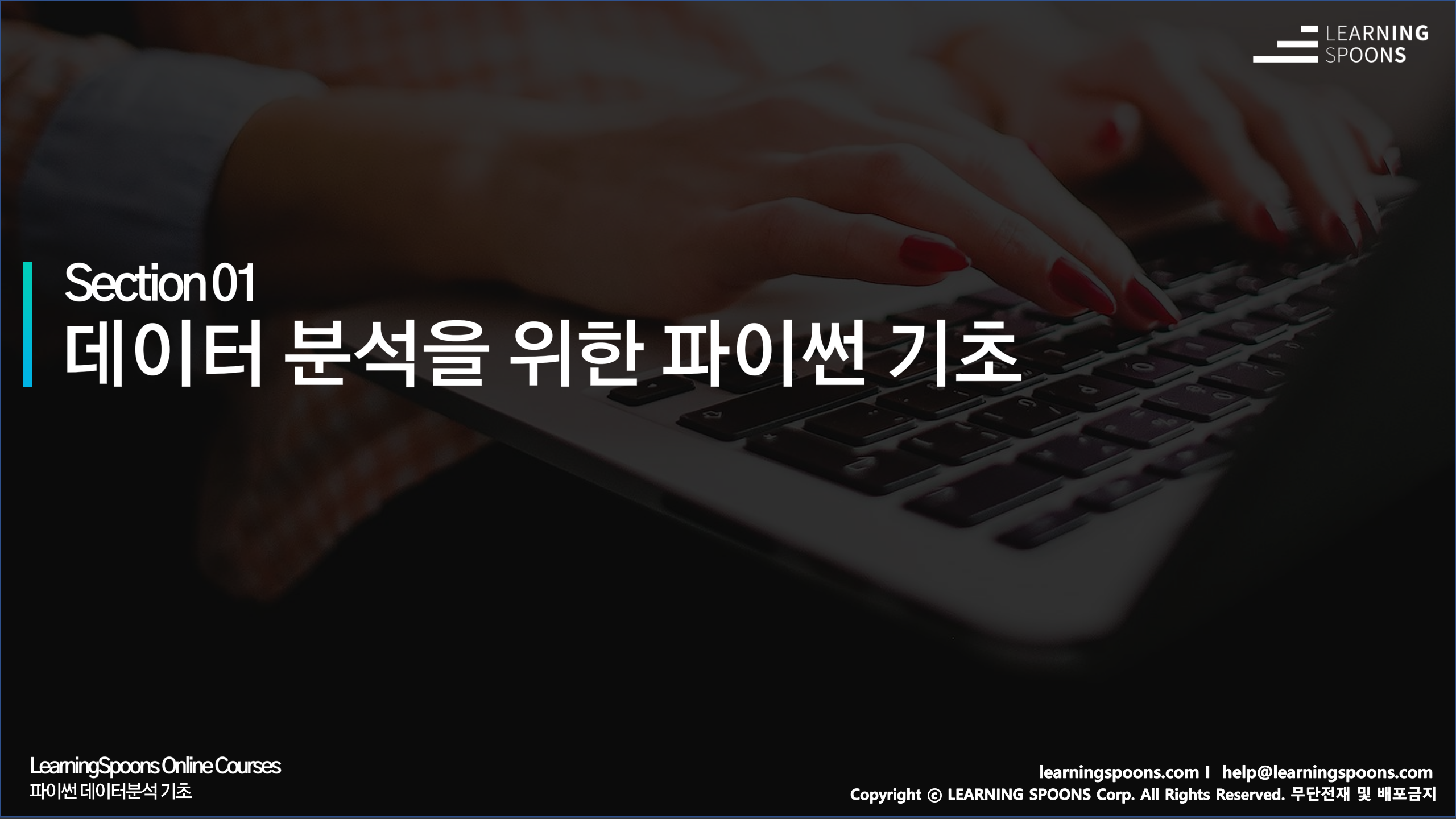
LearningSpoons Online

입문자를 위한 파이썬 데이터분석 & 시각화

장남수 강사

목차



LEARNING SPOONS

Section 01

데이터 분석을 위한 파이썬 기초

LEARNING SPOONS

Unit 1 - 1
파이썬 준비하기

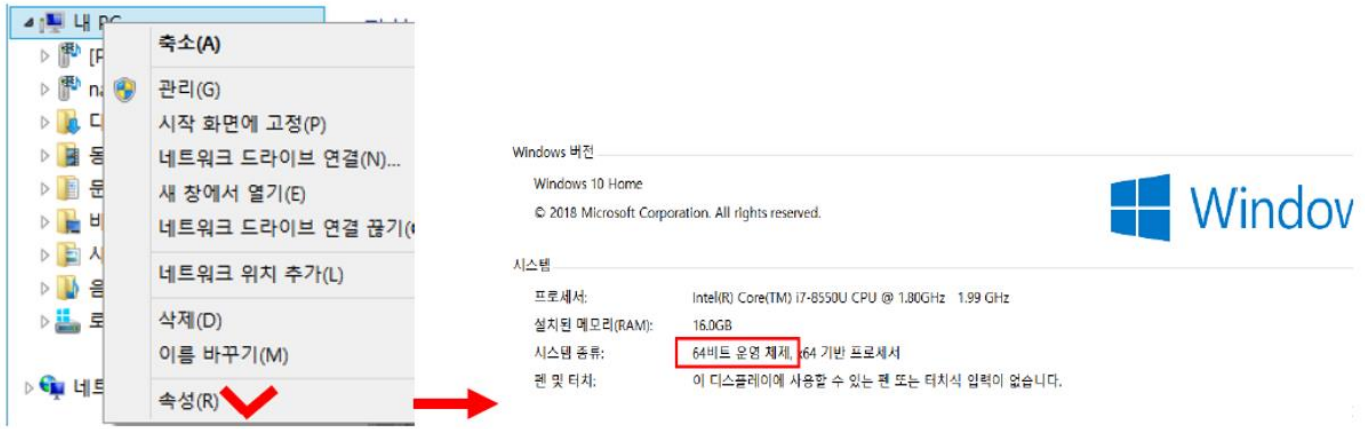
LEARNING SPOONS

파이썬 설치하기(아나콘다)

