Excel
Options for Smoothing Data Series

Why smoothing

A data series, particularly long series, can display an erratic behavior up and down from which it is hard to see a clear pattern. Is there a trend, up or down? Are there regular cycles? **Smoothing a series is particularly useful for line plots of NLP measures that present a great deal of variability, such as the results of sentiment analysis, sentence complexity, verb voices and more) plotted by sentence index.**

The Excel Analysis ToolPak add-ins

In order to perform data series smoothing in Excel you will need to add an add-in first.

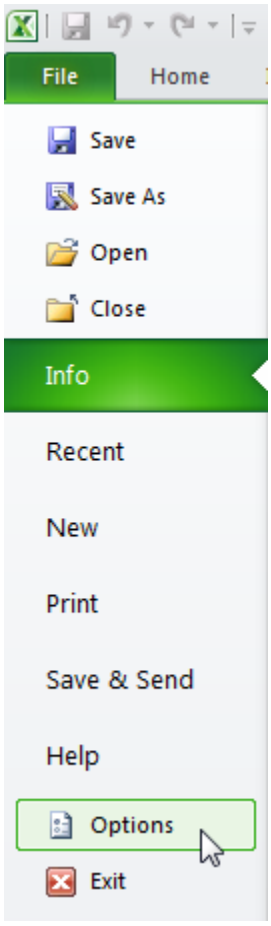
Add the Analysis ToolPak add-in to Excel

The Analysis ToolPak is an Excel add-in program that provides data analysis tools for financial, statistical and engineering data analysis.

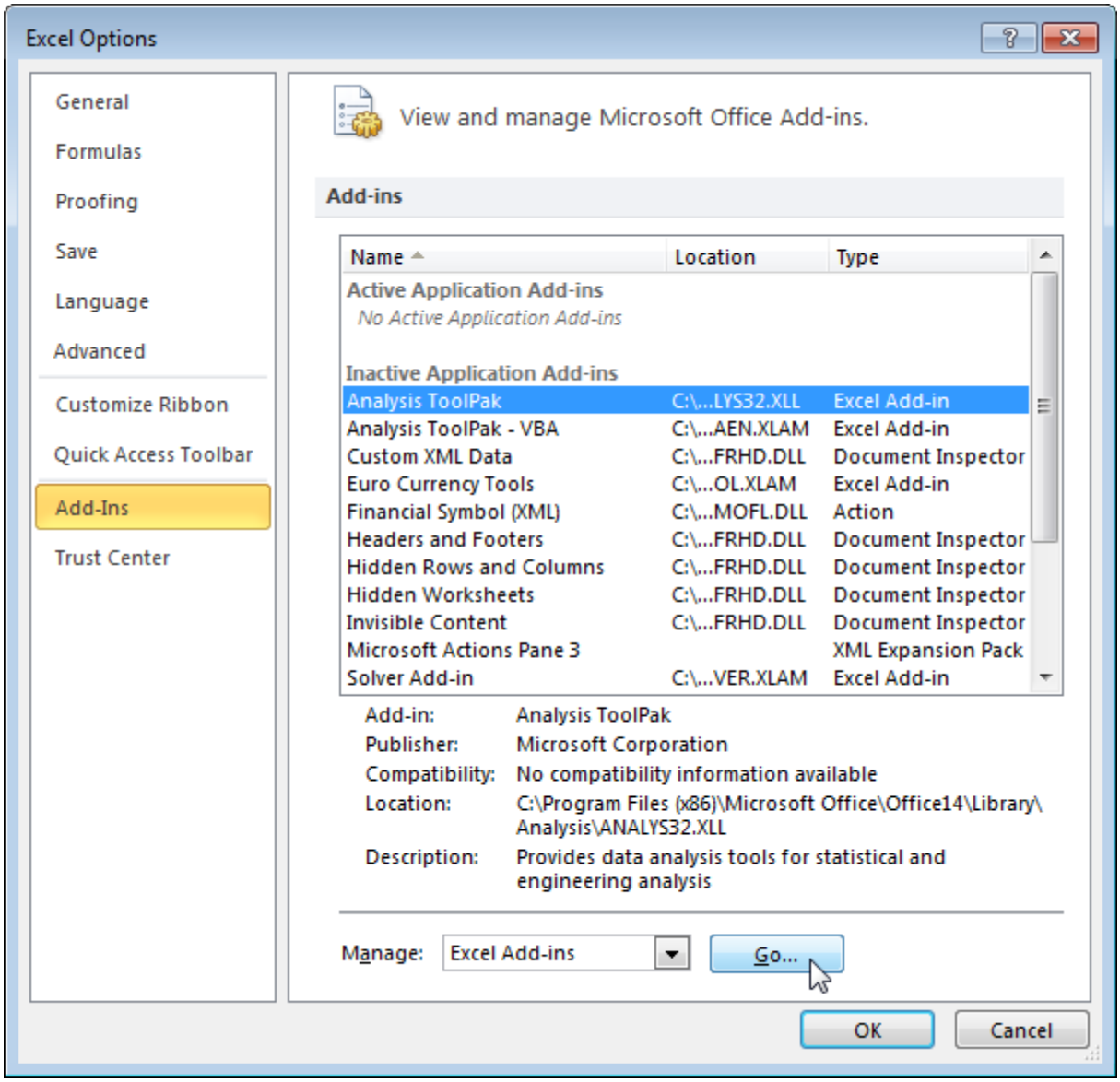
To load the Analysis ToolPak add-in, execute the following steps.

1. Click on the green File tab. The File tab in Excel 2010 or later replaces the Office Button (or File Menu) in previous versions of Excel.

