

Text Data Mining For Business Decisions

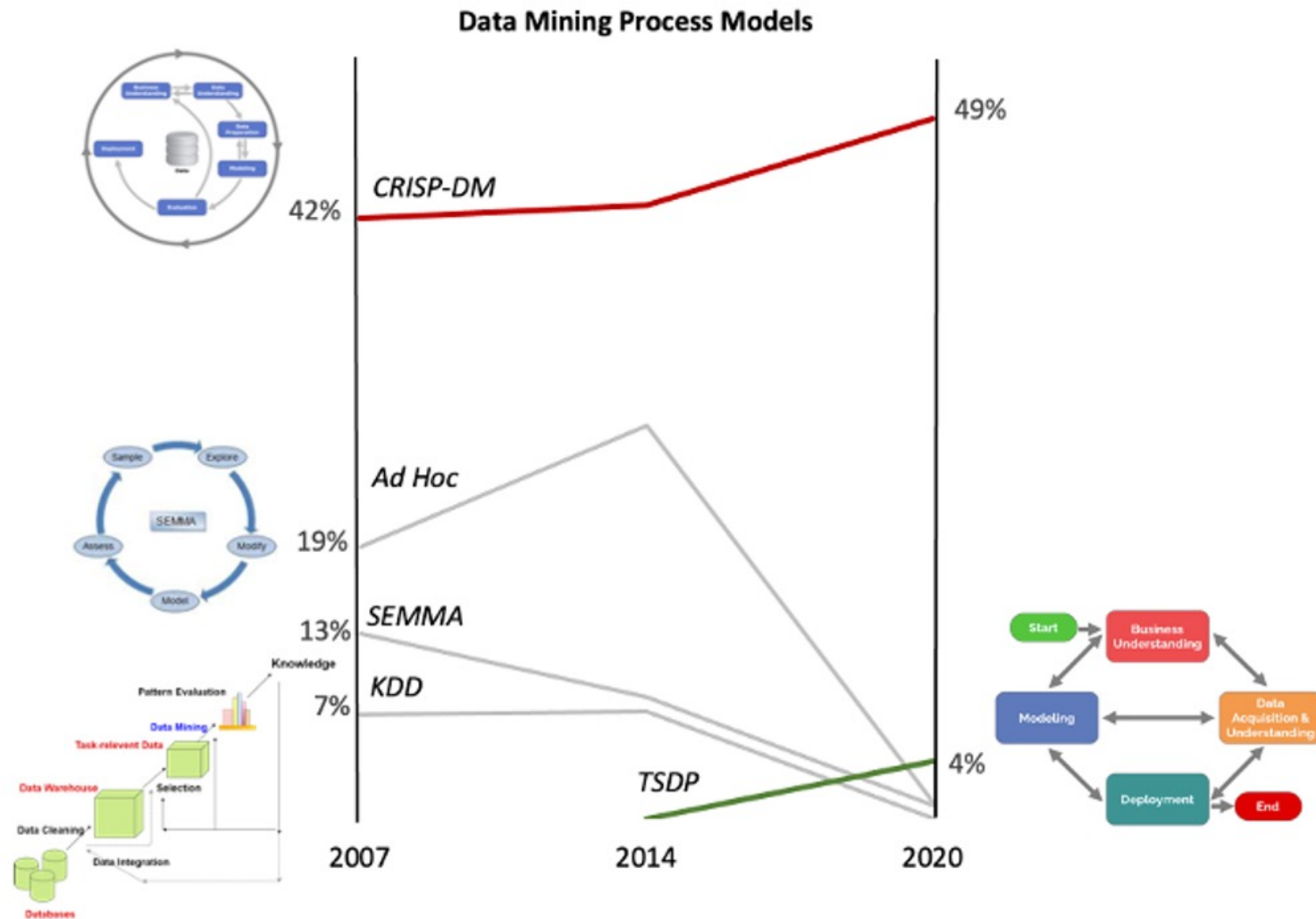
Module 2

Framing Analytical Questions

How to frame good analytical questions?

