1. Discovering

First we have to load the data in required format. In this step, the data is to be understood more deeply. Before implementing methods to clean it, we will definitely need to have a better idea about what the data is about. Wrangling needs to be done in specific manners, based on project criteria

2. Structuring

In this step we have to restructure given dataset in a manner that better suits the analytical method used, in most cases there will not be any structure to it. One column may become two, or rows may be split – whatever needs to be done for better analysis.

3. Cleaning

All datasets are sure to have some outliers, which can affect the results of the analysis. These will have to be cleaned, for the best results. In this step, the data is cleaned thoroughly for high-quality analysis. Null values will have to be changed, and the formatting will be standardized in order to make the data of higher quality.

4. Distributions of Feature Values

Looking at distributions of features is immensely useful in getting a feel for whether the values look sensible and whether there are any obvious outliers to investigate. We're interesting in focusing on whether distributions look plausible or wrong. Later on, we're more interested in relationships and patterns.

5. Validating

Validation rules refer to some repetitive programming steps which are used to verify the consistency, quality and the security of the data you have.

6. Publishing / Save the Data

The prepared data is published so that it can be used further down the line – that is its purpose after all. If needed, we also have to document the steps which were taken.