Chapter 12:
Programming Is Not Enough

Python modules and other programming tools are extremely useful and powerful in data science. However, they have their limitations. To be a good data scientist, you also need to have a deep understanding of the institutional details in your own field. That is, programming alone is not enough, and we need the institutional knowledge in specific fields (finance, accounting, marketing, management, etc.) for us to make intelligent business decisions.

In this chapter, I'll use stock price changes around dividend payments and stock splits as an example to demonstrate why a deep understanding of the institutional details in your field is extremely important. Specifically, I will explain how cash dividends and stock splits affect stock prices and stock returns. What we should do and what we should avoid if we want to forecast stock prices.

The "end of chapter exercise" for this chapter is a project in your business field that is related to predictive analytics. You'll lay out what problem you need to solve, what data you plan to use. You'll also explain what institutional details you'll need to know in order to solve the problem. After that, you'll explain how you plan to clean up the data, what different models you'll use to predict. You'll also split the data into train and test and compare the accuracy of different models and decide what model you'll use and what's the final conclusion of your project.

A Motivating Example

In Chapter 7, when we try to predict stock prices, we mentioned that we should predict stock returns first and then use the predicted return to forecast price. Here, we'll provide a detailed explanation on why that's the case.

In the process, you'll learn the institutional details you need to know on how stock prices behave around cash dividends and stock splits. As a result, you'll understand what's the implications of this on stock price forecasts.

Stock prices are adjusted for cash dividends and stock splits, and we'll discuss the two cases below one at a time.

Cash Dividends

Cash dividends are payments of cash by the firm to its shareholders. There are two types of cash dividends based on frequency: regular cash dividends and special cash dividends.

Regular cash dividends are usually paid out every quarter (in some cases, every year). This is the most common form of dividend payments. Shareholders expect the payments to continue in the future. For example, on July 20, 2004, Microsoft announced that it, "…will move from its current annual dividend of $0.16 per share to a quarterly dividend of $0.08 per share…"

Special cash dividends are one-time, large-scale, cash payments to shareholders. The payments won't be repeated. For example, Microsoft announced a special dividend of $3 per share on July 20, 2004. Another example is Rowan Companies, Inc. (NYSE: RDC) announced on Aug 2, 2005 that its Board of Directors had declared a special cash dividend of $.25 per share.

How Cash Dividends Affect Stock Prices

The issuing company declare in advance the cash dividend amount to be paid and the date when it will be paid. The last date when shares can be purchased to receive the dividend is called the *with-dividend date*. The next business day is called the *ex-dividend date*. If you buy shares on or after the ex-dividend date, you'll not receive the cash dividend payment.

On the ex-dividend date, the stock price will drop roughly by the amount of the cash dividend payment. For example, if the stock price is $20/share on the with-dividend date and the cash dividend amount is $2/share, the stock price should be around $18 the next business day after the $2 cash dividend is paid out.

Let's use a real-world example to look at some data to have a better understanding.

Diamond Hill Investment Group (traded on NASDAQ with ticker symbol DHIL) announced on Oct 27, 2020, that it would pay a special dividend of $12/share. The with-dividend date is Nov 23, 2020, and the ex-dividend date is Nov 24, 2020. You can learn more about the news announcement here https://www.prnewswire.com/news-releases/diamond-hill-investment-group-inc-reports-results-for-third-quarter-2020-301161078.html

Let's get the stock price information for the company and look at the price changes around Nov 24, 2020.

The script below, *cash_dividend.py*, shows us how to get the stock prices of a company during a certain time window.

```
import datetime as dt  #1
import pandas as pd
import pandas_datareader.data as web

start = dt.datetime(2020, 11, 16)  #2
end = start + dt.timedelta(days=14)

df = web.DataReader("DHIL", 'yahoo', start, end)  #3
pd.options.display.float_format = "{:,.2f}".format   #4
print(df)
```

We first import needed modules 1. We set the starting date as Nov 16, 2020, a few days before the ex-dividend date 2. The end date is set to 14 days after the starting date so that we cover a period around the ex-dividend date.

We use the `DataReader()` 3 method to retrieve stock price information. The first argument of the method is the stock ticker symbol, which is DHIL in this case. The second argument is where to retrieve data, and we go to Yahoo Finance to retrieve out data. The last two argument are the start and end dates of the dataset that you want to retrieve. At 4, we set the numerical values in the dataset to two digits after decimal.