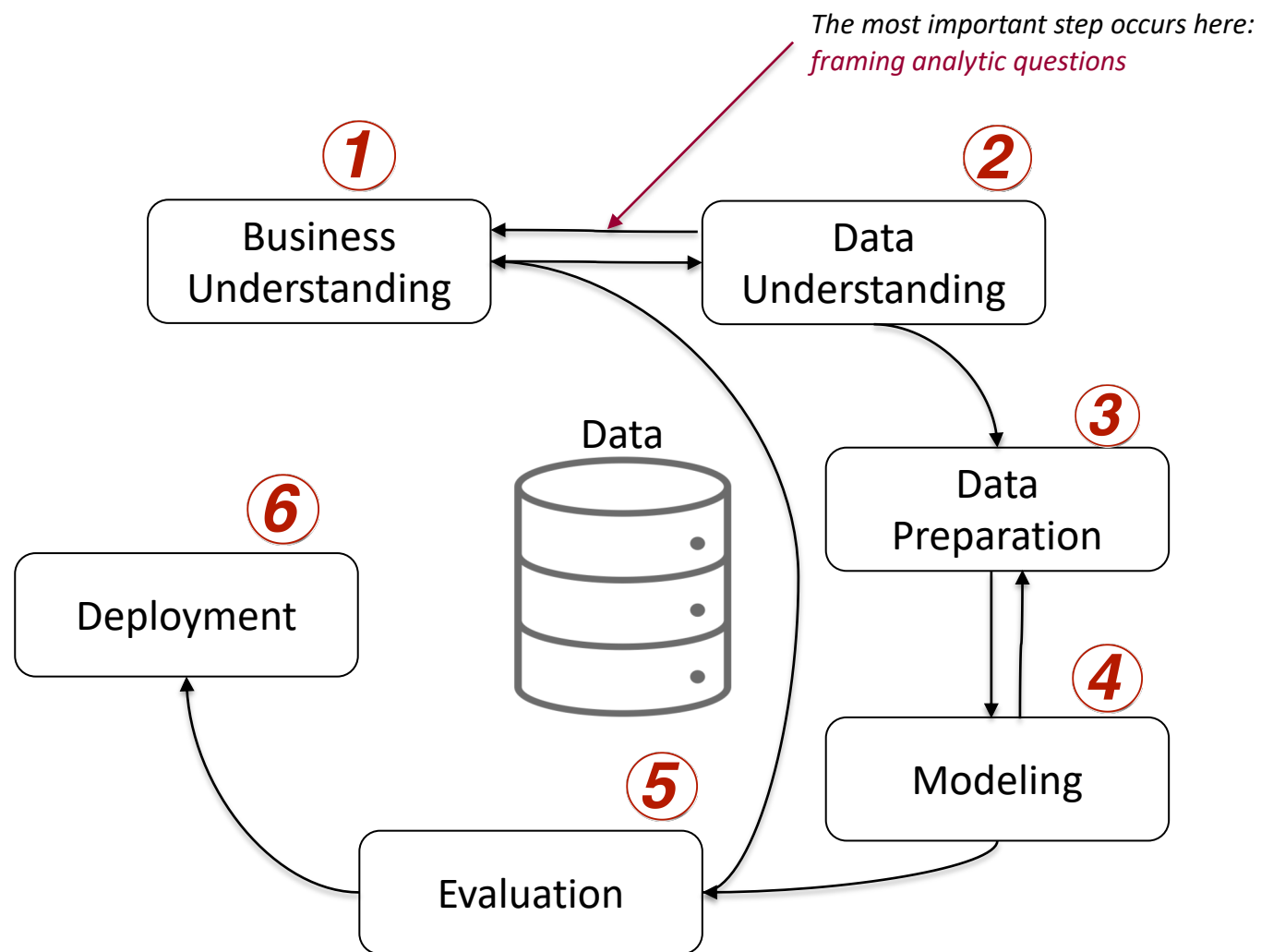


CRISP-DM is the Leading Process Methodology



CRISP-DM

Cross
Industry
Standard
Process for
Data Mining



The Critical Step in Text Data Mining is: Framing The Question

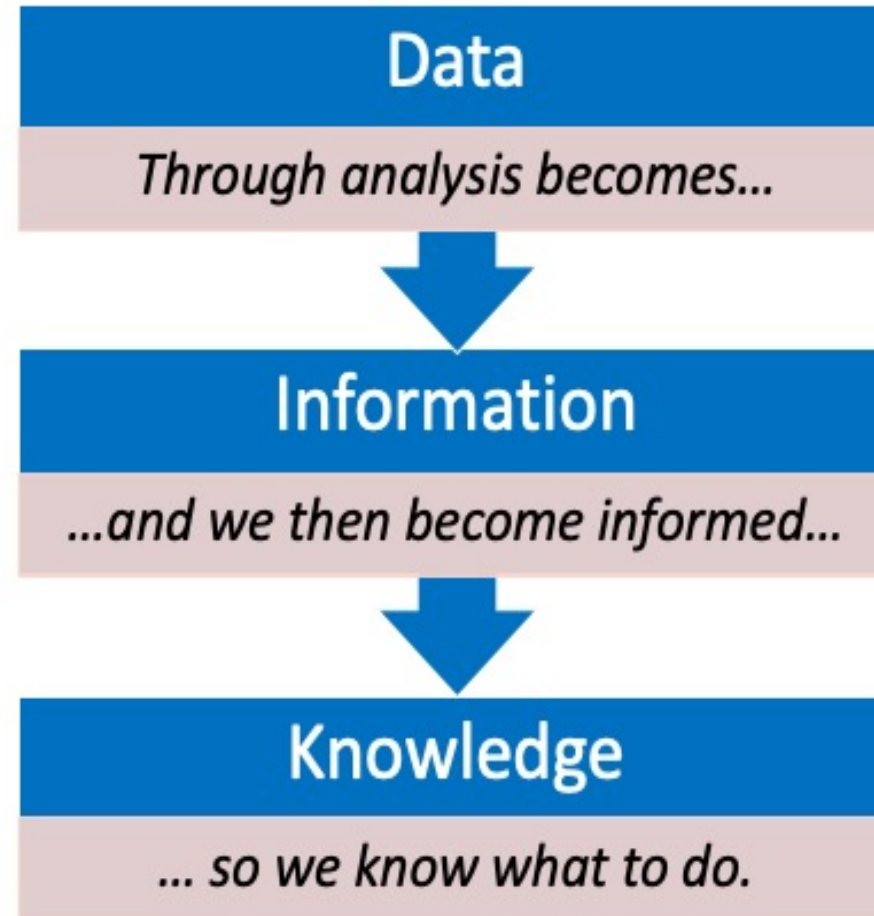
- Framing the data mining question and understanding the context.
- The data mining process starts with an analysis need, a question or a business objective.
 - This is possibly the most important step in the data mining process
 - Without a *well-defined statement of the problem*, it is impossible to select the right data set and pick the right data mining algorithm.

Framing the Question

- The Center for Data Insight has extended CRISP-DM with this most important step:
- Framing the Question(s) –
 - “One of the common misconceptions about data mining is that one can blindly set the algorithms loose on the data to find all interesting patterns that may be present.
 - “Data mining is not a magic panacea for curing all business ills. Rather it is a decision support tool that, when used in conjunction with business understanding, can provide a valuable means of gaining new business insights.
 - “Therefore, the *first step* in the KDD cycle must be to *determine between one and three questions* or goals to help direct the KDD focus”. (KDD= knowledge discovery in databases, another name for data mining)

