Week-9, Practice, Theory

Problem 1

Common data for questions 1, 2, 3 and 4

Consider the following file named `test.txt`

```
11111
22222
33333
44444
55555
66666
```

Question 1

Code-1 and Code-2 are two separate code-blocks that are executed independently of each other.

Code-1

```
file = open("test.txt",'r')
a = file.readline()
b = file.readlines()
c = b[2]
file.close()
```

Code-2

```
file = open("test.txt",'r')
d = file.readlines()
e = file.readline()
f = d[0]
file.close()
```

Which of the following expressions return `True` at the end of execution of above code blocks `code-1` and `code-2` ? It is MSQ type question.

(a) `a == f`

(b) `a == e`

(c) `b == d`

(d) `c == d[3]`

(e) `b[5] == d[5]`

(f) `e == ""`

Answer

(a), (d) and (f)

Solution

After executing of code-1 and code-1:-

```
a = 11111\n
b = ['22222\n', '33333\n', '44444\n', '55555\n', '66666']
c = 44444\n
d = ['11111\n', '22222\n', '33333\n', '44444\n', '55555\n', '66666']
e = ''
f = 11111\n
```

So only option (a), (d) and (e) will return True.

Question 2

```
def isFileClosed():
    f = None
    f = open("test.txt",'r')
    return f.closed
```

What will the function `isFileClosed()` return?

(a) `True`

(b) `False`

(c) `None`

(d) It will throw an error

Answer

(b)

Solution

`f.closed` checks whether the file is closed or not. In the above function, the file is not closed so it returns `False`

Question 3

```
def appendFile():
    f = open("test.txt",'a')
    f.writelines('77777')
    f.close()
    f = open("test.txt",'r')
    c = f.readlines()
    f.close()
    return c[6]
```

What will the function `appendFile()` return?

(a) `66666`

(b) `77777`

(c) `6666677777`

(d) It will throw an error.

Answer

(d)

Solution

`appendFile` function will append `77777` in the last line of the file which is `66666`. After appending it became `6666677777`. `c = f.readlines()` reads all lines of the file and returns a list of lines to `c`, so in the list `c` has only 6 lines (index 0 to 5) then due to this `return c[6]` will throw an error.

Question 4

Which of the following statements are true regarding the opening modes of a file? It is MSQ type question.

(a) For `f = open("test.txt",'r')`, if the file `test.txt` does not exist, an error occurs.

(b) For `f = open("test.txt",'w')`, if the file `test.txt` does not exist, an error occurs.

(c) For `f = open("test.txt",'r')`, if the file `test.txt` does not exist, a new file `test.txt` is created.

(d) For `f = open("test.txt",'w')`, if the file `test.txt` does not exist, a new file `test.txt` is created.

(e) For `f = open("test.txt",'w')`, if the file `test.txt` exists, the existing file is overwritten with the new file.

(f) For `f = open("test.txt",'a')`, if the file `test.txt` does not exist, a new file `test.txt` is created.

Answer

(a), (d) ,(e) and (f)

Solution

Self-explanatory

Problem 2

Common data for question 5, 6 and 7.

`the_road_not_taken.txt`

```
 1  Two roads diverged in a yellow wood,
 2  And sorry I could not travel both
 3  And be one traveler, long I stood
 4  And looked down one as far as I could
 5  To where it bent in the undergrowth;
 6
 7  Then took the other, as just as fair,
 8  And having perhaps the better claim,
 9  Because it was grassy and wanted wear;
10  Though as for that the passing there
11  Had worn them really about the same,
12
13  And both that morning equally lay
14  In leaves no step had trodden black.
15  Oh, I kept the first for another day!
16  Yet knowing how way leads on to way,
17  I doubted if I should ever come back.
18
19  I shall be telling this with a sigh
20  Somewhere ages and ages hence:
21  Two roads diverged in a wood, and I
22  I took the one less traveled by,
23  And that has made all the difference.
```

Question 5

What is the value stored in the variable `lineCount` at the end of execution? It is a Numerical Answer Type Question (NAT).

```python
1  f = open('the_road_not_taken.txt', 'r')
2  line = f.readline().strip()
3  lineCount = 0
4  while line:
5      lineCount += 1
6      line = f.readline().strip()
7  f.close()
```

Answer

5

Solution

line becomes null string character at line 6 after the `'\n'` is removed using strip. Hence, while stops at the fifth iteration where the `lineCount` is 5.

