Project Proposal – Pfizer Vaccine Tweets

# Project Goal:

Analyze real recent twitter data provided by Kaggle. Perform EDA on dataset to provide useful insights. Further utilize Amazon Comprehend on cleaned data and perform sentiment analysis.

# Team Members:

* Nicole Norelli – [nnorelli@smu.edu](mailto:nnorelli@smu.edu)
* Mingyang Nick YU – [nyu@smu.edu](mailto:nyu@smu.edu)
* Justin Ehly – [jehly@smu.edu](mailto:jehly@smu.edu)

# Objectives:

This project is aimed at analyzing real world twitter data with related to Pfizer Vaccine. The dataset is provided by Kaggle, and resource link is listed below. The project is aimed at providing useful insights for researchers and government officials with regard to the reactions of general public on Pfizer Vaccine, and sentiment on this particular vaccine brand. Such insight can potentially help direct government agencies across the globe on distribution, public education on the vaccine knowledge, and further advertisement necessary to ensure public confidence on taking the vaccine.

The analysis may include but not limited to location analysis on which country generate the most heat on discussion of Pfizer vaccine, hashtag analysis on what topics are people interested in sharing, sentiment analysis on whether general public are more positive, negative, neutral or mixed on the Pfizer Vaccine.

This project include learning objectives on deep data cleaning with R, EDA with R, natural language processing, Amazon cloud computing including S3, Comprehend and other visualization tools.

# Resources:

* <https://www.kaggle.com/gpreda/pfizer-vaccine-tweets>
* <https://docs.aws.amazon.com/comprehend/latest/dg/comprehend-dg.pdf>
* <https://www.red-gate.com/simple-talk/sql/bi/text-mining-and-sentiment-analysis-with-r/>