Finance Index crawler: US Non-farm Payrolls
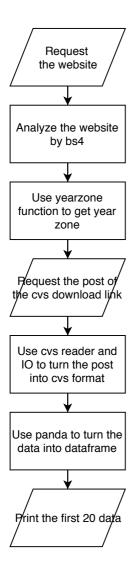
data: https://beta.bls.gov/dataViewer/view/timeseries/CES0000000001

Main Code

Request the website

↓

Analyze the website by bs4

↓

Use yearzone function to get year zone

↓

Request the post of the cvs download link

↓

Use cvs reader and IO to turn the post into cvs format

↓

Use panda to turn the data into dataframe

↓

Print the first 20 data

Function: yearzone

Input doc

↓

Set variables startyear & endyear

↓

Select nodes with name "option"

↓

for each item in nodes → Output [startyear. endyear]

↓

get the value of year the item as year and turn it into integer

↓

Startyear>year? — No →

Yes ↓

Startyear = year

↓

Endyear<year — No →

Yes ↓

Endyear = year