Distribution of Returns

Contents

1.0 Introduction - Distributions

In this guide we are going to use excel to examine past distributions of asset returns. Past price behaviour may not be a perfect guide to the future; however it is our only guide. Therefore it is by definition, useful to understand how price movements and returns have behaved in the past. With a sufficiently large dataset, it is probable that over a long period of time the asset behaves similarly to the way it did historically.

As traders/investors, this basic analysis can tell us whether the odds of exposing ourselves to an asset over a given time period provides us with a suitable risk-reward profile for us to make money. Furthermore, calculating a distribution of returns also allows us to investigate other important parameters such as expected return. Understanding price movements can give an insight into volatility of an asset which can contribute towards setting stop losses, targets and other risk management techniques. In short, the analysis can help us understand the probabilities of making money in an asset over a period of time, while displaying data that can be intuitively applied to establish effective risk management.

> *NB: In this guide we do not attempt to provide you with a rigorous lesson in theoretical statistics. For the purposes of trading and investing it is important to understand the basics of distributions, how they apply to publicly traded assets and how they can be used to approximate expected returns over given time periods. Most trading educators and brokers either do not know these techniques or choose not to advise them for risk of it conflicting with their own interest. If you would like to read further into the statistical side of distributions and how to calculate them, there are countless websites that can help you achieve this. You can also take advanced statistics classes in your local area to become a real expert!*

2.0 The Normal Distribution Example

Normal or "Gaussian" Distributions are most commonly found in nature and human behaviour. To give some practical examples, a Normal Distribution can explain the probability of being a certain height or IQ level. It is a bell shaped curve, symmetrical around the mean (arithmetic average) of the sample being analysed. Larger datasets will result in more accurate probability distributions and so in cases where a normal distribution applies, more data will show empirical analysis tending towards normality.

The Normal Distribution is the most commonly used probability distribution within the investment industry, largely because of the simplicity in its calculation. However, traders/investors should be aware that often it is not a perfect representation of asset returns – this will be tested and discussed further in the analysis later in the tutorial.

2.1 Human Height

As mentioned, human height is a classic example of a naturally normally distributed data. Remember, the larger the sample size, the higher the tendency towards normality. This distribution was generated from a sample of 500 people.

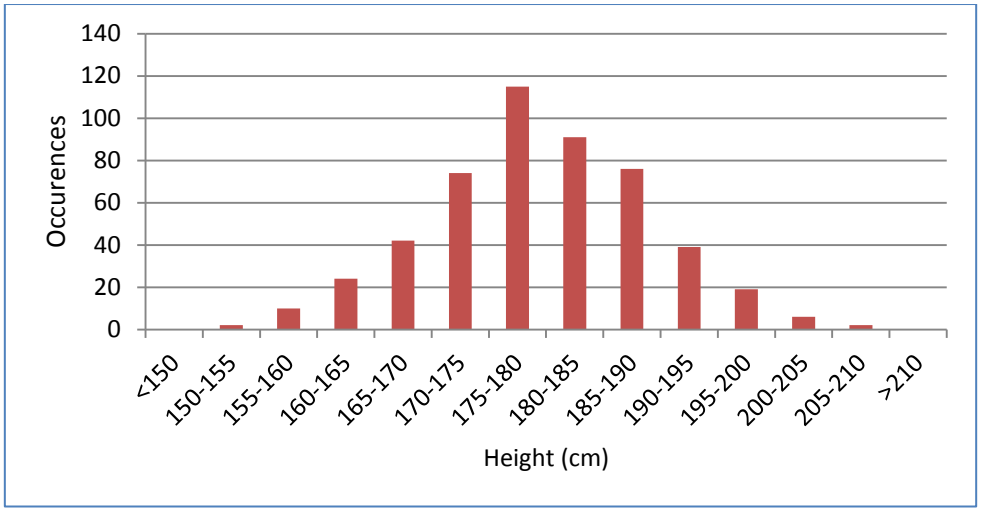


Figure 1 - Height Distribution Histogram

The mean of the data set used was measured as 180cm, with a standard deviation of 10cm. A standardised normal distribution can be plotted over the top of this histogram to assess its "goodness of fit."

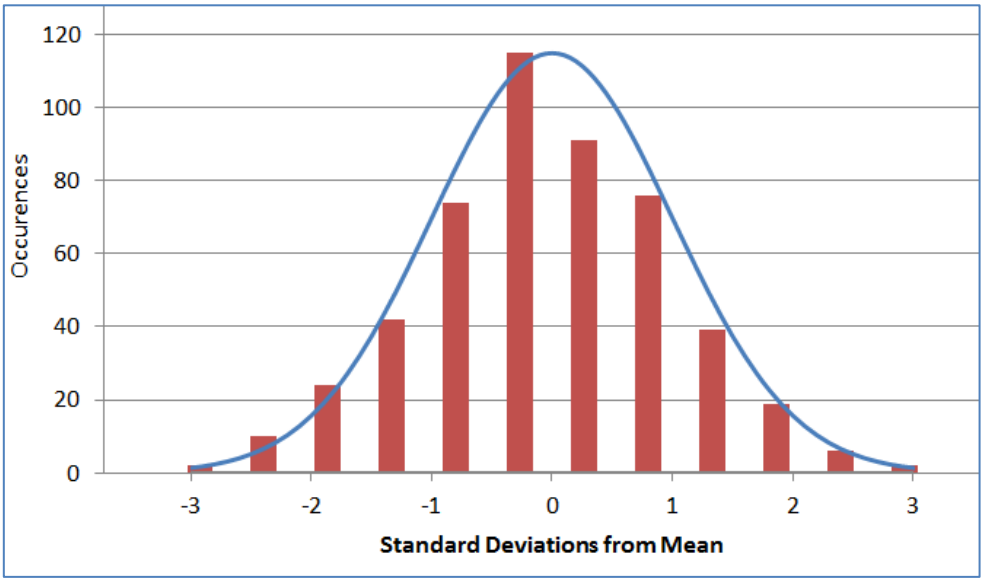


Figure 2 - Normal Distribution fitted over Empirical Height Distribution Histogram

We can see that a normal distribution fits the data fairly well, and therefore it can be argued that this type of distribution is a good predictor of height variation within the general population (not just the sample used). Note that the x-axis in the second graph has been "standardised" and displayed in terms of standard deviations. "0" standard deviations from the mean represents the mean height value itself, in this case 180cm, whereas plus and minus 1 standard deviation represent the mean plus (or minus) 10cm which is equal to 1 standard deviation.

In a perfect normal distribution, standard deviations from the mean account for the following frequency of occurences:

Standard deviations from the mean	Percentage of occurrences in data set
1	68.2%
2	95.4%
3	99.8%

E.g. 95.4% of values within a normally distributed data set are predicted to lie between +2 and -2 standard deviations from the mean. In the height example mentioned above, this would predict that 95.4% of the data

will lie between 160cm-200cm. So, by definition, 4.6% (100-95.4) of people are either smaller than 160cm or taller than 200cm. That equates to about 1 in 20 people having heights greater than 2 standard deviations away from the mean.

2.2 Normal Distribution Conclusion

It turns out that there are a lot of examples of data which can be assumed as normally distributed, with human height being one of the most obvious. As previously mentioned, asset returns are often estimated using the normal distribution, however due to the financial implications of these estimates it is important to draw your own conclusions from empirical data – which is why the rest of this tutorial is devoted to teaching how to plot and analyse your own empirical dataset of an asset.

3.0 Do It Yourself (DIY) Asset Analysis

After introducing probability distributions through the normally distributed human height example, a step-by-step method of how to conduct your own analysis on asset returns using empirical data will be explained. You will need **internet access** to obtain the data and **Microsoft Excel** to conduct the analysis.

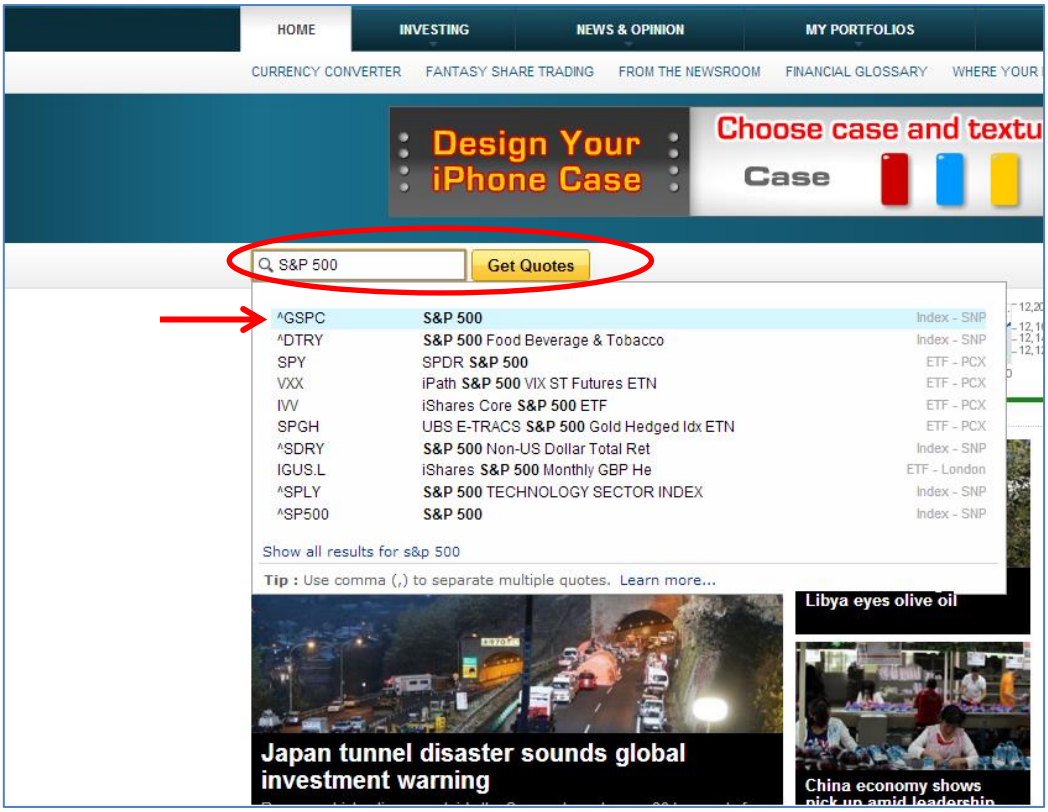
In this example we are going to look at some simple analysis of day trading the S&P500 index. It is commonly thought that around 95% of "day traders" (those that trade intra-day or over one day horizons) lose money. If this is the case, it is worthwhile looking at the actual numbers to find out why this is happening.

In this example, step-by-step instructions will be given on how to analyse the daily returns of the S&P500.

3.1 Finding the Data

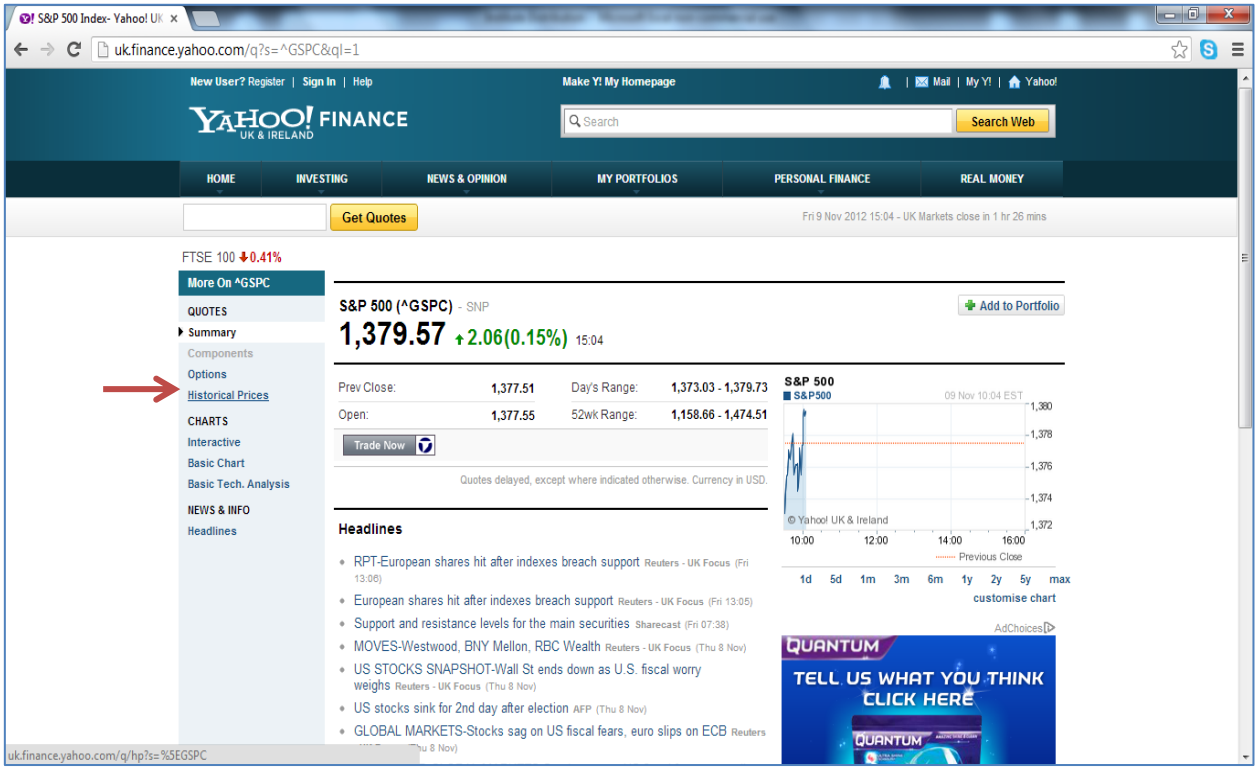
There are many sources of asset data on the internet. Yahoo Finance is highly recommended and will be used in the following example. It has comprehensive historical data and is simple to use.

a) *Go to http://finance.yahoo.com/*
b) *Enter "S&P 500" into the "Get Quote" box, and click on the appropriate drop-down item.*



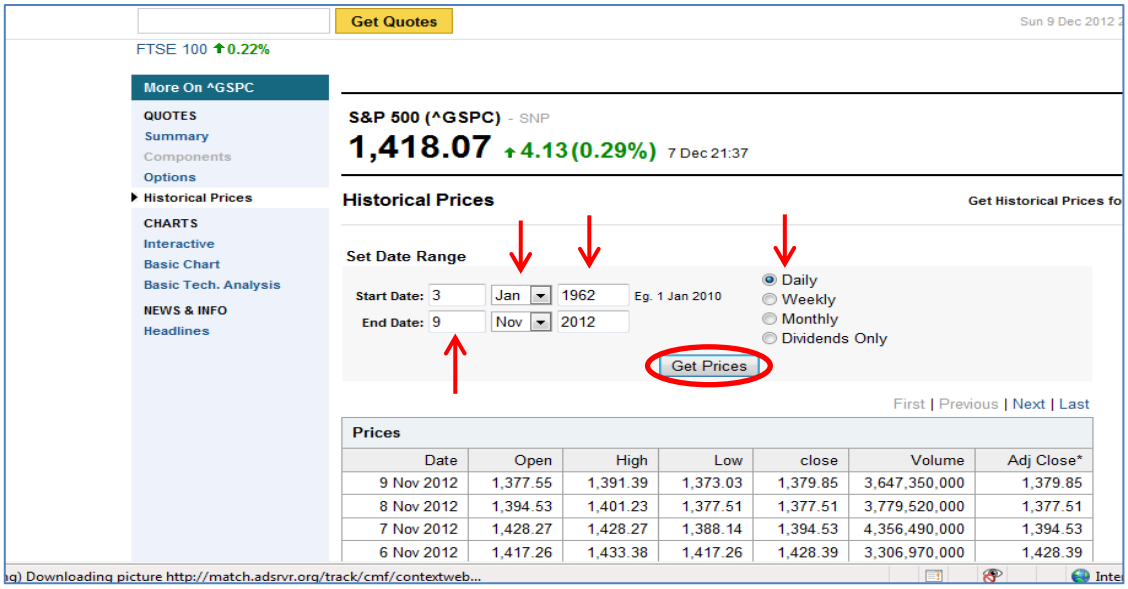
c) Once you have selected the S&P500, you will be presented with the "Summary Page."

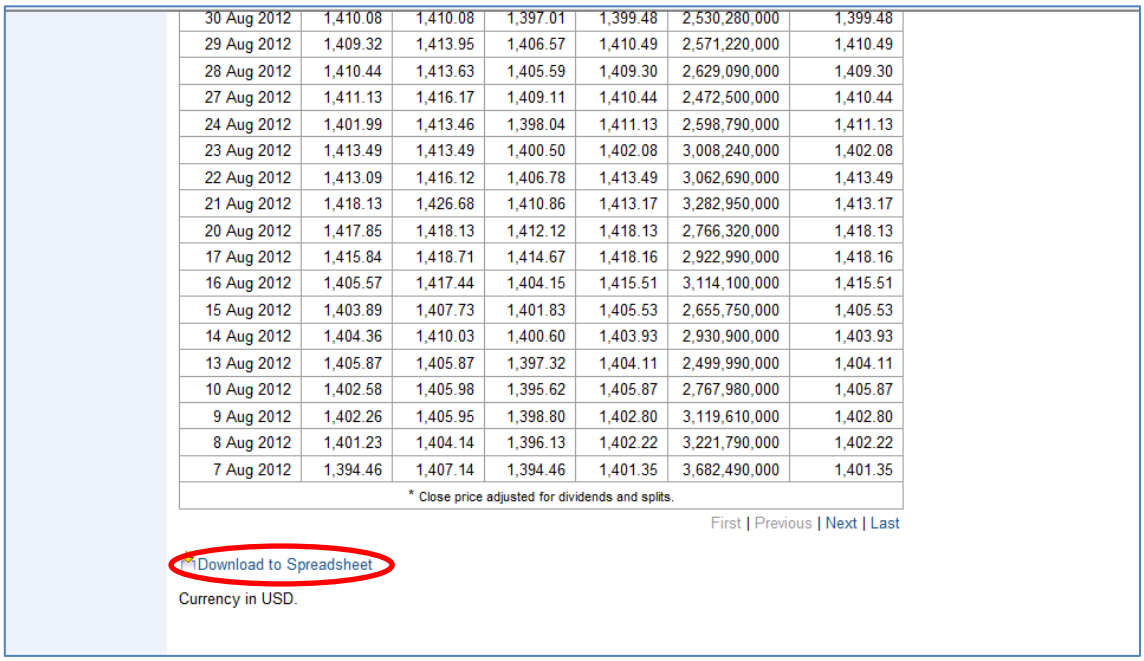
   a. *Now click on "Historical Prices" from the menu on the left-hand side.*



d) The "Historical Prices" page lets us view historical data for the S&P 500 (or a given asset), with adjustable time periods and price intervals available.

   a. *In this example we will use daily data from 3/1/1962 to 9/11/2012. Adjust the parameters as appropriate and click "Get Prices". Scroll down to the bottom of the webpage and click "Download to Spreadsheet".*





30 Aug 2012	1,410.08	1,410.08	1,397.01	1,399.48	2,530,280,000	1,399.48
29 Aug 2012	1,409.32	1,413.95	1,406.57	1,410.49	2,571,220,000	1,410.49
28 Aug 2012	1,410.44	1,413.63	1,405.59	1,409.30	2,629,090,000	1,409.30
27 Aug 2012	1,411.13	1,416.17	1,409.11	1,410.44	2,472,500,000	1,410.44
24 Aug 2012	1,401.99	1,413.46	1,398.04	1,411.13	2,598,790,000	1,411.13
23 Aug 2012	1,413.49	1,413.49	1,400.50	1,402.08	3,008,240,000	1,402.08
22 Aug 2012	1,413.09	1,416.12	1,406.78	1,413.49	3,062,690,000	1,413.49
21 Aug 2012	1,418.13	1,426.68	1,410.86	1,413.17	3,282,950,000	1,413.17
20 Aug 2012	1,417.85	1,418.13	1,412.12	1,418.13	2,766,320,000	1,418.13
17 Aug 2012	1,415.84	1,418.71	1,414.67	1,418.16	2,922,990,000	1,418.16
16 Aug 2012	1,405.57	1,417.44	1,404.15	1,415.51	3,114,100,000	1,415.51
15 Aug 2012	1,403.89	1,407.73	1,401.83	1,405.53	2,655,750,000	1,405.53
14 Aug 2012	1,404.36	1,410.03	1,400.60	1,403.93	2,930,900,000	1,403.93
13 Aug 2012	1,405.87	1,405.87	1,397.32	1,404.11	2,499,990,000	1,404.11
10 Aug 2012	1,402.58	1,405.98	1,395.62	1,405.87	2,767,980,000	1,405.87
9 Aug 2012	1,402.26	1,405.95	1,398.80	1,402.80	3,119,610,000	1,402.80
8 Aug 2012	1,401.23	1,404.14	1,396.13	1,402.22	3,221,790,000	1,402.22
7 Aug 2012	1,394.46	1,407.14	1,394.46	1,401.35	3,682,490,000	1,401.35

* Close price adjusted for dividends and splits.

First | Previous | Next | Last

Download to Spreadsheet

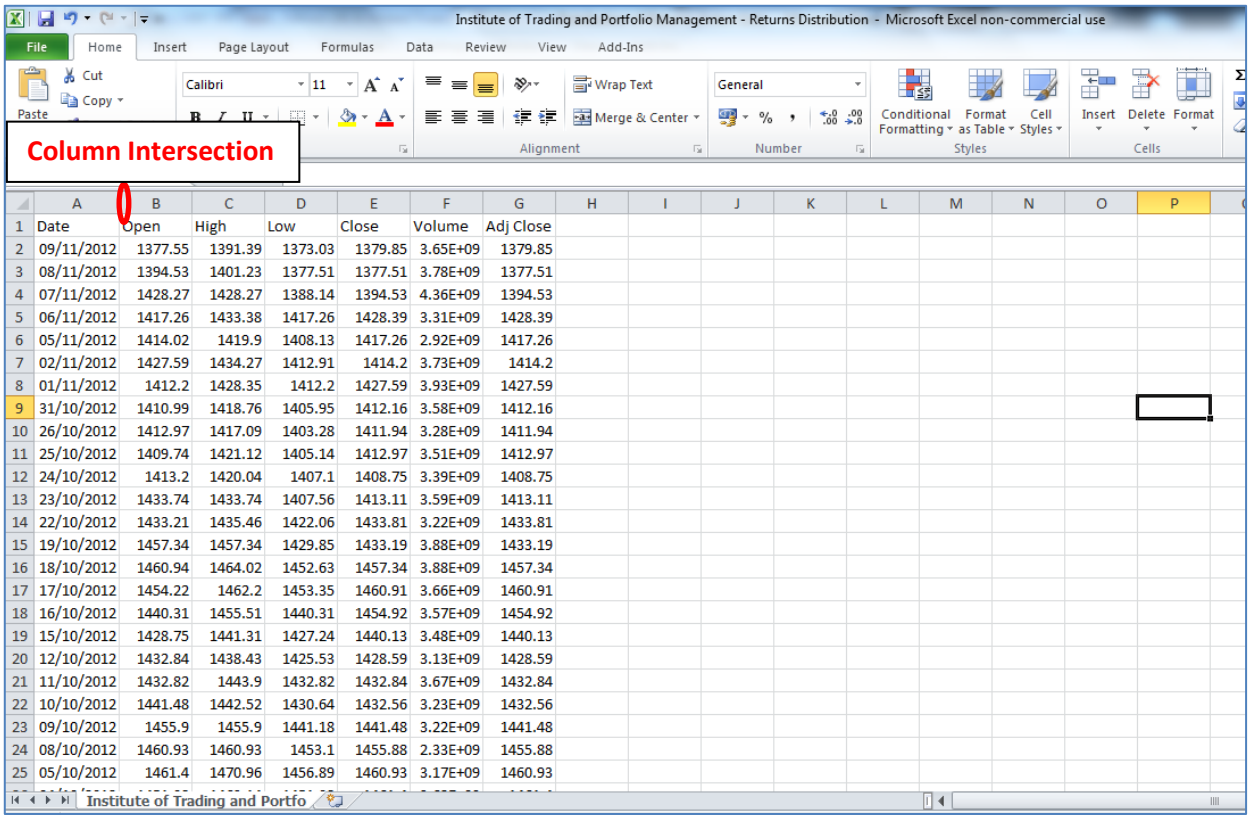
Currency in USD.

*Clicking “open spreadsheet”, will open the data in Excel straight away. You then need to save the excel sheet to a folder on your hard drive to permanently store it. Alternatively you can save the file straight to a location on your hard drive, and navigate to the file yourself and open it. Be sure to save the file after any work or edits carried out.

