

데이터 시각화 & 데이터 수집



0. Review

1. 데이터시각화

2. 데이터 수집

- code: https://github.com/zzhining/python_data_basic

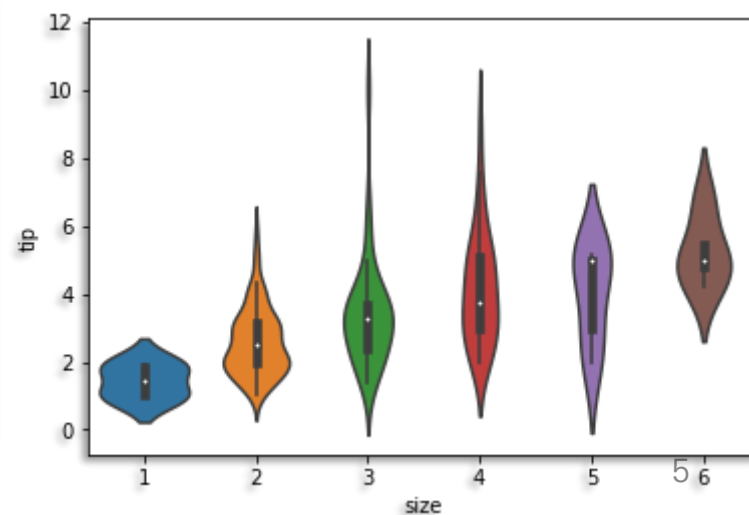
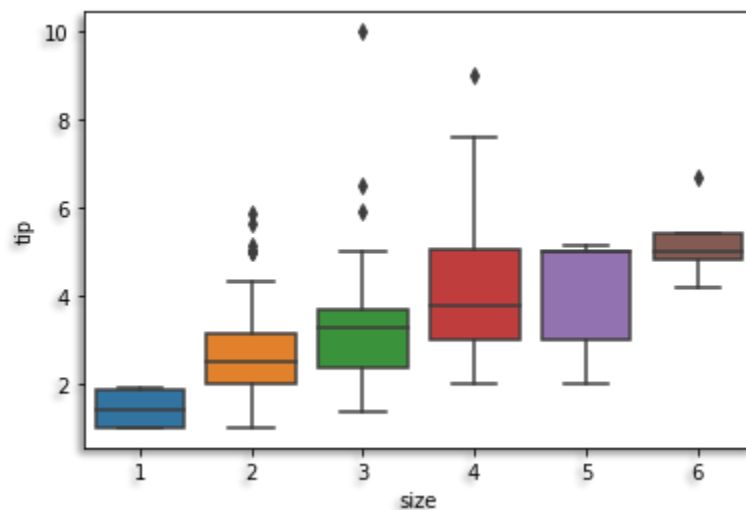
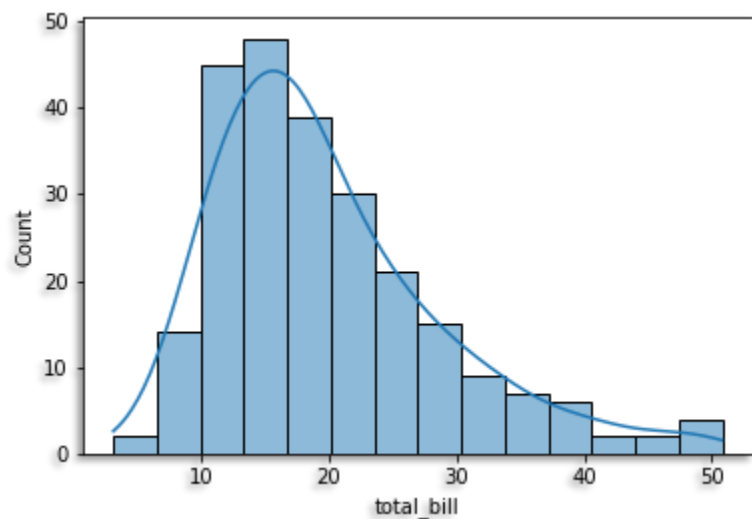
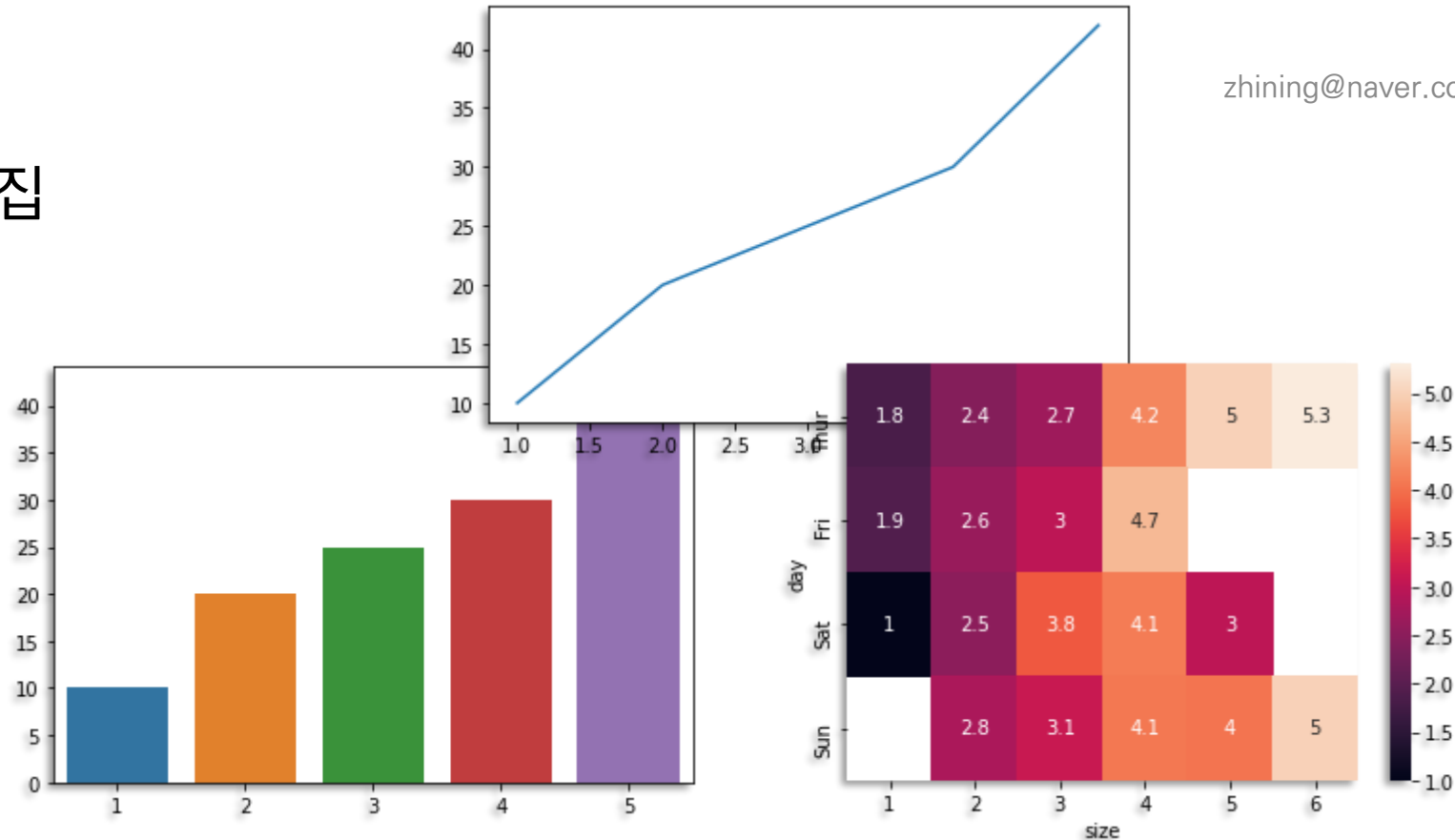
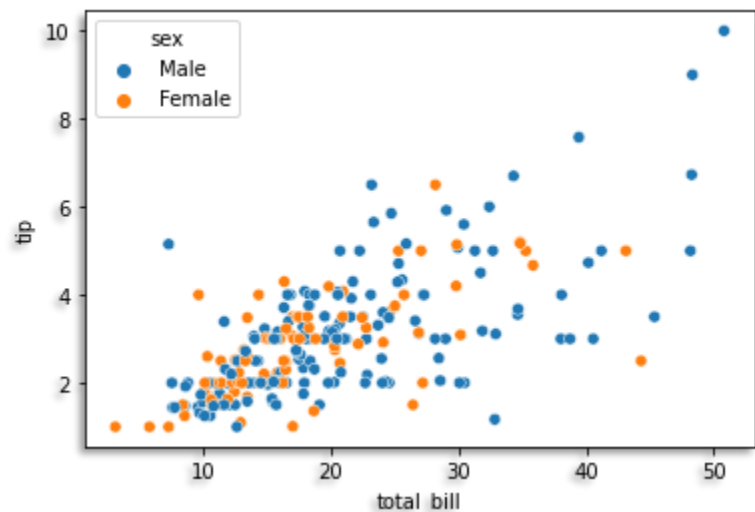
Today...

- 데이터 시각화 + 데이터 수집











today...

zhining@naver.com

■ 데이터 시각화 + 데이터 수집



■ 데이터 시각화 + 데이터 수집

				
MLB 핀치히터 LA다저스 볼캡 로알 블.. 940 MLB PINCH HITTER 91 LOSDOD D.. 33,000원	MLB 핀치히터 뉴욕 양키스 볼캡 네이.. 940 MLB PINCH HITTER 91 NEYYAN N.. 33,000원	무지 에센셜 반팔 티셔츠 3팩 멀티 AP 3PACK ESSENTIAL T 92 MUL 89,000원	MLB 팀 클래식 뉴욕 양키스 게임 볼.. TEAM CLASSIC 3930 NEYYAN GM 36,000원	MLB 핀치히터 디트로이트 타이거즈.. 940 MLB PINCH HITTER TEAM LOGO D.. 33,000원
				
MLB 어센틱 뉴욕 양키스 게임(홈/어.. ACPERF NEYYAN GM 49,000원	캡 이레이저 ACC CAP ERASER 3,500원	캡 리테이너 블랙 ACC CAP RETAINER83 BLACK 5,900원	캡 가드 ACC CAP GUARD 9,900원	MLB 팀 클래식 LA다저스 게임 볼캡 TEAM CLASSIC 3930 LOSDOD GM 36,000원

■ 데이터 시각화



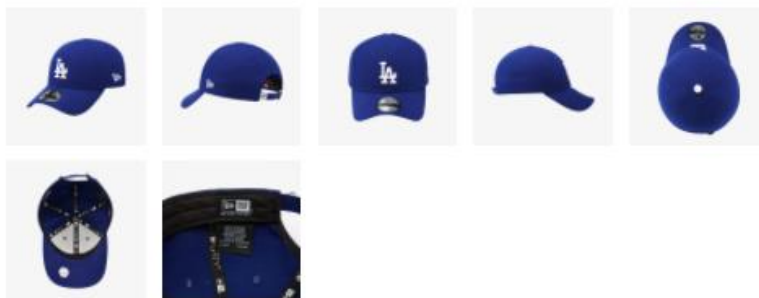
MLB 핀치히터 LA다저스 볼캡
940 MLB PINCH HITTER 91 LOSDOD D RYL
33,000원

재고확인



MLB 어센틱 뉴욕 양키스 게임
ACPERF NEYYAN GM
49,000원

재고확인



940 MLB PINCH HITTER 91 LOSDOD D RYL

MLB 핀치히터 LA다저스 볼캡 로얄 블루

33,000원 (5% 적립)

멤버십혜택

회원가입 즉시 사용 가능한 쿠폰을 확인하세요. ▼

제품코드 12024815

소재 100%폴리에스터

챙 길이 약 7cm

높이 약 11cm

특징 후면 사이즈 조절 가능

사이즈

옵션 선택 ▼

수량

1

+

-

로그인 or 비회원 구매

장바구니

N Pay 로 즉시 결제

▲ 네이버페이 추가 적립1%, 신용카드 무이자 7개월 등 결제수단별 혜택을 확인하세요. ▼

ACC CAP ERASER

3,500원

재고확인

ACC CAP RETAINER83 BLACK

5,900원

재고확인

ACC CAP GUARD

9,900원

재고확인

TEAM CLASSIC 3930 LOSDOD GM

36,000원

재고확인

■ 데이터 시각화

940 MLB PINCH HITTER 91 LOSDOD D RYL

MLB 핀치히터 LA다저스 볼캡 로얄 블루

	제품명	제품코드	수입원	출시일	소재	챙 길이	높이	특징	특이사항	깊이
0	MLB 핀치히터 LA다저스 볼캡 로얄 블루	12024815	뉴에라 캡 코리아	2018/12/18	100%폴리에스터	약 7cm	약 11cm	후면 사이즈 조절 가능	NONESEASON	NaN
1	MLB 핀치히터 뉴욕 양키스 볼캡 네이비	12024814	뉴에라 캡 코리아	2018/12/18	100%폴리에스터	약 7cm	약 11cm	NaN	NONESEASON	NaN
2	에센셜 무지 사이즈캡 블랙	12359641	뉴에라 캡 코리아	2018/09/12	100%울	약 7cm	NaN	NaN	18FW	약 11.5cm
3	MLB 핀치히터 디트로이트 타이거즈 볼캡 네이비	12033046	뉴에라 캡 코리아	2019/01/17	100%폴리에스터	약 7cm	약 11cm	NaN	NONESEASON	NaN
4	MLB 팀 클래식 뉴욕 양키스 게임 볼캡 네이비	10975804	뉴에라 캡 코리아	2016/06/14	95%폴리에스터, 5%스판덱스	약 7cm	NaN	NaN	NaN	약 12cm

