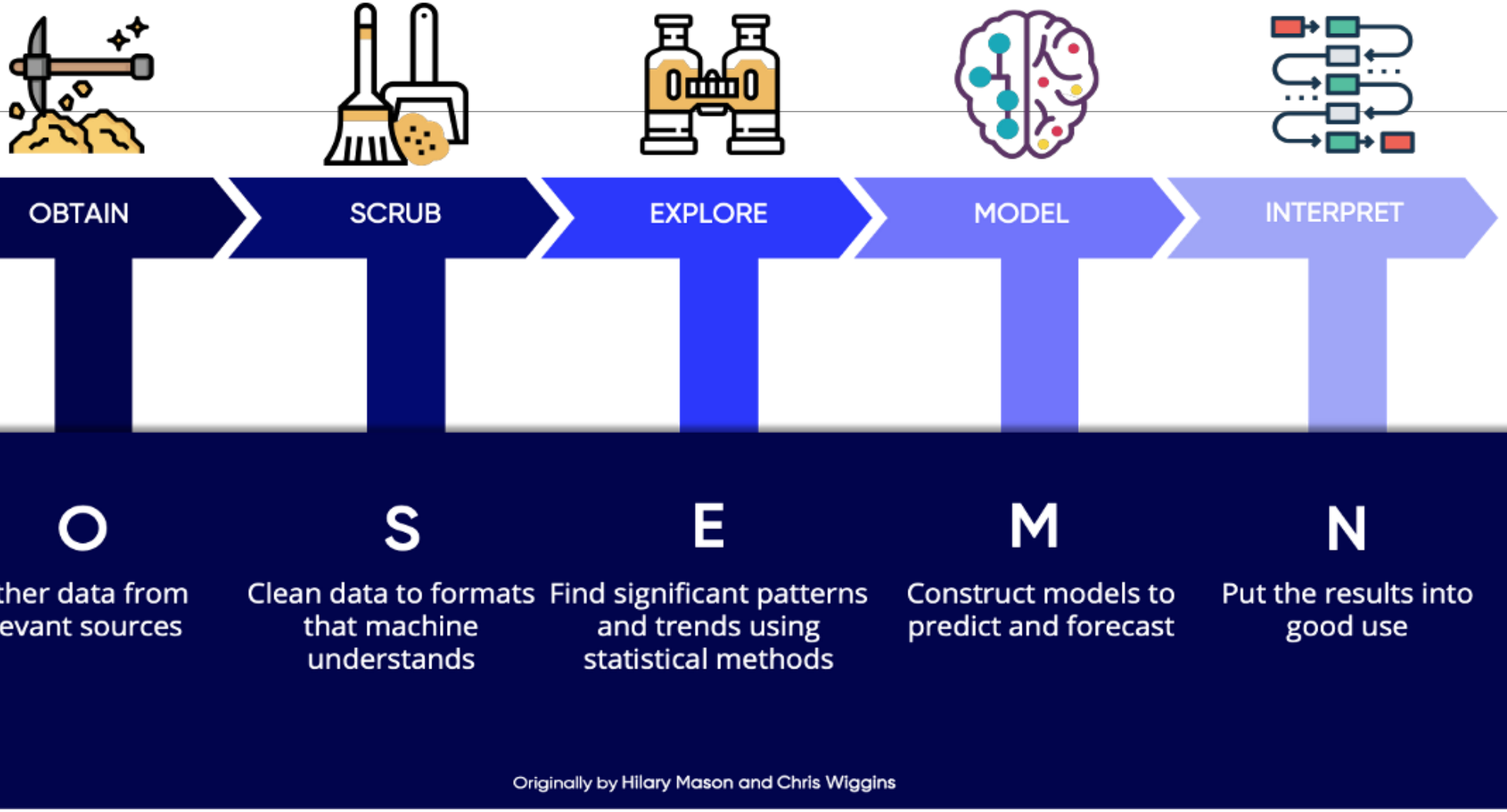


https://morioh.com/p/f805a21030f2



[Best Practice for Data Science Projects | by Murat Yalcin | Towards Data Science]

