

```
In [1]: import numpy as np
import pandas as pd
import json
import os
pd.options.mode.chained_assignment = None

from sklearn.utils import shuffle
```

```
In [2]: data = pd.read_csv('2023-04-14T0721_Grades-Trực quan hoá dữ liệu.csv')
```

```
In [3]: len(data)
```

```
Out[3]: 42
```

```
In [4]: data.columns
```

```
Out[4]: Index(['Student', 'ID', 'SIS User ID', 'SIS Login ID', 'Section',
'LTI Attendance (34474)', 'Roll Call Attendance (35546)',
'Assignment 1 (35967)', 'Assignment 2 (36357)', 'Assignment_3 (40427
)',
'Assignment_4 (42955)', 'Assignment_5 (43682)', 'Mid-term Exam (4451
9)',
'Assignment_6 (45740)', 'Bài tập Current Score',
'Bài tập Unposted Current Score', 'Bài tập Final Score',
'Bài tập Unposted Final Score', 'Bài tập Current Score.1',
'Bài tập Unposted Current Score.1', 'Bài tập Final Score.1',
'Bài tập Unposted Final Score.1',
'CC1 (Tủ Phòng Đào Tạo) Current Score',
'CC1 (Tủ Phòng Đào Tạo) Unposted Current Score',
'CC1 (Tủ Phòng Đào Tạo) Final Score',
'CC1 (Tủ Phòng Đào Tạo) Unposted Final Score',
'CC2 (Tủ Phòng Đào Tạo) Current Score',
'CC2 (Tủ Phòng Đào Tạo) Unposted Current Score',
'CC2 (Tủ Phòng Đào Tạo) Final Score',
'CC2 (Tủ Phòng Đào Tạo) Unposted Final Score',
'LT_BAI1 (Tủ Phòng Đào Tạo) Current Score',
'LT_BAI1 (Tủ Phòng Đào Tạo) Unposted Current Score',
'LT_BAI1 (Tủ Phòng Đào Tạo) Final Score',
'LT_BAI1 (Tủ Phòng Đào Tạo) Unposted Final Score',
'THI1 (Tủ Phòng Đào Tạo) Current Score',
'THI1 (Tủ Phòng Đào Tạo) Unposted Current Score',
'THI1 (Tủ Phòng Đào Tạo) Final Score',
'THI1 (Tủ Phòng Đào Tạo) Unposted Final Score', 'Current Score',
'Unposted Current Score', 'Final Score', 'Unposted Final Score'],
dtype='object')
```

```
In [9]: keep = ['Student']
```

```
In [10]: keep_df = data.filter(keep)
```

```
In [12]: keep_df = keep_df.drop([0,41])
```

```
In [25]: group = [1, 2, 3, 4, 5]
```

```
In [26]: for i, sub_df in enumerate(np.array_split(shuffle(keep_df), len(group))):
        keep_df.loc[sub_df.index, 'Group'] = str(group[i])
```

```
In [27]: keep_df
```

```
Out[27]:
```

	Student	Group
1	Anh, Bùi Huy Quyền	5
2	Anh, Vũ Hoàng	3
3	Chinh, Nguyễn Thị	5
4	Công, Nguyễn Chí	2
5	Công, Nguyễn Tiến	1
6	Cường, Hoàng Thế	1
7	Cường, Nguyễn Văn	3
8	Duy, Phạm Văn	5
9	Dương, Nguyễn Hoàng	4
10	Hiếu, Nguyễn Minh	2
11	Hiếu, Trần Minh	3
12	Hiệp, Nguyễn Tuấn	4
13	Huấn, Nguyễn Văn	1
14	Hưng, Nguyễn Văn	4
15	Hưng, Nguyễn Vũ Duy	5
16	Hải, Nguyễn Đăng Đức	5
17	Khiêm, Trịnh Gia	3
18	Kiên, Đoàn Trung	3
19	Linh, Nguyễn Lý Khánh	1
20	Linh, Đặng Văn Tuấn	2
21	Long, Nguyễn Hải	2
22	Long, Nguyễn Việt	4
23	Long, Nguyễn Văn	5
24	Miễn, Triệu Thị	2
25	Mạnh, Nguyễn Văn	4
26	Nga, Nguyễn Thị	4
27	Ngọc, Nguyễn Thị	1
28	Phát, Nguyễn	2
29	Phương, Trần Hữu Duy	1

30	Quang, Vi Đăng	2
31	Quyên, Đào Nguyễn Thảo	2
32	Quân, Nguyễn Đình	3
33	Thành, Trần Chí	3
34	Thắng, Nguyễn Việt	3
35	Thọ, Đặng Tiến	4
36	Trang, Hòa Thị Thu	5
37	Tuấn, Nguyễn Văn	1
38	Tài, Nguyễn Hữu	4
39	Tú, Lê Hoàng Ngọc	1
40	Vũ, Cao Đức	5

```
In [28]: keep_df.groupby('Group').size()
```

```
Out[28]: Group
1      8
2      8
3      8
4      8
5      8
dtype: int64
```

```
In [29]: keep_df.to_csv('list_group.csv')
```

```
In [ ]:
```