3.2 Analysing the Data in Excel

A few Excel notes before we start:

- *Columns are listed as A…Z etc running across the top of the spreadsheet.*
- *Rows are numbered down the left-hand side.*
- *A 'cell' is coded by the column-row reference grid – E.g. Cell P9 – as selected below.*
- *If data presents itself as "'####", do not worry, this just means the data text or number has too many characters to fit in the cell/column. Adjust the column width where this problem occurs. This can happen quite often in the "Date" column when importing from Yahoo Finance – adjust by clicking and dragging the intersection of column A and B to the right.*
- *To zoom in and out of the spreadsheet, hold control and use the mouse wheel, or equivalent.*



	A	B	C	D	E	F	G
1	Date	Open	High	Low	Close	Volume	Adj Close
2	09/11/2012	1377.55	1391.39	1373.03	1379.85	3.65E+09	1379.85
3	08/11/2012	1394.53	1401.23	1377.51	1377.51	3.78E+09	1377.51
4	07/11/2012	1428.27	1428.27	1388.14	1394.53	4.36E+09	1394.53
5	06/11/2012	1417.26	1433.38	1417.26	1428.39	3.31E+09	1428.39
6	05/11/2012	1414.02	1419.9	1408.13	1417.26	2.92E+09	1417.26
7	02/11/2012	1427.59	1434.27	1412.91	1414.2	3.73E+09	1414.2
8	01/11/2012	1412.2	1428.35	1412.2	1427.59	3.93E+09	1427.59
9	31/10/2012	1410.99	1418.76	1405.95	1412.16	3.58E+09	1412.16
10	26/10/2012	1412.97	1417.09	1403.28	1411.94	3.28E+09	1411.94
11	25/10/2012	1409.74	1421.12	1405.14	1412.97	3.51E+09	1412.97
12	24/10/2012	1413.2	1420.04	1407.1	1408.75	3.39E+09	1408.75
13	23/10/2012	1433.74	1433.74	1407.56	1413.11	3.59E+09	1413.11
14	22/10/2012	1433.21	1435.46	1422.06	1433.81	3.22E+09	1433.81
15	19/10/2012	1457.34	1457.34	1429.85	1433.19	3.88E+09	1433.19
16	18/10/2012	1460.94	1464.02	1452.63	1457.34	3.88E+09	1457.34
17	17/10/2012	1454.22	1462.2	1453.35	1460.91	3.66E+09	1460.91
18	16/10/2012	1440.31	1455.51	1440.31	1454.92	3.57E+09	1454.92
19	15/10/2012	1428.75	1441.31	1427.24	1440.13	3.48E+09	1440.13
20	12/10/2012	1432.84	1438.43	1425.53	1428.59	3.13E+09	1428.59
21	11/10/2012	1432.82	1443.9	1432.82	1432.84	3.67E+09	1432.84
22	10/10/2012	1441.48	1442.52	1430.64	1432.56	3.23E+09	1432.56
23	09/10/2012	1455.9	1455.9	1441.18	1441.48	3.22E+09	1441.48
24	08/10/2012	1460.93	1460.93	1453.1	1455.88	2.33E+09	1455.88
25	05/10/2012	1461.4	1470.96	1456.89	1460.93	3.17E+09	1460.93

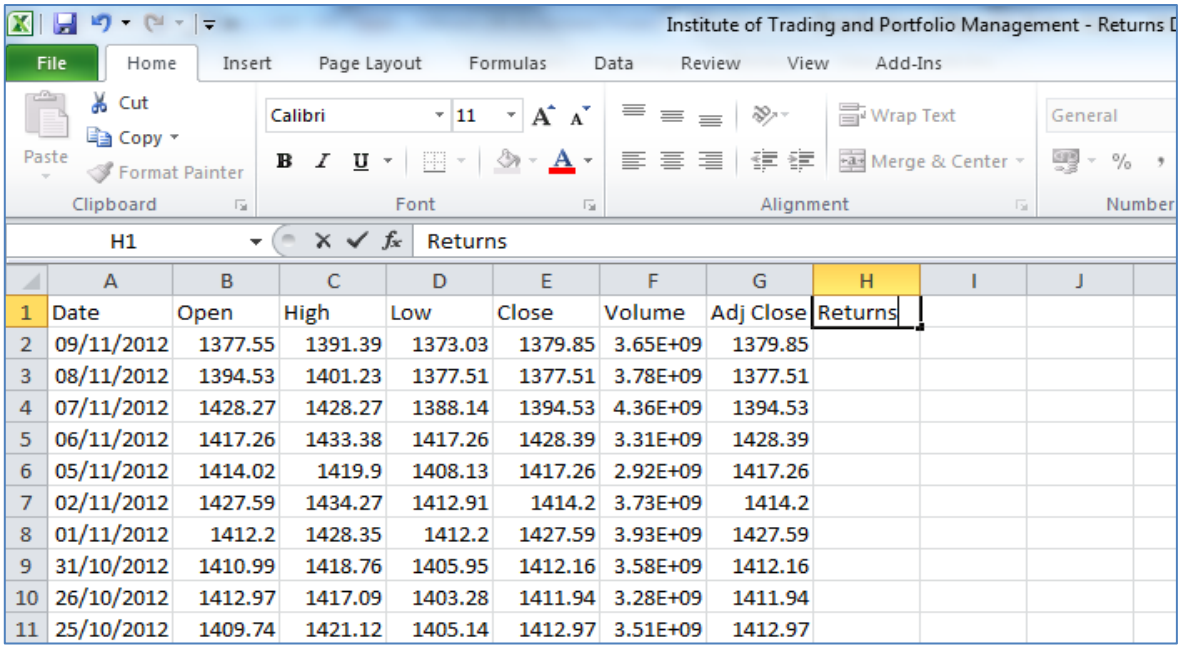
3.2.1 Initial Set-up

After opening up the spreadsheet you should be faced with seven columns of *Date, Open, High, Low, Close, Volume* and *Adjusted Close.* The number of rows will depend on the time period and data frequency as chosen earlier. *In this example there are 12,804 rows.*

With the data downloaded, saved, and open in Excel, we can now start setting up the data ready to analyse - Take the following steps to set up your data:

(i) Heading the Returns Column

*Select cell **H1** and type "Returns" to create a heading for the column.*



	A	B	C	D	E	F	G	H	I	J
1	Date	Open	High	Low	Close	Volume	Adj Close	Returns		
2	09/11/2012	1377.55	1391.39	1373.03	1379.85	3.65E+09	1379.85			
3	08/11/2012	1394.53	1401.23	1377.51	1377.51	3.78E+09	1377.51			
4	07/11/2012	1428.27	1428.27	1388.14	1394.53	4.36E+09	1394.53			
5	06/11/2012	1417.26	1433.38	1417.26	1428.39	3.31E+09	1428.39			
6	05/11/2012	1414.02	1419.9	1408.13	1417.26	2.92E+09	1417.26			
7	02/11/2012	1427.59	1434.27	1412.91	1414.2	3.73E+09	1414.2			
8	01/11/2012	1412.2	1428.35	1412.2	1427.59	3.93E+09	1427.59			
9	31/10/2012	1410.99	1418.76	1405.95	1412.16	3.58E+09	1412.16			
10	26/10/2012	1412.97	1417.09	1403.28	1411.94	3.28E+09	1411.94			
11	25/10/2012	1409.74	1421.12	1405.14	1412.97	3.51E+09	1412.97			

(ii) Daily Returns Formula

We are now going to calculate the daily returns. This will show us the percentage return if we bought on the open and sold on the close each day over the whole period of our dataset.

Mathematically the formula for returns over a single period can be presented like this:

$$(R_N) = \frac{CP_N - OP_N}{OP_N}$$

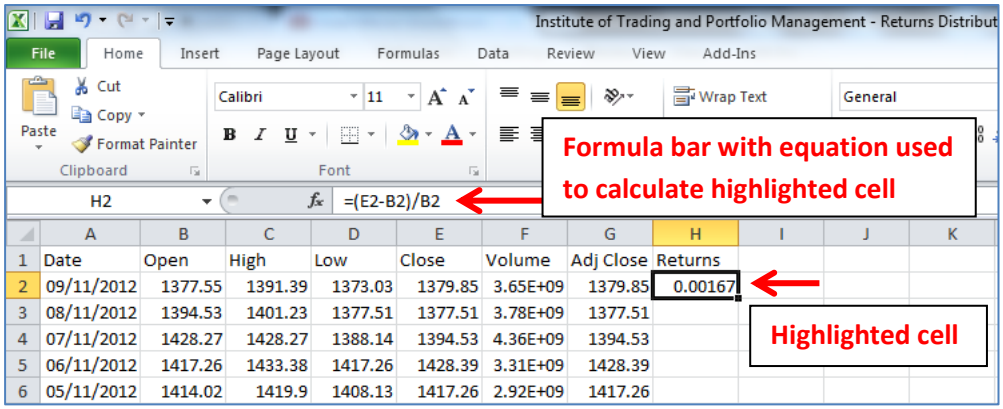
Where: R_N = Return in period N
CP_N = Closing Price of period N
OP_N = Opening Price of period N

To apply this formula in column H, we are going to enter it into *cell H2* and copy it down the whole column to apply it to every row in our dataset.

(iii) Entering the Returns Formula

In cell H2, type *=(E2-B2)/B2* *and press Enter*

This follows the formula outlined previously. E2 is our Closing Price for that day, B2 is our Opening Price. Make sure the brackets are included to avoid any mistakes. While entering the formula into cell H2, you may also manually click the cells E2 and B2 instead of typing the references in, however you must make sure the equation's form remains unchanged.



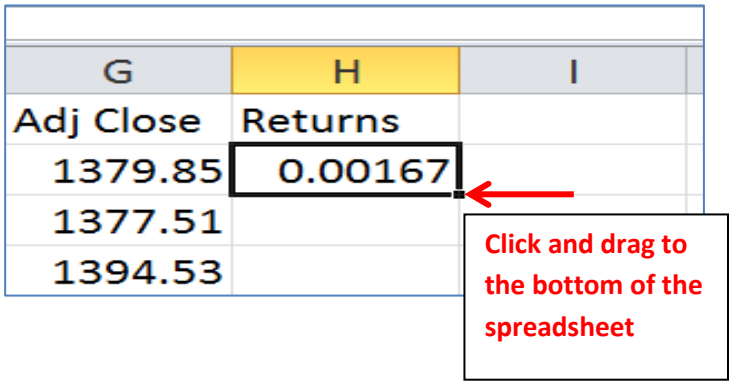
The cell H2 now shows the return for buying at the Open and selling the Close on the 9/11/2012. When a cell is highlighted, the formula entered to calculate it appears in in the formula bar as shown above.

You may note that we used the Close Price and not the Adjusted Close Price. The Adjusted Close Price may be used instead (it accounts for dividends and stock splits) depending on your preference.

(iv) Copying Down the Formula

There are two methods to do this:

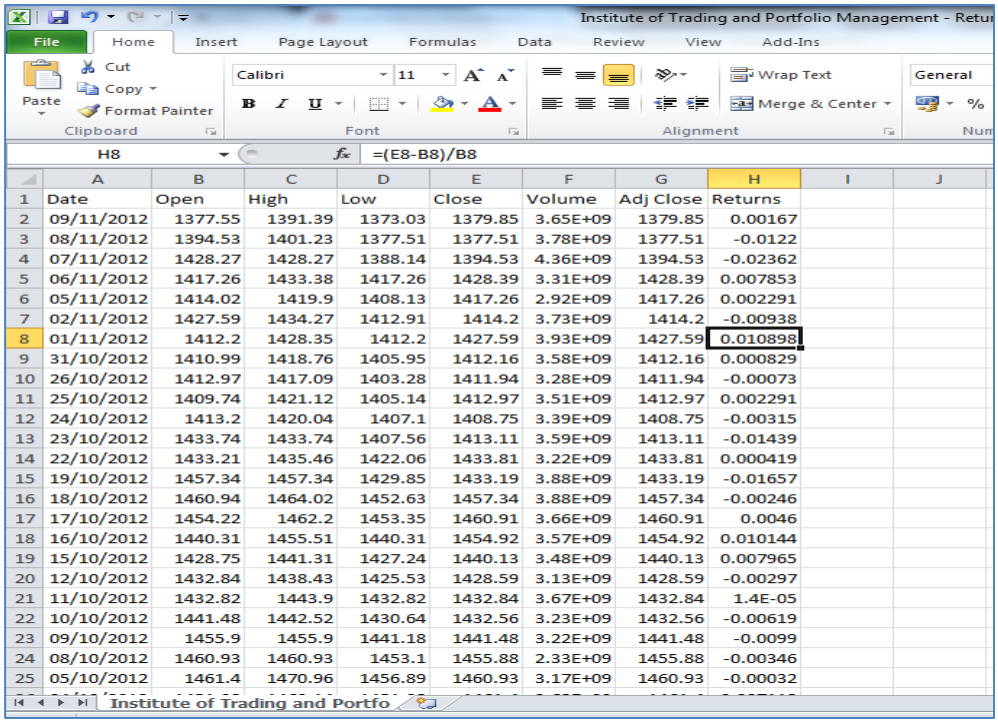
- EITHER
  - *Highlight the cell H2*
  - *Click and drag the small black dot on the bottom right-hand corner of the cell all the way down the spreadsheet to 3/1/1962*.



- OR
  - *Navigate to the bottom of the spreadsheet using the scrollbar*
  - *Select cell H12804 (the row with our earliest date, the last row in our spreadsheet)*
  - *Press CTRL+SHIFT+UP* - this will select all the cells H2 to H12804.
  - *Press CTRL+D* - copies down the formula in the top cell selected.

For very large sets of data like this one, the second method is much faster although may lead to more mistakes if you are not confident using excel.

REMEMBER: It makes sense to check the formula has copied down correctly by highlighting a cell in the column and checking the formula bar.



	A	B	C	D	E	F	G	H	I	J
1	Date	Open	High	Low	Close	Volume	Adj Close	Returns		
2	09/11/2012	1377.55	1391.39	1373.03	1379.85	3.65E+09	1379.85	0.00167		
3	08/11/2012	1394.53	1401.23	1377.51	1377.51	3.78E+09	1377.51	-0.0122		
4	07/11/2012	1428.27	1428.27	1388.14	1394.53	4.36E+09	1394.53	-0.02362		
5	06/11/2012	1417.26	1433.38	1417.26	1428.39	3.31E+09	1428.39	0.007853		
6	05/11/2012	1414.02	1419.9	1408.13	1417.26	2.92E+09	1417.26	0.002291		
7	02/11/2012	1427.59	1434.27	1412.91	1414.2	3.73E+09	1414.2	-0.00938		
8	01/11/2012	1412.2	1428.35	1412.2	1427.59	3.93E+09	1427.59	0.010898		
9	31/10/2012	1410.99	1418.76	1405.95	1412.16	3.58E+09	1412.16	0.000829		
10	26/10/2012	1412.97	1417.09	1403.28	1411.94	3.28E+09	1411.94	-0.00073		
11	25/10/2012	1409.74	1421.12	1405.14	1412.97	3.51E+09	1412.97	0.002291		
12	24/10/2012	1413.2	1420.04	1407.1	1408.75	3.39E+09	1408.75	-0.00315		
13	23/10/2012	1433.74	1433.74	1407.56	1413.11	3.59E+09	1413.11	-0.01439		
14	22/10/2012	1433.21	1435.46	1422.06	1433.81	3.22E+09	1433.81	0.000419		
15	19/10/2012	1457.34	1457.34	1429.85	1433.19	3.88E+09	1433.19	-0.01657		
16	18/10/2012	1460.94	1464.02	1452.63	1457.34	3.88E+09	1457.34	-0.00246		
17	17/10/2012	1454.22	1462.2	1453.35	1460.91	3.66E+09	1460.91	0.0046		
18	16/10/2012	1440.31	1455.51	1440.31	1454.92	3.57E+09	1454.92	0.010144		
19	15/10/2012	1428.75	1441.31	1427.24	1440.13	3.48E+09	1440.13	0.007965		
20	12/10/2012	1432.84	1438.43	1425.53	1428.59	3.13E+09	1428.59	-0.00297		
21	11/10/2012	1432.82	1443.9	1432.82	1432.84	3.67E+09	1432.84	1.4E-05		
22	10/10/2012	1441.48	1442.52	1430.64	1432.56	3.23E+09	1432.56	-0.00619		
23	09/10/2012	1455.9	1455.9	1441.18	1441.48	3.22E+09	1441.48	-0.0099		
24	08/10/2012	1460.93	1460.93	1453.1	1455.88	2.33E+09	1455.88	-0.00346		
25	05/10/2012	1461.4	1470.96	1456.89	1460.93	3.17E+09	1460.93	-0.00032		

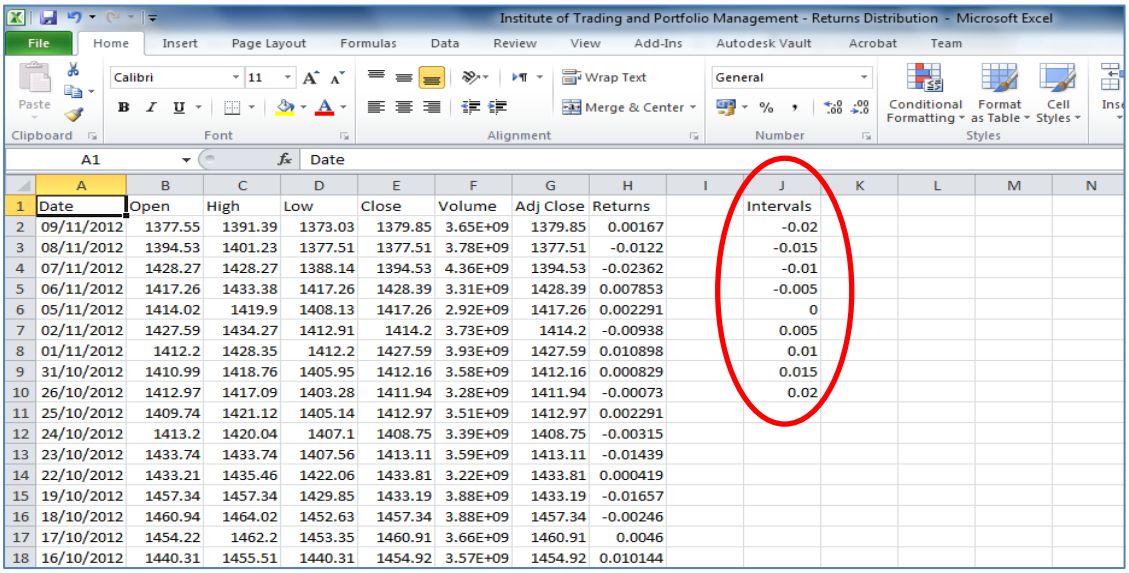
3.2.2 Returns Analysis

With the daily returns in place, we are now going to examine the distribution of these returns. We want to find out how many days the S&P 500 moves up or down 1%, 2%, or any other percentage. To do this we first need to determine between which percentage values we want to analyse. In this example we will analyse from -2% in 0.5% increments to +2%.

(i) **Setting up Intervals**

- *Select cell J1, and title it "Intervals".*
- *In cell J2, enter -0.02 (equivalent to -2%)*
- *In cell J3, enter -0.015*
- *Click and drag to highlight cells J2 AND J3, copy down the formula using the click and drag method mentioned earlier, stopping at cell J10.*

Note: By selecting two cells and copying down, Excel uses the difference in value between the two cells as the incremental change for copying down.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Date	Open	High	Low	Close	Volume	Adj Close	Returns		Intervals				
2	09/11/2012	1377.55	1391.39	1373.03	1379.85	3.65E+09	1379.85	0.00167		-0.02				
3	08/11/2012	1394.53	1401.23	1377.51	1377.51	3.78E+09	1377.51	-0.0122		-0.015				
4	07/11/2012	1428.27	1428.27	1388.14	1394.53	4.36E+09	1394.53	-0.02362		-0.01				
5	06/11/2012	1417.26	1433.38	1417.26	1428.39	3.31E+09	1428.39	0.007853		-0.005				
6	05/11/2012	1414.02	1419.9	1408.13	1417.26	2.92E+09	1417.26	0.002291		0				
7	02/11/2012	1427.59	1434.27	1412.91	1414.2	3.73E+09	1414.2	-0.00938		0.005				
8	01/11/2012	1412.2	1428.35	1412.2	1427.59	3.93E+09	1427.59	0.010898		0.01				
9	31/10/2012	1410.99	1418.76	1405.95	1412.16	3.58E+09	1412.16	0.000829		0.015				
10	26/10/2012	1412.97	1417.09	1403.28	1411.94	3.28E+09	1411.94	-0.00073		0.02				
11	25/10/2012	1409.74	1421.12	1405.14	1412.97	3.51E+09	1412.97	0.002291						
12	24/10/2012	1413.2	1420.04	1407.1	1408.75	3.39E+09	1408.75	-0.00315						
13	23/10/2012	1433.74	1433.74	1407.56	1413.11	3.59E+09	1413.11	-0.01439						
14	22/10/2012	1433.21	1435.46	1422.06	1433.81	3.22E+09	1433.81	0.000419						
15	19/10/2012	1457.34	1457.34	1429.85	1433.19	3.88E+09	1433.19	-0.01657						
16	18/10/2012	1460.94	1464.02	1452.63	1457.34	3.88E+09	1457.34	-0.00246						
17	17/10/2012	1454.22	1462.2	1453.35	1460.91	3.66E+09	1460.91	0.0046						
18	16/10/2012	1440.31	1455.51	1440.31	1454.92	3.57E+09	1454.92	0.010144						

(ii) Frequency Table

We can use these intervals to make a frequency table. This will allow us to see how many of our observations (daily returns) will have values in between our intervals and beyond. From there we can create a histogram to display our results graphically.

For this next section we are going to use an add-in called Analysis Toolpak, making our analysis much easier. In some versions of excel this is not included automatically. To load the Analysis Toolpak take the following steps:

For Excel 2003

1. On the Tools menu, click Add-Ins.
2. In the Add-Ins available box, select the check box next to Analysis Toolpak, and then click OK. Tip If Analysis Toolpak is not listed, click Browse to locate it.
3. If you see a message that tells you the Analysis Toolpak is not currently installed on your computer, click Yes to install it.
4. Click Tools on the menu bar. When you load the Analysis Toolpak, the Data Analysis command is added to the Tools menu.

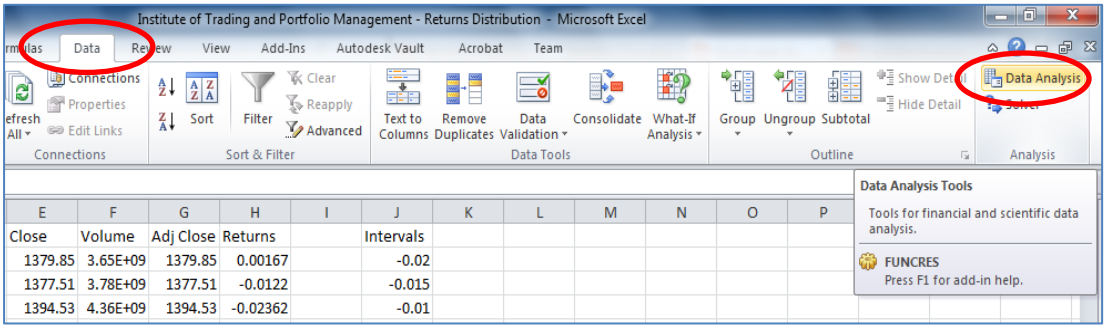
For Excel 2007, (2010)

1. Click (File)→Options, and then click the Add-Ins category.
2. Near the bottom of the Excel Options dialog box, make sure that Excel Add-ins is selected in the Manage box, and then click Go.
3. In the Add-Ins dialog box, select the check box for Analysis ToolPak and then click OK.
4. If Excel displays a message that states it can't run this add-in and prompts you to install it, click Yes to install the add-in.

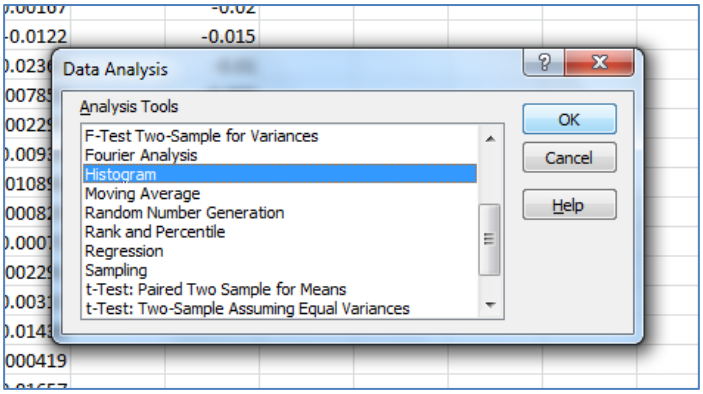
This example will be carried out in Excel 2010. In Excel 2010, Data Analysis is located in the Data ribbon. In Excel 2003, Data Analysis is located under Tools on the menu bar.

Back to the frequency table…

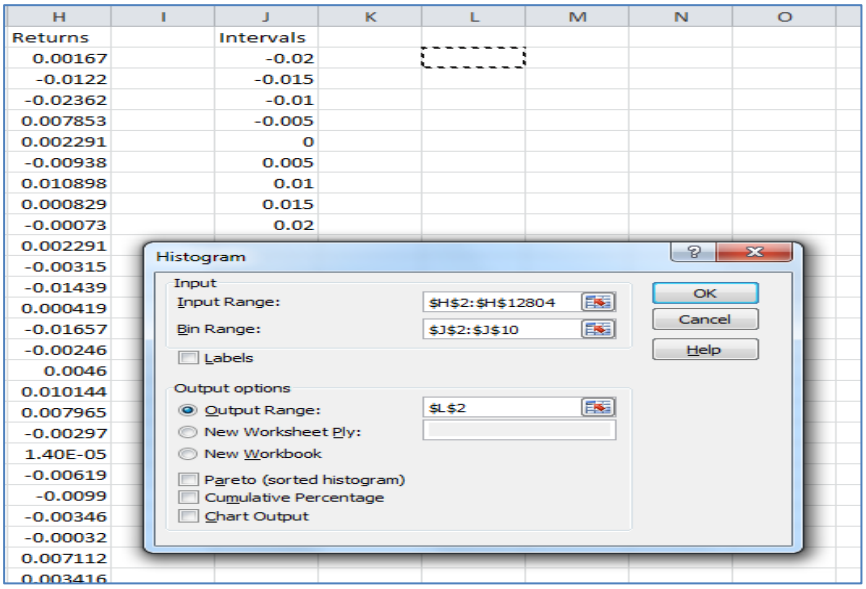
- *Locate Data Analysis and click on it to get to the Data Analysis menu.*

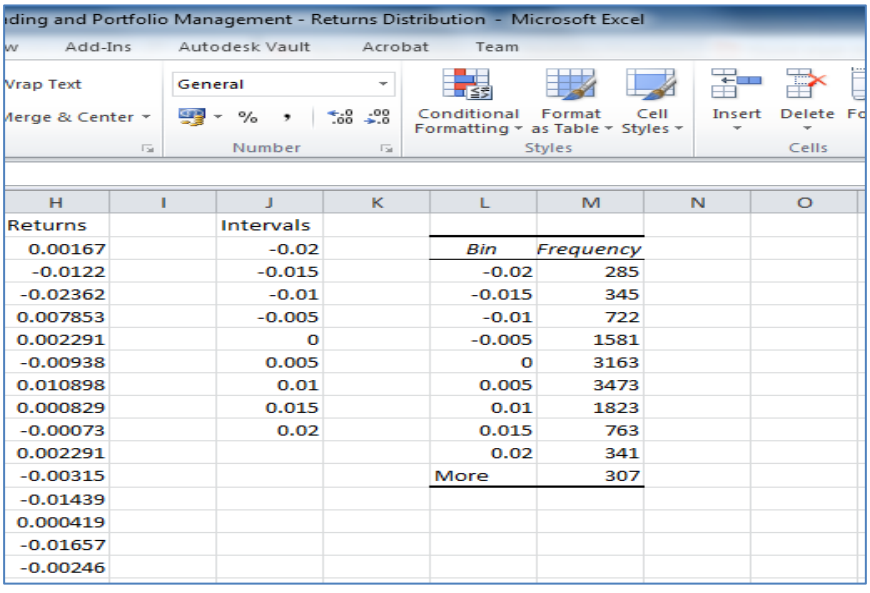


- *Select Histogram from the list and press OK*



- *Select the Input Range for the returns we want to analyse. Highlight cell H2, press CTRL+SHIFT+DOWN to select all the data in that column.*
- *Using the same method for the Bin Range, select our intervals by highlighting cell J2, and press CTRL+SHIFT+DOWN to highlight all the data in that column.*
- *Finally, select Output Range in the Output options. Be careful here as Excel likes to return you to the Input Range box once you have clicked this. MAKE SURE you DO NOT change the input range, and put the cursor in the Output Range box before selecting cell L2. Press OK.*





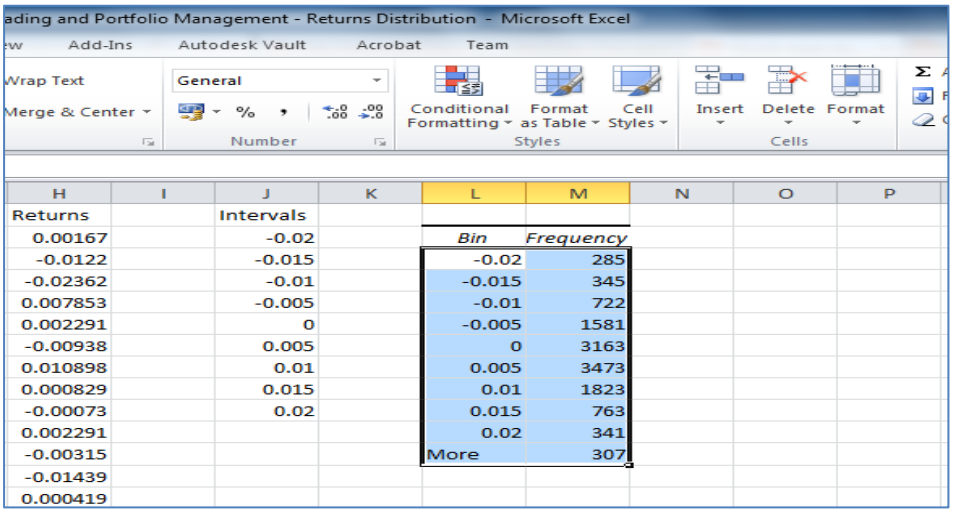
Your frequency table should look like the one above. From this we can plot a histogram to display this information graphically.

AN IMPORTANT NOTE: The frequency information displayed shows the number of days the S&P500 realises returns between its corresponding bin value and the next lesser bin value. For example, Cell M3 tells you how many days the S&P 500 has been down by -2% or more (285). Cell M3 tells you how many days the S&P 500 has been down between -2% and -1.5% (345).

