

Data Science - Assignment 1

Upload an .ipynb file of Assignment 1 to Canvas

Single-choice question (30 points)

1. What is the correct syntax to output "Hello World" in Python?

- A. p("Hello World")
- B. echo "Hello World"
- C. echo("Hello World");
- D. print("Hello World")

2. How do you insert COMMENTS in Python code?

- A. #This is a comment
- B. /This is a comment/
- C. //This is a comment

3. Which one is NOT a legal variable name?

- A. my-var
- B. Myvar
- C. _myvar
- D. my_var

4. How do you create a variable with the numeric value 5?

- A. Both the other answers are correct
- B. x= int(5)
- C. x=5

5. What is the correct file extension for Python files?

- A. .pyth
- B. .py
- C. .pyt
- D. .pt

6. How do you create a variable with the floating number 2.8?

- A. x=float(2.8)
- B. x=2.8
- C. Both the other answers are correct

7.

```
def C2F(c):  
    return c * 9/5 + 32  
print(C2F(100))  
print(C2F(0))
```

What will be the output of the above Python code?

A. 567.0

98.0

B. 212.0

32.0

C.314.0

24.0

8. What is the correct syntax to output the type of a variable or object in Python?

A. print(type(x))

B. print(typeof x)

C. print(typeof(x))

D. print(typeOf(x))

9. What is the correct way to create a function in Python?

A. def myFunction():

B. function myfunction0:

C. create myFunction():

10. In Python, 'Hello', is the same as "Hello"

A. False

B. True

11. Which operator is used to multiply numbers?

A. #

B. X

C. *

D. %

12. Which operator can be used to compare two values?

A. ==

B. <>

C. =

D. ><

13. Which of these collections defines a LIST?

A. {"apple", "banana", "cherry"}

B. ("apple", "banana", "cherry")

C. {"name": "apple", "color": "green"}

D. ["apple", "banana", "cherry"]

14. Which of these collections defines a DICTIONARY?

A. {"apple", "banana", "cherry"}

B. ["apple", "banana", "cherry"]

C. {"name": "apple", "color": "green"}

D. ("apple", "banana", "cherry")

15. What will be the output of the following Python code?

```
places = ['Bangalore', 'Mumbai', 'Delhi']
places1 = places
places2 = places[:]
places1[1]="Pune"
places2[2]="Hyderabad"
print(places)
```

A. ['Bangalore', 'Mumbai', 'Hyderabad']

B. ['Bangalore', 'Pune', 'Hyderabad']

C. ['Bangalore', 'Mumbai', 'Delhi']

D. ['Bangalore', 'Pune', 'Delhi']

16. How do you start writing an if statement in Python?

A. if x > y:

B. if (x > y)

C. if x > y then:

17. How do you start writing a while loop in Python?

A. while x> y:

B. while x > y{

C. x > y while {

D. while (x > y)

18. How do you start writing a for loop in Python?

A. for x in y:

B. for each x in y:

C. for x > y:

19. Which statement is used to stop a loop?

A. stop

B. return

C. break

D. exit

20. What will be the output of the following Python code snippet?

```
x = 'abcd'
for i in range(len(x)):
    print(x)
    x = 'a';
```

A. a

a

- a
- a
- B. abcd
 - abcd
 - abcd
 - abcd
- C. none of the mentioned
- D. a

Programming Exercise (70 points)

Problem 1. Take a list: (20 points)

```
a = [1, 2, 3, 3, 6, 8, 15, 20, 32, 59, 87]
```

and write a program that prints out all the elements of the list that are less than 10.

Extras:

- a. Instead of printing the elements one by one, make a new list that has all the elements less than 10 from this list in it and print out this new list.
- b. Write this in one line of Python.
- c. Ask the user for a number and return a list that contains only elements from the original list `a` that are smaller than that number given by the user.

Problem 2. Write a program that asks the user how many Fibonacci numbers to generate and then generates them. Take this opportunity to think about how you can use functions. Make sure to ask the user to enter the number of numbers in the sequence to generate. (Hint: The Fibonacci sequence is a sequence of numbers where the next number in the sequence is the sum of the previous two numbers in the sequence. The sequence looks like this: 1, 1, 2, 3, 5, 8, 13, ...) (20 points)

Problem 3. CSV File Reading and Writing Exercise (30 points)

1. Download the **Insurance** dataset and use Python's built-in `csv` package to load the data. ([CSC-405-605-705-Data-Science/upload_slides/insurance.csv at main · XiaochenLi-w/CSC-405-605-705-Data-Science](https://github.com/XiaochenLi-w/CSC-405-605-705-Data-Science/blob/master/upload_slides/insurance.csv))
2. Print:
 - The total number of records
 - The column names
3. Compute the following statistics:
 - a. The average value of the `age` column
 - b. The count of records for each **gender**
 - c. The average **age** for each gender
 - d. The average value of `label` for records where
 - `region = "southwest"` **and**
 - `smoker = "yes"`

4. Write all results from **Task c** into a file named `summary.txt`.

```
Average age: 39.21
```

```
Gender count:
```

```
female: 662
```

```
male: 676
```

```
Average age by gender:
```

```
female: 39.50
```

```
male: 38.92
```

```
Average label (region = southwest & smoker = yes): 32269.06
```