MEF UNIVERSITY BIG DATA ANALYTICS GRADUATE PROGRAM

BDA507 INTRO TO PROGRAMMING FOR BIG DATA (PYTHON)

ASSIGNMENT #4 (FALL 2017-18)

Instructor: Dr. Tuna Çakar

Due: November 26 (Sunday), 2017, 23:59 (sorry but no-late submission is accepted)

Guidelines for Assignments:

- The current assignment is an individual-based one. You can discuss the content with your friends but you must do it on your own and with your own codes/sentences.
- Please provide your comments properly for every task to be able to get full credit for your work.
- You should submit it via Blackboard system on-line. No e-mail submission is graded but you can send also via email for the potential problems related to Blackboard upload. You should send these additional e-mails to repository@mef.edu.tr
- This assignment includes five questions and each of these is worth a single point. You should upload your responses with the following procedure: Go to the calendar, choose the Assignment-4, then assignment submission (by the end of the page), then file submission, upload your python file and submit it. Again, no late assignment will be accepted.
- If there is a single unclear issue, please send an email to me via cakart@mef.edu.tr

Questions:

Q1. Please prepare a number array of 3 rows and 7 columns. The content should be the numbers from 9 to 100s with the incrementation value of 7. Then you should make a clone of this array with a different name. Then this copy array's values should be changed to remainder of 8. For instance, the initial array held the values: 9, 16, 23, 30, …. After being copied to a new array the values should be adjusted as such: 1, 0, 7, 6 …. (these are the remainder when the initial values are divided by 8).

Q2. Prepare a numpy array of 7 x 7 matrix with all its content having the value 9. Then please change the border content to 8 and the inside content to 7 separately. Please try to do it with the short cuts that we have done in the class (this is Part-A) as well as using 2 for loops (this is Part-B).

Q3. Please prepare the checkerboard pattern with using for loops as opposed to the one that we have done during our last class. The code for checkerboard pattern that we have done during class is as follows:

```
Print ("Checkerboard pattern:")
x = np.zeros((8,8),dtype=int)
x [1: : 2, : : 2] = 1
x [: : 2, 1 : : 2] = 1
print(x)
```

Q4. For Titanic dataset, please work on the "test.csv" file this time and do the following tasks (The file is also in the relevant folder for Week-4):

(a) Get the dataset via pandas library, (b) display the dimensions (rows and columns), (c) show the first 10 lines, (d) show the descriptive statistics both for numeric and categorical data, (e) change the survived column to categorical data (as "yes" and "no" values), (f) display the number of passengers that have used A-B-C-D-E-F cabins, (g) display the number of passengers whose age is greater than 40 and male (then count the females that are greater 40 years of age), (h) count the number of missing values for Age column.

Q5. Please plot the histograms for each of the columns in the dataset. You can decide on the details (such as figure sizes and bins) of the figures. The code that we have gone over in the class includes the relevant function you need to run.