

り込むかのどちらかになります。②の場合、2019年3月に退会したユーザーはカウントされないの、注意が必要です。どちらで集計するかは、目的にもよりますが、今回は、最新月に在籍していたユーザーということで、①を採用していきます。

まずは、対象ユーザーに絞り込みを行いましょ。また、絞り込んだデータが、正しく絞り込めているのかを確認するために、end_dateのユニーク件数も確認してみましょ。

```
customer_join["end_date"] = pd.to_datetime(customer_join["end_date"])
customer_newer = customer_join.loc[(customer_join["end_date"] >= pd.to_datetime("20190331")) | (customer_join["end_date"].isna())]
print(len(customer_newer))
customer_newer["end_date"].unique()
```

出力した結果、データ件数が2953件で、end_dateのユニークは、NaTと2019-03-31のみが表示されれば正しく絞り込みができています。NaTは、datetime型の欠損値という意味で、このデータにおいては退会していない顧客を示します。

それでは、会員区分、キャンペーン区分、性別毎に全体の数を把握してみましょ。

```
customer_newer.groupby("class_name").count()["customer_id"]

customer_newer.groupby("campaign_name").count()["customer_id"]

customer_newer.groupby("gender").count()["customer_id"]
```

■ 図3-5：最新月顧客の集計結果

ノック 24：最新顧客データの基礎集計をしよう

```
In [13]: customer_join["end_date"] = pd.to_datetime(customer_join["end_date"])
customer_newer = customer_join.loc[(customer_join["end_date"] >= pd.to_datetime("20190331")) | (customer_join["end_date"].isna())]
print(len(customer_newer))
customer_newer["end_date"].unique()

2953

Out[13]: array(['NaT', '2019-03-31T00:00:00.000000000'],
              dtype='datetime64[ns]')

In [14]: customer_newer.groupby("class_name").count()["customer_id"]

Out[14]: class_name
オールタイム    1444
デイトタイム    696
ナイト         813
Name: customer_id, dtype: int64

In [15]: customer_newer.groupby("campaign_name").count()["customer_id"]

Out[15]: campaign_name
入会費半額    311
入会費無料   242
通常        2400
Name: customer_id, dtype: int64

In [16]: customer_newer.groupby("gender").count()["customer_id"]

Out[16]: gender
F    1400
M    1553
Name: customer_id, dtype: int64
```

出力結果を見ると、会員区分、性別は、ノック 23 で全体を集計した際と比率が大きく変わっていません。これは、特定の会員区分や性別が退会しているわけではないと考えられます。キャンペーン区分に関しては、若干違いがあり、全体で集計した際には、通常で入会しているユーザーが72%であったのに対して、最新月の顧客は、通常で入会しているユーザーの比率が81%となっています。入会キャンペーンは、比率が変化するくらいに、何かしらの効果を生んでいることが推測できます。

会員や性別の区分が大きく変化しているわけではないので、ここからは利用履歴データの活用を検討していきましょう。

利用履歴データで分かることを考えていきます。

利用履歴データでは、顧客データと違い、時間的な要素の分析をすることができます。例えば、月内の利用回数がどのように変化しているのか、定期的にジムを活用しているユーザーなのか等が考えられます。

それでは、利用履歴データを集計して、顧客の把握に活用していきましょう。