Pandas Quiz 11/11/2021

1) How is a Pandas DataFrame different from a 2D Numpy array?

Both Pandas and Numpy offer many different ways of subsetting. 2D Numpy arrays can only contain values of the same basic type, a downside compared to Pandas if you're working on typical Data Science problems.

2) What are two characteristics that describe Pandas DataFrame?

Rows correspond to observations; columns correspond to variables, or properties, of these observations.

3) Which Pandas function do you use to import data from a comma-separated value (CSV) file into a Pandas DataFrame?

read\_csv() is the function you need. You can specify a ton of other arguments to customize the way the data is imported. To learn about those, have a look at [the documentation of read\_csv()](http://pandas.pydata.org/pandas-docs/stable/generated/pandas.read_csv.html).

4) Which technique should you use to select an entire row by its row label when accessing data in a Pandas DataFrame?

Square brackets are used to get specific columns from a Pandas DataFrame. iloc is used if you want to select a row based on its position in the DataFrame, and not based on its row label.

5) What will be output for the following code?

import pandas as pd

import numpy as np

s = pd.Series(np.random.randn(4))

print (s.ndim)

Returns the number of dimensions of the object. By definition, a Series is a 1D data structure, so it returns 1.

6) Point out the correct statement.

NaN is the standard missing data marker used in pandas

If data is a dict, if index is passed the values in data corresponding to the labels in the index will be pulled out.

7) Point out the wrong statement.

A DataFrame is like a fixed-size dict in that you can get and set values by index label   
A Series is like a fixed-size dict in that you can get and set values by index label.

8) Which of the following works analogously to the form of the dict constructor?

DataFrame.from\_items

DataFrame.from\_records takes a list of tuples or an ndarray with structured dtype.

9) The groupby method on DataFrames is associated with what type of operations?

By “group by” we are referring to a process involving one or more of the following steps:

**Splitting** the data into groups based on some criteria.  
**Applying** a function to each group independently.  
**Combining** the results into a data structure.

10) Which of the following describe the relationship between Matplotlib, Pandas and Seaborn? (Check all that apply.)

Pandas and Seaborn use Matplotlib for plotting but provide different, higher-level interfaces.

Plots created in Pandas and Seaborn can be further manipulated using functions from Matplotlib.

Seaborn is a Python data visualization library based on matplotlib. It provides a high-level interface for drawing attractive and informative statistical graphics.

Seaborn vs matplotlib is that seaborn is more agreeable in taking care of data frames in Pandas, while matplotlib is very much associated with Pandas and NumPy and goes about as a graphics package for Python data visualisation.

Seaborn vs matplotlib is that seaborn tries not to cover plots with the assistance of its default themes, while matplotlib is profoundly robust and customised.