

## 12회차 수업 데이터 분석 & 시각화 및 고찰

# 파이썬으로 배우는 **따릉이** 데이터 분석과 시각화

이 자료는 Elixirr의 사전 서면 승인 없이 외부에 배포하기 위해  
그 일부를 배포, 인용 또는 복제 할 수 없습니다.

© Copyright Elixirr



# 수업 일정

전체 수업은 13회로 구성된다.



- 따릉이 이용현황 파악
- 문제 정의
- 파이썬 및 사용할 라이브러리 소개



- 비주얼 스튜디오 코드 설치
- 따릉이 데이터 수집



- 파이썬 라이브러리
- 따릉이 데이터프레임 만들기



- 따릉이 데이터프레임 관찰하기



- 시간 개념에 따른 데이터 분석을 위한 컬럼 추가



- 장소적 특징에 따른 데이터 분석을 위한 컬럼 추가



- 시간 개념에 따른 데이터 분석 및 시각화-(1)



- 시간 개념에 따른 데이터 분석 및 시각화-(2)



- 장소 특징에 따른 데이터 분석 및 시각화-(1)



- 장소 특징에 따른 데이터 분석 및 시각화-(2)

# 수업 일정

---

전체 수업은 13회로 구성된다.



- 시간 개념 X 장소 특징에 따른 데이터 분석 및 시각화



- 주말과 평일에 이용건수가 많은 대여소 데이터 분석 및 시각화



- 총정리

1. 문제정의

2. 데이터 수집

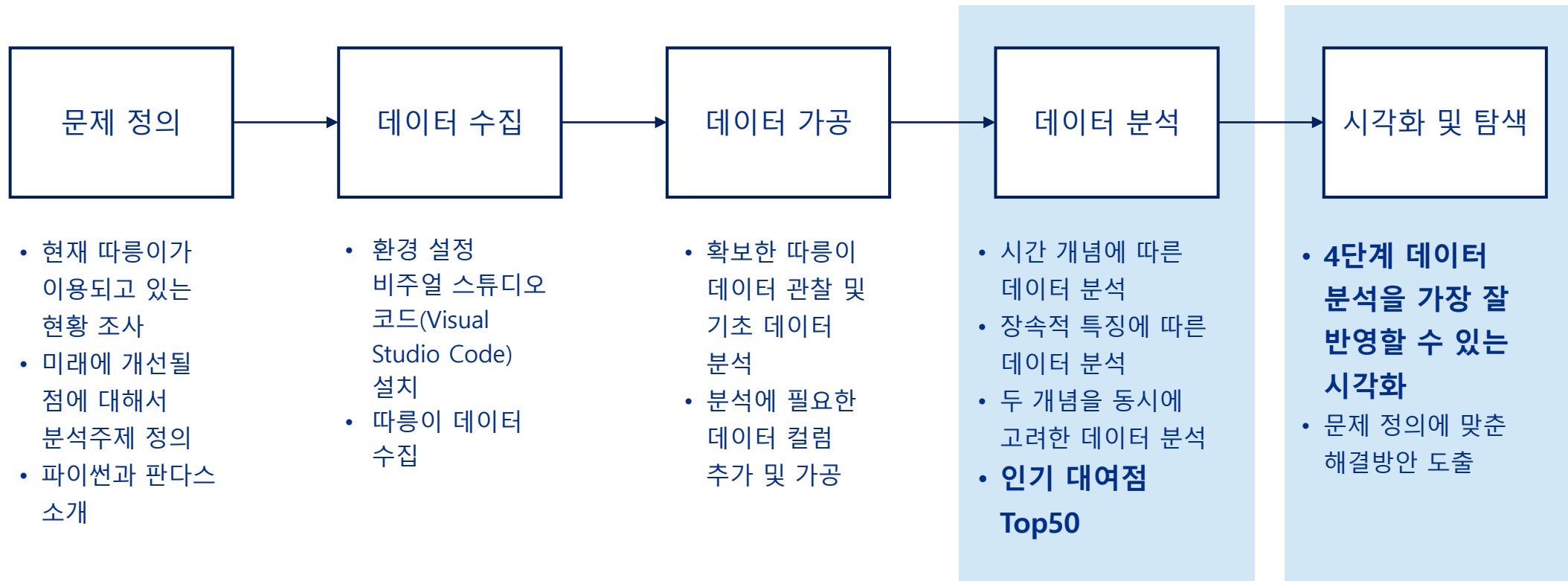
3. 데이터 가공

**4. 데이터 분석**

**5. 시각화 및 탐색**

데이터 분석 단계에 맞추어 따릉이 데이터 분석을 수행한다.

데이터 분석의 5단계





# 여기서 배울 내용은 ?

## 데이터 분석 및 시각화

1.문제정의

2.데이터수집

3.데이터 가공

4.데이터 분석

5.시각화 및 탐색

단계 1 : 시간 개념에 따른 따릉이 이용패턴 분석 및 시각화

단계 2 : 장소적 특징에 따른 따릉이 이용패턴 분석 및 지도 시각화

단계 3 : 시간 개념 x 장소적 특징 연관 분석 후 시각화

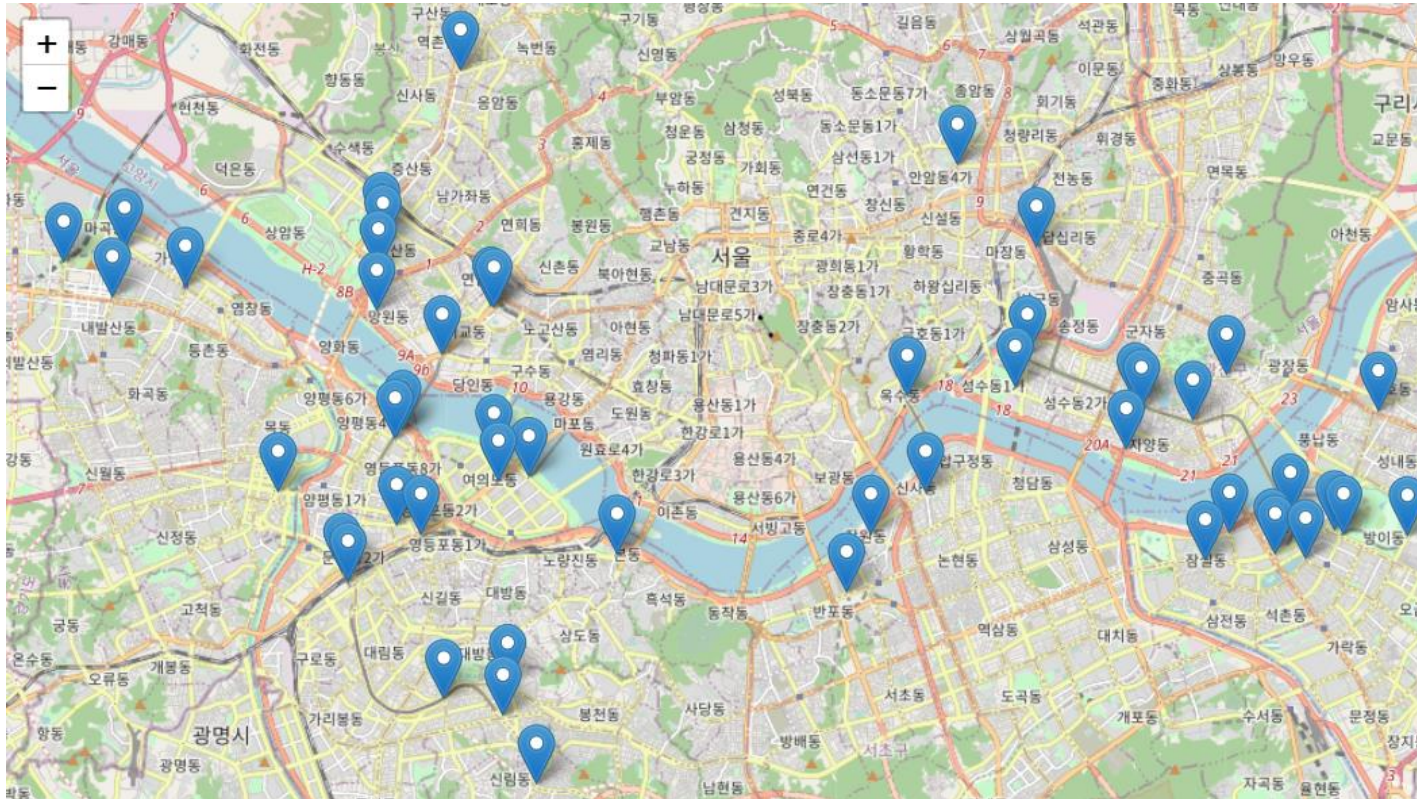
단계 4 : 주말과 평일에 인기 있는 대여소 상위 50개 지도에 표시해보기



