Python Advance Assignment 2

1. Explain three-dimensional data indexing.

Solution:- Array indexing and slicing are important parts in data analysis and many different types of mathematical operations. We always do not work with a whole array or matrix or Dataframe. Array indexing and slicing is most important when we work with a subset of an array.

```
In [18]: x=np.arange(45).reshape(3,3,5)

In [19]: x

Out[19]: array([[[ 0,  1,  2,  3,  4],
                 [ 5,  6,  7,  8,  9],
                 [10, 11, 12, 13, 14]],

                [[15, 16, 17, 18, 19],
                 [20, 21, 22, 23, 24],
                 [25, 26, 27, 28, 29]],

                [[30, 31, 32, 33, 34],
                 [35, 36, 37, 38, 39],
                 [40, 41, 42, 43, 44]]])

In [26]: x[0][2]

Out[26]: array([10, 11, 12, 13, 14])

In [27]: x[1][2][4]

Out[27]: 29

In [28]: x[2]

Out[28]: array([[30, 31, 32, 33, 34],
                [35, 36, 37, 38, 39],
                [40, 41, 42, 43, 44]])
```

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

- Solution:- 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. To initialize a series, use- **pd.Series()**

- A pandas dataframe is a two-dimensional data-structure that can be thought of as a spreadsheet. A dataframe can also be a combination of two or more series. To initialize a dataframe, use- **pd.DataFrame().**

3. What role does pandas play in data cleaning?

Solution:- Pandas is an incredibly powerful function when cleaning data or manipulating a DataFrame.it have lot of inbuilt function for data cleaning or manipulating. Some inbuilt function in pandas as give below

- .info() gives the information about the each column in dataset like datatype of column, null value in column
- .describe() gives the mathematical information about the each columns in dataset like mean, standard deviation, maxvalue , minvalue etc.
- .dropna() –we can easily remove the null value from the dataset the help of dropna function.
- .fillna() we can also fill the null value in dataset with the help of fillna function.

So Pandas offer a diverse range of built-in functions that can be used to clean and manipulate datasets prior to analysis. It can allow you to drop incomplete rows and columns, fill missing values and improve the readability of the dataset through category renaming.

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

Solution:-

```
In [1]: import numpy as np

In [4]: data=np.array([[1,2],[3,4],[2,3]])

In [5]: data

Out[5]: array([[1, 2],
               [3, 4],
               [2, 3]])

In [3]: import pandas as pd

In [7]: df=pd.DataFrame(data,columns=['A','B'])

In [8]: df

Out[8]:
```

	A	B
0	1	2
1	3	4
2	2	3

5. Explain the notion of pandas plotting.

Solution:- There are several useful libraries for doing visualization with Python, like matplotlib or seaborn. These libraries are intuitive and simple to use. There's also **pandas**, which is mainly a data analysis tool, but it also provides multiple options for visualization. Plotting with pandas is pretty straightforward.

```
In [12]: import pandas as pd

In [29]: df=pd.read_csv('train.csv')

In [40]: df.plot('Sex','Age')

Out[40]: <AxesSubplot:xlabel='Sex'>
```