Deliverable 1
Name: Zhu, Feng

1. Choice of dataset: NASDAQ Clean Edge Green Energy Index price

ESG investing has become a hot field in the financial world due to the increasing awareness of global warming and sustainability. Green stock prices have surged and green assets have delivered high returns in the past few years, however many researchers hold the opinion that the strong performance of green stocks will not last long, as the outperformance reflects unexpectedly strong increases in environmental awareness, not high expected returns. This year many ESG indexes and green stocks started to underperform. Due to my background in finance and my interest in sustainability and ESG investment, I decided to run a model to predict green stock prices, whether it will follow the previous outperforming trend or continue to underperform.

Here the dataset I choose is historical NASDAQ Clean Edge Green Energy Index price. I am not sure whether it's the optimal choice but many other ESG indexes were established last year which doesn't give me a sufficient data set to work on. The dataset I chose has relatively more data to train on. My second option can be picking one single green stock and using its historical trading price as a dataset.

2. Methodology: I will use NASDAQ Clean Edge Green Energy Index price as a dataset, each row includes open price, daily high, daily low, close price, adjusted close price and trading value.

Data Preprocessing: I think here the most valuable factors are adjusted close price and trading value. However, I have also considered adding other variables, especially news from the media regarding environmental issues and macroeconomics situations. I don't have such datasets and I need to check whether it's doable to construct such variables. I will also split the data: 70% for training and 30% for testing.

Machine learning model: LSTM Model. I chose this model due to its efficiency and good performance in time series data. I am open to your suggestions as well.

Evaluation Metric: RMSE (Root Mean Squared Error) and MAPE (Mean Absolute Percentage Error %) to measure forecast accuracy.

3. Application: I would like to integrate the model in a landing-page to illustrate stock price forecast results (3-month, 6-month, 1-year, and 3-year). I would like to use this empirical work to further my theoretical research in green premium and green assets pricing.