# 인기 있는 대여소 TOP50 지도 표시

## 데이터 분석 및 시각화

서울시에서 주말에 인기 있는 TOP50 대여소를 지도에 표시하고 싶다! 무엇부터 알아야 할까?



해리

우선 주말에 인기 있는 대여소 TOP50 을 구해야 겠지?



제니

우리는 대여소의 위도와 경도 데이터가 있으니까 대여소 하나씩 표시할 수 있을 것 같아.



론

50개의 대여소를 표시해야 하니까 똑같은 일을 50번 반복해야 할 것 같아.



우선 주말에 인기 있는 대여소  
TOP50 을 구해야 겠지?

## df.nlargest()

## 데이터 분석 및 시각화

**df.nlargest( n, columns='컬럼명' )**

1

2

3

- 1 **df.nlargest** : 데이터프레임의 특정 컬럼값에서 가장 큰 값을 n번째까지 구하는 명령어
- 2 **n** : 구하고 싶은 개수
- 3 **columns='컬럼명'** : 가장 큰 값을 얻고 싶은 특정 컬럼명

# 대여 대여소 x 주말구분 의 이용건수 계산

```
rent_bike = bikes.pivot_table( \
    index = ['대여 대여소명', '대여점위도', '대여점경도'], \
    columns = '주말구분', \
    values = '자전거번호', \
    aggfunc = 'count')
```

rent\_bike.head()

✓ 0.6s

		주말구분	주말	평일
대여 대여소명	대여점위도	대여점경도		
(구)신한은행 중랑교지점	37.60	127.08	391	832
(구)합정동 주민센터	37.55	126.91	357	715
3호선 매봉역 3번출구앞	37.49	127.05	443	1049
419민주묘지역	37.65	127.01	547	946
8삼호@ 2동 ( 간선도로)	37.49	127.05	111	316



# 인기 있는 주말 대여점 50개를 구하고 지도에 표시한다.

```
rent_bike.nlargest(50, '주말')
```

✓ 0.6s

			주말구분	주말	평일
대여 대여소명	대여점위도	대여점경도			
독섬유원지역 1번출구 앞	37.53	127.07	5777	6840	
여의나루역 1번출구 앞	37.53	126.93	5286	6241	
고속터미널역 8-1번, 8-2번 출구 사이	37.51	127.00	3932	4556	
봉림교 교통섬	37.48	126.93	3699	5986	
롯데월드타워(잠실역2번출구 쪽)	37.51	127.10	3231	5383	
마포구민체육센터 앞	37.56	126.90	3229	3893	
건대입구역 사거리(롯데백화점)	37.54	127.07	3169	4573	
홍대입구역 2번출구 앞	37.56	126.92	2922	5617	



