**Goal**

The goal of this project is to build a sentiment analysis model which will allow us to categorize words in the twitter on economy based on their sentiments, that is whether they are positive, negative or neutral using the lexicons from the tidytext package and get insights like to know what are the most frequent words used on twitter regarding economy through word count and word clouds which determine frequent words, to know top positive-negative words with the help of bing lexicon, to determine whether there is total positive or negative sentiment on economy.

**Abstract**

This sentiment analysis model is to understand whether people are optimistic or pessimistic on Economy, as it is generally considered by economists that employers pessimism on economy might create fears that things may go more worse and they will stop hiring and might even reduce work force in their respective companies and consumers on the other hand might want to save more and spend less as they are uncertain on their future financial prospect. These both actions will shrink economy as both supply and demand will be reduced.

This analysis model is particularly relevant as economies have started reopening across the world after more than a year gap due to covid-19 pandemic. Inflation is reported across various nations around the globe. Largest Economy United States is still recovering as US economy still did not reach pre-covid levels with also inflation fears where India( Most populated country after China) reported first time shrink in their middle income class population (shifting more towards poverty) after nearly two decades along with Whole sale price index grew 7.3% , EU also reported 8.2 % unemployment of its workforce in April.

The project is mainly based on Sentiment Analysis within R using tidytext package that comprises of sentiment lexicons like AFINN, BING and NRC. The visualization used in this project were created using R programming on R Studio.

**Packages Used**

The required text data was loaded into RStudio from the twitter along with the other packages that are required for this analysis like tidytext, textdata, worldcloud2, ggplot2, RColorBrewer, dplyr, rshape2, tidyverse, tidyr,tm, tidytext, snowballC.

**Conclusion**

* Negative sentiment is prevailing than the positive, i.e., people tweeting negatively on the economy of their respective countries or global scenario.
* The words like debt , hard ,crisis are used to describe pessimism on Economy and support, relief are used for optimism.
* Presence of words like health, pandemic, covid reflects pandemic is a vital factor determining people’s sentiment on economy

**Limitations**

* Twitter data obtained is not divided in to regions or countries, average of global purview on economy via twitter interface can be observed but not individual countries or region
* The above limitation will hinder any public policy recommendations via this model for nations
* As it is not a survey analysis , this analysis is limited for only purview global perspective on economies currently. A paper is attached for further reading where authors have analyzed in long run for presence of over optimism on inflation and unemployment and considerable effects.

**References:**

Dua, P., & Smyth, D. J. (1993). Survey Evidence on Excessive Public Pessimism About the Future Behavior of Unemployment. *Public Opinion Quarterly*, *57*(4), 566. https://doi.org/10.1086/269396

Kochhar, R. (2021, March 22). *In the pandemic, India's middle class shrinks and poverty spreads while China sees smaller changes*. Pew Research Center. https://www.pewresearch.org/fact-tank/2021/03/18/in-the-pandemic-indias-middle-class-shrinks-and-poverty-spreads-while-china-sees-smaller-changes/.

“Unemployment Statistics.” *Unemployment Statistics - Statistics Explained*, ec.europa.eu/eurostat/statistics-explained/index.php?title=Unemployment\_statistics.