Question 1 [10Marks]

1. Write a Pandas program to find and replace the missing values in a given DataFrame which do not have any valuable information like NaN, ? and –

```
df.replace({"?": np.nan, "--": np.nan})
df.fillna(0)
```

2. Write a Pandas program to rename columns of a Data Frame. **(max 2 lines)**

```
df.rename(columns={'Runtime (Minutes)':'Runtime', 'Revenue (Millions)':'Revenue' },inplace=True)
```

3. Write down a panda's function to get the Non-Null Count and Data Type for Every Column. **(1 line)**

df.info()

4. Which exception is used when Element/item is not found on webpage?

NoSuchElementException

5. Write down a function to rotate the image.

np.flipud(image1)

Question 2 [7 marks]

1. What will be the output produced by the following programming statements 1 & 2?

```
import pandas as pd
S1=pd.Series(data=[31,41,51])
print(S1>40) -->Statement1
print(S1[S1>40]) -->Statement2
```

Output:

Statement 1:

```
Statement1-
0    False
1     True
2     True
dtype: bool
```

Statement 2:

```
Statement2-
1    41
2    51
dtype: int64
```

2. Given two series S1 and S2

	S1		S2
A	39	A	10
B	41	B	10
C	42	D	10
D	44	F	10

Find the output for following python pandas statements? Write your output along with the indexes.

a. S1[: 2]*100

```
OUTPUT:
A    3900
B    4100
dtype: int64
```

b. S1 * S2

```
A    390.0
B    410.0
C      NaN
D    440.0
F      NaN
dtype: float64
```

c. S2[: : -1]*10

```
F    100
D    100
B    100
A    100
dtype: int64
```

3. Consider the following code

```
import pandas
s1=pandas.Series([2,3,4,5,6],index=['a','b','c','d','e'])
s1[1:5:2]=345.6
s1[2:4]= -14.65
print(s1)
```

What will be the output after executing the code.

```
a      2.00
b    345.60
c    -14.65
d    -14.65
e      6.00
dtype: float64
```

Question 3[17 marks]

1. Read the following code:

```
x = ['XX', 'YY']
   for i in a:
       i.lower()
   print(a)
```

What will be the output of this program?

a) ['XX', 'YY']
b) ['xx', 'yy']
c) [XX, yy]
d) None of these

2. Read the following statement:
print(0xA + 0xB + 0xC)

What will be the output of this statement?

a) 33
b) 63
c) 0xA + 0xB + 0xC

d) None of these

3. What happens when '2' == 2 is executed?

a) False

b) Ture

c) ValueError occurs

d) TypeError occurs

4. What will be the minimum number of arguments require to pass in pandas series?

a) 2

b) 3

c) 4

d) None of the above mentioned

5. During the execution of following code, what will be the response, we get

```
import pandas as pd
  s =pd.Series([1,2,3,4,5],index= ['a','b','c','d','e'])
  print(s['f'])
```

a) KeyError

b) IndexError

c) ValueError

d) None of the above mentioned

6. Observe the following code and identify what will be the output when we run following code:

```
import pandas as pd
  import numpy as np
  df = pd.DataFrame(np.array([[4,6,9],[5,1,3]]))
  print(df.shape)
```

a) SyntaxError: invalid syntax

b) KeyError

c) IndexError

d) None of the mentioned above

7. What will be output of following code?

```
import numpy as np
   array1=np.array([100,200,300,400,500,600,700])
   print(array1[1:5:2])
```

a) [200 300]

b) [200 700]

c) [200 400]

d) [200 500]

8. To display histogram with well-defined edge we can write:

a) df.plot(type = 'hist', edge = 'red')

b) df.plot(type = 'hist', edgecolor = 'red')

c) df.plot(type = 'hist', line = 'red')

d) df.plot(type = 'hist', linecolor = 'red')

