Pandas Advance-2

Assignment Questions

PW SKILLS

Consider following code to answer further questions:

```
import pandas as pd
course_name = [‘Data Science’, ‘Machine Learning’, ‘Big Data’, ‘Data Engineer’]
duration =  [2,3,6,4]
df = pd.DataFrame(data = {‘course_name’ : course_name, ‘duration’ : duration})
```

Q1. Write a code to print the data present in the second row of the dataframe, df.

Q2. What is the difference between the functions loc and iloc in pandas.DataFrame?

Q3. Reindex the given dataframe using a variable, reindex = [3,0,1,2] and store it in the variable, new_df then find the output for both new_df.loc[2] and new_df.iloc[2].

Did you observe any difference in both the outputs? If so then explain it.

Consider the below code to answer further questions:

```
import pandas as pd
import numpy as np
columns = ['column_1', 'column_2', 'column_3', 'column_4', 'column_5', 'column_6']
indices = [1,2,3,4,5,6]
#Creating a dataframe:
df1 = pd.DataFrame(np.random.rand(6,6), columns = columns, index = indices)
```

Q4. Write a code to find the following statistical measurements for the above dataframe df1:

(i) mean of each and every column present in the dataframe.

(ii) standard deviation of column, ‘column_2’

Q5. Replace the data present in the second row of column, ‘column_2’ by a string variable then find the mean of column, column_2.

If you are getting errors in executing it then explain why.

[Hint: To replace the data use df1.loc[] and equate this to string data of your choice.]

Q6. What do you understand about the windows function in pandas and list the types of windows functions?

Q7. Write a code to print only the current month and year at the time of answering this question.

[Hint: Use pandas.datetime function]

Q8. Write a Python program that takes in two dates as input (in the format YYYY-MM-DD) and calculates the difference between them in days, hours, and minutes using Pandas time delta. The program should prompt the user to enter the dates and display the result.

Q9. Write a Python program that reads a CSV file containing categorical data and converts a specified column to a categorical data type. The program should prompt the user to enter the file path, column name, and category order, and then display the sorted data.

Q10. Write a Python program that reads a CSV file containing sales data for different products and visualizes the data using a stacked bar chart to show the sales of each product category over time. The program should prompt the user to enter the file path and display the chart.

Q11. You are given a CSV file containing student data that includes the student ID and their test score. Write a Python program that reads the CSV file, calculates the mean, median, and mode of the test scores, and displays the results in a table.

The program should do the following:

- **Prompt the user to enter the file path of the CSV file containing the student data.**
- **Read the CSV file into a Pandas DataFrame.**
- **Calculate the mean, median, and mode of the test scores using Pandas tools.**
- **Display the mean, median, and mode in a table.**

Assume the CSV file contains the following columns:

- **Student ID: The ID of the student.**
- **Test Score: The score of the student's test.**

Example usage of the program:
Enter the file path of the CSV file containing the student data: student_data.csv

```
+-----------+--------+
| Statistic | Value  |
+-----------+--------+
| Mean      | 79.6   |
| Median    | 82     |
| Mode      | 85, 90 |
+-----------+--------+
```

Assume that the CSV file student_data.csv contains the following data:

```
Student ID,Test Score
1,85
2,90
3,80
4,75
5,85
6,82
7,78
8,85
9,90
10,85
```

The program should calculate the mean, median, and mode of the test scores and display the results in a table.

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.*