

Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

iHUB Ditya Sampark, IIT Roorkee
Ritvij Bharat Private Limited

Pandas

Led by : Shreyas Shukla

Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

Pandas is a library for Data Analysis built off of NumPy.

Fantastic documentation:

- <https://pandas.pydata.org/docs/>



Led by : Shreyas Shukla

Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

Using Pandas

- Tools for reading and writing data between many formats.
- Intelligently grab data based on indexing, logic, subsetting, and more.
- Handle missing data.
- Adjust and restructure data.

Led by : Shreyas Shukla

Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

- Series and DataFrames
- Conditional Filtering and Useful Methods
- Missing Data
- Group By Operations
- Combining DataFrames
- Text Methods and Time Methods
- Inputs and Outputs

Led by : Shreyas Shukla

Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

Let's get started!

Led by : Shreyas Shukla

Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

iHUB Divya Sampark, IIT Roorkee
Ritvij Bharat Private Limited

Series

Led by : Shreyas Shukla

Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

Series

- A data structure in Pandas that holds an array of information along with a named index.
- The named index differentiates this from a simple NumPy array.
- **Formal Definition:** One-dimensional ndarray with axis labels

led by : Shreyas Shukla

Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

Numeric index in Numpy array

Index	Data
0	1776
1	1867
2	1821

Led by : Shreyas Shukla

Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

Pandas Series adds on a labeled index

Labeled Index	Data
USA	1776
CANADA	1867
MEXICO	1821

Led by : Shreyas Shukla

Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

Data is still numerically organized

Numeric Index	Labeled Index	Data
0	USA	1776
1	CANADA	1867
2	MEXICO	1821

Led by : Shreyas Shukla

Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

- Let's see how to create a Pandas Series object.
- We'll also see some key properties and operations.
- Later on we will learn how to combine Series with a shared index to create a tabular data structure called a DataFrame.

Led by : Shreyas Shukla

Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

Let's code !!

iHUB DivyaSampark, IIT Roorkee
and
Ritvij Bharat Private Limited

Led by : Shreyas Shukla

Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

DataFrames

Led by : Shreyas Shukla

Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

- A DataFrame is a table of columns and rows in pandas that we can easily restructure and filter.
- **Formal Definition:** A group of Pandas Series objects that *share* the same index.

Led by : Shreyas Shukla

Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

Example of a Series

Index	Year
USA	1776
CANADA	1867
MEXICO	1821

Led by : Shreyas Shukla

Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

Example of Series with Same Index

Index	Year
USA	1776
CANADA	1867
MEXICO	1821

Index	Pop
USA	328
CANADA	38
MEXICO	126

Index	GDP
USA	20.5
CANADA	1.7
MEXICO	1.22

Led by : Shreyas Shukla

Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

Example of Series with Same Index

Index	Year
USA	1776
CANADA	1867
MEXICO	1821

Index	Pop
USA	328
CANADA	38
MEXICO	126

Index	GDP
USA	20.5
CANADA	1.7
MEXICO	1.22

Led by : Shreyas Shukla

Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

Index	Year	Pop	GDP
USA	1776	328	20.5
CANADA	1867	38	1.7
MEXICO	1821	126	1.22

Led by : Shreyas Shukla

Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

DataFrame is the main Pandas object we will work with and it is **extremely** useful!

iHUB DivyaSampark, IIT Roorkee
and
Ritvij Bharat Private Limited

Led by : Shreyas Shukla

Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

Conditional Filtering

Led by : Shreyas Shukla

Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

- In real world datasets, we don't filter based on position, but instead based on a **condition**.
-
- Conditional Filtering allows us to select **rows** based a condition on a column.

Led by : Shreyas Shukla

Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

Organizing Data

Index	Year	Pop	GDP
USA	1776	328	20.5
CANADA	1867	38	1.7
MEXICO	1821	126	1.22

Led by : Shreyas Shukla

Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

Columns are Features

Index	Year	Pop	GDP
USA	1776	328	20.5
CANADA	1867	38	1.7
MEXICO	1821	126	1.22

Led by : Shreyas Shukla

Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

Rows are instances of data

Index	Year	Pop	GDP
USA	1776	328	20.5
CANADA	1867	38	1.7
MEXICO	1821	126	1.22

Led by : Shreyas Shukla

Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

Index	Year	Pop	GDP
USA	1776	328	20.5
CANADA	1867	38	1.7
MEXICO	1821	126	1.22

Led by : Shreyas Shukla

Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

This allows to directly answer questions like

Index	Year	Pop	GDP
USA	1776	328	20.5
CANADA	1867	38	1.7
MEXICO	1821	126	1.22

Led by : Shreyas Shukla

Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

What countries have Pop greater than X?

Index	Year	Pop	GDP
USA	1776	328	20.5
CANADA	1867	38	1.7
MEXICO	1821	126	1.22

Led by : Shreyas Shukla

Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

What countries have Pop greater than 50?

Index	Year	Pop	GDP
USA	1776	328	20.5
CANADA	1867	38	1.7
MEXICO	1821	126	1.22

Led by : Shreyas Shukla

Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

`df["Pop"]`

Index	Year	Pop	GDP
USA	1776	328	20.5
CANADA	1867	38	1.7
MEXICO	1821	126	1.22

Led by : Shreyas Shukla

Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

`df["Pop"] > 50`

Index	Year	Pop	GDP
USA	1776	328	20.5
CANADA	1867	38	1.7
MEXICO	1821	126	1.22

Led by : Shreyas Shukla

Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

`df["Pop"] > 50`

Index	Year	Pop	GDP
USA	1776	328	20.5
CANADA	1867	38	1.7
MEXICO	1821	126	1.22

Led by : Shreyas Shukla

Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

`df["Pop"] > 50`

Index	Year	Pop	GDP
USA	1776	True	20.5
CANADA	1867	False	1.7
MEXICO	1821	True	1.22

Led by : Shreyas Shukla

Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

`df[df["Pop"] > 50]`

Index	Year	Pop	GDP
USA	1776	True	20.5
CANADA	1867	False	1.7
MEXICO	1821	True	1.22

Led by : Shreyas Shukla

Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

`df[df["Pop"] > 50]`

Index	Year	Pop	GDP
USA	1776	True	20.5
MEXICO	1821	True	1.22

Led by : Shreyas Shukla

Mastering Machine Learning with Python

(27th Aug 2024 - 18th Oct 2024)

Let's code !!

iHUB DivyaSampark, IIT Roorkee
and
Ritvij Bharat Private Limited

Led by : Shreyas Shukla