Question 1

Correct

Question text

If you try to read the file WHO POP TB same.xls into a dataframe using the code:

data = read\_excel('WHO POP TB same.xls')

What error would be shown if the file was in a different folder from that of the notebook?

Answer error

Feedback

Your answer is correct.

Correct answer:

**File not found** error or FileNotFound Error

This error would also be shown if the file name had been misspelt.

It indicates that the file cannot be found in the given location.

Question 2

Correct

Question text

Given a dataframe df with columns 'Country', 'Population (1000s)' and 'TB deaths' which of the following expressions will return a new dataframe consisting of just the 'Country' and 'Population (1000s)' columns?

Select one:



df[[Country, Population (1000s)]]



df['Country', 'Population (1000s)']



df[['Country', 'Population (1000s)']]



df[['Country', 'Population 1000s']]

Feedback

Your answer is correct.

Correct answer:

df[['Country', 'Population (1000s)']]

The list of columns inside the inner square brackets is used to select which columns you want in the new dataframe.  This option also shows the column names correctly, and as strings enclosed in quotes.

Question 3

Correct

Question text

At the end of the exercise notebook, calculate the mean and median of the death rate. Which statement is true?

Select one:



The mean is more or less double the median

The mean is 30.22 and the median is 15.53.



The various countries with low rates push the mean below the median.



The median is below 15 deaths per 100 thousand inhabitants



The mean is 15.53 deaths per 100 thousand inhabitants

Feedback

Your answer is correct.

Question 4

Correct

Question text

Given a dataframe df with columns 'Country', 'Population (1000s)' and 'TB deaths' which one of the following is true after this line of code is executed:

df.sort\_values('Population (1000s)')

Select one:



df will now be sorted.



The dataframe displayed will be in ascending order of population.



A dataframe will be displayed with the largest value for population in the 1st row.



The dataframe displayed will have only the index and 'Population (1000s)' columns.

Feedback

Your answer is correct.

Correct answer:

The dataframe displayed will be in ascending order of population.

The dataframe sort\_values() function returns a new sorted dataframe but doesn't change the original one.

sort\_values() doesn't select any columns so the dataframe displayed will have all the columns of the original dataframe.

The default for sort\_values() is ascending:  so the smallest population value will be in the first row and the largest in the final row.

Question 5

Correct

Question text

Which of the following statements will add a new column to a dataframe df called 'Diff\_From\_Average' with values of the difference between the population in thousands and the average population in thousands?

For example, for a particular row, if the population in 1000s is 3000 and the average is 4000 then the new column should have the value -1000. Assume the average required is the sum of the values divided by the number of values and choose the appropriate function.

Select one:



df['Diff\_From\_Average'] = df['Population (1000s)'].mean() - df['Population (1000s)']



df['Diff\_From\_Average'] = df['Population (1000s)'] - df['Population (1000s)'].mean()



df['Diff\_From\_Average'] = (df['Population (1000s)'] - df['Population (1000s)']).mean()



df['Diff\_From\_Average'] = df['Population (1000s)'] - df['Population (1000s)'].median()

Feedback

Your answer is correct.

Correct answer:

df['Diff\_From\_Average'] = df['Population (1000s)'] - df['Population (1000s)'].mean()

The correct function to calculate the average is mean, as this is the sum of the values divided by the number of values.  So the statement including median is not correct.