Index Rebalance: Portfolio Maximization

You are given a portfolio worth INR 50 Lakhs (1 Million = 10 Lakhs) on *day 0*. The total invested amount in all 51 stocks is INR 4167296. You have INR 832704 cash in hand, thus the portfolio worth is INR 50 Lakhs. Note that:

- You should invest (*buy* or *sell*) INR 25 Lakhs on *day 1* and again on *day 2*, thus a total of INR 50 Lakhs should be invested resulting in a portfolio of worth INR 1 Crore (1 Crore = 10 Million).
- You should invest in all 51 stocks.
- The investment should happen on the closing prices, i.e., if you are buying the k^{th} stock at day i, then the closing price of the k^{th} stock on day i should be considered.
- At the end of *day 2*, you should not have more than INR 5 Lakhs cash in hand.

The portfolio constituents should mimic the index constituent weights as closely as possible with a minimal cash component, if any. The index constituent weight, w_k, for the k^{th} stock is computed as:

$$w_k = \frac{e_k \times p_k}{\sum_{k=1}^{51} (e_k \times p_k)}$$

Here,

- e_k is the free float equity share of the k^{th} stock.
- p_k is the closing price of the k^{th} stock on *day 0*.
- $e_k \times p_k$ is the market capitalization of the k^{th} stock.

Dataset

We provide the zip file (*MD5* checksum is `b33a2f995a0effd91e08c06e84843e26`): index_rebalance_dataset.zip containing the following four files when unzipped:

- `stocks_info.csv` contains the information of all the 51 stocks, given by the following four columns:
  - `Symbol`: This is the stock symbol.
  - `Name`: This is the stock name.
  - `Industry`: This represents the industry.
  - `Free Float Equity Shares`: This is the value of free float equity shares.
- `portfolio.csv` contains the information of the base (*day 0*) portfolio, given by the following two columns:
  - `Symbol`: This is the stock symbol.
  - `Quantity`: The total number of shares of the stock on *day 0*.
- `stocks_closing_prices.csv` contains the information of the stocks closing prices for the *day 0*, *day 1*, and *day 2*, given by the following 4 columns:
  - `Symbol`: This is the stock symbol.
  - `Day_0`: The closing price of the stock on the *day 0*.

- `Day_1`: The closing price of the stock on the *day 1*.
- `Day_2`: The closing price of the stock on the *day 2*.

- `index_closing_prices.csv` contains the information of the index closing prices, given by the following two columns:
  - `Day`: This represents the i^{th} day, where $0 \leq i \leq 2$.
  - `Closing Price`: This is the index closing for the day.

Submission Details

You are required to upload the following three files:

- The output file, `index_constituents.csv` (max allowed size is `10MB`). The file should contain the quantity of shares of all the 51 shares on *day 1* and *day 2*.

  A valid output file has the following format:

```
Symbol,Quantity_Day_1,Quantity_Day_2
ACC,13,15
ADANIPORTS,118,120
AMBUJACEM,107,100
ASIANPAINT,63,60
AUROPHARMA,37,30
AXISBANK,235,246
BAJAJ-AUTO,19,21
.
.
.
TECHM,86,55
ULTRACEMCO,14,45
WIPRO,89,89
YESBANK,46,46
ZEEL,76,90
```

  Note that:

  - The first line of the output file should contain the header, with `Symbol`, `Quantity_Day_1`, and `Quantity_Day_2` as the column names separated by a `comma`.
  - There should be exactly 52 rows including the column header. The `Symbol` should be in alphabetically increasing order.
  - You shoud invest in all the 51 stocks.
  - You should not have more than INR 5 Lakhs cash in hand at the end of *day 2*.

- A *PDF* file (maximum allowed size is `4MB`) providing the findings and justification on the following topics:
  - Write a few lines about training dataset quality and any errors found in the training dataset.
  - Explain the data preprocessing steps.
  - Explain and justify the model you've chosen for calculating the index constituents.

- The source code of your approach for this task. Upload a *zip* file (maximum allowed size is `5MB`) with all relevant files to reproduce your results. The submitted file must have a `README` file with a detailed description about how to run the model to calculate the index constituents and generate the `index_constituents.csv`. Do not forget to include links to any external libraries or packages you use for the generation of your model.

  There is no limit on execution time, but the code should generate the output file: `index_constituents.csv`.

Evaluation

We calculate the total invested amount, A_i at the end of *day i* as:

$$A_i = \begin{cases} 4167296 & i = 0 \\ \sum_{k=1}^{51} Q_{k,\,1} \times P_{k,\,1} & i = 1 \\ \sum_{k=1}^{51} Q_{k,\,2} \times P_{k,\,2} & i = 2 \\ \sum_{k=1}^{51} Q_{k,\,2} \times P_{k,\,i} & 3 \leq i \leq 23 \end{cases}$$

Here,

- $Q_{k,\,i}$ is the total number of shares of the k^{th} stock at the end of the i^{th} day.
- $P_{k,\,i}$ is the closing price of the k^{th} stock at the end of the i^{th} day.

We calculate the total cash in hand, C_i at the end of *day i* as:

$$C_i = \begin{cases} 832704 & i = 0 \\ 75 \times 10^5 - A_1 & i = 1 \\ 10^7 - A_2 & i = 2 \\ C_2 & 3 \leq i \leq 23 \end{cases}$$

We calculate the portfolio value, W_i at the end of *day i* $(0 \leq i \leq 23)$ as:

$$W_i = A_i + C_i$$

We calculate the return, R_i for the index at the end of *day i* as:

$$R_i = \begin{cases} 5 \times 10^6 & i = 0 \\ R_0 \times \left(1 + \log_{10}\left(\frac{I_1}{I_0}\right)\right) + 25 \times 10^5 & i = 1 \\ R_1 \times \left(1 + \log_{10}\left(\frac{I_2}{I_1}\right)\right) + 25 \times 10^5 & i = 2 \\ R_{i-1} \times \left(1 + \log_{10}\left(\frac{I_i}{I_{i-1}}\right)\right) & 3 \leq i \leq 23 \end{cases}$$

Here, I_i is the index closing price at the end of i^{th} day.

Finally, we calculate the RMSE:

$$rmse = \sqrt{\frac{1}{21} \times \sum_{i=3}^{23} (W_i - R_i)^2}$$

Now, we calculate the normalized rmse:

$$normalized\_rmse = \frac{1}{10^7} \times rmse$$

Score is calculated as:

$$score = 1.0 - normalized\_rmse$$

Your leaderboard score will be $10^3 \times score$.

File Upload