# 주말에 인기 있는 대여소 TOP50

## 데이터 분석 및 시각화

# '주말' 컬럼만 추출한다.

```
rent_bike.nlargest(50, '주말')['주말']
```

✓ 0.6s

대여 대여소명	대여점위도	대여점경도	
독섬유원지역 1번출구 앞	37.53	127.07	5777
여의나루역 1번출구 앞	37.53	126.93	5286
고속터미널역 8-1번, 8-2번 출구 사이	37.51	127.00	3932
봉림교 교통섬	37.48	126.93	3699
롯데월드타워(잠실역 2번출구 쪽)	37.51	127.10	3231
마포구민체육센터 앞	37.56	126.90	3229
건대입구역 사거리(롯데백화점)	37.54	127.07	3169
홍대입구역 2번출구 앞	37.56	126.92	2922
잠실새내역 5번 출구 뒤	37.51	127.09	2632
당산육감문	37.54	126.90	2327
신도림역 1번 출구 앞	37.51	126.89	2304
IFC몰	37.53	126.93	2240

# 주말에 대여량이 많은 대여소 알기

```
주말대여점 50 = rent_bike.nlargest(50, '주말')['주말'].reset_index()  
주말대여점 50
```

✓ 0.4s

	대여 대여소명	대여점위도	대여점경도	주말
0	독섬유원지역 1번출구 앞	37.53	127.07	5777
1	여의나루역 1번출구 앞	37.53	126.93	5286
2	고속터미널역 8-1번, 8-2번 출구 사이	37.51	127.00	3932
3	봉림교 교통섬	37.48	126.93	3699
4	롯데월드타워(잠실역 2번출구 쪽)	37.51	127.10	3231
5	마포구민체육센터 앞	37.56	126.90	3229
6	건대입구역 사거리(롯데백화점)	37.54	127.07	3169
7	홍대입구역 2번출구 앞	37.56	126.92	2922
8	잠실새내역 5번 출구 뒤	37.51	127.09	2632

- rent\_bike 데이터프레임에서 '주말' 컬럼을 추출
- 결과 : 시리즈
- 인덱스 : 대여 대여소명, 대여점위도, 대여점경도
- 값(value) : 따릉이 이용건수

- reset\_index()를 하면 시리즈 -> 데이터프레임
- 인덱스 : 정수 인덱스
- 대여 대여소명, 대여점위도, 대여점경도 -> 컬럼
- 값(value) -> 주말컬럼



우리는 대여소의 위도와 경도 데이터가 있으니  
대여소 하나씩 표시할 수 있을 것 같아.