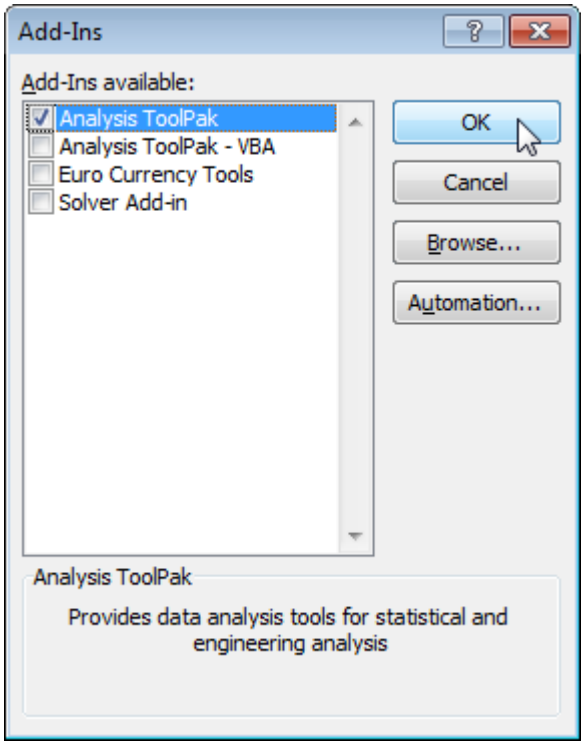
2. Click on Options.



3. Under Add-ins, select Analysis ToolPak and click on the **Go button**.



4. Check Analysis ToolPak and click on **OK button**.

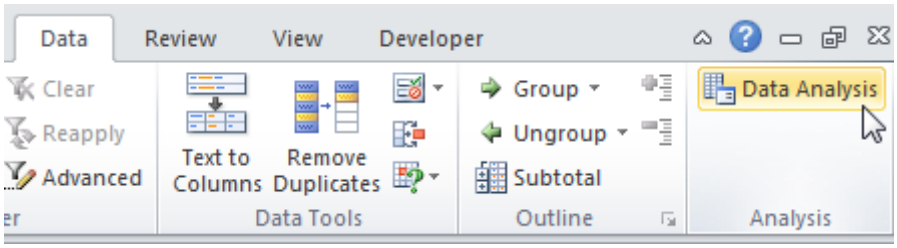


Two smoothing options in Excel

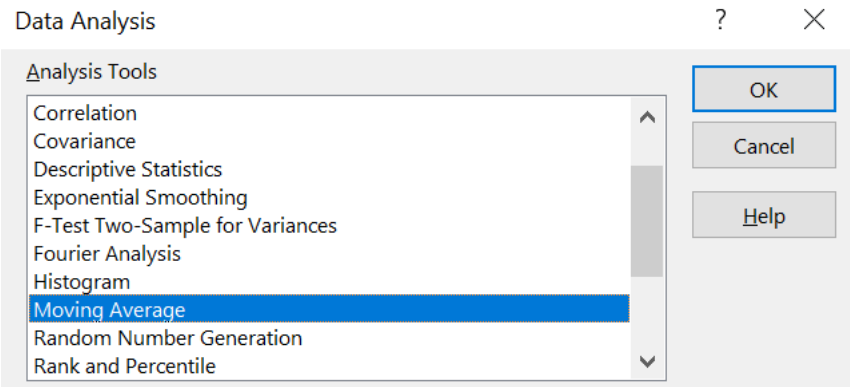
Excel offers two options for "smoothing" a series (i.e., removing the fine-grained variation between successive steps in a line plot, where these steps can be measured by time or sentence index):

1. Exponential smoothing
2. Moving average

Once you have installed the Analysis ToolPak, on the main page, click on Data; the Data Analysis tab will appear. Click on Data Analysis.



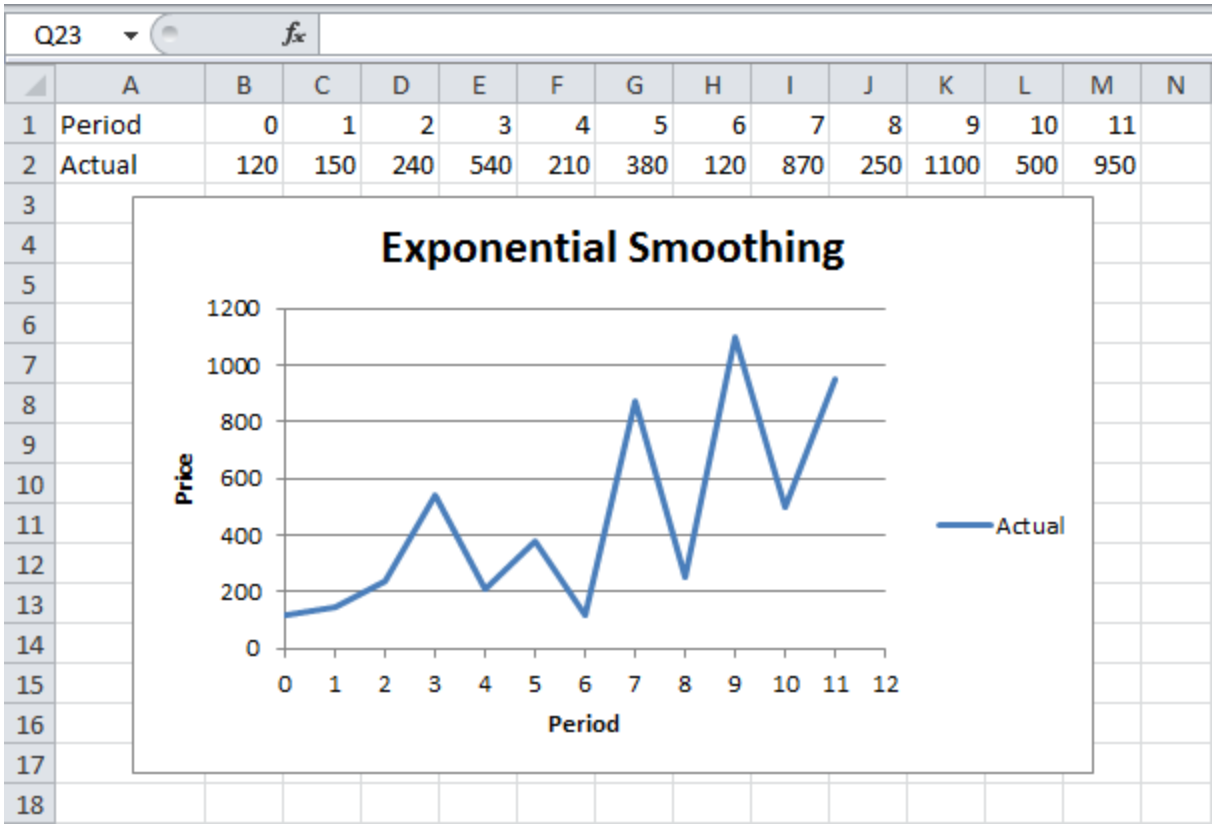
Both Exponential smoothing and Moving average options will be displayed among the Data Analysis options.