아나콘다(Anaconda)? 파이썬 + 주요 라이브러리 + 입/출력 프로그램

Step1) 설치 환경 확인하기

- 내 컴퓨터 OS 확인하기: Mac / Windows / Linux..
- 운영체계 BIT 확인하기 : 32 bit / 64 bit



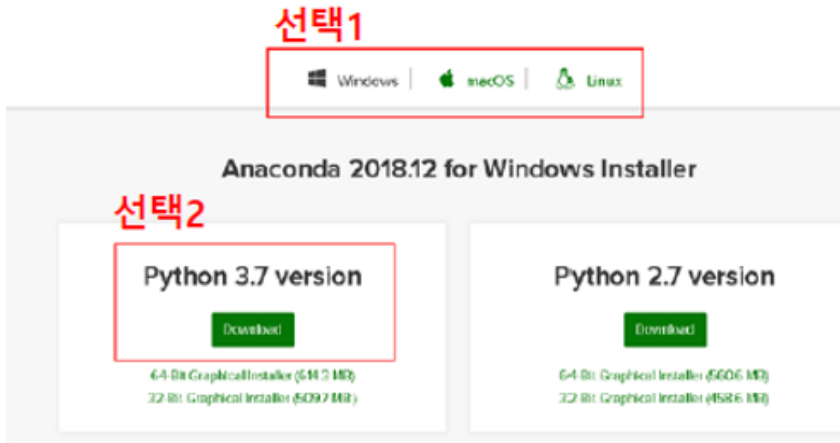
[운영체계 BIT 확인 방법] 윈도우 탐색기 → 내PC 우클릭 → 속성

LEARNING SPOONS

파이썬 설치하기(아나콘다)

Step2) 아나콘다 다운로드

- https://www.anaconda.com/download/
- **파이썬 3.x 버전**



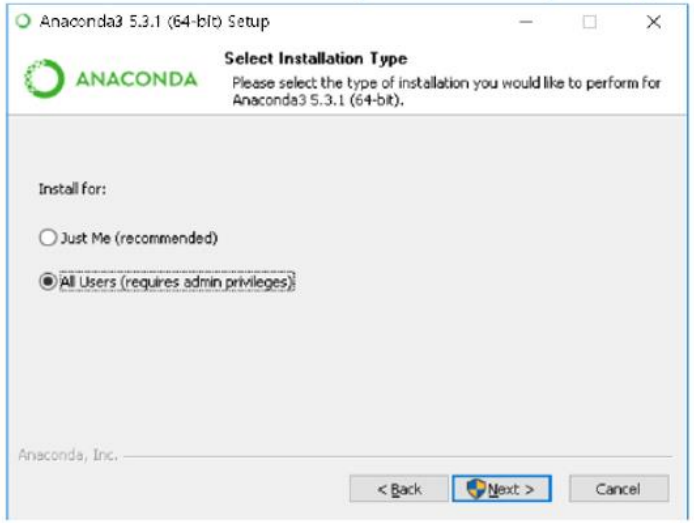
<Step 1에서 확인한 버전으로 다운로드>

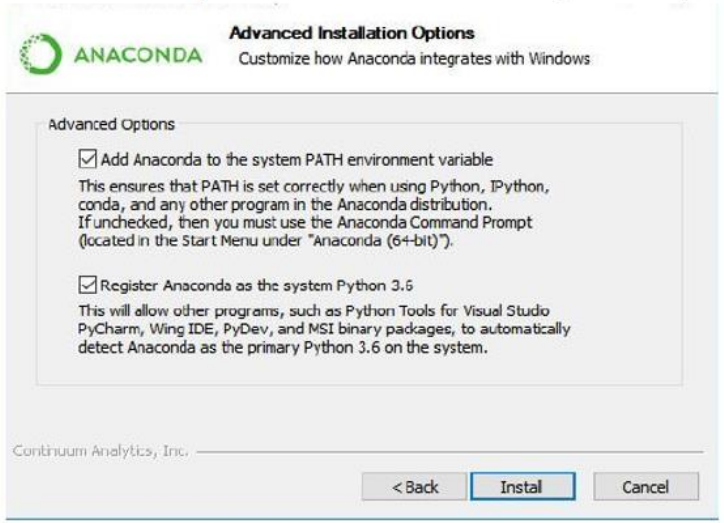
LEARNING SPOONS

파이썬 설치하기(아나콘다)

Step3) 아나콘다 설치

- Install for : All User(requires admin privileges)
- Advanced Options : 두 가지 모두 체크(윈도우 환경설정)





LEARNING SPOONS

쥬피터 노트북 실행하기

실행방법

- (Mac / Linux) 커맨드 창에서 jupyter notebook 실행
- (windows) 시작 메뉴에서 jupyter notebook

<인터넷 브라우저가 실행되면서 localhost:8888/tree 접속 >

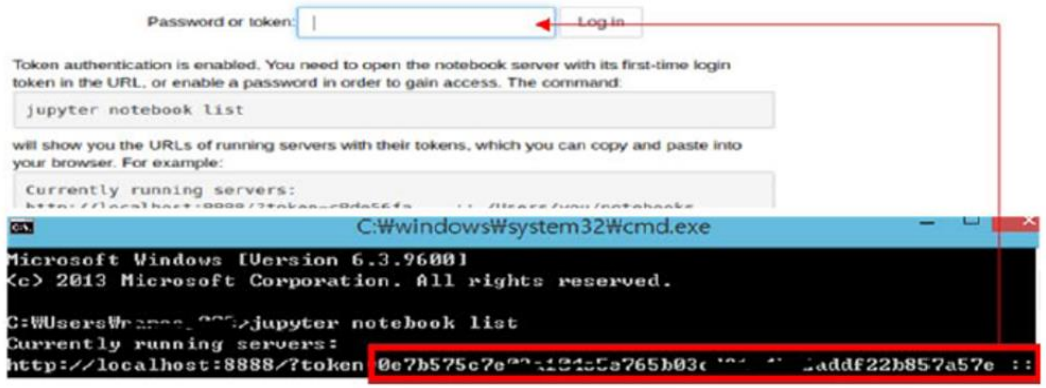
LEARNING SPOONS

쥬피터 노트북 실행하기

Q) 만약 토큰 값을 입력하라고 나온다면?

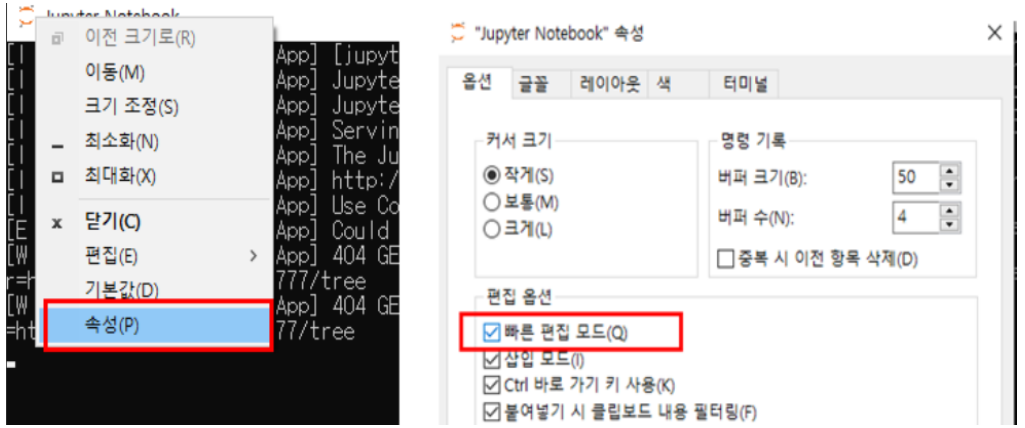
http://localhost:8888/?token=토큰값

토큰값 ← 복사&붙여넣기



Q) 만약 토큰값 복사하기가 되지 않는다면?