## folium.Marker()

## 데이터 분석 및 시각화

**folium.Marker( location=[위도, 경도], popup='이름', icon = folium.Icon() ).add\_to(map)**

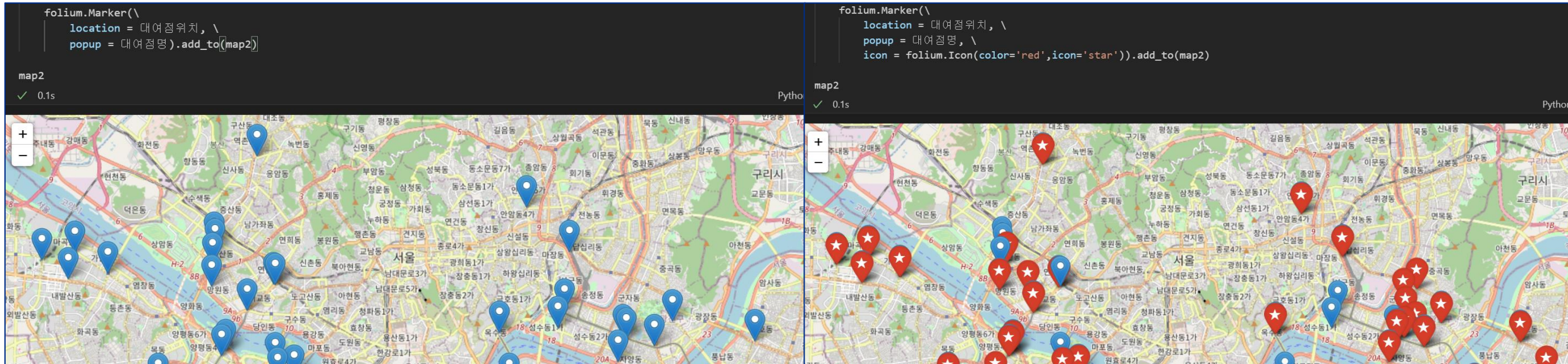
1

2

3

4

- 1 **folium.Marker** : 지도에서 특정 위치를 아이콘으로 표시
- 2 **location=[위도, 경도]** : 아이콘이 표시될 위치
- 3 **popup='이름'** : 아이콘으로 표시한 곳의 이름으로 아이콘을 클릭하면 이름이 나온다.
- 4 **icon = folium.Icon(color='red', icon='star')** : 지정하지 않으면 기본적으로 파란색 info 모양의 아이콘으로 표시하고 folium.Icon을 사용하여 다른 것을 지정할 수 있다.





50개의 대여소를 표시해야 하니까 똑같은 일을 50번 반복해야 할 것 같아.

## 파이썬 문법 : for 문 - (1)

for 문은 실행구문들을 목록안의 항목 수 만큼 반복하는 제어구조다.

# 목록의 값을 차례로 변수로 받아 실행한다.

```
for 변수 in 순회가능한 자료구조 또는 목록 :  
    실행구문1  
    실행구문2  
    ...
```

```
# 3.5.2.3.1 for문 연습1  
# 숫자 1,2,3 목록에서 차례로 a 의 값이 변하여 전달 되고 그 값이 출력된다.  
for a in [1,2,3] :  
    print (a)  
    print (a*a)
```

```
1  
1  
2  
4  
3  
9
```

1. 키워드 **in** 뒤에 기재된 자료 구조 또는 목록으로부터 항목을 처음부터 마지막까지 하나씩 꺼내어 변수에 넣고, for 블록의 실행구문들을 순서대로 실행한다.
2. 변수에 들어갈 값이 없게 되면 for 블록은 더 이상 실행되지 않고 빠져 나온다.





50개의 대여소를 표시해야 하니까 똑같은 일을 50번 반복해야 할 것 같아.

## 파이썬 문법 : for 문 - (1)

### 데이터 분석 및 시각화

```
for i in 주말대여점50.index:
```

```
    sub_lat = 주말대여점50.loc[i, '대여점위도']  
    sub_lon = 주말대여점50.loc[i, '대여점경도']  
    대여점위치 = [sub_lat, sub_lon]  
    대여점명 = 주말대여점50.loc[i, '대여소명']
```

#지도에 데이터 찍어서 보여주기

```
folium.Marker(  
    location = 대여점위치, \  
    popup = 대여점명).add_to(map2)
```

map2

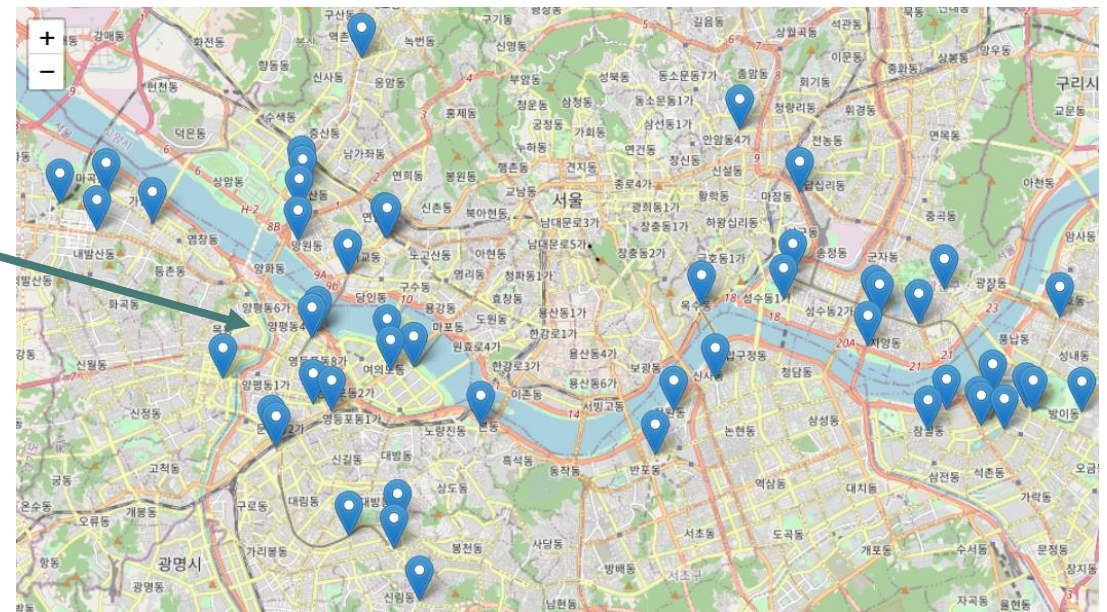
리스트의 이름 = [항목1, 항목2, ...]

