

- ① Getting Started with Numpy and Pandas
- ② NumPy
- ③ Pandas

1 Getting Started with Numpy and Pandas

2 NumPy

3 Pandas

- 1 Getting Started with Numpy and Pandas
- 2 NumPy
- 3 Pandas

What is NumPy?

- NumPy is the fundamental package for scientific computing in Python.
- It is a Python library that provides a multidimensional array object, various derived objects, and an assortment of routines for fast operations on arrays, including mathematical, logical, shape manipulation, sorting, selecting, basic linear algebra, basic statical operations, random simulation and much more.

Application of NumPy

- Mathematics
- Plotting
- Machine Learning

The Basics of NumPy

- `np.array()` - Create NumPy array
- `ndim` - Returns dimension of the NumPy array
- `shape` - Returns shape of the NumPy array
- `dtype` - Returns data type of that NumPy array
- `itemsize` - Returns size of that NumPy array in bytes

Concept of Slicing in NumPy

- `Array[row, :]` - Specific row and all columns
- `Array[:, column]` - Specific columns and all rows
- `Array[row, column]` - Specific row and specific column

Initializing NumPy Arrays

- `np.zeros()` - Create a array/matrix containing all '0's element.
- `np.ones()` - Create a arraye/matrix containing all '1's element.
- `np.random.full()` - Create a array/matrix eith containing all elements a specified value.
- `np.random.rand()` - Create a array/matrix containing all elements decimal valued randomly.
- `np.random.randint()` - Create a array/matrix containing all elements integer valued randomly.
- `astype()` - Change the datatype of the data

Reorganizing Arrays

- **Reshape()** - Change the shape of array/matrix
- **Vstack()** - Stack arrays in sequence vertically (row wise).
- **Hstack()** - Stack arrays in sequence horizontally (column wise).

- 1 Getting Started with Numpy and Pandas
- 2 NumPy
- 3 Pandas**

What is Pandas?

- Pandas is a Python library used for working with datasets.
- It has functions for analyzing, cleaning, and manipulating data.

What is Dataframe?

- A Pandas DataFrame is a 2 dimensional data structure, like a 2 dimensional array , or a table with rows and columns.

Loading DataFrame

```
1 import pandas as pd
2
3 # Loading dataset from a csv file
4 df = pd.read_csv('data_path\data_name.csv')
5
6 # Loading your own dataset
7 df = pd.DataFrame({'column1': [1, 2, 3], 'column2': ['Numpy', 'Pandas', 'Matplot']})
8
9 # Showing top-5 Rows
10
11 df.head(5)
12
13
14
15
```

Extract Data from Dataframe

- `df.head()`
- `df.tail()`
- `df.columns`
- `df.index`
- `df.info()`
- `df.size`
- `df.describe()`
- `df.shape`
- `df.dtypes`

Value at Specific Cell

- `df.at[row_index,column_name]`
- `df.iat[row_index,column_index]`

Fetch a Record

- `df.loc[rows]` - row can be integer or non-int value
- `df.iloc[rows]` - row must be integer

Find, Remove Null

- `isna()`
- `fillna()`
- `Dropna(inplace=True)`

Properties

- `max()`
- `min()`
- `mean()`
- `unique()`
- `nunique()`
- `value_counts()`

Set Operation Using Concat Method

- `pd.concat([df1, df2])`
- `df.append()`

End of NumPy and Pandas