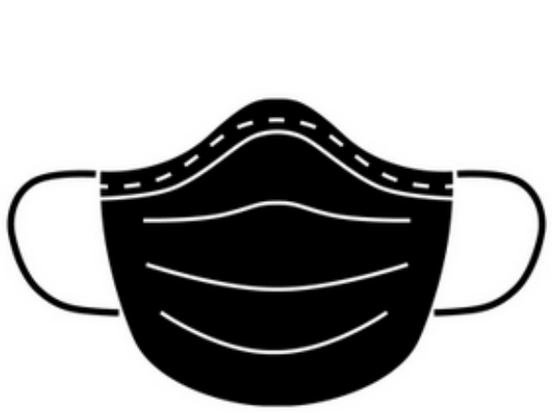
CS490DSC Data Science Capstone
Business Understanding

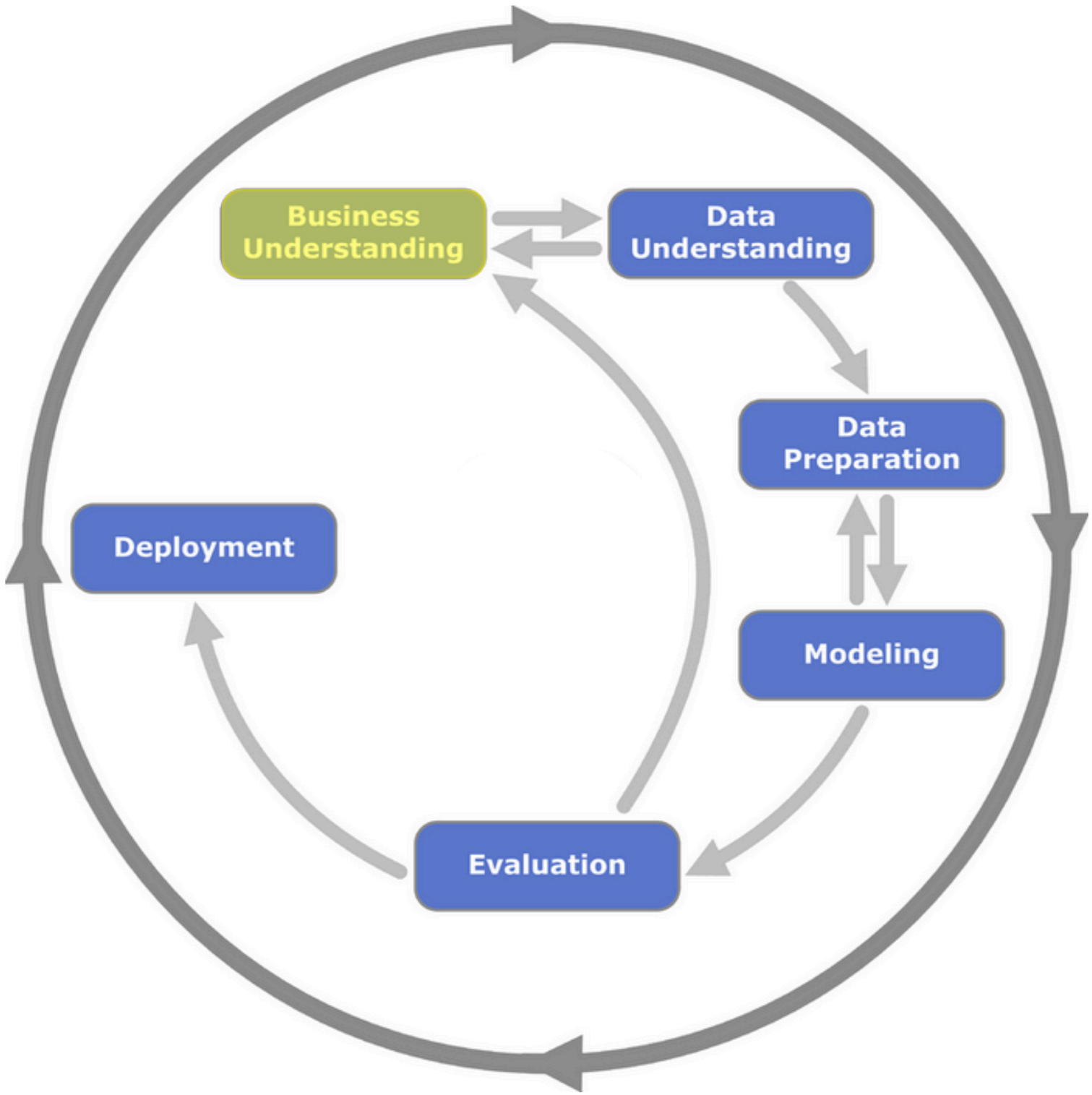
Jean Honorio
Purdue University



Important

- Please read this together with the case study
- The case study will discuss a fictitious health insurance company called the Amazing Health Network