예) 대여점위치 = [sub\_lat, sub\_lon]

주말대여점50

✓ 0.4s

	대여소명	대여점위도	대여점경도	주말
0	독성유원지역 1번출구 앞	37.53	127.07	5777
1	여의나루역 1번출구 앞	37.53	126.93	5286
2	고속터미널역 8-1번, 8-2번 출구 사이	37.51	127.00	3932
3	봉림교 교통섬	37.48	126.93	3699
4	롯데월드타워(잠실역2번출구 쪽)	37.51	127.10	3231





50개의 대여소를 표시해야 하니까 똑같은 일을 50번 반복해야 할 것 같아.

## 행 단위 데이터 추출

## 데이터 분석 및 시각화

loc 명령어 및 대괄호 [ ]를 사용해서 행 단위의 데이터를 추출합니다.

```
for i in 주말대여점50.index:
```

```
sub_lat = 주말대여점50.loc[i, '대여점위도']
```

```
sub_lon = 주말대여점50.loc[i, '대여점경도']
```

```
대여점위치 = [sub_lat, sub_lon]
```

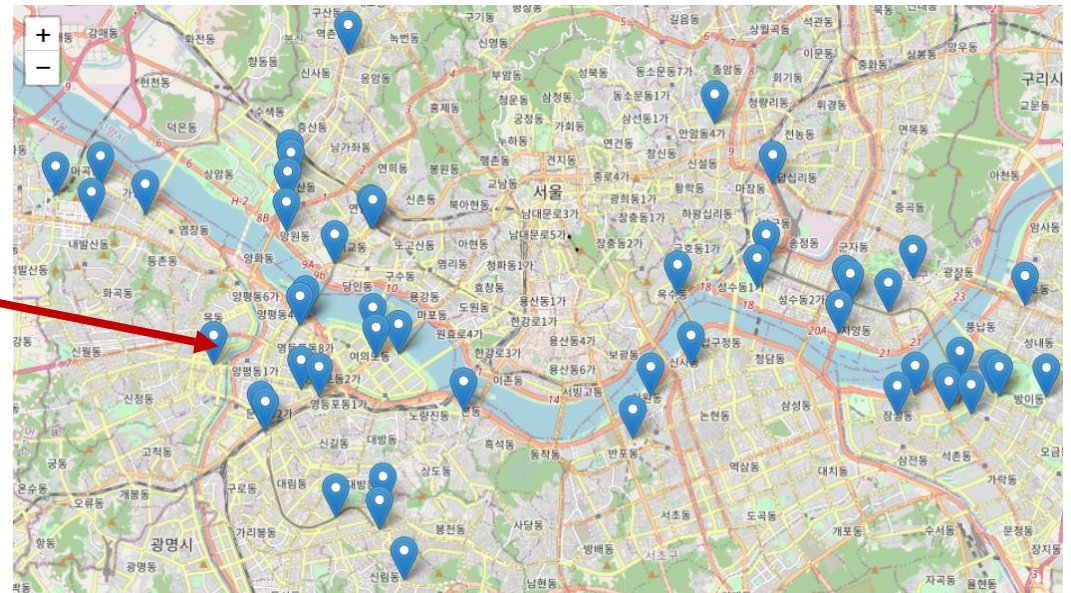
```
대여점명 = 주말대여점50.loc[i, '대여 대여소명']
```

```
#지도에 데이터 찍어서 보여주기
```

```
folium.Marker(\n    location = 대여점위치, \n    popup = 대여점명).add_to(map2)
```

