



(Python 라이브러리)



스마트인재개발원
Smart Human Resources Development

최 성 우 연구원



학습목표

- Numpy 라이브러리에 대해 알 수 있다.
- Numpy 라이브러리를 데이터 분석에 활용할 수 있다.

라이브러리 : 변수, 함수, 클래스, 모듈들을 모아 놓은 코드 집합

모듈 : 변수, 함수, 클래스들을 모아 놓은 코드 집합

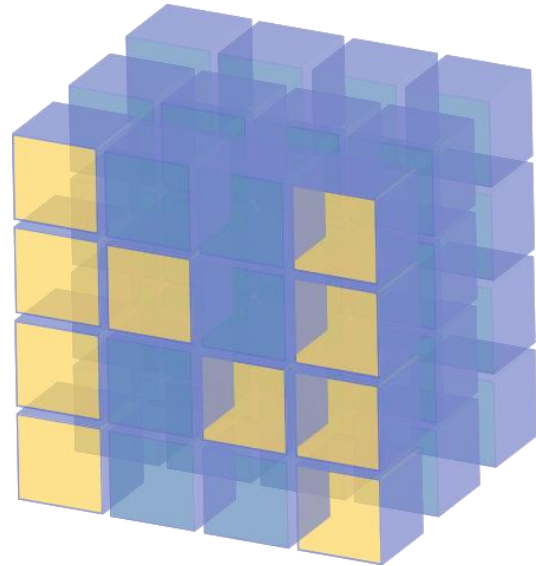
클래스 : 변수, 함수들을 모아 놓은 코드 집합

함수 : 하나의 특정 기능을 수행하기 위한 코드들의 집합



데이터 분석에 특화된 라이브러리

- Numpy : 고성능 수학, 과학계산을 위한 데이터 분석 라이브러리
- Pandas : 행과 열로 구성된 표 형식의 데이터 처리를 지원하는 라이브러리
- Matplotlib : 2D,3D 그래프로 시각화가 가능한 라이브러리



NumPy

Numerical Python 약자
Python의 수치 계산 및 해석용 라이브러리

Numpy의 주요 기능

- 빠르고 효율적인 데이터 산술연산을 제공하는 다차원배열 제공
(`ndarray` 클래스)

※ 배열(array) : 같은 자료형의 데이터들이 연속된 형태로 구성된 자료 구조로 다량의 데이터를 그룹으로 관리하며 손쉽게 계산하기 위해 사용

- 반복문 없이 전체 데이터 배열 연산이 가능한 표준 수학 함수 제공
(`mean()`, `sqrt()`, `std()` 등)

※ `sqrt` : square root(제곱근)

※ `std` : standard(표준편차)

numpy.ndarray 클래스

- 동일한 자료형을 가지는 값들이 배열 형태로 존재
- N 차원 형태로 구성이 가능(list와 유사)
ex) `[[1,2], [3,4], [5,6]]`
- 각 값들은 양의 정수로 색인(index)이 부여
- ndarray를 줄여서 array로 표현(nd는 n-dimension의 약자)

라이브러리 사용하기

- Numpy 라이브러리를 불러와서(import) 앞으로 np라는 이름으로 부름

```
1 import numpy as np
```




Pandas



Panel datas의 약자
행과 열로 이루어진 데이터들의 처리 지원용 라이브러리

요금제에 따른 한달제공데이터와 가격 부가세 선약적용가를 알려드리겠습니다 . 정식 명칭인 band 데이터 세이브는 통상명칭으로 299요금제 라고 하며 한달에 제공되는 데이터는 300M 입니다 가격을 부가세를 포함하여 32,890원이며 선약할인 (25%) 적용가는 24668원 입니다. 정식 명칭인 band 데이터 1.2G는 통상명칭으로 36요금제 라고 하며 한달에 제공되는 데이터는 1.2G 입니다 가격을 부가세를 포함하여 39,600원이며 선약할인 (25%) 적용가는 29,700원 입니다. 정식 명칭인 band 데이터 2.2G는 통상명칭으로 42요금제 라고 하며 한달에 제공되는 데이터는 2.2G 입니다 가격을 부가세를 포함하여 46,200원이며 선약할인 (25%) 적용가는 34,650원 입니다. 정식 명칭인 band 데이터 3.5G는 통상명칭으로 47요금제 라고 하며 한달에 제공되는 데이터는 3.5G입니다 가격을 부가세를 포함하여 51,700원이며 선약할인 (25%) 적용가는 38,775원 입니다. 정식 명칭인 band 데이터 6.5G는 통상명칭으로 51요금제 라고 하며 한달에 제공되는 데이터는 6.5G입니다 가격을 부가세를 포함하여 56,100원이며 선약할인 (25%) 적용가는 42,075원 입니다. 정식 명칭인 band 데이터 퍼펙트는 통상명칭으로 599요금제 라고 하며 한달에 제공되는 데이터는 11G + 매일 2GB 입니다 가격을 부가세를 포함하여 65,890원이며 선약할인 (25%) 적용가는 49,418원 입니다.

