

Lab 49 data Challenge

-Udit Anand

Contents

Problem Description	3
Input Provided	3
Solution	3
Project Files:.....	5
Important file and directories.....	5
Create a Python Virtual Environment	5
Run app	5
Improvements.....	5

Problem Description

1. Build a script that outputs the average return for an unknown number of Market Capitalization bins over an unknown time period. For your output, use 5 for the number of bins.
2. Generalize the solution to average over any factor, on bins of any factor.

Input Provided Time series data for public companies with a few relevant attributes added by another team - <https://github.com/lab49/DataLab49/>.

Solution

1. User provides input on the factor to average on. The default value is “Mkt Cap”.
2. The input factors available to select are:
 - a. Capitalization
 - b. Mkt Cap (Default)
 - c. FCF Yield
 - d. Momentum
 - e. Sales Growth 1Y
 - f. Sales Growth 5Y
 - g. EPS Revision
 - h. Date
3. User provides input on the number of bins. The default value is 5.
4. Based on the user input the chosen factor is divided into the chosen number of bins along with the corresponding returns.
5. The Aggregated values (sum, count and mean) are calculated for the binned returns. Please refer to Figure 1 for Steps 1-5.
6. For calculating the average return, we use the weighted average across the means calculated for specific bins. Weighted average is used to account for data concentration in bins, allocating more weight to bins with higher concentration. The following formula is used to calculate the weighted average.

$$\text{Weighted Average Return} = \sum_{i=1}^{i=\text{num_bin}} \text{mean}_{\text{bini}} * \text{count}_{\text{bini}} \div \sum_{i=1}^{i=\text{num_rows}} \text{count}$$

Figure 2 shows the average return computed

```
(lab49) udit@ubuntu:~/Desktop/Lab49/DataLab49/lab49/src$ python3 Main.py
[?] Please choose a factor. Default is Mkt Cap. : Mkt Cap
Capitalization
> Mkt Cap
FCF Yield
Momentum
Sales Growth 1Y
Sales Growth 5Y
EPS Revision
Date

[?] Please enter the number of bins. Default is 5. : 5
```

	Returns		
	sum	count	mean
Bin			
(-794.915, 159294.47]	290469.46	375609	0.773329
(159294.47, 318587.39]	699.67	1630	0.429245
(318587.39, 477880.31]	336.84	306	1.100784
(477880.31, 637173.23]	74.71	82	0.911098
(637173.23, 796466.15]	0.69	18	0.038333

Figure 1: Steps1-5

```
(lab49) udit@ubuntu:~/Desktop/Lab49/DataLab49/lab49/src$ python3 Main.py
[?] Please choose a factor. Default is Mkt Cap. : Mkt Cap
Capitalization
> Mkt Cap
FCF Yield
Momentum
Sales Growth 1Y
Sales Growth 5Y
EPS Revision
Date

[?] Please enter the number of bins. Default is 5. : 5
```

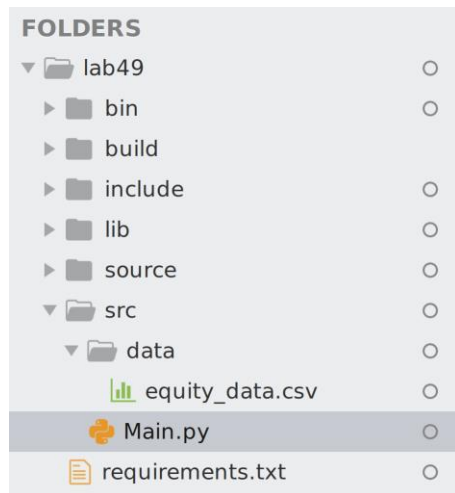
	Returns		
	sum	count	mean
Bin			
(-794.915, 159294.47]	290469.46	375609	0.773329
(159294.47, 318587.39]	699.67	1630	0.429245
(318587.39, 477880.31]	336.84	306	1.100784
(477880.31, 637173.23]	74.71	82	0.911098
(637173.23, 796466.15]	0.69	18	0.038333

Weighted Average Return : 0.772104

Figure 2: Weighted Average Return

Project Files:

A project directory called “Lab49” has been created to store code and other project specific information. The directory structure is as follows:



Important file and directories

- lab49: Parent directory
- lab49/src: Contains code and data used for the challenge
- lab49/src/Main.py: Contains code to the problem
- lab49/src/data: Contains any data files used for the challenge
- lab49/src/data/equity_data.csv: The data file provided for the challenge
- lab49/requirements.txt: Contains the project dependencies

Please note: The code is commented. Kindly refer to the lab49/src/Main.py for more details.

Steps to run code

Create a Python Virtual Environment

```
sudo pip install virtualenv  
  
virtualenv -p python3 <name of virtualenv>  
  
pip install -r requirements.txt
```

Run app

```
python3 main.py
```

Improvements

1. The distribution of the binning factor can be used to decide the size of the bin. This may reduce the computational time used to calculate the average returns over the binning factor. This approach will provide a more uniform number of elements in each bin.
2. The algorithm can be modified to compute the average returns over a specified time slot. Since as of now, the return is computed over the entire time frame of the data.

-
-
3. The algorithm can be further modified to provide the average return for specific ticker symbols.