

```
1 # duplicated()
  # drop_duplicates()
```

```
2 import pandas as pd
```

```
3 s = pd.Series([1, 2, 3, 2, 3, 3])
s
```

```
3 0    1
   1    2
   2    3
   3    2
   4    3
   5    3
dtype: int64
```

```
4 # 检查重复值 重复为ture 重复的第一个元素是False
  s.duplicated(keep='first')
```

```
4 0    False
   1    False
   2    False
   3     True
   4     True
   5     True
dtype: bool
```

```
5 # 删除重复值 保留第一个元素 也就是删除False
  s.drop_duplicates(keep='first')
```

```
5 0    1
   1    2
   2    3
dtype: int64
```

```
6 # 由后向前检查重复值
  s.duplicated(keep='last')
```

```
6 0    False
   1     True
   2     True
   3    False
   4     True
   5    False
dtype: bool
```

```
7 # 保留最后一个重复值 其它删除
  s.drop_duplicates(keep='last')
```

```
7 0    1
   3    2
   5    3
dtype: int64
```

```
8 # 检查重复值 只要重复都为False
  s.duplicated(keep=False)
```

```
8 0    False
   1     True
   2     True
   3     True
   4     True
```

```
5      True
dtype: bool
```

```
9  # 删除所有重复值
   s.drop_duplicates(keep=False)
```

```
9  0      1
   dtype: int64
```

```
10 df = pd.DataFrame(
    {
        'A': [1, 2, 2, 2, 3],
        'B': [4, 5, 6, 6, 6],
    }
)
df
```

10

| | A | B |
|---|---|---|
| 0 | 1 | 4 |
| 1 | 2 | 5 |
| 2 | 2 | 6 |
| 3 | 2 | 6 |
| 4 | 3 | 6 |

```
11 # 指定检查A列重复值
    df.duplicated(subset='A')
```

```
11 0      False
    1      False
    2       True
    3       True
    4      False
dtype: bool
```

```
12 # 指定检查B列重复值
    df.duplicated(subset='B')
```

```
12 0      False
    1      False
    2      False
    3       True
    4       True
dtype: bool
```

```
13 # 检查AB列相同的
    df.duplicated(subset=['A', 'B'])
```

```
13 0      False
    1      False
    2      False
    3       True
    4      False
dtype: bool
```

```
14 # 删除AB列相同的 保留第一个值
    df.drop_duplicates(subset=['A', 'B'])
```

14

| | A | B |
|--|---|---|
|--|---|---|

| | A | B |
|---|---|---|
| 0 | 1 | 4 |
| 1 | 2 | 5 |
| 2 | 2 | 6 |
| 4 | 3 | 6 |

```
15 # 删除AB列相同 保留第一个值
df.drop_duplicates(
    subset=['A', 'B'],
    keep='last'
)
```

15

| | A | B |
|---|---|---|
| 0 | 1 | 4 |
| 1 | 2 | 5 |
| 3 | 2 | 6 |
| 4 | 3 | 6 |

```
16 # 删除AB列相同 保留第最后一个值
df.drop_duplicates(
    subset=['A', 'B'],
    keep=False
)
```

16

| | A | B |
|---|---|---|
| 0 | 1 | 4 |
| 1 | 2 | 5 |
| 4 | 3 | 6 |

```
17 # 只要是重复的就删除
df.drop_duplicates(
    subset=['A', 'B'],
    keep=False,
    ignore_index=True
)
```

TypeError

Traceback (most recent call last)

```
<ipython-input-17-547e154ef6c6> in <module>
      3     subset=['A', 'B'],
      4     keep=False,
----> 5     ignore_index=True
      6 )
      7
```

TypeError: drop_duplicates() got an unexpected keyword argument 'ignore_index'

```
# 默认删除AB列相同的 保留第一个  
df.drop_duplicates()
```