Update #1 (1/13/21)

* Filter by most popular sites
* Decimal columns are GDELT NLP version, random letters are mapped
* Code on dashboard to run searches through dashboard
* Tone = Polarity
* Start from 2021; but go back as far as possible hopefully Mar 2020 (COVID)
* Run all data/features for today and then project for tomorrow
* Start with distil Roberta model
* Opening and closing, opening prices are important because of changes over night
* Readability
  + Hard to read 🡪 hide something
  + Easy to read 🡪 good performance
  + Look for opensource packages
  + 4th grade reading level, for news not necessarily
  + Word count, paragraph count, sentence length

Update #2 (1/27/22)

* Add sector and market cap to stock tickers
* Filter columns and websites
* Read in a day or week at a time
* Scrape 5-10 financial/popular websites