Advanced Time-Series Analysis with Pandas and Yahoo Finance Data

Data Visualization Course
Roi Yozevitch

May 21, 2023

In this assignment, you will apply advanced time-series functions from the Pandas library on stock market data obtained using the 'yfinance' library. You will download data for the following tickers: Apple (AAPL), Boeing (BA), Coca-Cola (KO), IBM, Disney (DIS), Microsoft (MSFT), Tesla (TSLA), and Google (GOOG) from January 1, 2010 to today.

1. Download the stock market data using the 'yfinance' library in Python. Print the first 10 rows of data.

2. Use the Pandas 'describe()' function to get the statistical summary of the DataFrame. What can you learn from it?

3. For each stock, calculate and plot the cumulative daily returns. What can you deduce about the performance of each stock?

4. Implement a simple moving average (SMA) trading strategy with a short window of 20 days and a long window of 100 days. For this, you should calculate the short-window (20 days) and long-window (100 days) moving averages. Then, create a signal when the short-window moving average crosses above the long-window moving average (buy signal) and when the short-window moving average crosses below the long-window moving average (sell signal). Discuss the potential effectiveness of this strategy for each stock.

5. Using the 'shift()' function in Pandas, compute and plot the 5 days lagged returns for each stock. How does it compare to the original daily returns?

6. For each stock, resample the data to a weekly frequency and calculate the weekly returns. Discuss the volatility of each stock.

7. Calculate the rolling correlation with a window of 120 days for each pair of stocks. The rolling correlation involves calculating the correlation between the returns of two stocks over a rolling window of 120 days, and then moving this window one day at a time over the entire time period. This results in a time series that shows how the correlation between each pair of stocks changes over time. After calculating the rolling correlations, visualize the results with a plot and discuss any interesting trends or anomalies you observe. Are there periods when the correlations are particularly high or low? If so, can you explain why? Does the correlation between some pairs of stocks change in a more volatile manner than others? Can you hypothesize why that might be the case?

8. Compute and plot the 14-day Relative Strength Index (RSI) for each stock. The RSI is a momentum oscillator that measures the speed and change of price movements. It is calculated using the formula: RSI = 100 - 100 / (1 + RS), where RS is the average gain over the given period divided by the average loss over the given period. A stock is generally considered overbought when the RSI is above 70 and oversold when it's below 30. Identify periods where the stocks were likely overbought and oversold.

9. Resample the data to monthly frequency and calculate the monthly rate of return. Then, calculate the volatility (standard deviation of returns) for each stock. Which stock would have been the riskiest to invest in?

10. Based on the data, if you had to build a portfolio of 3 stocks, which ones would you choose and why? Provide a detailed justification for your choices.

Please submit your Python Jupyter notebook with the solutions to these problems. Be sure to include any necessary comments in your code to explain what it is doing.