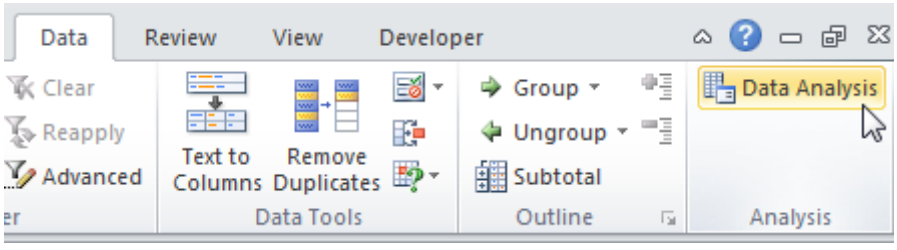
Exponential smoothing

Let’s now focus on Exponential smoothing by considering an example.

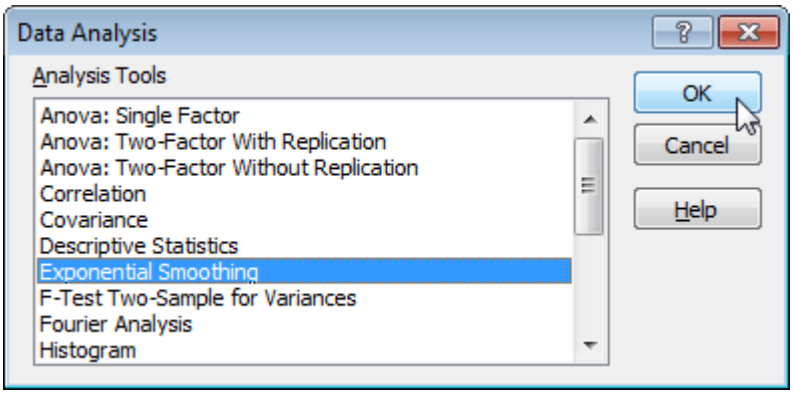
1. First, let's take a look at our time series.



