

## Unit 5 - Week 3

### How does an NPTE course work?

Week 0

## Week 2

### Week 3

- Read

☐ Pand

- ☐ Pandas Dataframes III
  - ☐ Control structures & Functions
  - ☐ Exploratory data analysis
  - ☐ Data Visualization-Part I
  - ☐ Data Visualization-Part II
  - ☐ Dealing with missing data
  - ☒ **Week 3: Lecture slides**
  - ☐ **Quiz : Practice Assignment 3**
  - ☐ **Quiz : Assignment 3**
  - ☐ Week 3 - FAQs
  - ☐ Week 3 Feedback
  - ☒ **Solution - Assignment 3**
- Week 4**
- Supporting material for Week**

## Download

The due date for submitting this assignment is 11/15/2023. 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

- 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?
- ☐ iloc

- Accepted Answers:  
all of the above
- 3) Which of the following function allows the use of 'Lambda expression' while querying the data?
- ☐ mask()
- ☐ apply()
- 1 point

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

- 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) is/are correct to extract the columns in the following sequence - *Time*, *TotalBill*, *Tips*?
- 1 point

- 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')

- Accepted Answers:
- ```
Data2 = pd.merge(Data, Data1, how='left')
```
- 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

- `Data3.groupby('Day')['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 command (s) gives the count of the Time ('Dinner' or 1 point

- 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
- 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

- 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
- 1 point

- Accepted Answers:  
Mode
- 12) Which of the following is not the right command to fill NaN values?
- 1 point

- No, the answer is incorrect.  
Score: 0
- Accepted Answers:  
`fillcolumn()`

- 25-30
- No, the answer is incorrect.  
Score: 0
- Accepted Answers:

- ☐ Saturday
- ☐ Sunday
- No, the answer is incorrect.

- ☐ 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:

- ☐ Friday
- ☐ Saturday
- ☐ Sunday
- ☐ Thursday

- ```
import copy
x = [5, 4, 3, 2, 1]
```

- `print('a =', a, 'b=', b)`
- ☐ a = [[5, 4, 3, 2, 1],[7, 8, 9]] b = [[5, 4, 3, 2, 1],[7, 8, 9]]

- 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. 1 point

- No, the answer is incorrect.  
Score: 0
- Accepted Answers:  
*DataFrame.duplicated()*