source of historical stock data [closed]

Asked 15 years, 8 months ago   Modified 7 years, 2 months ago

Viewed 187k times

243

votes

I'm trying to make a stock market simulator (perhaps eventually growing into a predicting AI), but I'm having trouble finding data to use. I'm looking for a (hopefully free) source of historical stock market data.

Ideally, it would be a very fine-grained (second or minute interval) data set with price and volume of every symbol on NASDAQ and NYSE (and perhaps others if I get adventurous). Does anyone know of a source for such info?

I found this question which indicates Yahoo offers historical data in CSV format, but I've been unable to find out how to get it in a cursory examination of the site linked.

I also don't like the idea of downloading the data piecemeal in CSV files... I imagine Yahoo would get upset and shut me off after the first few thousand requests.

I also discovered another question that made me think I'd hit the jackpot, but unfortunately that OpenTick site seems to have closed its doors... too bad, since I think they were exactly what I wanted.

I'd also be able to use data that's just open/close price and volume of every symbol every day, but I'd prefer all the data if I can get it. Any other suggestions?
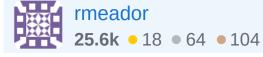
artificial-intelligence   simulation   finance   stocks

Share

edited May 23, 2017 at 11:55

Community Bot
1   1

asked Apr 16, 2009 at 3:01

rmeador
25.6k   18   64   104

5   @rmeador, Yahoo will not shut you off no matter how many requests you make, but Google will shut you off. I've been able to download about 4GB of EOD historical prices from Yahoo in about 5-6 hours without getting shut off. That's about 7,000 stocks with all of their EOD historical prices since they joined the market. See my answer for more information and sample source code. – Kiril Apr 24, 2010 at 17:37

1   I feel like EOD data isn't informative enough. If you want tick-by-tick quotes and trades, I believe polygon.io is the cheapest. – Quinton Pike Apr 24, 2017 at 16:06

I have found, this API has the cheapest and cleanest historical Price and Volume data I use for my backtesting. rapidapi.com/logicione/api/stock-price-and-volume-history – Watt Oct 1, 2023 at 20:09

Comments disabled on deleted / locked posts / reviews

18 Answers

Sorted by: Highest score (default) ⬍

232
votes

Let me add my 2¢, it's my job to get good and clean data for a hedge-fund, I've seen quite a lot of data feeds and historical data providers. This is mainly about US stock data.

To start with, if you have some money don't bother with downloading data from Yahoo, get the end of day data straight from **CSI data**, this is where Yahoo gets their EOD data as well AFAIK. They have an API where you can extract the data to whatever format you want. I think the yearly subscription for data is a few $100 bucks.

The main problem with downloading data from a free service is that you only get stocks that still exist, this is called [Survivorship Bias](#) and can give you wrong results if you look at many stocks, because you'll only include the ones that made it so far and not the ones that were de-listed.

For playing around with some intraday data I'd look into **[IQFeed](#)**, they provide several APIs to extract historical data, although they are mainly an outfit for real-time feeds. But here there are quite a few options, some brokers even provide historical data downloads via their APIs, so just pick your poison.

BUT usually all of this data is not very clean, once you really start back testing you'll see that certain stocks are missing or appear as two different symbols, or stock splits are not properly accounted for, etc. And then you realize that historical dividend data is need as well and so you start running in circles, patching data together from 100 different data sources and so on. So to start with a "discount" data feed will do, but as soon as you run more comprehensive backtests you might run into problems depending on what you do. If you just look at, let's say, the S&P 500 stocks this will not be so much a problem though and a "cheap" intraday feed will do.

What you will not find is free intraday data. I mean you might find some examples, I'm sure there's somewhere 5 years of MSFT tick data floating around but that will not get you very far.

Then, if you need the real stuff (level II order book, all ticks as they have happened at all exchanges) one "affordable", yet excellent option is **Nanex**. They'll actually ship you a drive with terabytes of data. If I remember right its about $3k-4K per year of data. But trust me, once you understand how hard it is to get good intraday data, you won't think this is very much money at all.

