# PYTHON – WORKSHEET 9 (PANDAS)

## Q1 to Q8 have only one correct answer. Choose the correct option to answer your question.

1. Which among the following options can be used to create a DataFrame in Pandas?
   1. An ndarray B) a python dictionary

C) A scalar value D) All of the above

Answer: (D )

1. A series is a one-dimensional array which is labelled and can hold any data type.
   1. True B) False

Answer (A )

1. Which of the following operation works with the same syntax as the analogous dictionary operations?
   1. Getting columns B) setting columns

C) deleting columns D) All of the above

Answer: ( D )

1. pandas.NA = = pandas.NA, will give which of the following result?
   1. <NA> B) True

C) False D) Error

Answer: ( C )

1. A panel is a container of data in pandas?
   1. 1 dimensional B) 2 dimensional

C) 3 dimensional D) infinite dimensions

Answer: ( C )

1. What will be the output of the following lines of code? import pandas as pd

import numpy as np

s = pd.Series(np.random.randn(4)) print(s.ndim)

* 1. Error B) 3

C) 2 D) 1

Answer: ( D )

1. Which of the following indexing capabilities is used as a concise means of selecting data from a pandas object??
   1. in B) iy

C) ix D) ipy

Answer: ( C )

1. All pandas data structures are mutable but not always mutable.
   1. size, value B) value, size

C) semantic, size D) None of these

Answer: ( C )

## Q9 and Q10 have multiple correct answers. Choose all the correct options to answer your question.

1. Select the correct statements from the following.
   1. A DataFrame is like a fixed-size dictionary in that you can get and set values by index label.
   2. Series can be passed into most NumPy methods expecting an ndarray.
   3. A key difference between Series and ndarray is that operations between Series automatically align the data based on label
   4. In pandas, Index values must be unique

Answer: ( B ) ( D )

1. Which of the following file formats are allowed for input output in pandas?
   1. JSON B) HTML

C) CSV D) TXT

Answer: ( C ) ( D )

## Q11 to Q15 are programming questions. Answer them in Jupyter Notebook.

1. Write a Pandas program to create and display a DataFrame from the following dictionary data and labels: exam\_data = {'name': ['Anastasia', 'Dima', 'Katherine', 'James', 'Emily', 'Michael', 'Matthew', 'Laura', 'Kevin',

'Jonas'],

'score': [12.5, 9, 16.5, np.nan, 9, 20, 14.5, np.nan, 8, 19],

'attempts': [1, 3, 2, 3, 2, 3, 1, 1, 2, 1],

'qualify': ['yes', 'no', 'yes', 'no', 'no', 'yes', 'yes', 'no', 'no', 'yes']}

labels = ['I’, ‘II’, ‘III’, ‘IV’, ‘V’, ‘VI’, ‘VII’, ‘VIII’, ‘IX’, ‘X’]

Program:

import pandas as pd

import numpy as np

exam\_data = {'name': ['Anastasia', 'Dima', 'Katherine', 'James', 'Emily', 'Michael', 'Matthew', 'Laura', 'Kevin', 'Jonas'],

'score': [12.5, 9, 16.5, np.nan, 9, 20, 14.5, np.nan, 8, 19],

'attempts': [1, 3, 2, 3, 2, 3, 1, 1, 2, 1],

'qualify': ['yes', 'no', 'yes', 'no', 'no', 'yes', 'yes', 'no', 'no', 'yes']}

labels = ['I’, ‘II’, ‘III’, ‘IV’, ‘V’, ‘VI’, ‘VII’, ‘VIII’, ‘IX’, ‘X’]

df = pd.DataFrame(exam\_data , index=labels)

print("Summary of the basic information about this DataFrame and its data:")

print(df.info())

1. Write a Pandas program to get the first 5 rows of the DataFrame created in Q11.

Program:

exam\_data = {'name': ['Anastasia', 'Dima', 'Katherine', 'James', 'Emily', 'Michael', 'Matthew', 'Laura', 'Kevin', 'Jonas'],  
'score': [12.5, 9, 16.5, np.nan, 9, 20, 14.5, np.nan, 8, 19],  
'attempts': [1, 3, 2, 3, 2, 3, 1, 1, 2, 1],  
'qualify': ['yes', 'no', 'yes', 'no', 'no', 'yes', 'yes', 'no', 'no', 'yes']}  
labels = ['I’, ‘II’, ‘III’, ‘IV’, ‘V’, ‘VI’, ‘VII’, ‘VIII’, ‘IX’, ‘X’]

1. Write a Pandas program to select the 'name' and 'score' columns of the DataFrame created in Q11.

Program:

exam\_data = {'name': ['Anastasia', 'Dima', 'Katherine', 'James', 'Emily', 'Michael', 'Matthew', 'Laura', 'Kevin', 'Jonas'],  
'score': [12.5, 9, 16.5, np.nan, 9, 20, 14.5, np.nan, 8, 19],  
'attempts': [1, 3, 2, 3, 2, 3, 1, 1, 2, 1],  
'qualify': ['yes', 'no', 'yes', 'no', 'no', 'yes', 'yes', 'no', 'no', 'yes']}  
labels = ['I’, ‘II’, ‘III’, ‘IV’, ‘V’, ‘VI’, ‘VII’, ‘VIII’, ‘IX’, ‘X’]

1. Write a Pandas program to select 'name' and 'score' columns in row indexes 3, 5, 6, 8 from the DataFrame created in Q11.
2. Write a Pandas program to select the rows where the score is between 15 and 20 (inclusive) from the DataFrame created in Q11.

Program:

exam\_data = {'name': ['Anastasia', 'Dima', 'Katherine', 'James', 'Emily', 'Michael', 'Matthew', 'Laura', 'Kevin', 'Jonas'],  
'score': [12.5, 9, 16.5, np.nan, 9, 20, 14.5, np.nan, 8, 19],  
'attempts': [1, 3, 2, 3, 2, 3, 1, 1, 2, 1],  
'qualify': ['yes', 'no', 'yes', 'no', 'no', 'yes', 'yes', 'no', 'no', 'yes']}  
labels = ['I’, ‘II’, ‘III’, ‘IV’, ‘V’, ‘VI’, ‘VII’, ‘VIII’, ‘IX’, ‘X’ labels = ['I’, ‘II’, ‘III’, ‘IV’, ‘V’, ‘VI’, ‘VII’, ‘VIII’, ‘IX’, ‘X’]