Question 6

Choose the correct function `getWordCount` to store the total number of words in the file in `wordCount`.

```python
f = open('the_road_not_taken.txt', 'r')
line = f.readline()
wordCount = 0
while line:
    wordCount += getWordCount(line.strip())
    line = f.readline()
f.close()
```

(a)

```python
def getWordCount(s):
    for i in ',:;\n':
        while i in s:
            s = s.replace(i, ' ')
    while ' '*2 in s:
        s = replace(' '*2, ' ')
    return s.count(' ')+1
```

(b)

```python
def getWordCount(s):
    for i in '.,:;!\n':
        while i in s:
            s = s.replace(i, ' ')
    while ' '*2 in s:
        s = s.replace(' ', '')
    if s != '':
        return s.strip().count(' ')+1
    else:
        return 0
```

(c)

```python
def getWordCount(s):
    for i in '.,:;!\n':
        while i in s:
            s = s.replace(i, ' ')
    while ' '*2 in s:
        s = s.replace(' '*2, ' ')
    return s.strip().count(' ')+1
```

(d)

```
def getWordCount(s):
    for i in '.,:;!\n':
        while i in s:
            s = s.replace(i, ' ')
    while ' '*2 in s:
        s = s.replace(' '*2, ' ')
    if s != '':
        return s.strip().count(' ')+1
    else:
        return 0
```

Answer

(d)

Solution

Option	Comment
(a)	Preceding and trial spaces will be counted .
(b)	Every spaces are removed, hence counting words becomes impossible.
(c)	If the null string is given then the incorrect output 1 is given.
(d)	Includes all the aspects.

Question 7

```python
f = open('the_road_not_taken.txt', 'r')
line = f.readline()
c = 0
while line:
    if line != '\n':
        c += 1
    line = f.readline()
f.close()
```

What does `c` represent at the end of execution?

(a) Number of lines in the file.
(b) Number of non-empty lines in the file excluding the new line character '\n'.
(c) Number of empty lines in the file.
(d) Number of duplicate lines in the file.

Answer

(b) Number of non-empty lines in the file.

Solution

`f.readline()` gives `'\n'` for empty lines. Hence, `c` holds the count of non-empty lines with reference to the common data.

Problem 3

Common data for question 8 and 9.

`file1.txt` is a file having some text and `file2.txt` does not exist before.

```python
f1 = open('file1.txt', 'r')
f2 = open('file2.txt', 'w')
line = ' '
for line in f1.readlines():
    f2.write(line.replace('\t', ' '*4).rstrip()+'\n')
f1.close()
f2.close()
```

Question 8

Choose all the correct statements. It is a Multiple Select Question (MSQ).

(a) All the tabs are always replaced by 4 spaces.
(b) All the spaces at the end of the lines are removed in `file2.txt`.
(c) All the spaces at the beginning of the lines are removed in `file2.txt`.
(d) The number of lines in `file1.txt` and `file2.txt` are the same.
(e) `file2.txt` has only one line.

Answer

(b), (d)

Solution

Option	Comment
(a)	Incorrect, `line.replace('\t', ' '*4)` is executed only one, hence it replaces the first found tab alone.
(b)	Correct, `rstrip()` removes all the trial spaces.
(c)	Incorrect, there is no manipulation of the preceding spaces.
(d)	Correct, All the lines are present, even the empty lines remains empty.

Question 9

Assume the following code-snippet is executed after the main code. What will be the evaluated value of the variable `check`?

```python
f1 = open('file1.txt', 'r')
f2 = open('file2.txt', 'r')
check = f1.readline() == f2.readline()
f1.close()
f2.close()
```

(a) True, if the first line of `file1.txt` and the second line of `file2.txt` are the same.
(b) True, if the first line of `file1.txt` and the first line of `file2.txt` have no tabs
(c) True, if the first line of `file1.txt` and the first line of `file2.txt` have no tabs and no spaces at the end.
(d) False, if the first line of `file1.txt` and the first line of `file2.txt` have no tabs and no spaces at the end.

Answer

(c) True, if the first line of `file1.txt` and first line of `file2.txt` have no tabs and no spaces at the end.