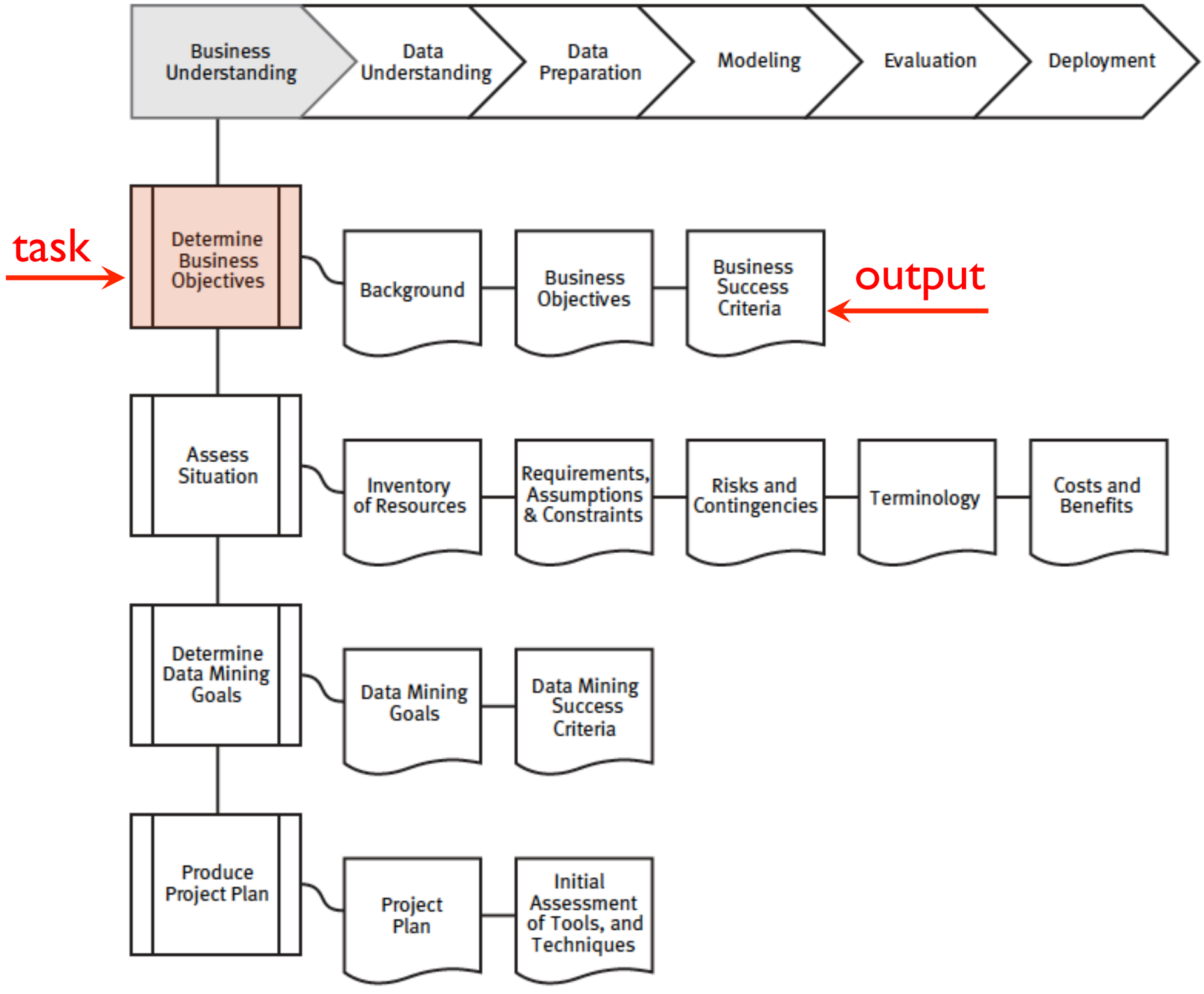
CRISP-DM


Business Understanding
Data Understanding
Data Preparation
Modeling
Evaluation
Deployment

