Assignment 1:

Due at 11:59pm on Friday 13th of November through Moodle.

Provide the Python script. Code should be well commented to show which question is being answered, and to describe what your code is doing.

Question 1

(a) There is a file containing weather data for three Brazilian stations in the brazil_weather csv file on Moodle. Read it in to Python as a Pandas DataFrame.

(b) List all problems you can find with the dataset that could be improved before analysing it.

(c) Use Python to make the dataset as tidy as possible for analysis ie. solve the problems you have identified. Make sure the dataset is being updated with the changes you make.

(d) Remove the mean hourly wind speed column. It is not needed for analysis.

(e) Save the cleaned dataset as a new csv file called brazil_weather_edit.csv for later analysis.

Question 2

Having tidied the dataset, answer the following questions. The questions should be answered using Python code to output the answer to the console, not by visual inspection.

(a) What was the hourly maximum temperature at station number 304 at 07:00 on 10/03/2010?

(b) What was the highest minimum temperature in the dataset? At which station number was this temperature measured?

(c) Output the values of row index 145 in the first 7 columns to the console.

(d) How many hourly total rainfall values in the dataset exceed 50mm?

(e) Find the daily total rainfall at station number 304 on 07/01/2009.

(f) Sort the dataset by hourly maximum temperature in descending order.

(g) Sort the dataset by hourly maximum temperature in descending order, then by minimum temperature in ascending order.

(h) Find all rows in the dataset in which hourly rainfall exceeded 60mm or the hourly maximum temperature was at least 39 °C.

(i) Output all rows that do not have any missing values.

(j) Find the mean of the non-missing hourly minimum temperatures at station number 178 in December 2007.

(k) Output the values of hourly maximum temperature when the mean hourly humidity is 100%.

(l) Which station number had the highest number of mean hourly humidity values equal to 100%?