

レシピ55 グループ分けの後でMultiIndex解消（P151）

- **groupby**を使うと、カラムや行に必ず**MultiIndex**ができる。カラム名がごちゃごちゃする
- MultiIndexの解除方法を示す

(1) サンプル作成

- 集約カラムが上段、集約関数は下段になっている

In [62]:

```
airline_info = flights.groupby(['AIRLINE', 'WEEKDAY'])\
    .agg({'DIST':['sum', 'mean'],
         'ARR_DELAY':['min', 'max']})\
    .astype(int)

airline_info.head(7)
```

Out[62]:

		DIST	ARR_DELAY		
		sum	mean	min	max
AIRLINE	WEEKDAY				
AA	1	1455386	1139	-60	551
	2	1358256	1107	-52	725
	3	1496665	1117	-45	473
	4	1452394	1089	-46	349
	5	1427749	1122	-41	732
	6	1265340	1124	-50	858
	7	1461906	1100	-49	626

(2) 行もカラムも2階層のMultiIndexになっている。これを1階層にまとめる

- カラム名はMultiIndexの `get_level_values` を使う
- 各レベルの出力を表示し、2階層分をまとめて新たなカラム値を作る

In [63]:

```
level0 = airline_info.columns.get_level_values(0)
level1 = airline_info.columns.get_level_values(1)

print(level0)
print(level1)

Index(['DIST', 'DIST', 'ARR_DELAY', 'ARR_DELAY'], dtype='object')
Index(['sum', 'mean', 'min', 'max'], dtype='object')
```

In [64]:

```
airline_info.columns = level0 + '_' + level1
airline_info.head(2)
```

Out[64]:

		DIST_sum	DIST_mean	ARR_DELAY_min	ARR_DELAY_max
AIRLINE	WEEKDAY				
AA	1	1455386	1139	-60	551
	2	1358256	1107	-52	725

(3) `reset_index()` で行ラベルを1階層にする

In [65]:

```
airline_info.reset_index().head(7)
```

Out[65]:

	AIRLINE	WEEKDAY	DIST_sum	DIST_mean	ARR_DELAY_min	ARR_DELAY_max
0	AA	1	1455386	1139	-60	551
1	AA	2	1358256	1107	-52	725
2	AA	3	1496665	1117	-45	473
3	AA	4	1452394	1089	-46	349
4	AA	5	1427749	1122	-41	732
5	AA	6	1265340	1124	-50	858
6	AA	7	1461906	1100	-49	626

(補足)

- MultiIndexにたくない場合は、gropubyの中でパラメータ `as_index=False` をセットすればよい
- ただし、グループ分けカラムが1つのときしか使えない（2つ以上はエラーになる）
- よって、利用価値は低い

In [66]:

```
# 変更前
flights.groupby(['AIRLINE'])[['DIST', 'ARR_DELAY']].agg('mean').head(2)
```

Out[66]:

		DIST	ARR_DELAY
AIRLINE			
AA	1114.347865		5.542661
AS	1065.884115		-0.833333

In [67]:

```
# MultiIndexを回避
flights.groupby(['AIRLINE'], as_index=False)[['DIST', 'ARR_DELAY']].agg('mean').head(2)
```

Out[67]:

	AIRLINE	DIST	ARR_DELAY
0	AA	1114.347865	5.542661
1	AS	1065.884115	-0.833333