unstructured data on wikipedia.org

Data scraping

s&p list 2005-2020.json

Data mining

company_overview_list.json
daily adjusted stock timeseries.json

Data processing

s&p_list_top20_2005-2020.json

Data interpretation

s&p_top20_vs._ s&p500 notebook

- structure of the current S&P 500 and its changes in last 15 years are scraped from wikipedia.org
- processed into a list of all companies which form the true S&P lists for each year observed
- data mining of the alphavantage.co via API
- each company gets its overview and stock time series

- ➤ list of 20 biggest companies by market cap for each year
- market cap is evaluated from the company's overview (number of stocks) and its stock split history
- performance of the adjusted list of top 20 S&P companies vs. The whole S&P 500 is measured
- interpretation of the results, better investment conclusion