커맨드창 메뉴바 마우스 우클릭 → "속성" → "빠른 편집 모드" 체크



LEARNING SPOONS

Unit 1 - 2

데이터 분석을 위한 파이썬 자료형

장남수 강사

LEARNING SPOONS

변수명 = 데이터

데이터 타입	내용	예시
int	정수	1, 2, 3, 4, 5
float	실수(소수점 존재)	3.14, 9.99
str	문자 나열	"파이썬", '대한민국'
list	순서대로 나열된 그룹	[1, 2, 3, 4, 5], ["가", "나", "다"]
dict	이름(key)별로 정의된 그룹	{ '이름' : '홍길동', '전화번호 ' : '010-0000-0000', '주소' : '대한민국 서울' }

LEARNING SPOONS

Unit 1 - 3

데이터 분석을 위한 기본 문법

장남수 강사

LEARNING SPOONS

파이썬 기본 문법

- 반복문
- 조건문
- 문자열포매팅
- 함수

LEARNING SPOONS

Section 02

데이터 분석을 위한 클래스 – pandas

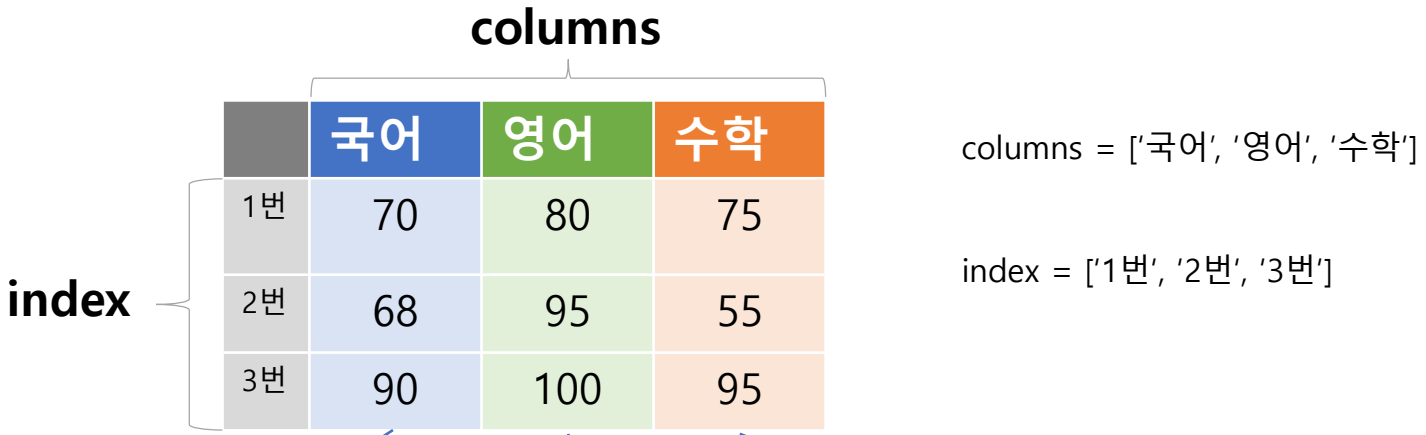
LEARNING SPOONS

Unit 2-1
판다스 기본 사용법

LEARNING SPOONS

판다스 데이터 구조

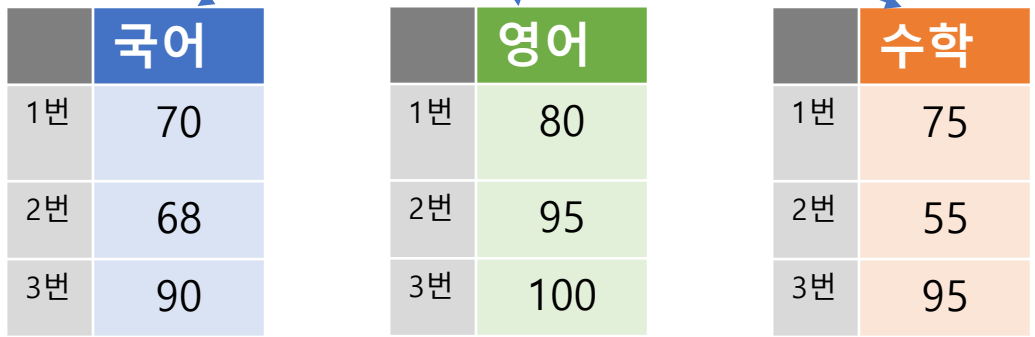
DataFrame



columns = ['국어', '영어', '수학']

index = ['1번', '2번', '3번']

Series



LEARNING SPOONS

판다스 기본 사용법

- 데이터 파일 읽기 : read_excel(), read_csv()
- 데이터 선택하기 : df.loc(), df.iloc()
- 인덱스 / 컬럼 변경하기 : columns/index, reset_index()

LEARNING SPOONS

Unit 2-2

데이터 병합하기

LEARNING SPOONS

데이터 추가/병합하기

- 컬럼 데이터 추가하기
- 두 데이터 병합하기: pd.merge( )

LEARNING SPOONS

데이터 병합하기 pd.merge()

- **"left"**: 왼쪽 데이터(A) 에 속한 데이터 기준
- **"right"**: 오른쪽 데이터(B)에 속한 데이터 기준
- **"inner"**: 양쪽 (A와 B)에 모두 속한 데이터 기준
- **"outter"**: 모든 데이터 기준

```
pd.merge( A , B ,  how = "left" ,  left_on = "A컬럼명 "  , right_on = "B컬럼명" )
                                   left_index = True "  ,  right_index = True"  )
```

왼쪽(A) 테이블 기준값 지정

오른쪽(B) 테이블 기준값 지정

A

	국어	영어	수학
1번	70	80	75
2번	68	95	55
3번	90	100	95

B

	과학	사회
1번	70	80
2번	80	85
4번	95	100
5번	90	70

LEARNING SPOONS

Unit 2-3
정리/집계하기

LEARNING SPOONS

데이터 정리/집계 하기

- 조건에 만족하는 데이터 살펴보기: df [조건]
- 특정 기준으로 테이블 변환하기: df.pivot_table( )
- 정렬하기 : df.sort_values( )

LEARNING SPOONS

Unit 2-4

실습) 영어 이름 트렌드 살펴보기

LEARNING SPOONS

Q) 남자는 James, 여자는 Mary 어때?