MLB 어센틱 뉴욕 양키스 게임

ACPERF NEYYAN GM

49,000원

ACC CAP ERASER

3,500원

ACC CAP RETAINER83 BLACK

5,900원

ACC CAP GUARD

9,900원

TEAM CLASSIC 3930 LOSDOD GM

36,000원

▲ 네이버페이 추가 적립1%, 신용카드 무이자 7개월 등 결제수단별 혜택을 확인하세요. ▼

재고확인 ▼ 🔍

재고확인 ▼ 🔍

재고확인 ▼ 🔍

재고확인 ▼ 🔍

재고확인 ▼ 🔍

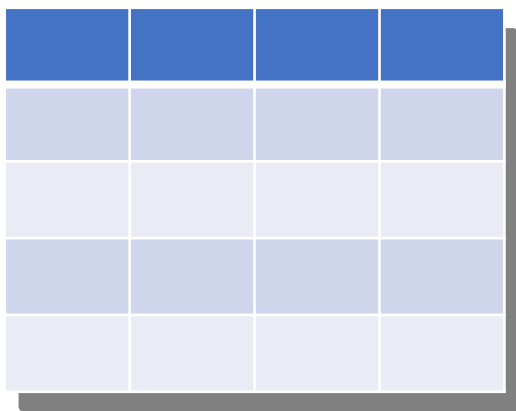
- 데이터 시각화 + 데이터 수집
 - 데이터 시각화
 - 그래프 그리는 방법
 - 데이터 수집
 - 웹 사이트의 데이터를 스크래핑(scraping)해서 데이터 프레임으로 만드는 방법

last time...

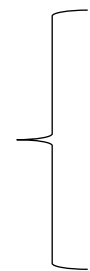
- 데이터 분석을 위한 파이썬 패키지

- 데이터 분석을 위한 파이썬 패키지
 - NumPy, Pandas

- 데이터 분석을 위한 파이썬 패키지
 - NumPy, Pandas



빠르게 연산!



- 수치연산
- 문자열 연산
- ...

주요 문법(입력 인자는 생략)	설명
np.array()	초기화할 값을 지정하여 배열 생성
np.zeros()	값을 0으로 초기화하여 배열 생성
np.ones()	값을 1로 초기화하여 배열 생성
np.arange()	수의 순차적인 증감을 이용하여 배열 생성
np.rand()	랜덤한 숫자로 배열 생성
reshape()	배열 변형
shape	배열의 형태 확인
dtype()	배열의 데이터 타입 확인
astype()	배열의 데이터 타입 변경
square()	제곱
sqrt()	제곱근(루트)
exp()	지수승

log()
add()
sum()
cumsum()
mean()
var()
std()
min()
max()
argmin()
argmax()
transpose()
np.save()

데이터프레임 생성

주요 문법	설명
pd.DataFrame(data, index, columns...)	리스트, 딕셔너리 등의 데이터를 데이터프레임 형태
pd.read_csv('csv파일명')	표 형태의 파일 읽어오기

데이터프레임 탐색

주요 문법(입력 인자는 생략)	설명
head()	처음 다섯 개의 로우 출력
tail()	마지막 다섯 개의 로우 출력
shape	데이터프레임의 구조 반환
describe()	데이터프레임의 주요 통계
info()	데이터프레임의 주요 정보
unique()	중복 제거된 유일 값 반환
value_counts()	유일 값의 개수

해키지

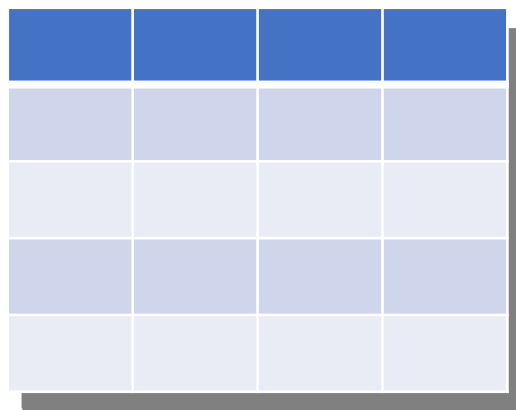
빠르게 연

데이터프레임 주요 연산

주요 문법(입력 인자는 생략)	설명
sort_values()	값 기준 정렬
sort_index()	인덱스 기준 정렬
iloc[n]	인덱스 번호로 선택(n번째 인덱스 로우 반환)
loc[name]	인덱스 이름으로 선택(인덱스 이름이 name인 로우 반환)
drop()	삭제
mean()	평균값 반환
max()	최대값 반환
min()	최소값 반환
sample()	전체 데이터 중 일부를 추출하여 반환
replace()	값 치환
apply()	함수 적용
get_dummies()	원핫인코딩 적용
isna()	결측치 여부 확인

주요 문법(입력 인자는 생략)	설명
fillna()	결측치 치환
dropna()	결측치 삭제
pd.merge()	두 개의 데이터프레임 조인
pd.concat()	두 개의 데이터프레임 이어 붙이기
to_csv()	저장

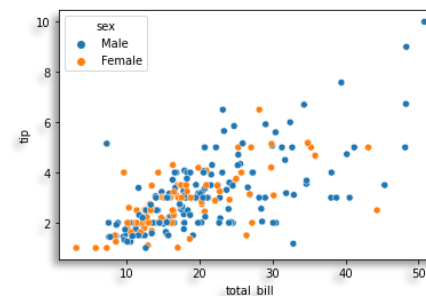
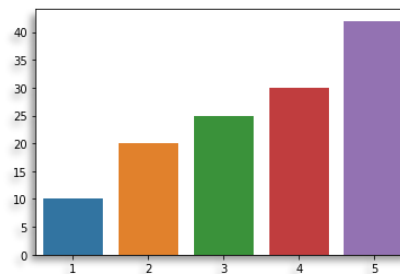
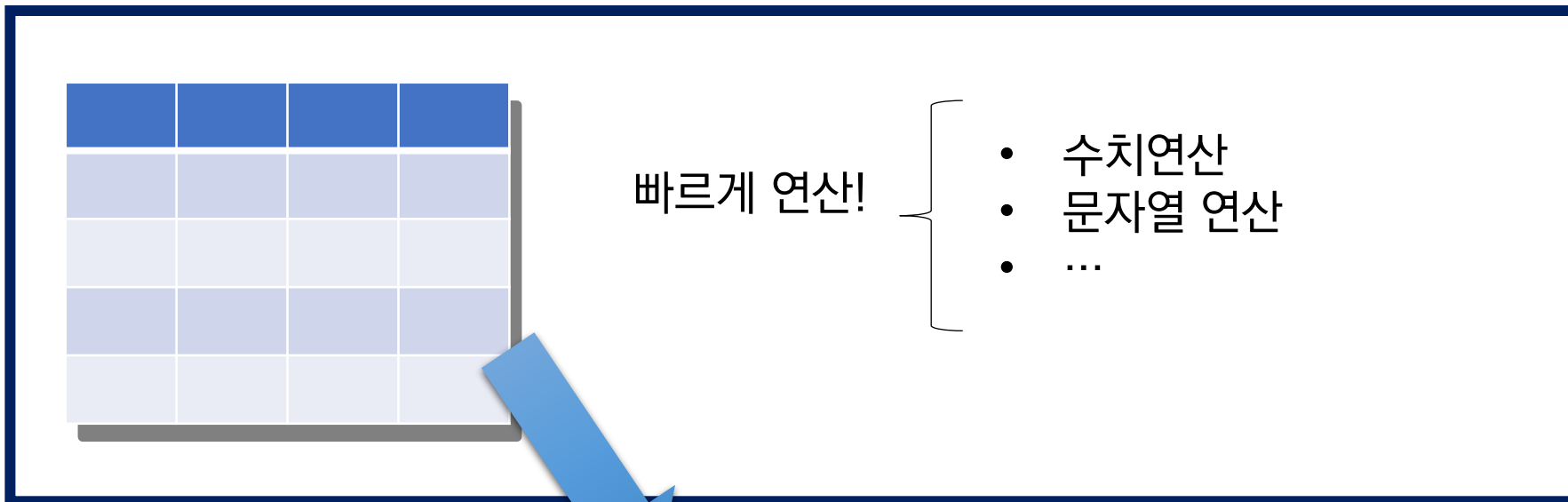
- 데이터 분석을 위한 파이썬 패키지
 - NumPy, Pandas



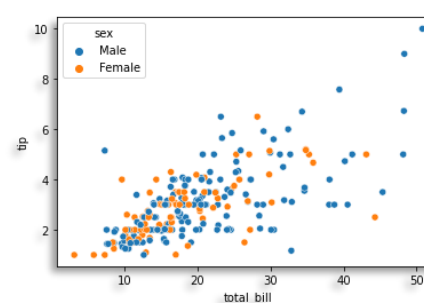
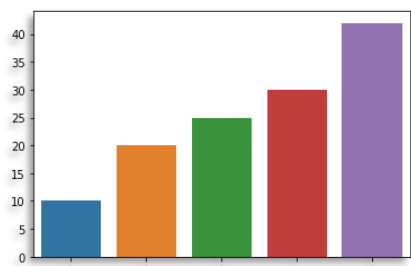
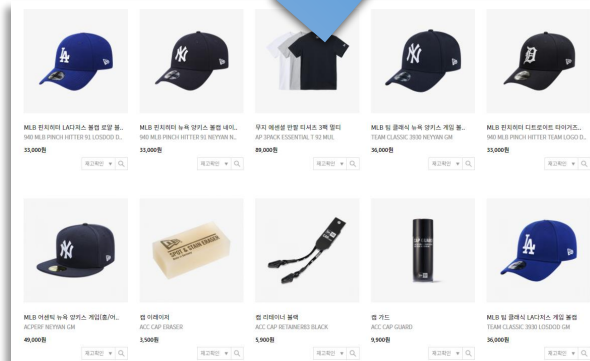
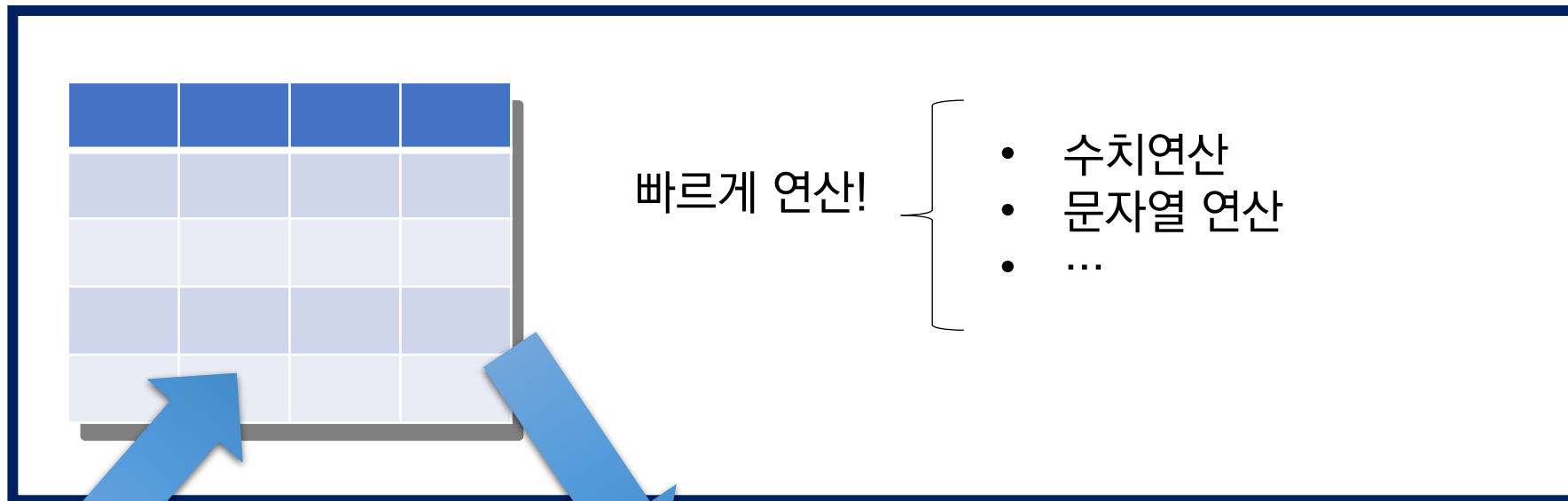
빠르게 연산!

