

Chapter. 01

데이터 분석이란?

| 데이터 분석 도구 소개

FAST CAMPUS
ONLINE

데이터 분석 TOOLS - Python I

강사. 김용담

I Application VS Programming



I 고객 유형 분석

Id	Product_Info_1	Product_Info_2	Product_Info_3	Product_Info_4	Product_Info_5	Product_Info_6	Product_Info_7	Ins_Age	Ht	Wt	BMI	Employment_Info_1	Employment_Info_2	Employment_Info_3
2	1	D3	10	0.076923077	2	1	1	0.641791045	0.581818182	0.148535565	0.323007976	0.028	12	1
5	1	A1	26	0.076923077	2	3	1	0.059701493	0.6	0.131799163	0.272287744	0	1	3
6	1	E1	26	0.076923077	2	3	1	0.029850746	0.745454545	0.288702929	0.428780429	0.03	9	1
7	1	D4	10	0.487179487	2	3	1	0.164179104	0.672727273	0.205020921	0.352437744	0.042	9	1
8	1	D2	26	0.230769231	2	3	1	0.417910448	0.654545455	0.234309623	0.424045645	0.027	9	1
10	1	D2	26	0.230769231	3	1	1	0.507462687	0.836363636	0.29916318	0.364886708	0.325	15	1
11	1	A8	10	0.166193846	2	3	1	0.373134328	0.581818182	0.173640167	0.376586717	0.11	1	3
14	1	D2	26	0.076923077	2	3	1	0.611940299	0.781818182	0.40376569	0.571611506	0.12	12	1
15	1	D3	26	0.230769231	2	3	1	0.52238806	0.618181818	0.184100418	0.36264306	0.165	9	1
16	1	E1	21	0.076923077	2	3	1	0.552238806	0.6	0.284518828	0.587795766	0.025	1	3
17	1	D3	26	0.128205128	2	3	1	0.537313433	0.690909091	0.309623431	0.521668453	0.05	9	1
18	1	D4	26	0.230769231	2	3	1	0.298507463	0.690909091	0.271966527	0.455050111	0.09	3	1
19	1	A2	26	0.102564103	2	3	1	0.567164179	0.618181818	0.163179916	0.320783966	0.075	9	1
20	2	D1	26	0.487179487	2	3	1	0.223880597	0.781818182	0.361924686	0.507514769	0.1	9	1
22	1	D4	26	0.487179487	2	3	1	0.328358209	0.636363636	0.142259414	0.264648223	0.16	3	1
23	1	A7	26	0	2	3	1	0.626865672	0.672727273	0.330543933	0.58127899	0.075	9	1
24	2	D4	26	0.487179487	2	3	1	0.208955224	0.745454545	0.246861925	0.360968696	0.1	14	1
25	1	D3	26	0.384615385	2	3	1	0.268656716	0.636363636	0.228033473	0.430949212	0.0378	9	1
26	1	D3	26	0.076923077	2	3	1	0.388059701	0.781818182	0.309623431	0.427393846	0.08	9	1
27	1	D4	26	0.487179487	2	3	1	0.223880597	0.6	0.138075314	0.285253828	0.055	9	1
29	1	D2	26	0.435897436	2	3	1	0.388059701	0.745454545	0.246861925	0.360968696	0.083	9	1

Source : <https://www.kaggle.com/c/prudential-life-insurance-assessment/data%9D>

I Application(MS Excel)

HomeInsertDrawPage LayoutFormulasDataReviewViewTell me

Paste

Calibri (Body)

12

B

I

U

<

- 손쉽게 데이터 분석을 시작할 수 있다.
- GUI(Graphic User Interface)를 이용하여 그래프, 통계 자료 등을 직관적으로 볼 수 있다.
- 실제 업무에서 사용하는 유저가 많다.
- 대용량 데이터에 취약하다. (100,000 rows 이상)
- 엑셀이 제공하지 않는 기능은 사용할 수 없다.
e.g. 머신러닝, 서버 구성, ...

I Programming with python(Pandas)

```
In [1]: 1 import pandas as pd
        2
        3 df = pd.read_csv("data/test.csv")
        4 df
```

```
Out[1]:
```

	Id	Product_Info_1	Product_Info_2	Product_Info_3	Product_Info_4	Product_Info_5	Product_Info_6	Product_Info_7	Ins_Age	Ht	...
0	1	1	D3	26	0.487179	2	3	1	0.611940	0.781818	...
1	3	1	A2	26	0.076923	2	3	1	0.626866	0.727273	...
2	4	1	D3	26	0.144667	2	3	1	0.582090	0.709091	...
3	9	1	A1	26	0.151709	2	1	1	0.522388	0.654545	...
4	12	1	A1	26	0.076923	2	3	1	0.298507	0.672727	...
...
19760	79093	1	D3	26	0.128205	2	3	1	0.537313	0.636364	...
19761	79099	1	D4	10	0.230769	2	3	1	0.134328	0.618182	...
19762	79102	1	A2	26	0.076923	2	3	1	0.850746	0.727273	...
19763	79125	1	D1	26	0.487179	2	3	1	0.522388	0.745455	...
19764	79129	1	D3	26	1.000000	2	1	1	0.343284	0.781818	...

19765 rows x 127 columns

- 대용량 데이터도 손쉽게 다룰 수 있다.
- 머신러닝, 서버 배포 등 원하는 기능들과 함께 구현할 수 있다.
- 학습의 난이도가 높다. (엑셀에 비해)
- GUI 환경이긴 하지만, 엑셀만큼 직관적이진 않다.

I 올바른 데이터 분석 도구의 선택



- 분석 목적을 확실히 하고, 어떤 기능들이 필요한지 리스트업한다.
- 해당 기능들을 가장 쉽게 사용할 수 있는 도구를 찾는다.
- IaaS(Infrastructure as a Service)들도 하나의 선택이 될 수 있다!

Source : <https://www.python.org>, <https://cloud.google.com/bigquery>

Chapter. 01

데이터 분석이란?

| 감사합니다.

FAST CAMPUS
ONLINE

데이터 분석 TOOLS - Python I

강사. 김용담