第3章 顧客の全体像を把握する 10本ノック

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で、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。

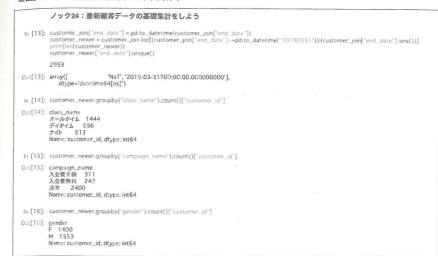
まずは、対象ユーザーに絞り込みを行いましょう。また、絞り込んだデータが、正しく絞り込めているのかを確認するために、end_dateのユニーク件数も確認してみましょう。

```
customer_join["end_date"] = pd.to_datetime(customer_join["end_date"])
customer_newer = customer_join.loc[(customer_join["end_date"] >= pd.to_dat
etime("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う。

■図3-5:最新月顧客の集計結果



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。