(iii) Frequency Histogram

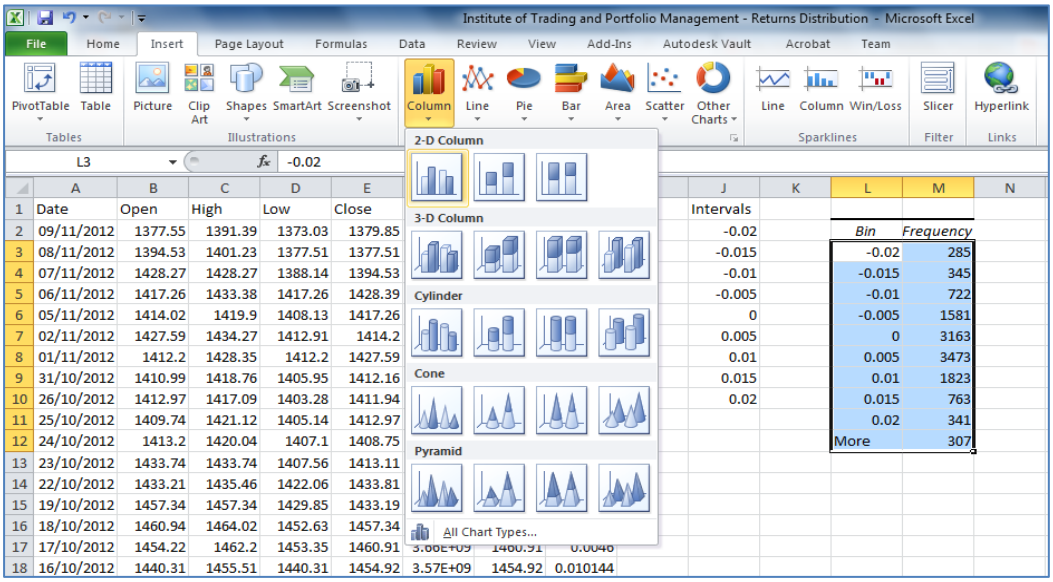
Creating a histogram using the data is very easily. Simply take the following steps:

- *Highlight the bins and histogram data, that is, cells L3 to M12.*

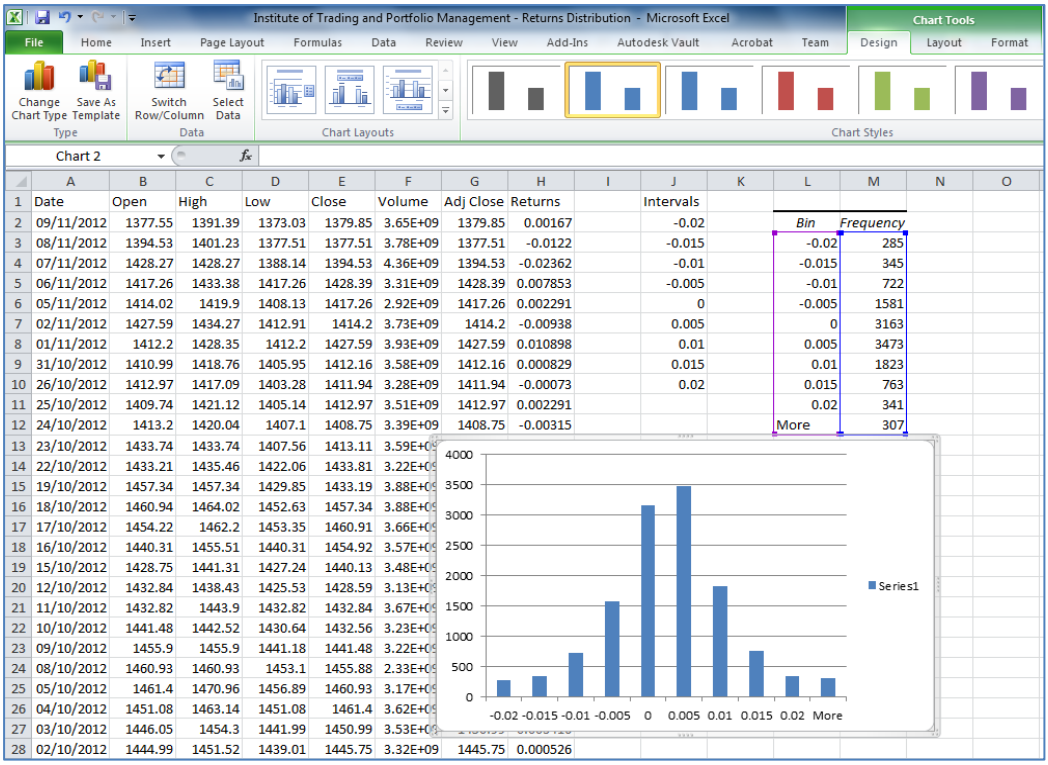


	H	I	J	K	L	M	N	O	P
1	Returns		Intervals						
2	0.00167		-0.02		*Bin*	*Frequency*			
3	-0.0122		-0.015		-0.02	285			
4	-0.02362		-0.01		-0.015	345			
5	0.007853		-0.005		-0.01	722			
6	0.002291		0		-0.005	1581			
7	-0.00938		0.005		0	3163			
8	0.010898		0.01		0.005	3473			
9	0.000829		0.015		0.01	1823			
10	-0.00073		0.02		0.015	763			
11	0.002291				0.02	341			
12	-0.00315				More	307			
13	-0.01439								
14	0.000419								

- *With that range of cells highlighted, go to Insert on the menu bar.*
- *Select Column in the ribbon.*
- *From the drop-down list of graph types, select the first of the 2-D options as highlighted below.*



	A	B	C	D	E	J	K	L	M	N
1	Date	Open	High	Low	Close	Intervals				
2	09/11/2012	1377.55	1391.39	1373.03	1379.85	-0.02		*Bin*	*Frequency*	
3	08/11/2012	1394.53	1401.23	1377.51	1377.51	-0.015		-0.02	285	
4	07/11/2012	1428.27	1428.27	1388.14	1394.53	-0.01		-0.015	345	
5	06/11/2012	1417.26	1433.38	1417.26	1428.39	-0.005		-0.01	722	
6	05/11/2012	1414.02	1419.9	1408.13	1417.26	0		-0.005	1581	
7	02/11/2012	1427.59	1434.27	1412.91	1414.2	0.005		0	3163	
8	01/11/2012	1412.2	1428.35	1412.2	1427.59	0.01		0.005	3473	
9	31/10/2012	1410.99	1418.76	1405.95	1412.16	0.015		0.01	1823	
10	26/10/2012	1412.97	1417.09	1403.28	1411.94	0.02		0.015	763	
11	25/10/2012	1409.74	1421.12	1405.14	1412.97			0.02	341	
12	24/10/2012	1413.2	1420.04	1407.1	1408.75			More	307	
13	23/10/2012	1433.74	1433.74	1407.56	1413.11					
14	22/10/2012	1433.21	1435.46	1422.06	1433.81					
15	19/10/2012	1457.34	1457.34	1429.85	1433.19					
16	18/10/2012	1460.94	1464.02	1452.63	1457.34					
17	17/10/2012	1454.22	1462.2	1453.35	1460.91					
18	16/10/2012	1440.31	1455.51	1440.31	1454.92	3.57E+09	1454.92	0.010144		



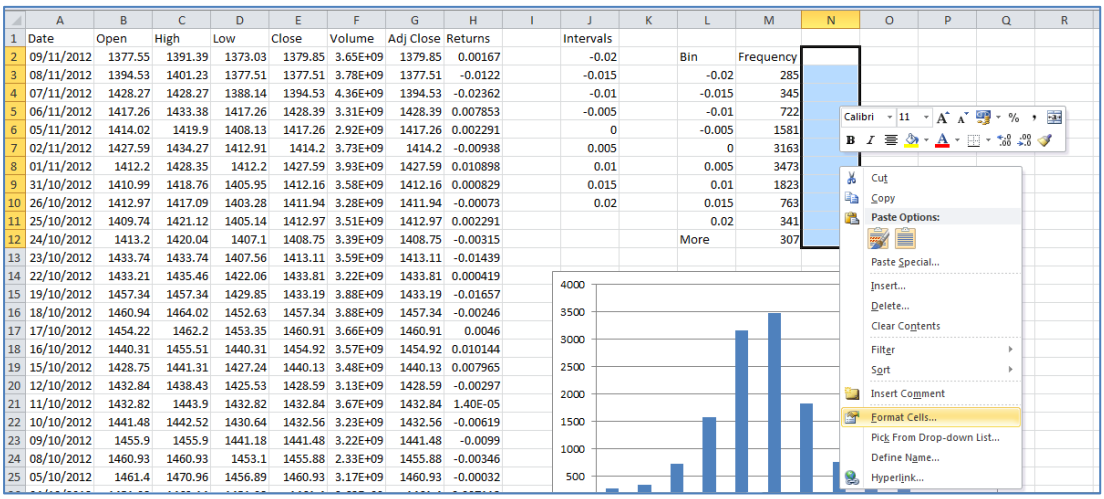
(iv) Editing the Histogram

Note earlier how we discussed the ranges that these frequencies fall into. In this section, we are going to tidy up the graph and display the ranges so our graphical representation is more accurate.

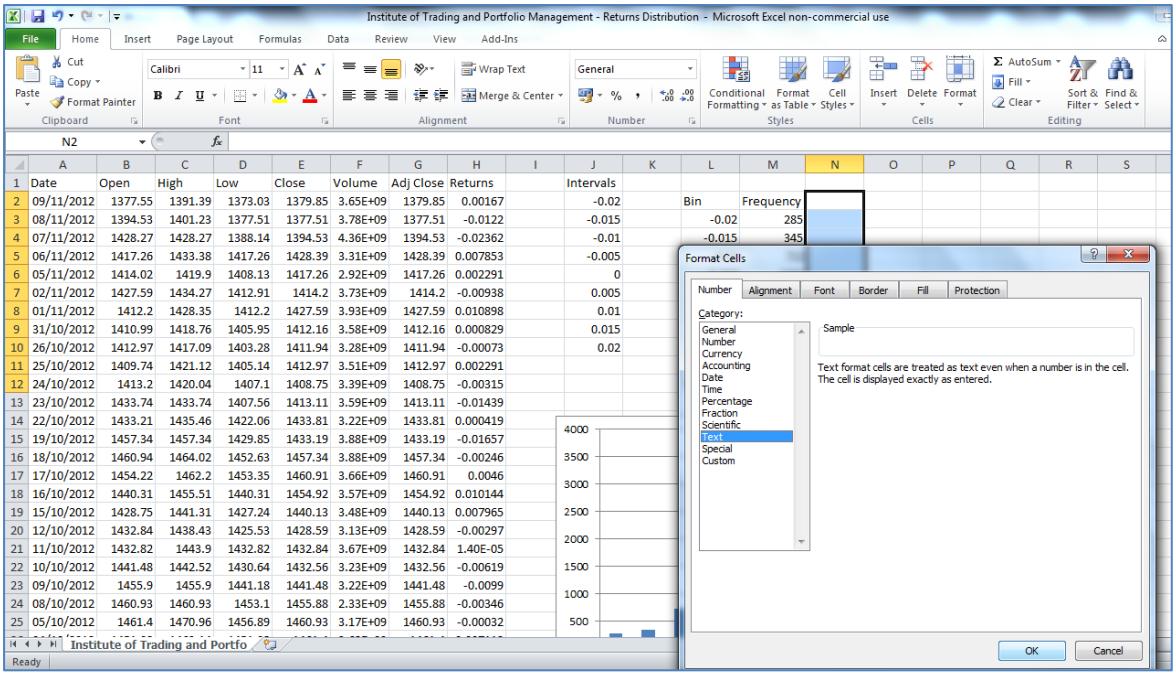
Changing the X-Axis

We are going to display the ranges in terms of percentages rather than decimal values. We will also enter ranges under which the corresponding data falls in the histogram. We will do this by creating another column solely for the purpose of displaying our X-Axis correctly. Remember that a Bin value of 0.02 represents a range of up to 2%, from the Bin value listed before (or from negative infinity if there is no previous intervals/bin values).

- *Select Cells N2 to N12 and right click and go to 'Format cells'*

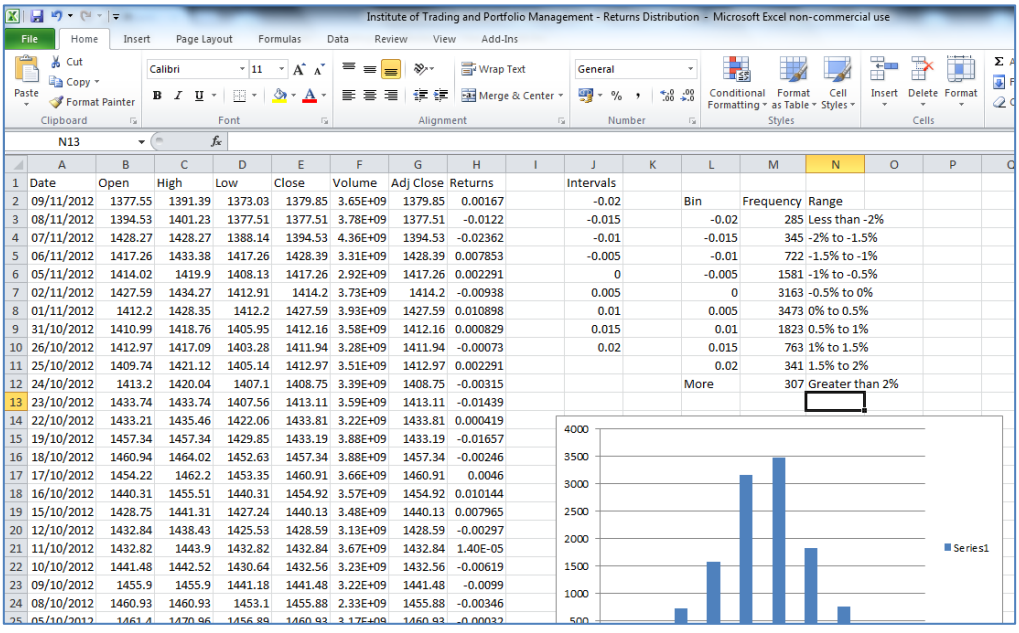


- *Select Category as 'Text' and press OK.*



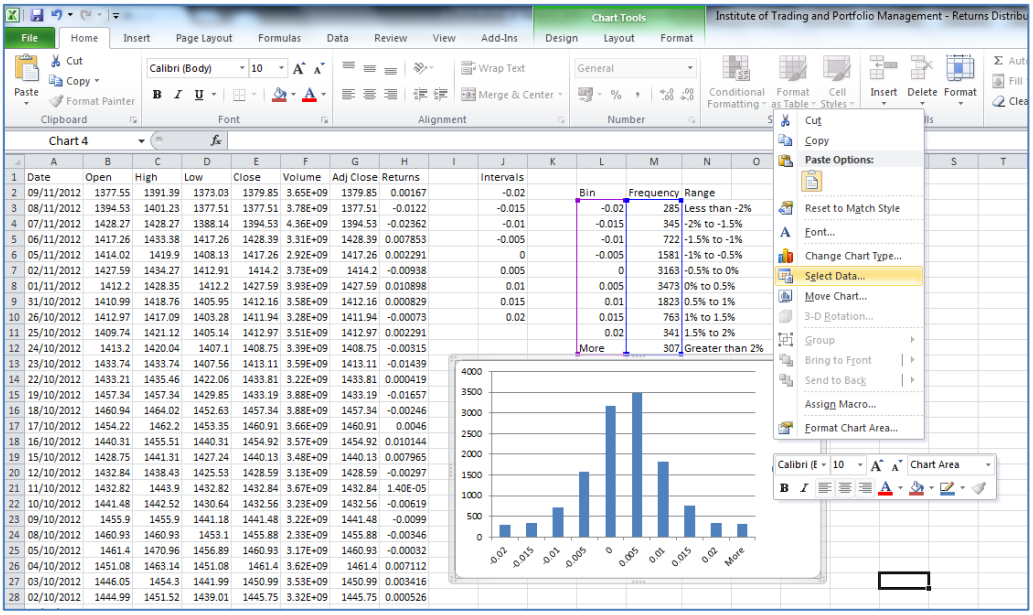
This allows us to enter numbers into these cells in a Text format for labelling purposes, without getting any errors.

- *Select Cell N2 and enter "Range"*
- *Select Cell N3 and enter "Less than -2%"*
- *Select Cell N4 and enter "-2% to-1.5%"*
- *Select Cell N5 and enter "-1.5% to -1%"*
- *Select Cell N6 and enter "-1% to -0.5%"*
- *Select Cell N7 and enter "-0.5% to 0%"*
- *Select Cell N8 and enter "0% to 0.5%"*
- *Select Cell N9 and enter "0.5% to 1%"*
- *Select Cell N10 and enter "1% to 1.5%"*
- *Select Cell N11 and enter "1.5% to 2%"*
- *Select Cell N12 and enter "Greater than 2%"*

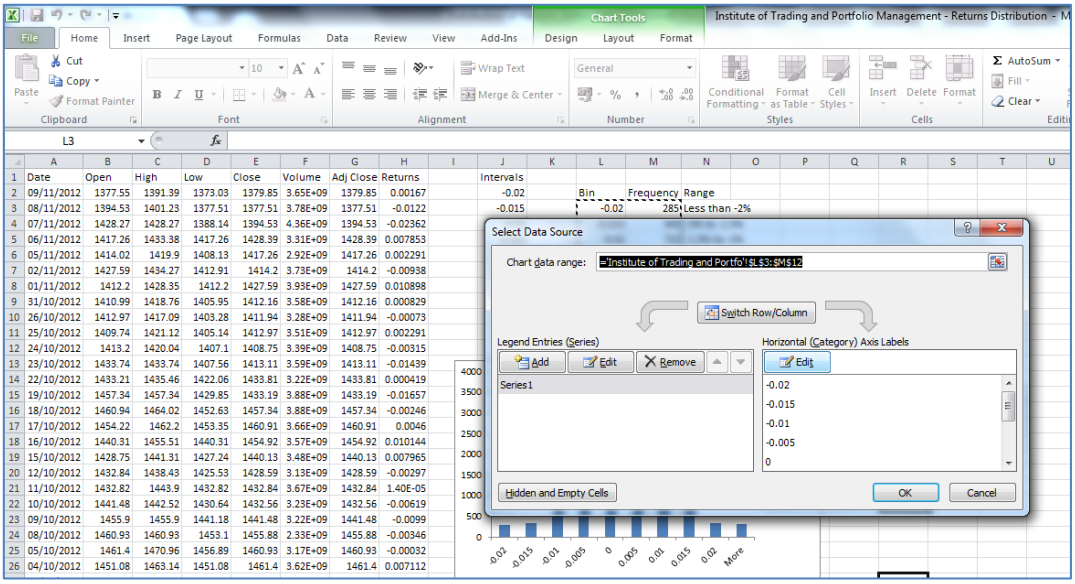


Now we are going to input these values as the X-Axis in our graph.

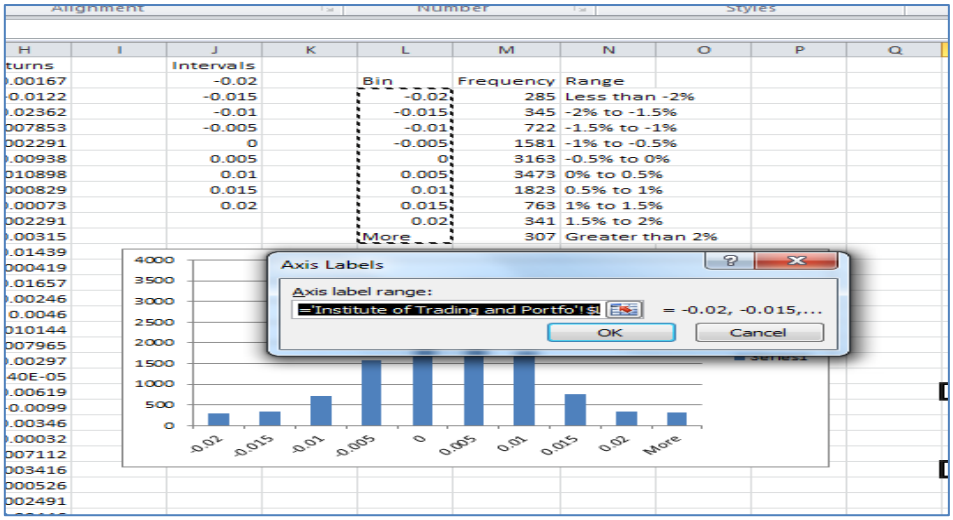
- *Right-click anywhere on the chart*



- *Left-click on 'Select Data'.*

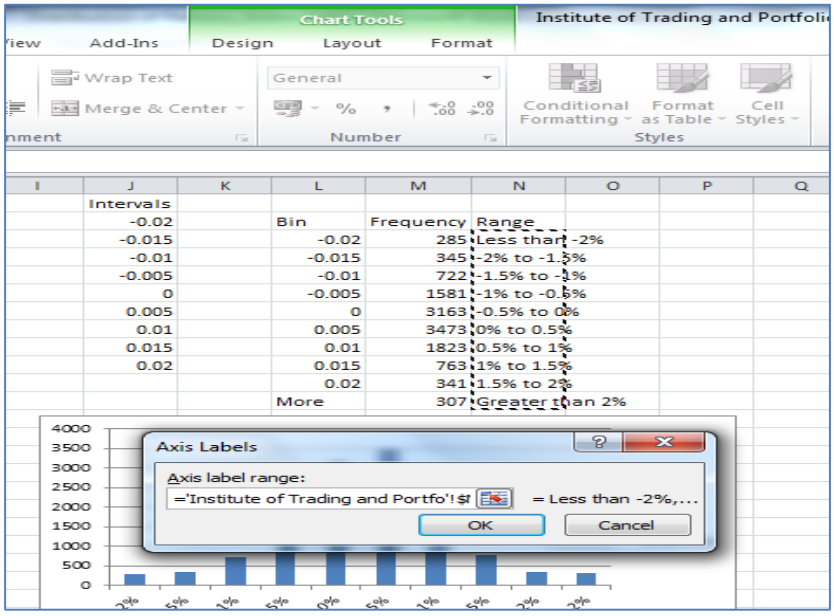


- *Left-click on 'Edit' under the Horizontal Axis Labels column*



Notice how the Bins values are highlighted as the current horizontal-axis labels. We are now going to change these labels to the ones we created in column N.

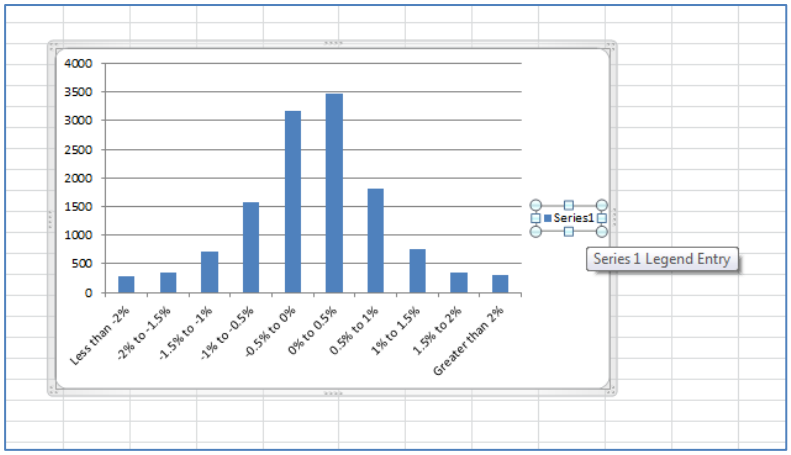
- *Select Cells N3 to N12 and press OK.*

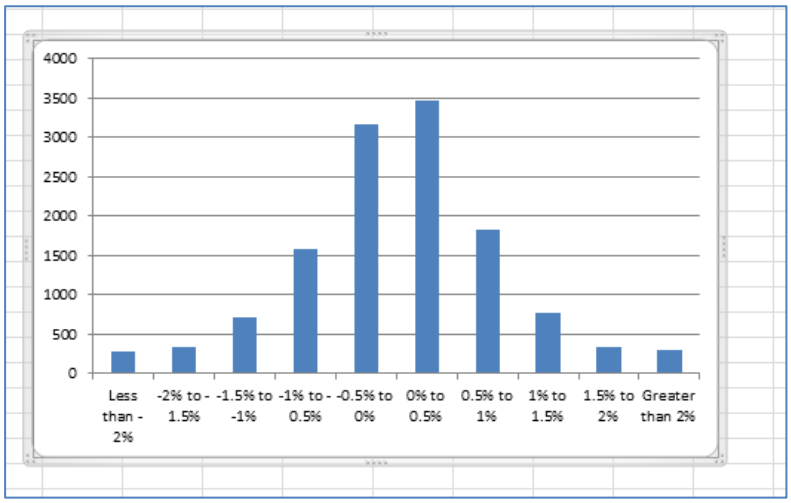


- *Press OK again to get out of the Select Data menu.*

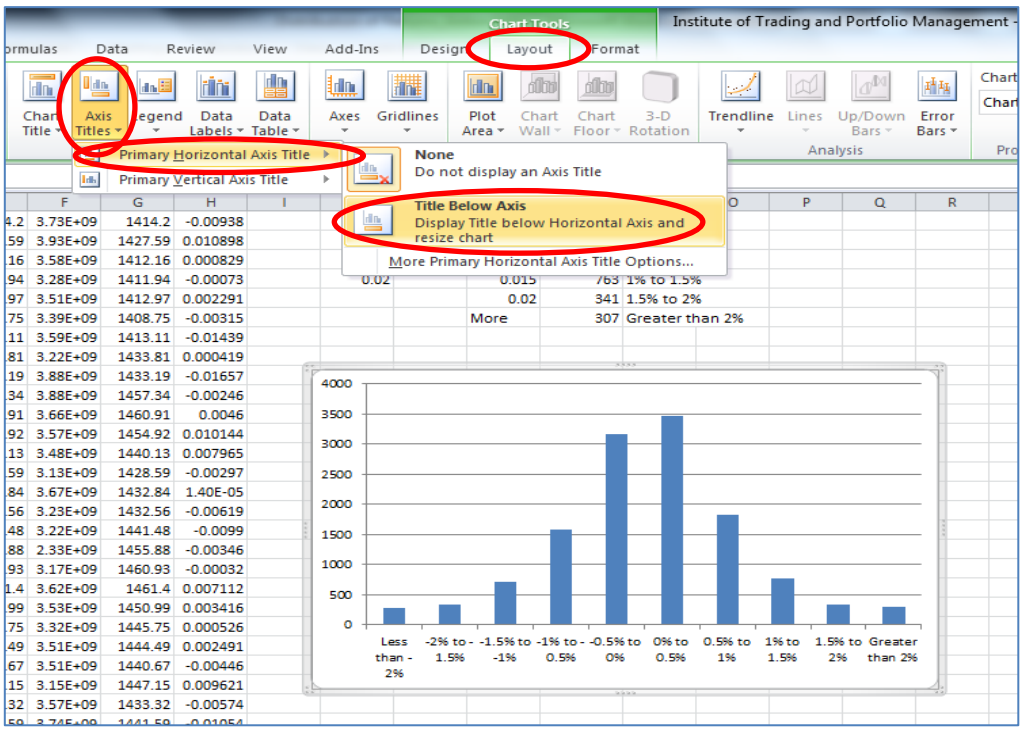
Next we are going to clean up the graph a bit more and label the Horizontal and Vertical Axes.