- 최근 영어 이름 트렌드 살펴보기

LEARNING SPOONS

Section 03

데이터 시각화를 위한 클래스 – seaborn, folium

LEARNING SPOONS

Unit 3-1

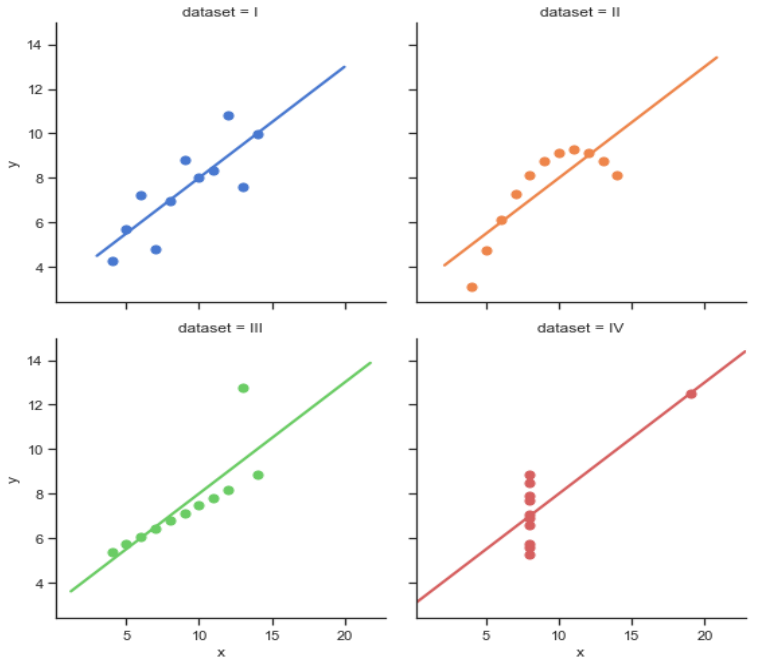
seaborn으로 시각화 하기

LEARNING SPOONS

데이터 시각화의 중요성

통계량이 모두 동일한 데이터 set ➔ ~~똑같다??~~

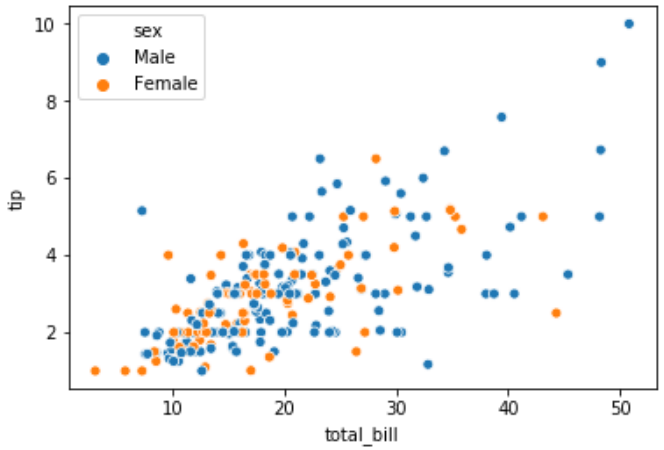
Property	Value
Mean of x	9
Sample variance of x	11
Mean of y	7.5
Sample variance of y	4.125
Correlation between x and y	0.816
Linear regression line	y = 3.00 + 0.500x



