X

swayam **(https://swayam.gov.in)**   (https://swayam.gov.in/nc_details/NPTEL)

santosh44kumar@yahoo.com ⌄

NPTEL (https://swayam.gov.in/explorer?ncCode=NPTEL) » **Python for Data Science (course)**

Announcements (announcements)   **About the Course (https://swayam.gov.in/nd1_noc20_cs36/preview)**

Ask a Question (forum)   Progress (student/home)   Mentor (student/mentor)

Unit 5 - Week 3

Course outline

How does an NPTEL online course work?

Week 0

Week 1

Week 2

Week 3

● Datasets (unit? unit=3&lesson=93)

● Reading data (unit? unit=3&lesson=20)

● Pandas Dataframes I (unit? unit=3&lesson=27)

● Pandas Dataframes II

Assignment 3

The due date for submitting this assignment has passed.   Due on 2020-02-19, 23:59 IST. As per our records you have not submitted this assignment.

Click here (https://drive.google.com/open?id=1RYTdwYUf7MuwmMJY3yuKU3EUxdybq8sA) to download the Data description & Data sets

1) Pandas features a number of functions for reading data as a DataFrame object. Which of the   **1 point** following commands are valid?

       I.   pd.read_txt()
      II.   pd.read_excel()
    III.   pd.read_jason()
    IV.   pd.read_table()

○ I and II

○ I, II and III

○ II and IV

○ I, II, III and IV

No, the answer is incorrect.
Score: 0

Accepted Answers:
II and IV

2) Which of the following is a valid indexing option with DataFrames?   **1 point**

○ iloc

○ loc

○ ix

○ all of the above

No, the answer is incorrect.
Score: 0

Accepted Answers:
all of the above

3) Which of the following function allows the use of *'Lambda expression'* while querying the data? **1 point**

○ mask()

○ apply()

○ aggregate()

○ groupby()

No, the answer is incorrect.
Score: 0

Accepted Answers:
apply()

4) While reading comma-separated values (csv) file into DataFrame., which of the following will   **1 point**
be used to set the first column as
        the index column?

○ index_col = False

○ index_col = 0

○ index_col = True

○ index_col = 1

No, the answer is incorrect.
Score: 0

Accepted Answers:
index_col = 0

5) Read the given dataset "*Tips.csv*" as a dataframe "*Data*". Which of the following command(s)   **1 point**
is/are correct to extract the columns in the
        following sequence - *Time, TotalBill, Tips*?

☐ df1=pd.DataFrame(Data, columns= ['Time', 'TotalBill', 'Tips'])

☐ df1=Data[['Time', 'TotalBill', 'Tips']]

☐ df1=Data.iloc[:,0:2]

☐ df1=Data.loc[:, ['Time', 'TotalBill', 'Tips']]

No, the answer is incorrect.
Score: 0

Accepted Answers:
df1=pd.DataFrame(Data, columns= ['Time', 'TotalBill', 'Tips'])
df1=Data[['Time', 'TotalBill', 'Tips']]
df1=Data.loc[:, ['Time', 'TotalBill', 'Tips']]

Supporting material for Week 4

Download Videos

6) Read the given excel sheet '*Tips1.xlsx*' as a dataframe '*Data1*'. Identify which of the following   **1 point** command (s) is/are correct to merge
the two data frames '*Data*' and '*Data1*' by columns?

○ Data2 = pd.concat(Data, Data1, join='outer')

○ Data2 = pd.DataFrame.join(Data, Data1, on=None, how='left')

○ Data2 = pd.DataFrame.append(Data,Data1)

○ Data2 = pd.merge(Data, Data1, how='left')

No, the answer is incorrect.
Score: 0
Accepted Answers:
Data2 = pd.merge(Data, Data1, how='left')

7) Copy the '*Data2*' dataframe as '*Data3*' (Data3 = Data2.copy()) and identify the command to   **1 point** find the total tips received across Day's from the dataframe '*Data3*'?

○ Data3.groupby(['Day', 'Tips']).aggregate(sum)

○ Data3.groupby('Day', 'Tips').aggregate(sum)

○ Data3.groupby('Day')[['Tips']].aggregate(sum)

○ Data3.groupby('Day', ['Tips'])['Tips'].aggregate(sum)

No, the answer is incorrect.
Score: 0
Accepted Answers:
Data3.groupby('Day')[['Tips']].aggregate(sum)

8) Copy the '*Data2*' dataframe as '*Data3*' (Data3 = Data2.copy()) and find which of the following   **1 point** command (s) gives the count of the Time ('Dinner' or 'Lunch') across gender?

