Discussion 1 : Problem Formulation

Background

There are SEVEN stages in the Data Science Pipeline, as you have already seen. Each stage consists of two modules -- the SCIENCE Module and the DATA Module. You can't consider one without the other; they are tied to each other in practice.

You can't draw upon a Practical Motivation unless you plan the Data Collection exercise too.

You can't formulate a Data Science problem without considering the Data Preparation aspect.
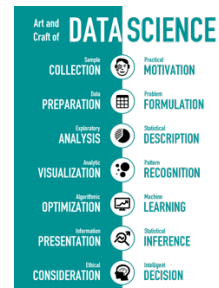
You can't arrive at a concise Statistical Description unless you have done Exploratory Analysis.

You can't Recognize Patterns/Structure in data without proper Analytic Visualization in place.

You can't blindly use Machine Learning unless you understand the Algorithmic Optimizations.

You can't just draw Inference from the data without being able to Communicate it well enough.

You can't of course take an Intelligent Decision unless you also consider its effects Ethically.

Discussion : Question 1

There are five basic questions Data Science can answer. Recap from LAMS, as follows.

- o How much? How many? – answered through Numeric Predictions, or Regression
- o Is it Type A or Type B? – answered through Prediction of Class, or Classification
- o How is this organized? – answered through Detection of Structure, or Clustering
- o Is it a weird behavior? – answered through Detection of Anomalies in given Data
- o What should be done next? – answered in AI through Adaptive Learning process

Provide ONE practical example for each of these five cases. You can cite examples from your academic discipline, practical experience, or simply from your day-to-day life. Justify why the example belongs to a certain case, as above.

Discussion : Question 2

Suppose that your parents (or some close relatives) are planning to buy a Condo in Singapore. Devise a strategy to judge for them if the price quoted by the Seller is expected, higher than expected, or lower than expected, for such a Condo.

Note : You must use Data Science to approach this problem. What type of data will you need and what will be the exact type of Data Science problem that you are planning to solve in this case? Discuss with your friends and present your view.

Discussion : Question 3

Suppose that your friend, majoring in Biology, wants to know how many types of Whales are there in the South China Sea during a specific time of the year. Devise a strategy to help him/her in identifying the types of Whales in this case.

Note : You must use Data Science to approach this problem. What type of data will you need and what will be the exact type of Data Science problem that you are planning to solve in this case? Discuss with your friends and present your view.