Setting up a Quant Trading Database

In order to develop your own quant trading strategies

The first step is to set up a database with historical price data

Set up a database with historical price data

Programmatically download historical price information

Set up a database with historical price data

Programmatically download historical price information

Insert the downloaded files into a database

Set up a database with historical price data

Programmatically download historical price information

Insert the downloaded files into a database

Clean up the data and put it in a ready-to-consume form

Set up a database with historical price data

Programmatically download historical price information

Insert the downloaded files into a database

Clean up the data and put it in a ready-to-consume form

Set up a database with historical price data

Programmatically download historical price information

Insert the downloaded files into a database

Clean up the data and put it in a ready-to-consume form

Programmatically download historical price information

Historical price information is available publicly

National Stock Exchange of India (NSE)

Yahoo Finance for International Stocks

Programmatically download historical price information

National Stock Exchange

Yahoo Finance

Python libraries like urllib2 and BeautifulSoup can be used to download / scrape information from websites

Programmatically download historical price information

National Stock Exchange

Yahoo Finance

Some websites like the NSE do not allow programs to directly download files

You need to make the NSE feel like it is a human and not a machine that's downloading the data

Programmatically download historical price information

National Stock Exchange

Yahoo Finance

There are different types of price information files

Programmatically download historical price information

National Stock Exchange

Yahoo Finance

On Yahoo Finance, you'll download 1 file for each security

Programmatically download historical price information

National Stock Exchange

NSE publishes 1 file each day for each type of security

Yahoo Finance

1 file for each security

Programmatically download historical price information

National Stock Exchange

1 file each day

Yahoo Finance

1 file for each security

Cash Markets file contains price movements for stocks that trade on the NSE

Programmatically download historical price information

National Stock Exchange

1 file each day

Cash Markets

Yahoo Finance

1 file for each security

Futures & Options file contains price movements for futures and options that trade on the NSE

Programmatically download historical price information

National Stock Exchange

1 file each day

Cash Markets

Futures & Options

Yahoo Finance

1 file for each security

The URL of these different types of files has a distinct format

Programmatically download historical price information

National Stock Exchange

1 file each day

Cash Markets

Futures & Options

Yahoo Finance

1 file for each security

The Indices file has the price information for various indices - NIFTY, BANKNIFTY etc

Programmatically download historical price information

National Stock Exchange

1 file each day

Cash Markets

Futures & Options

Indices

Yahoo Finance

1 file for each security

Using the type of security, date, we can construct the URL that corresponds to the file

Programmatically download historical price information

National Stock Exchange

Yahoo Finance

Download files from the last 10 years

Jan 1, 2006 onwards

Programmatically download historical price information

National Stock Exchange

Yahoo Finance

Some data might not be easily available for all 10 years

NIFTY index prices before 2013 will need to be manually downloaded

1 file per year

Step 1: Set up a database with historical price data

Programmatically download historical price information

Insert the downloaded files into a database

Clean up the data and put it in a ready-to-consume form

Step 1: Set up a database with historical price data

Programmatically download historical price information

Insert the downloaded files into a database

Clean up the data and put it in a ready-to-consume form

Insert the downloaded files into a database

In a MySQL database, create tables to hold prices for each type of security

NSE CM stocks

NSE Futures and Options

NSE Indices

International Stocks

Insert the downloaded files into a database

NSE CM stocks

NSE Futures and Options

NSE Indices

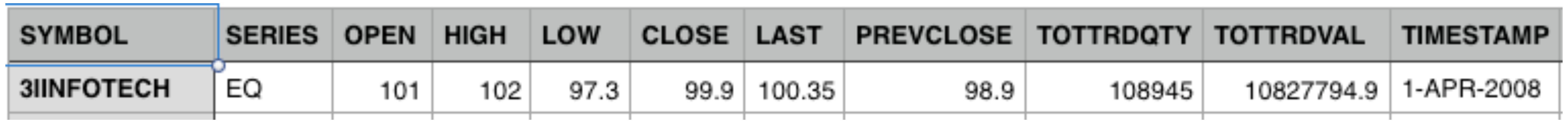
International Stocks

The downloaded files for each of these have different formats

Insert the downloaded files into a database

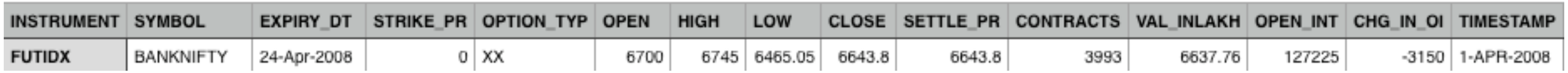
Different Formats

NSE CM stocks



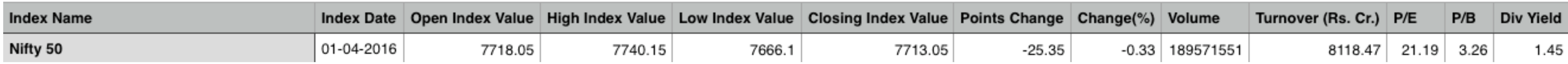
SYMBOL	SERIES	OPEN	HIGH	LOW	CLOSE	LAST	PREVCLOSE	TOTTRDQTY	TOTTRDVAL	TIMESTAMP
3IINFOTECH	EQ	101	102	97.3	99.9	100.35	98.9	108945	10827794.9	1-APR-2008

