Analysis with Programming Tools (ITP4869)

Written Test (2017/18)

Time Allowed: 1.5 Hours

(Total Marks : 50)

Student Name: _____ Student ID: _____

Answer any **FIVE** questions out of six. Each question contributes to 10 marks.

Q1 (a) One of the features of the Python language is *interpreted*. List **THREE** other features. (3 marks)

(b) Write a single statement to swap the values of two variables. (2 marks)

(c) Define a function, named `area`, that takes a radius as its argument and returns the area of the circle with this radius. (5 marks)

Q2 (a) Write a program that reads an integer, `n`, and prints three consecutive square numbers as a list, starting from the nth one. For example, the input 5 will lead to the list of square numbers, `[25, 36, 49]` (5 marks)

(b) Consider the list, `jargon`, as below. Sort the list, reverse it and then print the result. (5 marks)

```
jargon = ['python', 'spark', 'hadoop', 'rdd']
```

Q3 Define a Python class, `Rectangle`, with two attributes, `length` and `width`. Add a constructor to `Rectangle` to initialize a `Rectangle` object and add the following two methods: (10 marks)

(a) `area`: returns the area of a rectangle; and

(b) `perimeter`: returns the perimeter of a rectangle (i.e., total length of a rectangle).

Q4 Write Python scripts with NumPy to implement the following tasks. You may assume the import statement "`import numpy as np`" has been executed.

(a) Create a two dimensional 3×4 array named, `narr`, storing integers between 11 and 22. (3 marks)

(b) Take the elements from the last column of `narr` to form an array called `narr2`. (2 marks)

(c) Extract those elements from `narr` that are divisible by 3 and store in `narr3`. (2 marks)

(d) Transpose the array `narr2`, compute the square of each element in it, and store it in `narr4`. (3 marks)

Q5 Suppose you are given a CSV file, named `scores.csv`, storing the students' scores of three modules.

SID	Java	Python	Hadoop
18001	88	82	79
18002	82	75	74
18003	64	72	51
18004	53	85	60
18005	74	94	87

You are also given the following import statements:

```
import numpy as np
import pandas as pd
from pandas import Series, DataFrame
```

(a) Load the CSV file and store the content in DataFrame `df` with the `SID` column as its index column. (3 marks)

(b) Extract the students' scores with SID 18002 and 18005. (2 marks)

(c) Show students' records whose Python score > 80. (2 marks)

(d) Add a column, called "Total", to `df`, for storing the total score of the three modules of each student. (3 marks)

Q6 Consider the following three Series:

```
s1 = pd.Series([0, 1], index=['a', 'b'])
s2 = pd.Series([2, 3, 4], index=['c', 'd', 'e'])
s3 = pd.Series([5, 6], index=['f', 'g'])
```

You are also given the three import statements as in Question Q5.

(a) Concatenate these three Series to form a DataFrame, `df2`, with three columns labelled as '`cat1`', '`cat2`', and '`cat3`'. (3 marks)

(b) Replace all the `NaN` values in `df2` by letter '`X`' and store the result in `df3`. (2 marks)

(c) Extract the four rows with index '`c`' to '`f`' in `df3` and store them in `df4`. (2 marks)

(d) Update the column labels in `df4` by capitalizing them. (3 marks)

`df2`

	cat1	cat2	cat3
a	0.0	NaN	NaN
b	1.0	NaN	NaN
c	NaN	2.0	NaN
d	NaN	3.0	NaN
e	NaN	4.0	NaN
f	NaN	NaN	5.0
g	NaN	NaN	6.0

`df3`

	cat1	cat2	cat3
a	0	X	X
b	1	X	X
c	X	2	X
d	X	3	X
e	X	4	X
f	X	X	5
g	X	X	6

`df4`

	cat1	cat2	cat3
c	X	2	X
d	X	3	X
e	X	4	X
f	X	X	5

`Updated df4`

	Cat1	Cat2	Cat3
c	X	2	X
d	X	3	X
e	X	4	X
f	X	X	5

******* END OF PAPER *******