1. Pandas features a number of functions for reading data as a DataFrame object. Which of the 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

1. Which of the following is a valid indexing option with DataFrames?

 iloc

 loc

 ix

 all of the above

1. Which of the following function allows the use of *‘Lambda expression’* while querying the data?

 mask()

 apply()

 aggregate()

 groupby()

1. 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?

 index\_col = False

 index\_col = 0

 index\_col = True

 index\_col = 1

1. 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'] ]

1. 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')

1. Copy the '*Data2'* dataframe as '*Data3'  (Data3 = Data2.copy())*and identify the command to 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)

1. 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 '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)

1. Which of the following plot is a visual representation of the statistical five-number summary of a data?

 BoxPlot

 BarPlot

 Histogram

 ScatterPlot

1. 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

1. 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. Which of the following is not the right command to fill NaN values?

 fillna()

 ffill()

 bfill()

 fillcolumn()

1. 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

1. For the given dataframe “*Data3*” draw a bar chart for the variable “*Day*”. Identify the category with the maximum count

 Friday

 Thursday

 Saturday

 Sunday

1. 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)

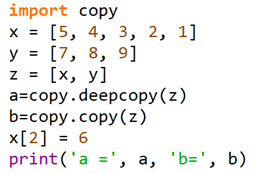
1. On which day sum of the total bill was maximum?

 Friday

 Saturday

 Sunday

 Thursday

1. What will be the output of ‘a’ and ‘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]]

1. 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.Isduplicate()

1. 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

1. 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