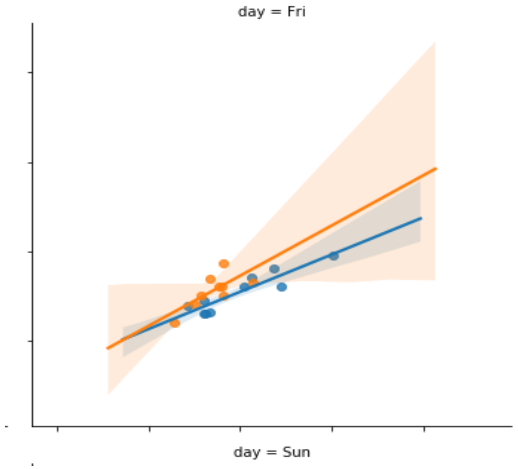
LEARNING SPOONS

데이터 타입별 시각화: 수치형 x 수치형

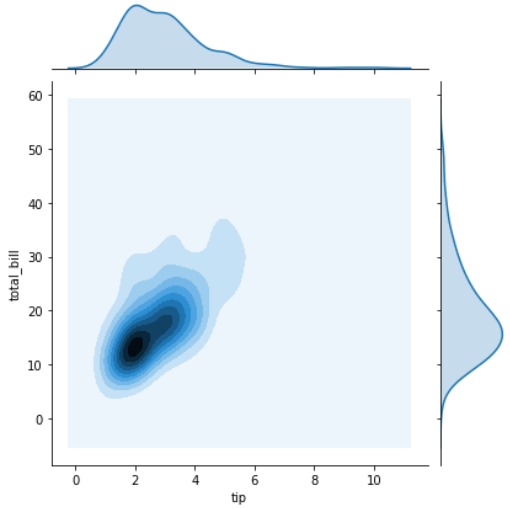
scatterplot



lmplot



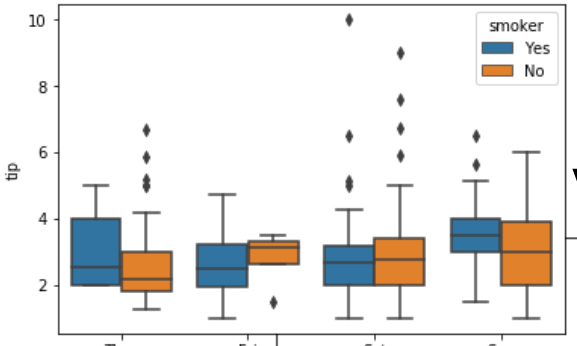
jointplot



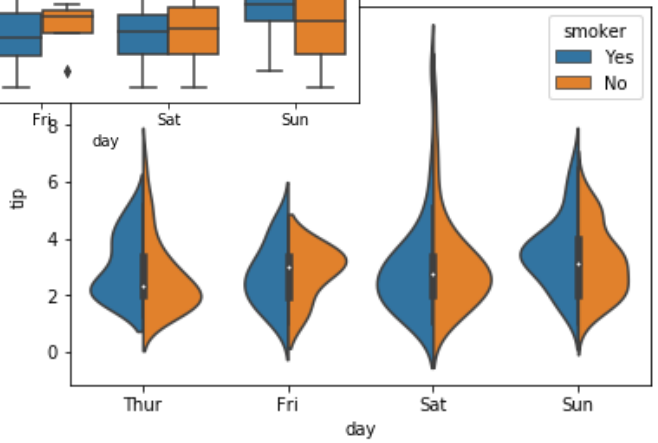
LEARNING SPOONS

데이터 타입별 시각화: 수치형 x 카테고리형

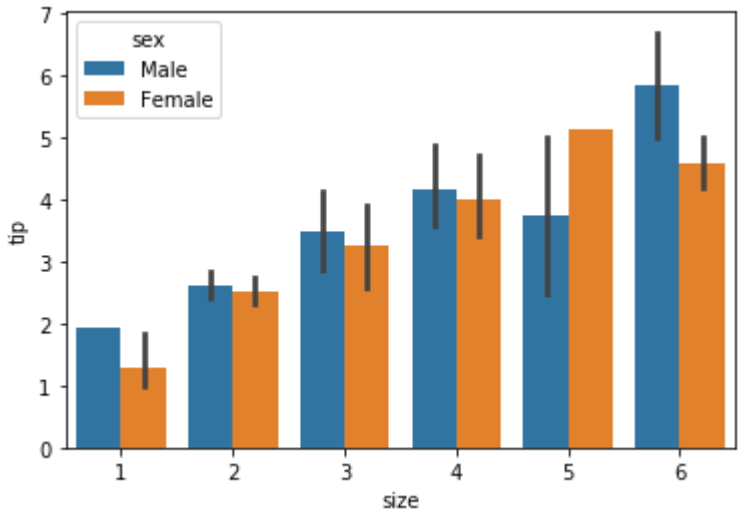
boxplot



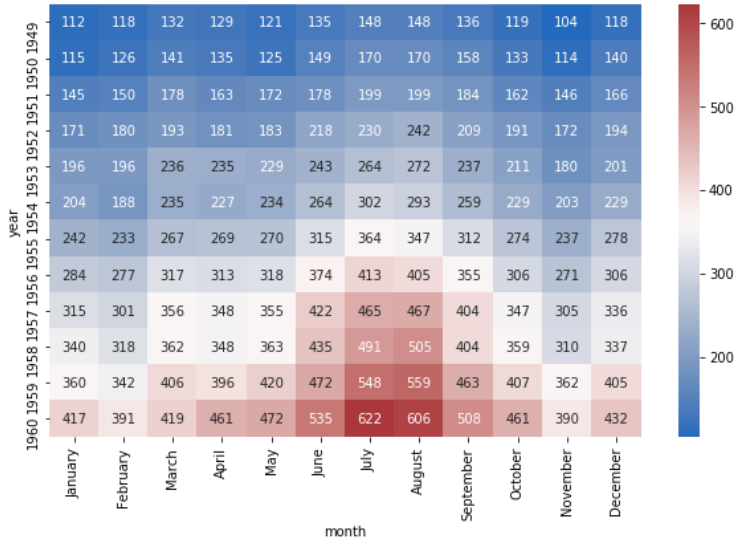
violinplot



barplot



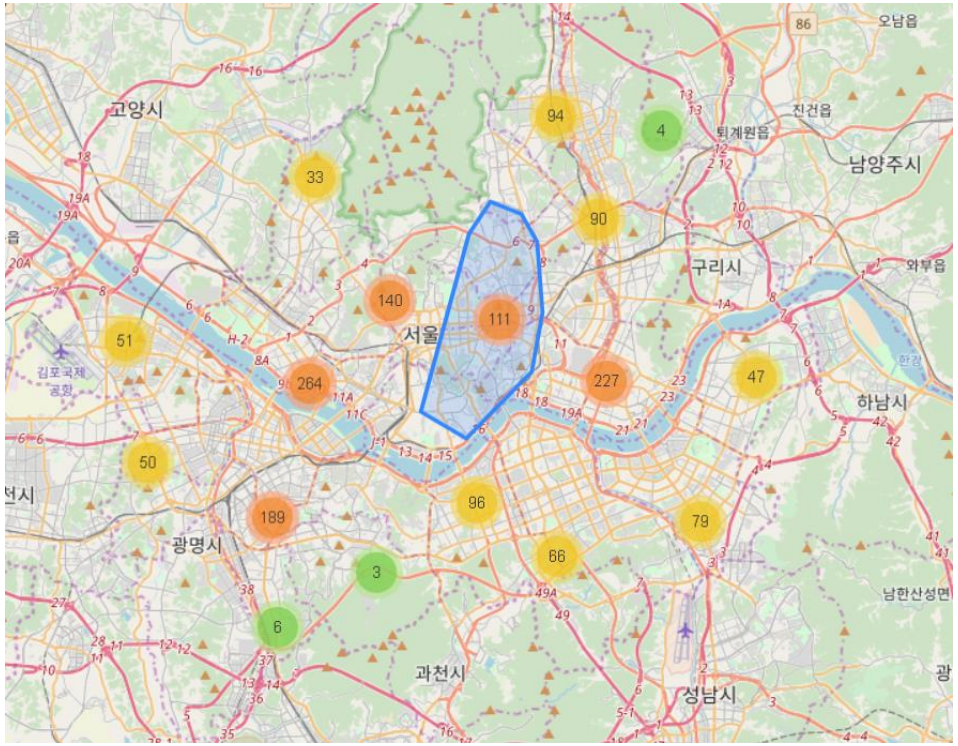
heatmap



LEARNING SPOONS

데이터 타입별 시각화: 수치형 x 위치정보

folium 라이브러리 활용



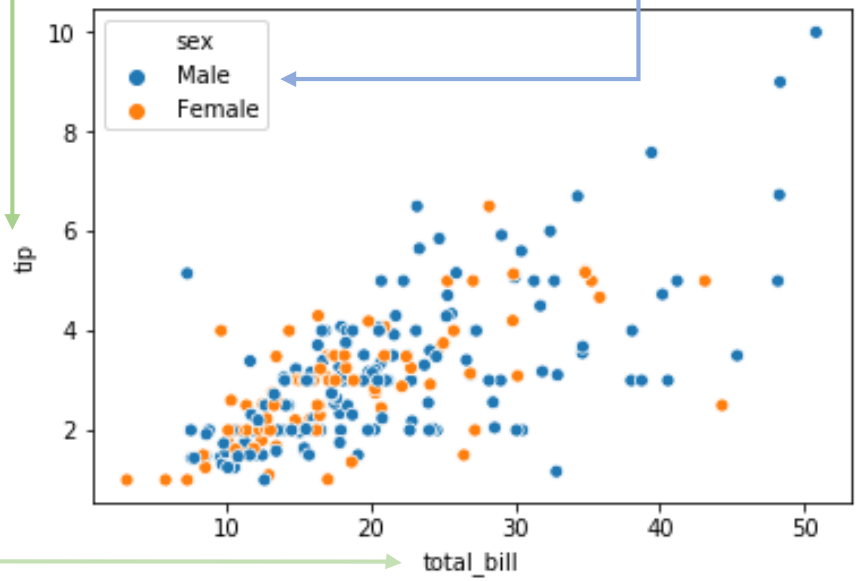
LEARNING SPOONS

seaborn 함수 기본 형태

```
sns.scatterplot ( data = 데이터프레임, x = 'total_bill', y = 'tip', hue = 'sex' )
```

