

# Reading Data From Different Sources

## 1. Reading JSON Data

Code:

```
import pandas as pd
from io import StringIO

Data = '{"employee_name": "James", "email": "james@gmail.com",
"job_profile": [{"title1": "Team Lead", "title2": "Sr. Developer"}]}'
df = pd.read_json(StringIO(Data))
df
```

Output:

```
employee_name email job_profile
0 James james@gmail.com [{'title1': 'Team Lead', 'title2': 'Sr. Developer'}]
```

Definition: Reads a JSON string or file and converts it into a DataFrame.

Syntax: `pd.read_json(path_or_buf, orient=None, ...)`

## 2. Reading CSV Data from a URL

Code:

```
df = pd.read_csv("https://archive.ics.uci.edu/ml/machine-learning-databases/
wine/wine.data", header=None)
df.head()
```

Output:

```
0 1 2 3 4 5 6 7 8 9 10 11 12 13
0 1 14.23 1.71 2.43 15.6 127 2.80 3.06 0.28 2.29 5.64 1.04 3.92 1065
1 1 13.20 1.78 2.14 11.2 100 2.65 2.76 0.26 1.28 4.38 1.05 3.40 1050
2 1 13.16 2.36 2.67 18.6 101 2.80 3.24 0.30 2.81 5.68 1.03 3.17 1185
3 1 14.37 1.95 2.50 16.8 113 3.85 3.49 0.24 2.18 7.80 0.86 3.45 1480
4 1 13.24 2.59 2.87 21.0 118 2.80 2.69 0.39 1.82 4.32 1.04 2.93 735
```

Definition: Reads a comma-separated values (CSV) file into a DataFrame.

Syntax: `pd.read_csv(filepath_or_buffer, sep=',', ...)`

## 3. Reading HTML Tables from a URL

Code:

```
url =
"https://www.fdic.gov/resources/resolutions/bank-failures/failed-bank-list/"
df = pd.read_html(url)
```

Output:

```
Bank Name City State Cert 0 Pulaski Savings Bank Chicago Illinois 28611
1 The First National Bank of Lindsay Lindsay Oklahoma 4134
2 Republic First Bank dba Republic Bank Philadelphia Pennsylvania 27332
3 Citizens Bank Sac City Iowa 8758
4 Heartland Tri-State Bank Elkhart Kansas 25851
```

```
Acquiring Institution Closing Date Fund
0 Millennium Bank January 17, 2025 10548
1 First Bank & Trust Co., Duncan, OK October 18, 2024 10547
2 Fulton Bank, National Association April 26, 2024 10546
3 Iowa Trust & Savings Bank November 3, 2023 10545
4 Dream First Bank, N.A. July 28, 2023 10544
```

Definition: Reads HTML tables from a specified URL or file and returns a list of DataFrames.  
Syntax: `pd.read_html(io, match=None, ...)`

## 4. Reading Excel Data

Code:

```
df_excel = pd.read_excel('data.xlsx')
df_excel
```

Output:

```
Name Age
0 Krish 32
1 Jack 34
2 John 31
```

Definition: Reads an Excel file into a DataFrame.  
Syntax: `pd.read_excel(io, sheet_name=0, ...)`

## 5. Reading Pickle Data

Code:

```
df_excel.to_pickle('df_excel')
pd.read_pickle('df_excel')
```

Output:

```
Name Age
0 Krish 32
1 Jack 34
2 John 31
```

Definition: Saves a DataFrame to a pickle file and reads it back.  
Syntax: `df.to_pickle(path)` and `pd.read_pickle(path)`