DS2500 Spring 2025 Practice for Exam 1

> *This practice exam contains some sample questions on the same topics that will appear on Exam #1. Please complete the questions on your own, and we'll publish the solutions on Piazza*

Exam #1 takes place on February 14th, 2025. It will be administered on paper, during our usual lecture time.

The exam is designed to be shorter than our 100-minute lecture time, but you may use the entire class time to complete it. Don't rush, take your time with each question, and double-check your solutions before handing it in.

For the exam, you may bring one 8.5x11-inch piece of paper, with anything written or typed on it (one side only). You will submit this cheat sheet along with your exam, and you will not be permitted to use any other materials or notes during the exam.

Part One -- Python Programming

Practice Question 1

Which of the options below best demonstrates how to access an object's attribute in Python? Assume your object is **nike** with attribute **material**

- A. nike.material
- B. nike.material()
- C. material.nike
- D. nike=>material
- E. nike(material)
- F. material(nike)

Practice Question 2

Which of the options below best describes what this code snippet does, assuming a variable named d already exists? Select only one.

```
d["abc"].append([4, 5, 6])
```

- A. Looks for an item with the string "abc" in *d*
- B. Adds a new key, value pair ["abc"], [4, 5, 6] to d, which is a dictionary
- C. Appends [4, 5, 6]; d["abc"] now contains a one-dimensional list
- D. Appends [4, 5, 6]; d["abc"] now contains a two-dimensional list
- E. Appends [4, 5, 6] to *d*, which is a list

Practice Question 3

What does the variable **lst** contain after the following code runs?

```
lst = []
for i in range(2):
    lst.append([j for j in range(i, i + 2)])
```

What does the following code snippet print?

```
def func(p1):
    return p1 ** 2
def main():
    x = 4
    func(x)
    print(x * 2)
    print(func(x * 2))
main()
```

Practice Question 5

What does the following code snippet print?

```
def func(lst):
    if len(lst) % 2 == 1:
        return len(lst) // 2
    else:
        return len(lst) // 2 - 1
def main():
    print(func([1, 2, 3, 4, 5]))
    print(func([1, 2, 3, 4]))
main()
```

What does the following code snippet print?

```
class MyClass:
    def __init__(self, a):
        self.a = "a"
    def __str__(self):
        return f"{self.a}!"
obj = MyClass("b")
print(obj)
```

Practice Question 7

Complete the Python function below. Its only parameter is a 2-dimensional list of numbers. Your function should compute the sum of each column of the given list, and return a 1-dimensional list of values. For example, **sum_cols([[4, 7], [3, 5]])** would return **[7, 12]**. You can assume that *lst* is non-empty, and that all rows have the same number of columns.

def sum_cols(lst):

Part Two -- Data Science Math

Practice Question 8

For which dataset below would it be more reasonable to compute the median, rather than the mean? Select the best option.

A. [85, 88, 90, 92, 94, 96, 98]
B. [5, 88, 90, 92, 94, 96, 198]
C. [0, 1]
D. [5, 10, 15, 21, 25]

Practice Question 9

If a dataset has a mean of 14.5, which of the following cannot possibly be the maximum value?

- A. 13.5
- B. 14.5
- C. 15
- D. 1004.5

Practice Question 10

Complete the Python function below. Its only parameter is a dictionary where the keys could be anything and the values are integers. The starter code in the function identifies the maximum value in the dictionary, but there could be more than one value that hits the max. Add on from here so that you return a list of key, value tuples for ALL max-values found in the dictionary.

For example, given {"a" : 5, "b" : 1, "c" : 5} this function should return [("a", 5), ("c", 5)].

You can assume the dictionary is non-empty.

```
def find_all_maxes(dct):
    max_val = max(list(dct.values()))
```



Imagine you have the following dataset representing the scores of students in a test. Apply min-max scaling to the given scores and fill in the table below.

Student	Score	Scaled Score
А	50	
В	0	
С	75	
D	100	
Е	25	

Part Three -- Data Science Data

Practice Question 12

Which distance metric would you use to compare two DNA sequences?

- A. Euclidean distance
- B. Manhattan distance
- C. Hamming distance
- D. Cosine similarity

Practice Question 13

Which of the following correlation coefficients represents the strongest relationship between two variables?

A. r = 0.85B. r = -0.95C. r = 0D. r = .5



Looking at this regression plot, what might be concerning?

- A. The sample size is too large
- B. There might be an influential outlier
- C. The slope is positive
- D. The y-intercept is too high

Practice Question 15

Consider two objects, X and Y, with the following features:

$$X = 3, 7, 12$$

 $Y = 5, 1, 12$

What is the ______ between X and Y? Write your answers in the boxes below.

- A. Euclidean Distance?
- B. Manhattan Distance?



C. Hamming Distance?



What does the following code snippet do? Select the best answer.

```
lst = [["1", "2", "-1.2"], ["", "8", "-1"]]
for i in range(len(lst)):
        for j in range(len(lst[i])):
            lst[i][j] = int(lst[i][j])
```

- A. Creates a new list with numeric values instead of strings
- B. Modifies the existing list with numeric values instead of strings
- C. Ends in a ValueError
- D. Ends in a NameError
- E. Ends in an IndexError

Practice Question 17

The code below is trying to convert all the values in a 2D list from string to numeric. But, it requires two fixes to work as expected. Which two changes do we need to make?

```
lst = [["1", "2", "-1.2"], ["", "8", "-1"]]
for i in range(len(lst)):
        for j in range(len(lst[i])):
            lst[i][j] = int(lst[i][j])
```

- A. Change for i in range(len(lst)): to for row in lst:
- B. Change for j in range(len(lst[i])): to for item in row:
- C. Add if lst[i][j] != "" inside the second loop
- D. Change lst[i][j] = int(lst[i][j]) to lst[i][j] = float(lst[i][j])
- E. Change lst[i][j] = int(lst[i][j]) to float(lst[i][j]) =
 int(lst[i][j])