☐ Data3.groupby(['Gender', 'Time'])['Time'].count().unstack()

☐ Data3.groupby('Gender')['Time'].aggregate(sum)

☐ pd.crosstab(index = Data3['Gender'], columns = Data3['Time'], normalize = False)

☐ Data3.pivot_table('Time', index='Gender', columns=Data3.Time.values, aggfunc=len)

No, the answer is incorrect.
Score: 0
Accepted Answers:
Data3.groupby(['Gender', 'Time'])['Time'].count().unstack()
pd.crosstab(index = Data3['Gender'], columns = Data3['Time'], normalize = False)
Data3.pivot_table('Time', index='Gender', columns=Data3.Time.values, aggfunc=len)

9) Which of the following plot is a visual representation of the statistical five-number summary of   **1 point** a data?

○ BoxPlot

○ BarPlot

○ Histogram

○ ScatterPlot

No, the answer is incorrect.
Score: 0
Accepted Answers:
BoxPlot

10)Which of the following statement is not true about histograms?   **1 point**

○ Represent the frequency distribution of categorical variables

○ It is a graphical representation of data using bars of different heights

○ Groups numbers into ranges and the height of each bar depicts the frequency of each range or bin

○ Represent the frequency distribution of numerical variables

No, the answer is incorrect.
Score: 0

Accepted Answers:
Represent the frequency distribution of categorical variables

11)If you have column with categorical variables, which will be the appropriate method to fill in the **1 point**
NaN's present in the column?

○ Mean

○ Median

○ Mode

○ None of the above

No, the answer is incorrect.
Score: 0

Accepted Answers:
Mode

12)Which of the following is not the right command to fill NaN values?   **1 point**

○ fillna()

○ ffill()

○ bfill()

○ fillcolumn()

No, the answer is incorrect.
Score: 0

Accepted Answers:
fillcolumn()

13)For the given dataframe "*Data3*" plot a histogram for the variable '*TotalBill*' to check which   **1 point**
range has the highest frequency.

○ 10-15

○ 15-20

○ 20-25

○ 25-30

No, the answer is incorrect.
Score: 0

Accepted Answers:
15-20

14)For the given dataframe "*Data3*" draw a bar chart for the variable "*Day*". Identify the category   **1 point**
with the maximum count

○ Friday

○ Thursday

○ Saturday

○ Sunday

No, the answer is incorrect.
Score: 0
Accepted Answers:
Saturday

15) Find the mean of the '*TotalBill*', '*Tips*' and '*Size*' across Days from the dataframe '*Data3*'?   *1 point*

☐ Data3.groupby('Day').aggregate('mean')

☐ Data3['Tips'].mean()

☐ Data3.groupby('Day').apply(lambda x: x.mean())

☐ Data3.groupby('Day').apply(mean)

No, the answer is incorrect.
Score: 0
Accepted Answers:
Data3.groupby('Day').aggregate('mean')
Data3.groupby('Day').apply(lambda x: x.mean())

16) On which day sum of the total bill was maximum?   *1 point*

○ Friday

○ Saturday

○ Sunday

○ Thursday

No, the answer is incorrect.
Score: 0
Accepted Answers:
Saturday

17) What will be the output of 'a' and 'b'?   *1 point*

```
import copy
x = [5, 4, 3, 2, 1]
y = [7, 8, 9]
z = [x, y]
a=copy.deepcopy(z)
b=copy.copy(z)
x[2] = 6
print('a =', a, 'b=', b)
```

○ a = [[5, 4, 3, 2, 1],[7, 8, 9]] b= [[5, 4, 3, 2, 1],[7, 8, 9]]

○ a = [[5, 4, 6, 2, 1],[7, 8, 9]] b= [[5, 4, 6, 2, 1],[7, 8, 9]]

○ a = [[5, 4, 6, 2, 1],[7, 8, 9]] b= [[5, 4, 3, 2, 1],[7, 8, 9]]

○ a = [[5, 4, 3, 2, 1],[7, 8, 9]] b= [[5, 4, 6, 2, 1],[7, 8, 9]]

No, the answer is incorrect.
Score: 0

Accepted Answers:
a = [[5, 4, 3, 2, 1],[7, 8, 9]] b= [[5, 4, 6, 2, 1],[7, 8, 9]]

18) In Pandas library, Dataframe class provides a member function to find duplicate rows based on **1 point** all columns. Identify the right option.

○ DataFrame.duplicateRows()

○ DataFrame.duplicated()

○ DataFrame.duplicateColumn()

○ DataFrame.Isduplicate()

No, the answer is incorrect.
Score: 0

Accepted Answers:
DataFrame.duplicated()

19) What does the following command do?   **1 point**

df.dropna(axis=0, how='all') ?

○ Drop rows if there are one or more missing values

○ Drop columns if there are one or more missing values

○ Drops rows if the entire row has missing values

○ Drops columns if they contain only missing values

No, the answer is incorrect.
Score: 0

Accepted Answers:
Drops rows if the entire row has missing values

20) Correlation between two variables X&Y is 0.85. Now, after adding the value 2 to all the values   **1 point** of X, the correlation

co-efficient will be

○ 0.85

○ 0.87

○ 0.65

○ 0.82

No, the answer is incorrect.
Score: 0

Accepted Answers:
0.85