2. On the Data tab, click Data Analysis.



3. Select Exponential Smoothing and click OK.

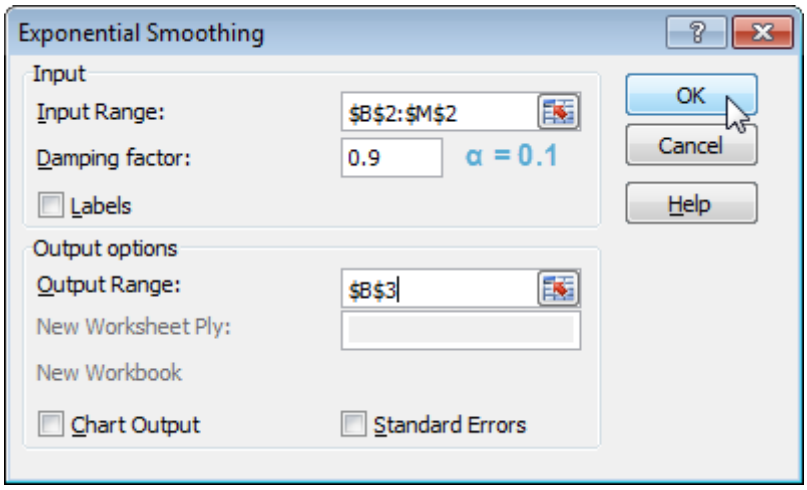


4. Click in the Input Range box and select the range B2:M2.

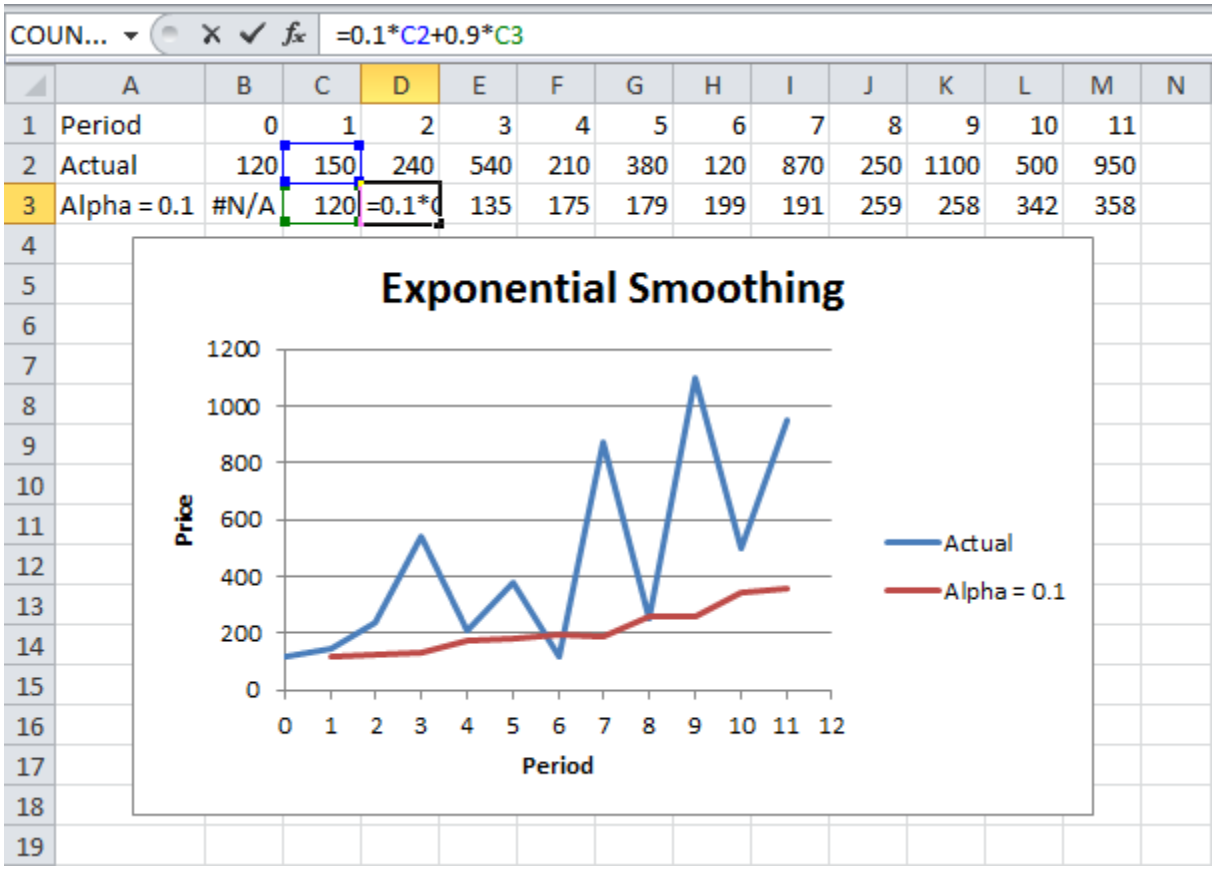
5. Click in the Damping factor box and type 0.9. Literature often talks about the smoothing constant α (alpha). The value (1- α) is called the damping factor.

6. Click in the Output Range box and select cell B3.

7. Click OK.



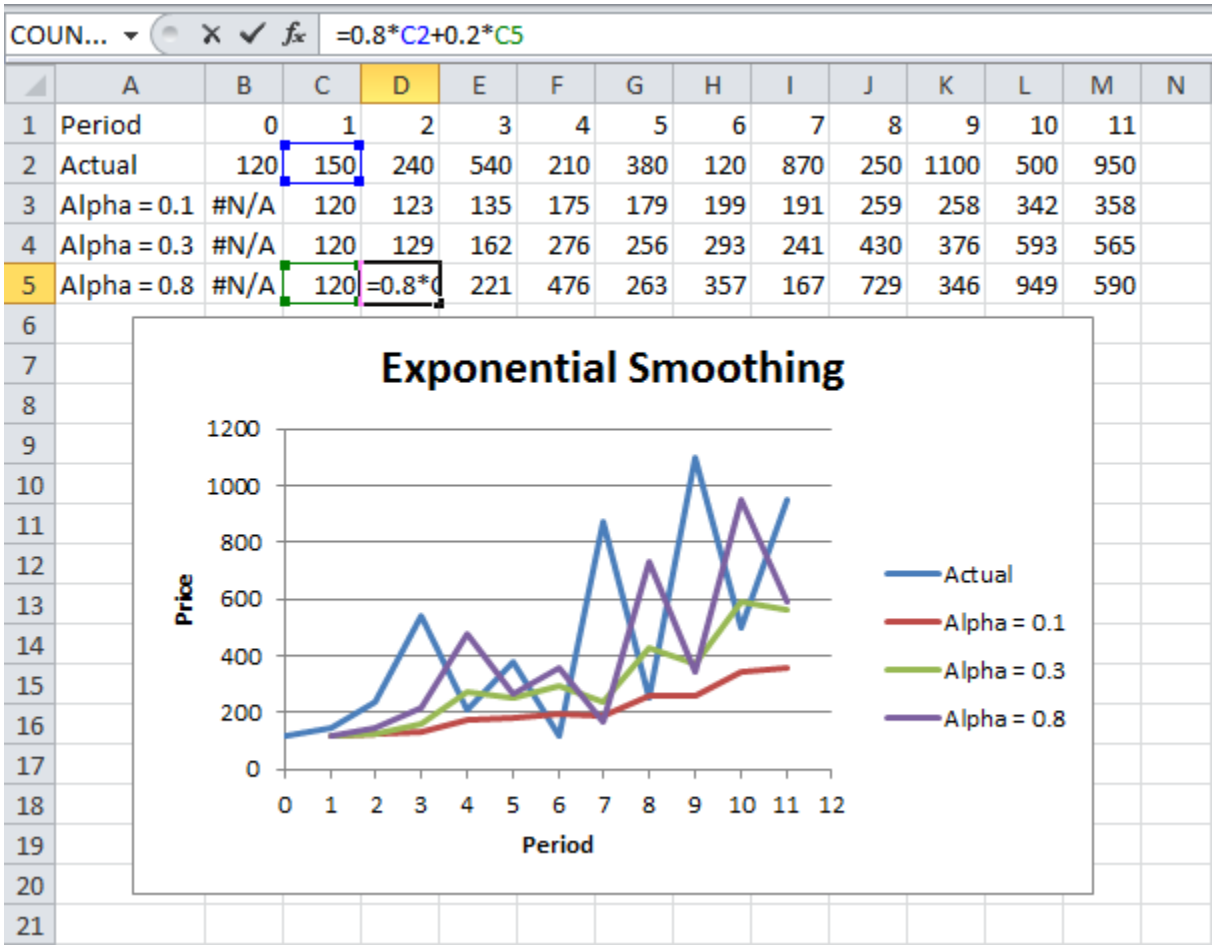
8. Plot a graph of these values.