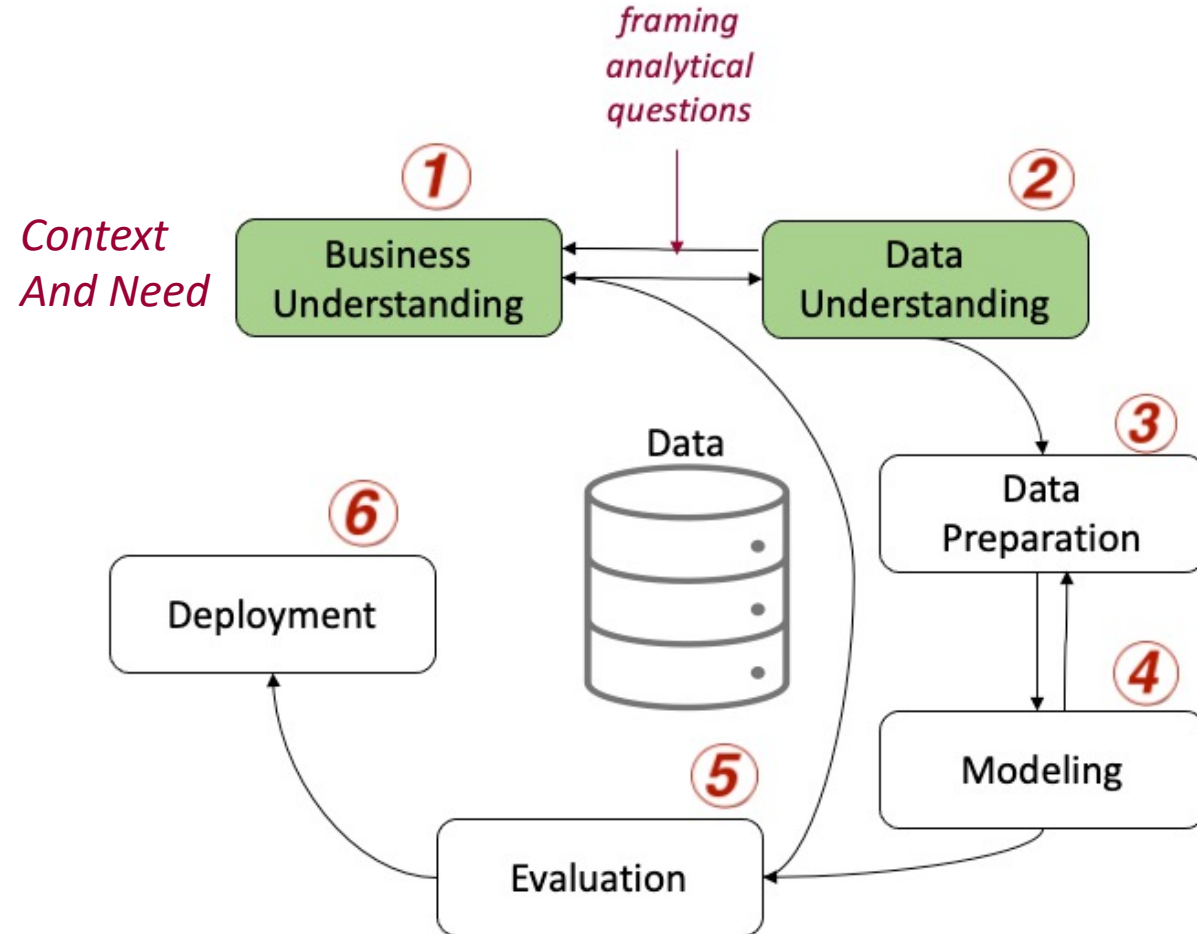
How Does Data Become Knowledge?

- **Data** – pieces of information, measurements, transactions, opinions, items we collect into a database and becomes the basis for answering questions
- **Information** – once we *analyze* data by answering questions and drawing conclusions, we *become informed*
- **Knowledge** – we can now make informed choices as to what to do – make recommendations on taking action, *we know what to do*