NSE Futures and Options



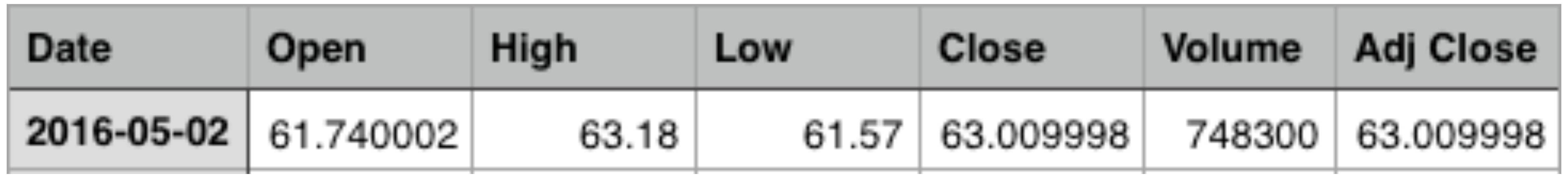
INSTRUMENT	SYMBOL	EXPIRY_DT	STRIKE_PR	OPTION_TYP	OPEN	HIGH	LOW	CLOSE	SETTLE_PR	CONTRACTS	VAL_INLAKH	OPEN_INT	CHG_IN_OI	TIMESTAMP
FUTIDX	BANKNIFTY	24-Apr-2008	0	XX	6700	6745	6465.05	6643.8	6643.8	3993	6637.76	127225	-3150	1-APR-2008

NSE Indices



Index Name	Index Date	Open Index Value	High Index Value	Low Index Value	Closing Index Value	Points Change	Change(%)	Volume	Turnover (Rs. Cr.)	P/E	P/B	Div Yield
Nifty 50	01-04-2016	7718.05	7740.15	7666.1	7713.05	-25.35	-0.33	189571551	8118.47	21.19	3.26	1.45

International Stocks



Date	Open	High	Low	Close	Volume	Adj Close
2016-05-02	61.740002	63.18	61.57	63.009998	748300	63.009998

Insert the downloaded files into a database

NSE CM stocks

NSE Futures and Options

NSE Indices

International Stocks

The tables corresponding to these files will be based on the format

Insert the downloaded files into a database

NSE CM stocks

NSE Futures and Options

NSE Indices

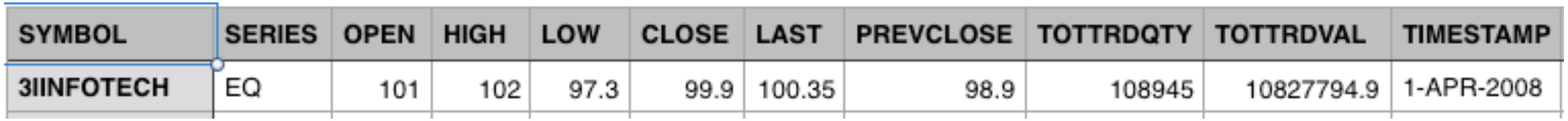
International Stocks

Once you have created the table, use a Python program to iterate through all the files and insert them into the right table

Insert the downloaded files into a database

NSE CM stocks

Let's take one example



SYMBOL	SERIES	OPEN	HIGH	LOW	CLOSE	LAST	PREVCLOSE	TOTTRDQTY	TOTTRDVAL	TIMESTAMP
3IINFOTECH	EQ	101	102	97.3	99.9	100.35	98.9	108945	10827794.9	1-APR-2008

This creates the cmStaging table with the columns corresponding to the CM files

```
create table cmStaging (
symbol varchar(256),
series varchar(256),
open float,
high float,
low float,
close float,
last float,
prevclose float,
tottrdqty float,
tottrdval float,
timestamp date,
totaltrades float,
isin varchar(256));
```

Insert the downloaded files into a database

NSE CM stocks

cmStaging

In Python, we can connect to the mySQL database and get a cursor to interact with the database

Then we iterate through all the downloaded files and insert each file into the table

Insert the downloaded files into a database

NSE CM stocks

cmStaging

We can bulk load a file into an sql table if we know how the columns in each map to each other

```
c.execute("Load data local infile %s
into table cmStaging fields terminated by %s
ignore 1 lines
(symbol,series,open,high,low,close,last,prevclose
,tottrdqty,tottrdval,@timestamp,totaltrades,isin)
SET timestamp = STR_TO_DATE(@timestamp, %s)",
(fileName,delimiter,dateString))
```

Insert the downloaded files into a database

NSE CM stocks

cmStaging

```
c.execute("Load data local infile %s
into table cmStaging fields terminated by %s
ignore 1 lines
(symbol,series,open,high,low,close,last,prevclose,tottrdqty,tott
rdval,@timestamp,totaltrades,isin)
SET timestamp = STR_TO_DATE(@timestamp, %s)",
(fileName,delimiter,dateString))
```

c is the cursor we use to execute queries on the database

Insert the downloaded files into a database

NSE CM stocks

cmStaging

```
c.execute("Load data local infile %s
into table cmStaging fields terminated by %s
ignore 1 lines
(symbol,series,open,high,low,close,last,prevclose,tottrdqty,tott
rdval,@timestamp,totaltrades,isin)
SET timestamp = STR_TO_DATE(@timestamp, %s)",
(fileName,delimiter,dateString))
```

c.execute will execute the given query and return results if any

Insert the downloaded files into a database

NSE CM stocks

cmStaging

```
c.execute("Load data local infile %s
into table cmStaging fields terminated by %s
ignore 1 lines
(symbol,series,open,high,low,close,last,prevclose,tottrdqty,tottrdval
,@timestamp,totaltrades,isin)
SET timestamp = STR_TO_DATE(@timestamp, %s)",
(fileName,delimiter,dateString))
```