Not to discourage you but to get good data is hard, so hard in fact that many hedge-funds and banks spend hundreds of thousands of dollars *a month* to get data they can trust. Again, you can start somewhere and then go from there but it's good to see it a bit in context.

Edit: The answer above is from my own experience. [This write-up from Caltech](#) about available data feeds will give more insights, and especially recommends **QuantQuote**.

Share

edited Aug 23, 2016 at 20:38

answered Jun 23, 2013 at 17:01

lukebuehler

4,081 ● 3 ● 25 ● 28

---

2   CSI is great but, just FYI, delisted stocks are now a premium service, no longer included in the basic packages. Just FYI.
– Olie Nov 1, 2014 at 23:29

---

1   CSI is great but price is expensive. If you use Unfair Advantage, you are bound to their application. It's boring to use it everyday to update you history. If you want to download

from http or ftp with CSI, you must pay about 200€ on month. Sorry, but it's too expensive. – davidxxx Mar 14, 2015 at 20:27 ✎

@davidh, CSI Unfair Advantage has an ActiveX API where you can export all the data of your subscription automatically. It takes about one day to write a robust exporter tool... If you know of a cheaper alternative with the same quality as CSI, feel free to post an alternative! – lukebuehler Mar 16, 2015 at 13:50 ✎

Note about QuantQuote: they review/process your order within 48 hours after purchasing. In case you thought you'd have immediate access. – fionbio May 22, 2016 at 18:21

95
votes

🔖

↺

THIS ANSWER IS NO LONGER ACCURATE AS THE YAHOO FEED HAS CEASED TO EXIST

Using Yahoo's CSV approach above you can also get historical data! You can reverse engineer the following example:

http://ichart.finance.yahoo.com/table.csv?s=YHOO&d=0&e=28&f=2010&g=d&a=3&b=12&c=1996&ignore=.csv

Essentially:

```
sn = TICKER
a = fromMonth-1
b = fromDay (two digits)
c = fromYear
d = toMonth-1
e = toDay (two digits)
```

```
f = toYear
g = d for day, m for month, y for yearly
```

The complete list of parameters:

```
a    Ask
a2   Average Daily Volume
a5   Ask Size
b    Bid
b2   Ask (Real-time)
b3   Bid (Real-time)
b4   Book Value
b6   Bid Size
c    Change & Percent Change
c1   Change
c3   Commission
c6   Change (Real-time)
c8   After Hours Change (Real-time)
d    Dividend/Share
d1   Last Trade Date
d2   Trade Date
e    Earnings/Share
e1   Error Indication (returned for symbol changed /
invalid)
e7   EPS Estimate Current Year
e8   EPS Estimate Next Year
e9   EPS Estimate Next Quarter
f6   Float Shares
g    Day's Low
h    Day's High
j    52-week Low
k    52-week High
g1   Holdings Gain Percent
g3   Annualized Gain
g4   Holdings Gain
g5   Holdings Gain Percent (Real-time)
g6   Holdings Gain (Real-time)
i    More Info
i5   Order Book (Real-time)
j1   Market Capitalization
j3   Market Cap (Real-time)
j4   EBITDA
```

Share

This page has a table special tags that can be used in the URL. – user443854 Apr 4, 2013 at 21:06

10 A big problem with getting the data from Yahoo, or whatever online service, is that you do not get delisted stocks, so you'll quickly run into the survivorship bias. Better follw Eric H.'s or my advice and go straight to CSI. – lukebuehler Jun 23, 2013 at 17:09

This page used to have a table special tags that can be used in the URL... "Yahoo contends that your use and distribution of the tool and the content located at ... constitutes a breach of sections 6, 12, and 18 of the Terms of Service (among other provisions), gives rise to unfair competition, and induces others to breach the Terms of Service. ... By interfering with these contractual and business relationships, you are potentially harming the ability of other users to obtain the benefits of the services provided at the Yahoo Finance site". Thumbs down on Yahoo – Mustapha George Aug 30, 2015 at 11:27

I added the remaining switches from my notes, that used to be found on that web page. Presenting these here does not appear to be in breach with the ToS found here: policies.yahoo.com/us/en/yahoo/terms/product-atos/apiforydn/… Yahoo must have been upset about the Excel data tool that was also available on that site. – Fredrik E Sep 2, 2015 at 14:09 ✏

50
votes

I know you wanted "free", but I'd seriously consider getting the data from csidata.com for about $300/year, if I were you.

