Final Exam   Spring 2020

Total 250 points

A.   Multiple choice (60 points)
Choose the best answer from each of the following questions. (each 2 points)

1.   In Python, which of the following will display this result?
     B = 66

a.   print("B = ", char("B"))   c.   print("B = ", ascii("B"))
b.   print("B = ", ord("B"))   d.   print(ord(66))

2. To determine the length of a string that's in a variable named city, you can use this code:

a.   len(city)   c.   length(city)
b.   city.len()   d.   city.length()

3. In Python, to delete all items from a dictionary you can

A).   use the pop() method without any arguments
B).   use the clear() method
C).   use the deleteAll() method
D).   use the del keyword on a dictionary item

4. The cost of insertion in an extremely unbalanced binary search tree can be as high as O(_____).
          a) n
          b) n log n
          c) n^2
          d) $n^2 logn$

5. Suppose you are given two sets in Python:
          set1 = {'Apples', 'Peaches', 'Pineapples'}
          set2 = {'Apples', 'Grapes', 'Peaches'}
   The result of set1- set2 is _____.
          a)   {Peaches}
          b)   {Apples}
          c)   { Pineapples }
          d)   {Apples, Grapes, Peaches, Pineapples}

6. The total time complexity of merge sort is O(___).
          a)   log n
          b)   n
          c)   n log n
          d) n^2

7. The extra memory required to perform merge sort is O(_____).
          a)   log n
          b)   n
          c)   n log n
          d) n^2

8. Each AVL subtree is allowed to be out of balance by ± _____.
          a)   1
          b)   2
          c)   3
          d)   4

9. To count the number of rows in a table, use the SQL construct _____.
   a) SELECT *
   b) SELECT TOP n *
   c) SELECT COUNT(TOP n)
   d) SELECT COUNT(*)
   e) SELECT COUNT *

10. What is the outcome of the following two Python statements
   m = 0.5
   type(m) == int
   a) m becomes 1 because round up
   b) m becomes 0 because round down
   c) it will return True
   d) it will return False because m is a float

11. After the execution of the following Python statement
   b = (3,4,5)
   b contains values 3, 4, and 5. Answer following

   What is the type of b?
   a) set
   b) dict
   c) list
   d) tuple

12   What is the output of the following python code?
   for i in range (1,10):
     if i==5 :
       break
     else :
       print (i, end=' ')

   a) 1 2 3 4
   b) 1 2 3 4 6 7 8 9
   c) 1 2 3 4 6 7 8 9 10
   d) 5

13. The first parameter of a class method is special. It is used in the implementation of the method, but not used when the method is called. Usually it is called _____ in python although any other identifier name can be used
   a) this
   b) me
   c) self
   d) xpointer

14. In python, to make a data field of class private, the variable in the class must be proceeded by
   a) two underscore letter
   b) one underscore letter
   c) declare as private using private keyword
   d) all the variable in python is default to private, so nothing need to be done.

15. Which of the following is not true when isinstance function is used in Python?
   a) isinstance(4, int)
   b) isinstance (4.5e+2, float)
   c) isinstance ((2,3), tuple)
   d) isinstance ((2,3), set)

16. If you have a class named Vehicle in Python, and you want to code a class named Truck that inherits the Vehicle class, you can begin by writing this code:
   a) class Truck(Vehicle):
   b) class Truck : Vehicle
   c) class Truck inherits Vehicle:
   d) class Truck extends Vehicle:

17. The __str__() method of the object class returns the
    a) name of the class and its attributes
    b) name of the class and its identifier
    c) name of the object and its methods
    d) arguments listed in the call

18. In a method of a class in Python, the first parameter, which is usually named self, refers to the current
    a. object
    b. class
    c. method
    d. parameter

19. To access the first three characters in a string that's stored in a variable named message, you can use this code:
    a) first_three = message[0:3]
    b) first_three = message[1:3]
    c) first_three = message.slice(0:2)
    d) first_three = message.split(0:2)

20. The _____ method adds an item to the end of a list.
    a. pop()   b.   append()
    c. insert()   d.   index()

21. Which of the following statements a), b) or c) is *false*?
    a. Scikit-learn conveniently packages the most effective machine-learning algorithms as evaluators.
    b. Each scikit-learn algorithm is encapsulated, so you don't see its intricate details, including any heavy mathematics.
    c. With scikit-learn and a small amount of Python code, you can create powerful models quickly for analyzing data, extracting insights from the data and making predictions.
    d. All of the above statements are *true*.

22. Which of the following statements about scikit-learn and the machine-learning models you'll build with it is *false*?
    a. It's difficult to know in advance which model(s) will perform best on your data, so you typically try many models and pick the one that performs best—scikit-learn makes this convenient for you.
    b. You'll rarely get to know the details of the complex mathematical algorithms in the scikit-learn estimators, but with experience, you'll be able to intuit the best model for each new dataset.
    c. It generally takes at most a few lines of code for you to create and use each scikit-learn model.
    d. The models report their performance so you can compare the results and pick the model(s) with the best performance.

