1 Q. Explain three dimensional data indexing .

Answer: A three dim. array is like a matrix with 3 dim. and can store values for three dependant variables.

It has many uses some of which being the color values for a particular image(rgb values), or if you want to store the price of an object given its three features, etc.

A 3 dimensional block of data. Sometimes it helps visualize it as something concrete.

1-Dimensional Array

You have 10 bricks. You lay them all in a row, end to end. It starts with brick 0 and ends with brick 9. You can reference any brick with brick[x].

2-Dimensional Array

You have 80 bricks. You lay them in rows of 10, each row adjacent to the other.

It starts with brick 0 row 0 and ends with brick 9 row 7. You can reference any brick with brick[x][y].

3-Dimensional Array

You have 720 bricks. You lay them in rows of 10, each row adjacent to the other up to 10, creating the same 10x8 array of bricks as in the 2D example. Let’s call that a “flat” of bricks. Now do this 8 more times with the flats of bricks behind each other, creating 10 stacked flats and a block of bricks 10x8x9.

You can reference any brick with brick[x][y][z].

That’s a 3 dimensional array of bricks.

solid shape has 3 dimensions. In Cartesian-coordinate system, these are normally represented respectively by x, y, z axes. This aids visualisation of few functions as solid objects for use as models for many real-time applications like Radars & Aero-space systems.

2 Q. What's the difference between a series and a dataframe?

Answer: Series is a type of list which can take integer values, string values, double value and more. Series can only contain single list with index, whereas dataframes can be made of more than one series or we can say that a dataframes is a collection of series that can be used to analyse the data.

Python provides a library called pandas that is popular with data scientists and analysts. Pandas enables users to manipulate and analyze data using sophisticated data analysis tools.

pandas provides two data structures that shape data into a readable form:

Series

Data frame

Series

A pandas series is a one-dimensional data structure that comprises of a key-value pair. It is similar to a python dictionary, except it provides more freedom to manipulate and edit the data.

Dataframe

A pandas dataframe is a two-dimensional data-structure that can be thought of as a spreadsheet. A dataframe can also be thought of as a combination of two or more series.

Series is a one dimensional object that can any datatype , it can be int , float , string or anything. In simpler words series are somewhat like a list , but here you can define your index too . Instead of starting from 0 to len(n-1) , you can define your own index like a, b ,c which are similar to keys in dictionary.

Now DataFrame is a two dimensional object. It has columns and rows. Each column may have different types of values including list , tuples , dictionaries and even other dataframes too.

3Q. What role does pandas play in data cleaning?

Answer:. When working with multiple data sources, there are many chances for data to be incorrect, duplicated, or mislabeled. If data is wrong, outcomes and algorithms are unreliable, even though they may look correct. Data cleaning is the process of changing or eliminating garbage, incorrect, duplicate, corrupted, or incomplete data in a dataset. There’s no such absolute way to describe the precise steps in the data cleaning process because the processes may vary from dataset to dataset. Data cleansing, data cleansing, or data scrub is that the initiative among the general data preparation process. Data cleaning plays an important part in developing reliable answers and within the analytical process and is observed to be a basic feature of the info science basics. The motive of data cleaning services is to construct uniform and standardized data sets that enable data analytical tools and business intelligence easy access and perceive accurate data for each problem.

Data cleaning with Pandas

Data scientists spend a huge amount of time cleaning datasets and getting them in the form in which they can work. It is an essential skill of Data Scientists to be able to work with messy data, missing values, inconsistent, noise, or nonsensical data. To work smoothly python provides a built-in module Pandas. Pandas is the popular Python library that is mainly used for data processing purposes like cleaning, manipulation, and analysis. Pandas stand for “Python Data Analysis Library”. It consists of classes to read, process, and write CSV data files. There are numerous Data cleaning tools present but, the Pandas library provides a really fast and efficient way to manage and explore data. It does that by providing us with Series and DataFrames, which help us not only to represent data efficiently but also manipulate it in various ways.

In this article, we will use the Pandas module to clean our dataset.

We are using a simple dataset for data cleaning i.e. iris species dataset.

4Q. How do you use pandas to make a data frame out of n-dimensional array?

Answer:

I want to be able to create n-dimensional dataframes. I've heard of a method for 3D dataframes using panels in pandas but, if possible, I would like to extend the dimensions past 3 dims by combining different datasets into a super dataframe

I tried this but I cannot figure out how to use these methods with my test dataset -> Constructing 3D Pandas DataFrame

Also, this did not help for my case -> Pandas Dataframe or Panel to 3d numpy array

I made a random test dataset with arbitrary axis data trying to mimic a real situation; there are 3 axis (i.e. patients, years, and samples). I tried adding a bunch of dataframes to a list and then making a dataframe with that but it didn't work :( I even tried a panel as in the 2nd link above but I couldn't get it to work either.

Rather than using an n-dimensional Panel, you are probably better off using a two dimensional representation of data, but using MultiIndexes for the index, column or both.

Panel, pandas’ data structure for 3D arrays, was always a second class data structure compared to the Series and DataFrame. To allow pandas developers to focus more on its core functionality built around the DataFrame, pandas removed Panel in favor of directing users who use multi-dimensional arrays to xarray.

5Q. Explain the notion of pandas plotting.

Answer:

Pandas uses the plot() method to create diagrams.

We can use Pyplot, a submodule of the Matplotlib library to visualize the diagram on the screen.

Scatter Plot

Specify that you want a scatter plot with the kind argument:

kind = 'scatter'

A scatter plot needs an x- and a y-axis.

In the example below we will use "Duration" for the x-axis and "Calories" for the y-axis.

Include the x and y arguments like this:

x = 'Duration', y = 'Calories'

Plotting with pandas

Pandas objects come equipped with their plotting functions. These plotting functions are essentially wrappers around the matplotlib library. Think of matplotlib as a backend for pandas plots.

The Pandas Plot is a set of methods that can be used with a Pandas DataFrame, or a series, to plot various graphs from the data in that DataFrame. Pandas Plot simplifies the creation of graphs and plots, so you don’t need to know the details of working with matplotlib.

Built-in visualization in pandas really shines in helping with fast and easy plotting of series and DataFrames.