DATA SCIENCE

TEST-1

1) **Which of the following statements is correct in this python code? (python-hard)**

```
class Name:
    def __init__(javatpoint):
        javajavatpoint = java
name1=Name("ABC")
name2=name1
```

a. It will throw the error as multiple references to the same object is not possible

b. id(name1) and id(name2) will have same value

c. Both name1 and name2 will have reference to two different objects of class Name

d. All of the above

Explanation: "name1" and "name2" refer to the same object, so id(name1) and id(name2) will have the same value.

2) **What is the method inside the class in python language**? **(python-easy)**

a. Object

b. Function

c. Attribute

d. Argument

Answer: (b) Function

Explanation: Function is also known as the method.

3) **Study the following function: round(4.576) (python-easy)**

What will be the output of this function?

a. 4

b. 5

c. 576

d. 5

Answer: (d) 5

Explanation: The round function is a built-in function in the Python language that round-off the value (like 3.85 is 4), so the output of this function will be 5.

4) **Study the following statements: (python-medium)**

```
>>> str1 = "javat"
>>> str2 = ":"
>>> str3 = "point"
>>> str1[-1:]
```

What will be the output of this statement?

a. t

b. j

c. point

d. java

Answer: (a) t

Explanation: The correct output of this program is "t" because -1 corresponds to the last index.

5) **Study the following program: (python-hard)**

```
class Std_Name:
    def __init__(self, Std_firstName, Std_Phn, Std_lastName):
        self.Std_firstName = Std_firstName
        self. Std_PhnStd_Phn = Std_Phn
        self. Std_lastNameStd_lastName = Std_lastName

Std_firstName = "Wick"
name = Std_Name(Std_firstName, 'F', "Bob")
Std_firstName = "Ann"
name.lastName = "Nick"
print(name.Std_firstName, name.Std_lastName)
```

What will be the output of this statement?

a. Ann Bob

b. Ann Nick

c. Wick Bob

d. Wick Nick

Answer: (d) Wick Nick

6**)** **Study the following statements: (python-easy)**

1. >>> **print(ord('h') - ord('z'))**

What will be the output of this statement?

a. 18

b. -18

c. 17

d. -17

Answer: (b) -18

Explanation: ASCII value of h is less than the z. Hence the output of this code is 104-122, which is equal to -18.

7) **Which of the following is correctly evaluated for this function? (python-medium)**

1. **pow(x,y,z)**

a. (x**y) / z

b. (x / y) * z

c. (x**y) % z

d. (x / y) / z

Answer: (c) (x**y) % z

Explanation: None

8) **Certainly! Here's another SQL multiple-choice question: (Sql-easy)**

Which SQL clause is used to filter the results of a query?

A) ORDER BY

B) GROUP BY

C) WHERE

D) HAVING

Answer: (c) WHERE

Explanation:

The `WHERE` clause in SQL is like a filter for your data. When you use it in a query, you're telling the database to only show you the rows that meet certain conditions. It's like asking, "Show me only the rows where something specific is true." It helps you narrow down and get exactly the data you're interested in from a larger set of information.

9**) Consider a table named "employees" with the following columns: (Sql-hard)**

- employee_id

- employee_name

- hire_date

- Salary

Which SQL query will retrieve the names of employees hired in the year 2023 with a salary greater than 50000?

A) SELECT employee_name FROM employees WHERE hire_date = '2023' AND salary > 50000;

B) SELECT employee_name FROM employees WHERE hire_date LIKE '2023%' AND salary > 50000;

C) SELECT employee_name FROM employees WHERE YEAR (hire_date) = 2023 AND salary > 50000;

D) SELECT employee_name FROM employees WHERE hire_date BETWEEN '2023-01-01' AND '2023-12-31' AND salary > 50000;

Answer: D

Explanation: In SQL, the BETWEEN keyword is used to specify a range, and in this case, we want to find employees hired in the year 2023. The condition hire_date BETWEEN '2023-01-01' AND '2023-12-31' ensures we're considering dates within that specific year. Additionally, we want employees with a salary greater than 50000, so we add AND salary > 50000 to the query. This ensures that only employees meeting both conditions (hired in 2023 and salary greater than 50000) are included in the result. So, the correct SQL query retrieves the names of employees hired in the year 2023 with a salary greater than 50000.

10) **Select customer_name, telephone (Sql-medium)**

From customers Where city in ('Jaipur', Delhi', 'Agra');

A) The customer_name and telephone of all customers.

B) The customer_name and telephone of all customers living in Jaipur, Delhi and Agra.

C) The customer_name and telephone of all customers living in either Jaipur Delhi or Agra.

D) The customer_name and telephone of all customers who are not living in Jaipur, Delhi or Agra.

Answer:

Option 3: The customer_name and telephone of all customers living in either Jaipur Delhi or Agra

11) **What will be the output of following code? (pandas-hard)**

import pandas as pd

series1 = pd.Series([10,20,30,40,50])

print (series1)

A)

```
0    10

1    20
```

```
2    30

3    40

4    50

dtype: int64
```

B)

```
1    10

2    20

3    30

4    40

5    50

dtype: int64
```

C)

```
0    10

1    20

2    30

3    40

4    50

dtype: float32
```

D) None of the above mentioned

Answer: A

Explanation:

Pandas Series is a one-dimensional labelled array that may carry data of any type. It is used to store text, numbers, and other data like integers, strings, float, python objects, etc. The index labels are used to refer to the labels on the axes as a whole. Pandas Series can be considered as a column in an Excel file.