Data Science Process

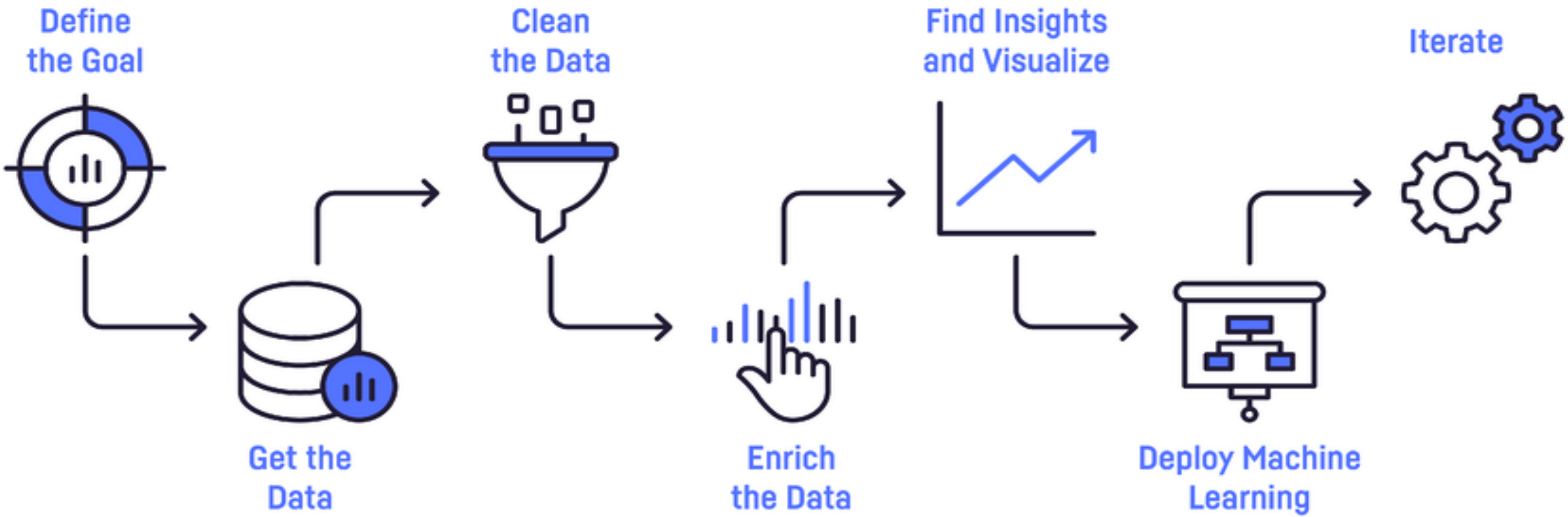


[5 Steps of a Data Science Project Lifecycle | by Dr. Cher Han Lau | Towards Data Science]

7 Fundamental Steps to Complete a Data Analytics Project

July 4, 2019

Data Basics | Alivia Smith



https://blog.dataiku.com/2019/07/04/fundamental-steps-data-project-success

Overview of the five steps in data science

Data science follows *a structured, step-by-step process that, when followed, preserves the integrity of the results*.

The five essential steps to perform data science are as follows:

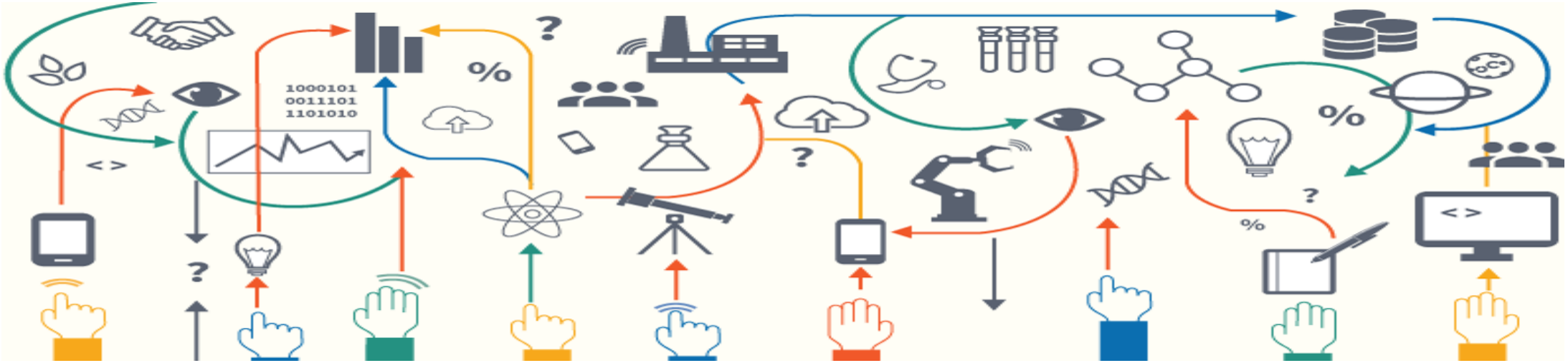
1. Asking an interesting question
2. Obtaining the data
3. Exploring the data
4. Modeling the data
5. Communicating and visualizing the results