	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Period	0	1	2	3	4	5	6	7	8	9	10	11
2	Actual	120	150	240	540	210	380	120	870	250	1100	500	950
3	Alpha = 0.1	#N/A	120	=0.1*C	135	175	179	199	191	259	258	342	358

Explanation: because we set alpha to 0.1, the previous data point is given a relatively small weight while the previous smoothed value is given a large weight (i.e. 0.9). As a result, peaks and valleys are smoothed out. The graph shows an increasing trend. Excel cannot calculate the smoothed value for the first data point because there is no previous data point. The smoothed value for the second data point equals the previous data point.

9. Repeat steps 2 to 8 for alpha = 0.3 and alpha = 0.8.



	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Period	0	1	2	3	4	5	6	7	8	9	10	11
2	Actual	120	150	240	540	210	380	120	870	250	1100	500	950
3	Alpha = 0.1	#N/A	120	123	135	175	179	199	191	259	258	342	358
4	Alpha = 0.3	#N/A	120	129	162	276	256	293	241	430	376	593	565
5	Alpha = 0.8	#N/A	120	=0.8*C	221	476	263	357	167	729	346	949	590

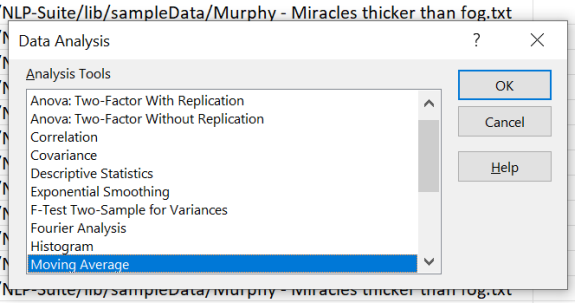
Conclusion

The smaller alpha (larger the damping factor), the more the peaks and valleys are smoothed out. The larger alpha (smaller the damping factor), the closer the smoothed values are to the actual data points.

Moving average

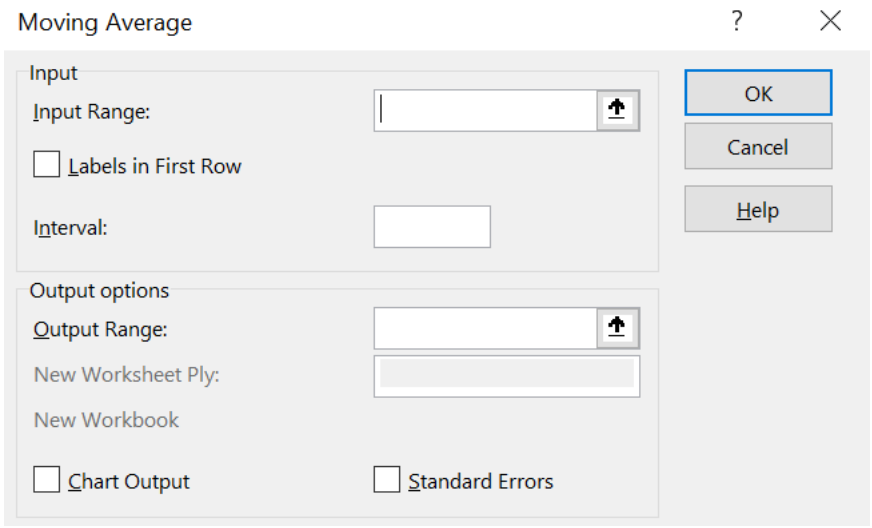
Let's now consider another example based on sentiment analysis for the Moving average option.

