2023.04.18

받은 데이터 목록 (공유 드라이브 data폴더에 올려놓음)

allstock.csv

* WRDS에서 다운로드
* 2000.01.01 ~ 2022.12.30의 모든 주식 daily price data
* Price가 음수인 경우: daily price의 종가데이터가 모종의 이유로 없는 경우 bid/ask average 값을 (-)표시와 함께 기록

Sp500\_list\_start\_end.csv (직관적 이름 추천받아용 …)

* WRDS에서 다운로드 - Historical S&P 500 Index Constituents
* <https://wrds-www.wharton.upenn.edu/pages/wrds-research/applications/python-replications/historical-sp-500-index-constituents/>
* 2000.01.01 ~ 2022.12.30 동안 S&P 500에 들어간 적 있었던 모든 기업의 permno, S&P500목록에 포함되었던 기간인 start-end date data
* 같은 permno더라도, S&P 목록에 포함되었다가 미포함 되었다가 다시 포함된 경우 다른 start-end date를 가지고 있으므로 data에 중복되어 있을 것으로 추정

오늘 한 내용>

1. WRDS에서 위 두개의 데이터를 다운받고 sp500에 포함된 전적이 있는 permno를 두번째 데이터에서 추려 첫번째 데이터의 daily price 추출
2. 해당 기업들의 daily price data를 행을 날짜로 열을 permno로 한 dataframe으로 두고, 2000.01~2022.12.30의 모든 날짜에 대해 S&P에 속해 있지 않았던 기간의 daily price를 NaN으로 처리.
3. 행별로 NaN 아닌 애들의 개수를 모두 더하면 전부 500이 나와야 하는데 그렇지 않은 상황 – 두번째 데이터의 중복 때문으로 추정
4. S&P 목록 데이터를 다시 찾거나, 중복되어 있는 애들을 하나의 열로 잘 합쳐 주어야 함. 일단 더 찾아보기로