Finally, we print out the dataset to have a look. The output is as follows:

```
             High    Low   Open  Close    Volume  Adj Close
Date
2020-11-16 160.91 153.58 153.58 159.47 36,000.00     147.51
2020-11-17 162.00 157.09 159.99 159.24 24,800.00     147.30
2020-11-18 161.12 157.12 161.12 157.12 20,500.00     145.34
2020-11-19 158.70 154.00 157.50 155.62 16,900.00     143.95
2020-11-20 158.70 154.96 156.27 156.89 21,100.00     145.12
2020-11-23 161.01 158.31 158.31 160.00 34,500.00     148.00
2020-11-24 151.15 145.67 150.79 145.88 28,400.00     145.88
2020-11-25 147.66 144.38 147.42 145.00 17,700.00     145.00
2020-11-27 145.49 140.50 144.54 143.81  9,300.00     143.81
2020-11-30 143.20 135.57 143.20 136.82 32,600.00     136.82
```

Let's first look at the closing price, the column with the heading *Close*. On the with-dividend date Nov 23, 2020, the closing price is $160/share. On the ex-divided date Nov 24, 2020, the

closing price is $145.88/share. The total price change is $14.12/share, and $12 is due to the fact that a $12 cash dividend is paid out that day. The remaining $2.12/share is the normal daily price fluctuation.

WARNING: The adjusted closing prices are likely have changed from the above numbers since Apple is paying quarterly dividends. As a result, the historical values of the adjusted closing prices changes from quarter to quarter, retroactively. And that's the whole point of what we are discussing here!!!

If you use the closing price to calculate stock return, you'll get a daily return of

```
(145.88 - 160)/160 = -8.825%
```

However, if you are a shareholder, even though the stock price changed from $160 to $145.88, your total return is more than -8.825%. This is because you received a $12/share cash dividend, so you real total return on Nov 24 should be

```
(145.88 - 160 + 12)/160 = -1.325%
```

This creates a problem: to calculate daily returns, you need to keep track of all cash dividend payments everyday and adjust for it when calculating returns.

Therefore, in most datasets, there is an added column called adjusted closing price, *Adj Close*. Once a cash dividend is paid, the dataset provider will deduct the dividend amount from the stock's historical prices. If you look at the adjusted closing price of DHIL on Nov 23, 2020, it's $148, even though on Nov 23, 2020, the closing price of the stock was $160. The $12 difference is due to the adjustment for the cash dividend payment to make the calculation of returns easy.

Takeaway: to calculate a stock's daily returns, we use the adjusted closing price instead of the closing price**.**

What Are Stock Splits (Stock Dividends)

Stock dividends are shares of additional stocks paid out as dividends. That is, instead of receiving a cash payment, shareholders receive additional shares of the company. For example, a 20% stock dividend means that a shareholder receives one new share for every five shares owned.

Stock splits are similar to a stock dividend, except at a much larger scale. The most common one is a 2-for-1 stock split, i.e., each old share is split into two new shares. The second most common one is a 3-for-2 stock split, i.e., every two original shares are split into three new shares.

Since stock dividends and stock splits are essentially the same, but at different scales, we'll use stock splits only as examples from now onwards. But you should understand that whatever we discuss below apply to stock dividends as well.

In principle, stock split should not affect the market value of the firm's common equity. Suppose before a stock split, stock price is $60/share, after a 2-for-1 split, each share should be $30. In reality the price may deviate from $30 slightly because lower prices may attract more small investors and increase liquidity of the stock, thus pushing the price higher. For our purpose, we can ignore the small deviation.

You probably read about stock splits quite often in the news. For example, Apple's stock has split five times since the company went public. The most recent one is a 4-for-1 split on August 28, 2020. You can read about it here https://www.marketwatch.com/story/3-things-to-know-about-apples-stock-split-2020-08-28. Before that, Apple conducted a 7-for-1 stock split on June 9, 2014, a 2-for-1 stock split on February 28, 2005, June 21, 2000, and June 16, 1987.

Tesla conducted a five-for-one stock split after the close of trading on Aug 28, 2020 (yes, the same day as Apple's stock split). See news announcement here https://ir.tesla.com/press-release/tesla-announces-five-one-stock-split.

How Stock Splits (Stock Dividends) Affect Stock Price

Once a stock splits its stock, all its historical prices will be changed. They are literally "changing the history." Here is how.

Let's use the August 28, 2020, Apple stock split as an example.

I have downloaded the Apple historical prices on July 25, 2019, for another purpose. Note that this is before the August 28, 2020, 4-for-1 stock split. The dataset, *AAPL_July25_2019.csv*, can be downloaded on Canvas. If you scroll down to the bottom, you'll see the price information as follows:

Date	High	Low	Open	Close	Volume	Adj Close
7/11/2019	204.39	201.71	203.31	201.75	20191800	201.75
7/12/2019	204	202.2	202.45	203.3	17595200	203.3
7/15/2019	205.87	204	204.09	205.21	16947400	205.21
7/16/2019	206.11	203.5	204.59	204.5	16866800	204.5
7/17/2019	205.09	203.27	204.05	203.35	14107500	203.35
7/18/2019	205.88	203.7	204	205.66	18582200	205.66
7/19/2019	206.5	202.36	205.79	202.59	20929300	202.59
7/22/2019	207.23	203.61	203.65	207.22	22277900	207.22
7/23/2019	208.91	207.29	208.46	208.84	17961600	208.84
7/24/2019	209.15	207.57	207.67	207.96	6828866	207.96

On July 24, 2019, the dataset shows that the high, low, open, close, and adjusted close prices are 209.15, 207.57, 207.67, 207.96, and 207.96, respectively.

