Data Wrangling Performed:

Datasets used:

1) News Data from the past 8 years with each row containing top 25 headlines
2) Stock Price data for the past 8 years for a company

Initially, I checked the datatypes of all the columns and the shape of the datasets. Then I researched online and found a research paper and an algorithm explaining the way to find out the subjectivity, objectivity and polarity of a give sentence or a paragraph. The algorithm was written in a way to read only a given sentence or a paragraph to find out the polarity. I modified the algorithm such that it accepts a dataframe and runs the algorithm across all the rows with 25 columns of headlines in conjunction. After running across the dataframe, it generates new columns with data for **subjectivity, objectivity, negative, positive, neutral** in terms of percentage for each row. At the end of the algorithm, a new dataframe is generated.

I will then **concatenate** the subjectivity, objectivity, negative, positive, neutral columns of one dataframe with another dataframe which has data about the stock prices for the past 8 years. After creating the new concatenated dataframe, I checked the shape and info of the dataframe and found some null values. I replaced the null values with the **mean** of the respective column the null values were present.

After this I used describe() to check the statistics of the dataframe and find out any outliers in the dataframe and luckily did not find any outliers in the dataframe. This newly created dataframe will be used for further evaluation and predictive analytics.