[**pandas**](https://so.csdn.net/so/search?q=pandas&spm=1001.2101.3001.7020) **根据某列去重**

*drop\_duplicates(subset=[‘comment’], keep=‘first’, inplace=True)*

参数：

* subset ： 列表的形式填写要进行去重的列名，默认为 None ，表示根据所有列进行。
* keep ： 可选参数有三个： first 、 last 、 False ， 默认值 first 。其中，  
  （1） first 表示： 保留第一次出现的重复行，删除后面的重复行。  
  （2） last 表示： 删除重复项，保留最后一次出现。  
  （3） False 表示： 删除所有重复项。
* inplace ：默认为 False ，删除重复项后返回副本。 True ，直接在原数据上删除重复项。

**✪ 栗子 ✪**  
首先创建一个 DataFrame 。

|  |
| --- |
| Python import pandas as pd dic = {'name':['a', 'b', 'c', 'd'], 'comment':['abc', '真棒', '真棒', '123']} df = pd.DataFrame(dic) df |



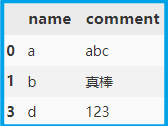
使用 drop\_duplicates 去除重复值，如果不指明 subset ，那么默认 **根据所有列来考虑** ，即当某两行数据所有列都重复时进行去重。

|  |
| --- |
| Python df.drop\_duplicates(keep='first', inplace=True) |



subset 默认为 None ，根据所有列考虑， 1 , 2 行虽然 comment 相同，但 name 不相同，故保留，使用时根据具体情况进行选择。  
  
  
现在设置 subset 为 comment 即可删除该列重复值。

|  |
| --- |
| Python df.drop\_duplicates(subset=['comment'], keep='first', inplace=True) |



注意:此时索引没有重置，如有需要可使用 reset\_index() 重置索引。

|  |
| --- |
| Python df.reset\_index(drop=True, inplace=True) |



**对于刚入门 Python 或是想要入门 Python 的小伙伴，可以通过下方 小卡片联系作者 ，一起交流学习，都是从新手走过来的，有时候一个简单的问题卡很久，但可能别人的一点拨就会恍然大悟，由衷的希望大家能够共同进步。另外还有本人整理的近千套模板，百本优质电子书资源，等你领取！**

👇🏻 **关注小卡片，一起学习Python，领取资料** 👇🏻

文章知识点与官方知识档案匹配，可进一步学习相关知识

[Python入门技能树](https://edu.csdn.net/skill/python/python-3-203)  [结构化数据分析工具Pandas](https://edu.csdn.net/skill/python/python-3-203)  [Pandas概览](https://edu.csdn.net/skill/python/python-3-203) 107735 人正在系统学习中



Python新视野



微信公众号



定期分享数据分析相关干货和其他有趣的内容