

## 3-8 クロス集計

### groupby メソッドで集約する

groupby() メソッドを使用して DataFrame を集約します。指定された列で集約されたオブジェクト DataFrameGroupBy が返されます。リスト 3.8.1 では type 列で集約されたオブジェクトを作成しています (結果はリスト 3.8.2)。

リスト 3.8.1: DataFrame の集約

```
import os
import pandas as pd
import numpy as np

base_url = 'https://raw.githubusercontent.com/practical-jupyter/sample-data/master/anime/'
anime_master_csv = os.path.join(base_url, 'anime_master.csv')
df = pd.read_csv(anime_master_csv)

grouped = df.groupby('type')
type(grouped)
```

リスト 3.8.2: DataFrame の集約 (結果)

```
pandas.core.groupby.DataFrameGroupBy
```

前節 (3-7: 統計量を求める) 同様の数学的、統計的なメソッドが使用できます (リスト 3.8.3 と表 3.29、リスト 3.8.4 と表 3.30)。詳細は pandas documentation API Reference<sup>注3.17</sup> を参照してください。

リスト 3.8.3: 集約されたデータの平均値

```
grouped.mean().round(1)
```

表 3.29: 集約されたデータの平均値 (結果)

type	anime_id	episodes	rating	members
Movie	14322.5	1.1	6.3	10654.0
Music	22495.1	1.1	5.6	1273.0
ONA	22738.0	6.8	5.6	4401.8
OVA	12207.7	2.5	6.5	6849.5
Special	16802.3	2.5	6.5	7424.6
TV	10929.6	37.5	6.9	41832.3

リスト 3.8.4: 集約されたデータの基本統計量

```
grouped.describe().round(1).head(16)
```

表 3.30: 集約されたデータの基本統計量 (結果)

type		anime_id	episodes	members	rating
Movie	count	2220.0	2220.0	2220.0	2220.0
	mean	14322.5	1.1	10654.0	6.3
	std	10925.7	2.2	31603.6	1.2
	min	5.0	1.0	13.0	2.5
	25%	4396.8	1.0	119.0	5.4
	50%	10677.5	1.0	489.5	6.5
	75%	24071.5	1.0	4239.0	7.3
	max	34201.0	100.0	466254.0	10.0
Music	count	485.0	485.0	485.0	485.0
	mean	22495.1	1.1	1273.0	5.6
	std	10175.0	1.3	4489.0	1.0
	min	731.0	1.0	24.0	3.3
	25%	12101.0	1.0	97.0	5.0
	50%	24903.0	1.0	226.0	5.6
	75%	31925.0	1.0	797.0	6.2
	max	34412.0	24.0	71136.0	8.4

#### NOTE

pandas のバージョンによっては describe() メソッドの表示形式が異なることがあります。

複数の要素で集約する場合、groupby() メソッドの引数にリストを渡します。リスト 3.8.5 では type 列と episodes 列に基づいてデータを集約し、平均を求めています (結果は表 3.31)。

リスト 3.8.5: 集約された複数列データの平均値

```
df.groupby(['type', 'episodes']).mean().round(1).head(20)
```

表 3.31: 集約された複数列データの平均値 (結果)

type	episodes	anime_id	rating	members
Movie	1	14320.0	6.3	10588.6
	2	13802.0	6.9	6638.9
	3	11339.3	6.7	53598.1
	4	15723.5	7.3	3566.5
	5	12558.3	6.1	3641.0
	6	8433.5	6.0	178.5
	7	13602.5	6.9	11989.5
	9	8928.0	6.2	267.0
	10	31020.0	6.9	57.0

注 3.17 GroupBy [URL](http://pandas.pydata.org/pandas-docs/stable/api.html#groupby) http://pandas.pydata.org/pandas-docs/stable/api.html#groupby



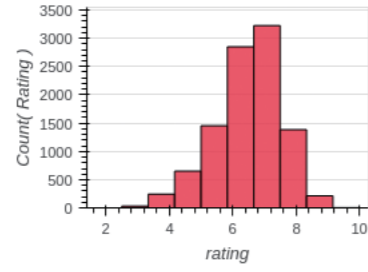


Genre

Movie icons

Genre: **Movie**

Æ q S { £



Genre

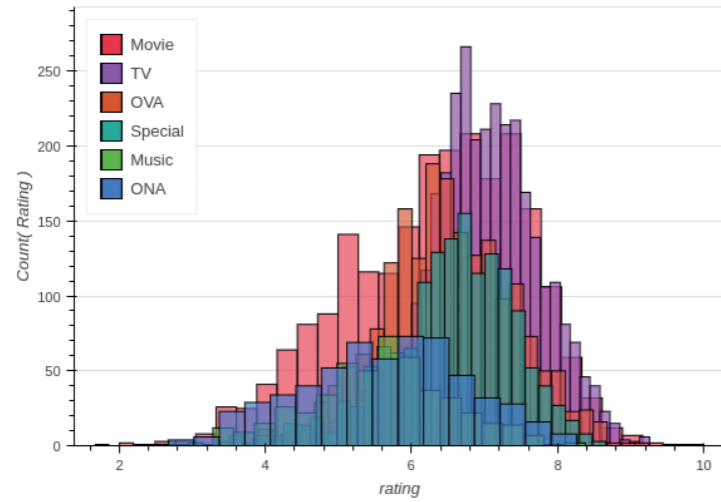


Genre

Movie icons

Genre: **Movie**

Æ q S { £



Genre



Genre

Movie icons

Genre: **Movie**

Æ { q£S



Genre

Movie icons

Genre: **Movie**

Genre

S { £

w

1

2

3

4

5

6

7

8

9

10

A