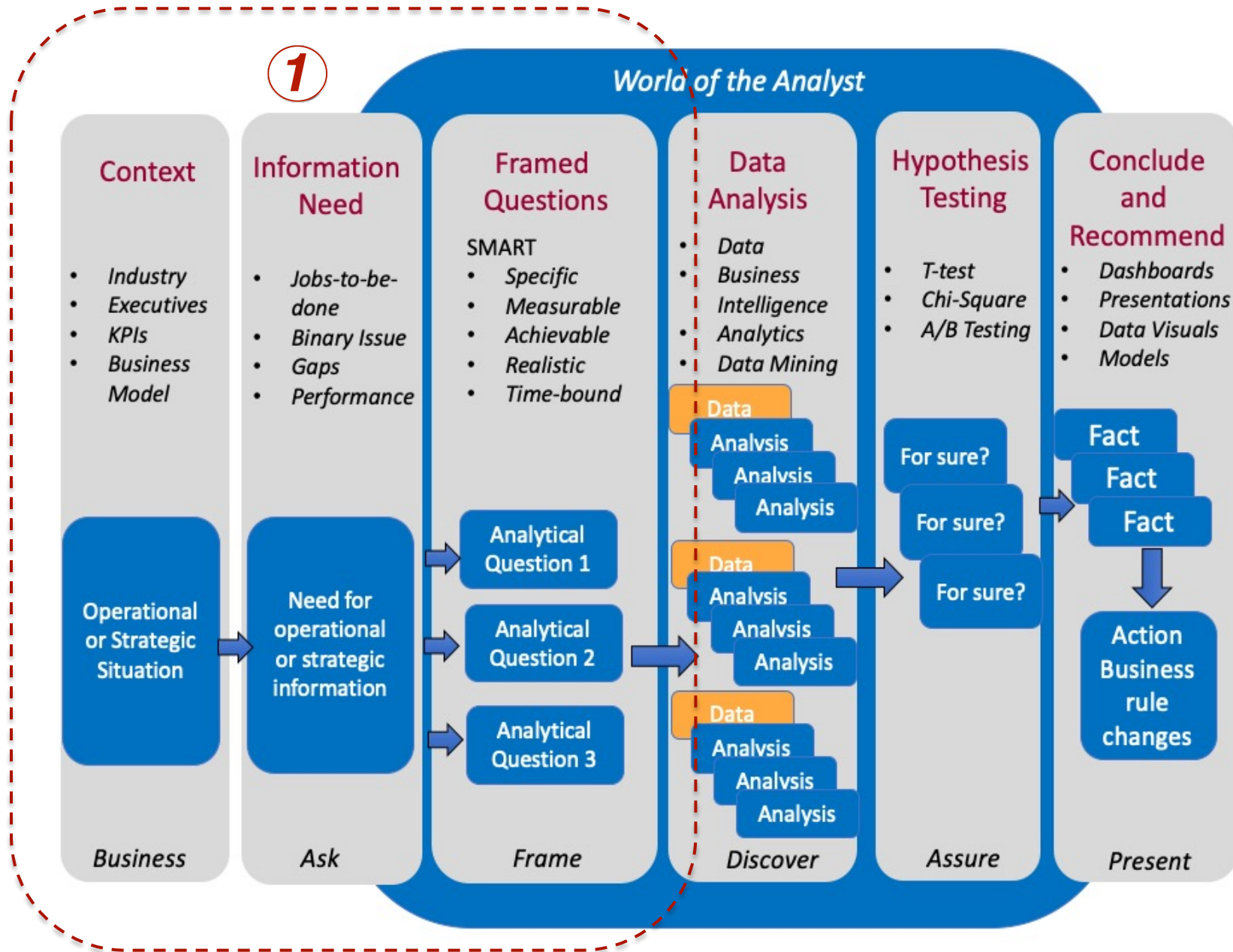


Where does this analysis fit in the Data Mining Process?

CRISP-DM
Cross
Industry
Standard
Process for
Data Mining



Step 1 of the CRISP-DM Model



World of the Analyst

