

Disclaimer: This is the first semester in which Professor Cannon has ever held a final exam for 1002, the result of which is that I and no one other than him has any idea what the content of the final will be like. As a result I cannot say with certainty that the questions you see on this document are going to be like those on the final, what I can guarantee is that I wrote these questions after a deep dive into the lectures from the context, meaning the content presented is stuff you should be aware of whether it is programming related or purely mathematical. I also plan to make this a full length practice exam so there will be 25 multiple choice questions, you will have the class time (technically slightly less) in order to complete the exam so please simulate the actual testing environment when you complete this practice. As always, Good luck!

COMS 1002 Computing in Context Practice Final Exam

Name: _____

Please Use **CAPITAL LETTERS** to fill in the boxes: A B C D E

1.	2.	3.	4.	5.
6.	7.	8.	9.	10.
11.	12.	13.	14.	15.
16.	17.	18.	19.	20.
21.	22.	23.	24.	25.

1. Consider the popular data science module 'Pandas'. How can we enable the use of pandas in our python programs?

- a. `import module pandas`
- b. `from pandas import pd`
- c. `import numpy as np`
- d. `import pandas_data`
- e. `import pandas`

2. Consider the head of the dataframe *df* below, how can we change the index to be the second column from the left?

	name	email	address
0	Donald	donald@magic.kingdom	NaN
1	Goofy	goofy@magic.kingdom	NaN
2	Ariel	ariel@magic.kingdom	NaN
3	Minnie	minnie@magic.kingdom	NaN
4	Mickey	mickey@magic.kingdom	NaN
5	Ursula	ursula@magic.kingdom	NaN
6	Captain Hook	NaN	The seven seas

- a. `df.set_index("email")`
- b. `df.index("email")`
- c. `df.set_index(email.toString())`
- d. `df = df.index("email")`
- e. `df = df.set_index("email")`

3. We have two methods meant for locating information within a dataframe: what are they and what is the primary difference between them?

- a. `.locate()` and `.ilocate()` the former uses labels to find the information and the latter uses integer based indexing
- b. `.loc()` and `.iloc()` the former uses labels to find the information and the latter uses integer based indexing
- c. `.loc[]` and `.iloc[]` the former uses labels to find the information and the latter uses integer based indexing
- d. There is only one way to locate values in a Pandas dataframe
- e. None of the above are correct

4. Consider the following code fragment, what is the output when the code is run in a proper development environment?

```
import numpy as np

a = np.array([1,2,3])
b = np.array([[4,5,6],[7,8,9]])
c = a*b
c = c.reshape(3,2)

print(c)
```

- a. $\begin{bmatrix} 4 & 10 & 18 \\ 7 & 16 & 27 \end{bmatrix}$
- b. $\begin{bmatrix} 4 & 10 & 18 \\ 7 & 16 & 27 \end{bmatrix}$
- c. This code will result in an error
- d. $\begin{bmatrix} 4 & 10 \\ 18 & 7 \\ 16 & 27 \end{bmatrix}$
- e. $\begin{bmatrix} 4 & 10 \\ 18 & 7 \\ 16 & 27 \end{bmatrix}$

5. Recall that we can define functions to help us determine the probability of certain events. In particular we can consider the fact that a standard coin flip can be modeled with a Bernoulli with a probability of success of 0.5. Now consider a biased coin where the probability of getting heads is 0.25, what is the probability that I flip 10 heads in a row?

- a. $(0.5)^{10}$
- b. $(0.75)^{10}$
- c. $(0.25)^{10}$
- d. $(0.25)^{10} / 2^{10}$
- e. $(0.75)^{10} / 2^{10}$

6. Recall your discussion on argument and parameter types in Python. Using that knowledge consider the following code segment and determine what the output is:

```
def foo(a, b, c, *args, d = 0, e = 100):  
    sum = 0  
    for value in args:  
        sum += (value * d)  
    sum -= e  
    sum *= b  
    sum /= a  
    return sum
```

foo(2, 5, 8, 9, 2, 150)

- a. -250.0
- b. 430.0
- c. -375.0
- d. 555.0
- e. The code produces an error

7. Recall from a previous economics lab that we wrote some functions in order to calculate the stock price S after T days using the following formula:

$$S = S_0 Y_1 Y_2 \dots Y_T$$

Specifically recall that you may have used `np.product` to compute the product of a list of values, which of the following is a proper alternative for that specific part of the code.

- a.
 `prod = 1`
 for `val` in `some_list`:
 `prod += val`
- b.
 `prod = 1`
 for `val` in `some_list`:
 `prod *= val`
- c.
 `prod = 1`
 for `val` in `some_list`:
 `prod **= val`
- d. `prod = np.sum(some_list)`
- e. None of the above are algorithmic substitutes

8. When discussing algorithms we have a systematic approach in order to describe how efficient they are. Recall Selection Sort from lecture what is the big- O of that sorting algorithm

- a. $O(n)$
- b. $O(1)$
- c. $O(n^3)$
- d. $O(\log n)$
- e. $O(n^2)$

9. What is the value of `y` after the following code executes:

```
x = "Columbia University in the City of New York"
y = x[4:12:3]
```

- a. "man"
- b. "mane"
- c. "uiU"
- d. "m v"
- e. There is an error in this code

10. Consider some standard pandas data frame `df` and determine what is printed to the console:

```
type(df.iloc[0])
```

- a. <class 'pandas.core.series.Series'>
- b. <class 'pandas.core.frame.DataFrame'>
- c. <class 'numpy.int64'>
- d. <class 'str'>
- e. Cannot be determined

11.