Phase 1: Business understanding

- We use the term "business" to refer to, for instance
  - a company, a corporation
  - a research group in a university
- This phase focuses on understanding the
  - project objectives
  - requirements from a business perspective
- Then converting this knowledge into
  - a data mining problem definition
  - a preliminary plan designed to achieve the objectives

Phase 1: Business understanding



1. Determine business objectives

- The first objective of the data analyst is to understand, from a business perspective, what the customer really wants to accomplish
- Often the customer has many competing objectives and constraints that must be properly balanced
- The analyst's goal is to uncover important factors, at the beginning, that can influence the outcome of the project
- **A possible consequence of neglecting this step is to expend a great deal of effort producing the right answers to the wrong questions**

1.1. Background

- Record the information that is known about the organization's business situation at the beginning of the project
  - Organization
  - Problem area
  - Current solution

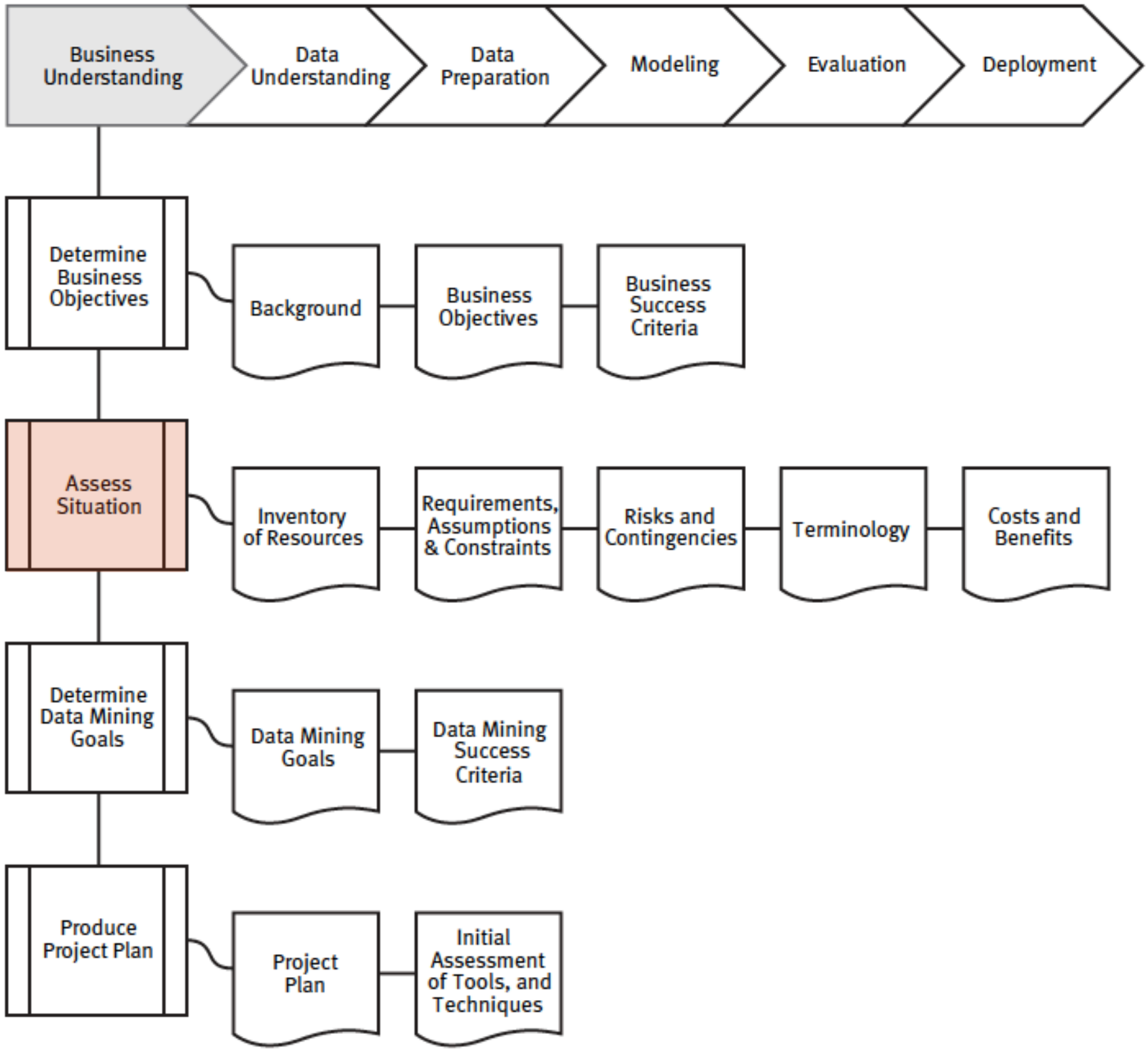
1.2. Business objectives

- Describe the customer's primary objective, from a business perspective
  - In addition to the primary business objective, there are typically other related business questions that the customer would like to address
- For example, the primary business goal might be to
  - keep current customers by predicting when they are prone to move to a competitor
- Examples of related business questions are
  - "How does the primary channel used (e.g., ATM, branch visit, Internet) affect whether customers stay or go?"
  - "Will lower ATM fees significantly reduce the number of high-value customers who leave?"

1.3. Business success criteria

- Describe the criteria for a successful or useful outcome to the project from the business point of view