It's what yahoo uses to supply their data.

It comes with a decent API, and the data is (as far as I can tell) very clean.

You get 10 years of history when you subscribe, and then nightly updates afterward.

They also take care of all sorts of nasty things like splits and dividends for you. If you haven't yet discovered the joy that is data-cleaning, you won't realize how much you need this, until the first time your ATS (Automated Trading System) thinks some stock is really really cheap, only because it split 2:1 and you didn't notice.

Share

edited Dec 3, 2013 at 13:40

Kristoffer la Cour
2,586 ● 3 ● 25 ● 36

answered Jun 22, 2009 at 10:51

Eric H.
2,574 ● 2 ● 23 ● 34

1   they have a ActiveX API which you can call with c++ code or C# or whatever in windows to get to your data. – lukebuehler Jun 23, 2013 at 17:05

1   Interesting. How does the splits and dividends handling differ from Yahoo? – Matthew Lock Jun 28, 2013 at 2:10

1   @MatthewLock I'm not 100% about this but I think Yahoo uses a different "corporate events" data provider and then just uses the raw CSI data adjusted by the corporate events. This somewhat old article gives some insight: amibroker.org/userkb/2007/09/23/yahoos-data-providers – lukebuehler Jul 8, 2013 at 2:45

2   CSI is great but, just FYI, delisted stocks are now a premium service, no longer included in the basic packages. Just FYI. – Olie Nov 1, 2014 at 23:32

17

votes

Intro:

From yahoo you can get EOD (end of day) historical prices, or real-time prices. The EOD prices are amazingly simple to download. See my blog for explanations on how to get the data and for C# code examples.

I'm in the process of writing a real-time data feed "engine" that downloads and stores the real-time prices in a database. The engine will initially be able to download historical prices from Yahoo and Interactive Brokers and it will be able to store the data in a database of your choice: MS SQL, MySQL, SQLite, etc. It's open source, but I'll post more information on my blog when I get closer to releasing it (within a couple of days).

Another option is [eclipse trader](...)... it allows you to record the historical data with granularity as low as 1 minute and stores the prices locally in a text file. It basically downloads the real-time data from Yahoo with a 15 minute delay. Since I wanted a more robust solution and I'm working on a big school project for which we need data, I decided to write my own data feed engine (which I mentioned above).

Sample Code:

Here is sample C# code that demonstrates how to download real-time data:

```csharp
public void Start()
{
    string url =
"http://finance.yahoo.com/d/quotes.csv?
s=MSFT+GOOG&f=snl1d1t1ohgdr";
    //Get page showing the table with the chosen
indices
    HttpWebRequest request = null;
    IDatabase database =
        DatabaseFactory.CreateDatabase(
        DatabaseFactory.DatabaseType.SQLite);

    //csv content
    try
    {
        while (true)
        {
            using (Stream file =
File.Create("quotes.csv"))
            {
                request =
(HttpWebRequest)WebRequest.CreateDefault(new
Uri(url));
                request.Timeout = 30000;
                using (var response =
(HttpWebResponse)request.GetResponse())
                    using (Stream input =
```

```
response.GetResponseStream())
            {
                CopyStream(input, file);
            }
        }
        Console.WriteLine("-------------------
---------------------------");

database.InsertData(Directory.GetCurrentDirectory()
+ "/quotes.csv");
```

Database:

On the database side I use an `OleDb` connection to the CSV file to populate a `DataSet` and then I update my actual database via the `DataSet`, it basically makes it possible to match all of the columns from the CSV file returned from Yahoo directly to your database (if your database does not support batch inserts of CSV data, like SQLite). Otherwise, inserting the data is a one-liner... just batch insert the CSV into your database.

