Drexel University
College of Computing and Informatics

INFO 212 – Data Science Programming I

Midterm Answers

B. Questions

1. **[5 points] The statement `a = ['foo', 2, [4, 5]]` defines**

   a. a tuple of length 3 **b.** a list of length 4

   c. a tuple of length 4 **d.** a list of length 3

2. **[5 points] Define a string `s` as `s = 'This_is_a_string'`. The statement `s[6] = 'f'` will**

   a. Replace the character `'i'` in the string to character `'f'`.

   b. Replace the character `'s'` in the string to character `'f'`.

   c. Replace the character `'_'` in the string to character `'f'`.

   d. Produce an error message.

3. **[5 points] What is the result of the following operation:**
   `('foo', 'bar') * 4`

   a. ('foo' * 4, 'bar' * 4) b. ('foo', 'bar', 'foo', 'bar', 'foo', 'bar', 'foo', 'bar')

   c. ('foo4', 'bar4') d. ('4foo', '4bar')

4. **[5 points] If you run the following two statements:**

   ```
   tup = (1, 2, [3, 4])
   a, b, c  = tup
   ```

   the value of the variable c will be:

   a. 3 b. [3, 4]

   c. 4 d. will get an error message.

5. **[5 points] Define a list: `seq = [7, 2, 3, 7, 5, 6, 0, 1]. seq[1:5]` is:**

a. [2, 3, 7, 5] b. [2, 3, 7, 5, 6]

c. [1, 5] d. [2]

6. **[5 points] Define a list: `seq = [7, 2, 3, 7, 5, 6, 0, 1]. seq[:4]` is:**

a. [7, 2, 3, 7, 5] b. [7, 2, 3, 7]

c. [5] d. 5

7. **[5 points] Define a list: `seq = [7, 2, 3, 7, 5, 6, 0, 1]. seq[-4:]` is:**

a. [7, 5, 6, 0, 1] b. [5]

c. [3, -2, -1, 3, 1, 2, -4, -3] d. [5, 6, 0, 1]

8. **Define a list: `seq = [7, 2, 3, 7, 5, 6, 0, 1]. seq[-4:0]` is:**

a. [7, 5, 6, 0, 1] b. []

c. [3, -2, -1, 3, 1, 2, -4, -3] d. [5, 6, 0, 1]

9. **[5 points] Which of the following CANNOT be a key of a dictionary?**

a. an integer b. a string c. a list

d. a tuple

10. **[5 points] Define a numpy array as:**

```
arr = np.array([[1, 2, 3], [4, 5, 6], [7, 8, 9], [10, 11, 12]])
```

Which of the following selects the last two rows?

a. `arr[2:0]`

b. `arr[-2:]`

c. `arr[:, -2:]`

d. `arr[:, 2:]`

11. [5 points] Define a numpy array as:

```
arr = np.array([[1, 2, 3], [4, 5, 6], [7, 8, 9], [10, 11, 12]])
```

Which of the following selects the last two columns?

a. `arr[2:0]`

b. `arr[-2:]`

c. `arr[:, -2:]`

d. `arr[:, 2:]`

12. [5 points] Define a numpy array as:

```
arr = np.array([[1, 2, 3], [4, 5, 6], [7, 8, 9], [10, 11, 12]])
```

Which of the following computes the mean for each column?

a. `arr.mean()`

b. `arr.mean(axis = 0)`

c. `arr.mean(axis = 1)`

d. `arr.mean(axis = 'columns')`

13. [5 points] Define a Pandas DataFrame `frame` as:

	Population	State	Year
0	1.5	Ohio	2000
1	1.7	Ohio	2001
2	3.6	Ohio	2002
3	2.4	Nevada	2001

4	2.9	Nevada	2002
5	3.2	Nevada	2003

Which of the following selects only the states with population greater than 3?

a. `frame.Population > 3` b. `frame[frame.Population > 3]`

c. `frame.Population > 3.State`

d. `frame[frame.Population > 3].State`

14. [5 points] Define a Pandas DataFrame `frame` as:

	Population	**State**	**Year**
0	1.5	Ohio	2000
1	1.7	Ohio	2001
2	3.6	Ohio	2002
3	2.4	Nevada	2001
4	2.9	Nevada	2002
5	3.2	Nevada	2003

Which of the following DOES NOT compute the mean of Population?

a. `frame.mean()` b. `frame.Population.mean()`

c. `frame.mean(axis = 'Population')` d. `frame.mean().Population`

15. [5 points] Define a Pandas DataFrame `frame` as:

	b	**d**	**e**
Utah	0.0	1.0	2.0
Ohio	3.0	4.0	5.0
Texas	6.0	7.0	8.0

Oregon	9.0	10.0	11.0

Define a Pandas Series `s` as:

b	0.0
d	1.0
e	2.0

Which of the following is the result of subtracting `s` from `frame` as `'frame - s'`?

a.

	b	**d**	**e**
Utah	0.0	1.0	2.0
Ohio	2.0	3.0	4.0
Texas	4.0	5.0	6.0
Oregon	9.0	10.0	11.0

b.

	b	**d**	**e**
Utah	0.0	0.0	0.0
Ohio	3.0	3.0	3.0
Texas	6.0	6.0	6.0
Oregon	9.0	9.0	9.0

c. It will produce an error message because you can't subtract a series from a data frame.

d.

	b	**d**	**e**
Utah	0.0	0.0	0.0
Ohio	3.0	4.0	5.0

Texas	6.0	7.0	8.0
Oregon	9.0	10.0	11.0