

파이썬으로 배우는 **따릉이** 데이터 분석과 시각화

10회차 시각화 및 고찰

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



- 문제 정의에 맞춘 해결방안 도출
- 총정리

1. 문제정의

2. 데이터 수집

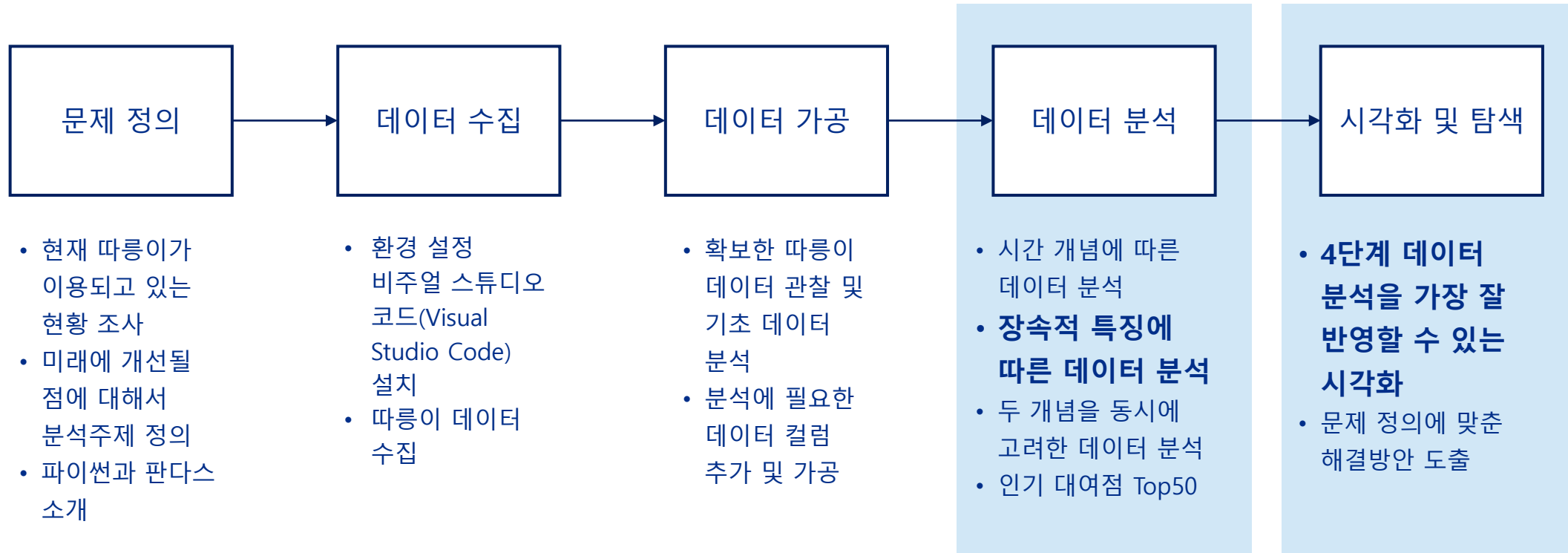
3. 데이터 가공

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 3 : 주말과 평일에 인기 있는 대여소 상위 50개 지도에 표시해보기

구별이용시간평균을 피벗테이블을 이용해서 만들고 데이터프레임으로 변환한다.

```
# 대여구별 이용시간 평균을 구한다.  
  
구별이용시간평균 = bikes.pivot_table(\  
    index = '대여구', \  
    values = '이용시간', \  
    aggfunc = 'mean')  
구별이용시간평균  
✓ 0.2s
```

	이용시간
대여구	
강남구	31.79
강동구	25.39
강북구	23.57
강서구	23.14
관악구	26.51
광진구	24.95
구로구	26.48
금천구	26.40

인덱스 → 컬럼

```
구별이용시간평균.reset_index(inplace=True)  
구별이용시간평균  
✓ 0.3s
```

	대여구	이용시간
0	강남구	31.79
1	강동구	25.39
2	강북구	23.57
3	강서구	23.14
4	관악구	26.51
5	광진구	24.95
6	구로구	26.48
7	금천구	26.40
8	노원구	24.26
9	도봉구	24.99
10	동대문구	23.78
11	동작구	31.66

구별이용시간평균

sort_values()를 사용하여 이용시간 컬럼을 기준으로 크기 순서로 나열한후 막대그래프로 시각화한다.

```
구별이용시간평균.sort_values(by='이용시간', \
                               ascending=True, \
                               inplace=True)
```