You can read more about the formatting of the url here: http://www.gummy-stuff.org/Yahoo-data.htm

Share

edited Apr 24, 2010 at 17:33

answered Apr 24, 2010 at 17:27

Kiril

40.3k ● 32 ● 175 ● 227

Does that actually provide real time data like you suggested? From the page, it does have this parameter "k1", but last time I checked, it still has some delay. – Antony Jun 13, 2011 at 3:04

@Antony most of the time there is a delay of some sort, so it just depends on how tolerant you are to the delays. Yahoo does say that they provide real time data, but it certainly isn't for all of the tickers. The tickers that are not real time are delayed by up to 15 minutes. Even if you get a co-located server on the exchange, there will STILL be **"some delay"**. So what kind of a delay are you willing to tolerate? – Kiril Jun 13, 2011 at 13:17

It's not reliable. Tell me, why doesn't this work right now for instance: real-chart.finance.yahoo.com/… used from finance.yahoo.com/q/… – mjs Jan 28, 2015 at 14:14

16

votes

A data set of every symbol on the NASDAQ and NYSE on a second or minute interval is going to be **massive**.

Let's say there are a total of 4000 companies listed on both exchanges (this is probably on the very low side since there are over 3200 companies listed on the NASDAQ). For data at a second interval, assuming there are 6.5 trading hours in a day, that would give you 23400 data points per day per company, or about 93,600,000 data points in total for that one day. Assuming 200 trading days in a year, thats about 18,720,000,000 data points for just one year.

Maybe you want to start with a smaller set first?

Share

answered Apr 16, 2009 at 3:13

matt b
140k ● 66 ● 284 ● 350

2   I was operating under the assumption that most of the companies would not be traded every second, so the number of data points would be significantly less. perhaps that's a bad

assumption. still, I was predicting on the order of 10s of GB per year... – rmeador Apr 16, 2009 at 3:16

One a couple months of stock data for like 10 symbols came on 3 DVDs. The data was compressed text as well. – Alan Apr 16, 2009 at 3:47

1 @rmeador thats true, but also some stocks have way more daily volume than there are seconds in a day, meaning they trade more than once a second - and not all trades are guaranteed to be at the same price. So you'd have to decide if you're interested in the price at an interval, or at trade – matt b Apr 16, 2009 at 3:49

2 If you want the whole thing, e.g., level II quotes of all exchanges etc its a few TBs for a year in a suuuuper compressed format (about 5GB per trading day). If you only store minute data its really little, about 10GBs for 10 years of all stocks... – lukebuehler Jun 23, 2013 at 17:22

8
votes

For survivorship bias free data, the only reliable source I have found is QuantQuote (http://quantquote.com)

Data comes in minute, second, or tick resolution, link to their historical stock data.

There was a suggestion for kibot above. I would do a quick google search before buying from them, you'll find lots posts like this with warnings about kibot data quality problems. It is also telling that their supposedly survivorship bias free sp500 only has 570 symbols for 14 years. That's pretty much impossible, sp500 changes by 1-2 symbols per *month*....

edited Sep 21, 2012 at 22:01

answered Sep 14, 2012 at 16:15

user788171

17.5k ● 43 ● 99 ● 127

---

3 kibot has only 3 free symbols. the rest have to pay! he is just doing advertisment – bouncingHippo Nov 21, 2012 at 15:47

quantquote's free daily data is undocumented: there are no column headers in the csv files, and no doc whatsoever. – user443854 Dec 5, 2012 at 22:04

there is documentation, the format is basically the same as their minute resolution datasets. – user788171 Jul 3, 2013 at 0:52

quantquote.com has lots of errors in the data – Ivelin Aug 7, 2014 at 3:28

They offer free minutely data for IBM since 1998. It was good enough for me. I only needed one symbol to test compression: kibot.com/buy.aspx – Sergei Rodionov Aug 25, 2017 at 16:10

8

votes

NASDAQ offers 10 years of historical EOD data for each symbol

http://www.nasdaq.com/aspx/historical_quotes.aspx?symbol=AAPL&selected=AAPL

You could automate the process of downloading this data.