- 수치연산
- 문자열 연산
- ...

- 데이터 분석을 위한 파이썬 패키지
 - NumPy, Pandas



- 데이터 분석을 위한 파이썬 패키지
 - NumPy, Pandas



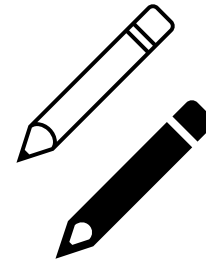
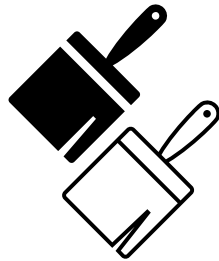
Ch4. 데이터 시각화

- seaborn & matplotlib

- seaborn & matplotlib



- seaborn & matplotlib



- seaborn & matplotlib



■ seaborn 사용법

- ① 데이터를 선택한다
- ② 그래프 종류를 선택한다
- ③ 그래프를 그린다

■ seaborn 사용법

- ① 데이터를 선택한다
- ② 그래프 종류를 선택한다
- ③ 그래프를 그린다

	total_bill	tip	sex	smoker	day	time	size
0	16.99	1.01	Female	No	Sun	Dinner	2
1	10.34	1.66	Male	No	Sun	Dinner	3
2	21.01	3.50	Male	No	Sun	Dinner	3
3	23.68	3.31	Male	No	Sun	Dinner	2
4	24.59	3.61	Female	No	Sun	Dinner	4

■ seaborn 사용법

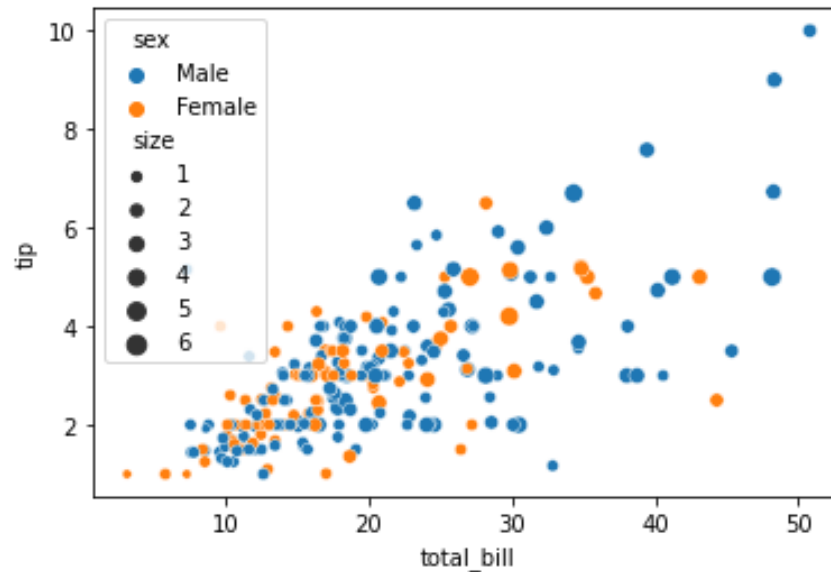
- import seaborn as sns

```
sns.scatterplot(x=df['total_bill'], y=df['tip'], hue=df['sex'], size=df['size'])
```

■ seaborn 사용법

- import seaborn as sns

```
sns.scatterplot(x=df['total_bill'], y=df['tip'], hue=df['sex'], size=df['size'])
```



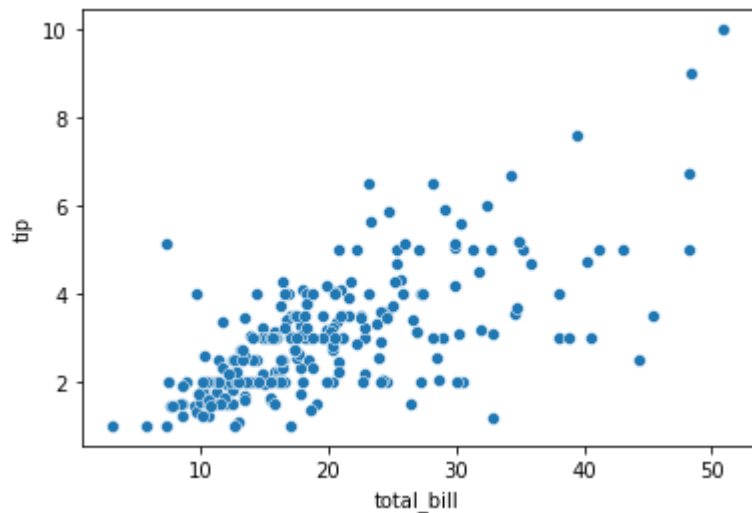
- total_bill과 tip의 상관관계?

	total_bill	tip	sex	smoker	day	time	size
0	16.99	1.01	Female	No	Sun	Dinner	2
1	10.34	1.66	Male	No	Sun	Dinner	3
2	21.01	3.50	Male	No	Sun	Dinner	3
3	23.68	3.31	Male	No	Sun	Dinner	2
4	24.59	3.61	Female	No	Sun	Dinner	4

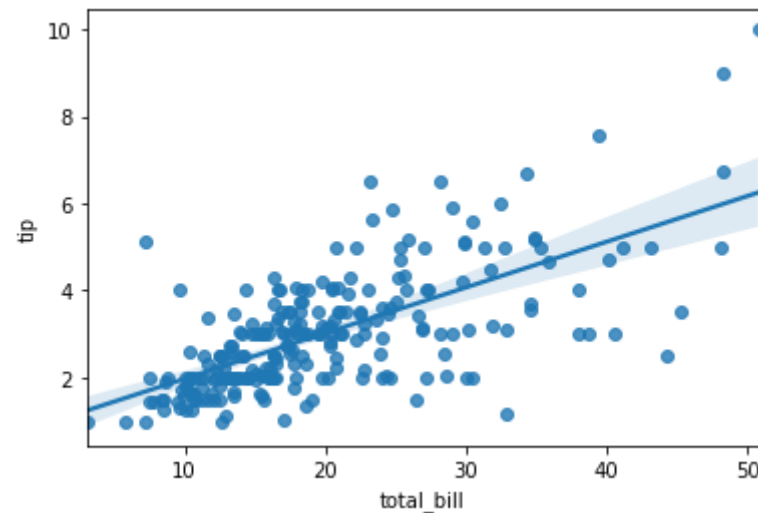
■ total_bill과 tip의 상관관계?

	total_bill	tip	sex	smoker	day	time	size
0	16.99	1.01	Female	No	Sun	Dinner	2
1	10.34	1.66	Male	No	Sun	Dinner	3
2	21.01	3.50	Male	No	Sun	Dinner	3
3	23.68	3.31	Male	No	Sun	Dinner	2
4	24.59	3.61	Female	No	Sun	Dinner	4

```
sns.scatterplot(x=df['total_bill'], y=df['tip'])
```



```
sns.regplot(x=df['total_bill'], y=df['tip'])
```



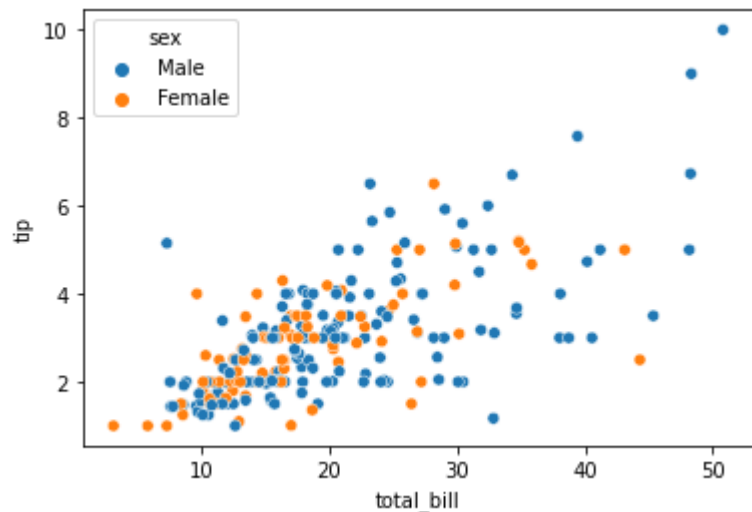
- total_bill과 tip의 상관관계?
& sex에 따른 분포!

	total_bill	tip	sex	smoker	day	time	size
0	16.99	1.01	Female	No	Sun	Dinner	2
1	10.34	1.66	Male	No	Sun	Dinner	3
2	21.01	3.50	Male	No	Sun	Dinner	3
3	23.68	3.31	Male	No	Sun	Dinner	2
4	24.59	3.61	Female	No	Sun	Dinner	4

- total_bill과 tip의 상관관계?
& sex에 따른 분포!

	total_bill	tip	sex	smoker	day	time	size
0	16.99	1.01	Female	No	Sun	Dinner	2
1	10.34	1.66	Male	No	Sun	Dinner	3
2	21.01	3.50	Male	No	Sun	Dinner	3
3	23.68	3.31	Male	No	Sun	Dinner	2
4	24.59	3.61	Female	No	Sun	Dinner	4

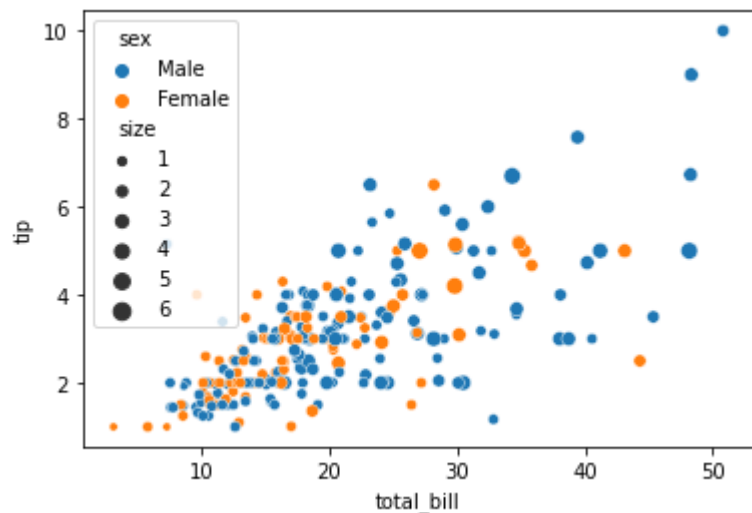
```
sns.scatterplot(x=df['total_bill'], y=df['tip'], hue=df['sex'])
```



- total_bill과 tip의 상관관계?
& sex에 따른 분포
& size에 따른 분포!

	total_bill	tip	sex	smoker	day	time	size
0	16.99	1.01	Female	No	Sun	Dinner	2
1	10.34	1.66	Male	No	Sun	Dinner	3
2	21.01	3.50	Male	No	Sun	Dinner	3
3	23.68	3.31	Male	No	Sun	Dinner	2
4	24.59	3.61	Female	No	Sun	Dinner	4

```
sns.scatterplot(x=df['total_bill'], y=df['tip'], hue=df['sex'], size=df['size'])
```



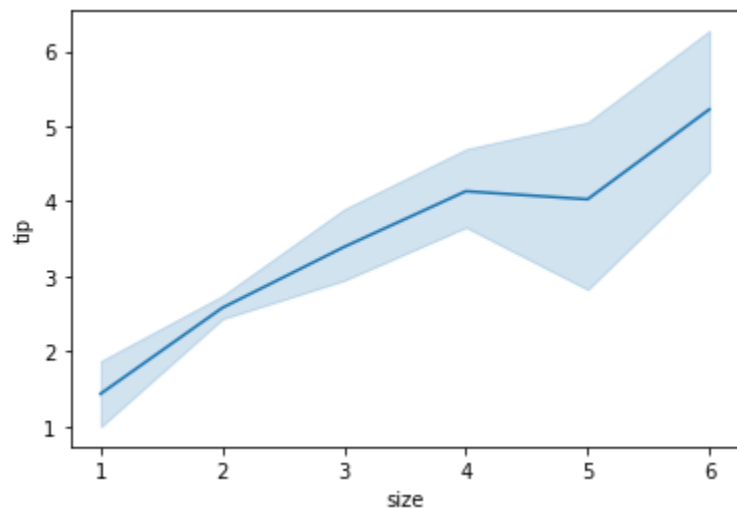
- size가 클수록 tip도 많나?

	total_bill	tip	sex	smoker	day	time	size
0	16.99	1.01	Female	No	Sun	Dinner	2
1	10.34	1.66	Male	No	Sun	Dinner	3
2	21.01	3.50	Male	No	Sun	Dinner	3
3	23.68	3.31	Male	No	Sun	Dinner	2
4	24.59	3.61	Female	No	Sun	Dinner	4