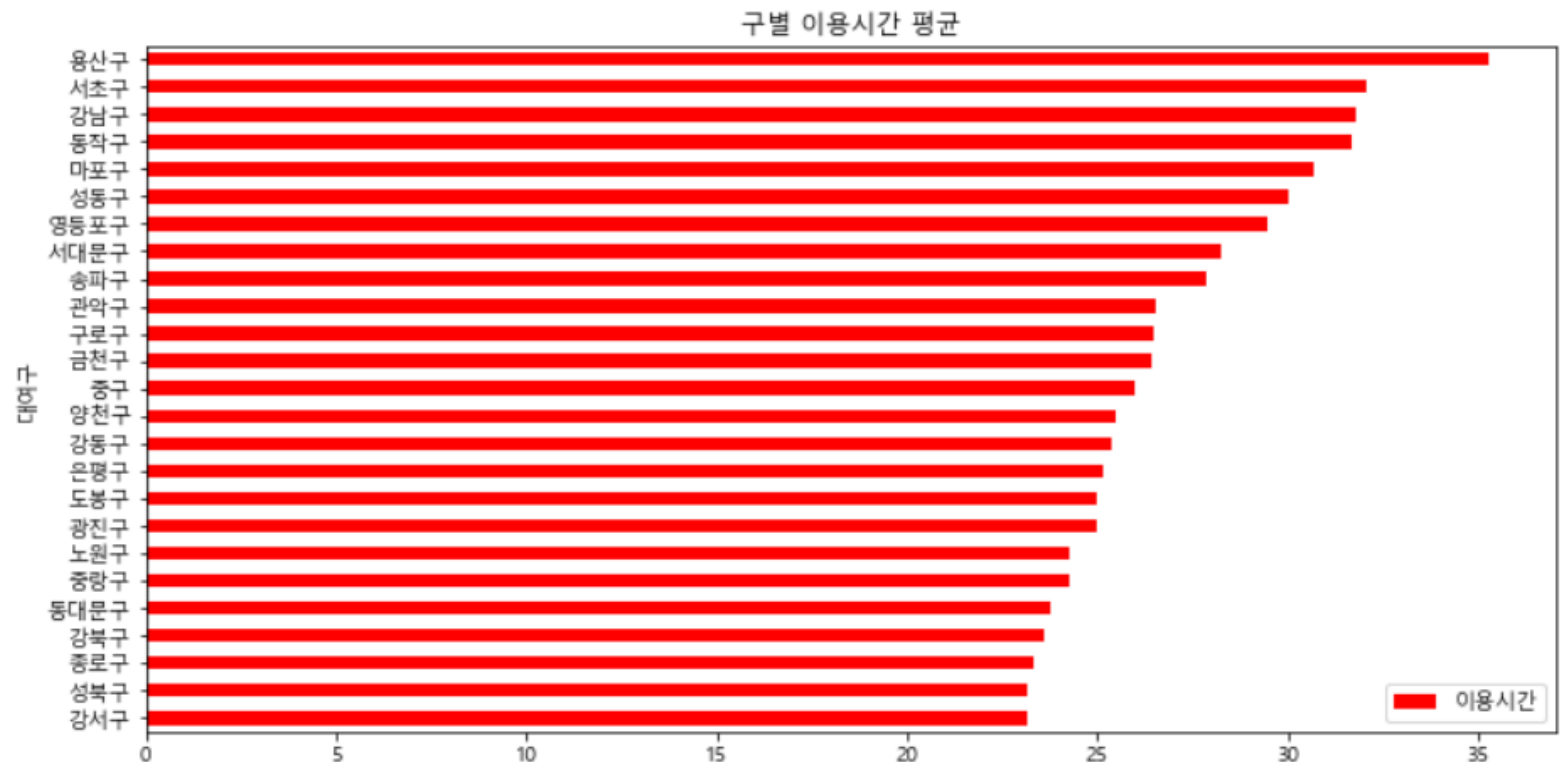
구별이용시간평균

✓ 0.4s

	대여구	이용시간
3	강서구	23.14
16	성북구	23.17
22	종로구	23.32
2	강북구	23.57
10	동대문구	23.78
24	중랑구	24.24
8	노원구	24.26
5	광진구	24.95
9	도봉구	24.99
21	은평구	25.16

```
# 막대그래프로 표시하기
```

```
구별이용시간평균.plot(kind='barh', title='구별 이용시간 평균', figsize=(12, 6), color='r');
```