요금제 정리표

정식명칭	통상명칭	한달 제공 데이터	가격(부가세 포함)	선약(25%)적용가
band 데이터 세이브	299요금제	300MB	32,890원	24,668원
band 데이터 1.2G	36요금제	1.2GB	39,600원	29,700원
band 데이터 2.2G	42요금제	2.2GB	46,200원	34,650원
band 데이터 3.5G	47요금제	3.5GB	51,700원	38,775원
band 데이터 6.5G	51요금제	6.5GB	56,100원	42,075원
band 데이터 퍼펙트	599요금제	11GB+매일2GB	65,890원	49,418원
band 데이터 퍼펙트S	69요금제	16GB+매일2GB	75,900원	56,925원

Pandas 에서 제공하는 데이터구조

- Series : 1차원 배열 형태로 방향이 하나인 데이터 구조
(인덱스(index) + 값(value)으로 구성)
- DataFrame : 2차원 배열 형태의 행과 열 2방향으로 이루어진 데이터 구조
(x축, y축을 가지는 2차원 표 형태로 출력)

Pandas 에서 제공하는 데이터구조

DataFrame

Series			Series				Series				
Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked	
PassengerId											
1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S
6	0	3	Moran, Mr. James	male	NaN	0	0	330877	8.4583	NaN	Q
7	0	1	McCarthy, Mr. Timothy J	male	54.0	0	0	17463	51.8625	E46	S
8	0	3	Palsson, Master. Gosta Leonard	male	2.0	3	1	349909	21.0750	NaN	S
9	1	3	Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)	female	27.0	0	2	347742	11.1333	NaN	S
10	1	2	Nasser, Mrs. Nicholas (Adele Achem)	female	14.0	1	0	237736	30.0708	NaN	C

Pandas 에서 제공하는 데이터구조

Index labels

Column names

Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S
0	3	Moran, Mr. James	male	NaN	0	0	330877	8.4583	NaN	Q
0	1	McCarthy, Mr. Timothy J	male	54.0	0	0	17463	51.8625	E46	S
0	3	Palsson, Master. Gosta Leonard	male	2.0	3	1	349909	21.0750	NaN	S
1	3	Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)	female	27.0	0	2	347742	11.1333	NaN	S
1	2	Nasser, Mrs. Nicholas (Adele Achem)	female	14.0	1	0	237736	30.0708	NaN	C
1	3	Sandstrom, Miss. Marguerite Rut	female	4.0	1	1	PP 9549	16.7000	G6	S

