Milestone 1: Data preparation

For the first part of your project, you need to prepare the data for your project.

Depending on your project, you may need to

- Synthetize your data: you need to simulate/generate your data.
    - You need to **submit the code for the data generation**.

- Get your data from a source, e.g. LHC, observatory, …
    - You need to **submit the code to extract and work with the data.**

- Extracted, e.g. Crawl from the internet
    - You need to **submit the code for extraction of data.**

Also, you need to clean up your data and put it in the right format, ready for feeding to the model you want to train. This means that you need to **submit a sample data (X, Y)** where both of these matrices contain only numerical values.

With this, you need to **submit a short report** that explains you data, e.g. how the data is collected, how many samples there is, what the features and the label(s) are, what they physically mean and etc.

You also need to **do some basic statistical analysis on your data**, e.g. find the histograms and correlation functions.

Finally, you need **to submit a list of articles** from the relevant literature and a brief paragraph of their significance/relevance to your project.