Preparing the dataset: