ROLL NO _____

DELHI SKILL AND ENTREPRENEURSHIP UNIVERSITY

DSEU Dwarka Campus, Sector-9, New Delhi -77; Website: dseu.ac.in
Semester - IV Sem
End Term Exam June 2023
Subject Name: Data Sciences using Python
Subject Code: BCA-EC401

Time: 1.5 HOURS **TOTAL MARKS: 25**

SECTION – A (All Questions Are Compulsory) **(5 * 1 = 5)**

Q1) Which of the following signifies usage of Code cell and Markdown cell in Jupyter Notebook respectively. **(T)**

a. Writing text, writing code
b. displaying visualizations, text display
c. writing code, writing text
d. writing equations, displaying data.

Q2) Separator characters used for separating values of CSV files are called as: **(T)**

a. Comma
b. Delimiter
c. Parameter
d. Spacer

Q3) In accordance with the given code, the DataFrame df has ___ rows, ___ columns, and ____ NaN values. **(A)**

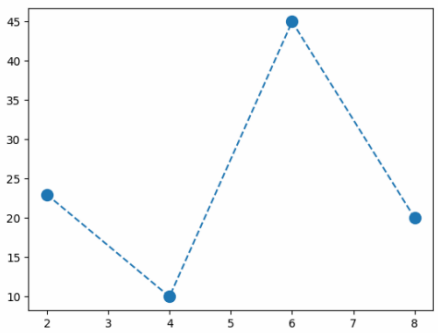
```
import pandas as pd
data=[ {'a':10, 'b':20}, {'a':5, 'b':10, 'c':20}, {'a':7, 'd':10, 'e':20}]
df=pd.DataFrame(data)
count_nan = df.isna().sum().sum()
print(df, '\n Count of Total NaN values is: ' + str(count_nan))
```

a. 3, 5, 7
b. 3, 4 ,5
c. 3, 3, 4
d. None of these

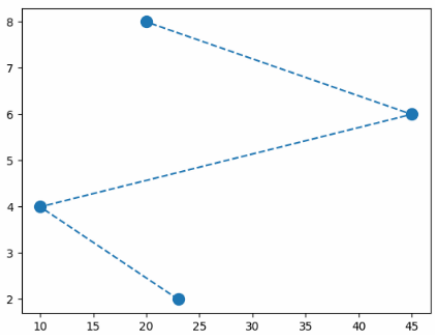
Q4) For the given code, identify the correct output. **(A)**

```
import pandas as pd
import matplotlib.pyplot as plt
x=[2, 4, 6, 8]
y=[23, 10, 45, 20]
plt.plot(y, x, linestyle ='dashed', marker='o')
plt.show()
```

a.



b.



c. No output
d. Error

Q5) What will be correct syntax for pandas series? **(T)**

a. pandas_Series(data, index, dtype)
b. pandas.Series(data, index, dtype, copy)
c. pandas.Series(data, index, dtype)
d. pandas_Series(data, index, dtype, copy)

SECTION - B (Attempt Any 5 Questions) **(5 * 2 = 10)**

Q1) Define 2 different types of Data Structures in Pandas with relevant examples. **(T)**
Q2) Explain magic functions. Which command is used to obtain list of magic functions in Jupyter console. **(T)**
Q3) What are the significant features of the pandas Library? **(T)**
Q4) Write a brief note outlining the Jupyter dashboard's features, including the role of kernel and checkpoints. **(T)**
Q5) Write 1 liner python codes to: **(A)**

a. Create a DataFrame(df) from a given list, list1= [[1,2,3,4],[5,6,7,8],[3,5,2,6]]
b. Read headers with only initial 2 rows, and last 1 row of df.
c. Change default column name to new names as col1, col2, col3, col4.
d. Slice out [7,8], [2,6] from df.

Q6) Write a Pandas program to convert a dictionary to a Pandas series. **(A)**
Q7) Write a Python code to create a pie chart for the given data on the popularity of programming languages and style it as follows: **(A)**

Programming languages: Java, Python, PHP, JavaScript, C#, C++
Popularity: 22.2, 17.6, 8.8, 8, 7.7, 6.7

a. Label the wedges and pull the wedge of most popular language at 0.2 from the center of the pie:
b. Add legends and also change the default start angle of pie to 140 degree.

Q8) Write a python code to create and join the two given DataFrames along columns and assign all data. **(A)**

DataFrame1

	student_id	name
0	S1	Tom
1	S2	Ryder
2	S3	Jensen

DataFrame 2

	student_id	name
0	S4	Nick
1	S5	James
2	S6	Benny

SECTION – C (Attempt Any 2 Questions) **(2 * 5 = 10)**

Q1) Explain the significance of uploading, streaming, and sampling real data with relevant examples. **(T)**
Q2) Draw the following two lines in a single plot with a **grid** along with the following style properties: **(A)**

line1 = [1, 5, 8, 9, 2, 0, 3, 10, 4, 7]
line2 = [3, 8, 9, 2, 1, 2, 4, 7, 6, 6]

i. Title: Demonstration of 2 lines in same plot, and label names: X- axis and Y- axis
ii. Set the ticks for both axes as [1, 2, 3, 4, 5, 6, 7, 8, 9, 10]
iii. Line 1 features: red color, *(star) marker, dotted line
iv. Line 2 features: green color, circle marker, dashed line
v. Show legends for both lines(as line 1, line 2) at the lower left location.

Q3) What is data cleaning? List out various types of bad data. For the given workout.csv file, Write Python codes for the following to: **(A)**

	Duration	Date	Pulse	Maxpulse	Calories
0	60	'2020/12/01'	110	130	409.1
1	60	Nan	117	145	479.0
2	300	'2020/12/03'	103	135	NaN
3	45	'2020/12/04'	109	175	282.4
4	45	'2020/12/05'	110	145	NaN

i. Remove all rows containing NULL values and display result.
ii. Replace the missing values in Calories column using mean.
iii. Discover and remove duplicate values in given dataset.