그래프 종류

	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



LEARNING SPOONS

heatmap 사용법

```
sns.heatmap( data = 데이터프레임,
             annot = True, fmt = '.1f',
             cmap = 'RdBu_r'
             )
```

히트맵에서 실제 숫자 표현,
fmt 옵션을 조정하여 표시 형태 정리

fmt = '.1f' → 소수점첫째자리까지 표시
fmt = '.2f' → 소수점 둘째자리까지 표시
fmt = '.0f' → 정수로 표시

히트맵 색상 컬러 조정
추천색상: Reds, Blues, vlag, Pastel1

LEARNING SPOONS

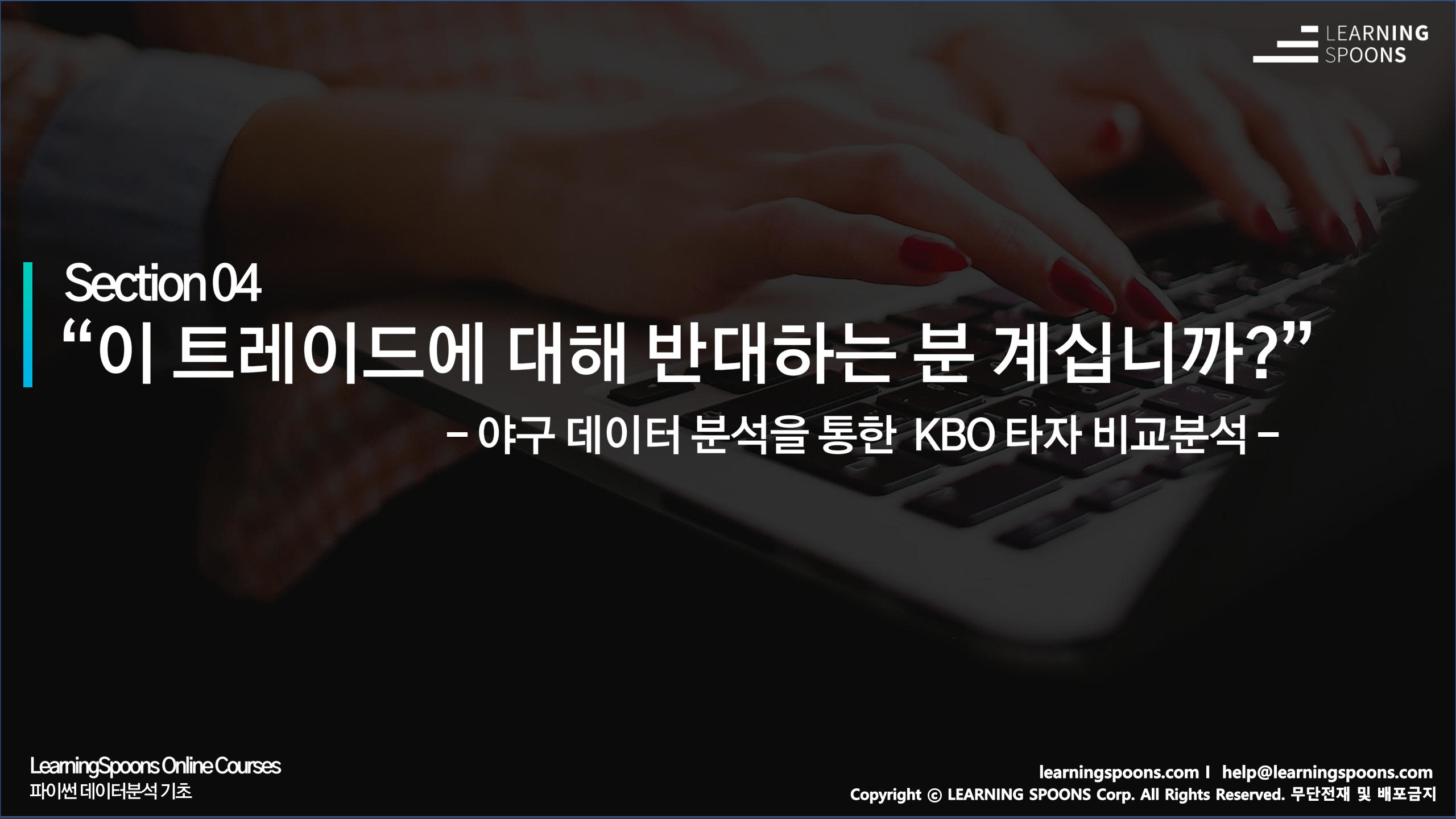
Unit 3-2

folium으로 지도시각화 하기

LEARNING SPOONS

지도시각화 by folium

- 지도 생성하기
- 마커 추가하기
- 원 추가하기
- 툴팁/팝업 정보 추가하기



LEARNING SPOONS

Section 04

“이 트레이드에 대해 반대하는 분 계십니까?”

- 야구 데이터 분석을 통한 KBO 타자 비교분석 -

LEARNING SPOONS

Unit 4-1
Who is the Best player?

LEARNING SPOONS

Q) 최고의 타자는?

지표	의미	계산식
타율	**타격에 성공해 살아나가는 정도**	= 타격 성공 횟수 / 타격 기회 수 **= 안타 수 / 타수**
출루율	**살아서 나가는 정도**	= 진루 성공 횟수 / 진루 기회 수 **= (안타+볼넷+몸에맞는볼) / (타수+볼넷+몸에맞는볼+희생플라이)**
장타율	**타격에 성공해 멀리 살아나가는 정도**	= 진루한 베이스 수 / 타격 기회 수 **= 루타 수 / 타수** * 타율에 거리 개념 추가, 2루타 = 1루타 x 2
OPS	**살아서 멀리 나가는 정도**	**= 출루율 + 장타율**

LEARNING SPOONS

Unit 4-2

여름에 힘 떨어지는 타자? vs 꾸준한 타자

LEARNING SPOONS

Q) 여름에 힘떨어지는 타자?? 꾸준한 타자?

