Introduction/Problem Statement:

The problem statement is to use historical data of the S&P 500 index to create a model that can accurately predict future high prices of the stock. The model will be built using a deep learning technique known as Long Short-Term Memory (LSTM) and will be trained on closing prices of the stock. The performance of the model will be evaluated by comparing its predictions with the actual high prices of the stock. The goal is to achieve a low mean squared error between the predicted and actual prices, indicating a high level of accuracy in the model's predictions. This will help stakeholders in making informed decisions about the stock market by having a better understanding of future stock prices.

Dataset Source/Information:

The dataset is historical stock price information for the S&P 500 index, which is an index of 500 stocks from different sectors of the US economy and is widely considered as an indicator of the US equities market. The dataset is obtained using the pandas_datareader library which extracts the information from Yahoo Finance databases. The closing price information is used in this dataset, but other information such as opening price, adjusted closing price, etc are also available. The dataset contains information about the high and low prices, open and close prices, and volume of trading activity. The adjusted values also factor in corporate actions such as dividends, stock splits, and new share issuance. The dataset is prepared using a utility function get_raw_data() which takes index ticker name as input, for S&P 500 index the ticker name is ^GSPC.

The columns in the dataset are as follows:

1. ***Date:*** Trading date
2. ***High:*** The high and low refer to the maximum and minimum prices in a given time period.
3. ***Low:*** The high and low refer to the maximum and minimum prices in a given time period. Open and close are the prices at which a stock began and ended trading in the same period.
4. ***Open:*** Open and close are the prices at which a stock began and ended trading in the same period.
5. ***Close:*** Open and close are the prices at which a stock began and ended trading in the same period.
6. ***Volume:*** Volume is the total amount of trading activity.
7. ***Adj Close:*** Adjusted values factor in corporate actions such as dividends, stock splits, and new share issuance.

Modeling:

This code is building a time series forecasting model using a deep learning technique known as Long Short-Term Memory (LSTM). The model is trained on historical high prices of a stock, and then used to make predictions on the stock's future high prices. The script starts by importing several libraries including Numpy, Pandas, and Keras, which is a deep learning library for building and training neural networks. The data used for the model is loaded from a CSV file and is then preprocessed for further analysis. The plot for the high, low, open and close of the stock is drawn with the help of Matplotlib and Seaborn library. The script then splits the data into a training set and a test set and creates a dataset for the model by using a sliding window approach, where each data point is a set of stock prices from the previous 'lag' number of days. A LSTM model is built using the Keras library, which is trained on the training set data. Then the model is used to make predictions on the test set data. The script plots the comparison of predicted vs actual prices and calculated mean squared error between the predictions and actual stock prices. Finally, the script uses this model to make predictions on a set of stock prices that were not used during the training and testing phases, and plots these predictions.

Limitations:

1. The model is only trained on historical data of the S&P 500 index, and its performance may not generalize well to other stocks or other markets.
2. The model is only trained on high prices of the stock, and it may not perform well on other features such as low prices or volume of trading activity.
3. The model is only trained on closing prices of the stock and not adjusted values which could have an impact on the stock prices.
4. The model uses a sliding window approach to create the dataset, which may not take into account other factors that could affect stock prices, such as news, events, or global market conditions.
5. The model uses only one feature, closing prices, which could be limiting for the model to make accurate predictions.
6. The model is only trained on the data of the S&P 500 index, which is an indicator of US equities, may not perform well on other markets or other indices.
7. The model is based on the historical data, and this data is limited by the time period of the data availability and it may not perform well on future data or out of sample data.
8. The model uses LSTM, which is a powerful method but it requires a large amount of data and computational resources to train. This could be a limitation if the dataset is small or if computational resources are limited.
9. The model only uses 10 epochs, which may not be enough for the model to fully converge, which could limit the accuracy of the predictions.
10. The model is based on a time series forecasting, which is a complex task and it's dependent on many factors like seasonality, trends, etc. So, it may not be accurate for all the cases.