23. Which of the following statements is *false*?
    a. The two main types of machine learning are supervised machine learning, which works with unlabeled data, and unsupervised machine learning, which works with labeled data.
    b. If you're developing a computer vision application to recognize dogs and cats, you'll train your model on lots of dog photos labeled "dog" and cat photos labeled "cat." If your model is effective, when you put it to work processing unlabeled photos it will recognize dogs and cats it has never seen before. The more photos you train with, the greater the chance that your model will accurately predict which new photos are dogs and which are cats.
    c. In this era of big data and massive, economical computer power, you should be able to build some pretty accurate machine learning models.

24. Which of the following statements is *false*?
    a. Even though k-nearest neighbors is one of the most complex classification algorithms, because of its superior prediction accuracy we use it to analyze the Digits dataset bundled with scikit-learn.
    b. Classification algorithms predict the discrete classes (categories) to which samples belong.
    c. Binary classification uses two classes, such as "spam" or "not spam" in an e-mail classification application. Multi-classification uses more than two classes, such as the 10 classes, 0 through 9, in the Digits dataset.
    d. A classification scheme looking at movie descriptions might try to classify them as "action," "adventure," "fantasy," "romance," "history" and the like.

25. Which of the following statements is *false*?
    a.   K-means clustering works through the data attempting to divide it into that many clusters.
    b.   As with many machine learning algorithms, k-means clustering is recursive and gradually zeros in on the clusters to match the number you specify.
    c.   K-means clustering can find similarities in unlabeled data. This can ultimately help with assigning labels to that data so that supervised learning estimators can then process it.
    d.   Given that it's tedious and error-prone for humans to have to assign labels to unlabeled data, and given that the vast majority of the world's data is unlabeled, unsupervised machine learning is an important tool.

26. Which of the following statements a), b) or c) is *false*?
    a.   You should first break your data into a training set and a testing set to prepare to train and test a model.
    b.   The function train_test_split from the sklearn.model_selection module simply splits in order the dataset's samples and target values into training and testing sets. This helps ensure that the training and testing sets have similar characteristics.
    c.   Function train_test_split provides the keyword argument random_state for *reproducibility*. When you run the code in the future with the same seed value, train_test_split will select the same data for the training set and the same data for the testing set. In machine-learning studies, this helps others confirm your results by working with the same randomly selected data.
    d.   All of the above statements are *true*.

27. For NumPy, the attribute _____ contains an array's number of dimensions and the attribute _____ contains a _____ specifying an array's dimensions:
    a.   dim, size, list
    b.   ndim, shape, tuple
    c.   dim, size, tuple
    d.   ndim, shape, list

28. For NumPy, which of the following statements a), b) or c) is *false*?
    a.   NumPy provides functions zeros, ones and full for creating arrays containing 0s, 1s or a specified value, respectively.
    b.   The first argument to the functions in Part (a) must be an integer or a tuple of integers specifying the desired dimensions. For an integer, each function returns a one-dimensional array with the specified number of elements. For a tuple of integers, these functions return a multidimensional array with the specified dimensions.
    c.   The array returned by NumPy function full contains elements with the second argument's value and type.
    d.   All of the above statements are *true*.

29. Assuming the following array grades:

```
import numpy as np
grades = np.array([[87, 96, 70], [100, 87, 90],
        [94, 77, 90], [100, 81, 82]])
```

Which of the following statements a), b) or c) is *false*?
    a.   The following code selects an element in grades by specifying a tuple containing the element's row and column indices in square brackets:

```
grades[0, 1]
```

    b.   To select a single row, specify only one index in square brackets, as in

```
grades[1]
```

    c.   To select multiple sequential rows, use slice notation, as in

```
grades[0:2]
```

    and to select multiple non-sequential rows, use a list of row indices, as in

```
grades[[1, 3]]
```

    d.   All of the above statements are ***true***.