Information Need

- *Jobs-to-be-done*
- *Binary Issue*
- *Gaps*
- *Performance*

Need for
operational
or strategic
information



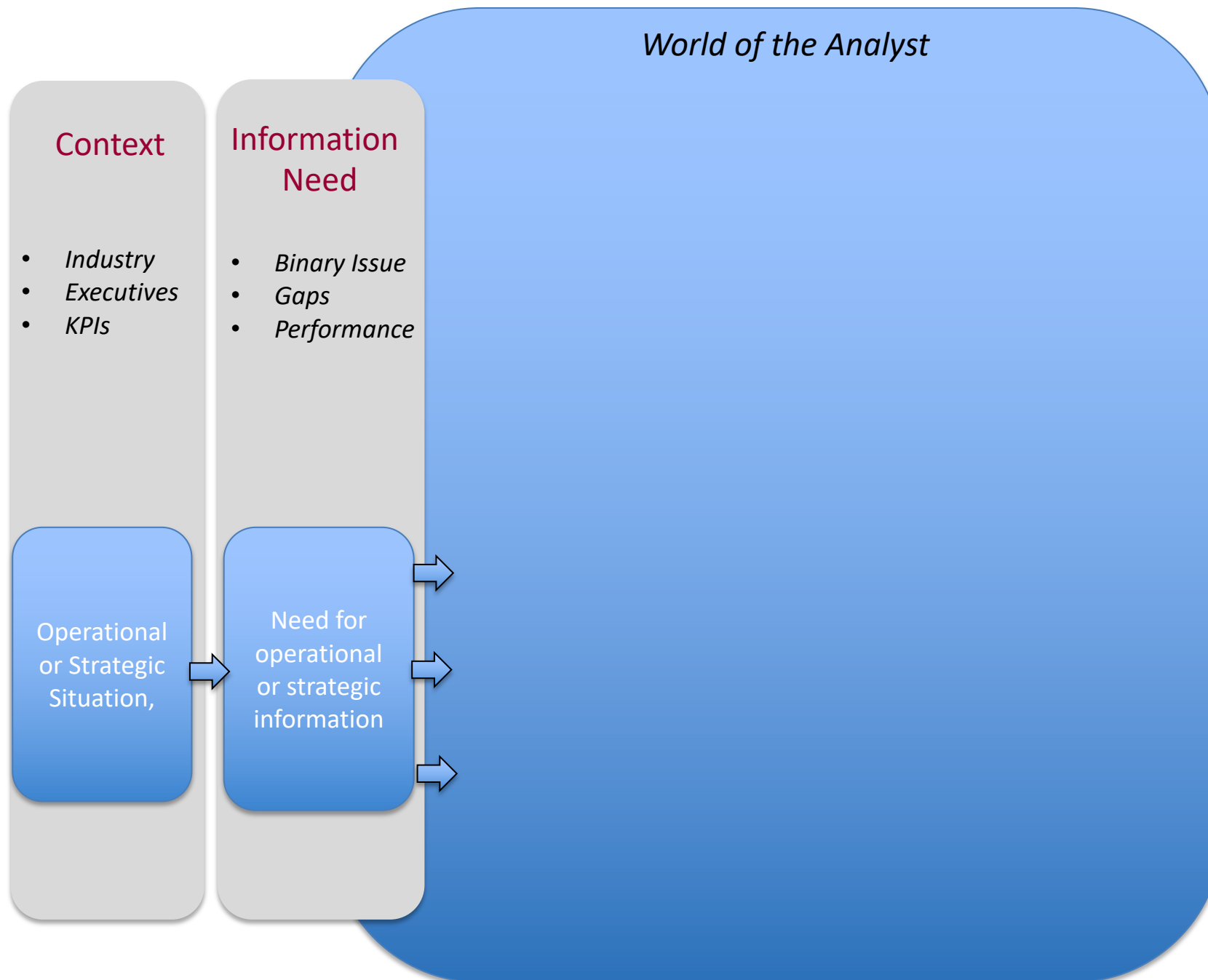
Business Understanding:

What is the Information **Need**?

- Everyone faces challenges.
- Things that, were they to be fixed or understood, would advance the goals they want to reach.
- What are the specific needs that could be fixed by intelligently using data?
- These needs should be presented in terms that are meaningful to the organization.
- If our method will be to build a model, the need is not to build a model.
- The need is to solve the problem that having the model will solve.
- Correctly identifying information needs is tough.

How do we Recognize an Information Need?

- Needs are often expressed by the client as a need to know something that affects them that they don't know what to do to fix
 - sometimes there is a problem
 - Often there is a gap in performance
 - Sometimes it comes as a question
 - Sometimes it is a statement of need
- It must be parsed – taken apart for meaning
- Must discover what is the **KPI** and the **Gap** in performance driving the **Need** to know



Business Understanding: Discovering the **Context** of Project

- Every project has a context, the defining frame that is apart from the particular problems we are interested in solving.
 - Who are the people with an interest in the results of this project?
 - What are they generally trying to achieve? The numbers they are trying to make?
 - What work, generally, is the project going to be furthering?

What are some of the questions to help us discover the context?

- Context analysis questions
 - What is the industry this organization is in? What are some of the pressures the kinds of organizations like this have in their industry? Why do they need to perform better?
 - Who are the people in the organization with an interest in the results of this project? Executives? Who will you be reporting results to?
 - What are they generally trying to achieve?
 - What work, generally, is the project going to be furthering?

Identify the KPI

- Key business performance indicators have often been identified for the team or business unit
- They may be part of a balanced score card formulation
- Performance is often expressed as
 - Increase...
 - Minimize...