Data



Mathdatas **plot library**

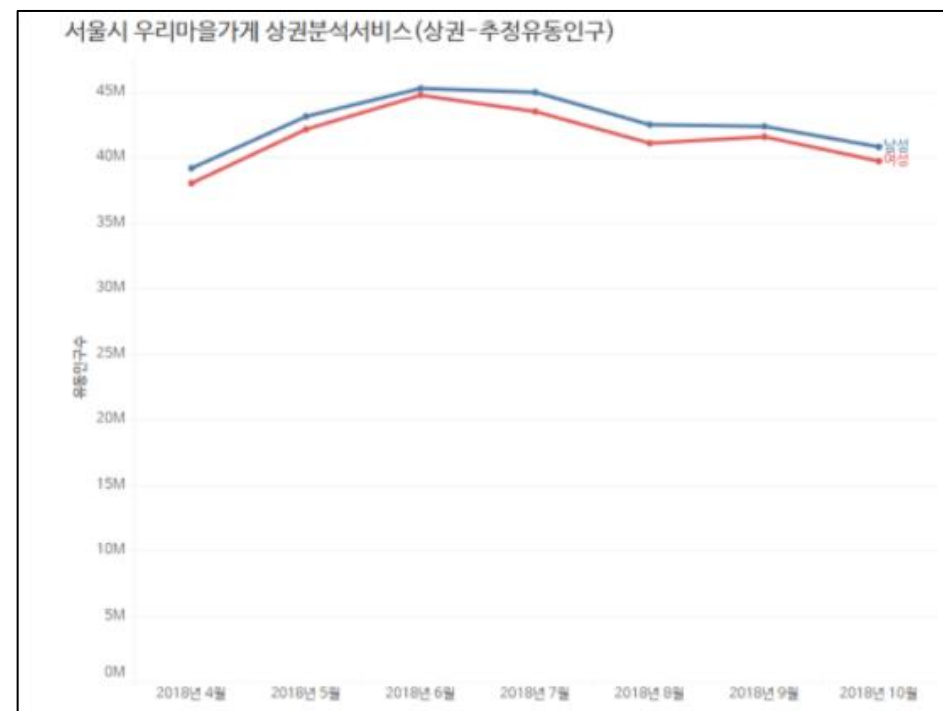
행과 열로 이루어진 데이터들의 처리 지원용 라이브러리

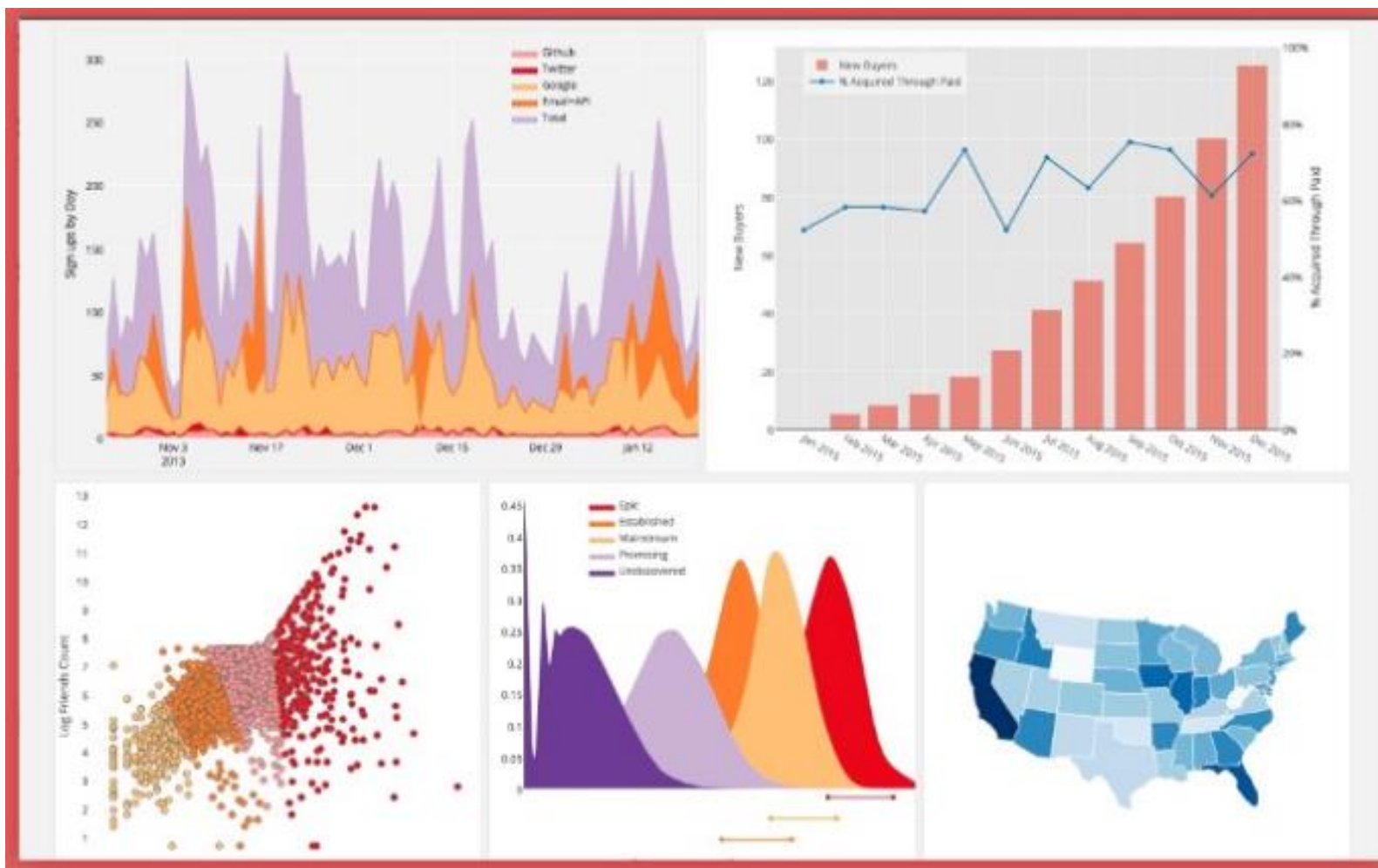
데이터 시각화가 필요한 이유

- 광범위하게 분산된 방대한 양의 자료를 **한눈에 볼 수 있도록** 도표나 차트 등으로 정리할 수 있음
- 시각화를 통해 **데이터의 특징을 쉽게 파악** 가능
- 분석 결과를 상대방에게 **효과적으로 전달** 가능