Nice source. They changed the request system, so now all requests are processed with JS (eg nasdaq.com/symbol/aapl/historical). Is there any way to automatize it? – Anton Tarasenko Sep 29, 2016 at 17:25

Yes, it is possible to automatize it. You need to use a sniffer such as Telerik and see how the data is obtained (via a POST). As long as you send the post variables right, you should get the data. – Nick_F Jun 9, 2017 at 2:21

1 Is this still available? Does it contain any intraday data? – Lee Dec 29, 2017 at 10:48

7
votes

Unfortunately historical ticker data that is free is hard to come by. Now that opentick is dead, I dont know of any other provider.

In a previous lifetime I worked for a hedgefund that had an automated trading system, and we used historical data profusely.

We used TickData for our source. Their prices were reasonable, and the data had sub second resolution.

6

votes

We have purchased 12 years of intraday data from Kibot.com and are pretty satisfied with the quality.
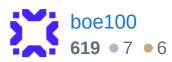
As for storage requirements: 12 years of 1-minute data for all USA equities (more than 8000 symbols) is about 100GB.

With tick-by-tick data situation is little different. If you record time and sales only, that would be about 30GB of data per month for all USA equities. If you want to store bid / ask changes together with transactions, you can expect about 150GB per month.

I hope this helps. Please let me know if there is anything else I can assist you with.

Share

answered Dec 30, 2009 at 9:25

boe100

619 ● 7 ● 6

> Both adjusted and unadjusted data is available. It is possible to update your data using an HTTP API or download new archives from FTP server daily. No betas or deltas are calculated.
> – boe100 Sep 25, 2013 at 8:20

6

votes

Let me add a source I just discovered, found [here](#).

It has lots of historical stock data in csv format and was gathered by Andy Pavlo, who according to his homepage is

an "Assistant Professor in the Computer Science Department at Carnegie Mellon University".

Share

answered Sep 21, 2014 at 1:37

Noam
617 ● 1 ● 6 ● 12

> This is great for anyone simply looking to mess around with a large enough set of historical stock market data.
> – Ciaran Gallagher Mar 7, 2017 at 21:53

> 1   Webpage down... – Brethlosze Nov 10, 2017 at 17:56

> 1   cs.cmu.edu/~pavlo/datasets/stocks – miodf Dec 3, 2017 at 13:40

5
votes

Mathematica nowoadays also offers access to both current and historical stock prices, see http://reference.wolfram.com/mathematica/ref/FinancialData.html , if you happen to have a copy of it.

Share

answered Dec 29, 2012 at 13:34

Tom Wenseleers
7,979 ● 7 ● 69 ● 113

> 5   the smallest time step is `day` – kirill_igum May 19, 2013 at 20:37

4
votes

You can use yahoo to get daily data (a much more managable dataset) but you have to structure the urls. See this link. You are not making lots of little requests you are

making a fewer large requests. Lot of free software uses this so they shouldn't shut you down.

EDIT: [This](#) guy does it, maybe you can have a look at the calls his software makes.

Share

answered Apr 16, 2009 at 3:23

jimconstable
2,388 ● 19 ● 28

at first I thought that link looked promising, but I can't seem to find how to specify historical data... it looks like it's all real-time. Am I missing something? – rmeador Apr 16, 2009 at 3:35

you are right. I have added another link of someone with software that does the historical stuff so I know it is possible. Maybe have a look at the calls his software makes.
– jimconstable May 27, 2009 at 4:09

4

votes

Yahoo is the simplest option to get preliminary free data. The link described in eckesicle's answer could be easily used in a python code, but you first need all the tickers. I'd use the NYSE for this example, but this can be used for different exchanges as well.