map2

주말대여점 50				
✓ 0.4s				
	대여 대여소명	대여점위도	대여점경도	주말
0	독섬유원지역 1번출구앞	37.53	127.07	5777
1	여의나루역 1번출구 앞	37.53	126.93	5286
2	고속터미널역 8-1번, 8-2번 출구 사이	37.51	127.00	3932



- loc 명령어와 대괄호를 사용하여 인덱스와 컬럼명을 전달한다.
- 인덱스에 해당하는 전체 행과 그 행 중에서 필요한 컬럼의 값을 가져온다.



# 나 지금 어느 단계를 공부하는 거지?

데이터 분석 및 시각화

1.문제정의

2.데이터수집

3.데이터 가공

4.데이터 분석

5.시각화 및 탐색

**단계 4 :** 시간 개념 x 장소적 특징 연관 분석 후 시각화

데이터 집계 -> `bikes.nlargest`

지도 시각화 : `folium.Marker`

반복 피하기 : `for문`



퀴즈를  
풀어봅시다

1. folium 라이브러리에 있는 명령어로 지도에 특정 위치를 아이콘으로 표시해주는 명령어는?

2. 실행구문들을 목록안의 항목 수 만큼 반복하는 제어구조는?

3. 데이터프레임에서 행을 추출하는 연산자는?

4. 주말대여점50에서 인덱스를 알려주는 명령어는 ?



인기 있는 대여소 TOP 50

주말에 인기 있는 대여소 TOP 50

평일에 인기 있는 대여소 TOP 50



GD쌤

이제부터 Visual Studio Code 실습 환경에서 지금까지 배운 내용을 실습해 보겠습니다.

앞에서 배웠던 내용을 Visual Studio Code에서 직접 실습해보면 더욱 이해하기 편리할 것입니다.

### 시간 개념에 따른 데이터 분석

일자별 따릉이 이용건수, 요일별 따릉이 이용건수  
대여시간대별 따릉이 이용건수  
대여시간대 x 요일 따릉이 이용건수  
대여시간대 x 주말구분 따릉이 이용건수

### 장소 특징에 따른 데이터 분석

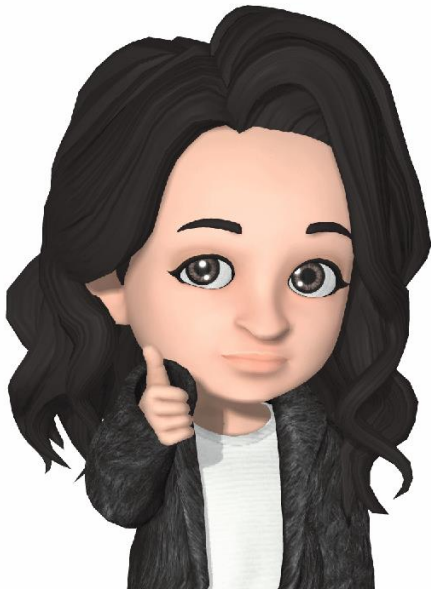
대여구 별 이용건수 분석  
대여구 별 이용시간 분석

### 시간 X 장소 데이터 분석

일자 x 대여구 데이터분석  
대여시간대 x 대여구 데이터분석  
인기 있는 대여소

## 수업 마무리

---



GD쌤

지금까지 12회차 수업내용을 배워 보았습니다.

다음 시간에는 13회차 수업내용으로 지금까지 수행한 데이터 분석을 총정리해 보겠습니다.

수고 많으셨어요. 다음 시간에 만나요.