Assignment 6

In this module, you will write two Python notebooks to create required outputs. You will also participate in module 6 forum discussion. Video 8 describes the requirements.

You will use markdown cells and be creative of summarizing/commenting your notebook. You will also add detailed comments in your Python code (using "#" or triple quote signs)

Notebook 1: visualization

Watch the lecture 3 (videos 6-7) and continue working on the visualization portion.

Requirements:

1. Create a histograms as close to the sample_plot_result.png as possible. "x-distribution" is the "average" scores, "y-distribution" is the math scores
2. Create a function – find_outlier() that takes a string as input and output the outliers in the corresponding column in the dataframe. The outliers are defined as values more than 1.8 times of the standard deviation from the mean of the entire column

*The data is from https://en.wikipedia.org/wiki/Programme_for_International_Student_Assessment PISA 2012 dataset. It's the first table on the Wikipedia page.

Notebook 2: twitter

Watch lecture 2 (video 5), collect real-time data from twitter, and provide basic analysis.

Requirements:

1. Collect real-time data using a search term that you are interested in
2. Choose one specific field in the collect tweets to describe and analyze the data. Use proper plots and write a paragraph in your notebook to report what you found. The "map-of-a-tweet.png" provides a description of the data fields in tweet.

Submissions:

You will export your notebook to both .html and .py formats. You will submit the following 2 files to Blackboard. In your html file, you should **include all the outputs** of your python script without error messages.

1. Firstname_Lastname_visualization.zip (zip the .html and .ipynb files)
2. Firstname_Lastname_twitter.zip (zip the .html and .ipynb files)

Attachments:

Sample_plot_result.png:	Sample plot
Outlier_function.png	Sample input and output of the find_outlier() function
Map-of-a-tweet.png:	Description of data fields in twitter tweets