However, if you retrieve the Apple stock price now using the following script *stock_split.py*:

```
import datetime as dt
import pandas as pd
import pandas_datareader.data as web

start = dt.datetime(2019, 7, 11)
end = dt.datetime(2019, 7, 24)
df = web.DataReader("AAPL", 'yahoo', start, end)
pd.options.display.float_format = "{:,.2f}".format
print(df)
```

When I ran the script on Jan 23, 2021, I got this result:

```
              High    Low   Open   Close     Volume  Adj Close
Date
2019-07-11 51.10 50.43 50.83  50.44  80767200      49.67
2019-07-12 51.00 50.55 50.61  50.83  70380800      50.05
2019-07-15 51.47 51.00 51.02  51.30  67789600      50.52
2019-07-16 51.53 50.88 51.15  51.12  67467200      50.34
2019-07-17 51.27 50.82 51.01  50.84  56430000      50.06
2019-07-18 51.47 50.92 51.00  51.42  74162400      50.63
2019-07-19 51.62 50.59 51.45  50.65  83717200      49.87
2019-07-22 51.81 50.90 50.91  51.81  89111600      51.01
2019-07-23 52.23 51.82 52.12  52.21  73420800      51.41
2019-07-24 52.29 51.79 51.92  52.17  59966400      51.37
```

Note that when you run it, the result may look slightly different if there is a cash dividend payment after Jan 23, 2021.

According to results from the script *stock_split.py*, on July 24, 2019, the high, low, open, close, and adjusted close prices are 52.29, 51.79, 51.92, 52.17, and 51.37, respectively.

How can the Apple stock have two different sets of prices on the same day?

The reason lies in the fact that there is a 4-for-1 stock split on Aug 28, 2020. After the split, Yahoo Finance and other finance data providers adjusted all prices of Apple on or before Aug 28, 2020 (dividing everything by 4 due to the 4-for-1 split). They are changing all the historical data for the stock (hence changing the history).

Implication For Predictive Analytics

For reasons we discussed above, when you want to forecast the stock price of a company, you need to forecast the returns first. This is because stock prices are not *statistically stationary*.

NOTE: A time-series variable x *is* statistically stationary *if the mean, variance, autocorrelation of the variable* x *is constant over time.*

As we have seen from examples above, the mean of the stock price changes dramatically around cash dividends and stock splits. In the example of the $12 special cash dividend for DHIL, the mean stock price was $160 before the dividend payment and $148 after the dividend payment. In the case of Apple's 4-for-1 stock split on August 28, 2020, the mean stock price was around $499.23 before the split (the actual closing price that day, again, see this article https://www.marketwatch.com/story/3-things-to-know-about-apples-stock-split-2020-08-28 for details). After the split, the mean stock price is $124.81.

On the other hand, the daily stock returns based on adjusted closing price is statistically stationary, because its means is always a number close to 0. Its variance and autocorrelation are also close to constant. Therefore, we can use our statistical tools to forecast the daily returns.

Once we have a prediction for next day's stock return, and we can forecast next day's stock price by using this formula:

```
Next day's price = today's price × (1 + predicted next day return)
```

Note here we assume that there is no cash dividend or stock splits today or the next day. But if there is one, you can predict how the price will behave and adjust your prediction accordingly.

Explore A Project in Your Field (Course Project)

Use what you have learned in this class, work on a project in your business field that relates to predictive analytics. Examples include forecasting the revenue of a business proposal, the stock price of a company, cost related to a project, probability of an account being audited, a firm being accused of fraud, and so on.

You are required to lay out what problem you need to solve, what data you plan to use. You'll also explain what institutional details you'll need to know in order to solve the problem. After that, you'll explain how you will clean up the data, what different models you'll use to predict. You'll also split the data into train and test and compare the accuracy of different models and decide what model you'll use based on that. Finally, state your conclusion (i.e., prediction).

Items you need to submit on Canvas, all in one large zip file:

1. An 8-15 page writeup. The writeup should specify: a) what variable you'll predict; b) how and where to collect data; c) institutional details you'll need to know in order to resolve the problem; d) steps to clean up the data; e) what different models you'll use to predict; f) how to split the data into train and test; g) how to compare the accuracy of different models; f) which model you'll use and why; h) your result.
2. Python scripts for all the steps involved;
3. The raw data you collected;
4. The cleaned-up data.