**Project Title: Statistical Analysis Between the Correlations of Sentiment and Daily Returns Varying by Business Sector**

**Team (1) members: Ly-Bach Truong**

**Project Description/Outline:**

At first, I intended to analyze the relationship between the daily return of stocks and their sentiment score (which is provided through Quantopian’s data pipeline) and create a trading algorithm based on sentiment, but while working through the project, I realized that an LSTM network would better suit my purposes.

Now, the aim of my project is to find correlations between the sentiment of various stocks and their corresponding daily returns, grouped by similarity (market cap, stock price, etc.) and varying by sector and I will also develop a forecasting model using an LSTM deep learning network, which will base its predictions on sequential data in order to gain daily returns.

**Data Sets to be Used:**

* Quantopian’s data pipeline that can be accessed through their in-browser Jupyter Notebook environment
* PsychSignal’s trader moods data set, which uses NLP to quantify trader sentiment with StockTwit’s data

**Intended Machine Learning Libraries:**

* Sci-Kit Learn for preprocessing data
* Keras for a model over sequential data and LSTM network

**Intended Additional Libraries**

* pandas
* Matplotlib.pyplot
* NumPy
* SciPy.stats
* Quantopian data pipeline (accessed by their Python library)