



# Lab 5

---

2021.04.03

# Context

---

- Basic of Data Analysis using Pandas
  - Data Load
  - Sorting
  - Indexing & Retrieving
  - Apply Functions
  - Grouping
  - Summary
  - Data Frame transformations

# Pandas 연습 문제

- 'telecom.churn.csv'를 읽어 자료들을 출력하고, 여러 함수들을 통해 결과 값을 출력하시오.

	State	Account length	Area code	International plan	Voice mail plan	Number vmail messages	Total day minutes	Total day calls	Total day charge	Total eve minutes	Total eve calls	Total eve charge	Total night minutes	Total night calls	Total night charge	Total intl minutes	Total intl calls
0	KS	128	415	No	Yes	25	265.1	110	45.07	197.4	99	16.78	244.7	91	11.01	10.0	3
1	OH	107	415	No	Yes	26	161.6	123	27.47	195.5	103	16.62	254.4	103	11.45	13.7	3
2	NJ	137	415	No	No	0	243.4	114	41.38	121.2	110	10.30	162.6	104	7.32	12.2	5
3	OH	84	408	Yes	No	0	299.4	71	50.90	61.9	88	5.26	196.9	89	8.86	6.6	7
4	OK	75	415	Yes	No	0	166.7	113	28.34	148.3	122	12.61	186.9	121	8.41	10.1	3

telecom.churn.csv

# Pandas 연습 문제

- 자료에 대한 정보들을 오른쪽과 같이 출력하시오.

- 행의 개수
- 열의 정보 (feature)
- feature들의 type
- etc

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3333 entries, 0 to 3332
Data columns (total 20 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   State                                3333 non-null   object
1   Account length                       3333 non-null   int64
2   Area code                           3333 non-null   int64
3   International plan                   3333 non-null   object
4   Voice mail plan                      3333 non-null   object
5   Number vmail messages               3333 non-null   int64
6   Total day minutes                   3333 non-null   float64
7   Total day calls                     3333 non-null   int64
8   Total day charge                    3333 non-null   float64
9   Total eve minutes                   3333 non-null   float64
10  Total eve calls                     3333 non-null   int64
11  Total eve charge                    3333 non-null   float64
12  Total night minutes                 3333 non-null   float64
13  Total night calls                   3333 non-null   int64
14  Total night charge                  3333 non-null   float64
15  Total intl minutes                  3333 non-null   float64
16  Total intl calls                    3333 non-null   int64
17  Total intl charge                   3333 non-null   float64
18  Customer service calls              3333 non-null   int64
19  Churn                              3333 non-null   bool
dtypes: bool(1), float64(8), int64(8), object(3)
memory usage: 498.1+ KB
None
```

# Pandas 연습 문제

- 각 feature들에 대한 통계치를 아래와 같이 출력하시오.

	Account length	Area code	Number vmail messages	Total day minutes	Total day calls	Total day charge	Total eve minutes	Total eve calls	Total eve charge	Total night minutes	Total night calls
count	3333.00	3333.00	3333.00	3333.00	3333.00	3333.00	3333.00	3333.00	3333.00	3333.00	3333.00
mean	101.06	437.18	8.10	179.78	100.44	30.56	200.98	100.11	17.08	200.87	100.11
std	39.82	42.37	13.69	54.47	20.07	9.26	50.71	19.92	4.31	50.57	19.57
min	1.00	408.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	23.20	33.00
25%	74.00	408.00	0.00	143.70	87.00	24.43	166.60	87.00	14.16	167.00	87.00
50%	101.00	415.00	0.00	179.40	101.00	30.50	201.40	100.00	17.12	201.20	100.00
75%	127.00	510.00	20.00	216.40	114.00	36.79	235.30	114.00	20.00	235.30	113.00
max	243.00	510.00	51.00	350.80	165.00	59.64	363.70	170.00	30.91	395.00	175.00

# Pandas 연습 문제

---

- 주어진 데이터에서 가입자 중 해지를 한 사람들의 비율을 계산하시오.  
(Churned User Proportion)
- 가입자 중 해지를 한 사람들의 다른 feature의 평균을 계산하시오.
- 가입자 중 해지를 한 사람들이 낮 시간 동안 휴대폰을 사용한 시간을 출력하시오.
- 해지를 하지 않고, 'International Plan'에 가입되어 있지 않은 사람들 중 국제 전화 통화 시간이 가장 긴 사람의 국제 전화 통화 시간을 출력하시오.

# Pandas 연습 문제

---

- 0~5번 가입자들의 'State' 열부터 'Area Code'까지의 정보를 출력하시오.
- 0~5번 가입자들의 0~2번 열 정보들을 출력하시오.
- 마지막 가입자의 모든 정보를 출력하시오.

# Pandas 연습 문제

---

- 각 feature들의 최댓값을 출력하시오.
- 거주지가 'W'로 시작하는 가입자들의 정보를 모두 출력하시오.
- 'International Plan'과 'Voice Mail Plan'의 값들을 Boolean 값으로 변경하시오.



# Pandas 연습 문제

---

- 가입자와 해지 한 사람들을 묶어서 낮, 저녁, 밤 시간대의 통화 시간에 대한 통계치를 출력하시오.
- (가입 사용자, 해지 사용자)와 ('International Plan' 가입 유, 무)에 따른 인원 수를 출력하시오.
- (가입 사용자, 해지 사용자)와 ('Voice Mail Plan' 가입 유, 무)에 따른 인원 비율을 출력하시오.

# Pandas 연습 문제

---

- 하루 낮, 저녁, 밤 시간대의 통화량의 평균을 지역 코드에 따라 출력하시오.
- 하루 전체 통화량을 feature로 설정하여 data frame에 추가하시오.
- 특정 열과 행을 삭제하고 출력하시오.