**The business understanding stage of the “**diaper manufacturing” **Team Data Science Process lifecycle**

This article outlines the goals, tasks, and deliverables associated with the business understanding stage of the Team Data Science Process (TDSP). This process provides a recommended lifecycle that you can use to structure your data-science projects. The lifecycle outlines the major stages that projects typically execute, often iteratively:

1. **Business understanding**
2. **Data acquisition and understanding**
3. **Modeling**
4. **Deployment**
5. **Customer acceptance**

Here is a visual representation of the TDSP lifecycle:

Data Science Lifecycle

Business Understanding

Start

Transform Binning, Temporal, Text,Image Feature Selection

Feature Engineering

On-Premises Vs Cloud Database Vs Files

Data Source

Pipeline

Data Acquisition & Understanding

Modeling

Streaming vs Batch Low vs High Frequency

Algorithms, Ensemble Parameter Tuning Retraining Model Management

Model Training

On-premises vs Cloud Database vs Data Lake vs Small vs Medium vs Big Data

Environment

Model Evaluation

Cross Validation Model Reporting A/B Testing

Wrangling, Exploration & Cleaning

Structured vs Unstructured Data Validation and Cleanup Visualization

Deployment

Customer Acceptance

End

## **Goals**

Specify the key variables that are to serve as the model targets and whose related metrics are used determine the success of the project.

Identify the relevant data sources that the business has access to or needs to obtain.

## **How to do it**

There are two main tasks addressed in this stage:

* **Define objectives**: Work with your customer and other stakeholders to understand and identify the business problems. Formulate questions that define the business goals that the data science techniques can target.
* **Identify data sources**: Find the relevant data that helps you answer the questions that define the objectives of the project.

### **Define objectives**

1. A central objective of this step is to identify the key business variables that the analysis needs to predict. We refer to these variables as the *model targets*, and we use the metrics associated with them to determine the success of the project. Two examples of such targets are sales forecasts or the probability of an order being fraudulent.
2. Define the project goals by asking and refining "sharp" questions that are relevant, specific, and unambiguous. Data science is a process that uses names and numbers to answer such questions. You typically use data science or machine learning to answer five types of questions:
   * How much or how many? (regression)
   * Which category? (classification)
   * Which group? (clustering)
   * Is this weird? (anomaly detection)
   * Which option should be taken? (recommendation)

Determine which of these questions you're asking and how answering it achieves your business goals.

1. Define the project team by specifying the roles and responsibilities of its members. Develop a high-level milestone plan that you iterate on as you discover more information.
2. Define the success metrics. For example, you might want to achieve a customer churn prediction. You need an accuracy rate of "x" percent by the end of this three-month project. With this data, you can offer customer promotions to reduce churn. The metrics must be **SMART**:
   * **S**pecific
   * **M**easurable
   * **A**chievable
   * **R**elevant
   * **T**ime-bound

### **Identify data sources**

Identify data sources that contain known examples of answers to your sharp questions. Look for the following data:

* Data that's relevant to the question. Do you have measures of the target and features that are related to the target?
* Data that's an accurate measure of your model target and the features of interest.

For example, you might find that the existing systems need to collect and log additional kinds of data to address the problem and achieve the project goals. In this situation, you might want to look for external data sources or update your systems to collect new data.

Artifacts

Here are the deliverables in this stage:

* [Charter document](https://github.com/Azure/Azure-TDSP-ProjectTemplate/blob/master/Docs/Project/Charter.md): A standard template is provided in the TDSP project structure definition. The charter document is a living document. You update the template throughout the project as you make new discoveries and as business requirements change. The key is to iterate upon this document, adding more detail, as you progress through the discovery process. Keep the customer and other stakeholders involved in making the changes and clearly communicate the reasons for the changes to them.
* [Data sources](https://github.com/Azure/Azure-TDSP-ProjectTemplate/blob/master/Docs/Data_Report/Data%20Defintion.md#raw-data-sources): The **Raw data sources** section of the **Data definitions** report that's found in the TDSP project **Data report** folder contains the data sources. This section specifies the original and destination locations for the raw data. In later stages, you fill in additional details like the scripts to move the data to your analytic environment.
* [Data dictionaries](https://github.com/Azure/Azure-TDSP-ProjectTemplate/tree/master/Docs/Data_Dictionaries): This document provides descriptions of the data that's provided by the client. These descriptions include information about the schema (the data types and information on the validation rules, if any) and the entity-relation diagrams, if available.