- This might be quite specific and able to be measured objectively, for example, reduction of customer churn to a certain level
  - Customer churn is the percentage of customers that stopped using your company's product or service during a certain time frame
- Or it might be general and subjective, such as “give useful insights into the brain region interactions”
  - It should be indicated who makes the subjective judgment.

Phase 1: Business understanding



2. Assess situation

- More detailed fact-finding about all of the resources, constraints, assumptions, and other factors that should be considered in determining the data analysis goal and project plan
- In the previous task, your objective was to quickly get to the crux of the situation
- Here, you want to expand upon the details

2.1. Inventory of resources

- List the resources available to the project, including
  - personnel: business experts, data experts, technical support, data mining experts
  - data: fixed extracts, access to live, warehoused, or operational data
  - computing resources: hardware platforms
  - software: data mining tools, other relevant software

2.2. Requirements, assumptions and constraints

- Requirements
  - schedule of completion
  - comprehensibility and quality of results
  - security and legal issues
  - Make sure that you are allowed to use the data
- Assumptions
  - assumptions about the data that can be verified during data mining
  - non-verifiable assumptions about the business related to the project
  - The latter might affect the validity of the results
- Constraints
  - availability of resources
  - size of dataset that it is practical to use for modeling

2.3. Risks and contingencies

- Risks or events that might delay the project or cause it to fail
- Corresponding contingency plans, what action will be taken if these risks or events take place

