COMP 247 Lab assignment #1: “Data pre-processing”

**Assignment due date:** End of week # 3

# Purpose:

The purpose of this Lab assignment is to:

1. To practice data pre-processing using Pandas library.
2. To practice data pre-processing using Numpy library.
3. To practice generating plots using Matplotlib

# General Instructions:

Be sure to read the following general instructions carefully:

1. This assignment must be completed individually by all the students.
2. If needed copy from original notebook.

## Pre-requisite to carrying out the assignment:

1. Jupyter Notebook
2. Download the attached three notebooks available with this assignment & practice executing each notebook.

**Assignment - exercises:** (100 marks)

Create a new notebook name it firstname\_data and carry out the below exercises:

**Exercise #1 – Pandas: (30 marks)**

Inspect the sections mentioned in the table below in the pandas notebook and carry out the requirements:

|  |  |  |
| --- | --- | --- |
|  | **Section in notebook for guidance** | **Requirements** |
| 1 | Init from dict | Inspect the code under this section, then add a cell to:   1. Create a new dictionary, name it your firstname where firstname \_fruits is your first name. 2. Add four items to the dictionary with names of your favorite fruits as keys and the respective color as values. 3. Convert the dictionary into a pandas series named firstname\_f. 4. Print out the second and third items. 5. Create a sub series named firstname\_f2 containing the second and third items. 6. Printout from the sub series the last item using iloc. |
| 2 | Handling time | Add a cell with the following logic:   1. Create a list containing four  rainfall amounts  of values 10, 23,24,30 name the list firstname\_amounts. 2. Using pandas create a date\_range for todays date/time (you can set any time) with four time intervals. 3. Create a series that combines both the list and date range name it firstname\_rainfall amounts\_today. 4. Plot as bar chart. |
| 3 | Pandas Multi - indexing | Make a copy of the dataframe d5 and name it fristname\_d5, carryout the following:   1. print out a dataframe containing all “private” columns 2. Swap the columns and rows (hint: look at transpose) |
| 4 | Querying | Use the query() to query the people dataframe  you created earlier and retrieve everything related to alice. |
| 5 | Operations on dataframes | Add a cell to create a dataframe containing grade for four students choose the name of the students and use the names as index. For columns create four columns to reflect the months April, May, June, July. Set grade items for each student for each month to be between 0 and 100.  Name the dataframe fristname\_grades. Carry out the following using pandas operations:   1. Print out the average for the month of April 2. Adjust all the grades by 2% (i.e. increase) 3. Printout the grades for the month of may that are higher than 50% 4. Group the failing students i.e. the students with average over four month below 50% |

**Exercise #2 – Numpy (45 marks):**

Inspect the sections mentioned in the table below in the numpy notebook and carry out the requirements:

|  |  |  |
| --- | --- | --- |
|  | **Section in notebook for guidance** | **Requirement** |
| 1 | np.functionname | Add a cell to create a function and name it my\_function\_firstname, *where firstname is your first name.*  Let the function return an integer value stored in one byte i.e. ‘int8’ of (4x)\*(3y). Where x is the number of rows and y is the number of columns.   Use np.fromfunction() to generate  three elements each are two by six using the  my\_fuction\_firstname. |
| 2 | Multi-dimensional arrays | Inspect the code under this section copy it, add a cell to extract values 16,17,18 |
| 3 | Iterating | Inspect the code under this section copy it, then add a cell to iterate over c and print the Boolean values for items equivalent to zeros. |
| 4 | vstack | Inspect the code under this section copy it, then add a cell to create a variable name it q5\_firstname where firstname is your firstname and vertically stack q1 and q2 and print the output. |
| 5 | concatenate | Inspect the code under this section copy it, then add a cell to create a variable name it q8\_firstname where firstname is your firstname , concatenate q1 and q3 and print the results. |
| 6 | Transpose | Inspect the code under this section copy it, then add a cell and create a variable named t\_firstname where firstname is your name, let the variable hold any ndaray size 2 by 7 with zero values, print the result then transpose and print the result. |
| 7 | Matrix multiplication | Inspect the code under this section copy it, then add a cell to create 2 ndarys name the first a1 and the second a2. Both arrays should contain numbers in the range 0 to 8, inclusive . Print a1 and a2. Reshape a1 to a 2 by 4. Reshape a2 to a 4 by 2. Create a new variable a3 \_first name where firstname is your first name which holds the dot product  of a1 and a2 name it a3 and print the output of a3\_firstname, then the shape of a3\_first name. |
| 8 | Matrix inverse and pseudo-inverse | Add a cell to create a new 4 by 4 ndaray with values between 0 and 15, name the variable that holds the array your first name, print the array and the inverse of the array. |
| 9 | Identity matrix | Add a cell to create a 4 by 4 identity array. |
| 10 | Determinant | Add a cell to create a 3 by 3 matrix with values generated randomly then printout the determinant of the matrix. |
| 11 | Eigenvalues and eigenvectors | Add a cell to create a 4 by 4 matrix with values generated randomly, assign the matrix to a variable named e\_firstname. Printout the Eigenvalue and eigenvectors of the matrix. |
| 12 | Solving a system of linear scalar equations | Add a cell to solve the following linear equations:  2x+4y+z =12  3x+8y+2z =16  X+2y+3z = 3  Check the results using the allcolse method. |

**Exercise #3 - Matplotlib: (25 marks)**

Inspect the sections mentioned in the table below in the matplotlib notebook and carry out the requirements:

|  |  |  |
| --- | --- | --- |
|  | **Section** | **Amendment** |
| 1 | Plotting your first graph | Add a cell at the end to generate a 2 D graph as follows:  x holds 1000 values between -4 and 4  z holds 1000 values between -5 and 5  y = x^2 + z^3 +6  plot x and y  name the plot(i.e.set the title) “Ploynomial\_firstname” where firstname is your firstname.  Give names for the x and y axis. |
| 2 | Subplots | Add a cell at the end to generate a plot using subplot2grid with the following characteristics:  A 4 by 4 grid.  On the first row plot the function x^2 in a dashed green line.  On the second-row plot two functions, the first function x^3 in yellow color and the second function x^4 spanning three columns in red color.  On the third-row plot two functions the first X^6 in a dashed blue color and the second is X=x in magna (pink) color.  On the fourth row plot one function^7 spanning all columns in dotted red.    Your figure should look something like this: |
| 3 | Drawing text | On the first graph showing the beautiful point add a new point name new point \_firstname and display the coordinates, your figure should look something like this: |
| 4 | Scatter | Add a cell to generate a scatter plot of x and y where each contains 300 numbers generated randomly between 3 and 100. Set the scale, alpha and colors as you see suitable |

Once you are done with all three exercises, name the notebook firstname\_data. Then download the notebook as:

1. Html
2. Pdf
3. Notebook
4. Python executable script

**Submission:**

Save all four into a zipped file named firstname\_data\_pre-processing.zip and submit to the assignment folder.

## Rubric

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Evaluation  criteria | **Not acceptable** | **Below**  **Average** | **Average** | **Competent** | **Excellent** |
|  | **0% - 24%** | **25%-49%** | **50-69%** | **70%-83%** | **84%-100%** |
| Requirements & Submission | Missing all requirements required | Some requirements are implemented. | Majority of requirements are implemented but some are malfunctioning. | Majority of requirements implemented. | All requirements are implemented Correctly. |

**End of assignment**