	Sentiment	Sentiment	Sentence	Sentence	Document	Document
2	2	neutral	1	Miracles t	1	C:/Users/rfranzo/Desktop/NLP-Suite/lib/sampleData/Murphy - Miracles thicker than fog.txt
3	2	neutral	2	It was the	1	C:/Users/rfranzo/Desktop/NLP-Suite/lib/sampleData/Murphy - Miracles thicker than fog.txt
4	1	negative	3	As I was w	1	C:/Users/rfranzo/Desktop/NLP-Suite/lib/sampleData/Murphy - Miracles thicker than fog.txt
5	1	negative	4	The mirac	1	C:/Users/rfranzo/Desktop/NLP-Suite/lib/sampleData/Murphy - Miracles thicker than fog.txt
6	1	negative	5	The anestl	1	C:/Users/rfranzo/Desktop/NLP-Suite/lib/sampleData/Murphy - Miracles thicker than fog.txt
7	2	neutral	6	He came t	1	C:/Users/rfranzo/Desktop/NLP-Suite/lib/sampleData/Murphy - Miracles thicker than fog.txt
8	2	neutral	7	My wife, K	1	C:/Users/rfranzo/Desktop/NLP-Suite/lib/sampleData/Murphy - Miracles thicker than fog.txt
9	1	negative	8	But even t	1	C:/Users/rfranzo/Desktop/NLP-Suite/lib/sampleData/Murphy - Miracles thicker than fog.txt
10	1	negative	9	Everything	1	C:/Users/rfranzo/Desktop/N
11	2	neutral	10	The distan	1	C:/Users/rfranzo/Desktop/N
12	2	neutral	11	I had plen	1	C:/Users/rfranzo/Desktop/N
13	2	neutral	12	I rememb	1	C:/Users/rfranzo/Desktop/N
14	1	negative	13	I looked it	1	C:/Users/rfranzo/Desktop/N
15	1	negative	14	I rememb	1	C:/Users/rfranzo/Desktop/N
16	3	positive	15	I recalled I	1	C:/Users/rfranzo/Desktop/N
17	1	negative	16	I knew, th	1	C:/Users/rfranzo/Desktop/N
18	1	negative	17	She did nc	1	C:/Users/rfranzo/Desktop/N
19	2	neutral	18	Just in cas	1	C:/Users/rfranzo/Desktop/N
20	1	negative	19	I rememb	1	C:/Users/rfranzo/Desktop/N

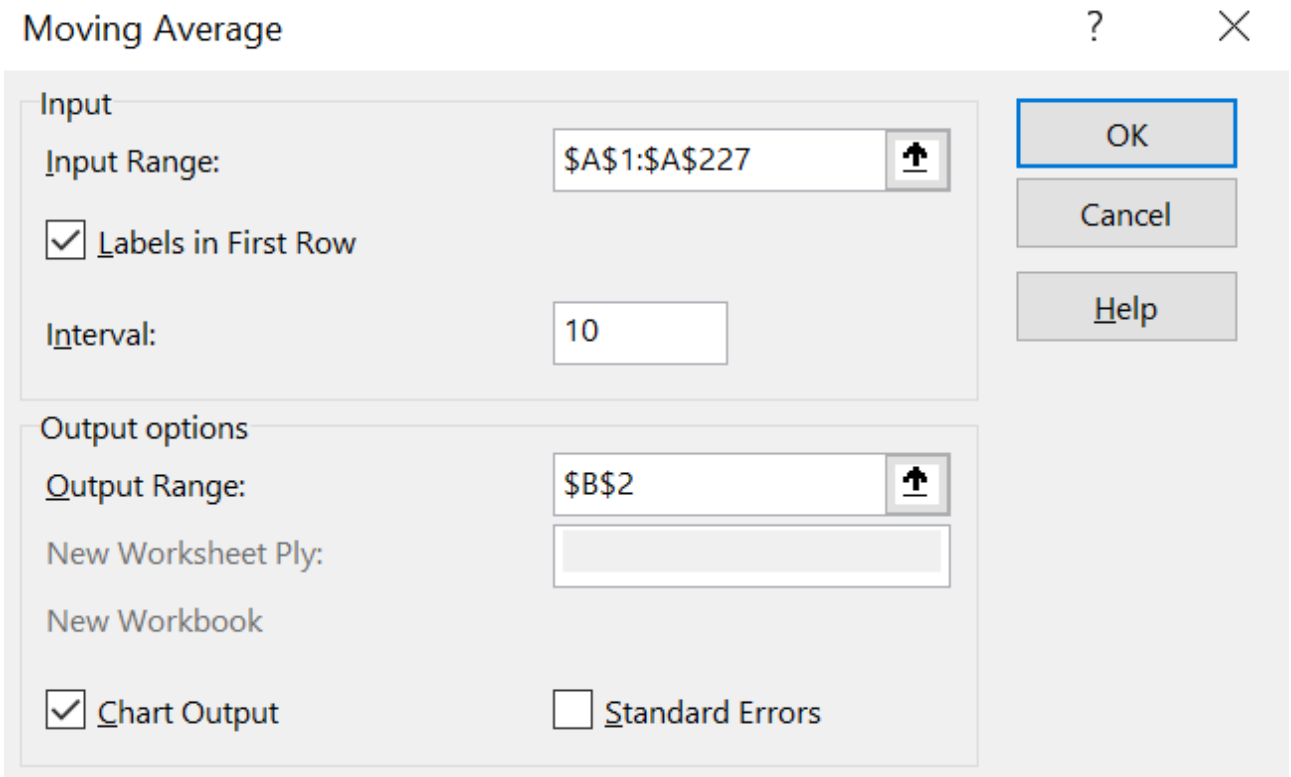


Before you get started, you need to insert a new column for the moving average output not to overwrite any of your data.

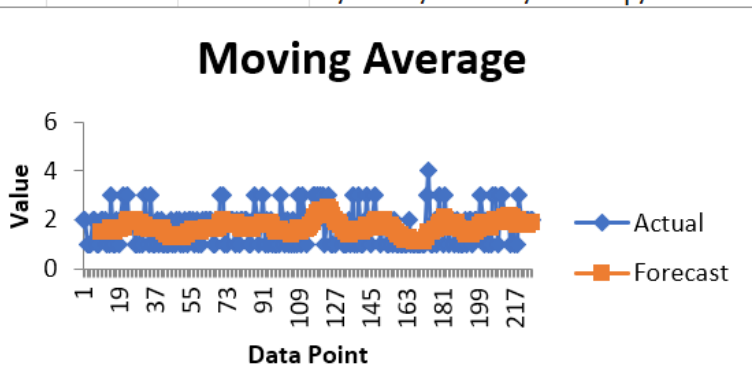
You will be required to enter the range of your data to be smoothed (B2:B227)



Ener the appropriate values, the rage of values in the column input.
Click the checkbox Labels in First Row, since we do have headers. In the Output Range start from row 2. The Chart Output option will automatically create a chart. Click OK to visualize the output.