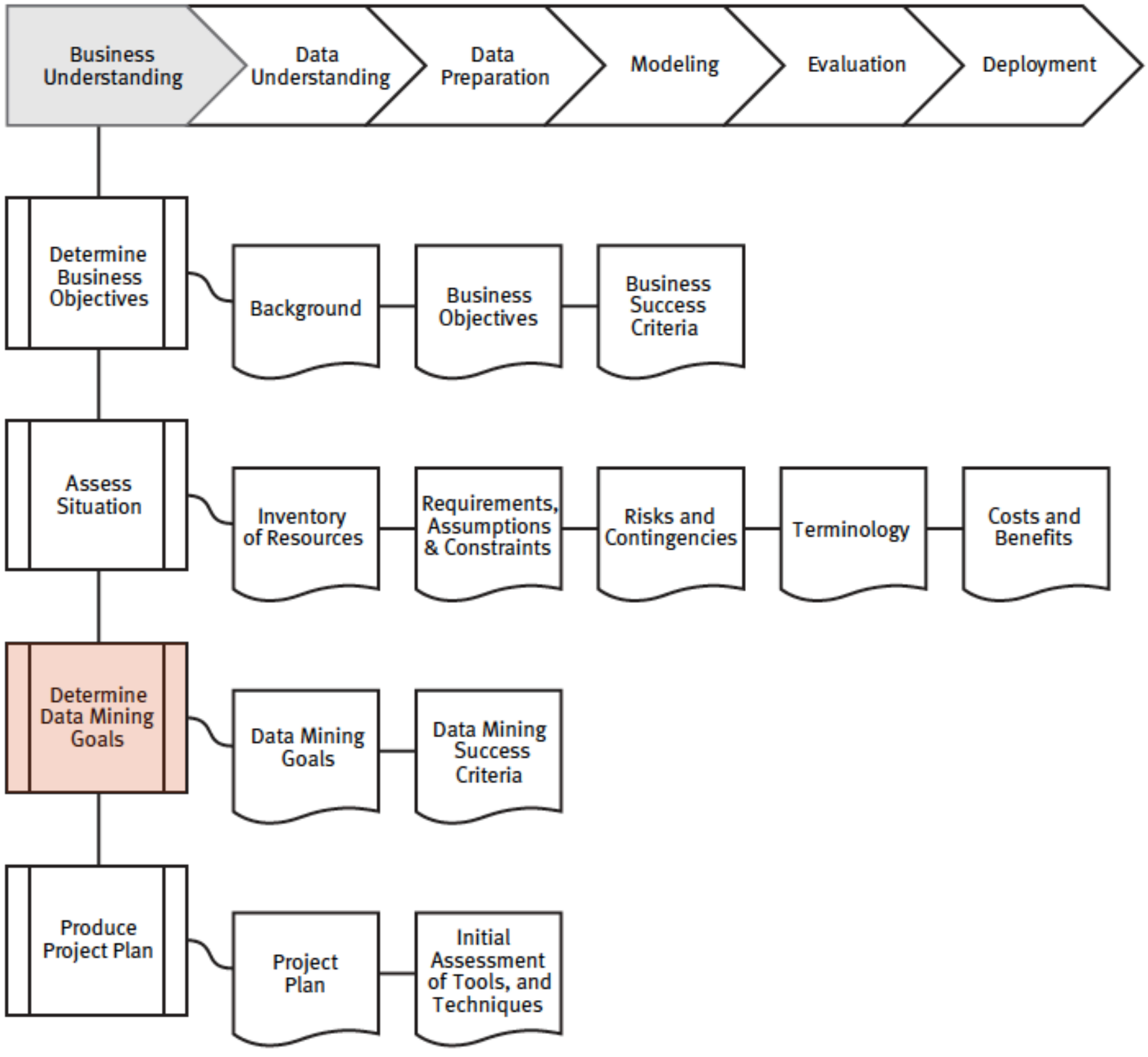
2.4. Terminology

- Glossary of terminology relevant to the project
- This may include two components:
  - A glossary of relevant business terminology, which forms part of the business understanding available to the project
  - A glossary of data mining terminology, illustrated with examples relevant to the business problem in question

2.5. Costs and benefits

- Cost-benefit analysis for the project, which compares the costs of the project with the potential benefits to the business if it is successful
- The comparison should be as specific as possible
  - For example, use monetary measures in a commercial situation

Phase 1: Business understanding



3. Determine data mining goals

- **A business goal states objectives in business terminology**
- **A data mining goal states project objectives in technical terms**
- For example, the business goal might be
  - "Increase sales to existing customers"
- A data mining goal might be
  - "Predict how many items a customer will buy, given their purchases over the past three years, demographic information (age, salary, city, etc.), and the price of the item"