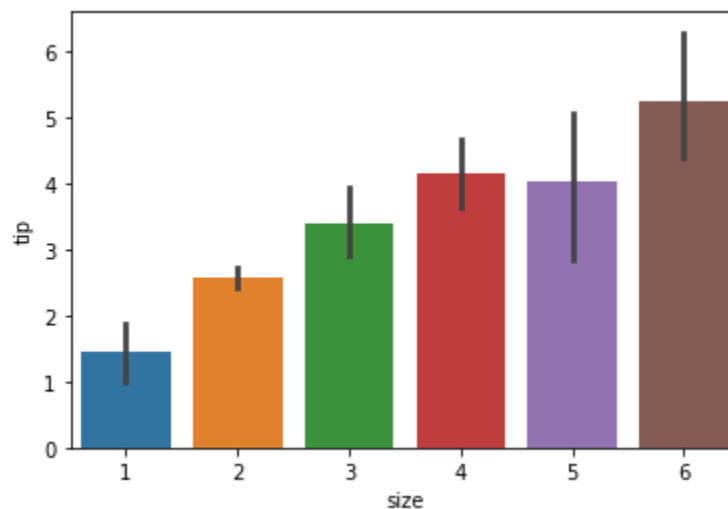
■ size가 클수록 tip도 많나?

	total_bill	tip	sex	smoker	day	time	size
0	16.99	1.01	Female	No	Sun	Dinner	2
1	10.34	1.66	Male	No	Sun	Dinner	3
2	21.01	3.50	Male	No	Sun	Dinner	3
3	23.68	3.31	Male	No	Sun	Dinner	2
4	24.59	3.61	Female	No	Sun	Dinner	4

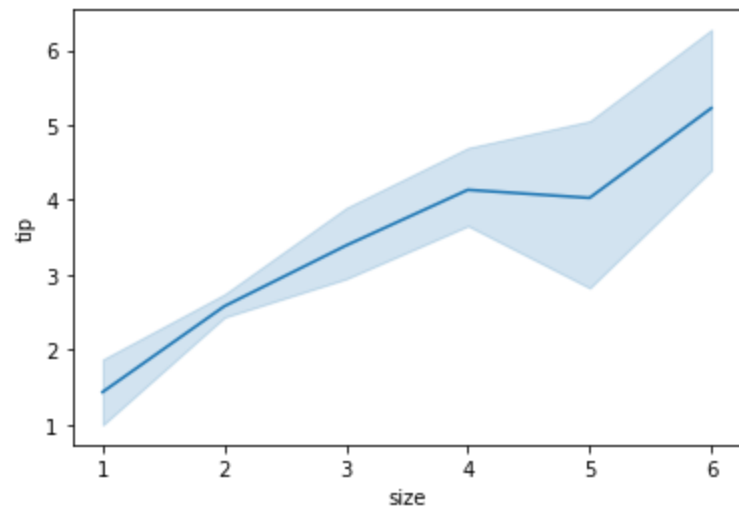
```
sns.lineplot(x=df['size'], y=df['tip'])
```



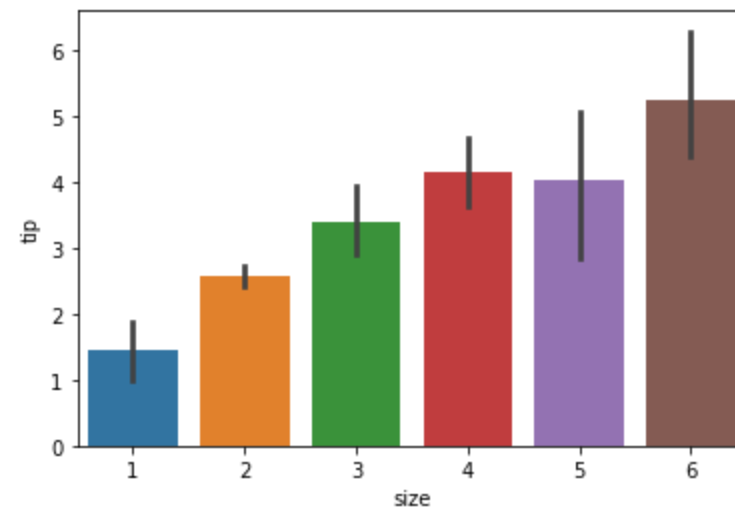
```
sns.barplot(x=df['size'], y=df['tip'])
```



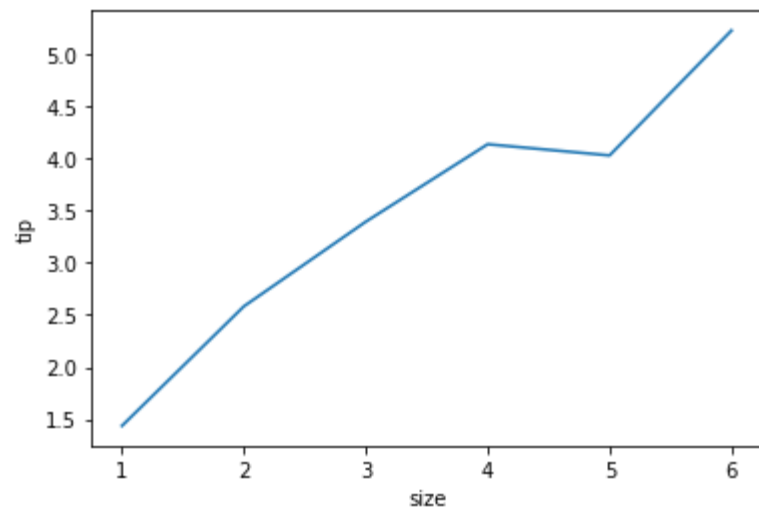
```
sns.lineplot(x=df['size'], y=df['tip'])
```



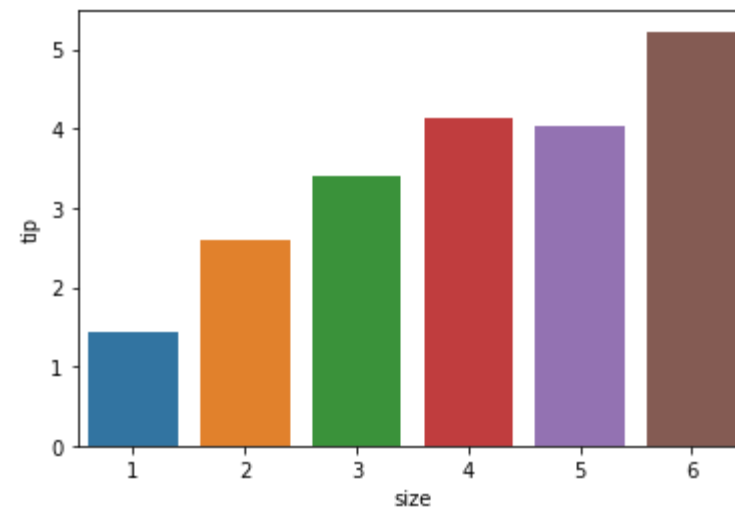
```
sns.barplot(x=df['size'], y=df['tip'])
```



```
sns.lineplot(x=df['size'], y=df['tip'], ci=None)
```



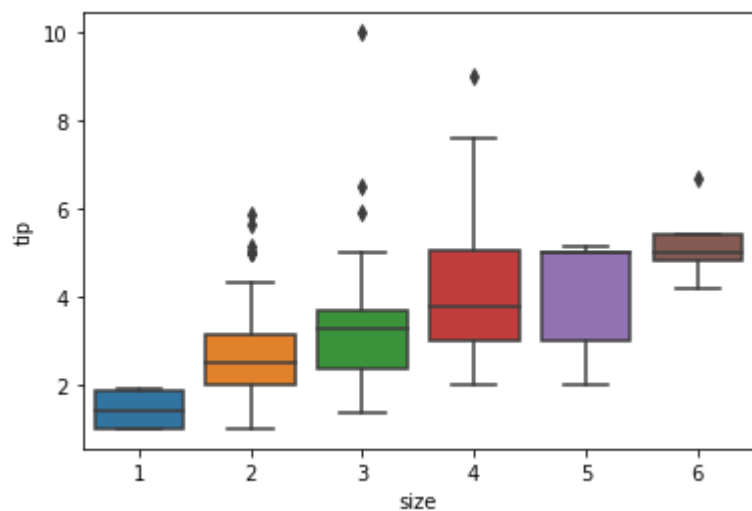
```
sns.barplot(x=df['size'], y=df['tip'], ci=None)
```



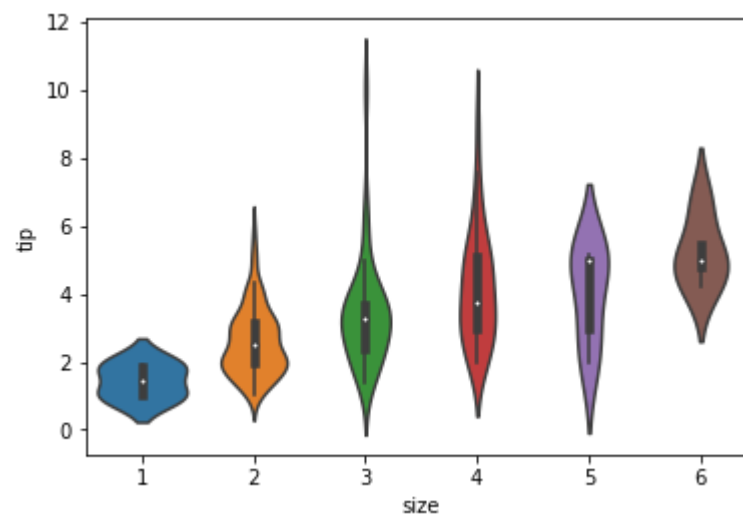
■ size가 클수록 tip도 많나?

	total_bill	tip	sex	smoker	day	time	size
0	16.99	1.01	Female	No	Sun	Dinner	2
1	10.34	1.66	Male	No	Sun	Dinner	3
2	21.01	3.50	Male	No	Sun	Dinner	3
3	23.68	3.31	Male	No	Sun	Dinner	2
4	24.59	3.61	Female	No	Sun	Dinner	4

```
sns.boxplot(x=df['size'], y=df['tip'])
```

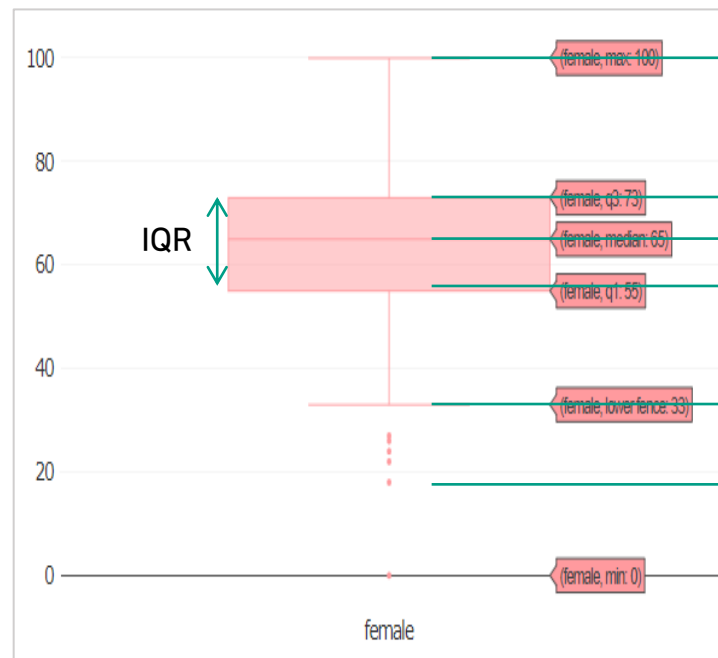
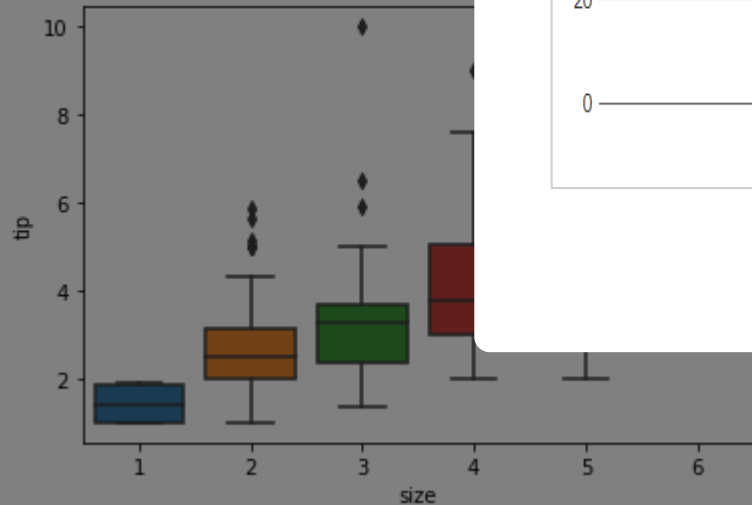


```
sns.violinplot(x=df['size'], y=df['tip'])
```



- size가 클수록 tip도 많나?