1. Ask an interesting question

- Treat this step as you would treat a **brainstorming** session.
- Start writing down questions regardless of whether or not you think the data to answer these questions even exists.
- The reason for this is twofold:
  - Firstly, don't be bias before searching for data.
  - Secondly, obtaining data might involve searching in both public and private locations and, therefore, might not be very straightforward.

2. Obtain the data

- Once you have selected the question you want to focus on, it is time to scour the world for the data that might be able to answer that question.
- As mentioned before, the data can come from a variety of sources; so, this step can be very creative!



3. Explore the data

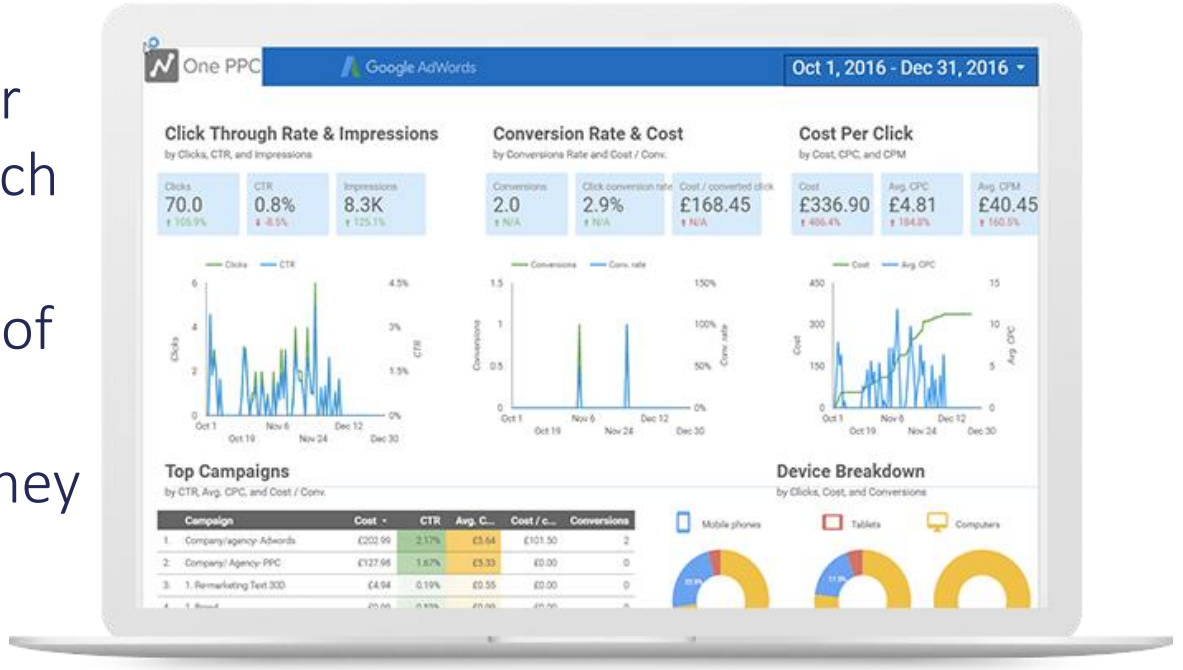
- Once we have the data, we use the lessons learned in Chapter 2, Types of Data, and begin to break down the types of data that we are dealing with.
- This is a **pivotal step** in the process.
- Once this step is completed, the analyst generally has spent several hours learning about the domain, using code or other tools to manipulate and explore the data, and has a very good sense of what the data might be trying to tell them.

4. Model the data

- This step involves the use of **statistical and machine learning models**.
- In this step, we are not only fitting and choosing models, but we are also implanting **mathematical validation metrics** in order to quantify the models and their effectiveness.

5. Communicate and visualize the results

- This is arguably the **most important step**.
- While it might seem obvious and simple, the ability to conclude your results in a digestible format is much more difficult than it seems.
- We will look at different examples of cases when results were communicated poorly and when they were displayed very well.



Watch this video: Experimental Designs | Data Science Summer School (4 hours)

@SalhaAlzahrani