Lesson 07: Missing Data and Cleaning

Data Science with Python – BSc Course

45 Minutes

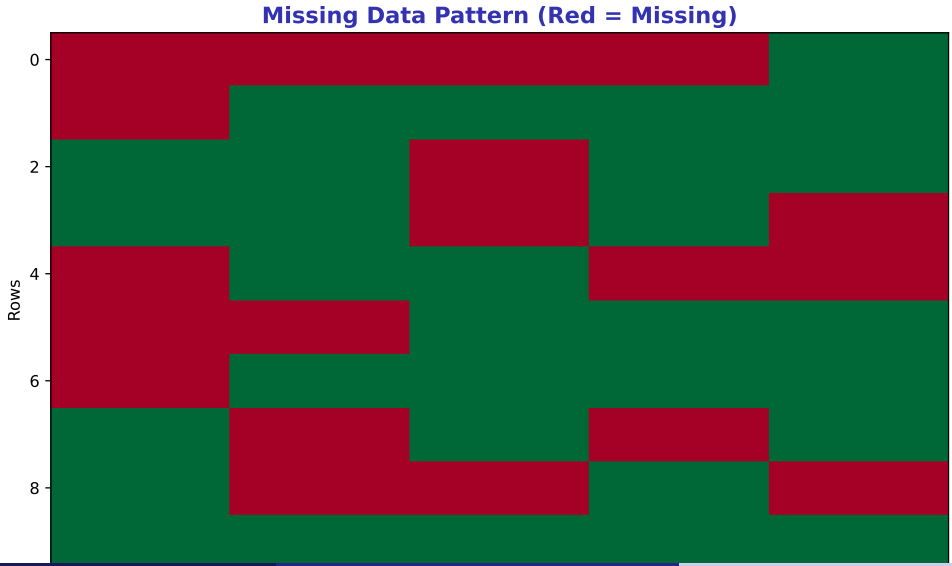
Learning Objectives

After this lesson, you will be able to:

- isna()/isnull() for detection
- fillna() methods (ffill, bfill, mean)
- dropna() to remove missing
- Handling duplicates
- Data type conversion

Finance application: Stock data processing and analysis

Missing Patterns

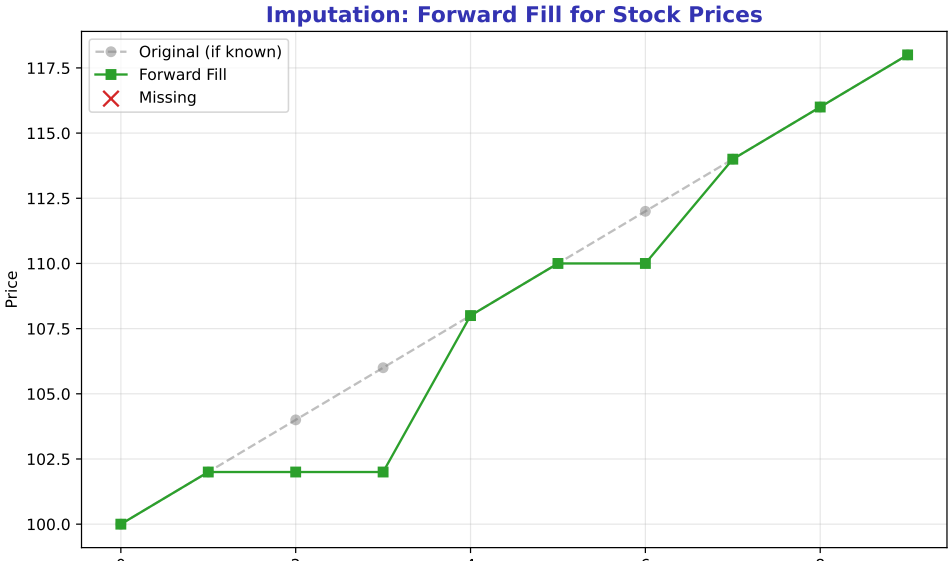


fillna() Methods Comparison

`fillna(0)`	Fill with constant value
`fillna(method="ffill")`	Forward fill (last valid)
`fillna(method="bfill")`	Backward fill (next valid)
`fillna(df.mean())`	Fill with column mean

Data Quality Checklist

- [ ] Check for missing values (isna())
- [ ] Identify duplicates (duplicated())
- [ ] Verify data types (dtypes)
- [ ] Check value ranges (describe())
- [ ] Validate dates (date parsing)
- [ ] Look for outliers


Imputation: Forward Fill for Stock Prices
Original (if known)
Forward Fill
Missing
Price
117.5
115.0
112.5
110.0
107.5
105.0
102.5
100.0

dropna() Behavior

Before

185	376	NaN
190	NaN	142
188	380	140

After dropna()

```
188  380  140
```

(1 row remains)

Detecting Duplicates

```
df.duplicated()
```

Returns boolean mask

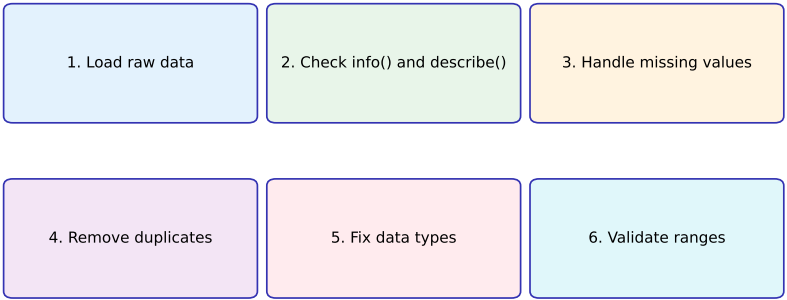
```
df.drop_duplicates()
```

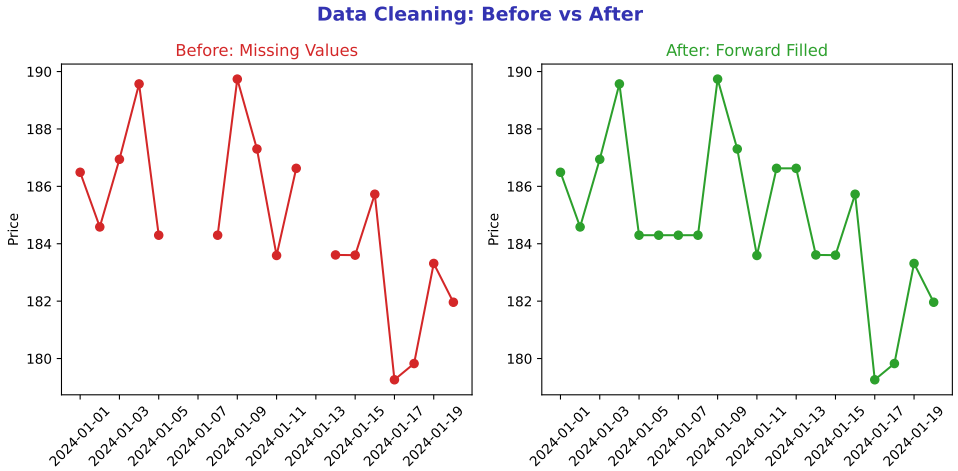
Removes duplicate rows

```
df.drop_duplicates(subset=['Date'])
```

Check specific columns only

Data Cleaning Workflow





Key concept for financial data analysis

Lesson Summary

Key Takeaways:

- isna()/isnull() for detection
- fillna() methods (ffill, bfill, mean)
- dropna() to remove missing
- Handling duplicates
- Data type conversion

Practice: Apply these concepts to the stock price dataset.