

Pandas Advance-1

Assignment Questions



Q1. List any five functions of the pandas library with execution.

Q2. Given a Pandas DataFrame df with columns 'A', 'B', and 'C', write a Python function to re-index the DataFrame with a new index that starts from 1 and increments by 2 for each row.

Q3. You have a Pandas DataFrame df with a column named 'Values'. Write a Python function that iterates over the DataFrame and calculates the sum of the first three values in the 'Values' column. The function should print the sum to the console.

For example, if the 'Values' column of df contains the values [10, 20, 30, 40, 50], your function should calculate and print the sum of the first three values, which is 60.

Q4. Given a Pandas DataFrame df with a column 'Text', write a Python function to create a new column 'Word_Count' that contains the number of words in each row of the 'Text' column.

Q5. How are DataFrame.size() and DataFrame.shape() different?

Q6. Which function of pandas do we use to read an excel file?

Q7. You have a Pandas DataFrame df that contains a column named 'Email' that contains email addresses in the format 'username@domain.com'. Write a Python function that creates a new column 'Username' in df that contains only the username part of each email address.

The username is the part of the email address that appears before the '@' symbol. For example, if the email address is 'john.doe@example.com', the 'Username' column should contain 'john.doe'. Your function should extract the username from each email address and store it in the new 'Username' column.

Q8. You have a Pandas DataFrame df with columns 'A', 'B', and 'C'. Write a Python function that selects all rows where the value in column 'A' is greater than 5 and the value in column 'B' is less than 10. The function should return a new DataFrame that contains only the selected rows.

For example, if df contains the following values:

	A	B	C
0	3	5	1
1	8	2	7
2	6	9	4
3	2	3	5
4	9	1	2

Your function should select the following rows: A B C

1 8 2 7

4 9 1 2

The function should return a new DataFrame that contains only the selected rows.

Q9. Given a Pandas DataFrame `df` with a column 'Values', write a Python function to calculate the mean, median, and standard deviation of the values in the 'Values' column.

Q10. Given a Pandas DataFrame `df` with a column 'Sales' and a column 'Date', write a Python function to create a new column 'MovingAverage' that contains the moving average of the sales for the past 7 days for each row in the DataFrame. The moving average should be calculated using a window of size 7 and should include the current day.

Q11. You have a Pandas DataFrame `df` with a column 'Date'. Write a Python function that creates a new column 'Weekday' in the DataFrame. The 'Weekday' column should contain the weekday name (e.g. Monday, Tuesday) corresponding to each date in the 'Date' column.

For example, if `df` contains the following values:

	Date
0	2023-01-01
1	2023-01-02
2	2023-01-03
3	2023-01-04
4	2023-01-05

Your function should create the following DataFrame:

	Date	Weekday
0	2023-01-01	Sunday
1	2023-01-02	Monday
2	2023-01-03	Tuesday
3	2023-01-04	Wednesday
4	2023-01-05	Thursday

The function should return the modified DataFrame.

Q12. Given a Pandas DataFrame `df` with a column 'Date' that contains timestamps, write a Python function to select all rows where the date is between '2023-01-01' and '2023-01-31'.

Q13. To use the basic functions of pandas, what is the first and foremost necessary library that needs to be imported?

Note: Create your assignment in Jupyter notebook and upload it in GitHub & share that github repository link through your dashboard. Make sure the repository is public.

