

Question 1 (50 pts)

Conduct a research on the Internet and write a paper to describe what is Data Science? Why is Python language a choice and used for data science? Describe or give examples on how Python is implemented in data analysis? Use Microsoft Word to write your paper and use the American Psychological Association (APA) format to support your arguments. Your paper should have a minimum of 200 words.

Question 2 (50 pts)

(INDEXING AND SLICING ARRAYS) Create an array containing the values 1–15, reshape it into a 3-by-5 array, then use indexing and slicing techniques to perform each of the following operations:

- Select row 2.
- Select column 4.
- Select rows 0 and 1.
- Select columns 2–4.
- Select the element that is in row 1 and column 4.
- Select all elements from rows 1 and 2 that are in columns 0, 2 and 4.

Question 3 (100 pts)

For this question, please use NumPy Python library. Make sure you have installed NumPy, Pandas, SciPy, and Matplotlib Python libraries. Use command pip (for Windows) or pip3 (for MacOS).

Write a program to load a given dataset called *Student_Grades.csv* into a NumPy array. Then determine the following items:

- Display all data on screen.
- Determine how many students were in the dataset?
- Display the number of rows and columns of your numpy array.
- Display the array data types.
- Display the following Descriptive Statistics of students' overall percentage scores:
 - Min score
 - Max score
 - Mean value
 - Median
 - Mode
 - Standard Deviation
 - 25% and 75% percentile
- Determine how many students achieved an A grade, B, C, D and F grades.
- Create a pie chart based on the above grade achievements (option f)

Write a Learning Report Summary

Using Microsoft Word, write a summary report (not a bullet items) with a minimum of 100 words explaining how you completed your assignment. *Please describe your responses, not just yes/no answers.*

1. Did you successfully get your assignment done? Did it run? Any error? Did you get the correct result? Did you test your program thoroughly?
2. How much time did you spend to complete your assignment?
3. Did you find the assignment easy or challenging for you?
4. Did you write the program yourself? Did you get any help from anyone?
5. When you encountered obstacles to complete your program, how did you resolve the issues? Did you use Google to get help? Describe how Google was able or not able to assist you?
6. What did you learn from doing this assignment?
7. Any other information you would like to share with your instructor? Make sure you provide program output on each option.

What to submit on blackboard?

- a) Your written paper on data science.
 - b) Program source code listings.
 - c) Program test-run results.
 - d) Pie chart output.
 - e) Learning Report Summary.
-

The following is a sample of Question #3 program test-run result:

Display data

```
[[ 15.    50.    50.    ... 207.   1442.    94.1868 ]
 [ 15.    50.    50.    ... 197.   1322.    86.34879]
 [ 15.    50.    50.    ... 210.   1510.    98.62834]
 ...
 [ 15.    50.    50.    ... 200.   1374.    89.74526]
 [ 15.    50.    50.    ... 195.   1287.    84.0627 ]
 [ 15.    50.    50.    ... 210.   1490.    97.32201]]
```

There are 33 students

Display numbers of rows and columns
(33, 32)

Display data type
float64

Descriptive Statistics ...

Min overall score: 29.65382
Max overall score: 99.86936
Mean: 87.66996787878789
Median: 92.88047
Std. Dev.: 15.553686425410683
Percentile (25%, 75%): [87.1979 96.73416]

Number of students achieved in each grade category:

21 A
8 B
1 C
1 D
2 F

Student Performance Pie Chart

