DSC 10 Practice Midterm Exam From Spring 2021

This was the midterm exam for DSC 10 in Spring 2021. This is a 50 minute exam. For this practice exam, you'll need to access datasets provided in a separate Jupyter notebook. Some questions ask you to manipulate this data. You'll can find the necessary Jupyter notebook on our course website, in the resources section.

This practice exam is different in format to the midterm you'll take, but it might still be good practice. We recommend primarily focusing on the practice exam from Winter 2021 available on Gradescope because it's much more similar in style and format to the exam you'll take.

For the real exam, you can reference anything besides other people, so we recommend taking this exam in the same way you'll take the real exam, without collaborating or communicating with anyone else. We also recommend having the DSC 10 Reference Sheet open while you take the exam. You can find that reference sheet on the course website, in the Resources section.

Question 1

I will complete this exam in a fair, honest, respectful, responsible and trustworthy manner. This means that I will complete the exam as if the professor was watching my every action. I will act according to the professor's instructions, and I will neither give nor receive any aid or assistance other than what is authorized. I know that the integrity of this exam and this class is up to me, and I pledge to not take any action that would break the trust of my classmates or professor, or undermine the fairness of this class.

- ○ I agree

Question 2

You would like to list numbers from 100 to 0, in descending order. Which code will give you the correct solution

- ○ `np.arange(100,-1)`
- ○ `np.arange(1,100,-1)`
- ○ `np.arange(100,-1,-1)`
- ○ `np.arange(0,100,-1)`

Question 3

You would like to build a function that takes in 2 integer arguments (I1, I2), makes a list of all integers between 1 and I1 and another list between 1 and I2, and returns a list of all possible products between members of the 2 lists. Which of the following would return the correct output:

```
def list_of_prods(int1,int2):
    list1 = np.arange(1,int1+1)
    list2 = np.arange(1,int2+1)
    allProds = list1*list2
    return allProds
```

```
def list_of_prods(int1,int2):
    list1 = np.arange(1,int1+1)
    list2 = np.arange(1,int2+1)
    allProds = np.array([])
    for I1 in list1:
        for I2 in list2:
            allProds = np.append(allProds,I1*I2)
    return allProds
```

Both A and B

Neither A or B

Question 4

Using the data and code you have been given, execute the code `streaming.get('show_id')`. This should result in an error. Which of the following best describes the reason for this error?

- `Show_id` is not the index of the dataframe
- You cannot use the `.get()` method on the index of a dataframe
- `show_id` information has been deleted from this dataframe
- None of the above

Question 5

Consider the notebook and data that you have been given. Which of the following best explains why we may want to be cautious using the command `streaming.set_index('country')`

- We've already set the index and this will overwrite the results in a way that we can't recover from.
- We won't be able to access rows by the row index
- The entries in the country column are not unique
- None of the above.

Question 6

Suppose you are given a dataframe called `Numbers` and a function called `squared`. You want to use '`squared`' to square a column in `Numbers` called x. Which of the following are the correct syntax for doing this using the apply method.

- Numbers.get('x').apply(squared())
- Numbers.index['x'].apply(squared)
- Numbers.get('x').apply(squared)
- apply.squared(Numbers.get('x')
- None of the above

Question 7

Using the data table you've been given, execute the following command: `streaming.count().groupby('country')`. This command results in an error. Choose the option which best describes why this error occurred.

- You cannot use `.groupby()` because `streaming` is not a dataframe.
- You cannot use `.count()` because there are not enough entries in `streaming`
- The aggregator, .`count()`, must follow the `.groupby()` statement
- None of the above

Question 8

Suppose you've been given 2 dataframes about dogs: `hair_types`, and `ear_types`. `hair_types` has columns, 'breed', 'color', and 'length', `ear_types` has columns 'dog_breed'', and 'shape'. You would like to combine these two dataframes. Which of the following would do so?

- hair_types.merge(ear_types, left_on='breed', right_index=True)
- hair_types.merge(ear_types, left_on='breed', right_on='dog_breed')
- hair_types.merge(ear_types, left_on='dog_breed', right_on='breed')
- merge(hair_types,ear_types, merge_on = 'breed')

Question 9

You roll a 3-sided die (with faces 1, 2,3) 4 times. What is the probability that you roll a 2 at least once.

- >0.8
- >0.5
- (1/3)^4
- None of the above

Question 10

You flip a fair coin 20 times. You want to know the probability of getting heads exactly 5 times. Which of the following could you do to obtain this answer?

- Calculate this probability by hand
- Do a simulation using `np.random.choice(['H','T'], 20)` and count how many 'H' I get.
- Do a simulation using `np.random.choice(['H','T'], 20)` and count how many 'H' I get many times in a `for`-loop and count the number of time I get 5 'H'
- A & C

Question 11

You want to draw 20 sample numbers uniformly from 1-100 (inclusively), with replacement. You write the following code: `np.random.choice(np.arange(100), 20,replace=True)`

Subsequent testing finds that your statistics from these samples are slightly downward biased (i.e. they are small by a tiny amount). Why?

- np.arange(100) will list numbers from 0-99
- You didn't take enough samples
- You needed to use the statement replace = False
- None of the above

Question 12

Using the given dataframe `taste_testing`, plot a scatterplot of the 3 foods with the amount of compound ‘a’ on the x-axis and the taste rating on the y-axis. Which of the follow accurately describes the way the addition of compound ‘a’ effects the rating of each food:

- Improves food_1
- reduces food_2
- increases food_3 slightly
- All of the above

Question 13

Using the given dataframe `streaming`, determine which country released the **4th** most entries on its own, overall.

- United States
- UK
- Japan
- India

Question 14

Use the `streaming` dataframe that you were given to answer the following question. How many titles were released only in Japan in 2018?

- 5
- 41
- 22
- 35