

Week 3: FAQs on Pandas

1. How to add Pandas to spyder?

Ans-Import pandas as pd

2. Does Python support .xls and .xlsx formats ?

Ans-Yes, Python supports excel files. To read an excel file, follow the steps given below.

Import pandas as pd

xls = pd.read_excel('file_name.xls')

xlsx = pd.read_excel('file_name.xlsx')

3. Instead of returning the data types for all columns, can the data type of single column be returned ?

Ans-Yes, the data type of a single column can be returned using the following command.
(df1-name of the existing dataframe)

df1['column_name'].dtypes

Refer to the [link](#) to know more.

4. Is Iris_data_sample inbuilt data file in pandas?

Ans-No. This is a different dataset.

5. How do you find the directory where pandas directory in my pc?

Ans-You can find where pandas library is stored by using the following commands:

import pandas

pandas.__path__

6. How do you get the number of unique data types in a given dataframe?

Ans-Pandas **dataframe.get_dtype_counts()** function returns the counts of **dtypes** in the given dataframe. It returns a pandas series object containing the counts of all data types present in the pandas object.

7. How is the memory usage reduced when the data type is converted from object to category ?

Ans-Refer to the [link](#) below for information on categorical data in pandas.

8. The output for the .txt file does not read “??” and “###” values, instead it is showing the average values for the applicable column values.

Ans-Use any of the following codes to read the text data into Spyder:

data_txt1=pd.read_table('Iris_data.txt',delimiter=" ")

```
data_txt2=pd.read_csv('Iris_data.txt',delimiter=" ")
```

Clear your console and environment variables and then try reading it again.

9. How to find out if there are any duplicate values in a DataFrame ?

Ans-To find out duplicated values, use the `pandas.DataFrame.duplicated()` function.

10. How to get the list of in-built functions (most useful) in Python?

Ans-There many useful functions in python. Please refer to documentation for each library.

11. How to import a .csv file?

Ans-To import .csv files, use the following command:

```
pandas.read_csv("path")
```

Refer to the [documentation](#) for a better understanding.

12. Are there any built in functions in python which display the statistical 5 point summary?

Ans-The function `describe()` from `pandas library` returns the statistical five-number summary.

13. What is difference between `pandas.DataFrame.drop()` and `pandas.DataFrame.dropna()` functions?

Ans-The function `pandas.DataFrame.drop()` removes rows or columns by specifying label names and corresponding axis, or by specifying directly index or column names

The function `pandas.DataFrame.dropna()` drops rows with missing values. Refer to the [documentation](#) for more on which values are considered missing.

14. How do you deal with special or junk characters in data?

Ans-By converting special or junk characters to nan, the data cleaning process is made easy. This is because the pandas libraries offers several functions that help in handling nan values. For instance the function `pandas.DataFrame.dropna()` is used to drop rows with nan values and `pandas.DataFrame.fillna()` is used to fill the nan values with the given value.

15. Is it necessary to import the library os for using the command cd?

Ans-No. The cd command works independent of any library.

16. What is categorical data?

Ans-Refer to the [documentation](#) to understand categorical data.

17. Which function is used to find the Spearman's correlation coefficient in python?

Ans-Refer to the [link](#) for information on Spearman correlation in python.

18. When there is only one 'nan' in the entire column then does the entire column become an object if the datatype was previously of category?

Ans-Yes. Refer to the [link](#) for further information.

19. In the function dataframe.select_dtypes(include=None,exclude=None) what do the arguments include and exclude mean and what is the difference between them?

Ans-The arguments include and exclude help you specify which datatypes you want to select from the dataframe. Refer to the [link](#) for further information.

20. When used with the apply() function, is lambda like a for or while loop where x takes all the values of the columns until it has filled missing values?

Ans-The lambda function returns a value based on the conditions specified and apply functions applies it across rows /columns. Refer to the [link](#) for an in depth understanding.

21. How are values between a specific index range accessed?

Ans-If df is the name of the dataframe, then `df.at()` is used for accessing single values from a dataframe. In order to access a range of values from a dataframe, the `df.loc()` or `df.iloc()` functions can be used. Refer to the [link](#) below for further information.

22. Can a dataframe be multi-dimensional?

Ans-No. A data frame consists of only 2 dimensions i.e. rows and columns. Each column stores data corresponding to a specific dimension.

23. How to select two or three columns from data frame?

Ans-There many ways to select multiple columns from a dataframe. Refer to links below to know how to select different multiple columns from a dataframe.

[Reference1](#)

[Reference 2](#)