This query will be run

Insert the downloaded files into a database

NSE CM stocks

cmStaging

```
c.execute("Load data local infile %s

into table cmStaging fields terminated by %s
ignore 1 lines
(symbol,series,open,high,low,close,last,prevclose,tottrdqty,tottrd
val,@timestamp,totaltrades,isin)

SET timestamp = STR_TO_DATE(@timestamp, %s)",
(fileName,delimiter,dateString))
```

We can use %s to pass in variables which are used to construct the query

Insert the downloaded files into a database

NSE CM stocks

cmStaging

```
c.execute("Load data local infile %s
into table cmStaging fields terminated by %s
ignore 1 lines
(symbol,series,open,high,low,close,last,prevclose,tottrdqty,to
ttrdval,@timestamp,totaltrades,isin)
SET timestamp = STR_TO_DATE(@timestamp, %s)",
(fileName,delimiter,dateString))
```

This will read the data in file specified fileName into the table cmStaging

Insert the downloaded files into a database

NSE CM stocks

cmStaging

```
c.execute("Load data local infile %s
into table cmStaging fields terminated by %s
ignore 1 lines
(symbol,series,open,high,low,close,last,prevclose,tottrdqty,tottrd
val,@timestamp,totaltrades,isin)
SET timestamp = STR_TO_DATE(@timestamp, %s)",
(fileName,delimiter,dateString))
```

This specifies the delimiter string and that the file contains a header

Insert the downloaded files into a database

NSE CM stocks

cmStaging

```
c.execute("Load data local infile %s
into table cmStaging fields terminated by %s
ignore 1 lines
(symbol,series,open,high,low,close,last,prevclose,tottrdqty,tottrdval,@timestamp,totaltrades,isin)
SET timestamp = STR_TO_DATE(@timestamp, %s)",(fileName,delimiter,dateString))
```

Here we map the columns of the file in order to the columns in the SQL table

Insert the downloaded files into a database

NSE CM stocks

cmStaging

```
c.execute("Load data local infile %s
into table cmStaging fields terminated by %s
ignore 1 lines
(symbol,series,open,high,low,close,last,prevclose,tottrdqty,tottrdval,@timestamp,totaltrades,isin)
SET timestamp = STR_TO_DATE(@timestamp, %s)",(fileName,delimiter,dateString))
```

The position in the list corresponds to a column in the csv

The name corresponds to the name in the SQL table

Insert the downloaded files into a database

NSE CM stocks

cmStaging

```
c.execute("Load data local infile %s
into table cmStaging fields terminated by %s
ignore 1 lines
(symbol,series,open,high,low,close,last,prevclose,tottrdqty,tottrdval,@timestamp,totaltrades,isin)
SET timestamp = STR_TO_DATE(@timestamp, %s)",(fileName,delimiter,dateString))
```

Strings and numbers are parsed automatically

Insert the downloaded files into a database

NSE CM stocks

cmStaging

```
c.execute("Load data local infile %s
into table cmStaging fields terminated by %s
ignore 1 lines
(symbol,series,open,high,low,close,last,prevclose,tottrdqty,tottrdval,@timestamp
,totaltrades,isin)
SET timestamp = STR_TO_DATE(@timestamp, %s)",
(fileName,delimiter,dateString))
```

“%d-%b-%Y”

This converts the timestamp column to a date based on the format specified

Insert the downloaded files into a database

NSE CM stocks

NSE Futures and Options

NSE Indices

International Stocks

We'll repeat this exercise with each security type

Insert the downloaded files into a database

NSE CM stocks

NSE Futures and Options

NSE Indices

International Stocks

The NSE CM and FO files need to be unzipped before they can be inserted into the database