- *Left-click on 'Series 1' in the graph and delete it by pressing delete on the keyboard.*



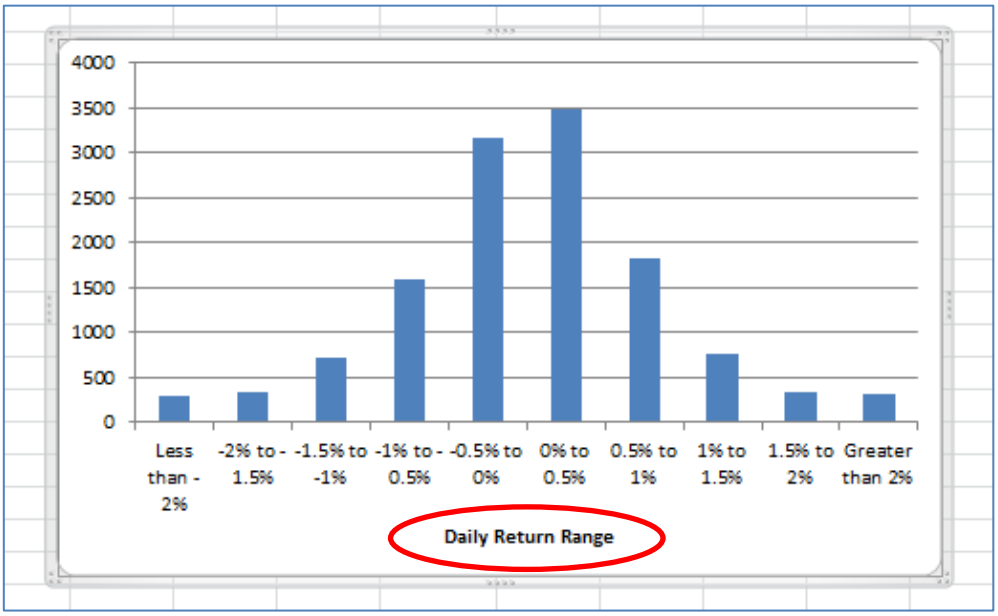


- *Next, left-click on the graph to select it and go to Layout → Axis Titles → Primary Horizontal Axis Title → Title Below Axis*

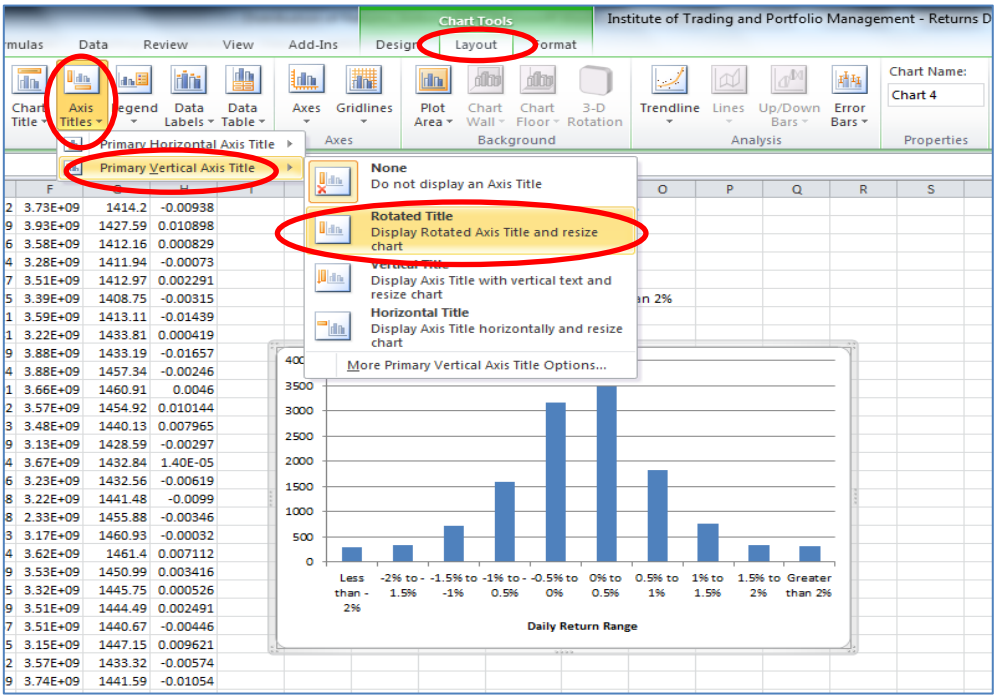


This inserts an editable Text Box below the horizontal axis as a label.

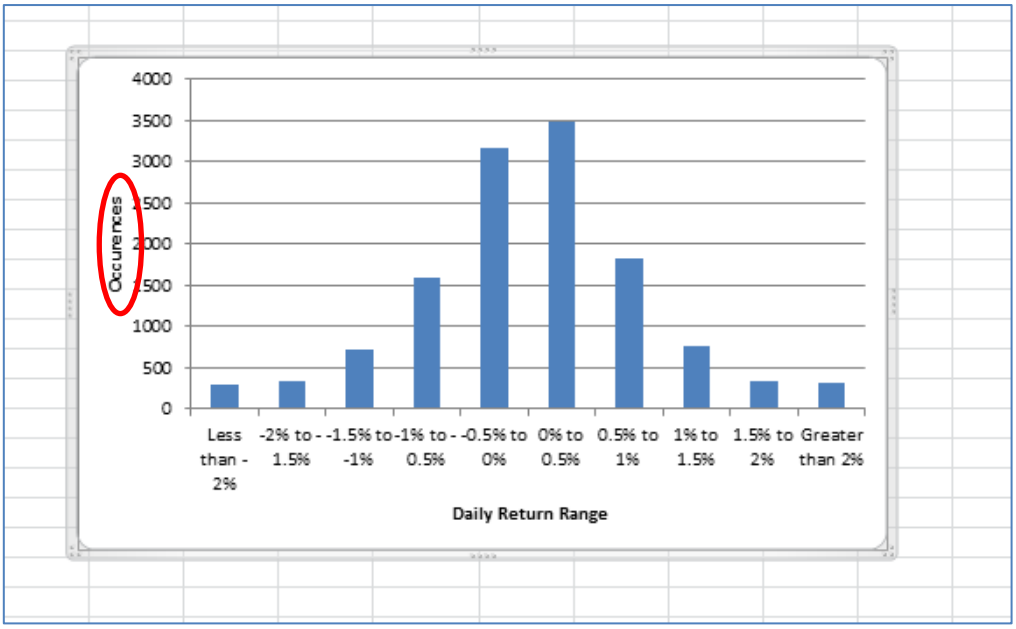
- *Select the label, called "Axis Label", delete its content and rename it 'Daily Return Range'.*



- *Now follow the same process as the last two bullet points but this time select the "Vertical Axis Title" and "Rotated Title" options.*



- *Rename the rotated title 'Occurences'*

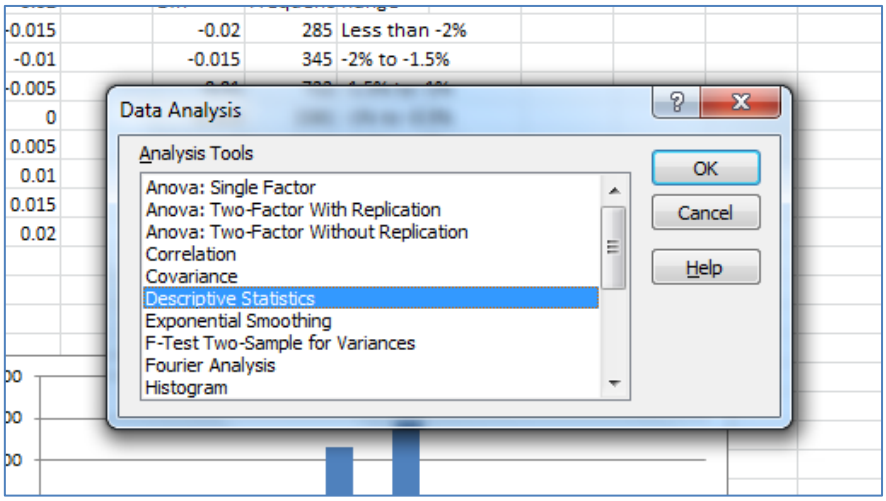


(v) Data Analysis

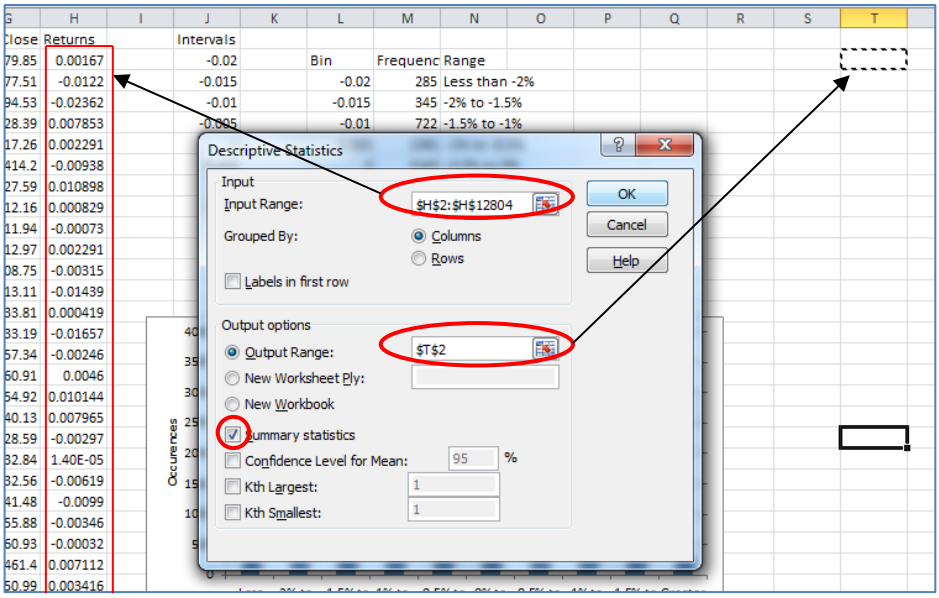
Now that we have our probability distributions displayed both graphically and in tabular format, we can start drawing some statistical conclusions from the data. The first thing to note is the shape of our histogram. It looks fairly similar to the normal distribution discussed earlier, with one key difference – the tails of the distribution (the extremities in price movements) have many more occurrences than a normal distribution would have. Traders call these 'fat tails'. It is important to be aware of how likely these volatile daily movements occur.

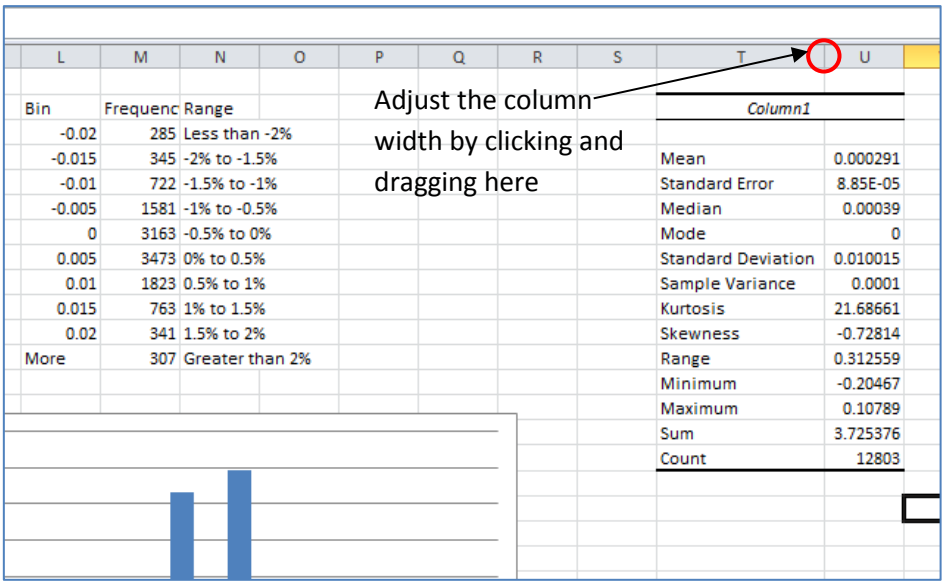
We are going to start with some summary statistics on our data. Excel makes these easily obtainable. The data analysis tool will be used again but this time to obtain descriptive statistics - simply follow these steps:

- *Navigate to Data → Data Analysis as before. This time select Descriptive statistics and press OK.*



- *Select the returns column (highlight H2, press CTRL+SHIFT+DOWN), select T2 for the output range and check the summary statistics box. Now press OK.*





- *Adjust the column width of T so you can see all the statistics in full*

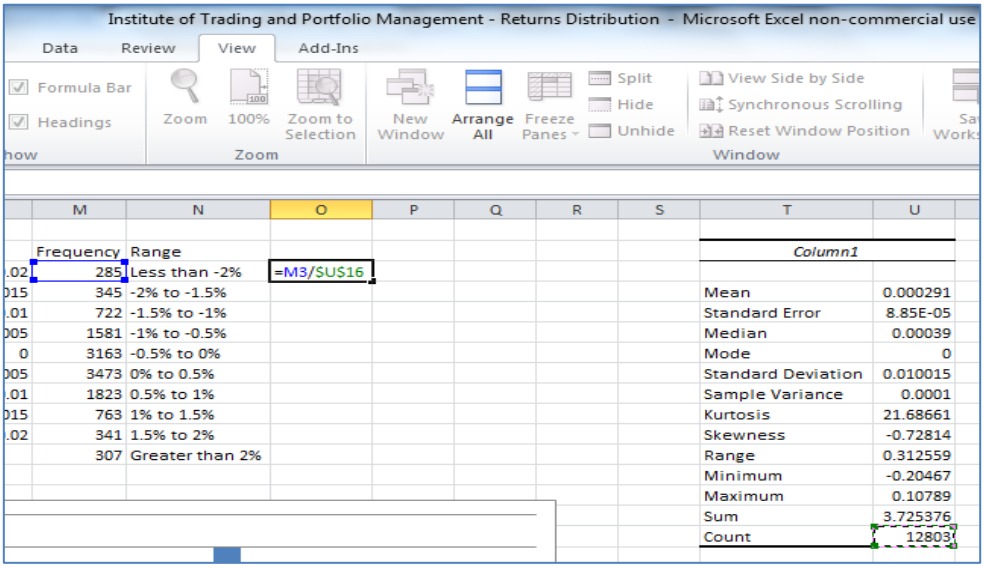
Let's talk through some of the key statistics shown here.

- **Mean** – This is the average daily return from the sample we have analysed. It is also known as the Expected Return as it is the amount we expect to make if we buy on the open and sell on the close on any given day. The value is 0.03%.
- **Median** – This is the middle number in our sample. As this is positive, we can intuitively see that there are more days of positive returns than negative. Furthermore, as it is greater than the mean, it is also implied that negative returns must be of higher magnitude than positive ones.
- **Standard Deviation** – If we assume that our distribution is normal we can say that 68.2% of data lies within one standard deviation either side of the mean, 95.4% lies within two and 99.8% within three, as discussed at the beginning of this tutorial.
- **Kurtosis** – Don't worry too much about this one. Just know that if this value is positive using Excel's method of calculation, then the distribution has 'fat tails' and is not strictly normally distributed.
- **Skewness** –Skew is another sign of a non-normal distribution. Negative skew means that negative values are more extreme than positive ones.

- **Range** – This tells you the difference between the highest positive and lowest negative return in the data set. The value of 0.312 (31.2%) in this sample comes from a lowest negative of about -20% (yes, in one day) and highest positive of roughly 10%. This is confirmed by looking at the **Maximum** and **Minimum** values. Beware of those extremes.
- **Sum** – The sum of all the returns we have calculated. Ignore it. Our method of calculation does NOT give return you would expect over the entire period if you invested from 1962 to 2012.
- **Count** – The total number of trading days analysed.

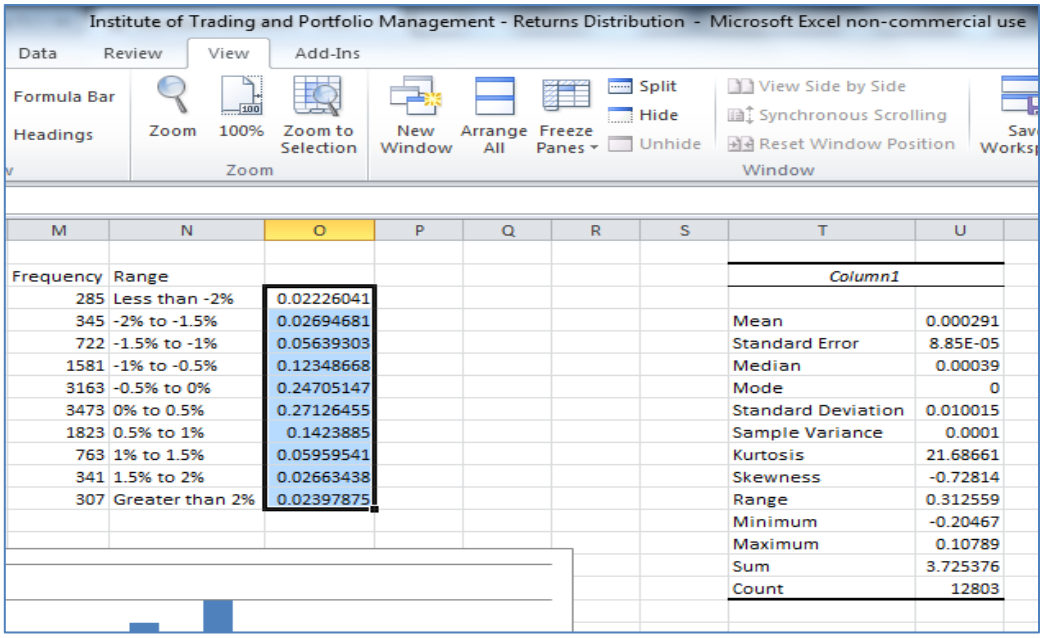
It would be useful to know the probability of different ranges of returns occurring given the historic data. Next we will use the numbers in our frequency table to create some easily understandable probability values for us to interpret. For each of the number of occurrences that falls in a given range, we are going to divide it by the total number of trading days so we can see the percentage of trading days that returns fall within the range specified.

- *Select Cell O3 and type the following equation: "=M3/U16" … Before you press enter press F4 on your keyboard and it should make the formula look like this "=M3/$U$16". If it does not, type this formula in manually or try again. Now press Enter.*



Adding dollar signs around U16 within the equation adds a 'lock' to that cell (U16). So when we want to copy the formula down, U16 is continuously referenced as the denominator in the equation. Simply, when we copy down the formula it will use the frequency on each row divided by the total number of trading days (count).

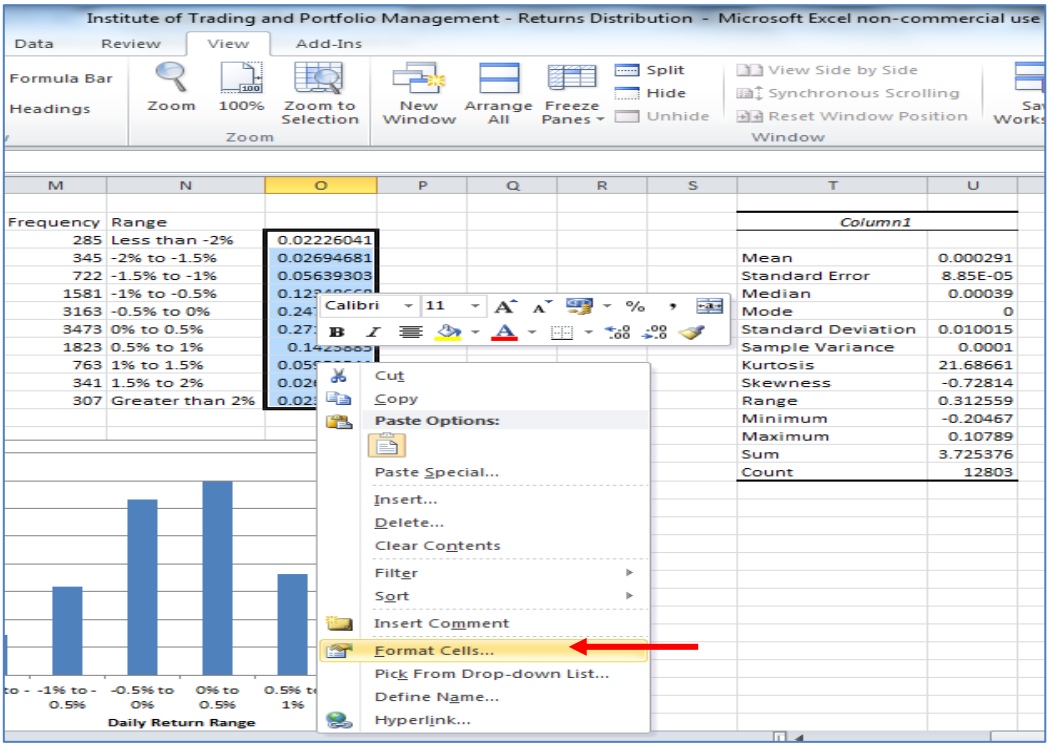
- *Next, highlight cells O3 to O12 and press CTRL+D to copy the formula down and apply to each of the cells selected.*



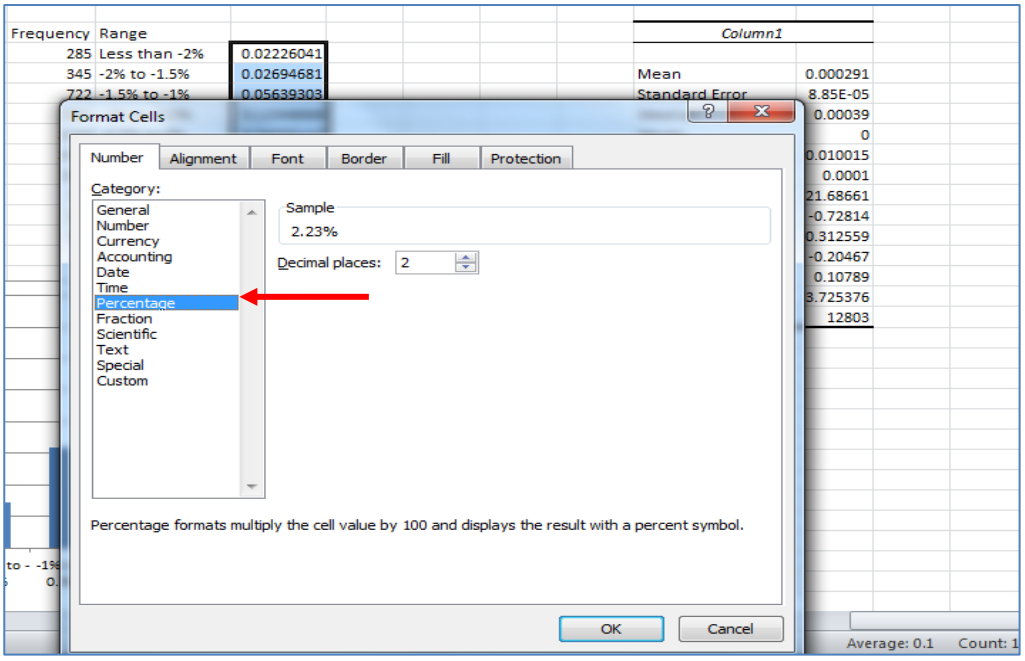
Frequency	Range	
285	Less than -2%	0.02226041
345	-2% to -1.5%	0.02694681
722	-1.5% to -1%	0.05639303
1581	-1% to -0.5%	0.12348668
3163	-0.5% to 0%	0.24705147
3473	0% to 0.5%	0.27126455
1823	0.5% to 1%	0.1423885
763	1% to 1.5%	0.05959541
341	1.5% to 2%	0.02663438
307	Greater than 2%	0.02397875

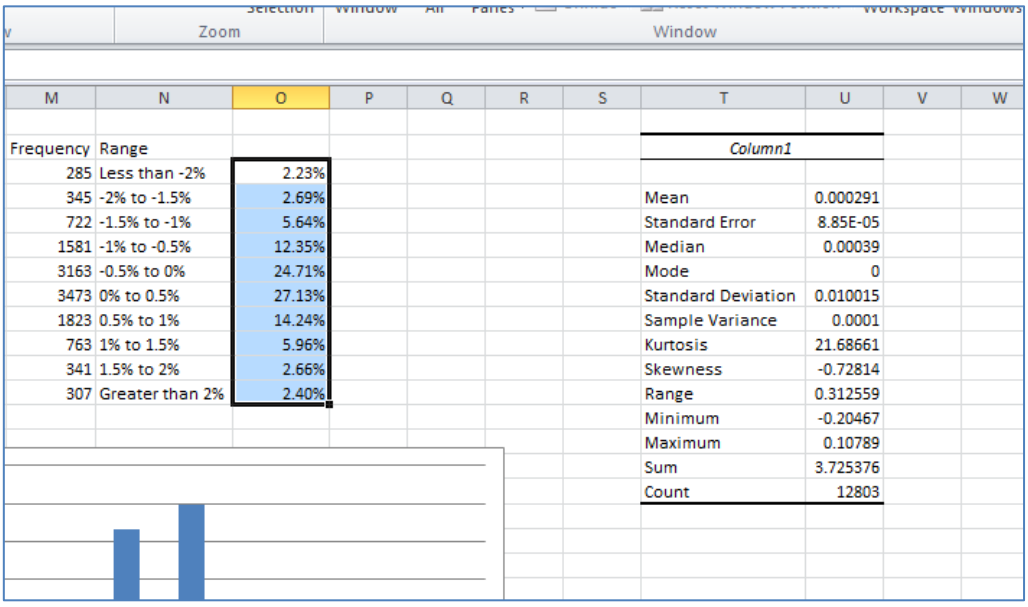
Column1	
Mean	0.000291
Standard Error	8.85E-05
Median	0.00039
Mode	0
Standard Deviation	0.010015
Sample Variance	0.0001
Kurtosis	21.68661
Skewness	-0.72814
Range	0.312559
Minimum	-0.20467
Maximum	0.10789
Sum	3.725376
Count	12803

- *Keep these cells highlighted and right-click on the highlighted area. Go to format cells.*



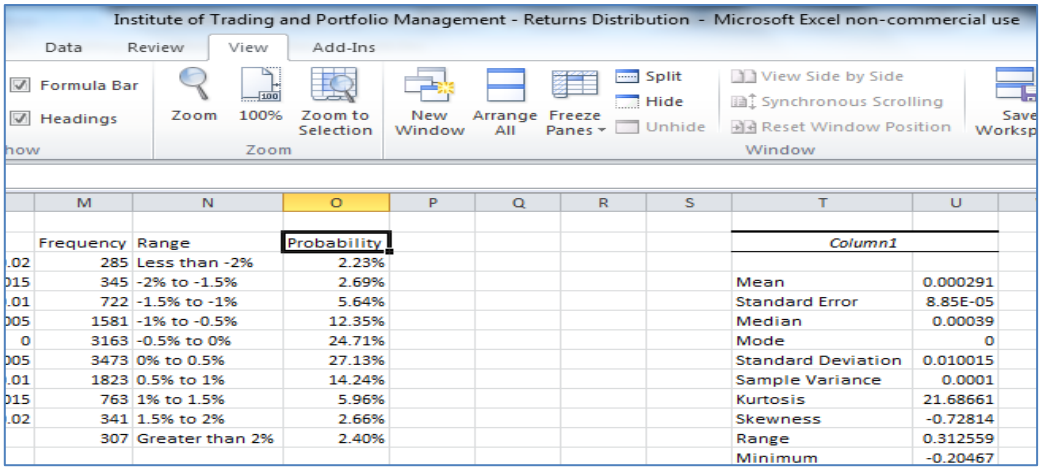
- *Click on 'Percentage' as the cell format and press OK.*





We now have a column of percentages showing the percentage of our past returns that fall within these ranges. Therefore we can argue that these percentages represent our percentage probability of each of these returns happening in the future, if we believe that historical returns provide a good prediction.