Solution

`f1.readline()` and `f2.readline()` reads the first line of `file1.txt` and `file2.txt` respectively. `check` will True when there are no trial spaces and tab characters because the preceding space will be removed and the first tab character is replaced by four spaces.

Problem 4

Question 10 , 11, 12 and 13 are based on `scores_dataset.csv`. The data is read into Pandas dataframe variable `data_df`. Assume `pandas` library is already imported into the program as `pd`.

Question 10

Which of the following options can be used to read a `scores_dataset.csv` file into a pandas dataframe variable `data_df`.

(a)

```
data_df = readcsv('scores_dataset.csv')
```

(b)

```
data_df = pd.read_csv('scores_dataset.csv')
```

(c)

```
data_df = pd.readcsv('scores_dataset.csv')
```

(d)

```
data_df = pd.readCsv('scores_dataset.csv')
```

Answer

(b)

Solution

The valid way to create dataframe in Pandas from a csv file is:

```
pd.read_csv('scores_dataset.csv')
```

Question 11

Given a Pandas dataframe variable `data_df`, match the following `Methods` to `Descriptions`.

Methods	Descriptions
1. `data_df.sort_values(by = ['Total'])`	A. Count non empty (non missing) values in column `'DateOfBirth'`.
2. `data_df.shape`	B. Group `score` dataset on a `'Gender'` column.
3. `data_df['DateOfBirth'].count()`	C. Arrange dataset `data_df` on the column `Total` in `asc`/`desc` order.
4. `data_df.groupby('Gender').groups`	D. Returns the sum of the values in the column `'Total'`.
5. `data_df['Total'].sum()`	E. Return a tuple representing the dimensionality of the dataframe `data_df`.

(a) 1-D, 2-E, 3-B, 4-A, 5-C

(b) 1-C, 2-A, 3-E, 4-D, 5-B

(c) 1-B, 2-A, 3-E, 4-C, 5-D

(d) 1-C, 2-E, 3-A, 4-B, 5-D

Answer

(d)

Solution

- `data_df.sort_values(by = ['Total'])`: `sort_values` function sorts the `data_df` on a given column
- `data_df.shape`: It gives the dimension of the data as tuple `(num_of_rows, num_of_columns)`
- `data_df['DateOfBirth'].count()`: It counts the number of elements in column `'DateOfBirth'` of `data_df`
- `data_df.groupby('Gender').groups`: Group index of score dataset `data_df` on `Gender` (`M`and `F`)
- `data_df['Total'].sum()`: Gives the sum of all the values in the column `'Total'`.

Question 12

Which of the following ways give the number of rows and the number of columns separated by comma in the variable `data_df`?

(a) `data_df.shape[0], data_df.shape[1]`

(b) `data.count().max(), len(data.columns)`

(c) `len(data), len(list(data))`

(d) None of these

Answer

(a)

Solution

Option (a) is correct. The function `shape()` gives a tuple where the first value represents the number of rows and the second value represents the number of columns of the score dataframe `data_df`.

Option (b) and option (c) is not correct as it is using incorrect variable `data` which is not defined. The following would have been valid ways to get the number of rows and number of columns:

- `data_df.count().max(), len(data_df.columns)`
- `len(data_df), len(list(data_df))`

Question 13

Which of the following holds true about the output of the following code. It is a Multiple Select Question (MSQ).

```
data_df[(data_df['Total'] > 180) & (data_df['Physics'].between(50, 100)) & (data_df['Chemistry'] < 50)].shape[0]
```

(a) Student's marks is below `50` in `'Physics'`

(b) Student's Total marks is more than `180`

(c) The output is `1`

(d) Students scored less than `50` in `Chemistry`

Answer

(b), (c), (d)

Solution

Options (b), (c), (d) are True.

- `(data_df['Total'] > 180)`: Extracts student records where a student has `'Total'` score more than 180
- `(data_df['Physics'].between(50, 100))`: Extracts student records where a student has `'Physics'` score between `50` and `100`
- `(data_df['Chemistry'] < 50)].shape[0]`: Extracts student records where a student has `'Chemistry'` score below `50`

Since all these conditions are joined with `&`, all must be `True` in the resulting records. Only one record satisfies this condition, which is given below.

SeqNo	Name	Gender	DateOfBirth	CityTown	Mathematics	Physics	Chemistry	Total
16	16	Tauseef	M	30 Dec	Trichy	87	86	43