Step 1: Set up a database with historical price data

Programmatically download historical price information

Insert the downloaded files into a database

Clean up the data and put it in a ready-to-consume form

Step 1: Set up a database with historical price data

Programmatically download historical price information

Insert the downloaded files into a database

Clean up the data and put it in a ready-to-consume form

Clean up the data and put it in a ready-to-consume form

There are 3 things we need to do before the data is ready to consume

Clean up the data and put it in a ready-to-consume form

3 things we need to do

1. Remove any duplicates present in the data

2. Adjust for any symbol changes over time

3. Adjust for corporate actions

Clean up the data and put it in a ready-to-consume form

3 things we need to do

1. Remove any duplicates present in the data
2. Adjust for any symbol changes over time
3. Adjust for corporate actions

1. Remove any duplicates present in the data

We need to remove any duplicate rows which are present in the MySql tables

1. Remove any duplicates present in the data

a. Create a duplicate table

b. Add a Unique index constraint on that table

c. Insert rows from the old table into the new table

Any duplicates will be ignored

Clean up the data and put it in a ready-to-consume form

3 things we need to do

1. Remove any duplicates present in the data

2. Adjust for any symbol changes over time

3. Adjust for corporate actions

Clean up the data and put it in a ready-to-consume form

3 things we need to do

1. Remove any duplicates present in the data

2. Adjust for any symbol changes over time

3. Adjust for corporate actions

2. Adjust for any symbol changes over time

All the price information is keyed by the Symbol of a security

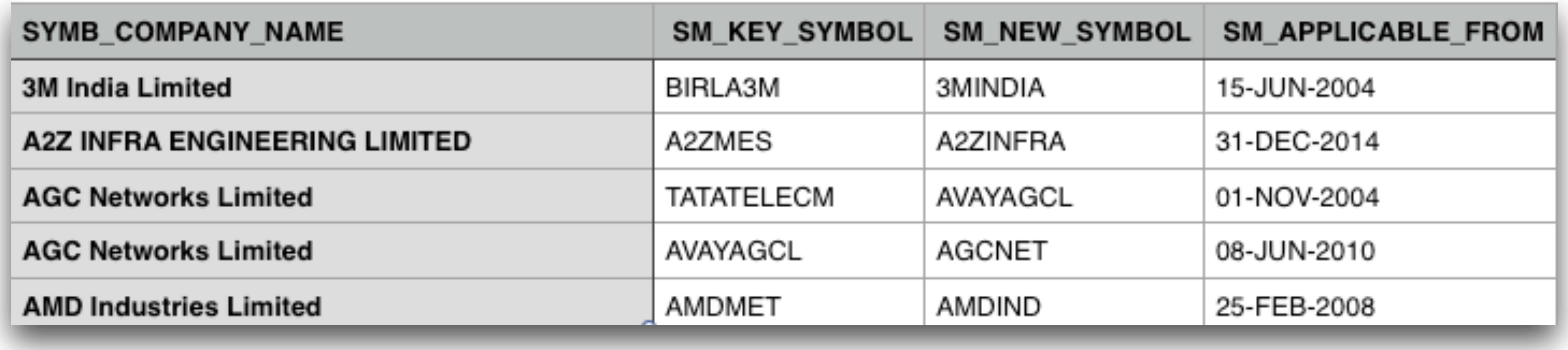
2. Adjust for any symbol changes over time

If the symbol of a company has changed in the past,

Then the old symbol name needs to be updated to the new symbol name

2. Adjust for any symbol changes over time

The NSE maintains a log of symbol changes in a csv file



SYMB_COMPANY_NAME	SM_KEY_SYMBOL	SM_NEW_SYMBOL	SM_APPLICABLE_FROM
3M India Limited	BIRLA3M	3MINDIA	15-JUN-2004
A2Z INFRA ENGINEERING LIMITED	A2ZMES	A2ZINFRA	31-DEC-2014
AGC Networks Limited	TATATELECM	AVAYAGCL	01-NOV-2004
AGC Networks Limited	AVAYAGCL	AGCNET	08-JUN-2010
AMD Industries Limited	AMDMET	AMDIND	25-FEB-2008

We can use this to update the symbol wherever it has changed

Clean up the data and put it in a ready-to-consume form

3 things we need to do

1. Remove any duplicates present in the data

2. Adjust for any symbol changes over time

3. Adjust for corporate actions

Clean up the data and put it in a ready-to-consume form

3 things we need to do

1. Remove any duplicates present in the data

2. Adjust for any symbol changes over time

3. Adjust for corporate actions

3. Adjust for corporate actions

Corporate actions such as splits, dividends etc affect the stock price

The price data needs to be adjusted for these actions

3. Adjust for corporate actions

Let's understand in detail
how stock splits affect prices

Stock Splits

A split occurs when a company divides its existing shares into multiple shares

Stock Splits

For example: in a 2-for-1 split

The investors now own 2 shares for each share they originally owned

Stock Splits

For example: in a 2-for-1 split

The total value of the shares remains the same

The number of shares doubles

The price of each share halves

Stock Splits

Why do companies do stock splits?

1. When the price of the share is too high for investors to easily buy lots of 100 shares

Stock Splits

Why do companies do stock splits?

2. To increase the liquidity of a stock

Liquidity represents how easy it is to buy or sell an asset

Stock Splits

Why do companies do stock splits?