- *Select Cell O2 and type 'Probability' to head the column. Press Enter.*



	M	N	O	P	Q	R	S	T	U
	Frequency	Range	Probability					*Column1*	
.02	285	Less than -2%	2.23%						
015	345	-2% to -1.5%	2.69%					Mean	0.000291
.01	722	-1.5% to -1%	5.64%					Standard Error	8.85E-05
005	1581	-1% to -0.5%	12.35%					Median	0.00039
0	3163	-0.5% to 0%	24.71%					Mode	0
005	3473	0% to 0.5%	27.13%					Standard Deviation	0.010015
.01	1823	0.5% to 1%	14.24%					Sample Variance	0.0001
015	763	1% to 1.5%	5.96%					Kurtosis	21.68661
.02	341	1.5% to 2%	2.66%					Skewness	-0.72814
	307	Greater than 2%	2.40%					Range	0.312559
								Minimum	-0.20467

Let's now take this a step further by adding a column for cumulative percentages, so we can start deducing more probabilities. At this stage it is worth noting that any returns with a value exactly equal to one of the intervals, for example -1.5, is counted as a daily return within the LOWER range. So, a return of -1.5% is counted in the range of "-2% to -1.5%", NOT "-1.5% to 1%". The same is true for -2%. An exact return of -2% is counted in the range of "Less than -2%" NOT in "-2% to -1.5%". This applies to the whole frequency distribution. Using this principle, let's continue:

- *In Cell P2, type 'Cumulative Percentage' to head the column.*
- *Adjust the column width to fit all the characters into the cell.*
- *In Cell P3, type "=O3" and press Enter.*

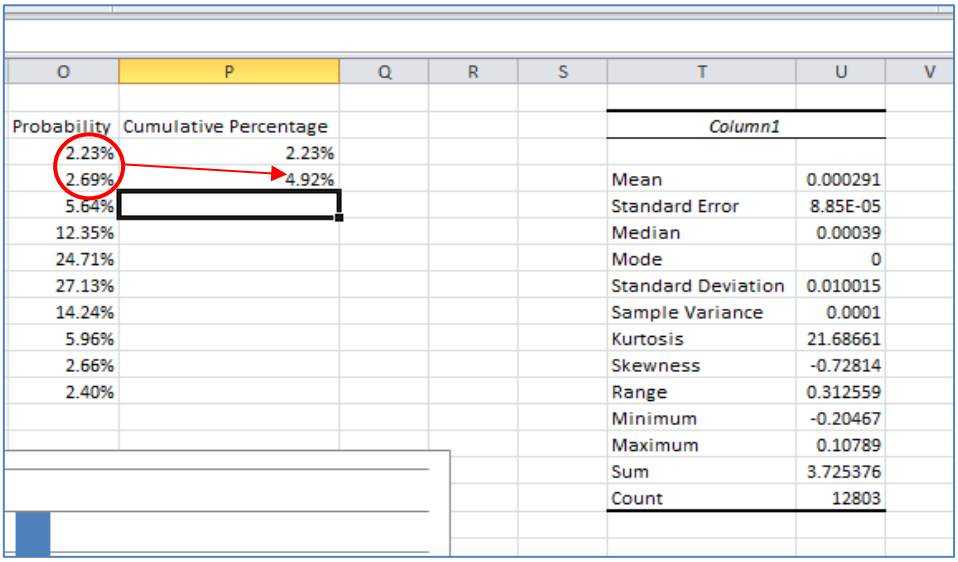
Here it gets a little trickier. In order to calculate cumulative percentages we are going to reference all the percentages we have calculated so far (by using the previous P cell) and adding the new percentage probability for the new range. All will become clear when we go through it. At this stage you should be here:

	N	O	P	Q	R	S	T	U	V
	Range	Probability	Cumulative Percentage				*Column1*		
	Less than -2%	2.23%	2.23%						
	-2% to -1.5%	2.69%					Mean	0.000291	
	-1.5% to -1%	5.64%					Standard Error	8.85E-05	
	-1% to -0.5%	12.35%					Median	0.00039	
	-0.5% to 0%	24.71%					Mode	0	
	0% to 0.5%	27.13%					Standard Deviation	0.010015	
	0.5% to 1%	14.24%					Sample Variance	0.0001	
	1% to 1.5%	5.96%					Kurtosis	21.68661	
	1.5% to 2%	2.66%					Skewness	-0.72814	
	Greater than 2%	2.40%					Range	0.312559	
							Minimum	-0.20467	
							Maximum	0.10789	
							Sum	3.725376	
							Count	12803	

- *In Cell P4, type "=P3+O4". Press Enter.*

Range	Probability	Cumulative Percentage
Less than -2%	2.23%	2.23%
-2% to -1.5%	2.69%	=P3+O4
-1.5% to -1%	5.64%	
-1% to -0.5%	12.35%	
-0.5% to 0%	24.71%	
0% to 0.5%	27.13%	
0.5% to 1%	14.24%	
1% to 1.5%	5.96%	
1.5% to 2%	2.66%	
Greater than 2%	2.40%	

Column1	
Mean	0.000291
Standard Error	8.85E-05
Median	0.00039
Mode	0
Standard Deviation	0.010015
Sample Variance	0.0001
Kurtosis	21.68661
Skewness	-0.72814
Range	0.312559
Minimum	-0.20467
Maximum	0.10789
Sum	3.725376
Count	12803



Now you can see what is happening when we calculate the cumulative percentage. Each of column P's cells equal the sum of all column O's cells up to the same row number.

- *Select Cells P4 to P12 and copy down. CTRL+D. <u>Do not include cell P3!!!</u> (it's formula is a different construction to the one we want to apply to the following rows)*

Frequency	Range	Probability	Cumulative Percentage
285	Less than -2%	2.23%	2.23%
345	-2% to -1.5%	2.69%	4.92%
722	-1.5% to -1%	5.64%	10.56%
1581	-1% to -0.5%	12.35%	22.91%
3163	-0.5% to 0%	24.71%	47.61%
3473	0% to 0.5%	27.13%	74.74%
1823	0.5% to 1%	14.24%	88.98%
763	1% to 1.5%	5.96%	94.94%
341	1.5% to 2%	2.66%	97.60%
307	Greater than 2%	2.40%	100.00%

Column1	
Mean	0.000291
Standard Error	8.85E-05
Median	0.00039
Mode	0
Standard Deviation	0.010015
Sample Variance	0.0001
Kurtosis	21.68661
Skewness	-0.72814
Range	0.312559
Minimum	-0.20467
Maximum	0.10789
Sum	3.725376
Count	12803

As you can see this is another useful way to convey percentages. From this we can start concluding things like:

- 94.94% of the time, daily returns yield less than or equal to 1.5%.
- 11.02%, that is 100%-88.98%, of the time daily returns are in excess of 1%.

Perhaps the MOST useful is the conclusion that:

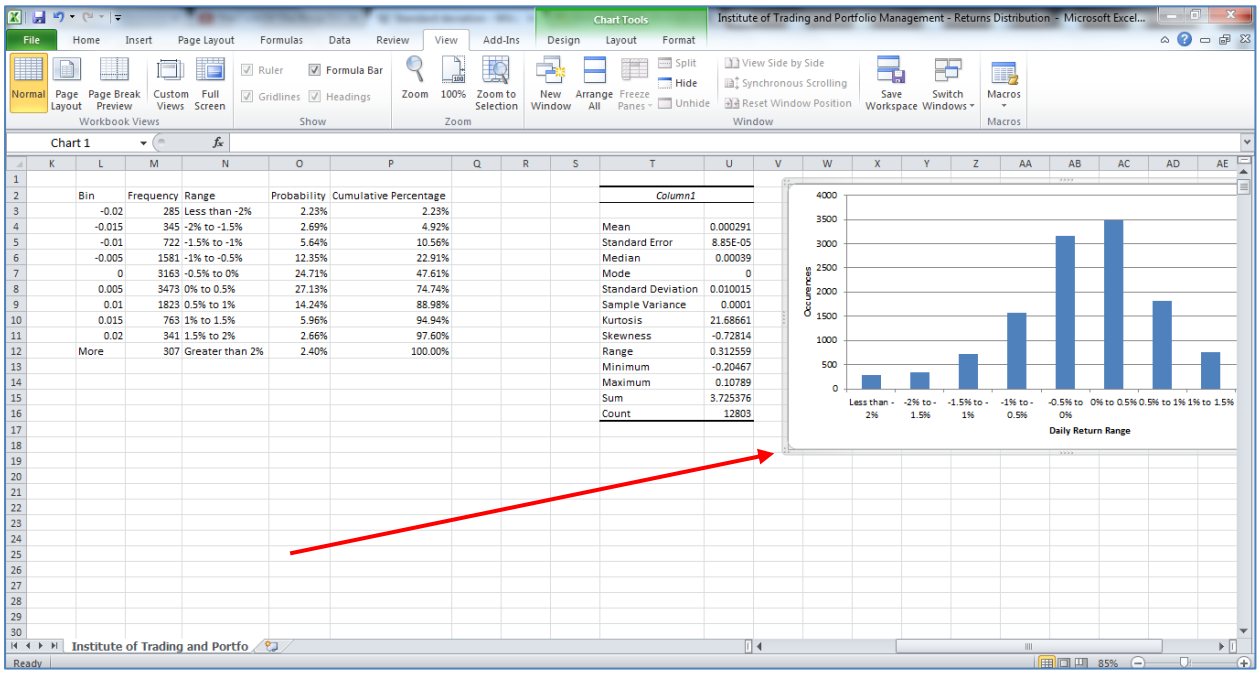
- 47.61% of the time, the S&P 500 has yielded negative or 0% daily returns. Meaning…
- 52.39% of the time, the S&P 500 has had positive daily returns.

Let's remember here, that despite the stats saying the majority of the time the S&P 500 yields positive returns, there are three reasons why you should not conclude (at least straight away) that this is a good investment:

1. Historical returns don't necessarily predict future returns.
2. It is possible to have a long string of losing days that are likely to make an investor blow up.
3. REMEMBER THE DESCRIPTIVE STATISTICS! Negative skew and 'fat tails' mean negative returns hit harder and are more regular in occurrence than is predicted by normal distribution. In short, these are the sort of returns that make the average retail trader blow up.

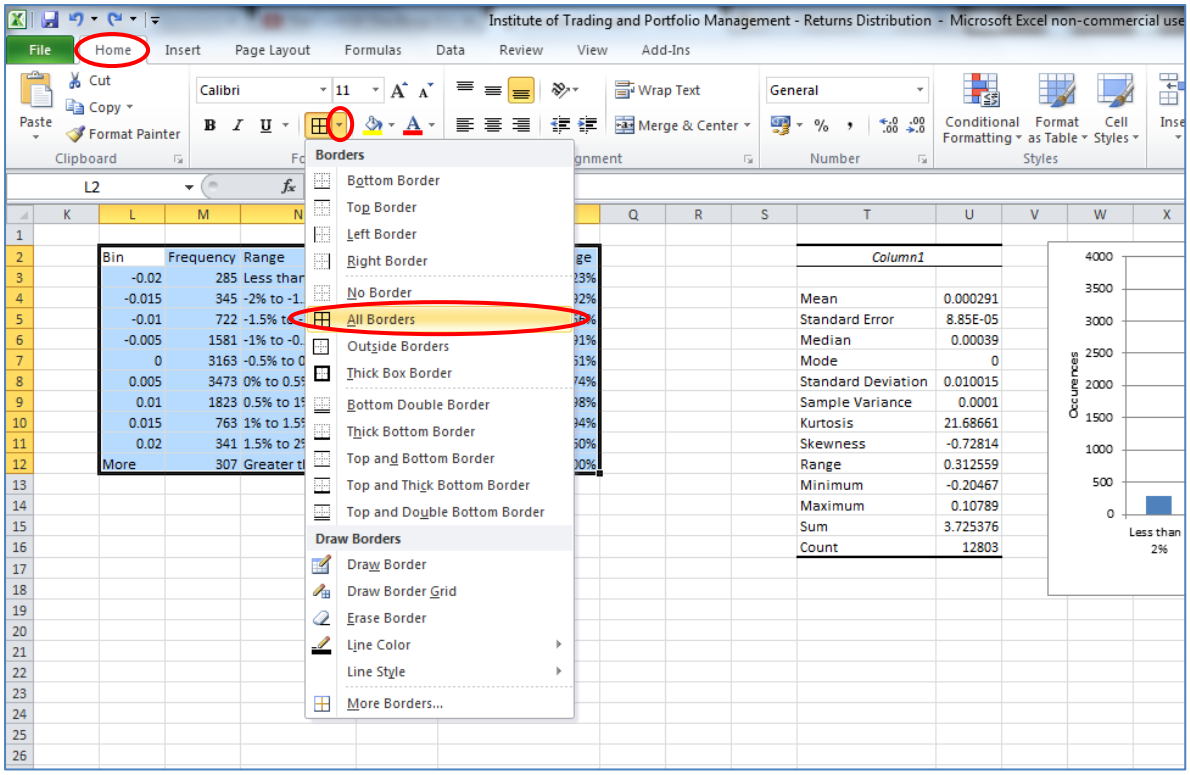
Let's rearrange the spreadsheet to make it a little more presentable and easier to continue working with.

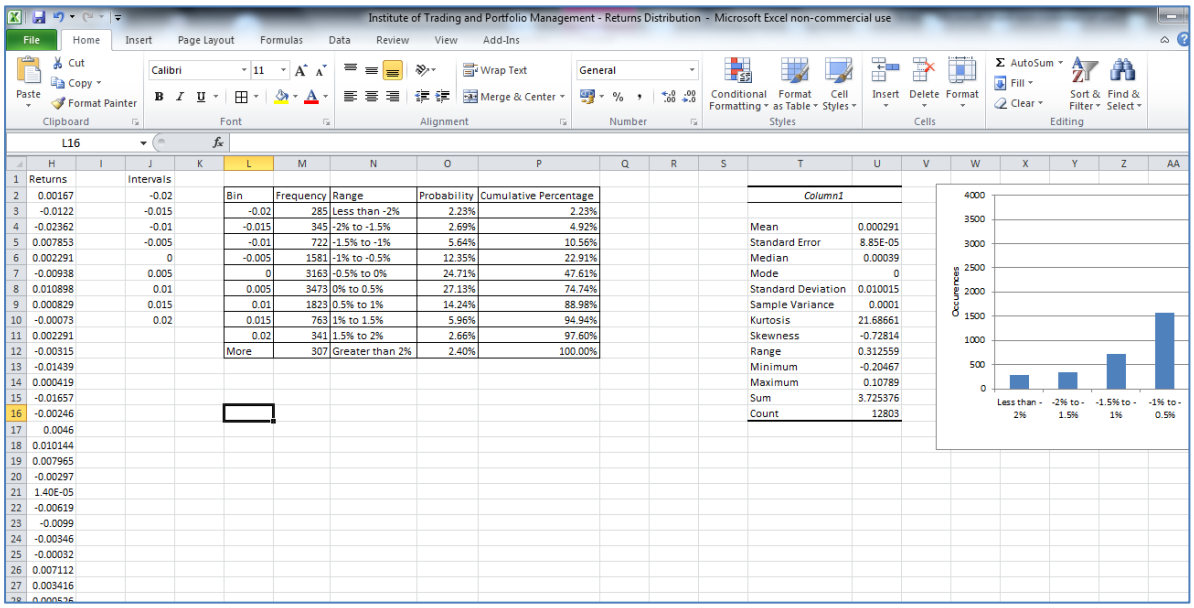
- *Click and drag the graph over to the right-hand side past the descriptive statistics into column V or W.*



Next, let's make the useful calculations we have done so far stand out a bit more in the spreadsheet.

- *Select L2 to P12 and outline all the cells in the range with a border by going to Home →Borders →All Borders. See the navigation below.*

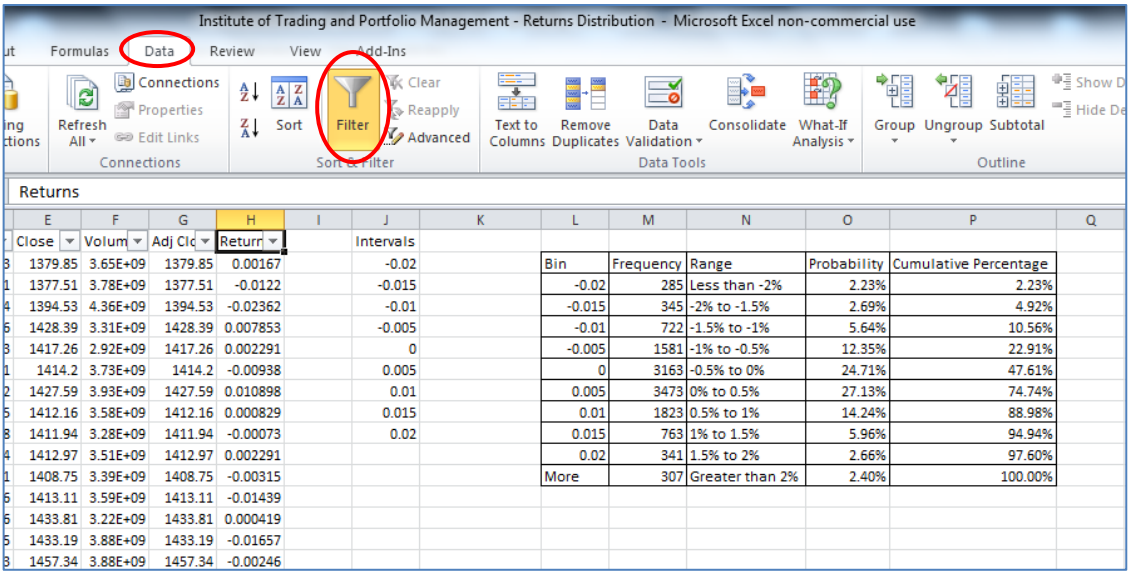




In order to establish some more detailed information on our data we are going to use the Filter toggle in Excel which will allow us to calculate statistics on certain filtered data from our sample. Sounds a little complicated, but stick with it and we will have some useful numbers to look at.

To add a filter to our data we need click on 'Returns' cell H1 and apply the filter.

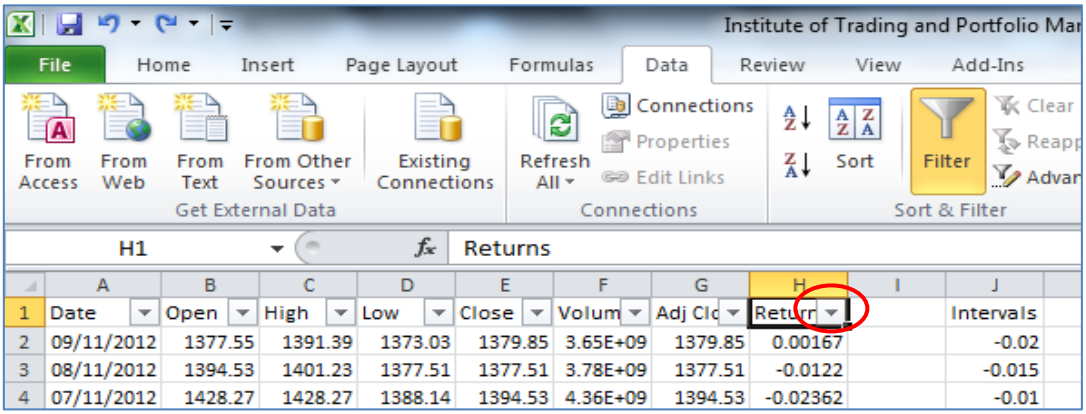
- *Select cell H1*
- *With this cell selected, add a filter by EITHER navigating to Data→Filter OR pressing CTRL+SHIFT+L.*



	E	F	G	H	I	J	K	L	M	N	O	P	Q
	Close	Volum	Adj Clo	Returr		Intervals							
3	1379.85	3.65E+09	1379.85	0.00167		-0.02							
1	1377.51	3.78E+09	1377.51	-0.0122		-0.015							
4	1394.53	4.36E+09	1394.53	-0.02362		-0.01							
6	1428.39	3.31E+09	1428.39	0.007853		-0.005							
8	1417.26	2.92E+09	1417.26	0.002291		0							
1	1414.2	3.73E+09	1414.2	-0.00938		0.005							
2	1427.59	3.93E+09	1427.59	0.010898		0.01							
5	1412.16	3.58E+09	1412.16	0.000829		0.015							
8	1411.94	3.28E+09	1411.94	-0.00073		0.02							
4	1412.97	3.51E+09	1412.97	0.002291									
1	1408.75	3.39E+09	1408.75	-0.00315									
6	1413.11	3.59E+09	1413.11	-0.01439									
5	1433.81	3.22E+09	1433.81	0.000419									
5	1433.19	3.88E+09	1433.19	-0.01657									
8	1457.34	3.88E+09	1457.34	-0.00246									

Bin	Frequency	Range	Probability	Cumulative Percentage
-0.02	285	Less than -2%	2.23%	2.23%
-0.015	345	-2% to -1.5%	2.69%	4.92%
-0.01	722	-1.5% to -1%	5.64%	10.56%
-0.005	1581	-1% to -0.5%	12.35%	22.91%
0	3163	-0.5% to 0%	24.71%	47.61%
0.005	3473	0% to 0.5%	27.13%	74.74%
0.01	1823	0.5% to 1%	14.24%	88.98%
0.015	763	1% to 1.5%	5.96%	94.94%
0.02	341	1.5% to 2%	2.66%	97.60%
More	307	Greater than 2%	2.40%	100.00%

Once you have activated the filter, a small arrow representing a drop-down menu will appear in the top cell of each of your data columns A to H. See the next picture for what this should look like.

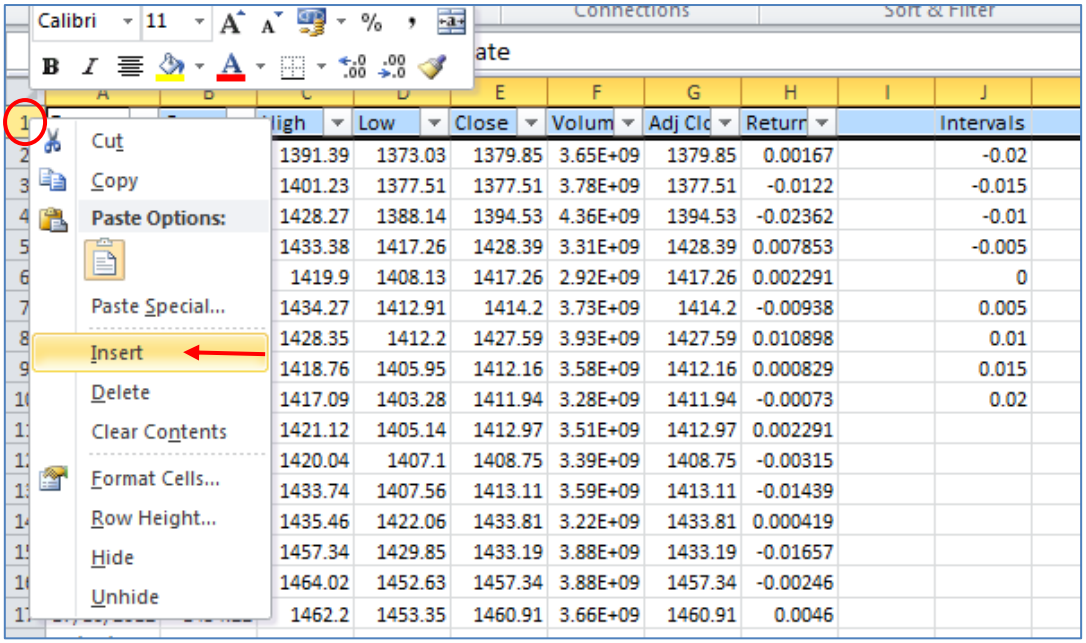


	A	B	C	D	E	F	G	H	I	J
1	Date	Open	High	Low	Close	Volum	Adj Clo	Retur		Intervals
2	09/11/2012	1377.55	1391.39	1373.03	1379.85	3.65E+09	1379.85	0.00167		-0.02
3	08/11/2012	1394.53	1401.23	1377.51	1377.51	3.78E+09	1377.51	-0.0122		-0.015
4	07/11/2012	1428.27	1428.27	1388.14	1394.53	4.36E+09	1394.53	-0.02362		-0.01

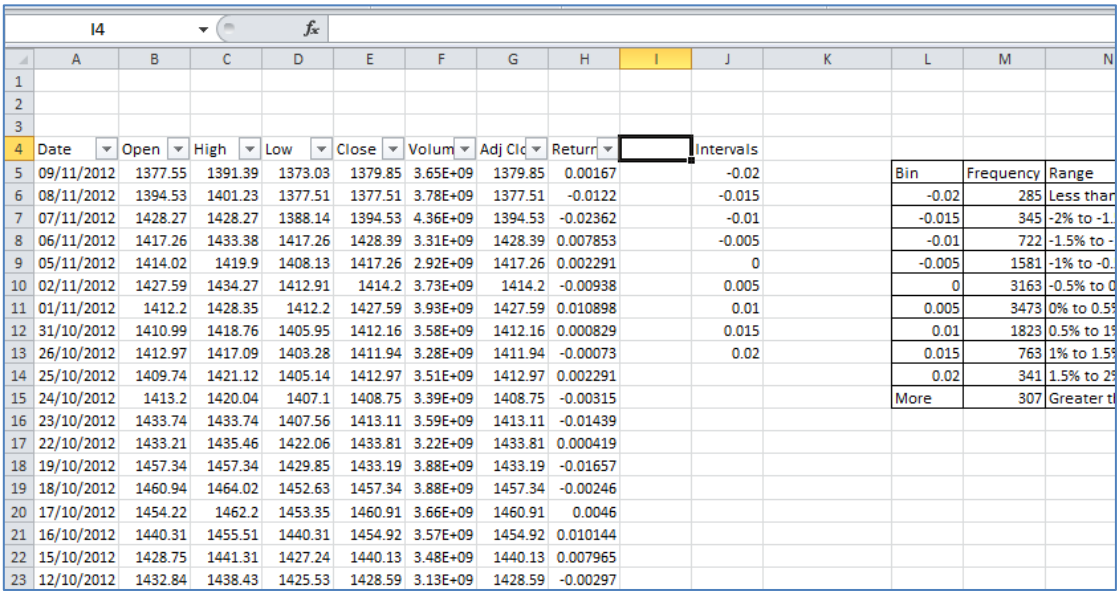
Clicking on one of these arrows will open a drop-down menu that will allow us to filter the data in that column.

Before we continue in this section, we are going to edit our spreadsheet further by adding some blanks rows above our data. The reason for this is so that when we conduct filter-analysis (which collapses rows temporarily) there will be a section of the spreadsheet that is unaffected.