`sns.boxplot(x=df['size'], y=df['tip'])`



종양값 + 1.5 x IQR보다 큰 데이터 중 가장 큰 값

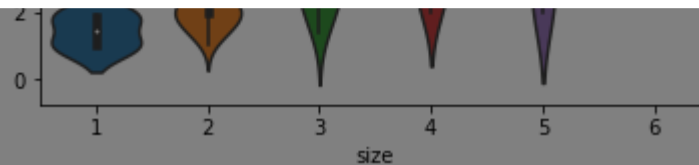
3사분위수

종양값

1사분위수

종양값 - 1.5 x IQR보다 큰 데이터 중 가장 작은 값

이상치(outlier)



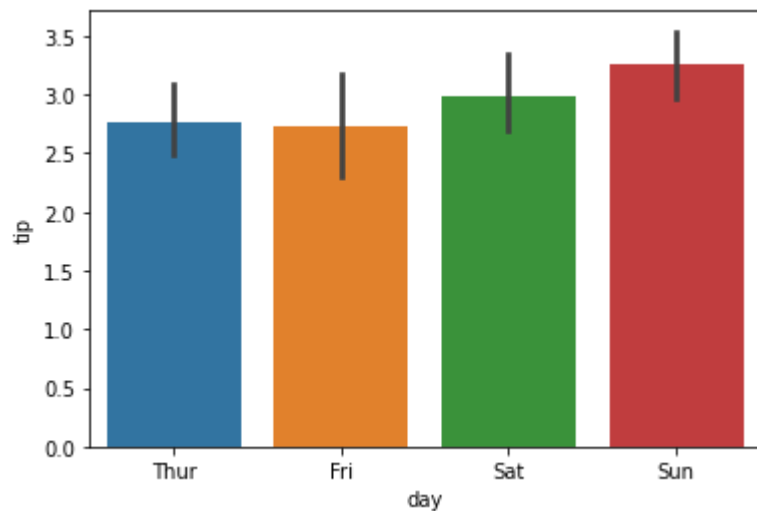
■ 요일별 팁의 분포?

	total_bill	tip	sex	smoker	day	time	size
0	16.99	1.01	Female	No	Sun	Dinner	2
1	10.34	1.66	Male	No	Sun	Dinner	3
2	21.01	3.50	Male	No	Sun	Dinner	3
3	23.68	3.31	Male	No	Sun	Dinner	2
4	24.59	3.61	Female	No	Sun	Dinner	4

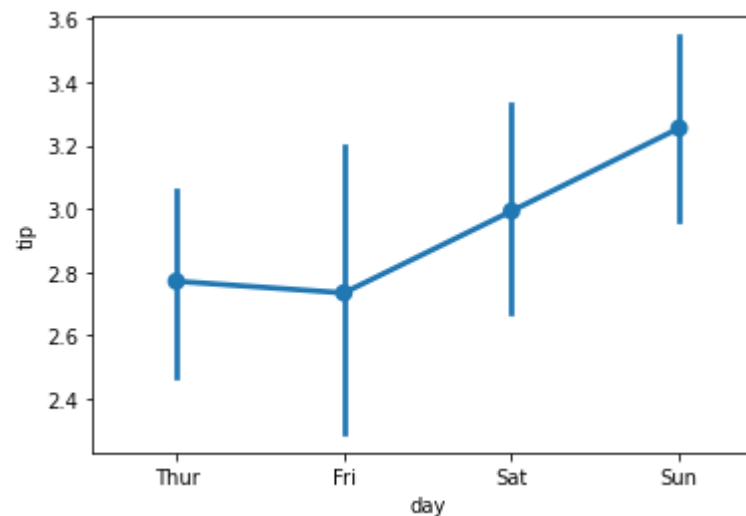
■ 요일별 팁의 분포?

	total_bill	tip	sex	smoker	day	time	size
0	16.99	1.01	Female	No	Sun	Dinner	2
1	10.34	1.66	Male	No	Sun	Dinner	3
2	21.01	3.50	Male	No	Sun	Dinner	3
3	23.68	3.31	Male	No	Sun	Dinner	2
4	24.59	3.61	Female	No	Sun	Dinner	4

```
sns.barplot(x=df['day'], y=df['tip'])
```



```
sns.barplot(x=df['day'], y=df['tip'])
```



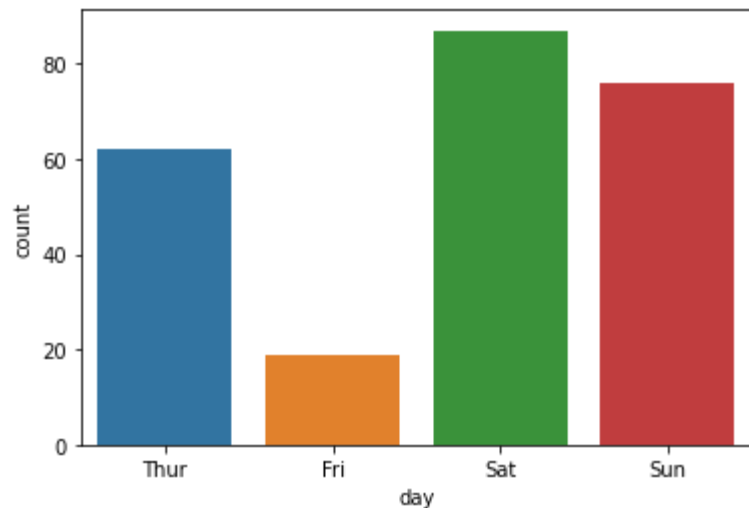
■ 요일별 방문자수

	total_bill	tip	sex	smoker	day	time	size
0	16.99	1.01	Female	No	Sun	Dinner	2
1	10.34	1.66	Male	No	Sun	Dinner	3
2	21.01	3.50	Male	No	Sun	Dinner	3
3	23.68	3.31	Male	No	Sun	Dinner	2
4	24.59	3.61	Female	No	Sun	Dinner	4

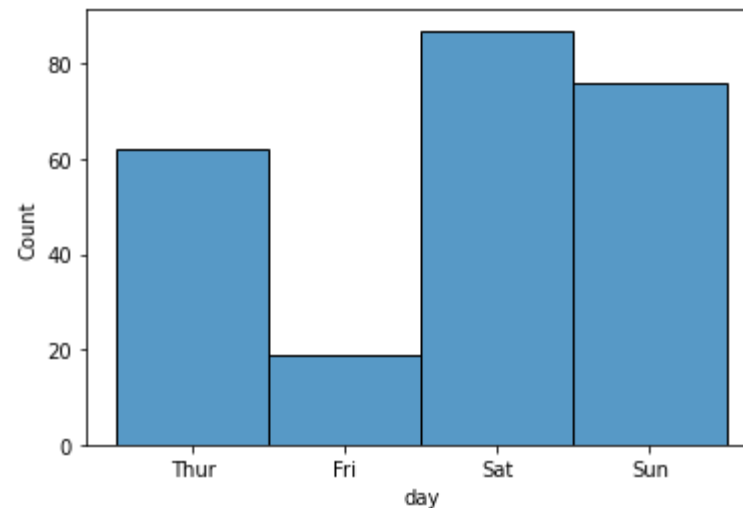
■ 요일별 방문자수

	total_bill	tip	sex	smoker	day	time	size
0	16.99	1.01	Female	No	Sun	Dinner	2
1	10.34	1.66	Male	No	Sun	Dinner	3
2	21.01	3.50	Male	No	Sun	Dinner	3
3	23.68	3.31	Male	No	Sun	Dinner	2
4	24.59	3.61	Female	No	Sun	Dinner	4

```
sns.countplot(x=df['day'])
```



```
sns.histplot(data=df['day'])
```



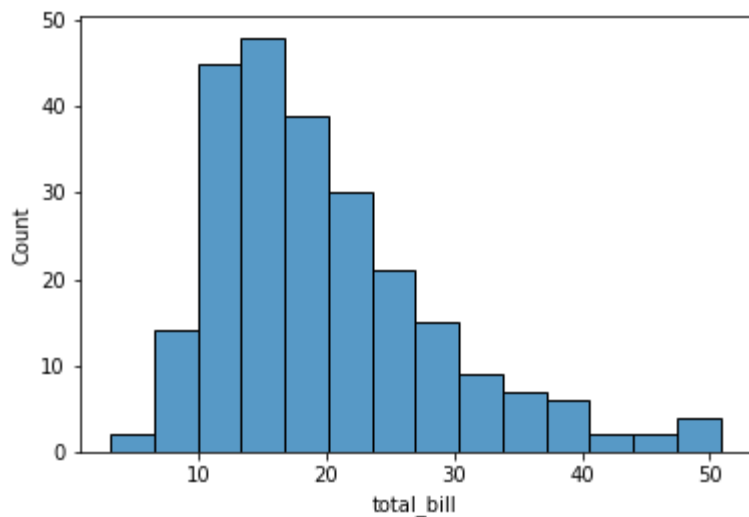
- total_bill의 분포

	total_bill	tip	sex	smoker	day	time	size
0	16.99	1.01	Female	No	Sun	Dinner	2
1	10.34	1.66	Male	No	Sun	Dinner	3
2	21.01	3.50	Male	No	Sun	Dinner	3
3	23.68	3.31	Male	No	Sun	Dinner	2
4	24.59	3.61	Female	No	Sun	Dinner	4

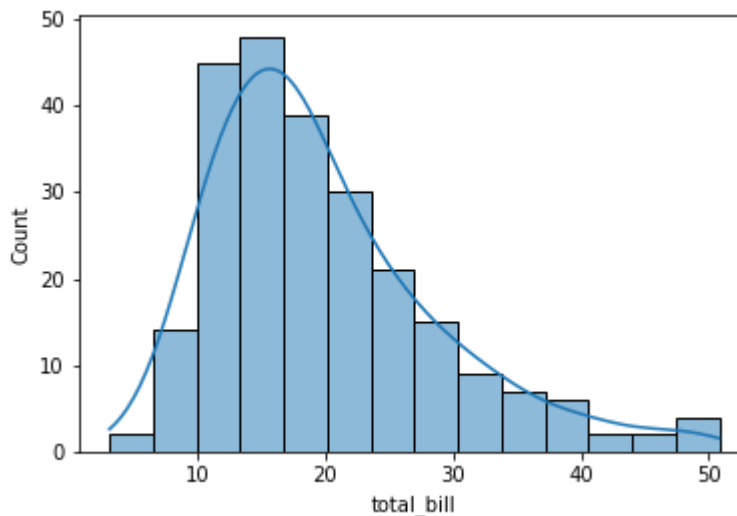
total_bill의 분포

	total_bill	tip	sex	smoker	day	time	size
0	16.99	1.01	Female	No	Sun	Dinner	2
1	10.34	1.66	Male	No	Sun	Dinner	3
2	21.01	3.50	Male	No	Sun	Dinner	3
3	23.68	3.31	Male	No	Sun	Dinner	2
4	24.59	3.61	Female	No	Sun	Dinner	4

```
sns.histplot(data=df['total_bill'])
```



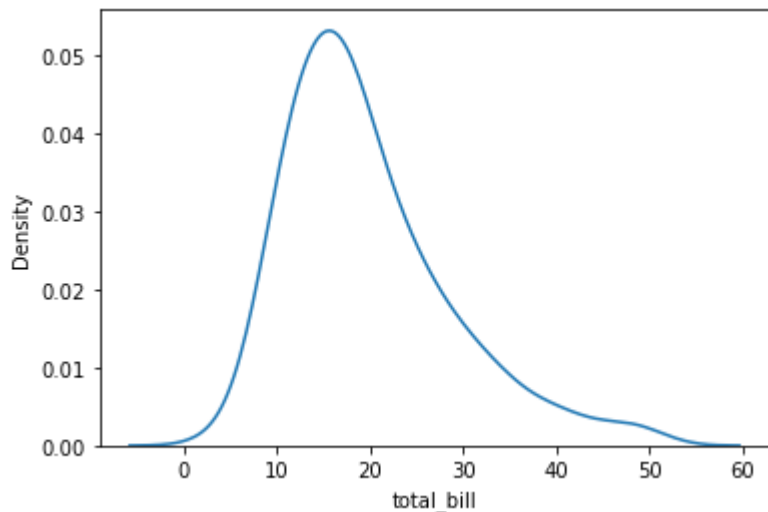
```
sns.histplot(data=df['total_bill'], kde = True)
```