2. To increase the liquidity of a stock

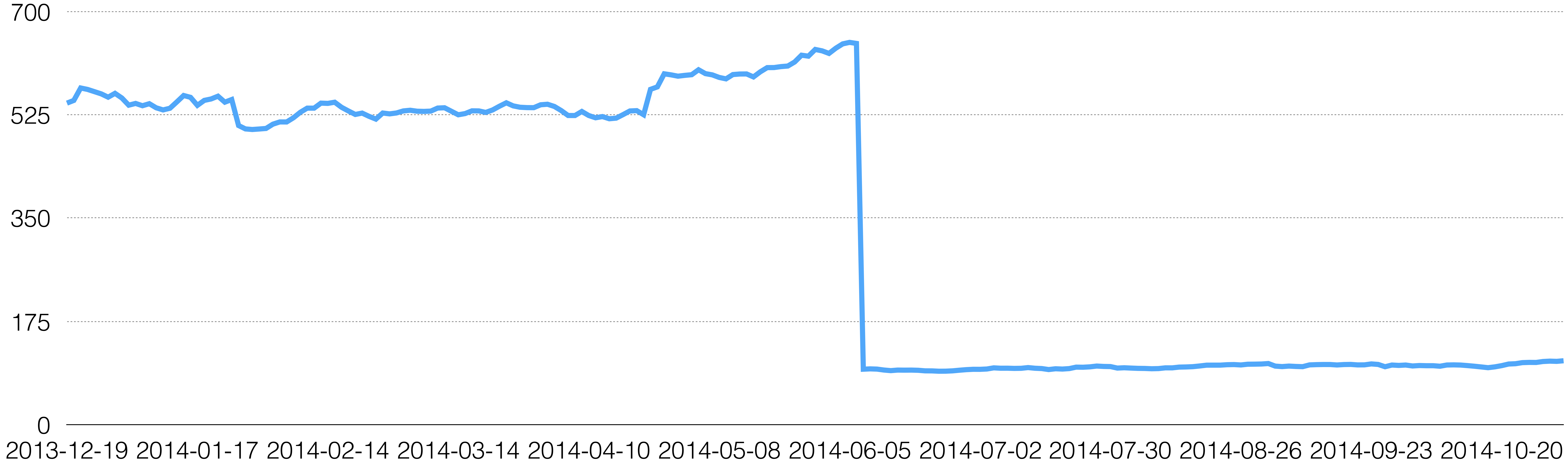
The larger the number of shares available, the higher the liquidity

Stock Splits

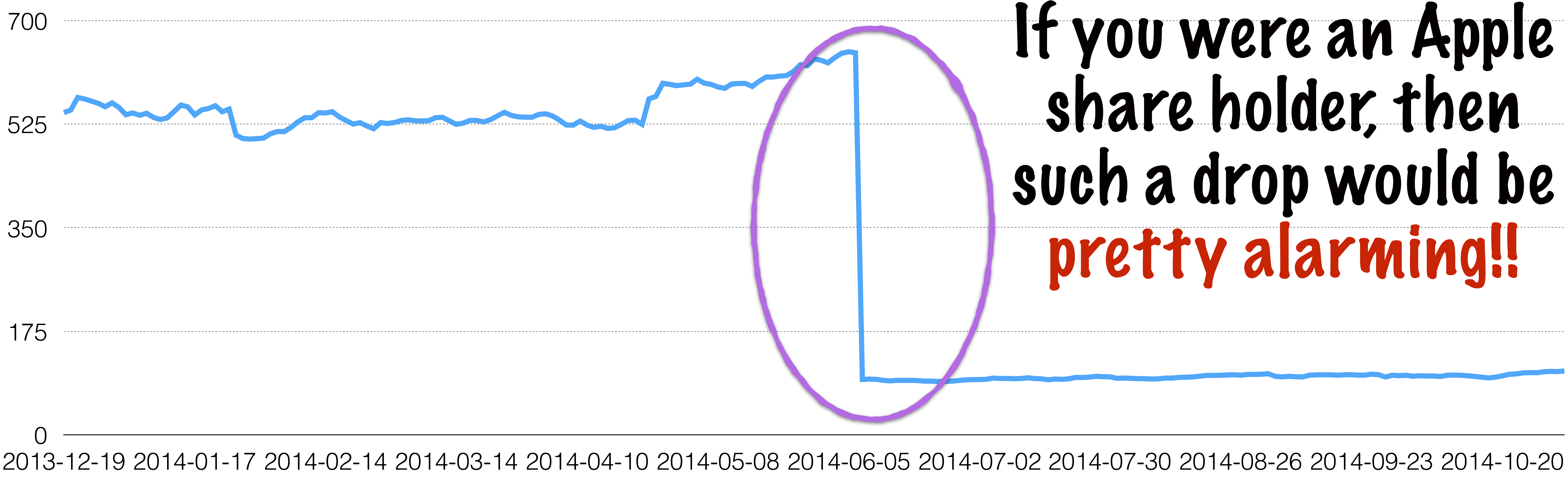
Why do we need to adjust for stock splits?