- *Right-click on the row1 box and choose add row.*
- *Repeat this twice so there are three additional rows at the top of the spreadsheet.*



	A	B	C	D	E	F	G	H	I	J
1			ligh	Low	Close	Volum	Adj Clc	Returr		Intervals
2			1391.39	1373.03	1379.85	3.65E+09	1379.85	0.00167		-0.02
3			1401.23	1377.51	1377.51	3.78E+09	1377.51	-0.0122		-0.015
4			1428.27	1388.14	1394.53	4.36E+09	1394.53	-0.02362		-0.01
5			1433.38	1417.26	1428.39	3.31E+09	1428.39	0.007853		-0.005
6			1419.9	1408.13	1417.26	2.92E+09	1417.26	0.002291		0
7			1434.27	1412.91	1414.2	3.73E+09	1414.2	-0.00938		0.005
8			1428.35	1412.2	1427.59	3.93E+09	1427.59	0.010898		0.01
9			1418.76	1405.95	1412.16	3.58E+09	1412.16	0.000829		0.015
10			1417.09	1403.28	1411.94	3.28E+09	1411.94	-0.00073		0.02
11			1421.12	1405.14	1412.97	3.51E+09	1412.97	0.002291		
12			1420.04	1407.1	1408.75	3.39E+09	1408.75	-0.00315		
13			1433.74	1407.56	1413.11	3.59E+09	1413.11	-0.01439		
14			1435.46	1422.06	1433.81	3.22E+09	1433.81	0.000419		
15			1457.34	1429.85	1433.19	3.88E+09	1433.19	-0.01657		
16			1464.02	1452.63	1457.34	3.88E+09	1457.34	-0.00246		
17			1462.2	1453.35	1460.91	3.66E+09	1460.91	0.0046		



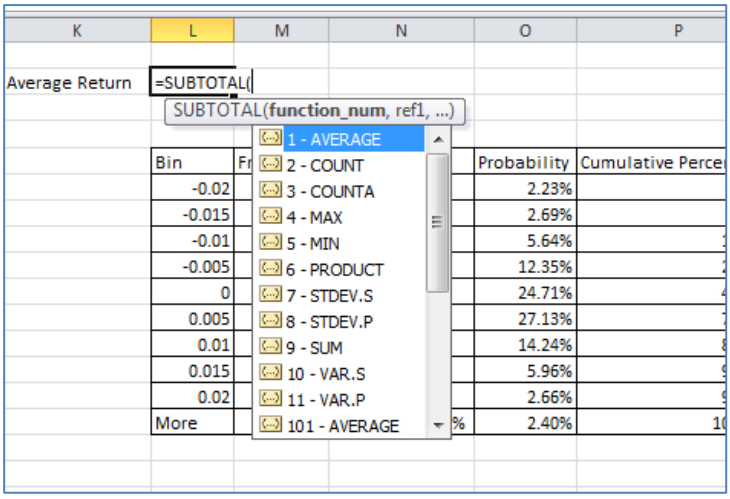
	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1														
2														
3														
4	Date	Open	High	Low	Close	Volum	Adj Clc	Returr		Intervals				
5	09/11/2012	1377.55	1391.39	1373.03	1379.85	3.65E+09	1379.85	0.00167		-0.02		Bin	Frequency	Range
6	08/11/2012	1394.53	1401.23	1377.51	1377.51	3.78E+09	1377.51	-0.0122		-0.015		-0.02	285	Less thar
7	07/11/2012	1428.27	1428.27	1388.14	1394.53	4.36E+09	1394.53	-0.02362		-0.01		-0.015	345	-2% to -1.
8	06/11/2012	1417.26	1433.38	1417.26	1428.39	3.31E+09	1428.39	0.007853		-0.005		-0.01	722	-1.5% to -
9	05/11/2012	1414.02	1419.9	1408.13	1417.26	2.92E+09	1417.26	0.002291		0		-0.005	1581	-1% to -0.
10	02/11/2012	1427.59	1434.27	1412.91	1414.2	3.73E+09	1414.2	-0.00938		0.005		0	3163	-0.5% to 0
11	01/11/2012	1412.2	1428.35	1412.2	1427.59	3.93E+09	1427.59	0.010898		0.01		0.005	3473	0% to 0.5
12	31/10/2012	1410.99	1418.76	1405.95	1412.16	3.58E+09	1412.16	0.000829		0.015		0.01	1823	0.5% to 1
13	26/10/2012	1412.97	1417.09	1403.28	1411.94	3.28E+09	1411.94	-0.00073		0.02		0.015	763	1% to 1.5
14	25/10/2012	1409.74	1421.12	1405.14	1412.97	3.51E+09	1412.97	0.002291				0.02	341	1.5% to 2
15	24/10/2012	1413.2	1420.04	1407.1	1408.75	3.39E+09	1408.75	-0.00315				More	307	Greater t
16	23/10/2012	1433.74	1433.74	1407.56	1413.11	3.59E+09	1413.11	-0.01439						
17	22/10/2012	1433.21	1435.46	1422.06	1433.81	3.22E+09	1433.81	0.000419						
18	19/10/2012	1457.34	1457.34	1429.85	1433.19	3.88E+09	1433.19	-0.01657						
19	18/10/2012	1460.94	1464.02	1452.63	1457.34	3.88E+09	1457.34	-0.00246						
20	17/10/2012	1454.22	1462.2	1453.35	1460.91	3.66E+09	1460.91	0.0046						
21	16/10/2012	1440.31	1455.51	1440.31	1454.92	3.57E+09	1454.92	0.010144						
22	15/10/2012	1428.75	1441.31	1427.24	1440.13	3.48E+09	1440.13	0.007965						
23	12/10/2012	1432.84	1438.43	1425.53	1428.59	3.13E+09	1428.59	-0.00297						

We are going to construct a cell which displays the average of all the data in the returns column, filtered or unfiltered. This will allow flexibility in changing the sample of returns we are looking at and instantaneously give us an average value of those returns.

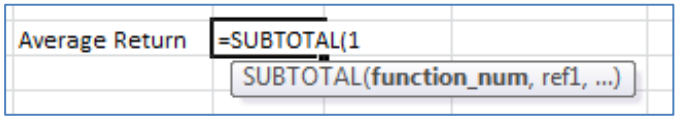
- *Select cell K2 and type "Average Return". Adjust the column width to fit in all the characters as in previous steps.*

G	H	I	J	K	L	M	N	O
				Average Return				
Adj Clo	Return		Intervals					
1379.85	0.00167		-0.02		Bin	Frequency	Range	Probability
1377.51	-0.0122		-0.015		-0.02	285	Less than -2%	2.23%
1394.53	-0.02362		-0.01		-0.015	345	-2% to -1.5%	2.69%
1428.39	0.007853		-0.005		-0.01	722	-1.5% to -1%	5.64%
1417.26	0.002291		0		-0.005	1581	-1% to -0.5%	12.35%
1414.2	-0.00938		0.005		0	3163	-0.5% to 0%	24.71%
1427.59	0.010898		0.01		0.005	3473	0% to 0.5%	27.13%
1412.16	0.000829		0.015		0.01	1823	0.5% to 1%	14.24%
1411.94	-0.00073		0.02		0.015	763	1% to 1.5%	5.96%
1412.97	0.002291				0.02	341	1.5% to 2%	2.66%
1408.75	-0.00315				More	307	Greater than 2%	2.40%
1413.11	-0.01439							
1433.81	0.000419							

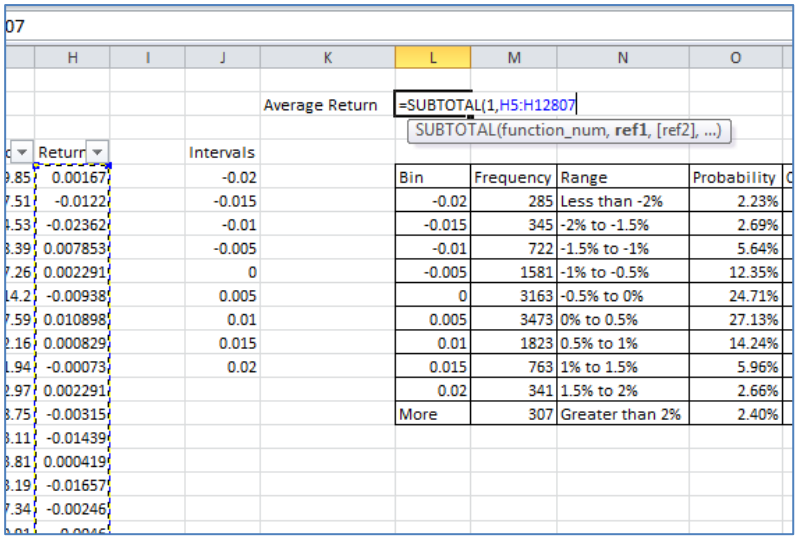
- *Select cell L2 and type "=SUBTOTAL("*



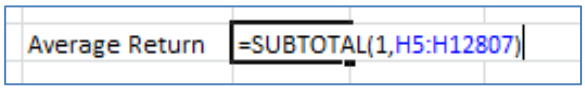
- *Double-click "1 – Average" from the drop-down menu that appears. Alternatively type "1".*

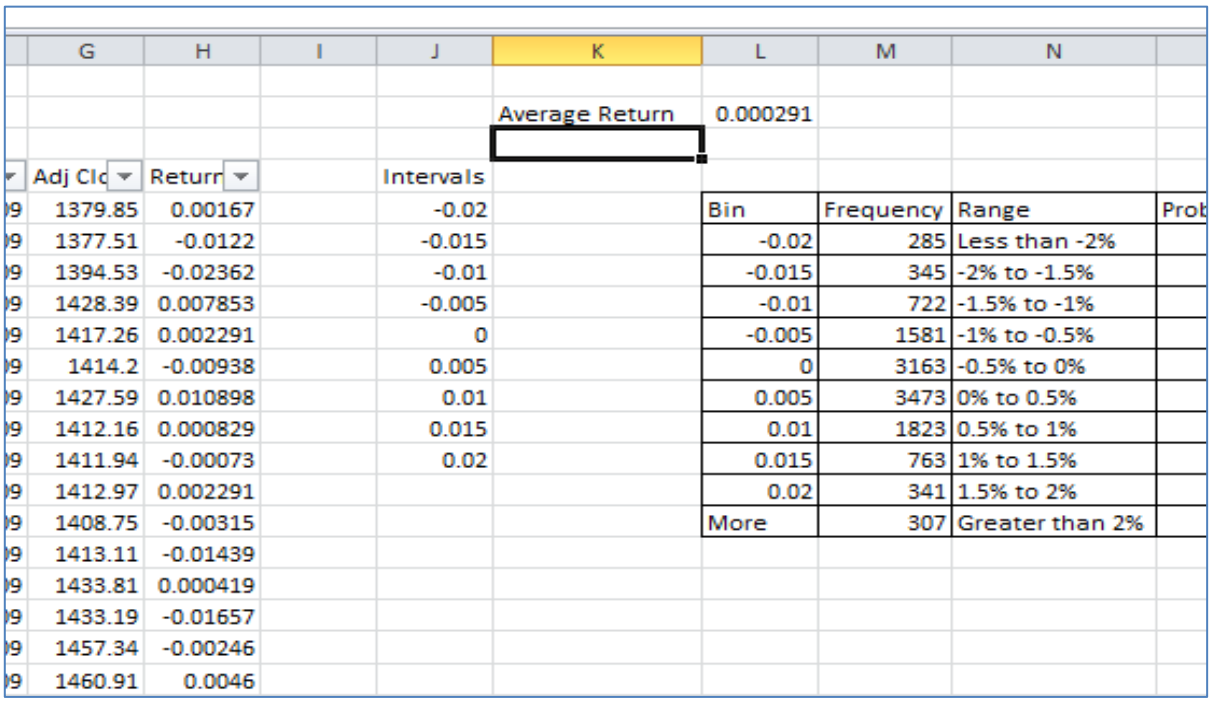


- *Next, add a comma to the equation. Type ",".*
- *Select cells H5 to the last returns cell in our filtered Returns column. Again, use CTRL+SHIFT+DOWN. This will shift your view to the bottom of the spreadsheet. Navigate back up to the top using the scrollbar on the right-hand side.*



- *Close the bracket to finish the equation and press Enter.*

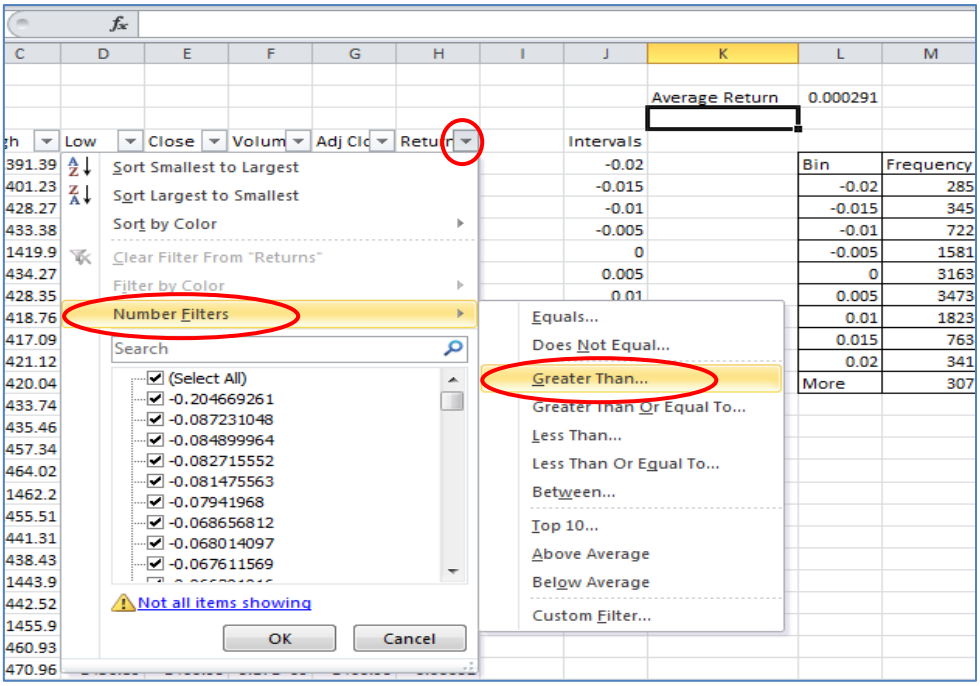


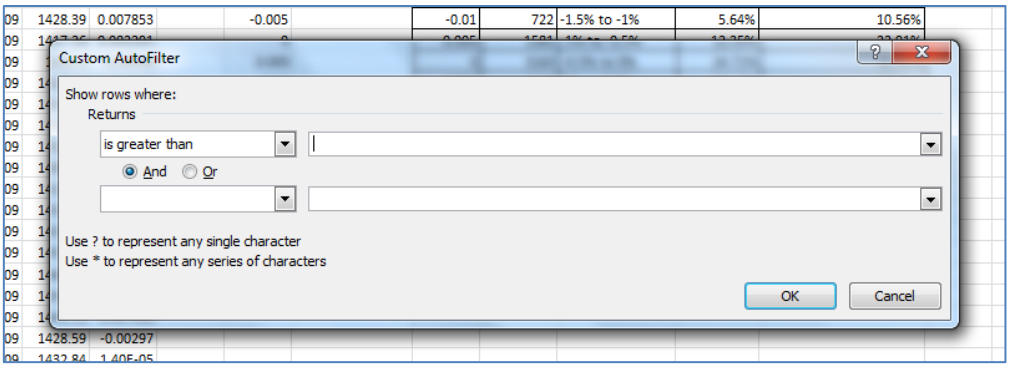


This gives an average return of 0.0291% using the current unfiltered data. This is the same value as calculated by the descriptive statistics for the overall mean. We can now filter the data so that the value in cell L2 changes depending on the data that is filtered out.

The first average return we will calculate is the average positive return. We need to filter our all returns that are negative or equal to 0% so that only positive returns are listed in the column. To do this:

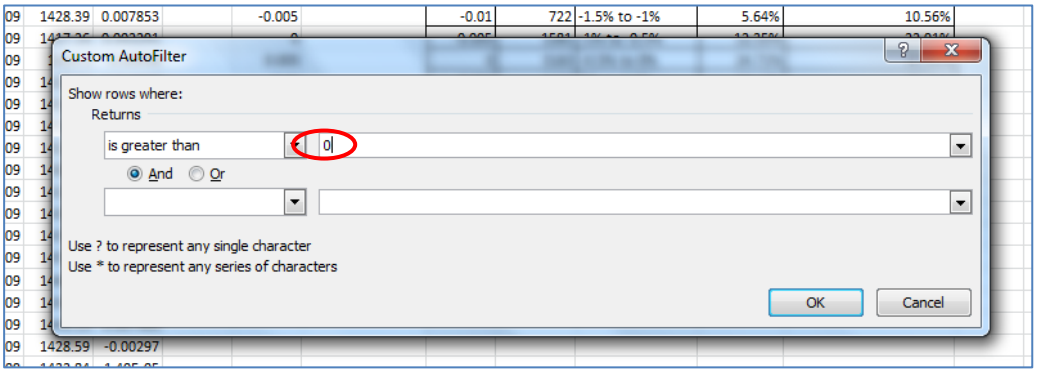
- *Click the filter arrow of the returns column and navigate to 'Number Filters'→ 'Greater Than' in the menu. Click the 'Greater Than' option to set up the filter.*

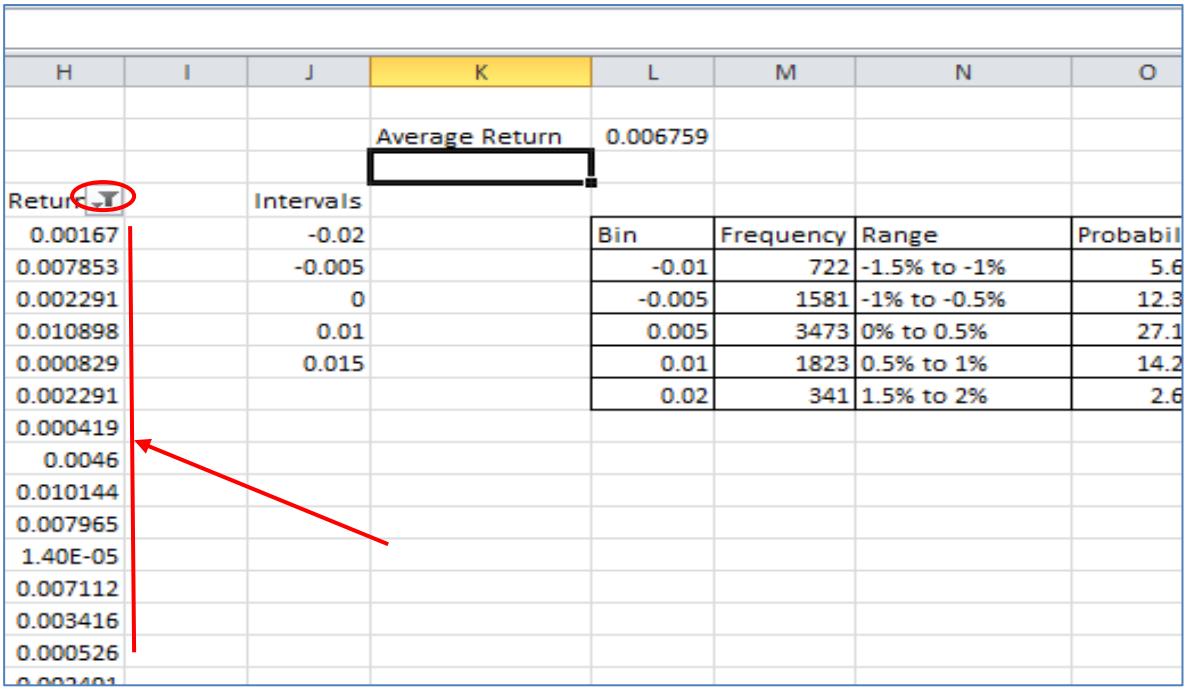




To filter out all non-positive returns, we need to set the filter to only keep values that are greater than 0.

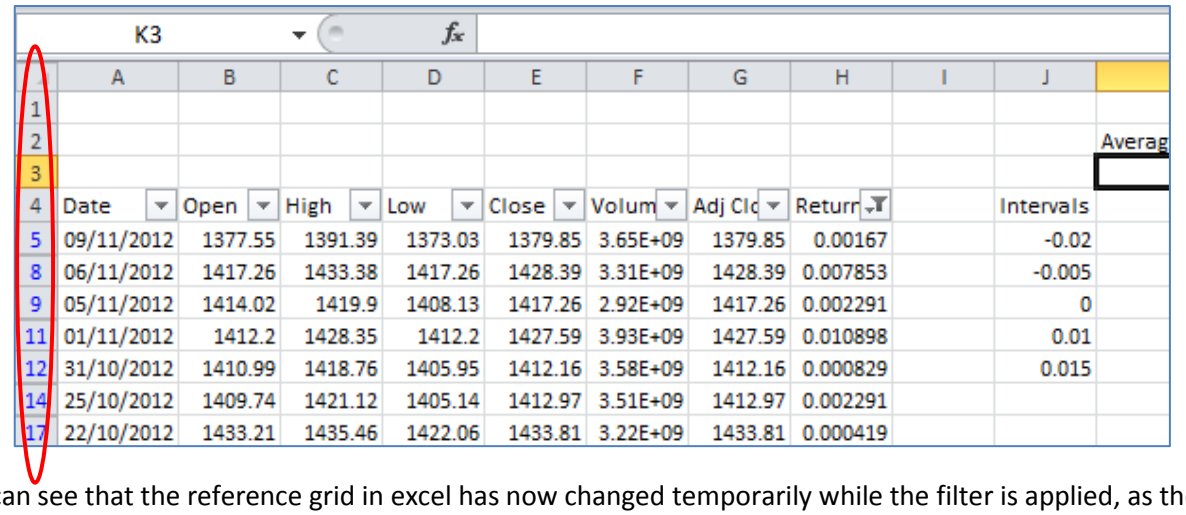
- *Enter "0" into the top right-hand box, opposite 'is greater than'. Press OK.*





H	I	J	K	L	M	N	O
			Average Return	0.006759			
Retur		Intervals					
0.00167		-0.02		Bin	Frequency	Range	Probabil
0.007853		-0.005		-0.01	722	-1.5% to -1%	5.6
0.002291		0		-0.005	1581	-1% to -0.5%	12.3
0.010898		0.01		0.005	3473	0% to 0.5%	27.1
0.000829		0.015		0.01	1823	0.5% to 1%	14.2
0.002291				0.02	341	1.5% to 2%	2.6
0.000419							
0.0046							
0.010144							
0.007965							
1.40E-05							
0.007112							
0.003416							
0.000526							

Note that the entire returns column has changed to positive values. Excel has filtered out the rows with negative or 0% daily returns and has applied it to the whole spreadsheet, this is why other data appears lost (temporarily) such as the frequency distribution. Do not worry, the filter can be undone and reapplied at any time. Also note that the drop-down menu symbol has changed to include a small filter symbol to show that the filter is currently active.

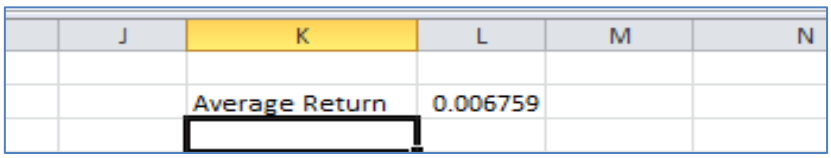


K3

	A	B	C	D	E	F	G	H	I	J
1										
2										Averag
3										
4	Date	Open	High	Low	Close	Volum	Adj Clo	Return		Intervals
5	09/11/2012	1377.55	1391.39	1373.03	1379.85	3.65E+09	1379.85	0.00167		-0.02
8	06/11/2012	1417.26	1433.38	1417.26	1428.39	3.31E+09	1428.39	0.007853		-0.005
9	05/11/2012	1414.02	1419.9	1408.13	1417.26	2.92E+09	1417.26	0.002291		0
11	01/11/2012	1412.2	1428.35	1412.2	1427.59	3.93E+09	1427.59	0.010898		0.01
12	31/10/2012	1410.99	1418.76	1405.95	1412.16	3.58E+09	1412.16	0.000829		0.015
14	25/10/2012	1409.74	1421.12	1405.14	1412.97	3.51E+09	1412.97	0.002291		
17	22/10/2012	1433.21	1435.46	1422.06	1433.81	3.22E+09	1433.81	0.000419		

You can see that the reference grid in excel has now changed temporarily while the filter is applied, as there are missing rows as shown in the picture above.

We can see that the 'Average Return' value has changed and now represents the average positive return for the data.

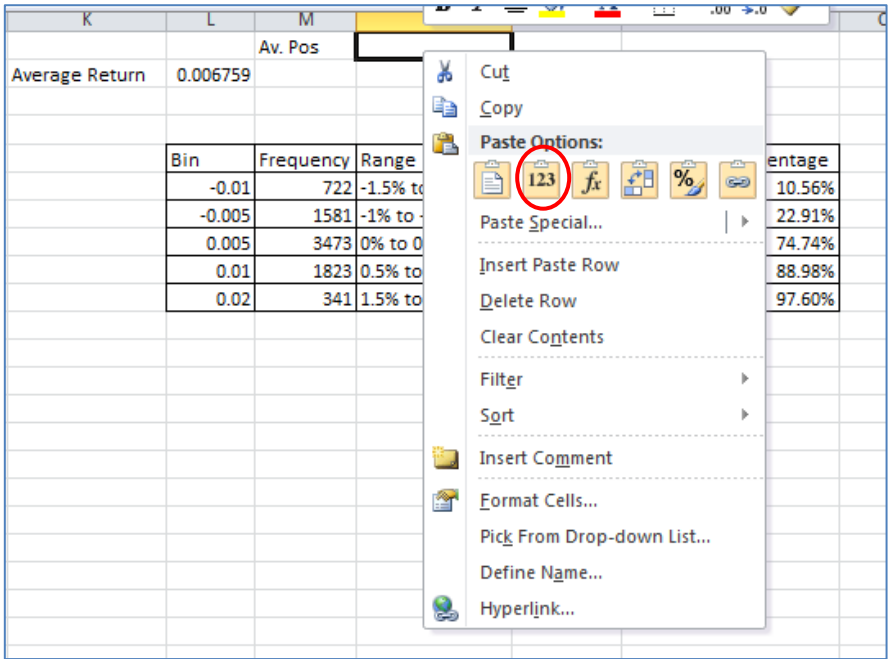


J	K	L	M	N
	Average Return	0.006759		

As this cell value changes depending on our variable, it is worth recording this number in a separate fixed cell.

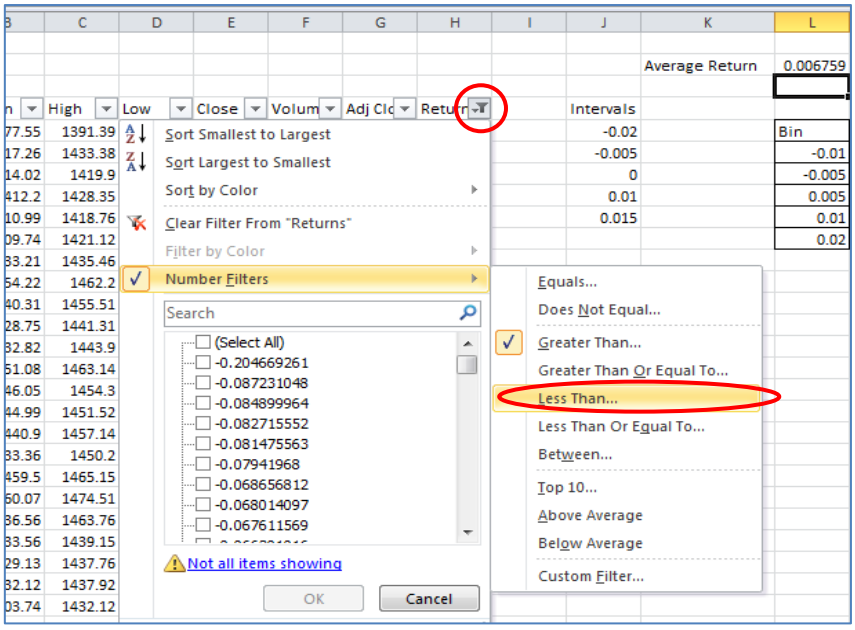
- *Select cell M1 and type "Av. Pos"*

- *Select cell L2 and press CTRL+C to copy the cell.*
- *Right-click on cell N1 and click on Paste Value in the "Paste Options" section.*

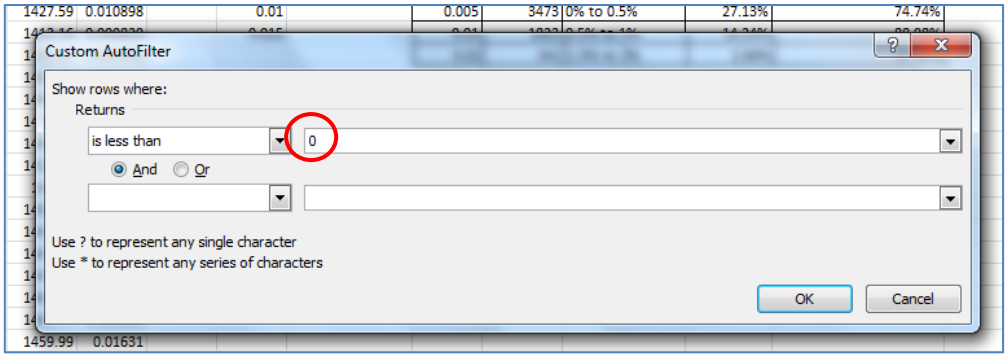


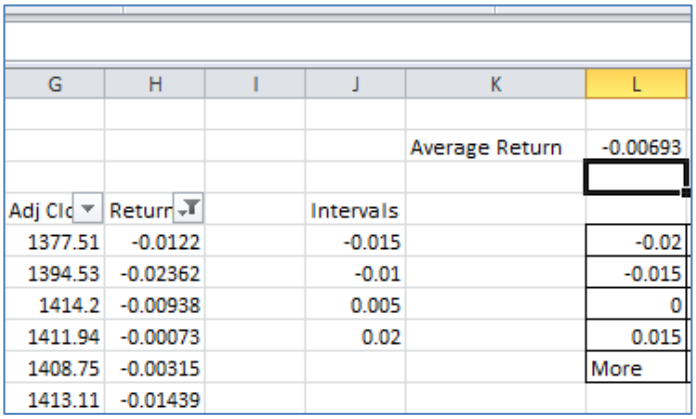
With the average positive return now calculated, we can change the filter to consider only the negative returns and thus find the average negative return.

