Task 3 Report

This task comes in 2 parts. The first being to create a function for generating a candlestick plot. The second being to generate a boxplot. While the generation of these plots are fairly trivial, the trick comes in formatting the data correctly to show the required windows.

In both cases, data is fetched by the same method as the previous task, using yfinance if a ticker label is passed in. Otherwise csv data may be used.

A screen shot of a computer program

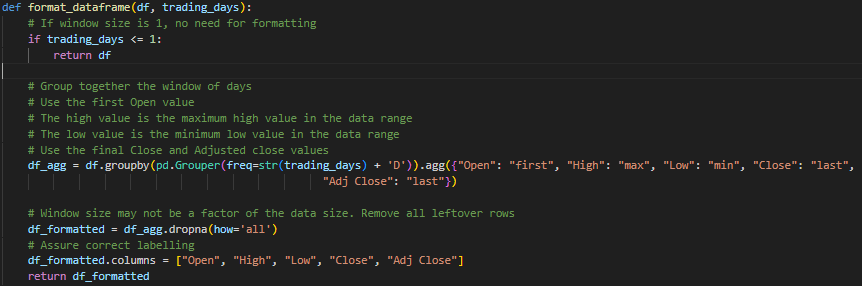
Description automatically generated

## Candlestick

The challenge in generating this chart is the aggregation of data. The case of candles representing 1 day are easy. Since the data is already fetched on a daily basis, we can plot the chart directly on the original data. However when a larger trading day size is chosen we need a method to aggregate the data. Initially I attempted doing this manually, looping over sections of data, calculating the maximum and minimum high and low values, taking the first open value and the final close value. Then passing this formatted data into a larger formatted array. While this did work, I investigated further to find a more streamlined way.

Fortunately groupby() and agg() proved useful here. I simply grouped the data into sections of n trading days. Then I called agg() passing in the same logic as before for the High, Low, Open, Close and Adjusted Close values.

Since these group sizes likely aren’t an exact factor of the original data size, dropna was used to remove rows with NaN values.

  
A screen shot of a computer program

Description automatically generated

## Boxplot

Unlike the candleplot, we don’t need to aggregate data, only trim data into sections of a single column. Somewhat arbitrarily I chose the Adjusted Close column.

Although more visually complex than perhaps necessary, I created a for loop with a range based on the window size, then trimmed the Adjusted Close column into sections. Storing the array of sections in the ‘boxplot\_data’ variable. After once again removing the NaN values, I created a basic matplotlib boxplot.

A screen shot of a computer code

Description automatically generated

## Output

A screenshot of a graph

Description automatically generated

A graph of a chart

Description automatically generated with medium confidence