9. What is the best call for finding multiple elements using XPath?

a) findElementByXpath

b) findElementsByXPath

c) findElementByCssSelector

d) Both B & C

10. Selenium command for entering text into text boxes?
a) sendKeys()
b) sendKey()
c) sendKey
d) SendsKeys()

11. Consider the below XPath and select the correct meaning

//li[@id='firstItem']//following::a
a) Find all the anchor tags and a list item whose id is 'firstItem'
b) Find a list item whose id is 'firstItem' and falls after an anchor tag
c) Find all the anchor tags in the entire DOM which are falling after the list item whose id is 'firstItem'

12. Which method to use to get an element from webpage?
a) getElement()
b) searchElement()
c) findElements()
d) findElement()

13. N-grams are defined as the combination of N keywords together. How many bi-grams can be generated from the given sentence: Quaid Azam is the father of our nation
a) 6
b) 7
c) 8
d) 5

14. Which of the following techniques can be used for the purpose of keyword normalization, the process of converting a keyword into its meaningful base form?
a) **Lemmatization**
b) Lowenstein distance
c) Morphing
d) Stemming

15. In NLP, the process of converting a sentence or paragraph into tokens is referred to as Stemming
a) True
b) False

16. Machine Translation is that converts -
a) Human language to machine language
b) One human language to another
c) Any human language to English
d) Machine language to human language

17. If the skewness value of data is 6.1 the data is:
a) Highly skewed
b) Fairly symmetrical
c) Moderately skewed
d) Not skewed

Question 4[7 marks]

Name any two text stemmers. **[2]**

Snowball Stemmer
Lancaster Stemmer
Porter Stemmer

What is the difference between Stemming and Lemmatization? **[2]**

Stemming usually operates on word without knowledge of the context.
Lemmatization usually considers words and the context of the word in the sentence

Write a code to remove all punctuations from the text.**[2]**

```
import string
punctuation = string.punctuation
for token in nltk.word_tokenize(text):
    if token not in punctuation:
        return token
```

Name any 2 image feature extraction techniques? **[1]**

Canny
Hog
Pixel Features
Sobel

Question 5[14 marks]

Given below is the code to create a Pandas Data Frame. Answer all the questions based on this data. Mention the output/ Error if any [No need for long explanation].

```
import pandas as pd
import numpy as np
x= np.arange(16).reshape(4,4)*2

data= pd.DataFrame(x,
                   index = ['Lahore', 'Karachi', 'Peshawar', 'Islamabad'],
                   columns = ['a', 'b',"c","d"])
```

Write the output of following operation when called on the "data" dataframe.

a) What will be the output of the following line of code [Show in table format]
data[:]

	a	b	c	d
Lahore	0	2	4	6
Karachi	8	10	12	14
Peshawar	16	18	20	22
Islamabad	24	26	28	30

b) What will be the output of the following line of code
data.loc['Lahore':'Peshawar']

	a	b	c	d
Lahore	0	2	4	6
Karachi	8	10	12	14
Peshawar	16	18	20	22

c) What will be the output of the following line of code
data.loc['Lahore':'Peshawar'][: -2]

	a	b	c	d
Lahore	0	2	4	6

d) What will be the output of the following line of code
data.iloc[-2:-1 , 2:3]

	c
Peshawar	20

Given below is a Pandas Data Frame named df. Answer all the questions based on this data. Mention the output/ Error if any [No need for long explanation].

id	country	sex	height	weight	discipline	gold	silver	bronze
736041664	None	male	1.72	64.0	athletics	0	0	0
532037425	None	female	1.68	56.0	fencing	0	0	0
435962603	CAN	male	1.98	79.0	None	0	0	1
521041435	MDA	male	1.83	80.0	None	0	0	0
33922579	NZL	male	1.81	71.0	cycling	0	0	0

a) What will be the output of the following line of code

df.isnull().any()

```
3]:  1  athletes.isnull().any()
3]: id             False
    nationality     True
    sex            False
    height         False
    weight         False
    sport           True
    gold           False
    silver         False
    bronze         False
    dtype: bool
```

b) What will be the output of the following line of code

df.isnull().sum(axis=1)

```
: 0    1
  1    1
  2    1
  3    1
  4    0
  dtype: int64
```

c) What will be the output of the following line of code

df.isnull().any(axis=1).sum()

```
: 1 athletes.isnull().any(axis=1).sum()
: 4
```