30. Which of the following statements is *false*?

    a.   You can quickly transpose a DataFrame's rows and columns—so the rows become the columns, and the columns become the rows—by using the T attribute.

    b.   T returns a transposed copy of the DataFrame.

    c.   Assuming the following grades DataFrame:

```
       Wally  Eva  Sam  Katie  Bob
Test1    87   100   94   100    83
Test2    96    87   77    81    65
Test3    70    90   90    82    85
```

rather than getting the summary statistics by student, you can get them by test. Simply call describe on grades.T, as in:

grades.T.describe()

    d.   To see the average of all the students' grades on each test, call mean on the T attribute:

grades.T.mean()

B. Short Answer: each 10 points (total 80 points)

1. Write a Python function which computer *prefix_sum* of a sequence. Let's define *prefix_sum* of a sequence is another sequence with each element is the sum from position 0 to its position. For example giving a sequence *a = [1,3,5,7,2,4,6,8]*, then *print (prefix_sum(a))* will output *[1, 4, 9, 16, 18, 22, 28, 36]*

2. Write a python list comprehension to generate the result: [(0, 0), (1, 1), (2, 4), (3, 9), (4, 16)]

3. (a)What is the time complexity for the following relation? (5 points)

$$T(n) = T(n-1) + 1$$

    (b)What is the time complexity for the following relation? (5 points)
$$T(n) = 2T(n/2) + n$$

4. Write a Python function that given a list A, it will return a list with all duplicate element removed and is sorted in ascending order. I.e. all the elements in the final result are unique and sorted. For example, giving *s = [1,3,5,3,4,6,7,8,2,3,4]*, then the *print(unique_and_sorted(s))* will output *[1, 2, 3, 4, 5, 6, 7, 8]*

5. (a) Write a recursive function for f which is define as f(x) = f(x-1) + f(x-2) + f(x//2) and the base cases are f(0) = 0, f(1) = 1, and f(2)= 2.
    (b) Write a decorator call it **dict1** for f so that it can speed up its process by dictionary lookup.
    Test run

```
f(5)
14

f(200)
131726174262146312323428892440010907514618
```

6. Write a python program which will ask the user to enter a sentence. It will process the sentence and output all the words ending with letters 'ed'. The following give a test run of the program and its outputs

```
Enter a sentence: He ended his speech saying 'Your team completed
the project beautifully'
The words ended with ed are:
ended
completed
```

7. Create a 2-by-3 numpy array containing the first six powers of 2 beginning with 2^0. Using Python list comprehension to create one-dimensional list and numpy reshape. Output the numpy array and then output the flatten array of this 2-by-3 array. The following is the output.

8. Write a Python script that uses a dictionary to determine and print the number of duplicated words in a sentence. Treat uppercase and lowercase letters the same and assume there is no punctuation in the sentence.
    Test run
    Assume the input text as following

```
text = ('this is sample text with several words '
        'this is more sample text with some different words')
```

    The output of your program will be

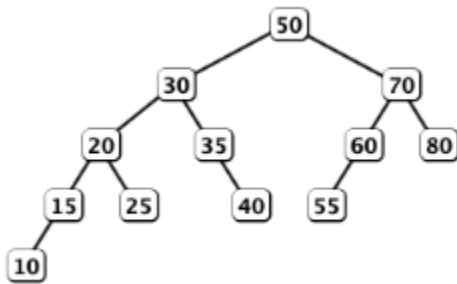| WORD	COUNT
is	2
sample	2
text	2
this	2
with	2
words	2

C. Answer the following questions

1. (30 points, each 10 points) Consider a binary tree which was represented using array. Index 0 is the root of the tree. The content is

0	1	2	3	4	5	6	7	8	9	10	11
96	66	74	37	32	39	20	26	18	28	29	3

(a) Draw this binary tree
(b) Explain why this tree is a max-heap.
(c) Draw the updated heap when 96 was removed

2. Giving the following AVL tree, what is the resultant AVL tree after delete 80? (10 points)

3. Write a class call it **Sentence** which can be initialized by giving a string as input sentence. The delimiter is space. This class contains an constructor, method **get_first_word()** which will return the first word; **get_last_word()** which will return the last word. A method **get_all_words()** will return all the words in list format. The method **replace (index, newWord)** which will replace the word at **index** with the parameter **newWord**. If the index is not valid, your program will output an error message. The following is the test run for this class. (20 points)

```
a = Sentence('I am the great guy')

a.get_first_word()

'I'

a.get_all_words()

['I', 'am', 'the', 'great', 'guy']

a.replace (3, 'handsome')

a.get_all_words()
['I', 'am', 'the', 'handsome', 'guy']

a.get_last_word()

'guy'

a.replace(10, 'pretty')
Error in replace: index out of range.

a.replace(-3, 'pretty')
Error in replace: index out of range.
```
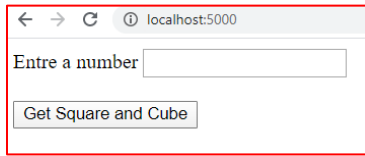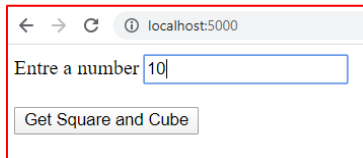
4. Write an HTML document and a simple flask server which when user enter localhost:5000 it will output the following (**30 points**)
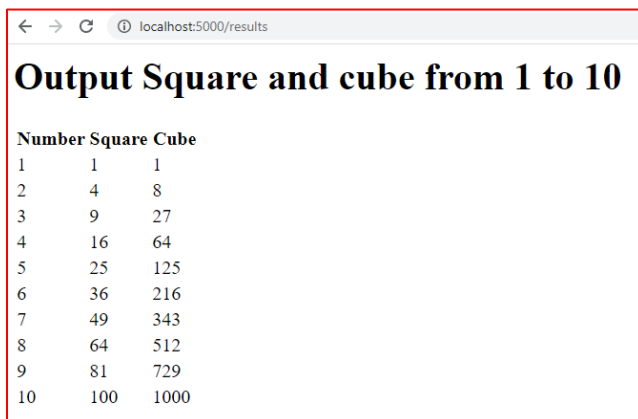