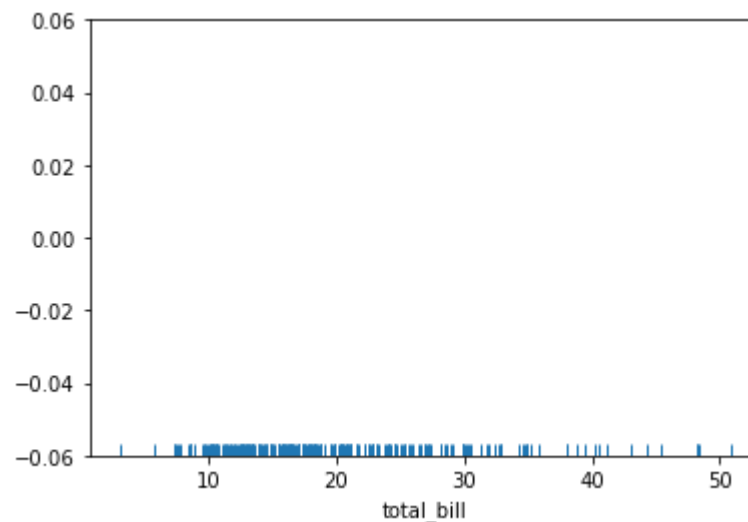
■ total_bill의 분포

	total_bill	tip	sex	smoker	day	time	size
0	16.99	1.01	Female	No	Sun	Dinner	2
1	10.34	1.66	Male	No	Sun	Dinner	3
2	21.01	3.50	Male	No	Sun	Dinner	3
3	23.68	3.31	Male	No	Sun	Dinner	2
4	24.59	3.61	Female	No	Sun	Dinner	4

```
sns.kdeplot(data=df['total_bill'])
```



```
sns.rugplot(data=df['total_bill'])
```



- 요일(day)별 size에 따른 tip 평균

	total_bill	tip	sex	smoker	day	time	size
0	16.99	1.01	Female	No	Sun	Dinner	2
1	10.34	1.66	Male	No	Sun	Dinner	3
2	21.01	3.50	Male	No	Sun	Dinner	3
3	23.68	3.31	Male	No	Sun	Dinner	2
4	24.59	3.61	Female	No	Sun	Dinner	4

■ 요일(day)별 size에 따른 tip 평균

	total_bill	tip	sex	smoker	day	time	size
0	16.99	1.01	Female	No	Sun	Dinner	2
1	10.34	1.66	Male	No	Sun	Dinner	3
2	21.01	3.50	Male	No	Sun	Dinner	3
3	16.99	3.31	Male	No	Sun	Dinner	2
4	24.59	3.61	Female	No	Sun	Dinner	4

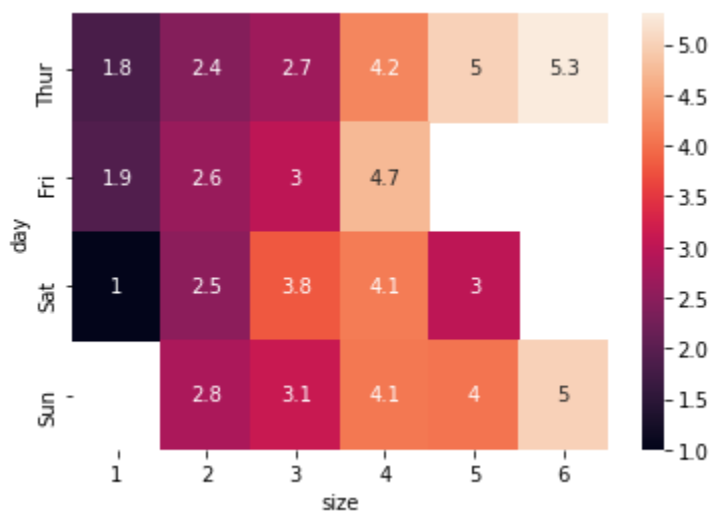
```
pivot_df = df.pivot_table("tip", "day", "size")
pivot_df
```

size	1	2	3	4	5	6
day						
Thur	1.83	2.442500	2.692500	4.218000	5.000000	5.3
Fri	1.92	2.644375	3.000000	4.730000	NaN	NaN
Sat	1.00	2.517547	3.797778	4.123846	3.000000	NaN
Sun	NaN	2.816923	3.120667	4.087778	4.046667	5.0

- 요일(day)별 size에 따른 tip 평균

	total_bill	tip	sex	smoker	day	time	size
0	size	1	2	3	4	5	6
1	day						
2	Thur	1.83	2.442500	2.692500	4.218000	5.000000	5.3
3	Fri	1.92	2.644375	3.000000	4.730000	NaN	NaN
4	Sat	1.00	2.517547	3.797778	4.123846	3.000000	NaN
	Sun	NaN	2.816923	3.120667	4.087778	4.046667	5.0

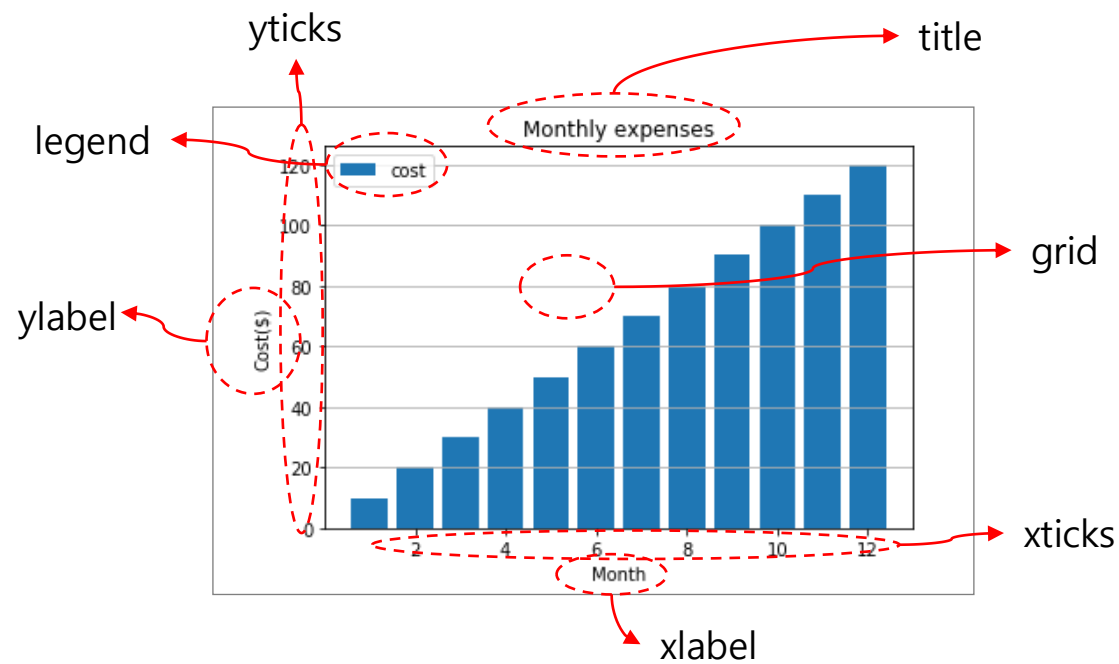
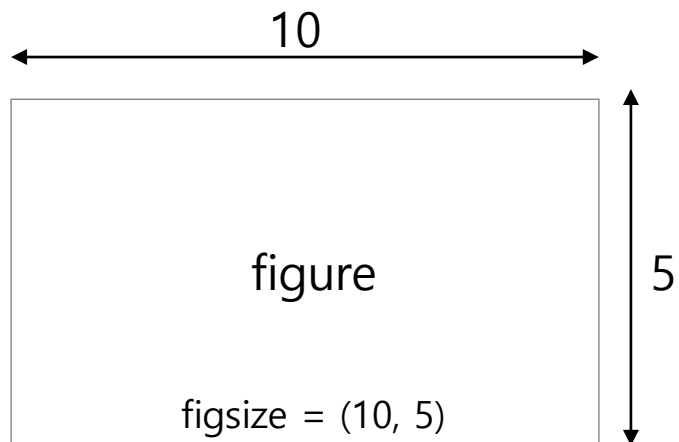
```
sns.heatmap(pivot_df, annot=True)
```



