Delete duplicate rows with the same value in all columns in pandas

Ask Question

▲

1

▼

☆

I have a dataframe with about a half a million rows. As I could see, there are plenty of duplicate rows, so how can I drop duplicate rows that have the same value in all of the columns (about 80 columns), not just one?

df:

```
period_start_time      id    val1    val2    val3
06.13.2017 22:00:00    i53   32      2       10
06.13.2017 22:00:00    i32   32      2       10
06.13.2017 22:00:00    i32   4       2       8
06.13.2017 22:00:00    i32   4       2       8
06.13.2017 22:00:00    i32   4       2       8
06.13.2017 22:00:00    i20   7       7       22
06.13.2017 22:00:00    i20   7       7       22
```

Desired output:

```
period_start_time      id    val1    val2    val3
06.13.2017 22:00:00    i53   32      2       10
06.13.2017 22:00:00    i32   32      2       10
06.13.2017 22:00:00    i32   4       2       8
06.13.2017 22:00:00    i20   7       7       22
```

python   pandas   dataframe

asked Jun 26 '17 at 12:33

jovicbg
117   8   28

did you try calling `df.drop_duplicates()` ? –
EdChum Jun 26 '17 at 12:35 ✎

1 Answer

▲

4

Use `drop_duplicates` :

```
df = df.drop_duplicates()
print (df)
```

```
   2   06.13.2017 22:00:00   i32       4
   5   06.13.2017 22:00:00   i20       7
```

answered Jun 26 '17 at 12:35

jezrael
359k   26   326   404