Lesson 06: Selection and Filtering

Data Science with Python – BSc Course

Data Science Program

45 Minutes

Learning Objectives

After this lesson, you will be able to:

- Select columns using bracket and dot notation
- Access rows with iloc (position) and loc (label)
- Filter data using boolean conditions
- Combine multiple conditions with & and —

Finance Application: Screen stocks by price, volume, and other criteria.

Selection and filtering extract relevant data for analysis

Column Selection Methods

```
df['AAPL']
```
Single column (Series)

```
df[['AAPL', 'MSFT']]
```
Multiple columns (DataFrame)

```
df.AAPL
```
Attribute access (simple names)

```
df.loc[:, 'AAPL':'GOOGL']
```
Range of columns

iloc vs loc

iloc (Integer Location)

Position-based indexing

```
df.iloc[0]
df.iloc[0:5, 1:3]
```

Uses: 0, 1, 2, ...

loc (Label Location)

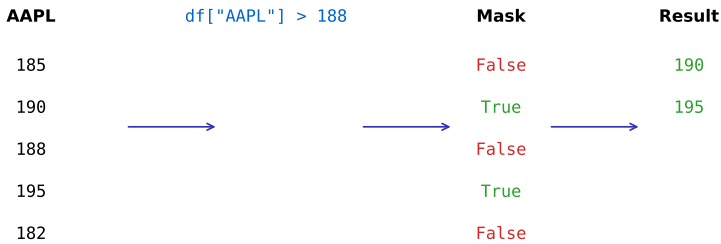
Label-based indexing

```
df.loc['2024-01-02']
df.loc[:, 'AAPL']
```

Uses: dates, names

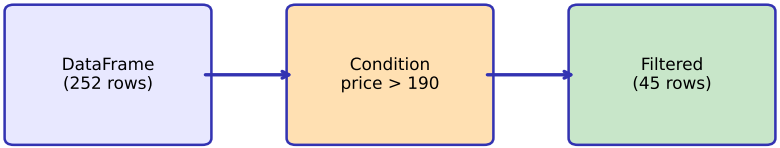
iloc: exclusive end | loc: inclusive end

Boolean Masking



Boolean mask filters rows where condition is True

Conditional Filtering Flow




```
df_filtered = df[df["AAPL"] > 190]
```

Multiple Conditions

AND: &

```
(df["AAPL"] > 185) &
(df["MSFT"] > 380)
```

OR: |

```
(df["AAPL"] > 200) |
(df["MSFT"] > 400)
```

NOT: ~

```
~(df["AAPL"] > 190)
```

Always use parentheses around each condition!

Chained Filtering with query()

Traditional

```
df[(df["AAPL"] > 185) &
   (df["Volume"] > 1e6)]
```

query() Method

```
df.query("AAPL > 185 and
          Volume > 1e6")
```

query() is more readable for complex filters

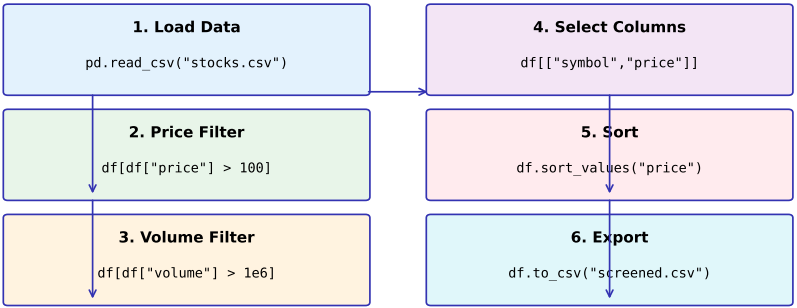
Membership: isin()

```
df[df["Symbol"].isin(["AAPL", "MSFT", "GOOGL"])]
```

Selection Methods Comparison

Method	Use Case	Returns
`df["col"]`	Single column	Series
`df[["col1","col2"]]`	Multiple columns	DataFrame
`df.iloc[0]`	Row by position	Series
`df.loc["date"]`	Row by label	Series
`df[df.col > x]`	Filter rows	DataFrame

Stock Screening Workflow



Combine filters to build powerful stock screeners

Hands-on Exercise (25 min)

Build a stock screener:

1. Load stock data from CSV
2. Select only AAPL and MSFT columns
3. Filter rows where AAPL ¿ 185
4. Filter rows where AAPL ¿ 185 AND MSFT ¿ 375
5. Use query() for the same filter
6. Select first 10 trading days using iloc
7. Sort by AAPL price descending

Stock screeners are fundamental tools in finance

Lesson Summary

Key Takeaways:

- df["col"] selects single column as Series
- iloc uses integer positions; loc uses labels
- Boolean conditions create True/False masks
- Combine conditions with & (and) and — (or)
- query() is cleaner for complex filters

Next Lesson: Missing Data and Cleaning

Week 1 complete! You can now load, explore, and filter data