Lesson 05: DataFrames Introduction

Data Science with Python – BSc Course

Data Science Program

45 Minutes

Learning Objectives

After this lesson, you will be able to:

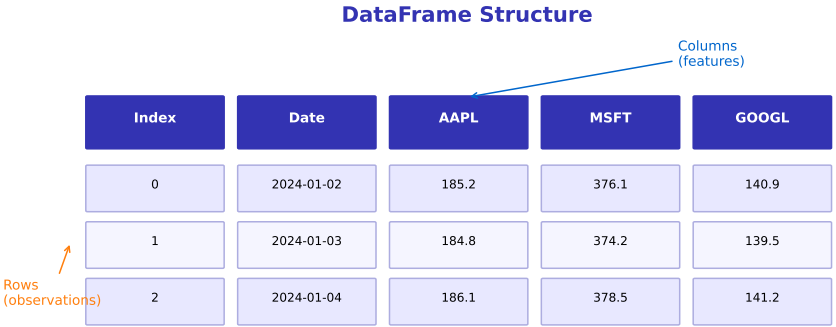
- Import pandas and create DataFrames
- Load data from CSV files
- Explore data with head(), tail(), info(), describe()
- Understand DataFrame structure (index, columns, values)

Finance Application: Load and explore stock price data.

---

pandas is THE library for data manipulation in Python

DataFrame Structure



2D labeled data structure with rows and columns

DataFrames are 2D labeled data structures

Series vs DataFrame

Series (1D)

Single column

0	185.2
1	184.8
2	186.1

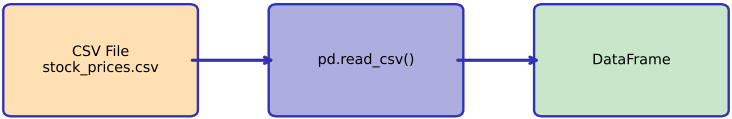
DataFrame (2D)

Multiple columns

	AAPL	MSFT	VOL
0	185.2	376.1	1.2M
1	184.8	374.2	1.1M
2	186.1	378.5	1.3M

DataFrame = Collection of Series sharing an index

Loading CSV Data



Common Parameters:

```
filepath: "data/prices.csv"
index_col: "Date"
parse_dates: True
usecols: ["AAPL", "MSFT"]
```

Viewing Data: head() and tail()

`df.head(3)`

First 3 rows

```
2024-01-02  185.2
2024-01-03  184.8
2024-01-04  186.1
```

`df.tail(3)`

Last 3 rows

```
2024-12-27  195.8
2024-12-30  196.2
2024-12-31  197.1
```

Default: 5 rows | Customize: head(10), tail(20)

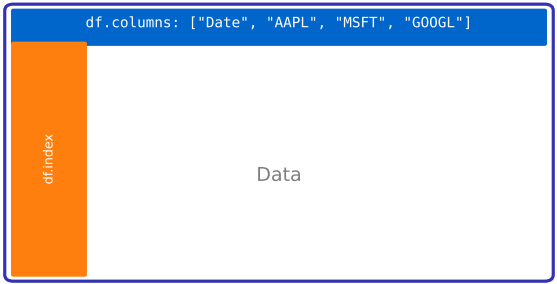
DataFrame Info: df.info()

```
<class pandas.DataFrame>
RangeIndex: 252 entries, 0 to 251
Data columns (5 columns):
  Date    252 non-null datetime64
  AAPL    252 non-null float64
  MSFT    252 non-null float64
  GOOGL   250 non-null float64  (2 missing)
memory usage: 10.0 KB
```

Summary Statistics: df.describe()

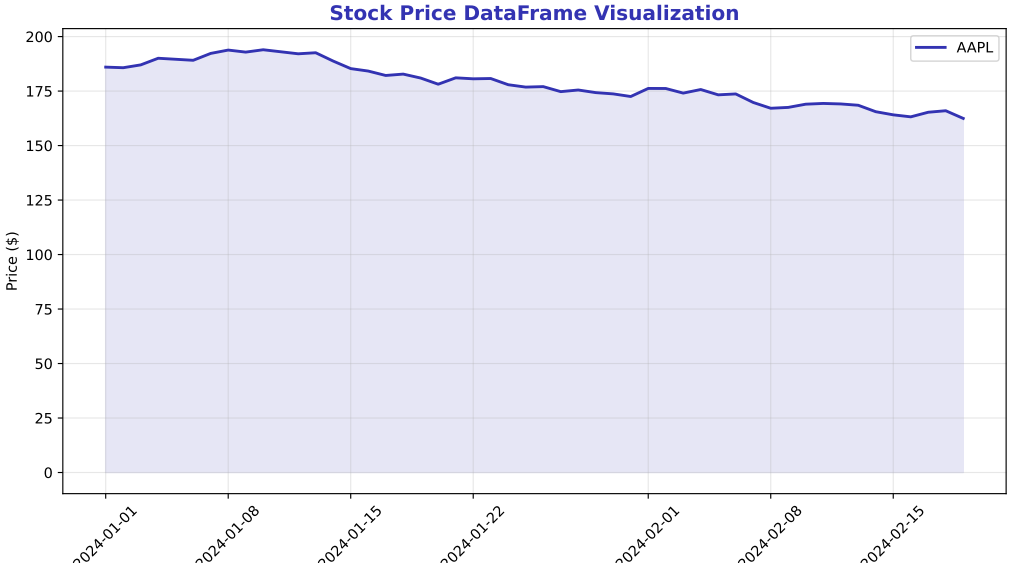
Stat	AAPL	MSFT
count	252	252
mean	189.5	385.2
std	8.2	12.5
min	175.1	355.8
25%	183.4	375.6
50%	188.9	384.1
75%	195.2	394.8
max	210.3	420.5

Index and Columns



df.shape: (252, 4) | df.dtypes: column data types

Stock Data Example



Hands-on Exercise (25 min)

Explore stock price data:

1. Load the stock data:
   `df = pd.read_csv("../datasets/stock_prices.csv")`
2. View first 10 rows: `df.head(10)`
3. Check data types: `df.info()`
4. Get statistics: `df.describe()`
5. Access column names: `df.columns`
6. Check shape: `df.shape`
7. Find which stock has highest mean price

Exploration before analysis prevents costly mistakes

Lesson Summary

Key Takeaways:

- pandas DataFrame is the core data structure
- pd.read_csv() loads CSV files easily
- head()/tail() show first/last rows
- info() shows data types and missing values
- describe() provides statistical summary

Next Lesson: Selection and Filtering

Loading data is step 1 – now we'll learn to slice it