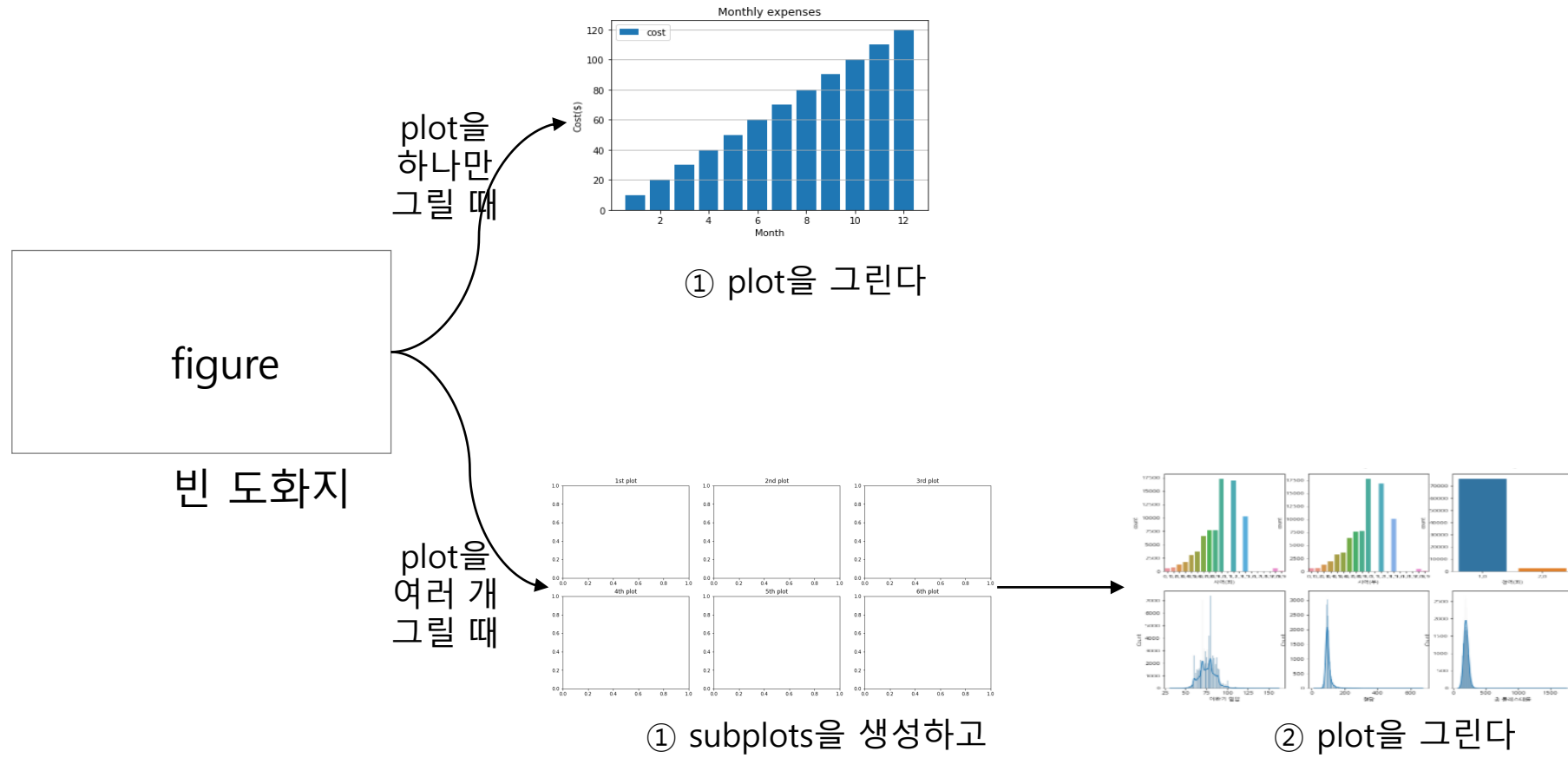
- seaborn & matplotlib



- matplotlib
 - 도화지 사이즈

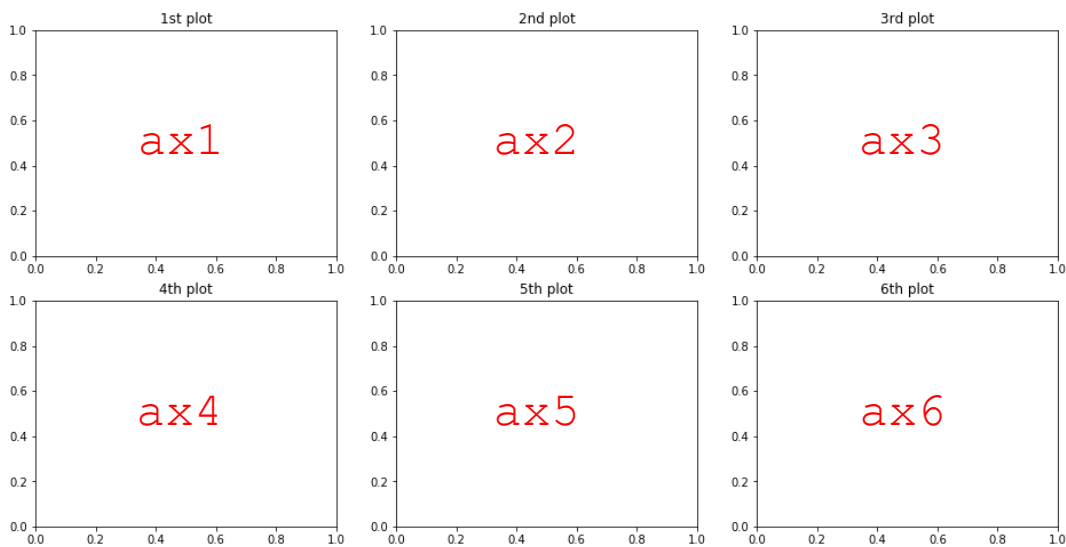


- matplotlib
 - 여러 개의 그래프



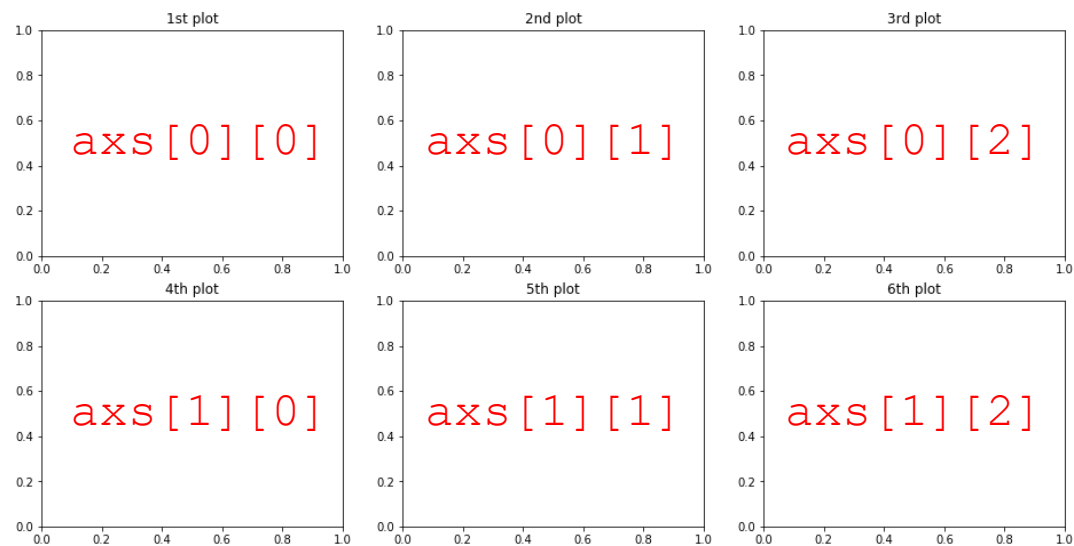
- matplotlib
 - 여러 개의 그래프

```
fig, ((ax1, ax2, ax3), (ax4, ax5, ax6)) = plt.subplots(nrows=2, ncols=3)
```



- matplotlib
 - 여러 개의 그래프

```
fig, axs = plt.subplots(nrows=2, ncols=3)
```



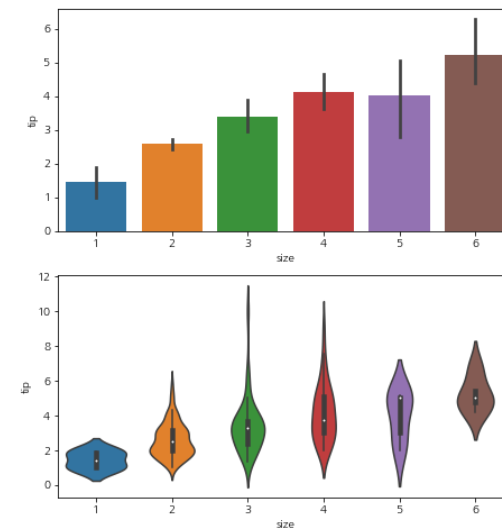
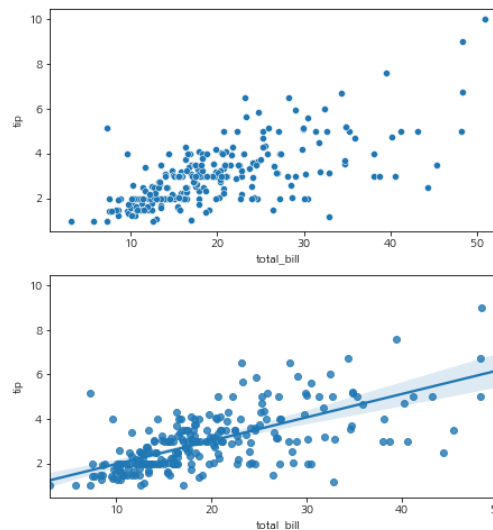
■ matplotlib

```
import seaborn as sns
import pandas as pd
import matplotlib.pyplot as plt
```

```
df = sns.load_dataset("tips")
```


```
fig, ax = plt.subplots(nrows=2, ncols=2)
fig.set_size_inches(16,8)
```

```
sns.scatterplot(x=df["total_bill"], y=df["tip"], ax=ax[0][0])
sns.barplot(x=df['size'], y=df['tip'], ax=ax[0][1])
sns.regplot(x=df["total_bill"], y=df["tip"], ax=ax[1][0])
sns.violinplot(x=df['size'], y=df['tip'], ax=ax[1][1])
```




Ch5. 데이터 수집


■ 데이터 수집




MLB 핀치히터 LADA저스 볼캡 로얄 블루
940 MLB PINCH HITTER 91 LOSDOD D RYL
33,000원




MLB 핀치히터 뉴욕 양키스 볼캡 네이비
940 MLB PINCH HITTER 91 NEYYAN N.
33,000원



MLB 어센틱 뉴욕 양키스 게임(홈/어..
ACPERF NEYYAN GM
49,000원



캡 이레이저
ACC CAP ERASER
3,500원



940 MLB PINCH HITTER 91 LOSDOD D RYL
MLB 핀치히터 LADA저스 볼캡 로얄 블루

33,000원 (5% 적립)

멤버십혜택 회원가입 즉시 사용 가능한 쿠폰을 확인하세요. ▼

제품코드 12024815

소재 100%폴리에스터

행 길이 약 7cm

높이 약 11cm

특징 후면 사이즈 조절 가능

사이즈 옵션 선택 ▼

수량 1

로그인 or 비회원 구매

장바구니 Pay로 즉시 결제

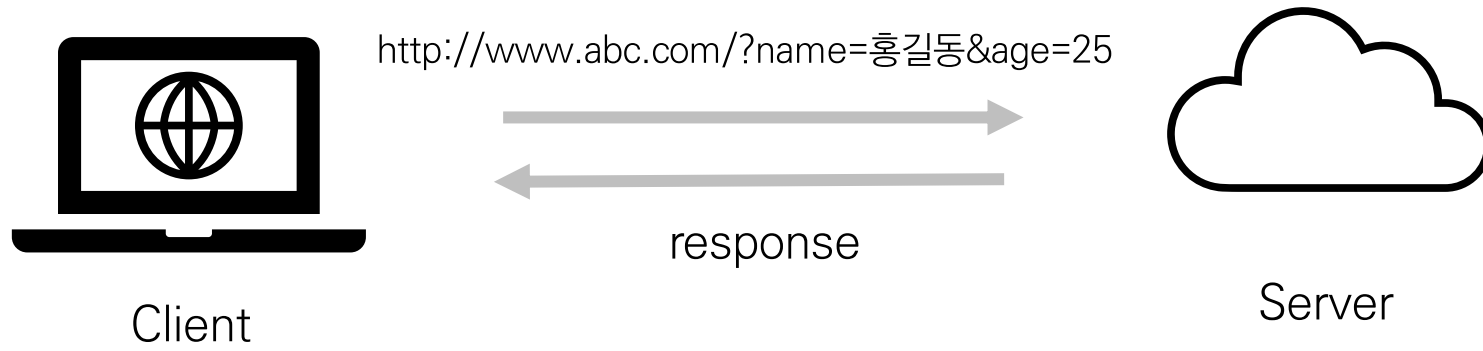
▲ 네이버페이 추가 적립1%, 신용카드 무이자 7개월 등 결제수단별 혜택을 확인하세요. ▼

소재	챙 길이	높이	특징	특이사항	깊이
100%폴리에스터	약 7cm	약 11cm	후면 사이즈 조절 가능	NONESEASON	NaN
100%폴리에스터	약 7cm	약 11cm	NaN	NONESEASON	NaN
100%울	약 7cm	NaN	NaN	18FW	약 11.5cm

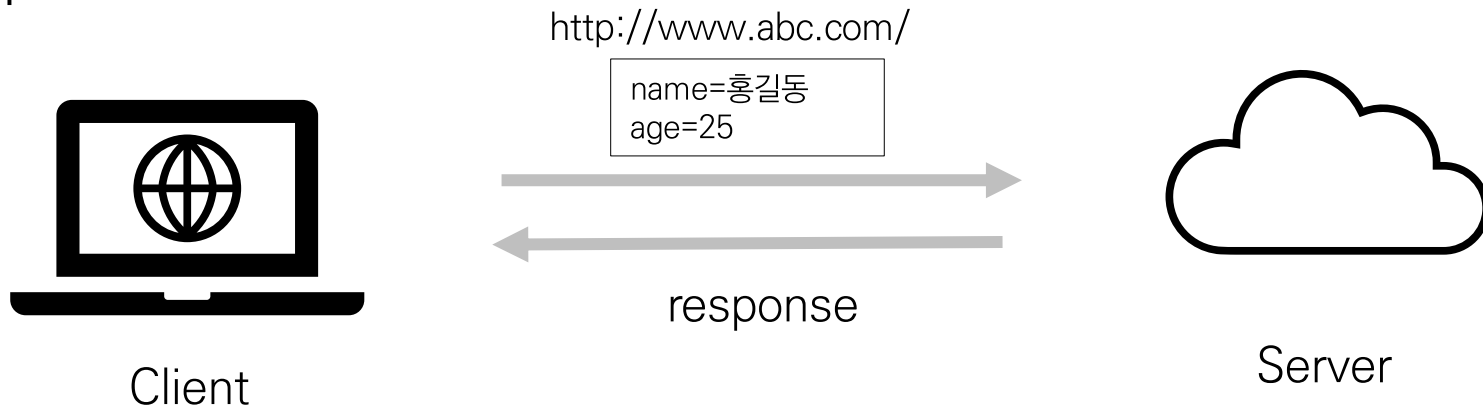
2	에									
3	MLB 핀치히터 디트로이트 타이거즈 볼캡 네이비	12033046	뉴에라 캡 코리아	2019/01/17	100%폴리에스터	약 7cm	약 11cm	NaN	NONESEASON	NaN
4	MLB 팀 클래식 뉴욕 양키스 게임 볼캡 네이비	10975804	뉴에라 캡 코리아	2016/06/14	95%폴리에스터, 5%스판덱스	약 7cm	NaN	NaN	NaN	약 12cm



GET



POST





파이썬

전체 이미지 동영상 도서 뉴스 더보기

검색결과 약 5,250,000개 (0.46초)

https://ko.wikipedia.org › wiki › 파이썬

파이썬 - 위키백과, 우리 모두의 백과사전

파이썬(영어: **Python**)은 1991년 프로그래머인 귀도 반 로섬이 발표한 고급 프로그래밍 언어로 플랫폼에 독립적이며 인터프리터식, 객체지향적, 동적 ...

파일 확장자: py, .pyc, .pyd, .pyo

설계자: 귀도 반 로섬

발표일: 1991년 2월 20일

패러다임: 프로그래밍 패러다임: 객체 지향

개요 · 역사 · 문법 · 자료형

https://docs.python.org › tutorial

파이썬 자습서 — Python 3.9.4 문서

파이썬은 배우기 쉽고, 강력한 프로그래밍 언어입니다. 효율적인 자료 구조들과 객체 지향 프로그래밍에 대해 간단하고도 효과적인 접근법을 제공합니다. 우아한 문법 ...

3. 파이썬의 간략한 소개 · 2. 파이썬 인터프리터 사용하기 · 1. 입맛 돋우기 · 9. 클래스

https://wikidocs.net › ...

1 파이썬 시작하기 - 왕초보를 위한 Python - WikiDocs

우리가 앞으로 배울 **파이썬** 언어는, 배우기 쉬우면서도 프로그램을 빨리 개발할 수 있고, 기능 뛰어나답니다. 이 강좌는 프로그래밍을 처음 하는 분들을 위해서 **파이썬** ...