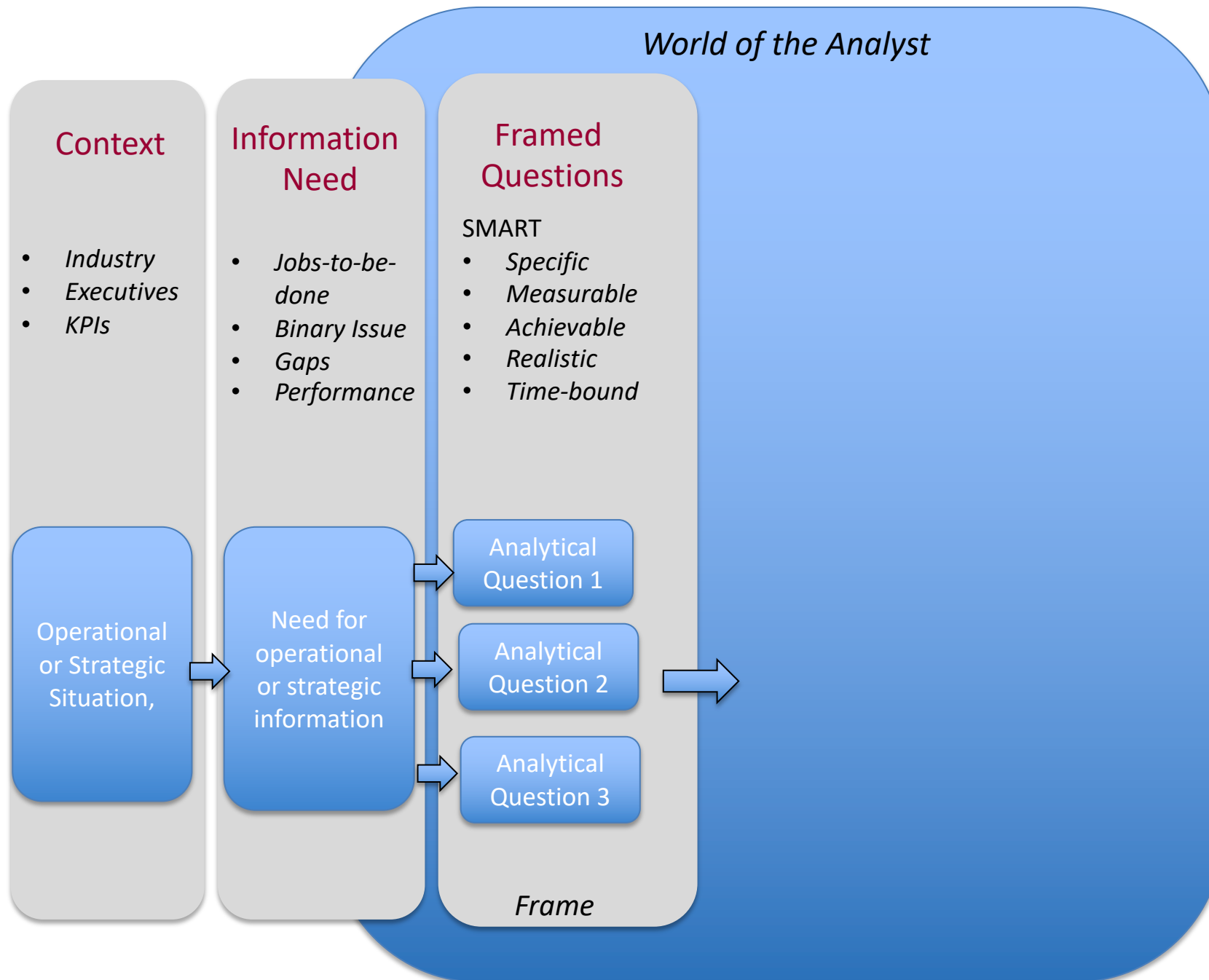
The Gap Question

- The Gap = the difference between where are we today on a KPI and where do we need to be at the end of the measurement period, our goal.
- We must identify the gap that must be resolved to satisfy the information need

Do Your Homework

- What is the information need?
- What is the context of the assignment?
- What are the KPIs and the gaps?