- *Change the filter from 'Greater Than' to 'Less Than'.*



- *Set the 'Less Than' filter to 0 and press OK.*





G	H	I	J	K	L
				Average Return	-0.00693
Adj Clc	Returr		Intervals		
1377.51	-0.0122		-0.015		-0.02
1394.53	-0.02362		-0.01		-0.015
1414.2	-0.00938		0.005		0
1411.94	-0.00073		0.02		0.015
1408.75	-0.00315				More
1413.11	-0.01439				

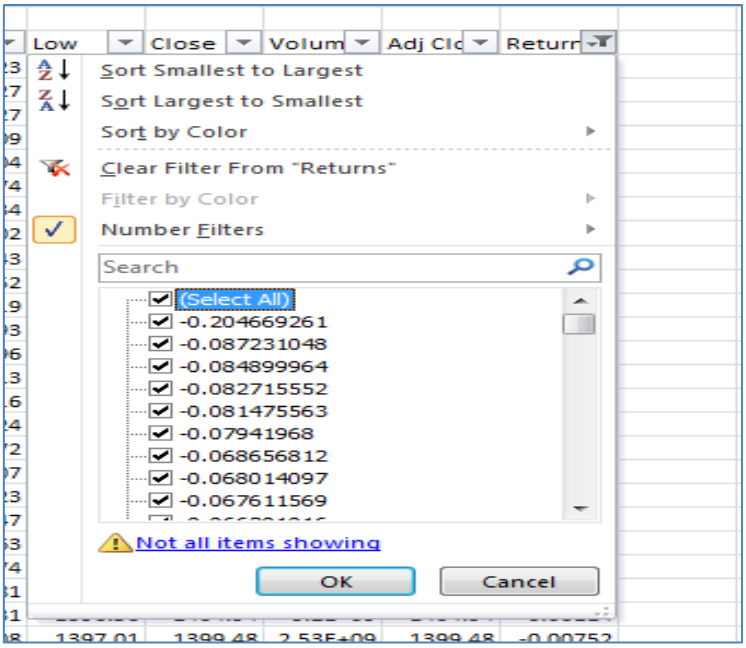
All the values in the returns column have changed to negatives. The variable average return cell (L2) now gives us the average negative return for the S&P 500 in our historical dataset. This value should be recorded in the same way as the average positive return.

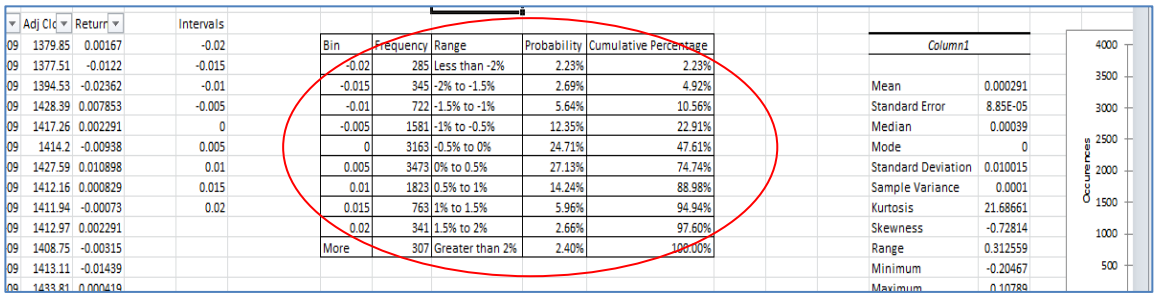
- *Select cell M2 and type "Av. Neg". Press Enter.*
- *Select cell L2 and press CTRL+C to copy the cell.*
- *Paste the cell as a value, as shown previously, into cell N2.*

H	I	J	K	L	M	N	O	P
					Av. Pos	0.00675916		
			Average Return	-0.00693	Av. Neg	-0.00692779		
Returr		Intervals						
-0.0122		-0.015		-0.02	285	Less than -2%	2.23%	2.23%
-0.02362		-0.01		-0.015	345	-2% to -1.5%	2.69%	4.92%
-0.00938		0.005		0	3163	-0.5% to 0%	24.71%	47.61%
-0.00073		0.02		0.015	763	1% to 1.5%	5.96%	94.94%
-0.00315				More	307	Greater than 2%	2.40%	100.00%
-0.01439								
-0.01657								
-0.00246								

We now have the average Up/Down returns for our historical S&P 500 dataset. The approach used here allows you to conduct your own analysis using average returns for any configurable filter to the dataset. To get the spreadsheet back to normal, with no rows collapsed, change the filter by doing the following:

- *Select the returns filter and check the box called select all. Press OK.*





Adj Clc	Returr	Intervals
1379.85	0.00167	-0.02
1377.51	-0.0122	-0.015
1394.53	-0.02362	-0.01
1428.39	0.007853	-0.005
1417.26	0.002291	0
1414.2	-0.00938	0.005
1427.59	0.010898	0.01
1412.16	0.000829	0.015
1411.94	-0.00073	0.02
1412.97	0.002291	
1408.75	-0.00315	
1413.11	-0.01439	

Bin	Frequency	Range	Probability	Cumulative Percentage
-0.02	285	Less than -2%	2.23%	2.23%
-0.015	345	-2% to -1.5%	2.69%	4.92%
-0.01	722	-1.5% to -1%	5.64%	10.56%
-0.005	1581	-1% to -0.5%	12.35%	22.91%
0	3163	-0.5% to 0%	24.71%	47.61%
0.005	3473	0% to 0.5%	27.13%	74.74%
0.01	1823	0.5% to 1%	14.24%	88.98%
0.015	763	1% to 1.5%	5.96%	94.94%
0.02	341	1.5% to 2%	2.66%	97.60%
More	307	Greater than 2%	2.40%	100.00%

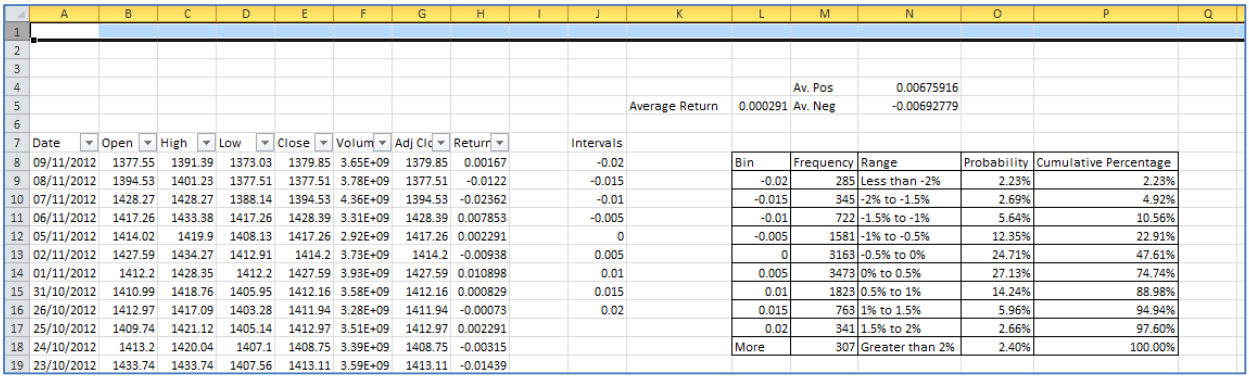
Column1	
Mean	0.000291
Standard Error	8.85E-05
Median	0.00039
Mode	0
Standard Deviation	0.010015
Sample Variance	0.0001
Kurtosis	21.68661
Skewness	-0.72814
Range	0.312559
Minimum	-0.20467
Maximum	0.10789

The rows should now be fully expanded and back to normal – shown by a full frequency table.

Before we continue, we are going to add some cells that tell us how many positive, negative and flat trading days there are in our dataset. To do that, we need to add a count cell that will sum the number of days in the returns column as it is filtered.

We start by adding more space on the spreadsheet.

- *Add another three rows to the top of the spreadsheet as before. Right-click on row 1 and click insert. Repeat three times.*

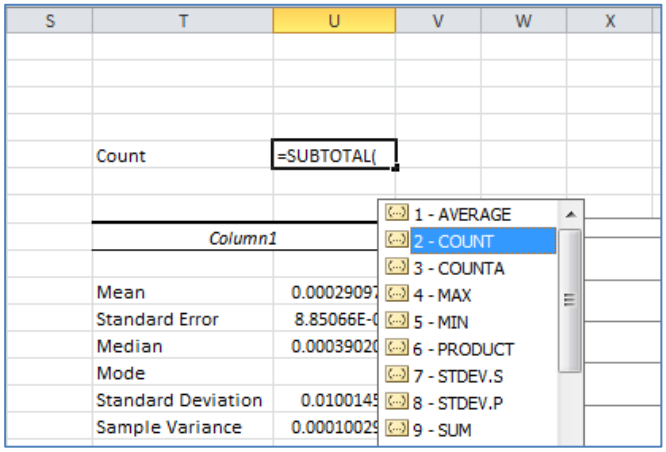


													Av. Pos	0.00675916		
										Average Return	0.000291		Av. Neg	-0.00692779		
Date	Open	High	Low	Close	Volum	Adj Clc	Returr		Intervals							
09/11/2012	1377.55	1391.39	1373.03	1379.85	3.65E+09	1379.85	0.00167		-0.02		Bin	Frequency	Range	Probability	Cumulative Percentage	
08/11/2012	1394.53	1401.23	1377.51	1377.51	3.78E+09	1377.51	-0.0122		-0.015		-0.02	285	Less than -2%	2.23%	2.23%	
07/11/2012	1428.27	1428.27	1388.14	1394.53	4.36E+09	1394.53	-0.02362		-0.01		-0.015	345	-2% to -1.5%	2.69%	4.92%	
06/11/2012	1417.26	1433.38	1417.26	1428.39	3.31E+09	1428.39	0.007853		-0.005		-0.01	722	-1.5% to -1%	5.64%	10.56%	
05/11/2012	1414.02	1419.9	1408.13	1417.26	2.92E+09	1417.26	0.002291		0		-0.005	1581	-1% to -0.5%	12.35%	22.91%	
02/11/2012	1427.59	1434.27	1412.91	1414.2	3.73E+09	1414.2	-0.00938		0.005		0	3163	-0.5% to 0%	24.71%	47.61%	
01/11/2012	1412.2	1428.35	1412.2	1427.59	3.93E+09	1427.59	0.010898		0.01		0.005	3473	0% to 0.5%	27.13%	74.74%	
31/10/2012	1410.99	1418.76	1405.95	1412.16	3.58E+09	1412.16	0.000829		0.015		0.01	1823	0.5% to 1%	14.24%	88.98%	
26/10/2012	1412.97	1417.09	1403.28	1411.94	3.28E+09	1411.94	-0.00073		0.02		0.015	763	1% to 1.5%	5.96%	94.94%	
25/10/2012	1409.74	1421.12	1405.14	1412.97	3.51E+09	1412.97	0.002291				0.02	341	1.5% to 2%	2.66%	97.60%	
24/10/2012	1413.2	1420.04	1407.1	1408.75	3.39E+09	1408.75	-0.00315				More	307	Greater than 2%	2.40%	100.00%	
23/10/2012	1433.74	1433.74	1407.56	1413.11	3.59E+09	1413.11	-0.01439									

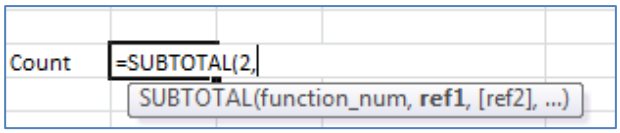
In order to count the number of occurrences within different standard deviations from the mean we must first set up another variable cell that will count the number of values in the returns column when different filters are applied. The method will be very similar to last time:

- *Select cell T5 and type "Count". Press Enter.*

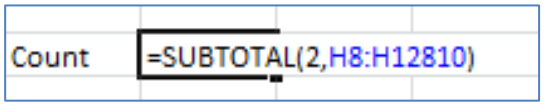
- *Select cell U5 and type "=SUBTOTAL("*

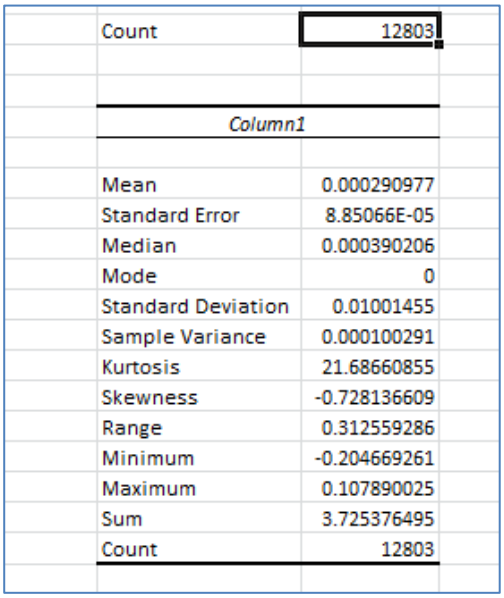


- *Double-click "2 – Count" and the type ",".*



- *Select all the values in the returns column by highlighting cell H8 and pressing CTRL+SHIFT+DOWN.*
- *Scroll back up to the top of the spreadsheet with the scrollbar on the right-hand side. Now finish the equation with a closed bracket and press Enter.*





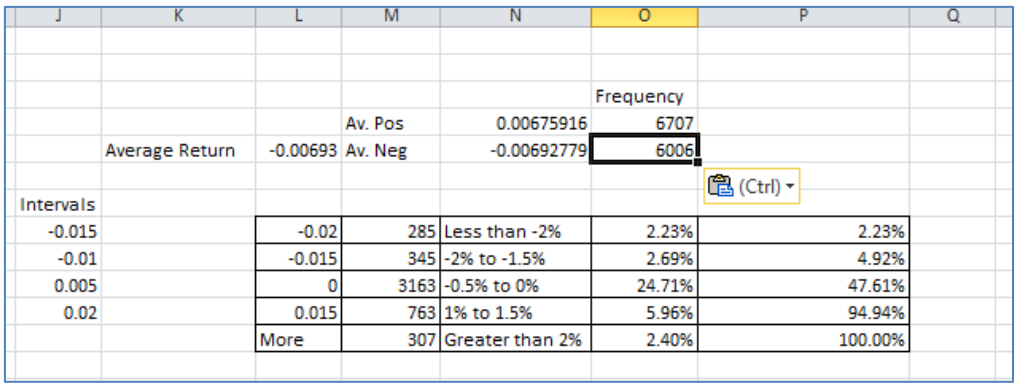
Count	12803
Column1	
Mean	0.000290977
Standard Error	8.85066E-05
Median	0.000390206
Mode	0
Standard Deviation	0.01001455
Sample Variance	0.000100291
Kurtosis	21.68660855
Skewness	-0.728136609
Range	0.312559286
Minimum	-0.204669261
Maximum	0.107890025
Sum	3.725376495
Count	12803

The count shows there are 12,803 trading days in the whole dataset which can be confirmed by comparing this number to the preliminary descriptive statistics.

We will now enter count data for positive, negative and flat trading days next to our average returns data.

- *In cell O3 type "Frequency". Press Enter.*
- *Filter the returns column to show positive data (greater than 0) and copy the count value from cell U5 and paste it as a value into O4.*

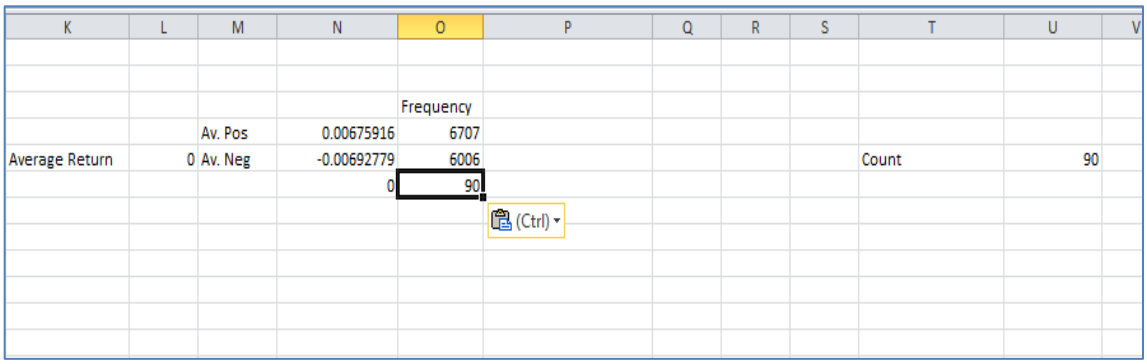
- *Similarly, follow on from this by filtering the returns column to show negative returns data only (Less than 0) and copy the count value from cell U5 and paste it into cell O5 as a value.*



					Frequency		
			Av. Pos	0.00675916	6707		
	Average Return	-0.00693	Av. Neg	-0.00692779	6006		
						(Ctrl)	
Intervals							
-0.015		-0.02	285	Less than -2%	2.23%	2.23%	
-0.01		-0.015	345	-2% to -1.5%	2.69%	4.92%	
0.005		0	3163	-0.5% to 0%	24.71%	47.61%	
0.02		0.015	763	1% to 1.5%	5.96%	94.94%	
		More	307	Greater than 2%	2.40%	100.00%	

We need to account for trading days that yielded 0 returns so that we can account for the whole dataset.

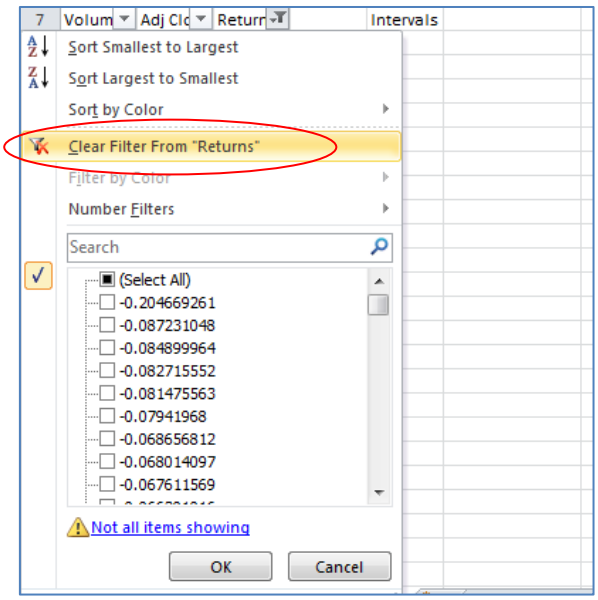
- *In cell N6 type "0". Press Enter.*
- *Next, filter the returns column again. This time create a filter that is 'equal to' 0. Copy the count number in cell U5 and paste it as a value into cell O6.*



				Frequency							
		Av. Pos	0.00675916	6707							
Average Return	0	Av. Neg	-0.00692779	6006					Count	90	
			0	90							
					(Ctrl)						

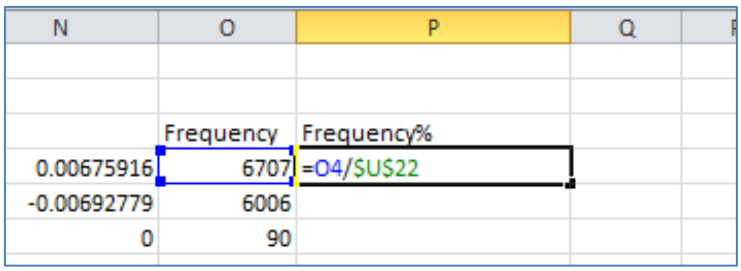
Now we know the number of positive, negative and flat trading days within the period we are analysing. Be sure to clear the returns filter before continuing.

- *Click on the returns filter and select "Clear Filter From Returns".*



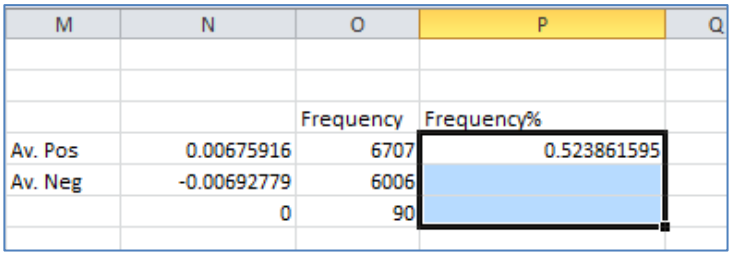
In order to understand these count values more easily, we will display them in percentage terms of the total trading day count in our entire dataset.

- *In cell P3 type "Frequency%". Press Enter.*
- *In cell P4 type "O4/U2". Before pressing enter, press F4 on your keyboard while the cursor is next to U2 within the equation you have just written. This should add dollar signs around U2 which will fix this cell reference when we copy this formula down or across. Press Enter.*

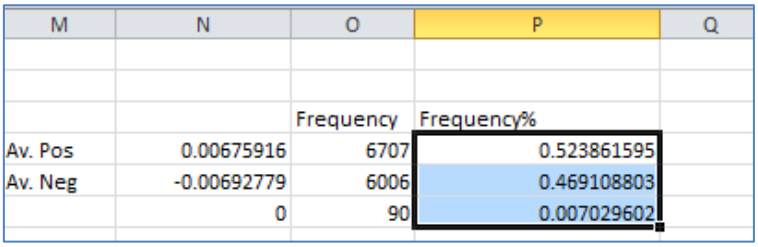


N	O	P	Q
	Frequency	Frequency%	
0.00675916	6707	=O4/$U$22	
-0.00692779	6006		
0	90		

- *Select cells P4 to P6 and press CTRL+D to copy the formula down.*

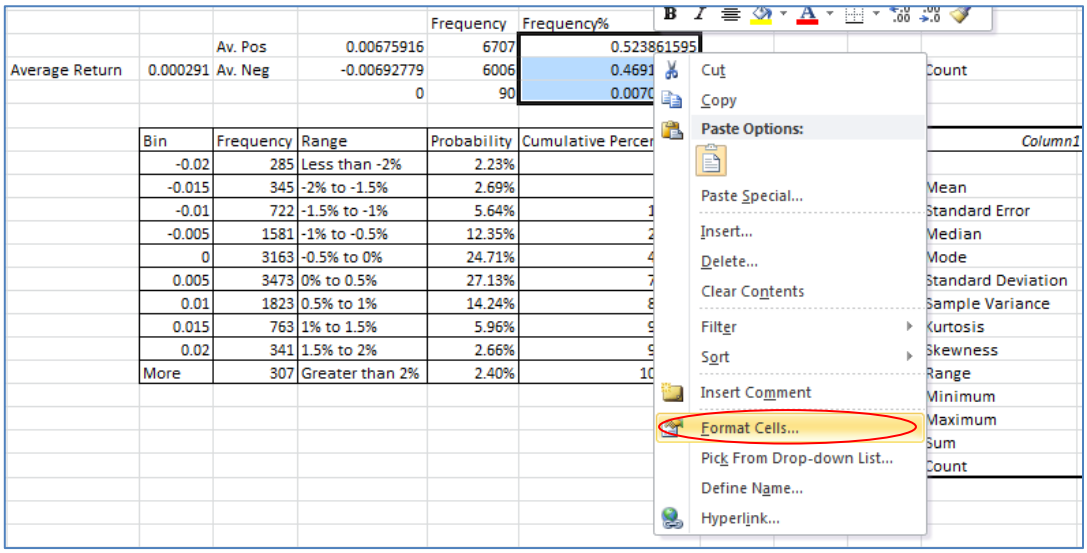


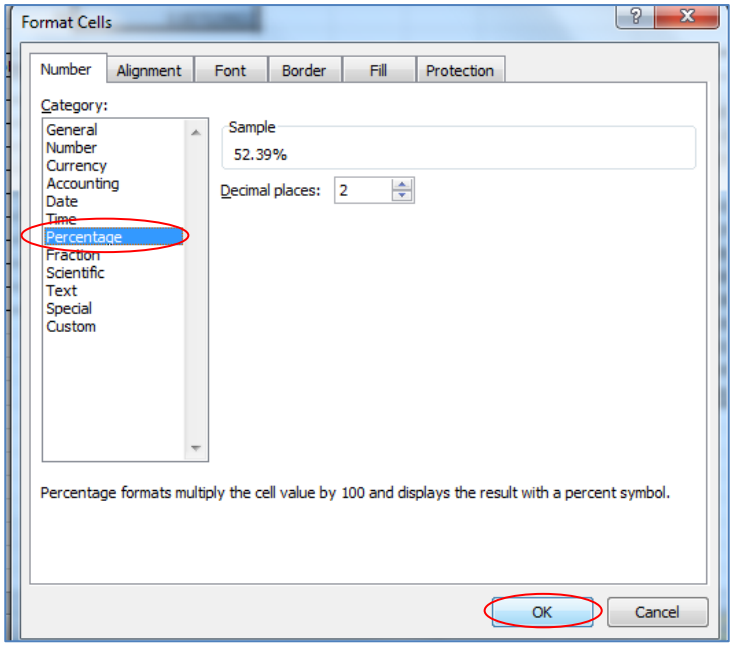
M	N	O	P	Q
		Frequency	Frequency%	
Av. Pos	0.00675916	6707	0.523861595	
Av. Neg	-0.00692779	6006		
	0	90		



M	N	O	P	Q
		Frequency	Frequency%	
Av. Pos	0.00675916	6707	0.523861595	
Av. Neg	-0.00692779	6006	0.469108803	
	0	90	0.007029602	

- *With cells P4 to P6 still selected, right-click within the selection and click on "Format Cells" and change the 'Category' (cell format) to "Percentage". Press OK*





	K	L	M	N	O	P	Q
					Frequency	Frequency%	
			Av. Pos	0.00675916	6707	52.39%	
	Average Return	0.000291	Av. Neg	-0.00692779	6006	46.91%	
				0	90	0.70%	

Bin	Frequency	Range	Probability	Cumulative Percentage
-0.02	285	Less than -2%	2.23%	2.23%
-0.015	345	-2% to -1.5%	2.69%	4.92%
-0.01	722	-1.5% to -1%	5.64%	10.56%
-0.005	1581	-1% to -0.5%	12.35%	22.91%
0	3163	-0.5% to 0%	24.71%	47.61%
0.005	3473	0% to 0.5%	27.13%	74.74%
0.01	1823	0.5% to 1%	14.24%	88.98%
0.015	763	1% to 1.5%	5.96%	94.94%
0.02	341	1.5% to 2%	2.66%	97.60%
More	307	Greater than 2%	2.40%	100.00%

Finally, we find the average daily positive return (out of all trading days in the dataset) should we only go long the S&P500 on positive days, and not participate on other trading days. This reflects the average positive return weighted by the frequency of positive trading days. We conduct the same method for the average daily negative returns weighted by the frequency of negative trading days.