데이터 시각화가 필요한 이유

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	기준_년월	상권_코드	상권_코드	종_유동인_남성_유동	여성_유동	연령대_10	연령대_20	연령대_30	연령대_40	연령대_50	연령대_60	시간대_1	시간대_2	시간대_3	시간대_4	시간대_5	시간대_6	
2	201810	1744	회무영로1	78736	36725	42011	3479	23194	21866	13353	9456	7388	14050	12060	11338	12975	18614	9699
3	201810	1743	회경로3길	208177	105723	102453	16490	115942	27240	17922	16829	13755	27389	26504	35324	38543	54123	26294
4	201810	1742	회경로23길	120515	62362	58153	11978	70973	13749	9723	8428	5663	16687	13564	20086	22201	32313	15664
5	201810	1741	회경로23길	21067	11373	9693	269	4446	4489	3702	4482	3678	4414	4037	3231	2788	4004	2592
6	201810	1740	회경로14길	18377	10004	8374	488	6372	3099	2829	2907	2682	3903	3025	2531	2559	4000	2359
7	201810	1739	회경로117	8062	4490	3572	335	3180	1319	983	1105	1139	1184	1108	1156	1231	2150	1233
8	201810	1738	회경로10길	14772	8462	6309	1487	7638	1913	1479	1332	922	1727	1307	2023	2579	4598	2538
9	201810	1737	후암로57길	36888	20949	15939	733	10550	9529	6793	5586	3697	1800	6476	9424	7942	8592	2654
10	201810	1736	후암로4길	32179	16016	16163	2236	7500	7644	5470	4615	4714	4410	7462	4792	4685	6371	4460
11	201810	1735	후암로35길	14054	7353	6700	547	2998	3177	2754	2456	2121	3143	2077	2495	1939	2952	1447
12	201810	1734	후암로30길	14143	7753	6390	477	3202	3411	2862	2405	1786	2436	2122	2499	2494	3133	1458
13	201810	1733	후암로23길	11459	5787	5671	489	2340	2469	2310	1961	1889	2632	1842	1839	1548	2342	1257
14	201810	1732	후암로13길	13264	7141	6123	1106	2798	2938	2834	1922	1666	1647	2301	2601	2188	3146	1380
15	201810	1731	효창원로9	103291	49963	53328	7518	26029	19906	18653	18073	13112	17226	21545	16503	16456	19809	11752
16	201810	1730	효창원로9	45433	12194	33238	5117	21255	4464	4912	5364	4321	4500	7658	12153	11273	7399	2450
17	201810	1729	효창원로9	72294	28879	43416	4166	24262	13195	11349	11701	7621	15225	12812	14202	13455	11532	5067
18	201810	1728	효창원로8	25508	13481	12027	2153	4902	5830	5059	4009	3555	3976	5403	4520	4230	4792	2586
19	201810	1727	효창원로7	30280	13283	16997	2663	8800	5299	5114	4581	3823	4283	5727	5979	5473	5916	2900
20	201810	1726	효창원로6	18139	8503	9636	1083	3844	3716	3603	2929	2963	3775	3933	3142	2745	2904	1641
21	201810	1725	효창원로6	19930	10979	8950	1099	4554	4518	3112	3441	3207	2288	4766	4002	3545	3629	1700
22	201810	1724	효창원로3	41896	20979	20917	1830	7885	9506	7883	7780	7012	7700	7392	5978	6536	9074	5215
23	201810	1723	효창원로74길	25161	13463	11698	810	4886	6402	5826	4069	3168	3977	5046	4972	4460	4494	2211
24	201810	1722	효창원로68길	19906	11006	8900	174	3609	5839	4435	3167	2682	1612	3903	4821	3916	4153	1500
25	201810	1721	효창원로67길	51149	26749	24400	737	11648	13996	11012	7255	6502	6136	10622	10760	9693	9580	4358
26	201810	1720	효창원로40길	16914	9673	7241	3273	2786	2975	3143	2518	2218	1573	3337	4068	3159	3251	1526
27	201810	1719	효창원로34길	34159	17027	17132	2600	8567	6604	5305	5344	5739	3352	5740	7418	6835	7517	3297
28	201810	1718	효창원로29길	55361	27610	27751	4470	14961	11662	9477	7842	6949	7793	9443	10925	9972	11510	5719
29	201810	1717	회나무로4	96187	46335	49853	2196	43305	25298	12893	7849	4646	19159	8800	10698	13581	25777	18165
30	201810	1716	회나무로1	32032	16130	15901	50	9712	10360	5115	3811	2984	3430	5199	5958	6654	7355	3435







다음시간에는?