

## Cluster Sampling Exercise

One commonly used sampling method is cluster sampling, in which a population is split into clusters and all members of some clusters are chosen to be included in the sample.

Let's learn how to perform cluster sampling on a pandas DataFrame in Python.

**Suppose a company that gives city tours wants to survey its customers. Out of ten tours, they give one day, they randomly select four tours and ask every customer to rate their experience on a scale of 1 to 10.**

The following code shows how to create a pandas DataFrame to work with:

```
import pandas as pd
import numpy as np

#make this example reproducible
np.random.seed(0)

#create DataFrame
df = pd.DataFrame({'tour': np.repeat(np.arange(1,11), 20),
                    'experience': np.random.normal(loc=7, scale=1,
size=200)})

#view the first five rows of DataFrame
df.head()
```

	tour	experience
0	1	8.764052
1	1	7.400157
2	1	7.978738
3	1	9.240893
4	1	8.867558

And the following code shows how to obtain a sample of customers by randomly selecting four tours and including every member in those tours in the sample:

```
#randomly choose 4 tour groups out of the 10
clusters = np.random.choice(np.arange(1,11), size=4,
replace=False)

#define sample as all members who belong to one of the 4 tour
groups
cluster_sample = df[df['tour'].isin(clusters)]

#view the first six rows of the sample
cluster_sample.head()
```

	tour	experience
40	3	5.951447
41	3	5.579982

```
42    3    5.293730
43    3    8.950775
44    3    6.490348
```

```
#find how many observations came from each tour group
cluster_sample['tour'].value_counts()
```

```
10    20
6     20
5     20
3     20
Name: tour, dtype: int64
```

From the output we can see that:

- 20 customers from tour group #10 were included in the sample.
- 20 customers from tour group #6 were included in the sample.
- 20 customers from tour group #5 were included in the sample.
- 20 customers from tour group #3 were included in the sample.

Thus, this sample is composed of 80 total customers that came from 4 different tour groups.