Announcements

About the Course

Ask a Question

Progress

Mentor

NPTEL » Python for Data Science

Course outline

Unit 5 - Week 3

How does an NPTEL online

course work? Week 0 Week 1 Week 2 Week 3 Datasets Reading data Pandas Dataframes I I and II Pandas Dataframes II I, II and III Pandas Dataframes III II and IV I, II, III and IV Control structures & Functions Exploratory data analysis Score: 0 Accepted Answers: Data Visualization-Part I II and IV Data Visualization-Part II Dealing with missing data lloc Week 3: Lecture slides Oloc Quiz : Practice Assignment 3 ○ ix all of the above Quiz : Assignment 3 Week 3 - FAQs Score: 0 Accepted Answers: Week 3 Feedback all of the above Solution - Assignment 3 Week 4 mask() Supporting material for Week apply() aggregate() groupby() **Download Videos** Score: 0 Accepted Answers: apply() the index column? index_col = False \bigcirc index_col = 0 index_col = True

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 to download the Data description & Data sets 1) Pandas features a number of functions for reading data as a DataFrame object. Which of the following commands are valid? 1 point pd.read_txt() pd.read_excel() pd.read_jason() IV. pd.read_table() No, the answer is incorrect. 2) Which of the following is a valid indexing option with DataFrames? 1 point No, the answer is incorrect. 3) Which of the following function allows the use of 'Lambda expression' while querying the data? 1 point No, the answer is incorrect. 4) While reading comma-separated values (csv) file into DataFrame., which of the following will be used to set the first column as 1 point index_col = 1 No, the answer is incorrect. Score: 0 Accepted Answers: $index_col = 0$ Read the given dataset "Tips.csv" as a dataframe "Data". Which of the following command(s) 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']] 6) Read the given excel sheet 'Tips1.xlsx' as a dataframe 'Data1'. Identify which of the following 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') 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) '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 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? 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 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? 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 range has the highest frequency. **10-15** 15-20 20-25 25-30 No, the answer is incorrect. Score: 0 Accepted Answers: 15-20

1 point 1 point 7) Copy the 'Data2' dataframe as 'Data3' (Data3 = Data2.copy()) and identify the command to find the total tips received across Day's from the 1 point 8) Copy the 'Data2' dataframe as 'Data3' (Data3 = Data2.copy()) and find which of the following command (s) gives the count of the Time ('Dinner' or 1 point 1 point 1 point 1 point 1 point 1 point 14) For the given dataframe "Data3" draw a bar chart for the variable "Day". Identify the category with the maximum count 1 point 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'? 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? Friday Saturday Sunday Thursday No, the answer is incorrect. Score: 0 Accepted Answers: Saturday 17) What will be the output of 'a' and 'b'? 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] = 6print('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 all columns. Identify the right option. DataFrame.duplicateRows() DataFrame.duplicated() DataFrame.duplicateColumn() DataFrame.lsduplicate() No, the answer is incorrect. Score: 0 Accepted Answers: DataFrame.duplicated() 19) What does the following command do? 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 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

1 point 1 point 1 point 1 point 1 point 1 point