I used [this wiki page](#) to download all company tickers with the following script (I'm not a very talented Pythonist, sorry if this code isn't very efficient):

```
import string
import urllib2
from bs4 import BeautifulSoup

global f

def download_page(url):
    aurl = urllib2.urlopen(url)
    soup = BeautifulSoup(aurl.read())

    print url

    for row in soup('table')[1]('tr'):
        tds = row('td')
        if (len(tds) > 0):
            f.write(tds[1].string + '\n')


f = open('stock_names.txt', 'w')

url_part1 =
'http://en.wikipedia.org/wiki/Companies_listed_on_the_
url = url_part1 + '(0-9)'
download_page(url)

for letter in string.uppercase[:26]:
    url_part2 = letter
    url = url_part1 + '(' + letter + ')'

    download_page(url)

f.close()
```

For downloading each ticker I used another quite similar
script:

```
import string
import urllib2
from bs4 import BeautifulSoup
```

```
global f

url_part1 =
'http://ichart.finance.yahoo.com/table.csv?s='
url_part2 =
'&d=0&e=28&f=2010&g=d&a=3&b=12&c=1996&ignore=.csv'

print "Starting"

f = open('stock_names.txt', 'r')
file_content = f.readlines()
count = 1;
print "About %d tickers will be downloaded" %
len(file_content)

for ticker in file_content:
    ticker = ticker.strip()
    url = url_part1 + ticker + url_part2

    try:
        # This will cause exception on a 404
        response = urllib2.urlopen(url)

        print "Downloading ticker %s (%d out of
%d)" % (ticker, count, len(file_content))

        count = count + 1
        history_file =
open('C:\\Users\\Nitay\\Desktop\\Historical Data\\'
+ ticker + '.csv', 'w')
        history_file.write(response.read())
        history_file.close()
```

Notice that the major downside to this method is that different data is available for different companies - Companies that don't have data existing in the requested dates (newly listed) will get you a 404 page.

Also keep in mind that this method is only good for preliminary data - If you really want to test your algorithm
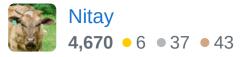
you should pay a bit and use a trusted data supplier like CSIData or others

Share

edited Jun 20, 2020 at 9:12

Community Bot

1 ● 1

answered Nov 15, 2013 at 8:05

Nitay

4,670 ● 6 ● 37 ● 43

Putting a global declaration inside global namespace is unnecessary, good response anyway though. – Luke Taylor Dec 17, 2015 at 2:38

1   Service down... – Brethlosze Nov 10, 2017 at 17:56

3

votes

I'd crawl finance.google.com (for the quotes) - or finance.yahoo.com.

Both these will return html pages for most exchanges around the world, including historical. Then, it's just a matter of parsing the HTML to extract what you need.

I've done this in the past, with great success. Alternatively, if you don't mind using Perl - there are several modules on CPAN that have done this work for you - i.e. extracting quotes from Google/Yahoo.

For more, see Quote History

Share

answered Apr 16, 2009 at 3:12

3 votes

I use the eodData.com. Its pretty decently priced. For 30 dollars a month you get 30 days of 1,5 and 60 minute bars for all US exchanges and 1 year of EOD data for most others.

Share

answered Aug 10, 2012 at 0:13

Dez Udezue
759 ● 1 ● 7 ● 16

2 votes

Why not model a fake stock market with Brownian Motion?

Plenty of resources for doing it. Easy to implement.

http://introcs.cs.princeton.edu/java/98simulation/

Share

answered Jun 4, 2011 at 8:36

Michael Thamm
87 ● 1

1   :-) to make it more real you'd need to create Fractional Brownian motion and even that is not quite real, for the most realistic fake market data you'd need to have also a fractal time dimension... needles to say it gets quite complicated. Better just buy real market data... – lukebuehler Jun 24, 2013 at 15:38

1   It also doesn't help that stock movement is not lognormal :) – Paul Milovanov Mar 23, 2016 at 2:19

1

vote

A former project of mine was going to use freely downloadable data from [EODData](#).

Share

answered Aug 25, 2009 at 6:15

Shaggy Frog
27.6k ● 16 ● 92 ● 129

1

vote

Take a look at the Mergent Historical Securities Data API - http://www.mergent.com/servius

Share

answered Oct 18, 2010 at 23:17

Eugene Osovetsky
6,541 ● 2 ● 40 ● 61