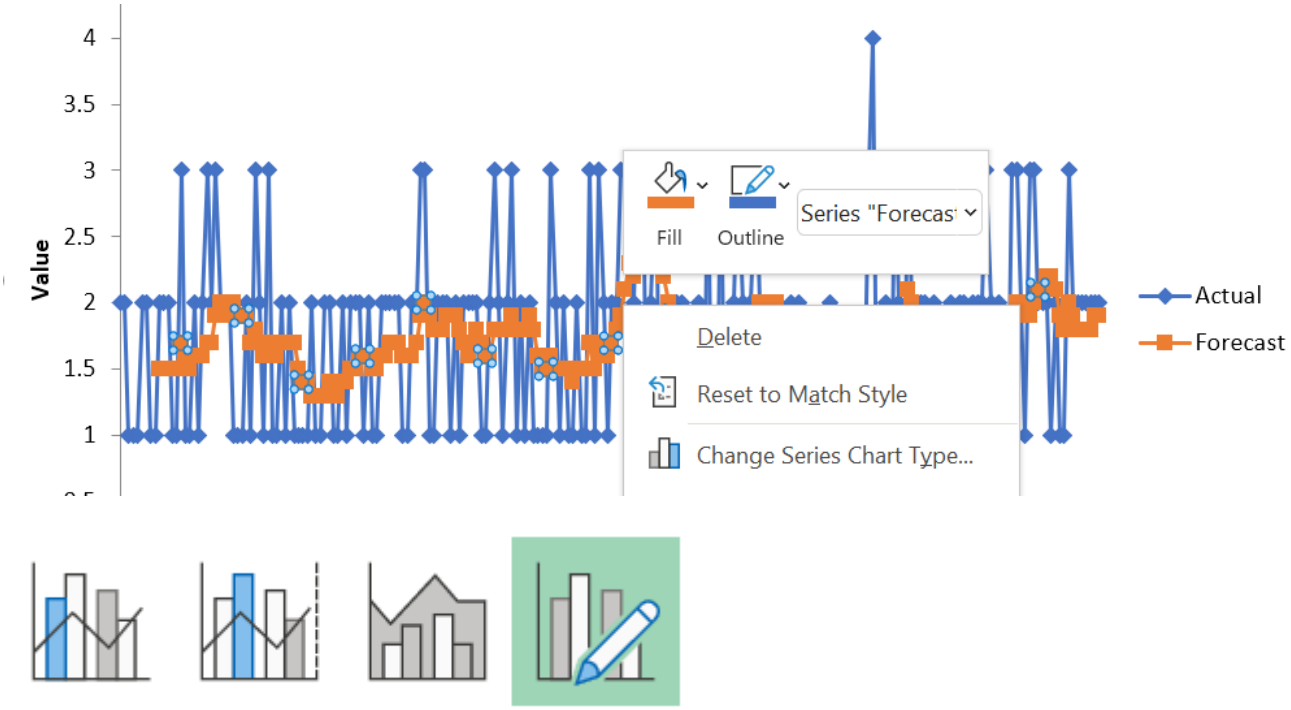


	Sentiment	Moving av	Sentiment	Sentence I	Sentence	Document	Document
2	2	#N/A	neutral	1	Miracles tl	1	C:/Users/rfranzo/Desktop/NLP-Suit
3	2	#N/A	neutral				
4	1	#N/A	negative				
5	1	#N/A	negative				
6	1	#N/A	negative				
7	2	#N/A	neutral				
8	2	#N/A	neutral				
9	1	#N/A	negative				
10	1	#N/A	negative				
11	2	1.5	neutral				
12	2	1.5	neutral				

