Getting Started With

Pandas

HiiT ...Your Premium Training Partner | World Class IT Services since 1996

Pandas Library

- Probably the most important package in python data analysis
- It encapsulates data in a more abstract way, making it easier to manipulate, document, and understand the transformations you make in the base datasets.
- To get started with pandas, we need to be comfortable with its two workhorse data structures namely: **Series** and **DataFrames**

HiiT ...Your Premium Training Partner | World Class IT Services since 1996

python

How To Import Pandas Library

```
import pandas as pd
```

HiiT ...Your Premium Training Partner | World Class IT Services since 1996

Panda Series

Series

- A Series is a one – dimensional array containing an array of data and an associated array of data labels called its index.

How to make a series data structure

Given a list

age = [20, 21, 22, 23, 24, 25]

Then

data = pd. Series(age)

```
In [36]: Data

Out[36]: 0    20
         1    21
         2    22
         3    23
         4    24
         5    25
         dtype: int64
```

Accessing Elements of Series

- The values in a series data structure can be accessed only by their index value using the loc [] and iloc [] methods

```
In [36]: Data

Out[36]: 0    20
         1    21
         2    22
         3    23
         4    24
         5    25
         dtype: int64
```

Such that

Data.loc [“C”] = 22

And

Data.iloc[4] = 24

HiiT ...Your Premium Training Partner | World Class IT Services since 1996

Useful Series Methods

- data.head()
- data.describe()
- data.mean()
- data.prod()
- data.sum()
- data.tail()

HiiT ...Your Premium Training Partner World Class IT Services since 1996

Pandas DataFrame

python

DataFrames

- A DataFrame represents a tabular spreadsheet like data structure containing an ordered collection of columns each of which can be of a different value type (number, string, boolean)
- Unlike the Series, the DataFrame has a row and a column index. It can be thought of as a dictionary of series
- In contrast to the Series structure, the DataFrame can be accessed by either their index value or column name.

How to make a DataFrame

A. From External File Source

In the real world, a dataset is often read into python via an external source that curated it. These files can come in different formats.

File Type	Function in pandas
.csv	read_ csv
.html	read_ html
.xml	read_ xml
.sql	read_ sql
.excel	read_ excel
.jason	read_ json

How to make a DataFrame

B. Creating DataFrames Yourself

We can certainly create a data frame yourself by inputting data which takes its data input argument and converts it into a DataFrame.

How to make a dataframe data structure

Given a dictionary

student = {“Age”:[20, 21, 22, 23, 24, 25]}

Then

student_age = pd.DataFrame(student)

```
In [210]: students_age
Out[210]:
```

	Age
0	20
1	21
2	22
3	23
4	24
5	25