12) **What will be output of following code? (pandas-medium)**

```
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. [200 400]

Answer: C) [200 400]

13) Amongst which of the following is / are used to analyze the data in pandas.(Pandas- Easy)

A. Data frame
B. Series
C. Both A and B
D. None of the mentioned above

Answer: C) Both A and B

14) **Amongst which Python library is similar to Pandas? (Numpy-Easy)**

A. NPy
B. RPy
C. NumPy
D. None of the mentioned above

Answer: C) NumPy

15) **What will be the output of the following Python code? (Numpy-medium)**

```
from numpy import random

x = random.randint(100)

print(x)
```

A. 56
B. 26
C. 40

D. All of the mentioned above

Answer: D) All of the mentioned above

Explanation:

In the above code, `random.randint(100)` function has been used which is used to create any integer number till 100.

16) **Observe the code and identify the outcome: (Numpy-Hard)**

```
from numpy import random

x = random.binomial(n=100, p=0.5, size=10)

print(x)
```

A. [41 53 50 52 60 47 50 50 50 46]
B. [50 52 60 47 50 50 50 46]
C. [41 53 50 52 60 47 50]
D. None of the mentioned above

Answer: A) [41 53 50 52 60 47 50 50 50 46]

Explanation:

In the above code, binomial distribution has been used so the outcome of the code will be [41 53 50 52 60 47 50 50 50 46].

17) **Which is the correct command to install Matplotlib? (Matplotlib-Easy)**

A. pip install matplotlib
B. pip install matplotlib.pz
C. pip install matplotlib.*
D. All of the above

Answer: A) pip install matplotlib

Explanation:

If Python and PIP are installed in your system, then you can easily install the matplotlib library by using the below-given command - `pip install matplotlib`

18) **11. Which is the correct code statement to draw a plot without a line? (Matplotlib-Hard)**

A. plt.plot(xpoints, ypoints)
B. plt.plot(xpoints, ypoints, 0)
C. plt.plot(xpoints, ypoints, False)
D. plt.plot(xpoints, ypoints, 'o')

Answer: D) plt.plot(xpoints, ypoints, 'o')

Explanation:

You can draw a plot without lines by specifying the string notation parameter `'o'`, which means 'rings'. The correct code statement to draw a plot without a line is:

```
plt.plot(xpoints, ypoints, 'o')
```

19) **How to activate a style (for example, you want to activate 'fivethirtyeight' style) in Matplotlib? (Matplotlib-Medium)**

A. plt.style.use('fivethirtyeight')
B. plt.style.apply('fivethirtyeight')
C. plt.style('fivethirtyeight')
D. plt.style.activate('fivethirtyeight')

Answer: A) plt.style.use('fivethirtyeight')

Explanation:

To activate a style in Matplotlib, you can use `plt.style.use()` function. Consider the below code statement to apply a style named "fivethirtyeight" –

```
plt.style.use('fivethirtyeight')
```

20) You have a dataset with a column named 'Price' that contains numeric values. Some of these values have extra spaces and special characters. Which pandas method can you use to clean and convert this

column to a numeric format, removing any non-numeric characters and spaces? (Data Cleaning –Medium)

A) df['Price'].clean()

B) df['Price'].convert_numeric()

C) df['Price'] = pd.to_numeric(df['Price'], errors='coerce')

D) df['Price'].remove_nonnumeric()

Answer: C

Explanation:

This option uses the pd.to_numeric function from pandas to convert the 'Price' column to numeric format. The errors='coerce' parameter is used to replace any non-numeric values with NaN (Not a Number). This helps clean the column by handling special characters and spaces, making it suitable for numeric operations.

21) You have a DataFrame named df with a column 'Email' containing email addresses. How can you remove duplicate email entries from the 'Email' column?(Data cleaning-Medium)

A) df['Email'].unique()

B) df['Email'].drop_duplicates()

C) df['Email'].remove_duplicates()

D) df.drop_duplicates('Email')

Answer: B

Explanation:

Option B uses the drop_duplicates () method from pandas, which is designed specifically for removing duplicate entries in a DataFrame. Here's a breakdown:

df ['Email']: Selects the 'Email' column.

.drop_duplicates (): Removes duplicate entries from the selected column.es ('Email')

22) Which pandas method is used to drop missing values from a DataFrame? (Data cleaning-Easy)

A) df.remove_missing()

B) df.drop_na()

C) df.dropna()

D) df.clean_missing()

Answer: C

Explanation:

The dropna() method in pandas is used to remove missing values (NaN) from a DataFrame.

23) What does the str.strip() method do in pandas? (Data Cleaning-Easy)

A) Removes duplicates from a string column.

B) Trims leading and trailing whitespaces from string values.

C) Converts strings to uppercase.

D) Splits a string into a list of substrings.

Answer: B

Explanation: The str.strip() method is used to remove leading and trailing whitespaces from string values in a pandas DataFrame.

24) You have a dataset with a column containing free-text comments. How would you perform sentiment analysis on these comments using Python? (Data Cleaning –Hard)

A) Apply a pre-trained machine learning model from scikit-learn.

B) Use regular expressions to extract keywords related to sentiment.

C) Utilize a natural language processing (NLP) library like spaCy or NLTK for sentiment analysis.

D) Convert the text into numerical vectors using TF-IDF and then apply a machine learning model.

Answer: C

25) You are dealing with a time-series dataset, and some data points appear to be outliers. How would you handle these outliers in a time-series context? (Data Cleaning-Hard)

A) Remove the entire time series containing outliers.

B) Apply a rolling mean or median to smooth out extreme values

C) Impute outliers with the mean value of the time series.

D) Use a statistical test, such as the Grubbs' test, to identify and remove outliers.

Answer: B