

# Week 8 Yahoo Finance Data

Group 22

April 29, 2021

## 1 Obtain Data

- Select companies via Yahoo Industry Filter
- Automatically crawl stock prices with Python

## 2 Issues

- Same query with different results
- Companies without geocode
- Profile

## 1 Obtain Data

- Select companies via Yahoo Industry Filter
- Automatically crawl stock prices with Python

## 2 Issues

- Same query with different results
- Companies without geocode
- Profile

# Get companies via Yahoo Industry Filter

- Example:

The screenshot shows the 'Applied Filters for Stocks screener' interface. The currency is set to USD. The filters applied are:

- Price (Intraday) greater than 0
- Sector is Basic Materials
- Industry is Agricultural Inputs
- Exchange is NasdaqGS, NYSE, Nasdaq, NasdaqGM, and NasdaqCM

There are buttons to 'Add another filter', 'Find Stocks', and 'Save As'. On the right, the 'Estimated results' are shown as 18.

- URL: [https://finance.yahoo.com/sector/ms\\_basic\\_materials](https://finance.yahoo.com/sector/ms_basic_materials)
- copy the table and use regex to get the company symbols
- Industry: agriculture, travel, energy

## 1 Obtain Data

- Select companies via Yahoo Industry Filter
- Automatically crawl stock prices with Python

## 2 Issues

- Same query with different results
- Companies without geocode
- Profile

# Automatically crawl stock prices with Python

- Input = the *txt* file of one industry
- Output1 = a dictionary {symbol: stock price history, ...}
- Output2 = csv for each company
- History = from 1970-1-1 to 2021-1-1

## 1 Obtain Data

- Select companies via Yahoo Industry Filter
- Automatically crawl stock prices with Python

## 2 Issues

- Same query with different results
- Companies without geocode
- Profile

# Same query with different results

- However, same query may return different table when filtering.

Applied Filters for Stocks screener Currency in USD

Price (Intraday) **greater than**

Sector is **Basic Materials** × +

Industry is **Agricultural Inputs** × +

Exchange is **NasdaqGS** × **NYSE** × **Nasdaq** × **NasdaqGM** × **NasdaqCM** × +

[+ Add another filter](#)

[Find Stocks](#) [Save As](#)

Estimated results  
**18**

Applied Filters for Stocks screener Currency in USD

Price (Intraday) **greater than**

Sector is **Basic Materials** × +

Industry is **Agricultural Inputs** × +

Exchange is **NasdaqCM** × **Nasdaq** × **NasdaqGS** × **NYSE** × **NasdaqGM** × +

[+ Add another filter](#)

[Find Stocks](#) [Save As](#)

Estimated results  
**208**



# Potential solutions

- try many times, and do intersection
- crawl data from the profile of each company to security its industry

## Yield10 Bioscience, Inc. (YTEN)

NasdaqGS - NasdaqGS Real Time Price. Currency in USD

★ In watchlist

Advertisement

**10.55** +0.07 (+0.67%)

As of 10:49AM EDT. Market open.

[Summary](#) [Chart](#) [Conversations](#) [Statistics](#) [Historical Data](#) [Profile](#) [Financials](#) [Analysis](#) [Options](#) [Holders](#) [Sustainability](#)

### Yield10 Bioscience, Inc.

19 Presidential Way

Woburn, MA 01801

United States

617 583 1700

<http://www.yield10bio.com>

Sector(s): **Basic Materials**

Industry: **Agricultural Inputs**

Full Time Employees: 25



## 1 Obtain Data

- Select companies via Yahoo Industry Filter
- Automatically crawl stock prices with Python

## 2 Issues

- Same query with different results
- Companies without geocode
- Profile

# Companies without geocode

- Listed in the U.S., but not located in the U.S.
- Two climate datasets are only about the U.S (wildfire, earthquake)
- Potential solutions:
  - don't care about the geocode
  - use profile like the previous

## Yield10 Bioscience, Inc. (YTEN)

NasdaqGS - NasdaqGS Real Time Price. Currency in USD

★ In watchlist

**10.55** +0.07 (+0.67%)

As of 10:49AM EDT. Market open.

Advertisement

Summary Chart Conversations Statistics Historical Data **Profile** Financials Analysis Options Holders Sustainability

### Yield10 Bioscience, Inc.

19 Presidential Way

Woburn, MA 01801

United States

617-583-1700

<http://www.yield10bio.com>

Sector(s): **Basic Materials**

Industry: **Agricultural Inputs**

Full Time Employees: 25

## 1 Obtain Data

- Select companies via Yahoo Industry Filter
- Automatically crawl stock prices with Python

## 2 Issues

- Same query with different results
- Companies without geocode
- Profile

- Many companies don't have profile page in Yahoo Finance now.
- Can't access all the pages once.
- Potential solutions:
  - use what we have, what we can ensure the industry and location
  - still works on it