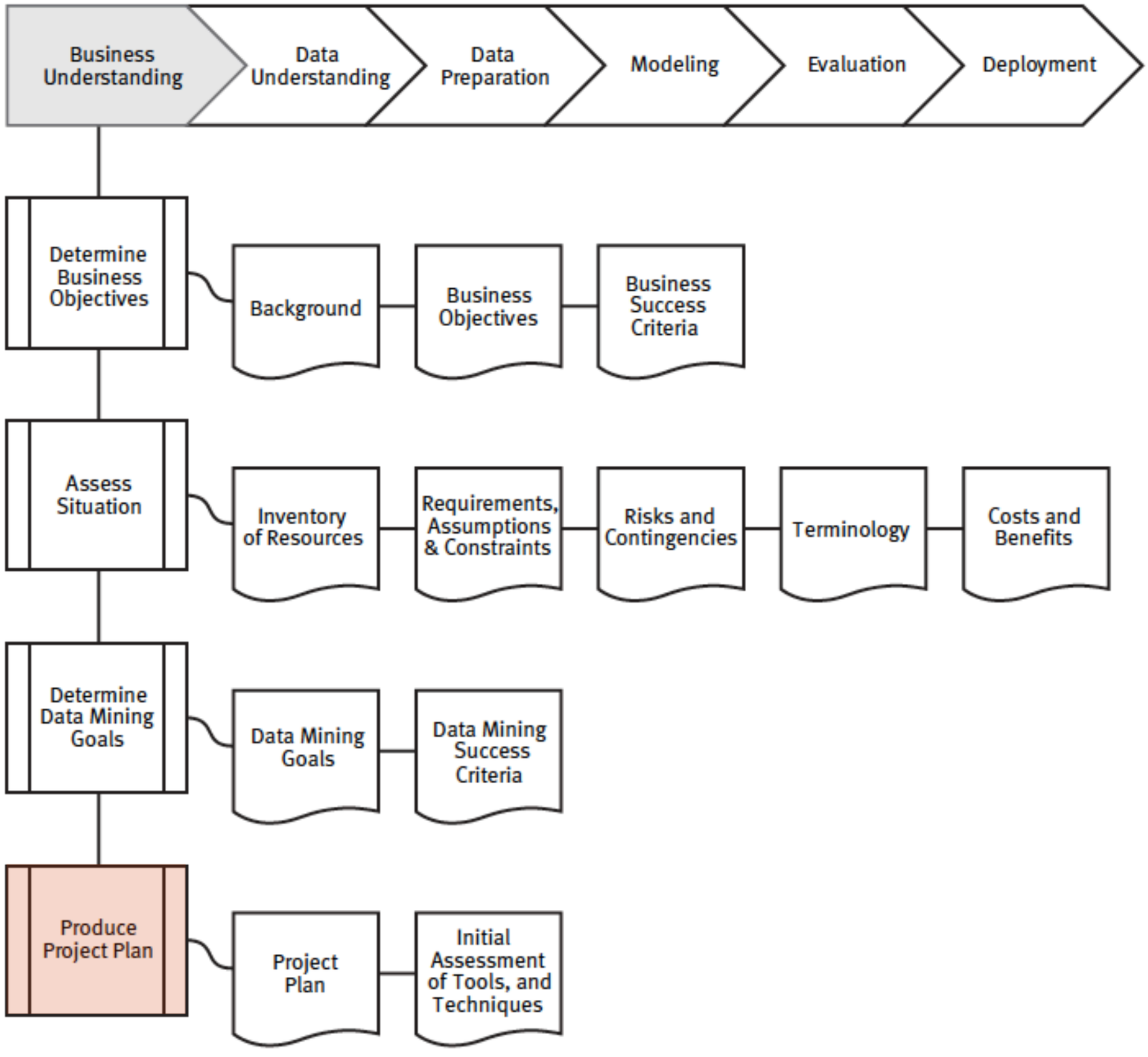
3.1. Data mining goals

- **Describe the intended outputs of the project that enable the achievement of the business objectives**

3.2. Data mining success criteria

- Define the criteria for a successful outcome to the project in technical terms
- For example
  - a certain level of predictive accuracy (e.g., in classification or regression)
  - propensity to purchase with a given degree of "lift." (e.g., in pattern discovery, association rules, frequent itemsets)
- As with business success criteria, it may be necessary to describe these in subjective terms
  - The person or persons making the subjective judgment should be identified

Phase 1: Business understanding



4. Produce project plan

- Describe the intended plan for achieving the data mining goals and thereby achieving the business goals
- The plan should specify the steps to be performed during the rest of the project, including the initial selection of tools and techniques

4.1. Project plan

- List the stages to be executed in the project, together with their duration, resources required, inputs, outputs, and dependencies
- Where possible, make explicit the large-scale iterations in the data mining process
  - For example, repetitions of the modeling and evaluation phases
- As part of the project plan, it is also important to analyze dependencies between time schedule and risks
  - Mark results of these analyses explicitly in the project plan, ideally with actions and recommendations if the risks are manifested

4.1. Project plan

- The project plan is a dynamic document
- At the end of each phase
  - a review of progress and achievements is necessary
  - a corresponding update of the project plan is recommended
- Specific review points for these updates are part of the project plan

4.2. Initial assessment of tools and techniques

- For example, you select a data mining tool that supports various methods for different stages of the process
- It is important to assess tools and techniques early in the process since the selection of tools and techniques may influence the entire project