Stock Splits

Let's say you are looking at a time series for Apple stock prices in 2014



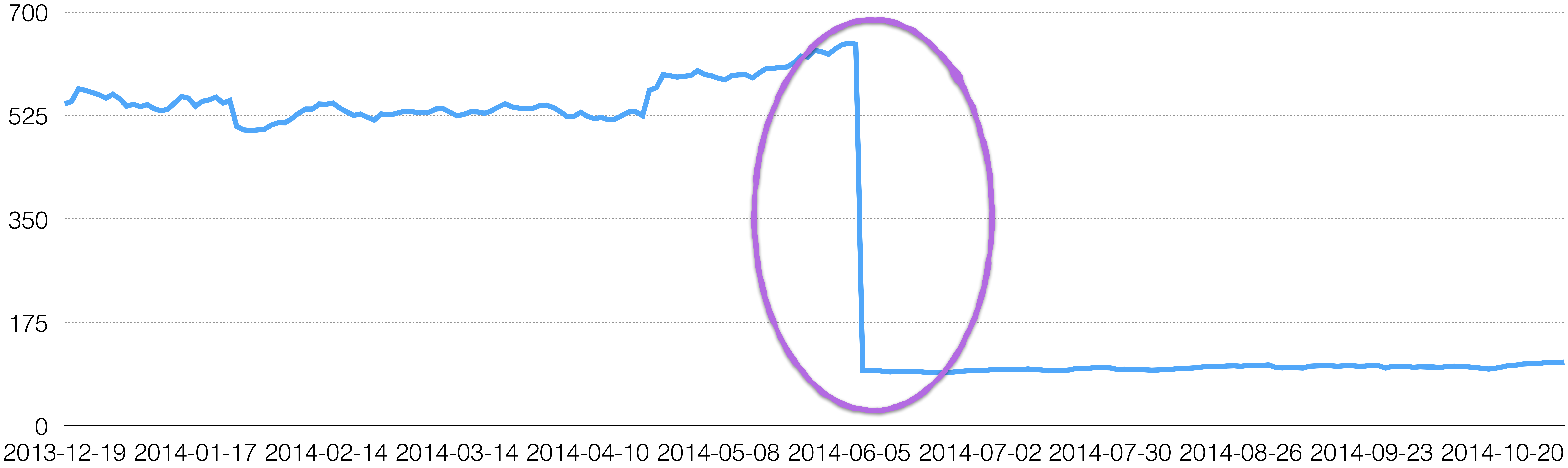
Stock Splits



If you were an Apple share holder, then such a drop would be **pretty alarming!!**

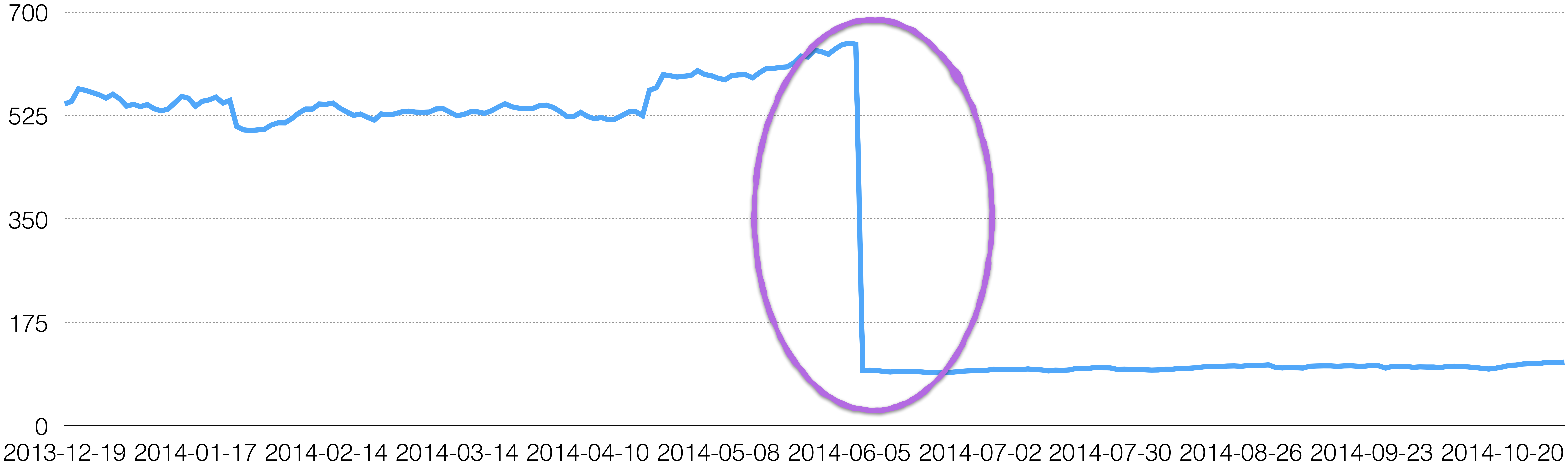
It looks as if the stock price dropped **by >80%** in June 2014

Stock Splits



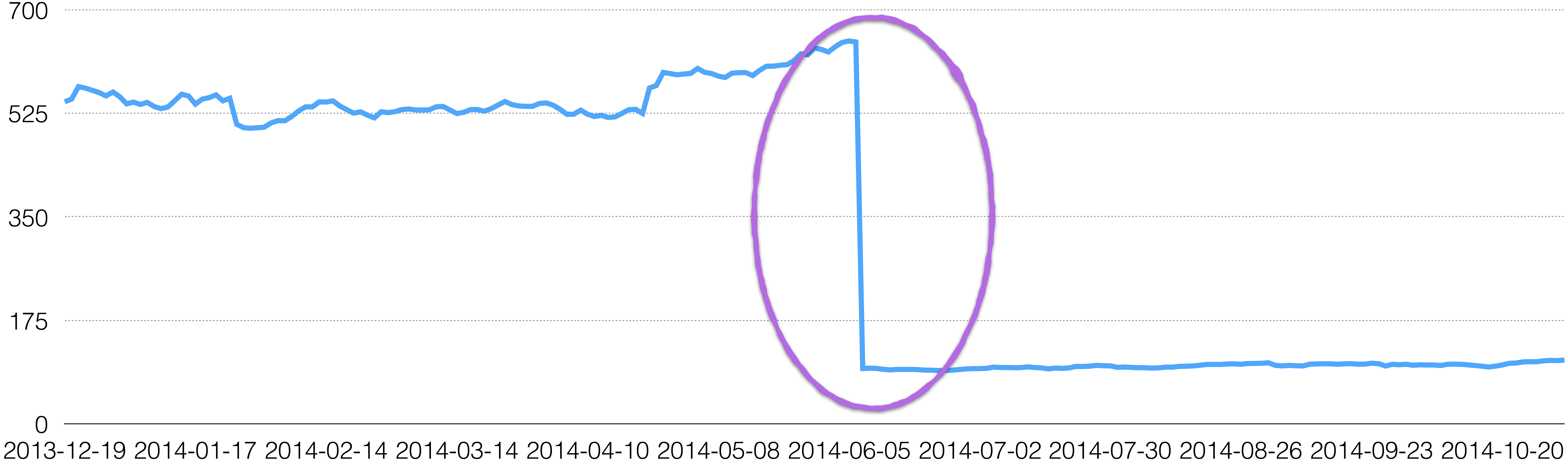
This drop had nothing to do with Apple's performance

