# John Hanlon

# GA-DAT-NYC-43

# Final Project: Part 2 - Project Design Writeup and Approval Template

### Project Problem and Hypothesis

This project will utilize a Long-Short Mean Reversion trading strategy to outperform the **iShares North American Tech ETF**. The project will use a machine learning classifier to suggest if a given security will increase or decrease in value. **The hypothesis is that the returns of the ETF’s components will track similarly to the performance of the ETF; any securities that are outperforming the ETF’s returns will revert to the mean, as will those securities underperforming.** Performances is measured by daily yield. I believe the DJIA will be a strong indicator of whether a security will increase or decrease in value, as it is often used as an indicator of investor sentiment.

### Datasets

The datasets will be pulled from Quandl. Additionally, Python script may be executed in Quantopian or with the Zipline API. The Zipline API allows for paper trading and back-testing, which will provide a monetary return value.



### Domain knowledge

I have previously worked on the trading floor and also attended several Quantopian seminars. The Quantopian seminars helped to formulate the benefits a long-short strategy in order to hedge against market risks. The long-short strategy attempts to be market neutral, given the opportunity for vast diversification, however given the scope of this project, this strategy will not be as diversified. I aim to rank the 266 constituents of the ETF by their monthly yield. We will short the top 25 highest returning securities and go long the bottom 25 securities.

Benchmarks would be the yield for the DJIA, SP500, iShares NA Tech ETF. We are looking to see if we can produce higher returns using this strategy.

### Project Concerns

* I want to understand if I have the concepts outlined correctly regarding how I would utilize machine learning classification to determine if a given security is above or below the benchmark return
* I want to confirm that this strategy makes sense
* I want to confirm if there are additional features I should consider (I know there are); perhaps metrics relating to each company’s fundamentals could be incorporated into the ranking system?
* For the ranking scheme, perhaps a linear regression to predict returns?
* Is there a difference between ‘factors’ and ‘features’ in quantities analysis?
* Risks to the model in the lack of features/factors
* The risk of learning I am wrong is the validation of passive management!

### Outcomes

I expect the SVM score to (hopefully) be around 70%. Additionally, I hope for the monetary return values to be higher than that of the index, for equal amounts of initial capital. I do not believe this model needs to be overly complicated, we need a script to pull the daily OHLC data for the securities and ETF in scope, US GDP data (this is most likely only monthly, may have to resample the data), DJIA data, SP500 data. With the data, we will then see if the returns of the ETF constituents correlate with returns of the ETF. If so, we will use a ML classifier to determine if the return for each security greater than that of the ETF. We will then need to determine a ranking scheme after we have classified our securities to indicate which securities should be put short and held long.

In order to be a success, I would like to see that the strategy can breakeven. A great outcome would be to outperform the ETF and/or the SP500. Bare minimum success would be to get the classifier and ranking scheme to work and produce meaningful output, even if the strategy results in a loss.