Detecting Duplicates in Stock Data

Index	Date	Price	Volume	Duplicate?
0	2024-01-01	$100.0	1,000	No
1	2024-01-02	$101.5	1,500	No
2	2024-01-02	$101.5	1,500	Yes
3	2024-01-03	$102.0	2,000	No
4	2024-01-03	$102.3	2,100	No
5	2024-01-03	$102.0	2,000	Yes
6	2024-01-04	$103.0	1,800	No
7	2024-01-05	$104.5	2,200	No
8	2024-01-05	$104.5	2,200	Yes

Total Rows: 9 | Duplicates: 3 | Unique: 6

df.duplicated() marks all rows after first occurrence as duplicates

Duplicate Handling Methods

duplicated()	->	**3 duplicates found**

Marks duplicates (keeps first)

duplicated(keep="last")	->	**3 duplicates found**

Marks duplicates (keeps last)

duplicated(keep=False)	->	**6 involved in duplication**

Marks ALL duplicates

drop_duplicates()	->	**6 rows remain**

Removes duplicate rows

drop_duplicates(subset=["Date"])	->	**5 rows remain**

Removes by Date only