

## ノック25： 利用履歴データを集計しよう

まずは、簡単に時間的な要素を取り入れていきましょう。

今回は、月利用回数の**平均値**、**中央値**、**最大値**、**最小値**と定期的に利用しているかのフラグを作成し、顧客データに追加していきましょう。まずは集計を行います。

最初に、顧客ごとの月利用回数を集計したデータを作成していきます。

```
uselog["usedate"] = pd.to_datetime(uselog["usedate"])
uselog["年月"] = uselog["usedate"].dt.strftime("%Y%m")
uselog_months = uselog.groupby(["年月", "customer_id"], as_index=False).count()
uselog_months.rename(columns={"log_id": "count"}, inplace=True)
del uselog_months["usedate"]
uselog_months.head()
```

■図3-6：月/顧客毎利用回数の集計結果

### ノック25：利用履歴データを集計しよう

```
In [31]: uselog["usedate"] = pd.to_datetime(uselog["usedate"])
uselog["年月"] = uselog["usedate"].dt.strftime("%Y%m")
uselog_months = uselog.groupby(["年月", "customer_id"], as_index=False).count()
uselog_months.rename(columns={"log_id": "count"}, inplace=True)
del uselog_months["usedate"]
uselog_months.head()
```

```
Out[31]:
```

	年月	customer_id	count
0	201804	AS002855	4
1	201804	AS009013	2
2	201804	AS009373	3
3	201804	AS015315	6
4	201804	AS015739	7

1行目、2行目で、年月(201804など)のデータを作成し、年月かつ顧客ID毎にgroupbyで集計を行っています。集計はlog\_idのカウントを取れば良いので、余分なusedateは削除してあります。

顧客AS002855が2018年4月に4回利用していることがわかります。ここから、顧客毎に絞り込み、平均値、中央値、最大値、最小値を集計します。

```
uselog_customer = uselog_months.groupby("customer_id").agg(["mean", "median", "max", "min"])[["count"]]
uselog_customer = uselog_customer.reset_index(drop=False)
uselog_customer.head()
```

■図3-7：顧客毎の月内利用回数の集計結果

```
In [45]: uselog_customer = uselog_months.groupby("customer_id").agg(["mean", "median", "max", "min"])[["count"]]
uselog_customer = uselog_customer.reset_index(drop=False)
uselog_customer.head()
```

```
Out[45]:
```

	customer_id	mean	median	max	min
0	AS002855	4.500000	5.0	7	2
1	AS008805	4.000000	4.0	8	1
2	AS009013	2.000000	2.0	2	2
3	AS009373	5.083333	5.0	7	3
4	AS015233	7.545455	7.0	11	4

1行目で、平均値、中央値、最大値、最小値をgroupbyで集計しています。2行目は、groupbyをした影響でindexに入っているcustomer\_id列をカラムに変更して、indexの振り直しを行っています。顧客AS002855は、平均値4.5、中央値5、最大値7、最小値2であることがわかります。

これで、顧客毎の月内の利用回数の集計ができました。

次は、定期的にジムを利用している場合のフラグ作成を行きましょう。

## ノック26： 利用履歴データから定期利用フラグを作成しよう

ジムの場合、習慣化が継続の重要なファクターの1つであると考えられます。そこで、定期的にジムを利用しているユーザーを特定してみましょう。定期的をどう定義にするかによりますが、毎週同じ曜日に来ているかどうかで判断してみます。

月によって定期的に利用しているかのばらつきもあるかとは思いますが、ここ