# Creation of the initial dataset

Creating the initial dataset is a critical step in many data-driven projects, especially those involving machine learning and data analysis. To create the initial dataset, you need to identify the data sources, gather the relevant data, and structure it in a format that can be easily processed by the tools you plan to use. This may involve cleaning, transforming, and normalizing the data to ensure that it is consistent and ready for analysis. Additionally, you should consider any legal, ethical, or privacy concerns that may apply to the data, such as obtaining consent or de-identifying sensitive information. A well-curated initial dataset is essential for accurate and reliable analysis, modeling, and decision-making.

In this Cement Material Compressive Strength Experiment Data, there are 109 null value rows, which we have dropped and added 10 extra rows using one of Data Augmentation Technique, by producing some noise of existing rows. I will be using this dataset for further analysis.