If user entered a number greater than 1 it will output a table on user's browser.

In the following, the user enters 10 and press the button 'Get Square and Cube'
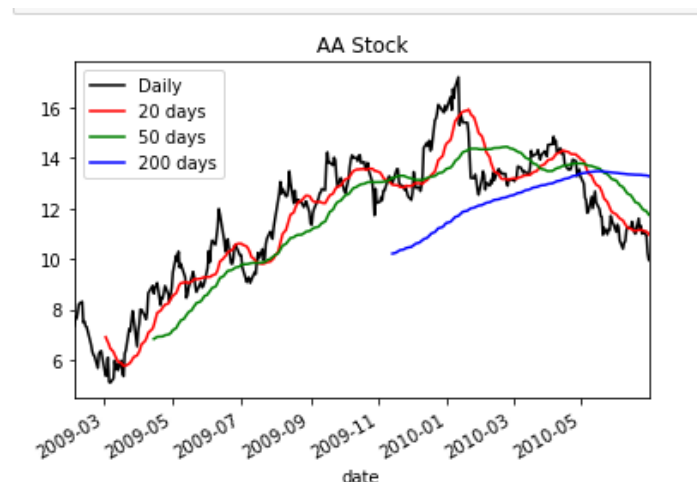

The server will return a table which compute the squares and cubes of numbers from 1 to the number entered by the user. In our example from 1 to 10. Here is the screen shot when user enters 10.

Output Square and cube from 1 to 10

Number	Square	Cube
1	1	1
2	4	8
3	9	27
4	16	64
5	25	125
6	36	216
7	49	343
8	64	512
9	81	729
10	100	1000

5. Write a pandas program which will read in stock.csv file and plot its daily, 20 days, 50 days and 200 days moving average as shown below. The dates are from Feb, 2009 to June, 2010 (20 points)

Step 1: Create a Word file or PDF file and call it **yourname-final-exam-sol,** where **yourname** is yourFirstnameLastname.

Step 2: Write your name on the first line of the first page of the file from Step1

Step 3: Put "MC answers" on a separate line

Step 4: Create a table form like the following. Put your answer under the column 'Answer'

Question	Answer
1	
2	
3	
4	
…	
26	
27	
28	
29	
30	

Step 5: For Short answer questions, start a new page on your yourname-exam2-sol file and mark carefully SA1, SA2,.. ,SA8 and put the answers under each. Make sure you answer them according to the order. I.e., Answer SA1, SA2, SA3, SA4, SA5, SA6, SA7, SA8 in that order.

Note: Since these short answers are Python programs except q3, so create a Python notebook and name it **SA-yourname.ipynb** and then put your answers into this notebook. Where **yourname** is yourfirstNameLastName

Step 6: Create a folder and call it **SRC-yourname** and put the python notebook of Step 5 into this folder.

Step 7: For "Answer the following Questions", start a new page on your yourname-final-exam-sol file. Put the answers of each question in that order. Since Q1 has sub questions from a to c, so mark them carefully as Q1.a, Q1.b, Q1.c and answer them in that order. For Q1, Q2 drawing of binary tree, heap and AVL tree, it is okay for you to use hand-drawing and import pictures into your solution file if you don't have time to create nice diagrams using software.

Step 8: Create a Python notebook for Q3 and put your answer into this notebook. Call this notebook as **Q3-yourname.ipynb**. Put Q3-*yourname*.ipynb into folder **SRC-yourname** created at Step 6.

Step 9: Create an html document and call it **Q4-yourname.html** which contains the answer of Q4 html. Name your flask server code in **Q4-yourname.ipynb** and put it into folder **SRC-yourname** created at Step 6. If you use render_template module then create a folder 'templates' and put the html file into this folder.

Step 10: Create a **Q5-*yourname*.ipynb** for Q5 and put it into folder **SRC-*yourname*** created at Step 6.

Step 11: Zip both **_yourname_-final-exam-sol** file and **SRC-*yourname*** folder into **_yourname_.zip** and submit the **_yourname_.zip** to Blackboard before the due time. Note Blackboard will be automatically closed so budget extra time for submission process.