1. Problem statement: Restate the initial project that you proposed in deliverable one in 2 – 3 sentences. Be sure to refer back to this problem statement in the following questions.
   * The aim of our project is to analyze text from news media to identify political leaning. We will use a logistic regression model and data drawn from a wide array of political articles to accurately model this bias.
2. Data Preprocessing: Confirm the dataset you are working with. State any changes from the initial dataset you chose. Discuss the content of the dataset (number of samples, labels, etc.). Describe and justify your data preprocessing methods (did you delete or modify any data? If so, why?).
   * We will be working with [this](https://www.kaggle.com/datasets/mayobanexsantana/political-bias/data) dataset, with minor tweaks made for formatting and to remove empty entries or extra characters. The final dataset after preprocessing is made up of about 3.5 thousand entries consisting of article titles, content, bias, tokenized text, and BERT embeddings. The tokenization and BERT feature extraction will make working with this dataset easier and more fruitful