Data Science Interview Round 1

Programming Assignment

The given dataset is the daily close of a major U.S. market index. Review the data and submit your python code in a clearly commented Jupyter Notebook along with a brief Powerpoint presentation.

Your submission be evaluated on the following points:

- Completeness
- Technical accuracy
- Presentation skills

Please answer the following questions:

- Examine the daily volume data
  - Identify the optimal number of clusters for daily volume data

- Examine the daily open, close, and volume data
  - Compute the fractional difference between current day open and previous day open (Parameter 1)
  - Compute the fractional difference between current day open and current day close (Parameter 2)
  - Compute the fractional difference between current day volume and previous day volume (Parameter 3)
  - How can you cluster the above data set (Parameters 1, 2, 3)?
  - Hint: Fractional difference of opening price = [Opening price (today) / Opening price (yesterday)] - 1

- There is a saying on Wall Street, “Santa brings the Christmas rally”.
  - For the data set, compute the monthly returns i.e., fractional difference between the month’s closing price and opening price
  - Use decision tree to classify if investing in any month can be a profitable strategy
  - What are the error metrics of your model?
  - Hint: Monthly returns = [Closing price (last day) / Closing price (first day)] - 1