**Exercise**

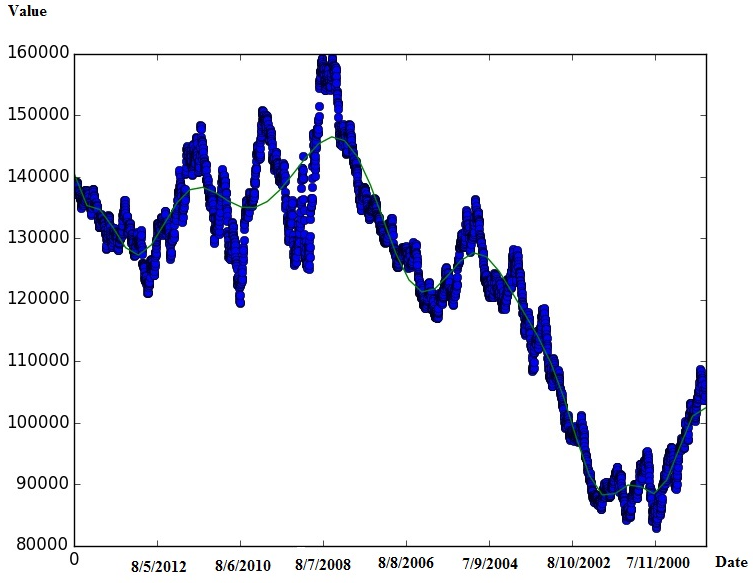
**#1 - data manipulation:**

This is a bit of a pure data manipulation/processing exercise. We have provided a sample of some real world data (TrendData.csv) of a value over time.

1. We would like you to come up with the best solution of segmenting the data by date based on trends in the Value column (an example input and output has been given). Please write some form of code/procedure or script to do this, the method should be easily repeatable for new data. The output should be a csv file with the following columns StartDate, EndDate, StartValue, EndValue. Start and End Date define the bounds of the segment and we ask you to include the Value (value from the input data) at the start and end of the trend for convenience. Note: Example\_input.csv and Example\_output.csv only represent the file structures your code should handle, and they are not related at all.

2. List your assumptions and explain why you chose them.

I used scatterplots to show the trend of data and I fitted a curve on the data to find the minimum and maximum points in the data, where the gradient is zero. Then, I represented the start date, end date, start value, and end value of the data by demonstrating the minimum and maximum of the data.



3. Are there any improvements you would potentially make to your approach / what

other information would be helpful in refining/choosing your approach?

This approach works great and shows the trends of the data very well, however, it is not so accurate as it doesn't show all the minimum and maximum points. We could improve it by increasing the degree of polynomial to fit an accurate curve. It is easy to change this point in my code, I just didn’t want to make it more accurate.