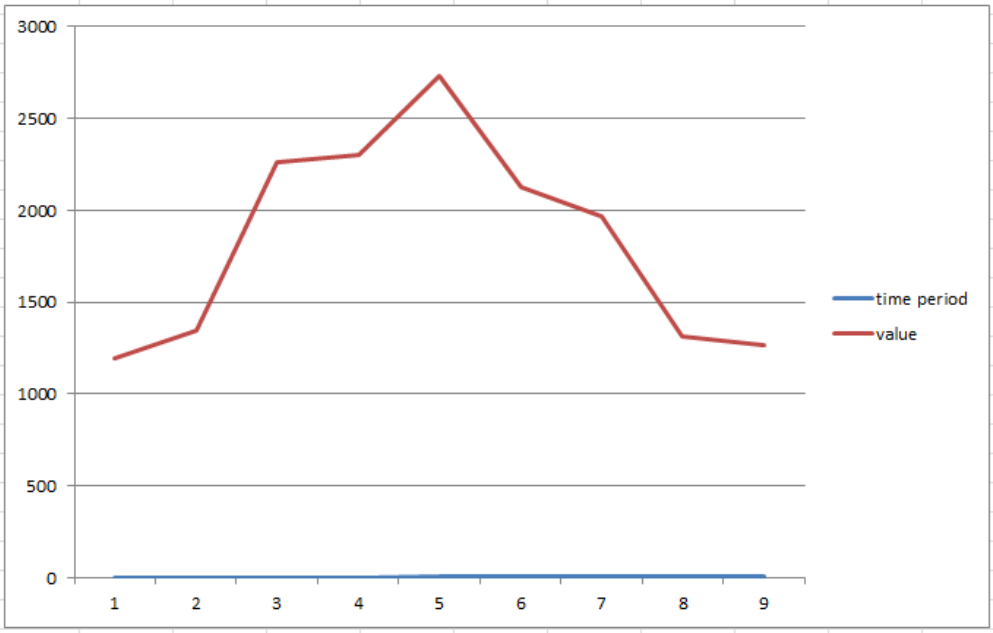
Project: Exponential Smoothing

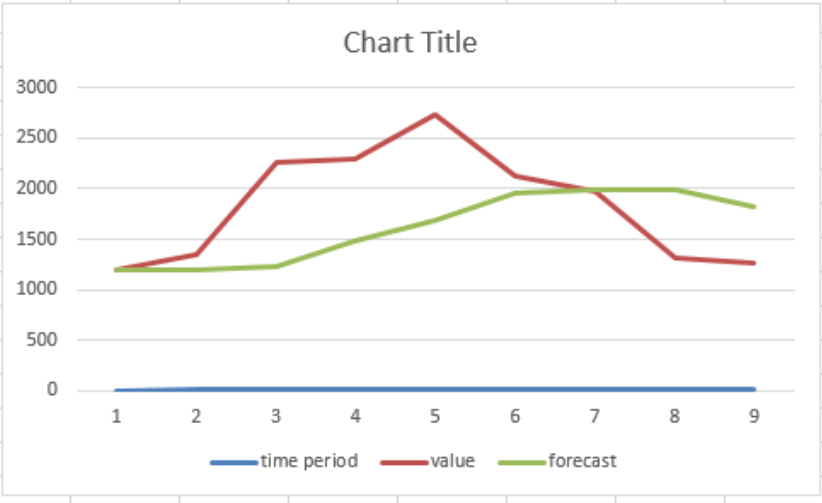
Many time series values give spiked results

	A	B
1	time perio	value
2	1	1195
3	2	1349
4	3	2265
5	4	2304
6	5	2732
7	6	2130
8	7	1971
9	8	1312
10	9	1268



This is difficult to develop a forecast on due to the erratic nature. Exponential smoothing can smooth this out so predictions are more stable:

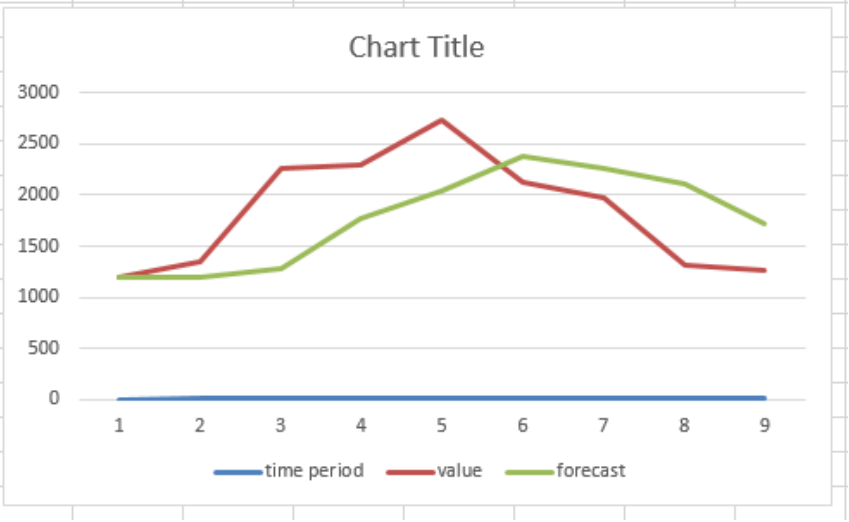
alpha =		0.25
time period	value	forecast
1	1195	1195
2	1349	1195
3	2265	1233.5
4	2304	1491.375
5	2732	1694.531
6	2130	1953.898
7	1971	1997.924
8	1312	1991.193
9	1268	1821.395



Where alpha = 0.25

Or perhaps this model:

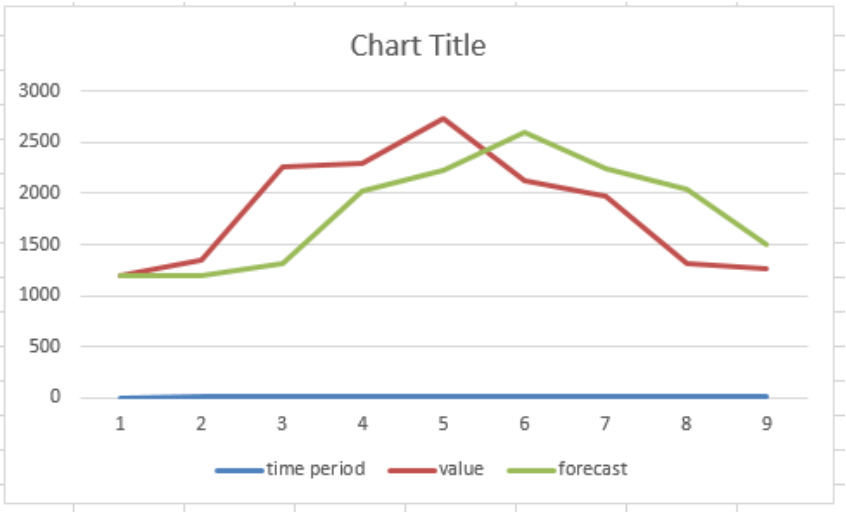
alpha =		0.5
time period	value	forecast
1	1195	1195
2	1349	1195
3	2265	1272
4	2304	1768.5
5	2732	2036.25
6	2130	2384.125
7	1971	2257.063
8	1312	2114.031
9	1268	1713.016



Where alpha = 0.5

Or perhaps this model:

alpha =		0.75
time period	value	forecast
1	1195	1195
2	1349	1195
3	2265	1310.5
4	2304	2026.375
5	2732	2234.594
6	2130	2607.648
7	1971	2249.412
8	1312	2040.603
9	1268	1494.151



Where alpha = 0.75

What do you believe will happen if alpha is equal to 1?

Formula for Exponential Modeling:

$$F_{t+1} = \alpha Y_t + (1-\alpha)F_t$$

where

F_{t+1} = forecast for time period t + 1

Y_t = time series value

F_t = forecast for time period t

α = smoothing constant (between 0 and 1)

note : $F_1 = Y_1$

In this project your program will use the exponential smoothing to help you predict a value in the future. In addition your program should come up with the linear regression equation to predict the same value as was done in exponential smoothing.