1.1 파이썬 준비 · 1.2 정수를 계산하기 · 1.3 변수 · 1.4 리스트(list)

https://namu.wiki › Python



Elements Console Sources Network Performance Memory Application

```
<!DOCTYPE html>
<html itemscope itemtype="http://schema.org/SearchResultsPage" lang="ko">
  <head>...</head>
  <body jsmodel="TvHxbe" class="srp vasq" jscontroller="aCZVp" marginheight="3" topmargin="3"
    jsaction="rcuQ6b:npT2md;YUC7He:.CLIENT;IVKTfe:.CLIENT;HiCeld:.CLIENT;KsNBn:.CLIENT;sbTXNb:.CLIE
    T;xjhTIf:.CLIENT;O2vyse:.CLIENT;Ez7VMc:.CLIENT;qqf0n:.CLIENT;me3ike:.CLIENT;IrNywb:.CLIENT;Z94jB
    f:.CLIENT;A8708b:.CLIENT;YcfJ:.CLIENT;A6SDQe:.CLIENT;LjVEjd:.CLIENT;VM8bg:.CLIENT;hWT9Jb:.CLIE
    T;WCulWe:.CLIENT;NTJodf:.CLIENT;szjOR:.CLIENT;PY1zjf:.CLIENT;wnJTPd:.CLIENT;JL9QDc:.CLIENT;kw1xh
    c:.CLIENT;qGMTIf:.CLIENT;aeBrn:.CLIENT" id="gsr">
    <style>...</style>
    <div id="_f8V8YL7VI4e7wAORrKjQCw1">...</div>
    <noscript>...</noscript>
    <h1 class="Uo8X3b">접근성 링크</h1>
    <div jscontroller="r36a9c" class="wYq63b">...</div>
    <div id="_f8V8YL7VI4e7wAORrKjQCw3"></div>
    <div class="CvDJxb big" jscontroller="ZyRBae" jsaction="rcuQ6b:npT2md" id="searchform" style=
    "position: absolute; top: 20px;">
      <div id="_f8V8YL7VI4e7wAORrKjQCw5">...</div>
      <div class="sfbg" style="margin-top:-20px">...</div>
      <form class="tsf" action="/search" id="tsf" data-submitfalse="q" method="GET" name="f"
        role="search">
        <div jsmodel="vWnDde" jsdata="MuIEvd;_;CNIBnE">
          <div jscontroller="mvYTse" jsmodel="TM1YFc" class="A8SBwf" jsdata="LVplcb;_;" jsaction=
            "1X6RWd:w3Wsmc;DkpM0b:d3sQLd;IQOavd:dFyQEf;XzZZPe:jI3wzf;Aghsf:AVsnlb;iHd9U:Q7Cnrc;f5EH
            e:G0jgYd;vmxUb:j3bJnb;R2c50:LuRugf;R3Yrj:DURtdb;qicKjd:ANDidc;NOg9L:HLgh3;uGoIkd:epUokb;z
            LdLw:eaGB5;rcuQ6b:npT2md">
              <div class="logo">...</div>
              <div class="RNNXgb" jsname="RNNXgb">
```

html body#gsr.srp.vasq div#main.main div#cnt.big div#rcnt.GyAeWb

Console What's New

Highlights from the Chrome 89 update

Debugging support for Trusted Type violations

Breakpoint on Trusted Type violations and link to more information in the Issues tab.

Capture node screenshot beyond viewport

Capture node screenshot for a full node including content below the fold.





파이썬

전체

이미지

동영상

도서

뉴스

더보기

검색결과 약 4,530,000개 (0.39초)

② 마우스 선택

https://ko.wikipedia.org › wiki › 파이썬 ▾

파이썬 - 위키백과, 우리 모두의 백과사전

h3.LC201b.DKV0Md 360 × 30.6

Color ■ #1A0DAB

Font 20px "Apple SD Gothic Neo", arial, sans-s...

Margin 0px 0px 3px

Padding 5px 0px 0px

ACCESSIBILITY

Contrast Aa 12.42 ✔

Name 파이썬 - 위키백과, 우리 모두의 백과사전

Role heading

Keyboard-focusable ↻

파이썬은 배우기 쉽고, 강력한 프로그래밍 언어입니다. 효율적인 자료 구조들과 객체 지향 프로그래밍에 대해 간단하고도 효과적인 접근법을 제공합니다. 우아한 문법 ...

① 기능 활성화

③ 선택부분 코드 표시

Elements Console Sources Network Performance Memory Application >>

```

<div class="hlw0c">
  <div class="g">
    <h2 class="Uo8X3b">웹 검색결과</h2>
    <div data-hveid="CAYQAA" data-ved="2ahUKEwiTtbSrQZbwAhVFyIsBHfaDCakQFSgAMAB6BAGGEEA">
      <div class="tF2Cxc">
        <div class="yuRUBf">
          <a href="https://ko.wikipedia.org/wiki/%ED%8C%8CEC%9D%B4%EC%8D%AC" data-ved="2ahUKEwiTtbSrQZbwAhVFyIsBHfaDCakQFjAAegQIBhAD" ping="/url?sa=t&source=web&rct=j&url=https://ko.wikipedia.org/wiki/%25ED%258C%258C%25EC%259D%25B4%25EC%258D%25AC&ved=2ahUKEwiTtbSrQZbwAhVFyIsBHfaDCakQFjAAegQIBhAD">
            <br>
            <h3 class="LC201b DKV0Md">파이썬 - 위키백과, 우리 모두의 백과사전</h3> == $0
          <div class="TbwUpd NJjxre">...</div>
        </a>
        <div class="B6fmyf">...</div>
      </div>
      <div class="IsZvec">...</div>
      <div jscontroller="m6a01" id="eob_16" jsdata="fxg5tf;_;AZJDTc" jsaction="rcuQ6b:npT2md" data-ved="2ahUKEwiTtbSrQZbwAhVFyIsBHfaDCakQ2Z0BMAB6BAGGEBU">...</div>
    </div>
  </div>
</div>
  <div class="g">...</div>
</div>
  <div class="g">...</div>

```

파이썬



🔍 전체 🖼️ 이미지 📺 동영상 📖 도서 📰 뉴스 : 더보기

검색결과 약 5,250,000개 (0.46초)

<https://ko.wikipedia.org> > wiki > 파이썬 ▾

파이썬 - 위키백과, 우리 모두의 백과사전

파이썬(영어: **Python**)은 1991년 프로그래머인 귀도 반 로섬이 발표한 고급 프로그래밍 언어로, 플랫폼에 독립적이며 인터프리터식, 객체지향적, 동적 ...

파일 확장자: py, .pyc, .pyd, .pyo

설계자: 귀도 반 로섬

발표일: 1991년 2월 20일

패러다임: 프로그래밍 패러다임: ?

개요 · 역사 · 문법 · 자료형

<https://docs.python.org> > tutorial ▾

파이썬 자습서 — Python 3.9.4 문서

파이썬은 배우기 쉽고, 강력한 프로그래밍 언어입니다. 효율적인 자료 구조들과 객체 래밍에 대해 간단하고도 효과적인 접근법을 제공합니다. 우아한 문법 ...

3. 파이썬의 간략한 소개 · 2. 파이썬 인터프리터 사용하기 · 1. 입맛 돋우기 · 9. 클래스

<https://wikidocs.net> > ... ▾

1 파이썬 시작하기 - 왕초보를 위한 Python - WikiDocs

우리가 앞으로 배울 파이썬 언어는, 배우기 쉬우면서도 프로그램을 빨리 개발할 수 있 뛰어나답니다. 이 강좌는 프로그래밍을 처음 하는 분들을 위해서 파이썬 ...

1.1 파이썬 준비 · 1.2 정수를 계산하기 · 1.3 변수 · 1.4 리스트(list)

① URL정보

<div class="TbwUpd NJjxre"> ...

② 제목

<h3 class="LC20lb DKV0Md"> ...

③ 본문 일부

 ...

④ 부가정보

<div class="FyYA1e lThcWe"> ...

- request
 - html 받아오기
- BeautifulSoup
 - html 태그 기반 데이터 파싱
- selenium
 - 동적으로 브라우저 제어

■ request

```
from bs4 import BeautifulSoup
import requests
import pandas as pd
response = requests.get('http://www.neweracapkorea.com/shop/shopbrand.html?xcode=031&.. 생략...')
response.content
```

```
b'\n<!DOCTYPE html PUBLIC "-//W3C//DTD HTML 4.01//EN"
"http://www.w3.org/TR/html4/strict.dtd">\n<html>\n<head>\n<meta http-equiv="CONTENT-TYPE"
content="text/html; charset=EUC-KR">\n<meta http-equiv="X-UA-Compatible" content="IE=edge"
/>\n<meta name="naver-site-verification"
content="a3e4e840fe9e419b8e34b1064f4d0c9a46c54efd"/>\n<meta name="google-site-verification"
content="6Nl8EyimZc-dfQImXCJKEsojYPUxowwcTvw1kQufiG8" />\n<meta name="naver-site-verification"
content="a3e4e840fe9e419b8e34b1064f4d0c9a46c54efd"/>\n
... 생략...
```

■ BeautifulSoup

```
soup = BeautifulSoup(response.content, "lxml")  
soup
```

```
<!DOCTYPE html PUBLIC "-//W3C//DTD HTML 4.01//EN" "http://www.w3.org/TR/html4/strict.dtd">  
<html>  
<head>  
<meta content="text/html; charset=utf-8" http-equiv="CONTENT-TYPE"/>  
<meta content="IE=edge" http-equiv="X-UA-Compatible"/>  
<meta content="a3e4e840fe9e419b8e34b1064f4d0c9a46c54efd" name="naver-site-verification"/>  
<meta content="6Nl8EyimZc-dfQImXCJKEsojYPUxowwcTvw1kQufiG8" name="google-site-verification"/>  
<meta content="a3e4e840fe9e419b8e34b1064f4d0c9a46c54efd" name="naver-site-verification"/>  
<meta content="website" property="og:type"/> <meta content="뉴에라 공식 온라인 스토어"  
property="og:title"/>  
<meta content="MLB,NBA,NFL 미국 3대 스포츠 리그의 공식 선수용 모자 및 다양한 라이프 스타일  
아이템" property="og:description"/>  
... 생략 ...
```

■ BeautifulSoup

```
cap_info = soup.findAll('div', {'class':'tb-center'})
```

```
cap_info[0]
```

```
<div class="tb-center">
<div class="thumb">
<a
href="/shop/shopdetail.html?branduid=3353566&xcode=031&mcode=002&scode=001&type=Y&so
rt=sellcnt&cur_code=031002&GfDT=Z2Z3UQ%3D%3D"></a>
</div>
<ul class="info">
<li class="dsc">MLB 핀치히터 LA다저스 볼캡 로얄 블..</li>
<li class="product_etc">940 MLB PINCH HITTER 91 LOSDOD D..</li> <li class="pdt10"></li>
<li class="price">33,000원</li>
<li class="info_icon" id="info_icon">
<span id="btn_opt01"></span><span id="btn_opt02"><a class="btn-
overlay-show" href="javascript:viewdetail('060006000266', '1', '');"></a></span></li>
<li class="icon"><span class="MK-product-icons"></span></li>
</ul>
</div>
```

■ BeautifulSoup

```
#url
cap_url = cap_info[0].find('a').get('href')
print(cap_url)

#상품명
name = cap_info[0].find('li', {'class':'dsc'}).text
print(name)

#가격
price = cap_info[0].find('li', {'class':'price'}).text
print(price)
```

```
/shop/shopdetail.html?branduid=3353566&xcode=031&mcode=002&scode=001&type=Y&sort=sel
lcnt&cur_code=031002&GfDT=Z2Z3UQ%3D%3D
MLB 핀치히터 LA다저스 볼캡 로얄 블..
33,000원
```


■ BeautifulSoup

```
name_list = []
price_list = []
url_list = []

for cap in cap_info:
    name = cap.find('li', {'class':'dsc'}).text
    price = cap.find('li', {'class':'price'}).text
    url = cap.find('a').get('href')
    print("이름: {}, 가격: {}".format(name, price))
    name_list.append(name)
    price_list.append(price)
    url_list.append(url)
```

이름: MLB 핀치히터 LA다저스 볼캡 로얄 블..., 가격: 33,000원

이름: MLB 핀치히터 뉴욕 양키스 볼캡 네이..., 가격: 33,000원

이름: MLB 팀 클래식 뉴욕 양키스 게임 볼..., 가격: 36,000원

이름: 에센셜 무지 사이즈캡 블랙, 가격: 49,000원

이름: MLB 핀치히터 디트로이트 타이거즈..., 가격: 33,000원

■ BeautifulSoup

```
df = pd.DataFrame({"이름": name_list, "가격" : price_list, "url": url_list})  
df.head()
```

	이름	가격	url
0	MLB 핀치히터 LA다저스 볼캡 로얄 블..	33,000원	/shop/shopdetail.html?branduid=3353566&xcode=0...
1	MLB 핀치히터 뉴욕 양키스 볼캡 네이..	33,000원	/shop/shopdetail.html?branduid=3353567&xcode=0...
2	MLB 팀 클래식 뉴욕 양키스 게임 볼..	36,000원	/shop/shopdetail.html?branduid=4633&xcode=031&...
3	에센셜 무지 사이즈캡 블랙	49,000원	/shop/shopdetail.html?branduid=3353027&xcode=0...
4	MLB 핀치히터 디트로이트 타이거즈..	33,000원	/shop/shopdetail.html?branduid=3353722&xcode=0...

- selenium
 - 동적으로 브라우저 제어



request로 HTML을 요청	Javascript 실행 전 HTML
Selenium으로 HTML을 요청	Javascript 실행 완료한 HTML

Q&A

THANK YOU