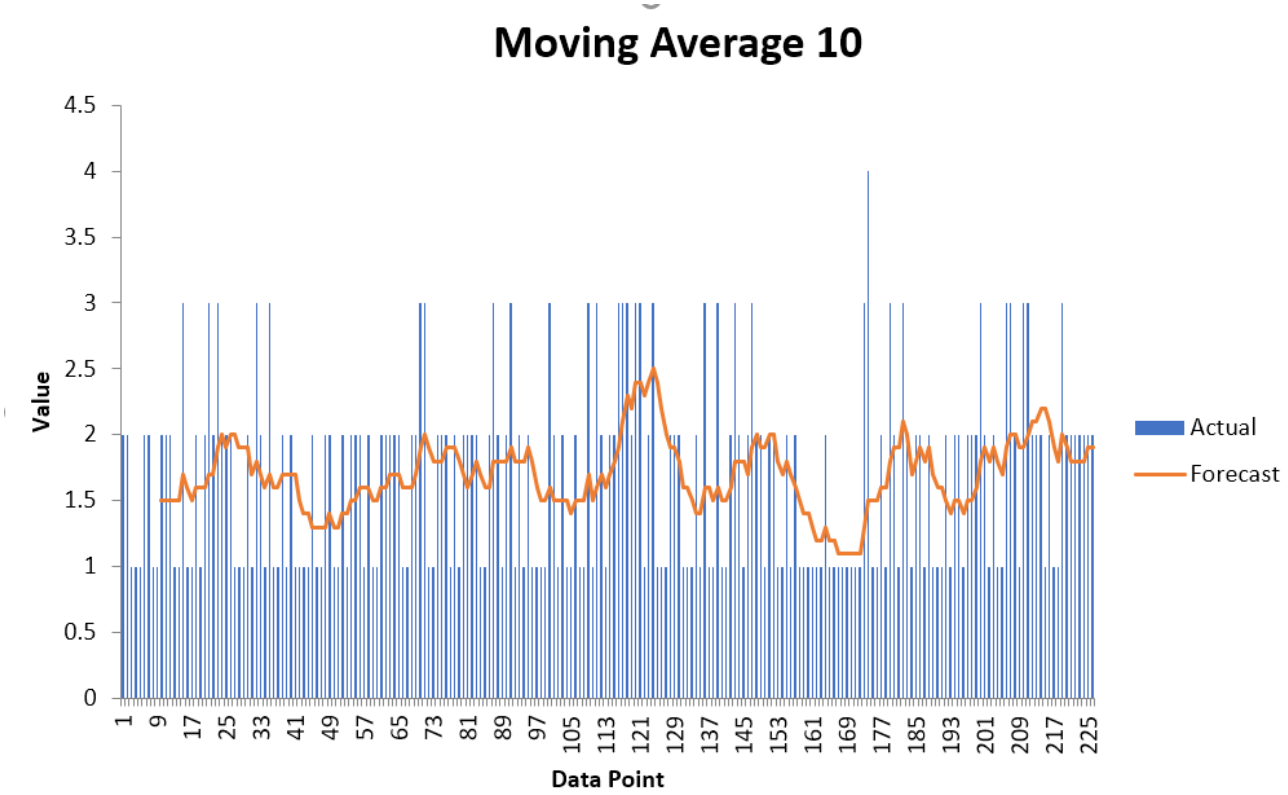


Two tings to notice.

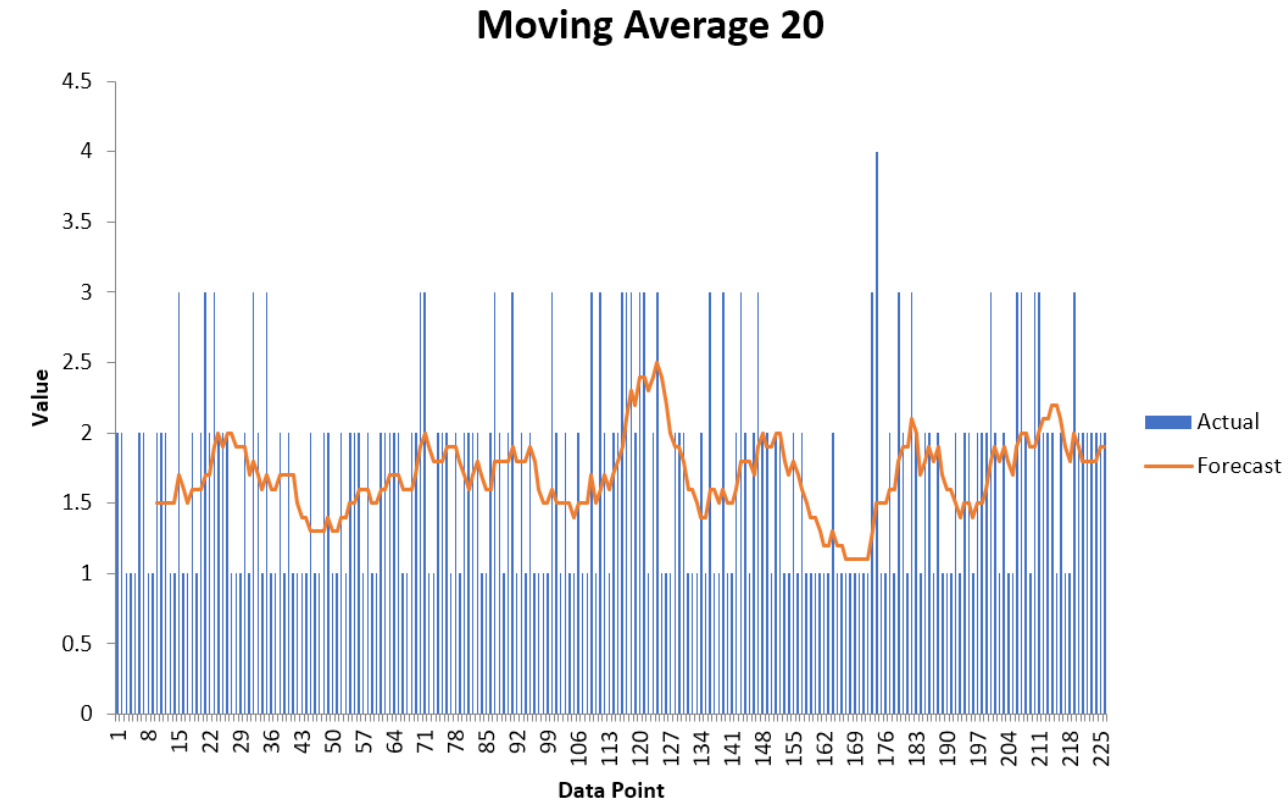
1. The #N/A indicate that values are Not Available since with an interval of 10 the first available value would be at 11.
2. The chart itself is pretty messy. Click on the chart, expand it, and right click on it to select Change Series Chart Type to change it to a simpler line chart.



Click on the first options and voila'. Here is your chart (after changing the chart title by adding 10 as your interval option).



The same chart with a moving average based on an interval of 20 gives you an even smoother series.



For a good video of various options, watch https://www.youtube.com/watch?v=cP7VTj9xFqM