- 월별 실적 변화 비교 분석하기

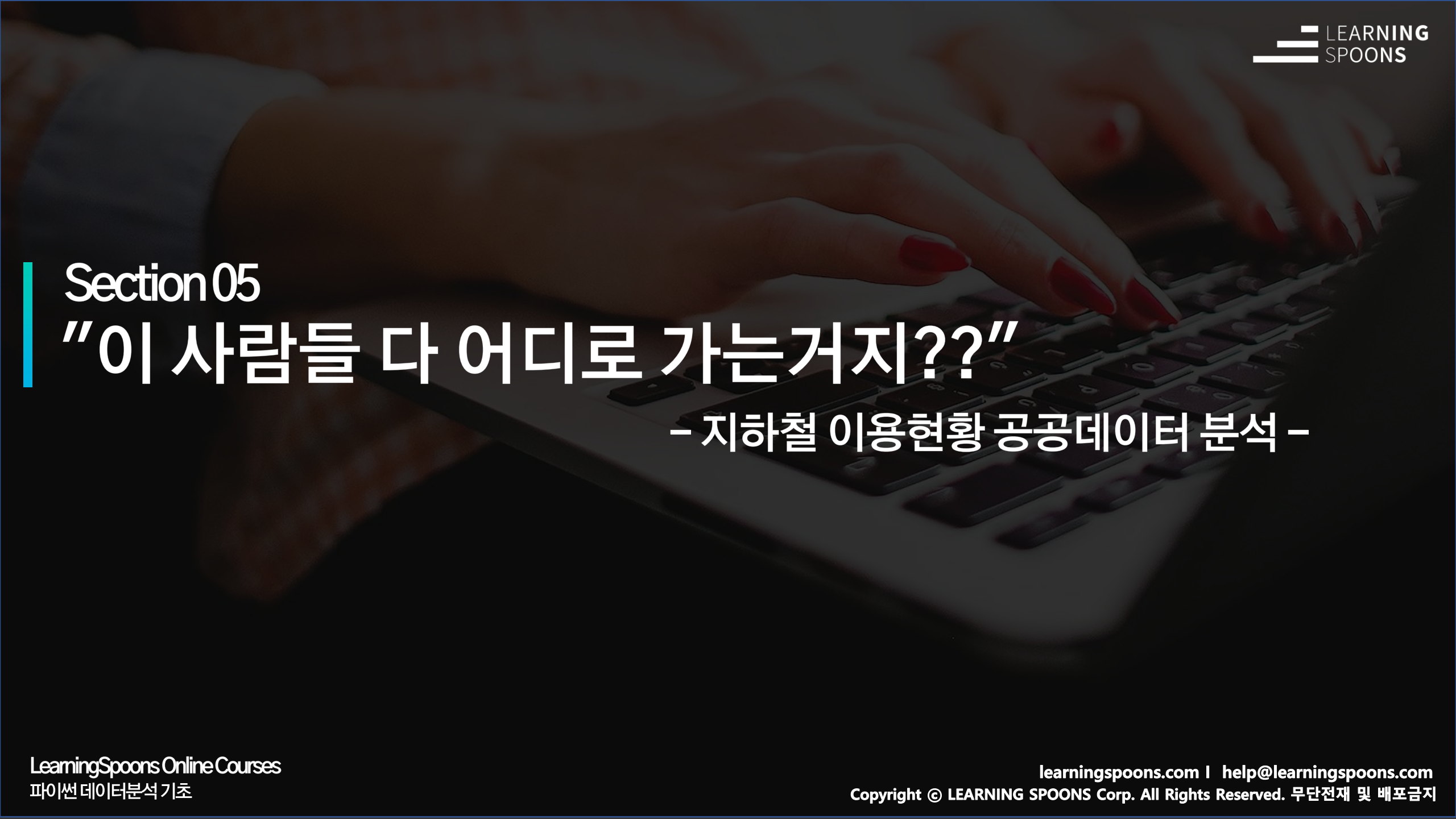
Unit 4-3

왜 우리 팀만 만나면 잘하는거야?

LEARNING SPOONS

Q) 우리팀을 상대로 가장 강한 타자는?

- 팀별 실적 비교 분석하기



LEARNING SPOONS

Section 05

"이 사람들 다 어디로 가는거지??"

- 지하철 이용현황 공공데이터 분석 -

LEARNING SPOONS

Unit 5-1

여러 개의 엑셀 파일 통합 정리하기

장남수 강사

LEARNING SPOONS

엑셀 파일 통합하기

- 내 컴퓨터 파일 조회하기(os)
- 반복문을 통한 파일 열기 / 병합하기
- 저장하기

LEARNING SPOONS

Unit 5-2

일자 별 승객 수 살펴보기

LEARNING SPOONS

Q) 언제가 가장 승객수가 많을까?