Stock Splits



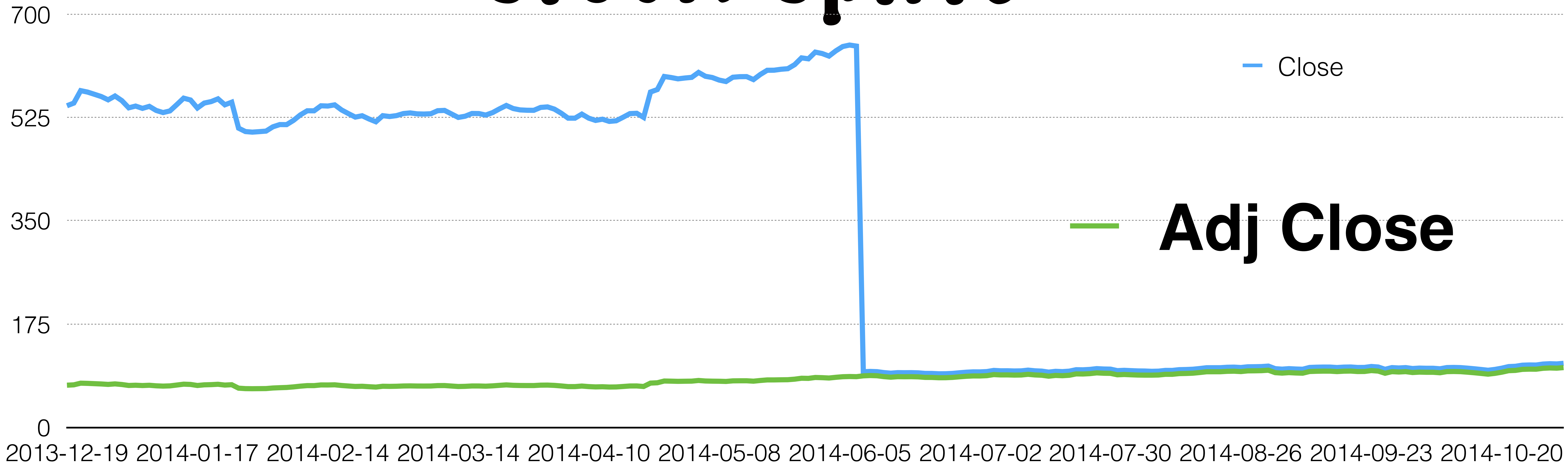
In fact, Apple products were selling well and continued to sell well

Stock Splits



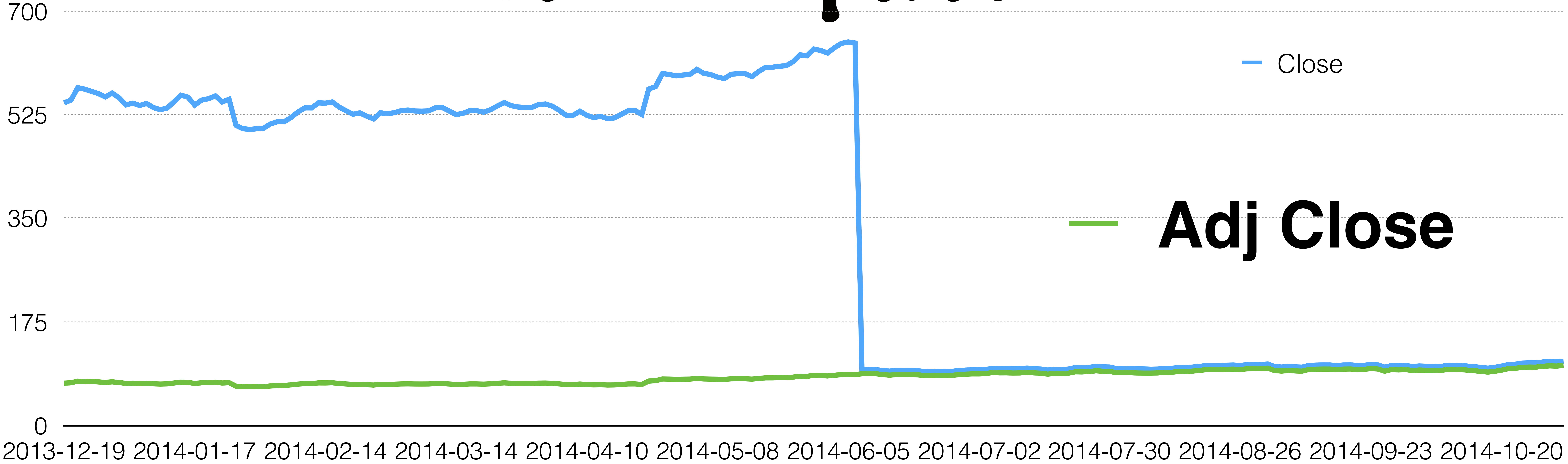
The drop was due to a 7-for-1 Stock split

