**Marcellus Journal:**

**Week - 1**

I joined Marcellus on 8th May 2023. Mayank was appointed as my mentor by Arindam Sir. The very first day, they made me feel extremely comfortable in the office. I was given an official laptop along with an official email id. I introduced myself to all the team members and tried to settle in. The very first day itself, Arindam Sir took me and Mayank to Starbucks and we had a delightful conversation about life and career. I was very impressed by Arindam Sir and his brilliance along with the fact that he was devoting a part of his precious time to an intern. That meant a lot to me.

I was told that I needed to work on mimicking the Fear and Greed Index by CNN based on which we would develop a short-term trading strategy to optimize the cash portion of Marcellus’ portfolio. I started reading about the FGI index and learned that it depended on 7 market indicators namely Market Momentum, Stock Price Strength, Stock Price Breadth, Put and Call Options, Market Volatility, Safe Haven Demand and Junk Bond Demand.

Mayank provided me with the necessary raw data extracted from Bloomberg and over the course of the week, both of us brainstormed on the project. Since this was based on CNN’s proprietary formula, we could not find any source that could directly give us the formula and thus, it was all about our brainstorming. I used ChatGPT to get an estimate of how we can proceed on the project and saved that as a document.

Gradually, I started to make user defined functions separately for each indicator. By the end of the week, I had made all the 7 indicators’ functions. We normalized each indicator score between 0 to 100, using MinMax Scaling. However, we realized that each indicator had a different mean, median and standard deviation. Since the FGI index is a linear combination of all the 7 indicators, we want all of them to have a similar distribution. Arindam Sir suggested to use the z-score normalization for all the features. Therefore, I calculated the z scores of the indicators by subtracting each value from its mean and dividing it by standard deviation. This resulted in better results. Now, the various indicators had a mean around zero and their distribution was similar to the normal distribution. I used matplotlib to plot histograms f the distribution of each indicator along with the FGI index, which was calculated as an average of all the 7 indicator z scores. FGI index was minmax scaled to a value between 0 and 100 where a higher value signified greed and a lower value signified fear.

Now, we started to backtest this data. A 5-day moving average of the FGI index was now used as the signal. Arindam Sir told me to calculate the monthly returns by taking the FGI index as a signal with lags. 22 trading days were taken as a month. Therefore, monthly return meant return on S&P500 index between T and T+22. In case there is a lag x, return was taken as return on S&P500 index between T+x and T+x+22. A dataframe was created with 0\_day\_lag, 3\_day\_lag and so on. The descriptive statistics of this dataframe were obtained. It was observed that a 5-day lag generated better returns relatively.

**Week-2**

Monday morning, Arindam Sir took a review meet to understand the work that had been done the previous week. I was able to make him understand what I had done and he seemed to be impressed by the work done. He asked me to show the distributions of each 20% range of returns based on only long positions every day irrespective of the signal. Based on this, we would decide what kind of strategy to adopt. I was also asked to create user defined functions for the backtesting that can take the data, number of days return to be calculated, lags, as well as number of days moving avg of FGI as input parameters and return dataframe accordingly.

Over the course of week, I designed the user defined backtesting function which can take parameters such as number of days of return, lags, number of days of moving avg and the data as input and provide the necessary dataframe. I also created functions to sort the monthly returns with lags based on the FGI values. Then, using matplotlib, I created histograms for each 20% range of the FGI values, displayed the mean, median, count\_positive, count\_negative, count\_days of the monthly returns. I repeated the process for daily returns.

After this, Arindam Sir asked me to go through all the data and create an algorithm that can time the market as accurately as possible. We needed to develop a signal which can tell us when we should be invested in the market and when not to. I spoke to Krishnan regarding the application of ML algorithms.

**Week-3**

I started studying about various ML algorithms like Logistic Regression, Decision Trees, etc. I also planned to do a Principal Component Analysis of the various features to identify the important ones for the model. I made a correlation matrix among all the 7 indicators as well. I deployed a Decision Tree model using the FGI scores as the sole feature. The data was split into training and testing data where 70% data was used for training and the remaining was used for testing. I labelled the returns more than 0.01 as 1 and 0 otherwise, implying that we would go long on the index if the monthly forward looking return is greater than 1%. The model showed decent results with a precision of 0.58 and a recall of 0.87. Then the predicted y values were put in the original dataframe and the returns against the long indicators were observed. They had a median of 0.0208 compared to the 0.0157 originally and at 75% mark, it was 0.0407 compared to the 0.0363 originally. Best results were seen for 3\_Day\_Lag returns.

* Recall: The ability of a model to find all the relevant cases within a data set. Mathematically, we define recall as the number of true positives divided by the number of true positives plus the number of false negatives.
* Precision: The ability of a classification model to identify only the relevant data points. Mathematically, precision is the number of true positives divided by the number of true positives plus the number of false positives.