Business and Data Understanding

In Business understanding and data understanding phase we will try to define the goals of the project by understanding the business needs and the data available to us. We will iterate between focusing on the business and exploring what data is available. This iteration typically involves specifying the business problem and then exploring if the appropriate data are available to develop a data-driven solution to the problem. If the data is available, then we will proceed; if not, we will have to identify an alternative problem to tackle. During this stage of the project, we will spend a great deal of time collecting data from different sources and try to understand whether it can fulfill our need.

Business Understanding

Understanding the project objectives and requirements from a business perspective, then converting this knowledge into a problem definition and a preliminary plan designed to achieve the objectives. This step includes-

- **Determine Business Objective:** Understand what the customer really wants to accomplish. Understand the objective and constraints that must have to be properly balanced. Uncover important factors that can influence the project outcome.

  Deliverables
  - **Background Report:** Record of the information known about the system
  - **Business Objectives Report:** Description of 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.
  - **Business Success Criteria Report:** Description of the criteria for a successful or useful outcome to the project from the business point of view.

- **Assess Situation:** Find detailed facts about all of the resources, constraints, assumptions, and other factors that should be considered in determining the data analysis goal and project plan.

Deliverables

- **Inventory of Resources Report:** List of the resources including data and software.
- **Requirements, Assumption and Constraints Report:** List of all the requirements of the project including schedule of completion, comprehensibility and quality of results, security and legal issues.

  List of the assumptions about the data that can be verified during data mining. List of the latter if it will affect the validity if the result.

  List of the constraints on the project. This is the constraints on the availability of resources and also technological constraints such as size of the dataset.
- **Risks and Contingencies Report:** List of the risks or events that might delay the project or cause it to fail. List of the corresponding contingency plans, what action will be taken if these risks or events take place.
- **Terminologies Report:** A compiled glossary of terminologies relevant to the project. It includes glossary of business terminologies and glossary of technical terminologies
- **Costs and Benefits Report:** A specific cost-benefit analysis for the project that will compare the costs of the project with the potential benefits to the business if it is successful.

- **Determine Project Goals:** Specify the project goals or project objectives in technical terms. That is what the project will do in business perspective.

  Deliverables

  - **Project Goals Report:** The description of the intended output of the project.
  - **Project Success Criteria Report:** The description of the criteria for the successful outcome to the project in technical terms. A certain level of predictive accuracy report of the project.

- **Produce Project Plan:** Description of the intended plan for achieving the project goals and thereby achieving the business goals. The plan will specify the steps to be performed during the rest of the project, including the initial selection of tools and techniques.

  Deliverables

  - **Project Plan Report:** List of the stages to be executed in the project, together with their duration, resources required, inputs, outputs, and dependencies with detailed

description. The plan will include the evaluation strategy that will be followed during the evaluation phase. This is a dynamic document and will be updated if necessary at the end of each phase. This update will take place after reviewing the progress and achievements.

    - **Initial Assessment of Tools and Techniques Report:** An initial assessment report of tools and techniques that should be performed.

Data Understanding

Collecting data and proceeds with activities to become familiar with the data, identifying data quality, discovering first insights into the data, and/or detecting interesting subsets to form hypotheses regarding hidden information. This step includes-

- **Collect Initial Data:** Acquire the data listed in the project resources. If the data source is multiple then integrate the data. Use the necessary tools to load and understand the data.

  Deliverables
    - **Initial Data Collection Report:** List of dataset(s) acquired, together with the locations, the methods used to acquire them and the problems encountered while collecting the data. Record of the resolutions achieved for the encountered problem.

- **Describe Data:** Examine the “gross” or “surface” properties of the acquired data and report on the results.

  Deliverables
    - **Data Description Report:** Description of the data that has been acquired including the format of the data, the quality of the data, the identities of the fields any other surface features which have been discovered. Evaluation notes of whether the data acquired satisfies the relevant requirements.

- **Explore Data:** Distribute the key attributes, relationships between pairs or small numbers of attributes. Explore properties of significant sub-populations and do simple statistical analyses. It may address the project goals.

  Deliverables
    - **Data Exploration Report:** The description of the result of data exploration. It includes first findings or initial hypothesis and their impact on the remainder of the project. May contain graphs and plots to indicate data characteristics.

- **Verify Data Quality:** Examine the quality of the data, addressing questions such as: Is the data complete (does it cover all the cases required)? Is it correct, or does it contain errors, and if there are errors, how common are they? Are there missing values in the data? If so, how are they represented, where do they occur, and how common are they?

  Deliverables
  - **Data Quality Report:** List of the results of the data quality verification; if quality problems exist, list of the possible solutions.

Tasks	Sub-tasks	Week-1	Week-2	Week-3	Week-4	Week-5
Business Understanding	Determining business objectives					
	Assessing situation					
	Determining project goals					
	Producing project plan					
Data Understanding	Collecting initial data					
	Describing data					
	Exploring data					
	Verifying data