- 월별 / 일자별 / 요일별 승객수 비교

LEARNING SPOONS

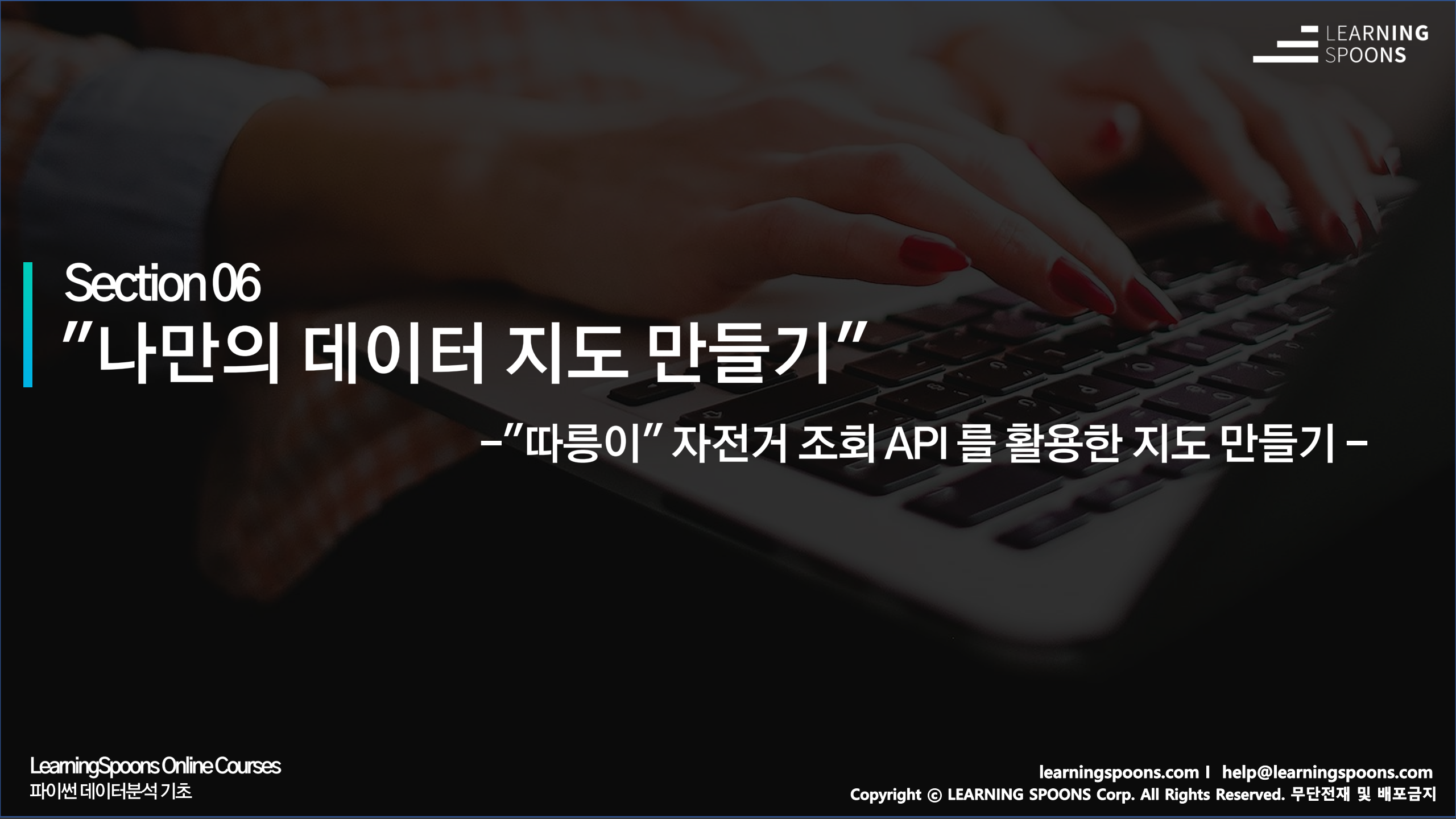
Unit 5-3

지하철역별 승객 수 살펴보기

LEARNING SPOONS

Q) 사람들은 어디서 탈까?? 어디로 갈까??

- 역별 승하차 인원 비교



LEARNING SPOONS

Section 06

"나만의 데이터 지도 만들기"

-"따릉이" 자전거 조회 API 를 활용한 지도 만들기 -

learningspoons.com I help@learningspoons.com

LEARNING SPOONS

Unit 6-1

오픈API 신청하기

LEARNING SPOONS

데이터 수집 방법

- 이미 가지고 있는 데이터 준비/정리
- 웹에 있는 자료 Copy&Paste → 크롤링
- 회사/기관/개인에게 자료 요청하고 받기 → API

LEARNING SPOONS

온라인 데이터 수집 방법

- 웹에 있는 자료 Copy&Paste → 크롤링
- 회사/기관/개인에게 자료 요청하고 받기 → API

LEARNING SPOONS

Q) API는 어떻게 사용할 수 있나요?

- 회사/기관/개인이 만든 규칙 활용

  → 주소 / 명령어 등의 기준에 따라 요청

- 일반적으로 API 사용량 등을 관리하기 위해 사용승인 필요

  → API Key 를 획득 한 후, 함께 전달 필요

LEARNING SPOONS

공공데이터 API 사용 신청하기

1. 서울 열린 데이터광장(http://data.seoul.go.kr/) 회원가입/로그인

2. 서울특별시 공공자전거 실시간 대여정보

   (http://data.seoul.go.kr/dataList/OA-15493/A/1/datasetView.do)

3. 인증키 신청

LEARNING SPOONS

공공데이터 API 사용 신청하기

4. 가입 신청서 작성

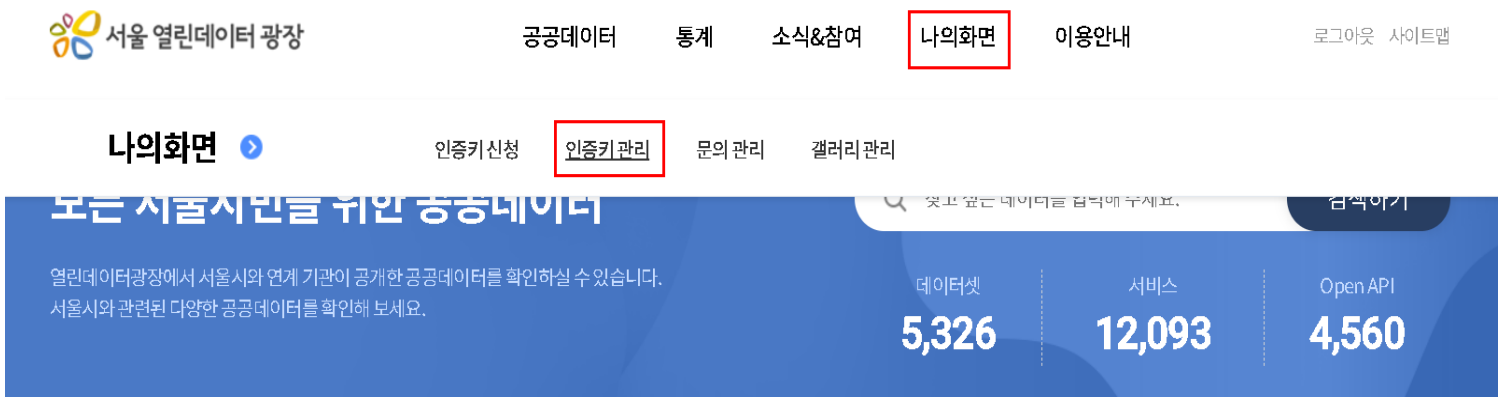
- 사용URL: localhost
- 이메일: 본인 이메일 작성
- 활용용도 및 내용 등록

LEARNING SPOONS

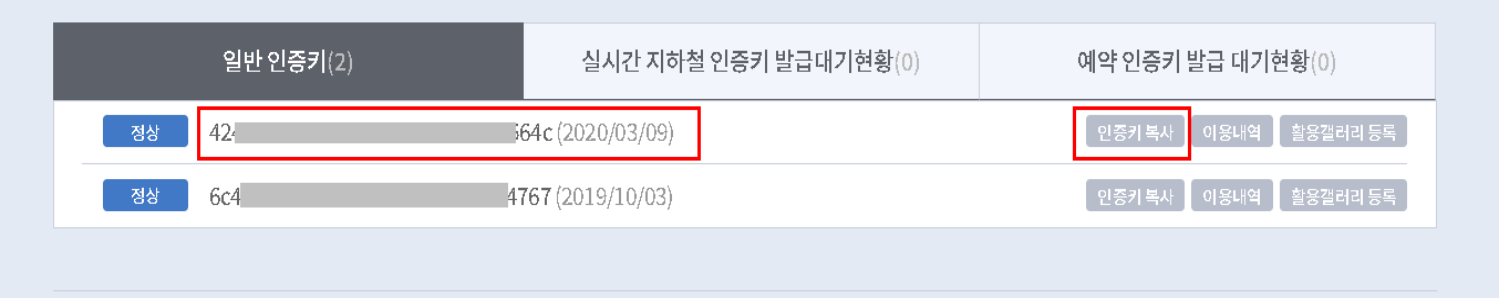
공공데이터 API 사용 신청하기

5. 인증키 확인

- 사용URL: localhost
- 이메일: 본인 이메일 작성
- 활용용도 및 내용 등록



6. 인증키 복사하기



LEARNING SPOONS

Unit 6-2

API를 통한 데이터 받기

LEARNING SPOONS

API 활용하여 데이터 수집하기

- 데이터 요청하기: requests.get(url)
- json 데이터 정리하기
- 엑셀파일에 저장하기

LEARNING SPOONS

Unit6-3
따릉이 지도 만들기

LEARNING SPOONS

따릉이 지도 만들기

- 데이터 불러오기
- 지도 시각화