# Python Exercise

## Installation

### Windows Users

1. Install Anaconda package if you are a new python user,especially on windows. Anaconda package can be installed from - https://www.continuum.io/downloads. Download Python3.

2. Managing packages in Anaconda - http://conda.pydata.org/docs/using/pkgs.html

3. Use the anaconda terminal to manage your packages. Packages can be downloaded in the following way - conda install package-name

4. Install pip using conda install pip. You can now install packages using the following - pip install package-name

5. To run jupyter notebooks, use the command 'jupyter notebook' in the Anaconda terminal. For more instructions on how to use jupyter notebooks - http://jupyter.org/

### Mac/Linux Users

By default, python 2.7 and python 3.5 will be available for you. In the terminal, you can start python console using 'python' (Python 2.7) and 'python3'(Python 3.5). I would suggest using Python 3.5. The following instructions are specific for Python3.5.

If you do not have python3 installed, refer - https://www.macobserver.com/tmo/article/how-to-upgrade-your-mac-to-python-3

Installing packages (use terminal for the following commands) -

1. pip3 install --upgrade pip

2. pip3 install package-name

To install jupyter (use terminal) -

1. pip3 install jupyter

2. Use 'jupyter notebook' to start working with jupyter notebook

3. More details - http://jupyter.org/

If you find the above process confusing, follow the instructions for windows users.

## Practice questions

1. Create a list named my\_list in python with the following data points -

45.4 44.2 36.8 35.1 39.0 60.0 47.4 41.1 45.8 35.6

* 1. Print the 5th element in the list
  2. Append 55.2 to my\_list
  3. Remove the 6th element in the list
  4. Iterate over the list to print data points greater than 45

1. Introduction to numpy –
   1. Import the numpy library using the following command – **import numpy**
   2. Declare numpy array with the same data points as in my\_list using **numpy.array()**
   3. Compute the mean and standard deviation using numpy.mean() and numpy.std() of the above array
   4. Use logical referencing to get only those values that are less than 45
   5. Compute the max and min of the array using numpy.max() and numpy.min()
2. Introduction to pandas –
   1. Import the pandas library – **import pandas**
   2. Read the IRIS dataset into **iris** using pandas.read\_csv(). Data file –



* 1. Using iris.head(), display the head of the dataset
  2. Use DataFrame.drop() to drop the **id** column
  3. Subset dataframe to create a new data frame that includes only the measurements for the setosa species
  4. Use DataFrame.describe() to get the summary statistics
  5. Use DataFrame.plot() to plot the boxplot for all 4 numeric columns/iris measurements
  6. Use DataFrame.groupby() to create grouped data frames by Species and compute summary statistics using DataFrame.describe()
  7. Use DataFrame.boxplot() to plot boxplots by Species
  8. Plot a scatter matrix plot using the seaborn library. Use the following to load and plot –
     1. **Import seaborn**
     2. **Seaborn.pairplot(dataframe,by=’column\_name’)**