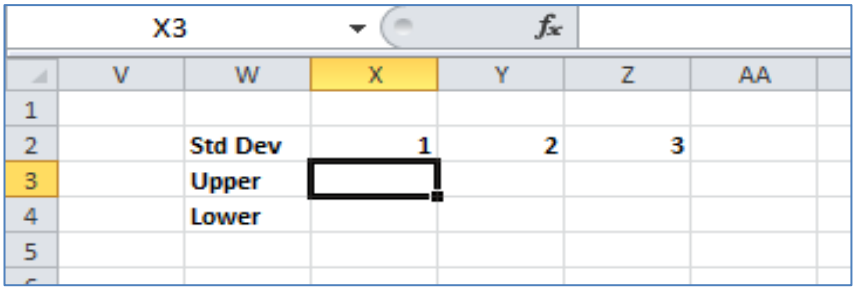
- *In cell Q3 type "Av Return". Press Enter.*
- *In cell Q4 type "N4*P4". Press Enter.*
- *In cell Q5 type "N5*P5". Press Enter.*
- *Select cells Q4 and Q5 and format them to display their values as percentages.*

K	L	M	N	O	P	Q	R	S
				Frequency	Frequency%	Av Return		
		Av. Pos	0.00675916	6707	52.39%	0.35%		
Average Return	0.000291	Av. Neg	-0.00692779	6006	46.91%	-0.32%		
			0	90	0.70%			

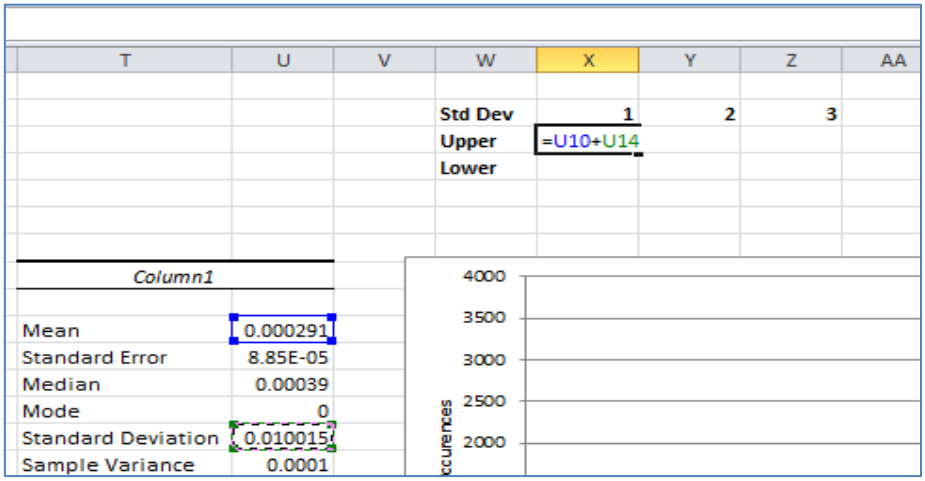
The final data analysis this tutorial will show is a standard deviation occurrences table. This will compare the actual occurrences within 1,2 and 3 standard deviations from the mean against the number of occurrences predicted by a normal distribution of the same data.

To analyse the number of trading days with returns within 1,2 and 3 standard deviations from the mean, a simple table needs to be constructed that calculates the upper and lower bound of 1,2 and 3 standard deviations from the mean. The upper bound returns of 1 standard deviation from the mean will be the mean value plus one standard deviation. The lower bound of 1 standard deviation will be the mean value minus one standard deviation. Continuing this, the upper bound return of 2 standard deviations from the mean will be the mean value plus two standard deviations. You see how this is going…

- *Type "Std Dev" into cell W2.*
- *Type "Upper" into cell W3.*
- *Type "Lower" into cell W4.*
- *Type "1" into cell X2.*
- *Type "2" into cell Y2.*
- *Type "3" into cell Z2.*
- *Highlight all these cell and press CTRL+B to put the text into BOLD.*

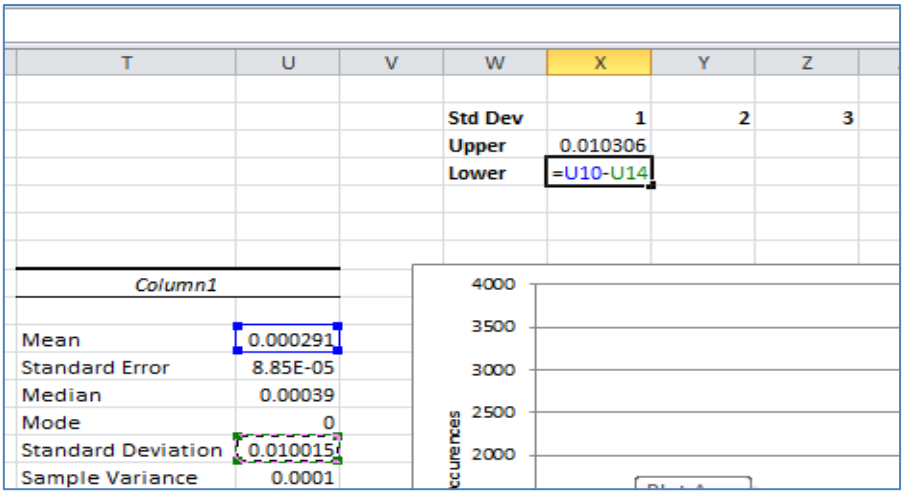


- *Type "=U10+U14" into cell X3. Press Enter.*



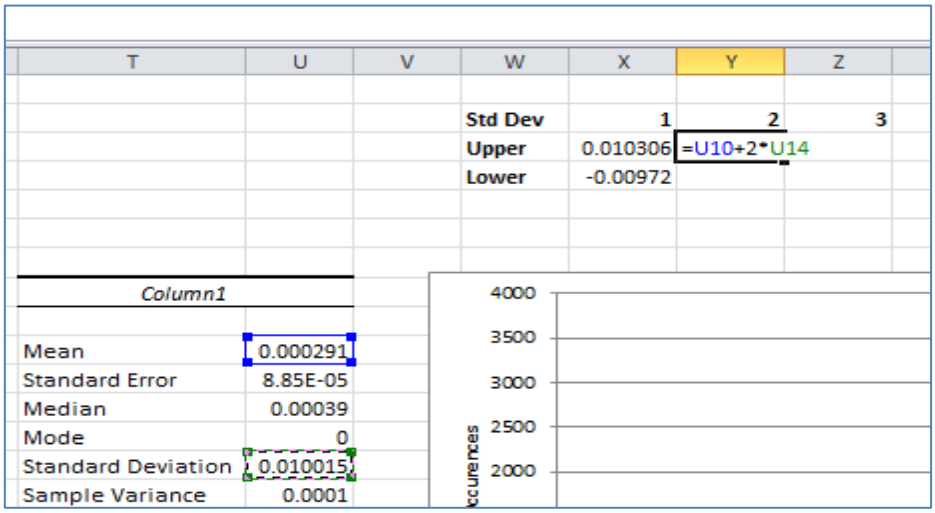
Mean plus 1 standard deviation = upper bound of 1 standard deviation from the mean.

- *Type "=U10-U14" into cell X4. Press Enter.*



Mean minus 1 standard deviation = lower bound of 1 standard deviation from the mean.

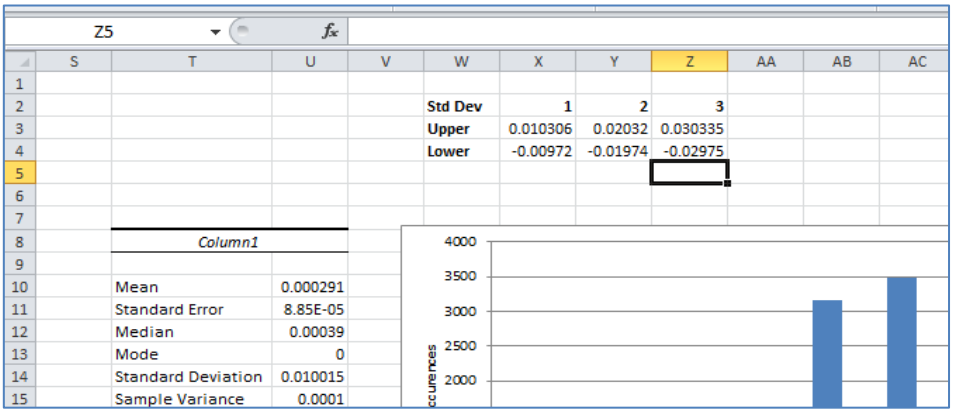
- *Type "=U10+2*U14" into cell Y3. Press Enter.*



Mean plus 2 standard deviations = upper bound of 2 standard deviations from the mean.

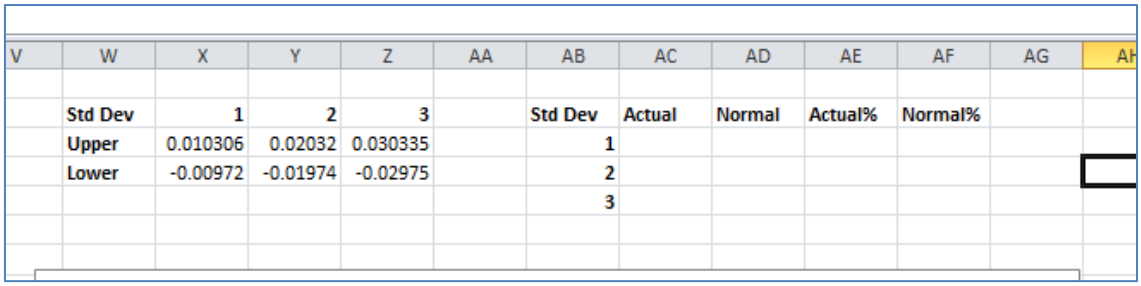
Carry on this method to fill out the rest of the table.

- *Type "=U10-2*U14" into cell Y4. Press Enter.*
- *Type "=U10+3*U14" into cell Z3. Press Enter.*
- *Type "=U10+3*U14" into cell Z4. Press Enter.*



Now we are ready to create our standard deviations occurrences table. This will show the percentage of trading days where the returns lie within 1,2 and 3 standard deviations from the mean and compare them to those predicted by normal distribution.

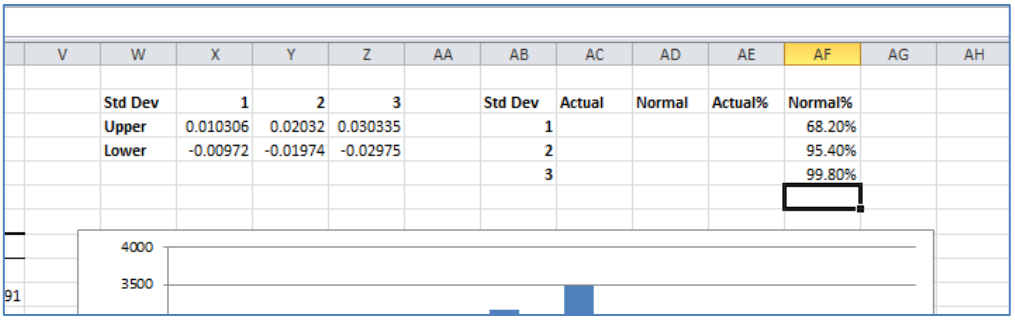
- *Type "Std Dev" into cell AB2. Press Enter.*
- *Type "Actual" into cell AC2. Press Enter.*
- *Type "Normal" into cell AD2. Press Enter.*
- *Type "Actual%" into cell AE2. Press Enter.*
- *Type "Normal%" into cell AF2. Press Enter.*
- *Type "1" into cell AB3. Press Enter.*
- *Type "2" into cell AB4. Press Enter.*
- *Type "3" into cell AB5. Press Enter.*



V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH
	Std Dev	1	2	3		Std Dev	Actual	Normal	Actual%	Normal%		
	Upper	0.010306	0.02032	0.030335		1						
	Lower	-0.00972	-0.01974	-0.02975		2						
						3						

The returns filter can be used again to find the number of occurrences between the upper and lower bounds of the "standard deviation from the mean" measures. This will show the real (or actual) count. The normal distribution percentages for 1,2 ad 3 standard deviations from the mean have been outlined earlier in this tutorial. They can be used to find the predicted number of occurrences in the dataset that fall within 1,2 and 3 standard deviations if the distribution was normal, by multiplying their percentages by the overall count of all the trading days in the dataset (12,803).

- *Type "68.2%" into cell AF3.*
- *Type "95.4%" into cell AF4.*
- *Type "99.8%" into cell AF5.*



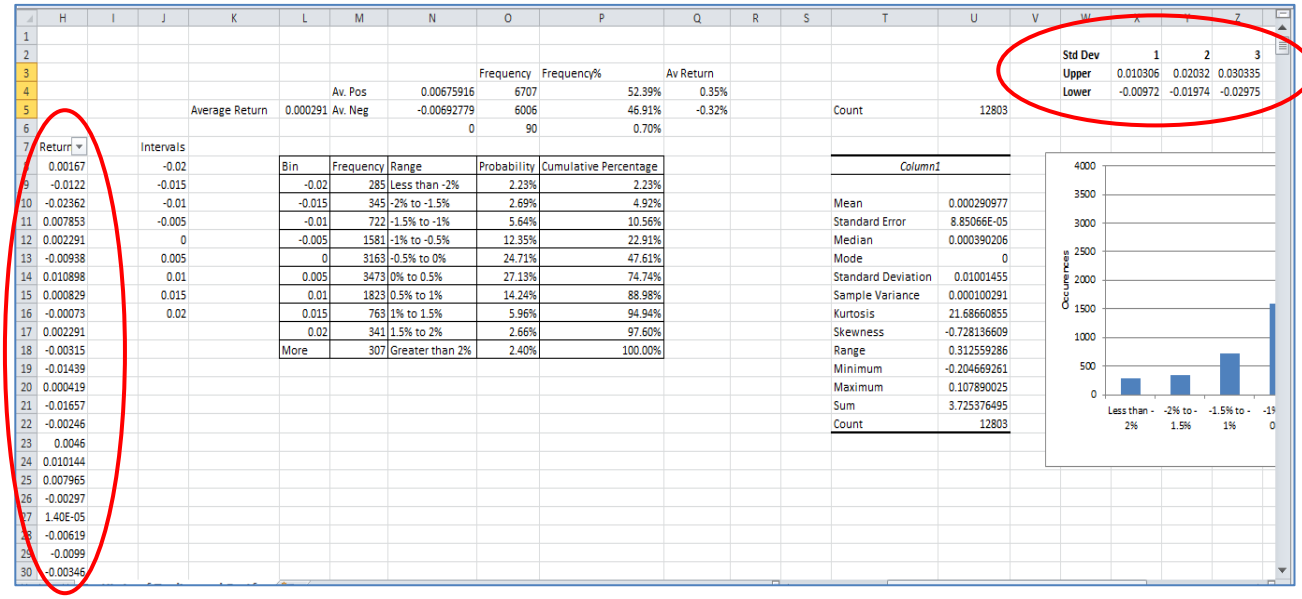
V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH
	Std Dev	1	2	3		Std Dev	Actual	Normal	Actual%	Normal%		
	Upper	0.010306	0.02032	0.030335		1				68.20%		
	Lower	-0.00972	-0.01974	-0.02975		2				95.40%		
						3				99.80%		

- *Type "=AF3*U22" into cell AD3.*
- *Type "=AF4*U22" into cell AD4.*
- *Type "=AF5*U22" into cell AD5.*

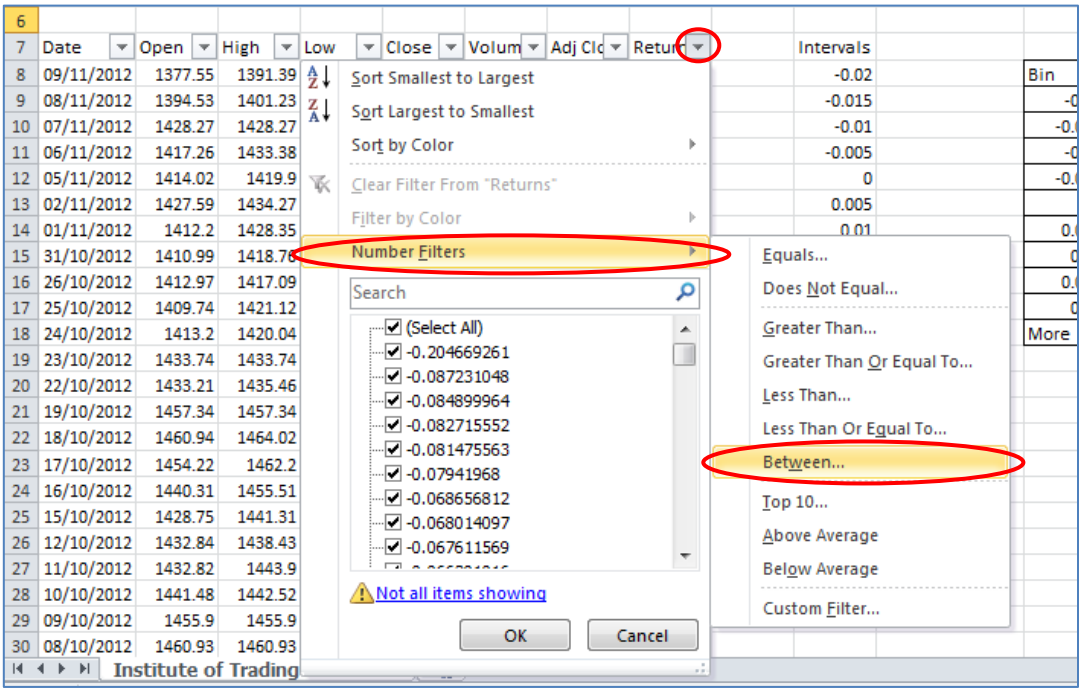
	AB	AC	AD	AE	AF	AG
	Std Dev	Actual	Normal	Actual%	Normal%	
	1		8731.646		68.20%	
	2		12214.06		95.40%	
	3		12777.39		99.80%	

The predicted occurrences (estimated by a normal distribution) are shown. Now we will calculate the actual values and see how they compare.

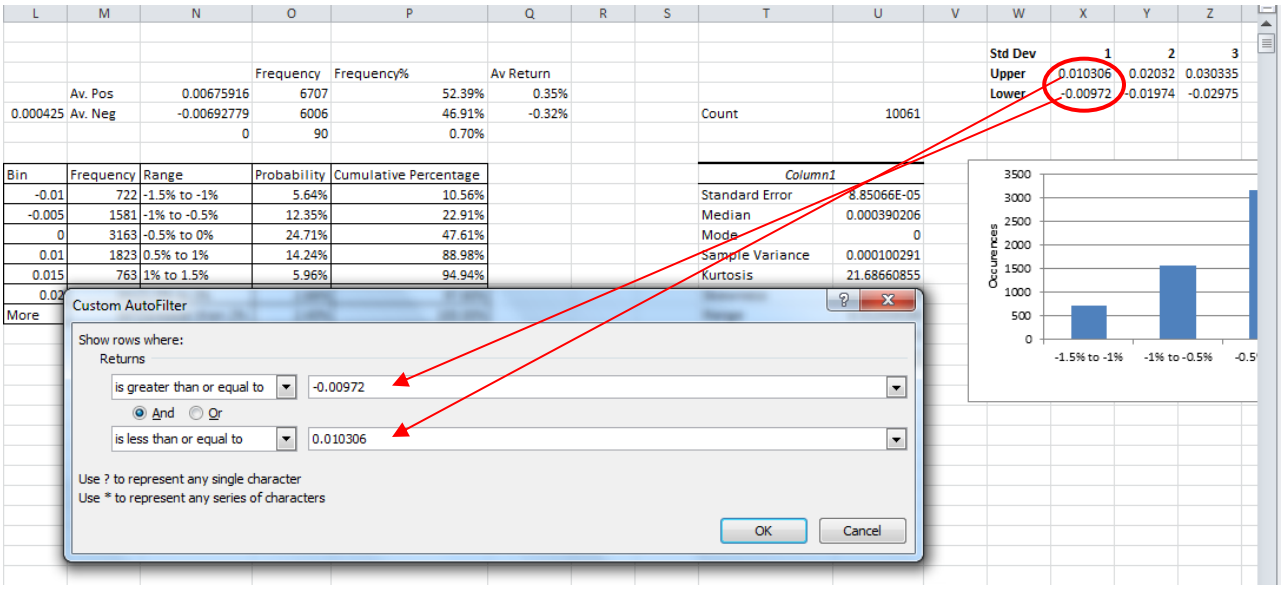
- *Position the view of the spreadsheet so you can see both the returns column AND the standard deviation lower and upper bounds. See below.*



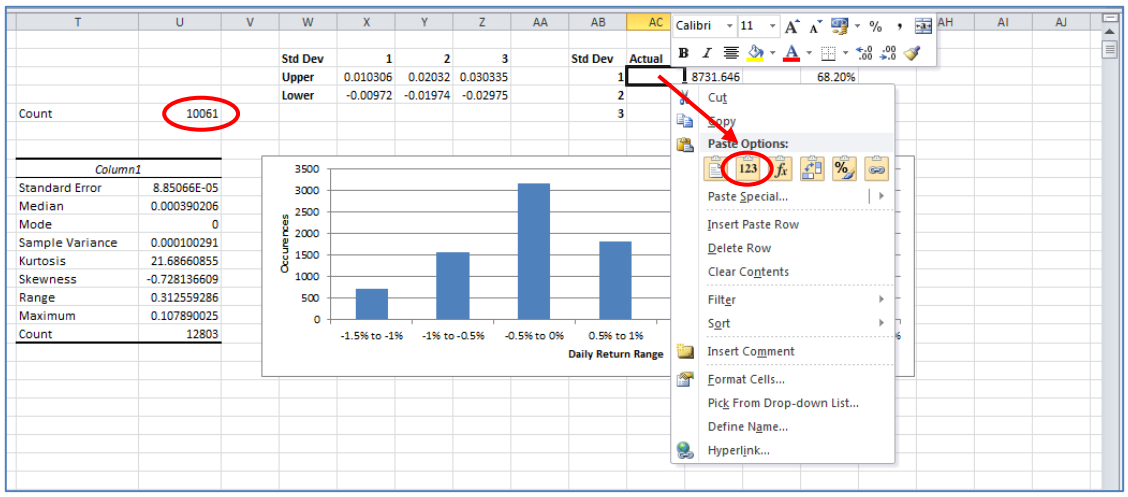
- *Select the returns filter and choose 'Number Filters'→'Between'.*



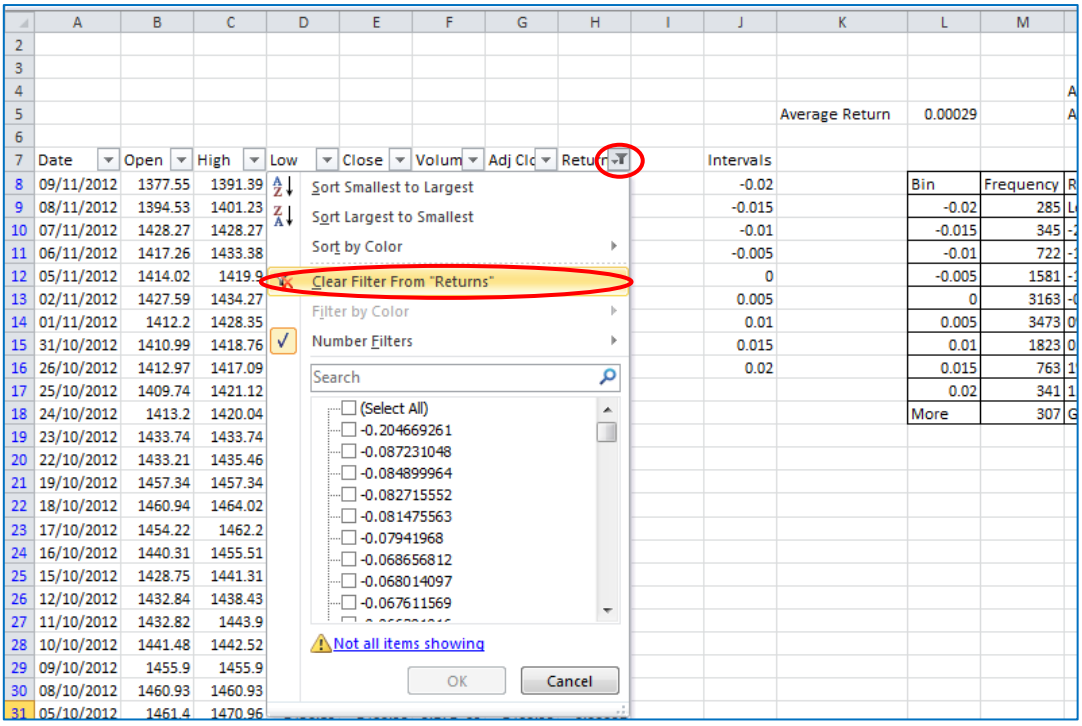
- *Manually type in the lower bound of 1 standard deviation from the mean into the 'Greater than or equal to' box and the upper bound into the 'Less than or equal to box. Press OK. See below.*

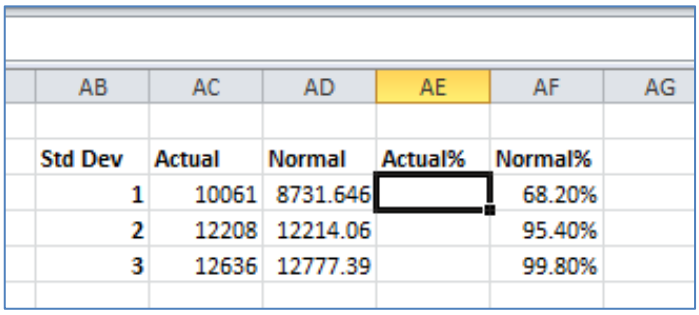


- *View the count number (cell U5) and copy it with CTRL+C, then paste it as a value into cell AC3 (the actual occurrences within 1 standard deviation from the mean). Paste the value by right-clicking and choosing paste as value under paste options. See below.*



- *Repeat the process above; remember to start with a view which shows the returns column and the upper and lower bounds so that you can easily type them in.*
- *When finished with this procedure, remove the filter from the returns column by selecting the "Clear Filter From Returns" option in the Returns Filter drop-down menu.*





Std Dev	Actual	Normal	Actual%	Normal%
1	10061	8731.646		68.20%
2	12208	12214.06		95.40%
3	12636	12777.39		99.80%

Finally, to get the actual percent of occurrences that lie within these ranges we take the actual count for each range and divide it by the total count for the whole dataset.

- *Type "=AC3/U22" in AE3.*
- *Type "=AC4/U22" in AE4.*
- *Type "=AC5/U22" in AE5.*
- *Format cells AE3 to AE5 by clicking and dragging over them, right-click 'Format Cells' → 'Percentage'.*

Std Dev	Actual	Normal	Actual%	Normal%
1	10061	8731.646	78.58%	68.20%
2	12208	12214.06	95.35%	95.40%
3	12636	12777.39	98.70%	99.80%

The simple conclusion from this table is that normal distribution predicts more occurrences than were actually counted within 2 and 3 standard deviations. Turning that on its head; more actual occurrences are unaccounted for by normal distribution outside 3 standard deviations. Since one standard deviation is roughly equal to 1% in this dataset, it means that returns of more than 3% or less than -3% occur more frequently than normal distribution predicts. This empirical analysis just backs up the basic descriptive statistics we described earlier.

4.0 Summary

In this document we have discussed how to construct a distribution of returns for the S&P 500 in Excel. The distribution allows us to interpret probabilities of a range of returns occurring. Descriptive statistics were discussed with some key interpretations for your future analysis. Expected returns and average positive and negative returns were calculated as well as comparisons drawn between a standard normal distribution and that of the S&P 500. Key conclusions were drawn about playing the S&P 500 daily and risks such as extreme price movements were highlighted. In fact, for this last point, commission was not even taken into account which makes it even more likely that there is no money to gain by day trading in this manner. The material covered in this tutorial will allow you to draw some key conclusions and interpret basic statistical data, while translating them into underlying probabilities of gaining certain returns. Methods put across were intended to provide some flexibility should you want to explore further analysis, with the data analysis and filter tools in Excel being particularly useful. It is worth experimenting with Excel to draw more personalised conclusions about your data and this tutorial should have provided a basis for you to accomplish that.