

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

customer_join.groupby("gender").count()["customer_id"]

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

■図3-4：顧客データの集計

ノック23：顧客データの基礎集計をしよう

```
In [8]: customer_join.groupby("class_name").count()["customer_id"]
Out[8]: class_name
オールタイム  2045
デイトタイム  1019
ナイト       1128
Name: customer_id, dtype: int64

In [9]: customer_join.groupby("campaign_name").count()["customer_id"]
Out[9]: campaign_name
入会費半額    650
入会費無料   492
通常        3050
Name: customer_id, dtype: int64

In [10]: customer_join.groupby("gender").count()["customer_id"]
Out[10]: gender
F    1983
M   2209
Name: customer_id, dtype: int64

In [11]: customer_join.groupby("is_deleted").count()["customer_id"]
Out[11]: is_deleted
0    2842
1    1350
Name: customer_id, dtype: int64
```

第1部でも取り扱ったgroupbyを用いて集計を行っています。今回は、customer_id毎にカウントを行っています。

上から見ていくと、会員クラスはオールタイムがおおよそ半数を占め、次いでナイト、最後にデイトタイムとなっています。キャンペーンに関しては、通常時の入会が多く、入会キャンペーン中に入会したのはおおよそ20%となっています。男女比に関しては、わずかに男性(M)の方が多い結果となっていることがわかります。最後に、退会しているユーザーは1350人で、2019年3月時点で在籍しているユーザーは2842人であることがわかりました。

このように見ていくと、いろんな仮説や気になる点が出てくるかと思います。

例えば、キャンペーンはいつ行われていたのか、性別と会員クラスの関係、今年度の入会人数など、多く浮かんでくるかと思います。こういった仮説や疑問点は、集計して確認するのももちろんのこと、現場の人にヒアリングすることで理解が進むことが多々あります。積極的にヒアリングをしていくと良いと思います。

ここでは、試しに入会人数を集計してみましょう。

入会人数は、start_date列が2018年4月1日から2019年3月31日までのユーザーとなります。第1部でも取り扱ったように、まずはstart_dateをdatetime型に変換した後、customer_startというデータフレームに該当ユーザーのデータを格納した上で、数を数えましょう。

```
customer_join["start_date"] = pd.to_datetime(customer_join["start_date"])
customer_start = customer_join.loc[customer_join["start_date"] > pd.to_datetime("20180401")]
print(len(customer_start))
```

実行すると、1361が出力され、入会会員は1361人で、退会した人数よりも11人多いですが、全体の会員数は変わらない結果となることがわかりました。

他にも気になる部分はありますが、一旦、ここで区切り、最新月の顧客情報を見てみましょう。気になる部分は、どんどん手を動かして自分なりに集計してみると良いと思います。

⑩ ノック24：最新顧客データの基礎集計をしてみよう

ここでは、最新月の顧客データの把握を行います。理由としては、現在の顧客データには既に離脱しているユーザーも含まれているため、単月の実態とは違いがあります。そこで、最新月の顧客データを集計することで、現状顧客の全体像を把握します。

まずは、最新月のユーザーのみに絞り込みを行います。

最新月に絞り込むためには、①2019年3月(2019年3月31日)に退会したユーザーもしくは、在籍しているユーザーで絞り込むか、②is_deleted列で絞