

# Pandas

Date \_\_\_\_\_  
Page \_\_\_\_\_

## Introduction

- Pandas:- Pandas is an open-source library for data analysis and manipulation data.

### → Why used Pandas

- Large data ko easily handle karta hai
- missing data ke manage karta hai
- Pandas Tabular data ke analysis ke liye use hata hai

### → Installation

- PIP install Pandas.

### → Key Data Structures.

#### 1. Series

A one-dimensional labeled array capable of holding any data type (integer, float, string, etc)

#### Creating using:

```
import Pandas as pd  
s = pd.Series([1, 2, 3, 4])
```

#### Series Properties

1. values

2. index

3. data  
dtype

2.

## DataFrame

- A Two-dimensional labeled data structure with columns of potentially different types.

Created using:

```
df = pd.DataFrame ({'A': [1, 2], 'B': [3, 4]})
```

→ Data Input/Output

1. CSV

Reading:

```
df = pd.read_csv(file.csv)
```

Writing:

```
df.to_csv(output.csv, index=False)
```

2. Excel

Reading:

```
df = pd.read_excel(file.xlsx)
```

Writing

```
df.to_excel(output.xlsx, index=False)
```

### 3. JSON

Reading:

```
df = pd.read_json('file.json')
```

Writing:

```
df.to_json('output.json')
```

### 4. Parquet

Reading:

```
df = pd.read_parquet('file.parquet')
```

Writing

```
df.to_parquet('output.parquet')
```

Dataframe important attributes

Shape → row & column

Columns → column names

index → row index

dtype → datatype



Viewing & Understanding

Data

1. Methods to view Data

heads() :- Display the first 5 rows of the dataset

Used to:

- Understand the structure of data
- Check column name and values

\* `tail()`

- Display the last 5 rows of the datasets

Used to:

- Check the ending records
- See the most recent entries

\* `sample()`

- Display random rows from the datasets
- Used to:
- Get a random overview of data
- Avoid bias from only first or last rows

## (2) Data information `info()`

\* Provides a summary of the entire dataset

- Total number of rows and columns
- Column names
- Data type of each column
- Number of non-null (non-empty) value

## describe()

- Gives statistical summary of numerical columns.

includes:

- count
- mean (average)
- minimum value
- maximum value
- 25%, 50%, 75% Percentiles

### 3. Checking missing values

- isnull()
- Checks whether a value is missing or not

#### → OUTPUT

- True → Value is missing
  - False → Value is Present
- 
- isnull().sum()
  - Counts the total number of missing values in each column