

これで、会員期間の列を追加できました。
それでは、いよいよ利用履歴も加味した形で顧客の分析を行っていきます。
まずは、いつものように各種統計量を把握しましょう。

🔪 ノック29： 顧客行動の各種統計量を把握しよう

まずは、全体の数を押さえていきましょう。
データ加工により追加した、mean、median、max、minはノック7で扱ったdescribeを用いて数値を把握しましょう。routine_flgに関しては、それぞれのflg毎に顧客数を集計してみましょう。

```
customer_join[["mean", "median", "max", "min"]].describe()
```

■図3-12：各種統計量の計算結果

ノック29：顧客行動の各種統計量を把握しよう				
In [32]:	customer_join[["mean", "median", "max", "min"]].describe()			
Out[32]:				
	mean	median	max	min
count	4192.000000	4192.000000	4192.000000	4192.000000
mean	5.333127	5.250596	7.823950	3.041269
std	1.777533	1.874874	2.168959	1.951565
min	1.000000	1.000000	1.000000	1.000000
25%	4.250000	4.000000	7.000000	2.000000
50%	5.000000	5.000000	8.000000	3.000000
75%	6.416667	6.500000	9.000000	4.000000
max	12.000000	12.000000	14.000000	12.000000

実行すると、ノック7の時と同様に、件数、平均値などが表示されます。ここで用いた列の名前がmean、median、max、minであるため、少し混乱しやすいので注意してください。

列名のmeanは顧客の月内平均利用回数であり、行にあるmeanは顧客の月内平均利用回数の平均です。つまり、表示されている値は、一人当たりの月内平均利用回数です。平均値、中央値に大きな違いはなく、顧客一人当たりおよそ5回程度の月内利用回数となっていることがわかります。

続けて、routine_flgを集計してみましょう。

```
customer_join.groupby("routine_flg").count()["customer_id"]
```

実行すると、0が779に対して、1が3413となっており、定期的に利用しているユーザーの方が多いことがわかります。

最後に、会員期間の分布を見ておきましょう。

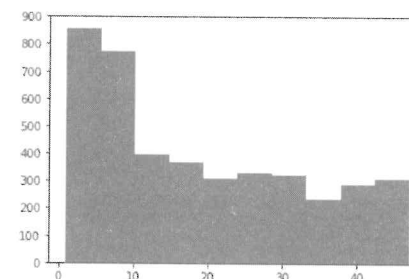
分布は、数字ではわかりにくいので、matplotlibを使ってヒストグラムを作成します。

```
import matplotlib.pyplot as plt
%matplotlib inline
plt.hist(customer_join["membership_period"])
```

■図3-13：会員期間の分布

```
In [35]: import matplotlib.pyplot as plt
%matplotlib inline
plt.hist(customer_join["membership_period"])

Out[35]: (array([857., 774., 395., 368., 311., 331., 323., 237., 288., 308.]),
array([ 1.,  5.6, 10.2, 14.8, 19.4, 24., 28.6, 33.2, 37.8, 42.4, 47. ]),
<a list of 10 Patch objects>)
```



実行するとヒストグラムが表示されます。第1部でも用いたmatplotlibですが、ここではhist()を用いてヒストグラムを作成しています。この結果を見ると、会員期間が10ヶ月以内のユーザーの方が多く、10ヶ月以上はほぼ横ばいで分布していることがわかります。これは、短期でユーザーが離れていく業界であることを示唆しています。