Part 1: Obtain the stock price inform from www.nasdaq.com. In this part, go to www.nasdaq.com, choose a company to analyze, click the "historical quotes" link on the left side after picking a company. Pick the stock prices for at least 8 months picking one data point out of each month as close to the first of the month as possible. The x values will be from 1 to 8 (where 1 indicates the first month looked at) while the y values will be the stock price. For example:

01/09/2014	189.02	189.5	186.55	187.38	4,321,330
01/08/2014	189.33	189.4175	187.26	187.97	4,603,156
01/07/2014	186.39	190.35	186.38	189.71	5,932,330
01/06/2014	187.15	187.355	185.3	186	4,067,790
01/03/2014	185.83	187.35	185.3	186.64	4,063,215
01/02/2014	187.21	187.4	185.2	185.53	4,546,489
12/31/2013	186.49	187.79	186.3	187.57	3,619,745
12/30/2013	185.32	186.7	184.67	186.41	3,018,390

Would give the data point: x = 1, y = 187.21. For the second data point the list is scrolled to find to find the first historical quote for the next month:

02/06/2014	173.97	174.85	173.79	174.67	4,292,241
02/05/2014	172.19	174.97	172.19	174.24	4,712,296
02/04/2014	173.53	173.75	172.36	172.84	4,349,776
02/03/2014	176.02	176.02	172.72	172.9	7,186,803
01/31/2014	176.11	177.835	175.34	176.68	5,193,390

Would give the data point: x = 2, y = 176.02. And so on. Keep in mind that the values for x (1, 2, …, 8) DO NOT have to correspond to Jan, Feb, etc. x = 1 merely indicates the first month that you decided to analyze. From there the months should proceed sequentially. Once the data is gathered then you should have a list of values such as this example:

Time period	1	2	3	4	5	6	7	8
Value	31.25	32.35	34.15	33.12	37.25	30.19	42.13	44.17

This table should be presented in an Excel document explaining why the choice was made for the company chosen (why would a model based on the first of the month possible make sense) and the time frame.

Part 2: Write a Python program that asks the user for the information from part 1 and performs exponential smoothing based on it. The perfect program will allow the user to input alpha, display the graph of the original data and the "smoothed data" and have the user verify if this model is appropriate. If it is not then it should loop asking for new entries for alpha until the user indicates the model is appropriate. At this point, it should use the exponential smoothing model to predict time period 9 (x = 9). Read the explanation above closely to understand what exponential smoothing provides for the next month based on the previous month.

Part 3: In the same Python program, the information from NASDAQ should be used to develop a linear regression model that is used to predict time period 9. It should show the correlation coefficient to indicate the strength of the model. No other tool is necessary for this project to test the appropriateness of using a linear regression model.

Category	20 pts	15 pts	10 pts	5 pts	0 pts
Data Collection Criteria - Company Identified and Choice Explained - Each data value obtained from first entry for the month chosen - Data was collect from sequential months - Data was delivered in an Excel spreadsheet	All 4 criteria met	Only 3 criteria was met	Only 2 criteria was met	Only 1 criteria was met	None of the criteria was met
Exponential Smoothing Logic	Logic is 100% correct	Logic is 75% correct	Logic is 50% correct	Logic is 25% correct	Logic does not follow the exponential smoothing formula

Visual Display of Smoothed Data	Python program prints the original data and the smoothed data (not necessarily in the same graph)	Python program does not print the graphs but writes the R program to a file which can then be opened into R and run.	Python program does not print the graphs but outputs the appropriate R lines of code that can be Edit-Copied and Edit-Pasted into R for viewing for graph with no re-formatting necessary in R.	Python program does not print the graphs but outputs the data that can be Edit-Copied and Edit-Pasted into R with manual reformatting of the data. (ex: 5, 8, 7, 10 Which then has to be formatted in R as x <- c(5,8,7,10)	No attempt is made to display the graphs.
Regression Logic	Logic is 100% correct	Logic is 75% correct	Logic is 50% correct	Logic is 25% correct	Logic does not follow the regression logic from the last unit in the course.

Structure and Design Criteria: - The appropriate flow mechanism is used in the program (while loop, etc.) for ease of use of the user - Code is placed in a library for code re-use - Code is documented where appropriate. - Code is "readable" (appropriate variable names and structured programming techniques used).	All 4 criteria met	Only 3 criteria was met	Only 2 criteria was met	Only 1 criteria was met	None of the criteria was met

Deliverables

- Excel document with data and explanation of company choice
- Python program
- Python library (if a library is implemented)
- R programs (any appropriate R programs)