

それでは、離れていくユーザーの特徴はどういったものなのかを、次に調べてみましょう。

### ノック30： 退会ユーザーと継続ユーザーの違いを 把握しよう

これまでは全体の数字を把握してきました。本章の最後に、これまでの知見を頭に入れながら、退会ユーザーの特徴を探していきましょう。これまでのノックでやってきたことを、退会ユーザーと継続ユーザーに分けて比較しましょう。

まずは、退会ユーザーと継続ユーザーの数は、**ノック23**で集計済みで、退会ユーザーが1350人で、継続ユーザーが2840人になります。

退会ユーザーと継続ユーザーを分けて、describeで比較してみましょう。

```
customer_end = customer_join.loc[customer_join["is_deleted"]==1]
customer_end.describe()
```

```
customer_stay = customer_join.loc[customer_join["is_deleted"]==0]
customer_stay.describe()
```

#### ■3-14：統計量の集計結果

ノック30：退会ユーザーと継続ユーザーの違いを把握しよう								
In [40]: customer_end = customer_join.loc[customer_join["is_deleted"]==1] customer_end.describe()								
Out[40]:	is_deleted	price	mean	median	max	min	routine_flg	membership_period
count	1350.0	1350.000000	1350.000000	1350.000000	1350.000000	1350.000000	1350.000000	1350.000000
mean	1.0	8595.555556	3.865474	3.621852	6.461481	1.821481	0.456298	8.026667
std	0.0	1849.163652	1.246385	1.270847	2.584021	0.976361	0.498271	5.033692
min	1.0	6000.000000	1.000000	1.000000	1.000000	1.000000	0.000000	1.000000
25%	1.0	6000.000000	3.000000	3.000000	4.000000	1.000000	0.000000	4.000000
50%	1.0	7500.000000	4.000000	4.000000	7.000000	2.000000	0.000000	7.000000
75%	1.0	10500.000000	4.666667	4.500000	8.000000	2.000000	1.000000	11.000000
max	1.0	10500.000000	9.000000	9.000000	13.000000	8.000000	1.000000	23.000000
In [41]: customer_stay = customer_join.loc[customer_join["is_deleted"]==0] customer_stay.describe()								
Out[41]:	is_deleted	price	mean	median	max	min	routine_flg	membership_period
count	2842.0	2842.000000	2842.000000	2842.000000	2842.000000	2842.000000	2842.000000	2842.000000
mean	0.0	8542.927516	6.030238	6.024279	8.471147	3.020690	0.984166	23.970443
std	0.0	1977.189779	1.553587	1.599765	1.571048	2.030488	0.124855	13.746761
min	0.0	6000.000000	3.166667	3.000000	5.000000	1.000000	0.000000	1.000000
25%	0.0	6000.000000	4.833333	5.000000	7.000000	2.000000	1.000000	12.000000
50%	0.0	7500.000000	5.833333	5.500000	8.000000	3.000000	1.000000	24.000000
75%	0.0	10500.000000	7.178030	7.000000	10.000000	5.000000	1.000000	35.000000
max	0.0	10500.000000	12.000000	12.000000	14.000000	12.000000	1.000000	47.000000

実行すると、退会ユーザーは、月内の利用回数の平均値、中央値、最大値、最小値いずれも継続ユーザーよりも低く出ています。特に、平均値や中央値は1.5倍程度の違いが出ていることがわかります。一方で月内最大利用回数の平均値に関しても継続ユーザーの方が高く出ていますが、退会ユーザーでも6.4程度はあります。routine\_flgの平均値に大きく差が出ており、継続ユーザーは0.98と多くのユーザーが定期的に利用していることが伺えますが、退会ユーザーは、0.45となり、およそ半分くらいのユーザーはランダムに利用していると考えられます。

このように、行動データを紐解いていくと、退会ユーザーとの違いが見えてきそうな気がしませんか。さらに、期間で絞ったり、会員の在籍期間ごとに深掘りしていくとさらに見えてくると思います。

最後に、ここまで使用したcustomer\_joinをcsv出力しておきましょう。

```
customer_join.to_csv("customer_join.csv", index=False)
```



これで、10本のノックは終了です。

ここでは、まずはデータを理解するという作業をご理解いただけましたでしょうか。機械学習は、データをもとにコンピュータが推論を行います。そのため、どのようなデータで学習させるかという部分がとても重要となります。そのためにも、できる限りデータを理解し、特徴となりそうな変数を掴んでおくのが重要です。

次章では、いよいよ機械学習を行っていきます。