Stock Splits



To adjust for the split, we need to divide all the prices before June 2014 by 7

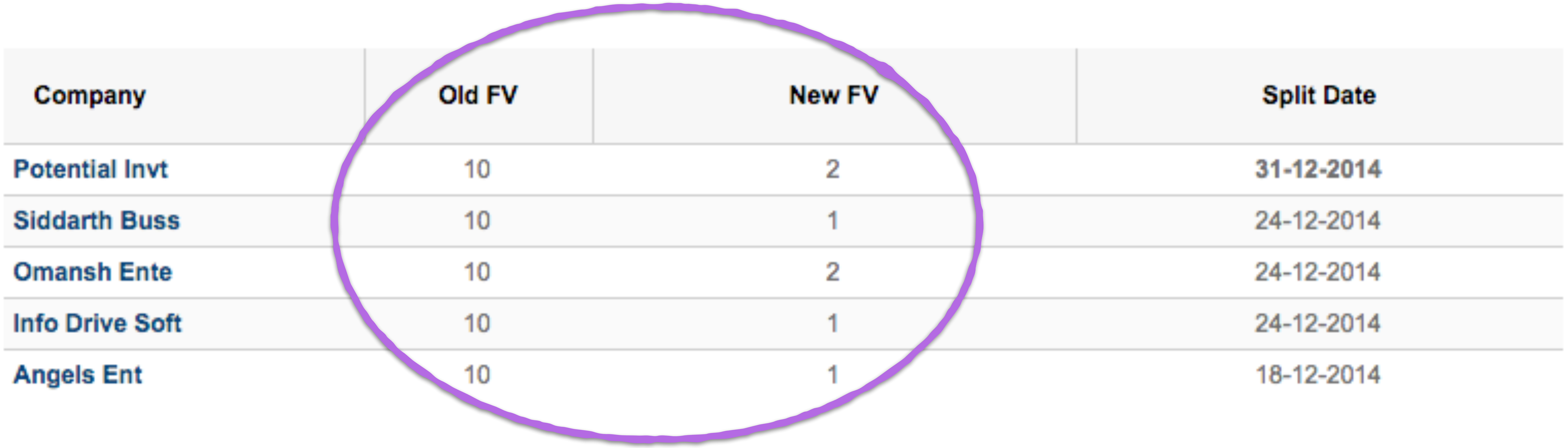
Stock Splits



The adjusted close price reflects the **true performance** of the stock

Stock Splits

Stock split announcement data is available online

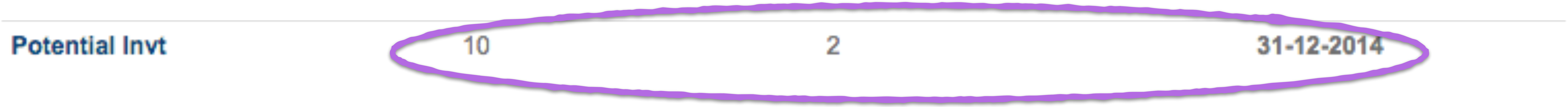


Company	Old FV	New FV	Split Date
Potential Invt	10	2	31-12-2014
Siddarth Buss	10	1	24-12-2014
Omansh Ente	10	2	24-12-2014
Info Drive Soft	10	1	24-12-2014
Angels Ent	10	1	18-12-2014

The prices need to be adjusted using the ratio of old face value to new face value

Stock Splits

Stock split announcement data is available online



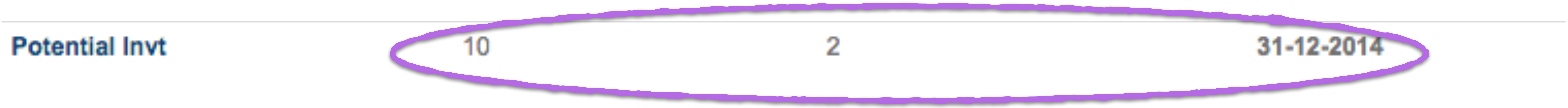
Potential Invt	10	2	31-12-2014

This means that there will be

A 5-for-1 stock split
on 31 Dec, 2014

Stock Splits

Stock split announcement data is available online



Potential Invt	10	2	31-12-2014

To adjust for the split

Divide all prices before 31 Dec, 2014 by 5

Clean up the data and put it in a ready-to-consume form

3 things we need to do

1. Remove any duplicates present in the data
2. Adjust for any symbol changes over time
3. Adjust for corporate actions

Once these 3 actions are done we are ready to build trading strategies using our data