THE CRITICAL STEP IN ANALYTICS: FRAMING ANALYTICAL QUESTIONS



First, Dig Deeper

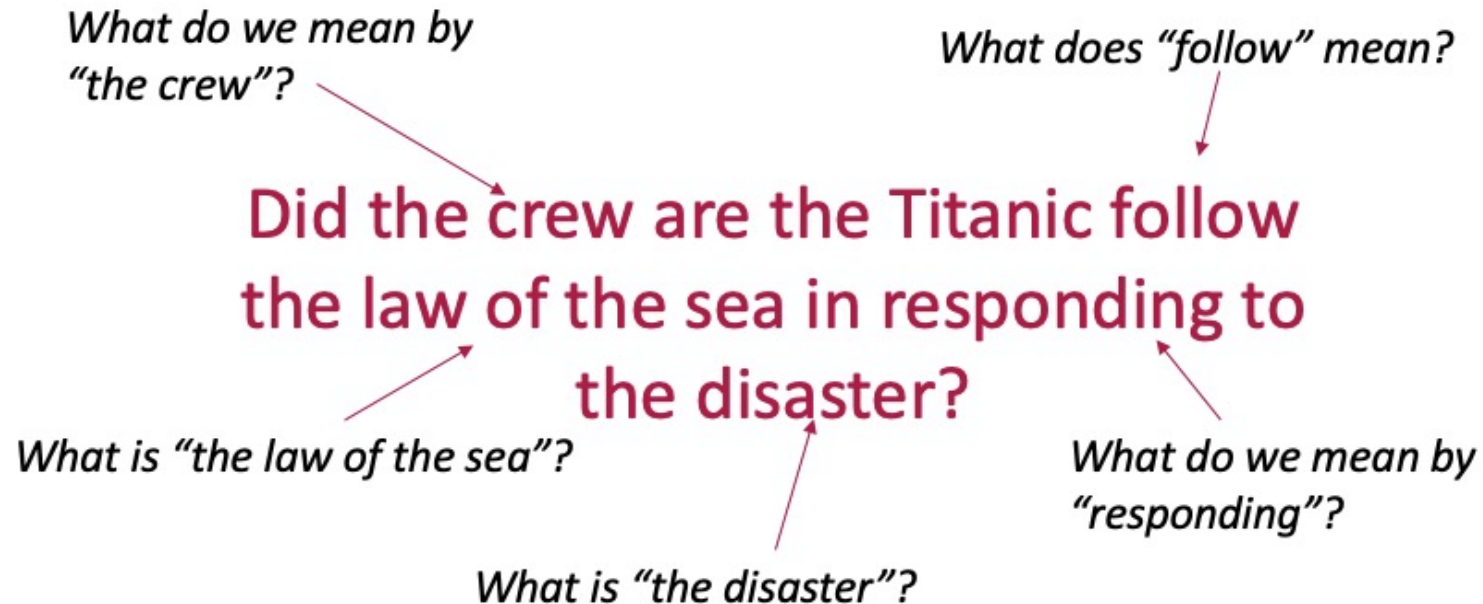
- Investigate the information need. Why do they need to know that?
- What job are they trying to accomplish by knowing that better?
- Usually executives and managers are driven by corporate/business unit/team goals that need to be accomplished – they are the responsible ones
- What are those goals – “their numbers”?
- What are the *KPIs* and CSFs and the metrics used to know they are achieved
- These are driving the *need to know* the *Gap* between *what is* and *what must be*
- Identify what has to be *increased*, what has to be *minimized*?

Then *Parse* the Information Request

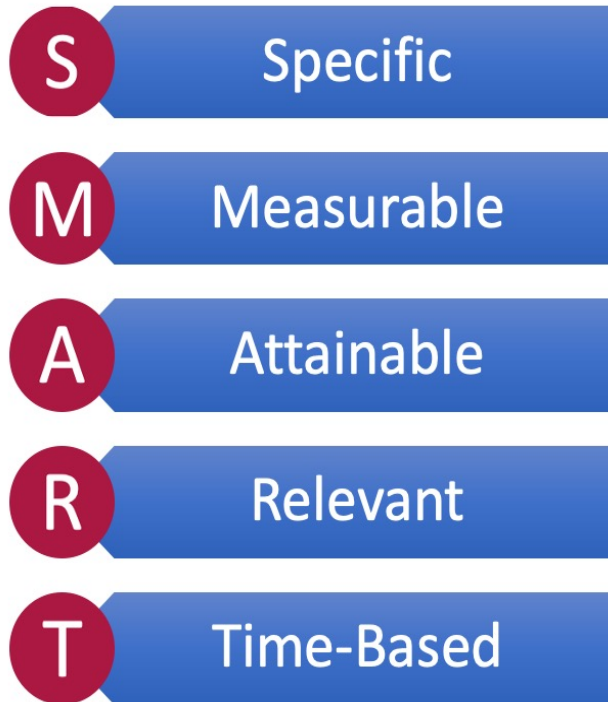
- Break the request down into meaningful elements
- Identify the central job they are trying to do as well as some of the major tasks
- Identify the timeframe for informing as well as the nature of the requester (position, level of responsibility, organizational context, department)

Parsing the information need and three framed analytical questions

Parsing the Information Need



Use SMART to Frame High-Quality Questions



- *Specific*
the framed question must be focused and detailed.
- *Measurable*
the framed question must be *computable*.
- *Attainable*
the framed question must be able to be answered by the techniques known to the analyst who will do the analysis.
- *Relevant*
the answers to framed question must apply to the business.
- *Time-related*
some element of time should be considered in the analysis

These are the Key Step In Framing Questions

1. Start with the need question and say to yourself:
To be able to answer that I will need to know...
2. Come up with as many questions as you can think of
3. Prioritize the questions
 - Can you answer it in the time given?
 - If you don't answer it, will you be able to still give a good enough answer? If yes, the ignore it, if not prioritize it
 - Do you have the data to answer it?
 - Do you know how to answer it yourself?
4. Reduce the list to the *top three questions* to start your analysis.