choropleth() 명령어는 판다스의 데이터프레임에 있는 지리 데이터와 그에 해당하는 수치 데이터분포를 지도에 색깔로 표시해 준다.

```
import json

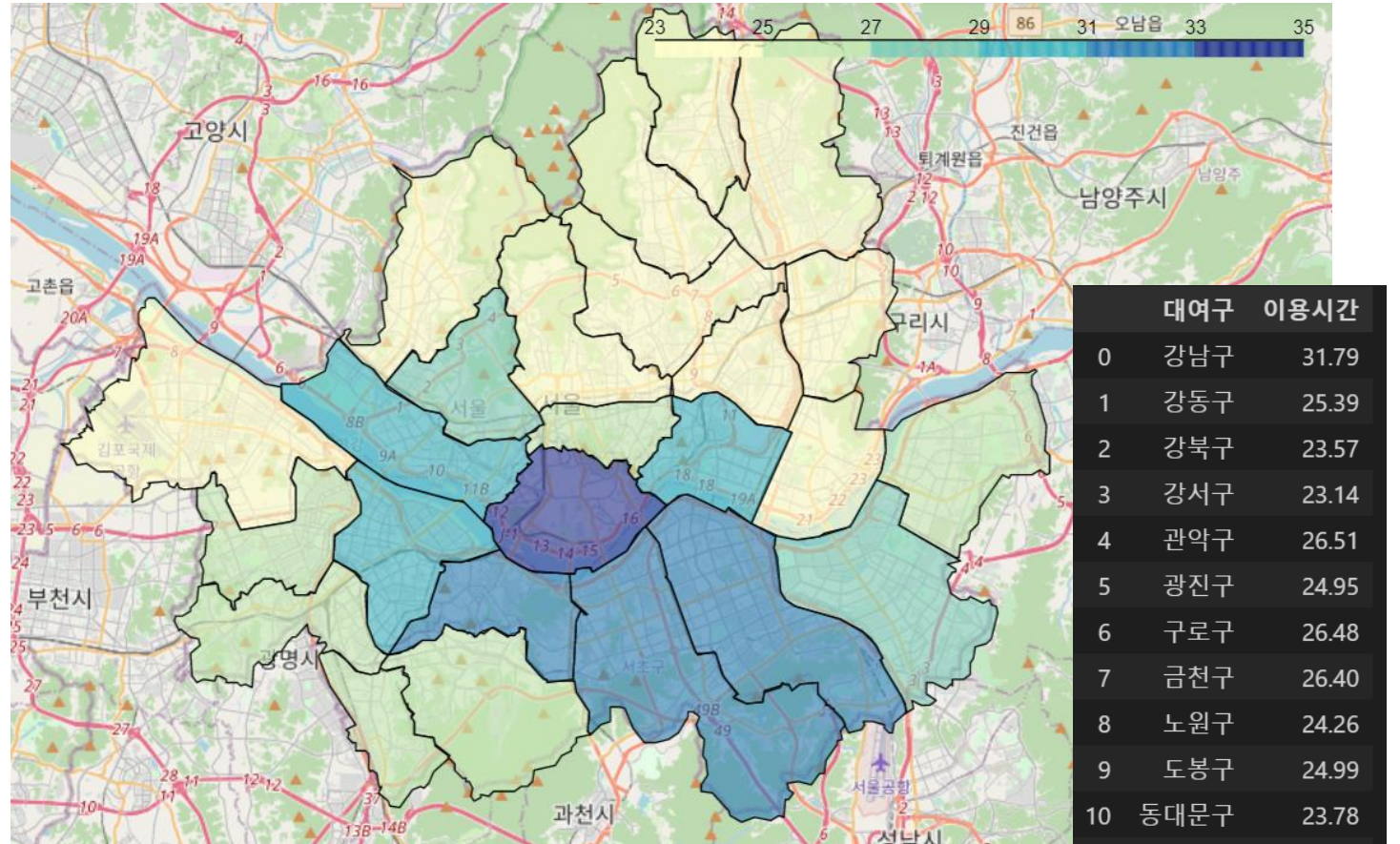
with open('./data/seoul.json', 'r', encoding='utf-8') as f:
    서울시지도 = json.load(f)

서울시지도

# 지도의 중심위치를 정한다.
lat = bikes['대여점위도'].mean()
lon = bikes['대여점경도'].mean()
center = [lat, lon]
map1 = folium.Map(location = center, zoom_start = 11)
map1

folium.Choropleth(
    geo_data = 서울시지도,
    data = 구별이용시간평균,
    columns = ['대여구', '이용시간'],
    fill_color = 'YlGnBu',
    key_on = 'feature.properties.name').add_to(map1)

map1
```





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 2 : 분석 명령어를 사용해서 장소적 특징에 따른 따릉이
이용패턴 분석 및 지도 시각화

데이터 집계 -> `df.pivot_table()`

인덱스 재설정 -> `df.reset_index()`

특정 컬럼 기준으로 정렬 및 시각화 -> `df.sort_values()`

구단위 지도 시각화 -> `folium.Map()`, `folium.choropleth()`



퀴즈를
풀어봅시다

1. 중심위치와 배율을 조정해서 지도를 초기화해서 보여주는 명령어는 ?

2. 판다스의 데이터프레임에 있는 지역 데이터와 그에 해당하는 수치 데이터분포를 지도에 색깔로 표시해 주는 명령어는 ?

3. 파이썬의 딕셔너리 스타일로 데이터를 표현한 파일로 키-값으로 이루어져 있으며 사람이 쉽게 읽을 수 있도록 구성되어 있는 파일은 ?

4. 데이터프레임의 특정 컬럼의 데이터 값을 기준으로 크기 순서로 나열하고 오름차순 (ascending)을 기본으로 적용하는 명령어는 ?

장소 특징에 따른 데이터 분석

대여구 별 이용건수 분석

대여구 별 이용시간 분석

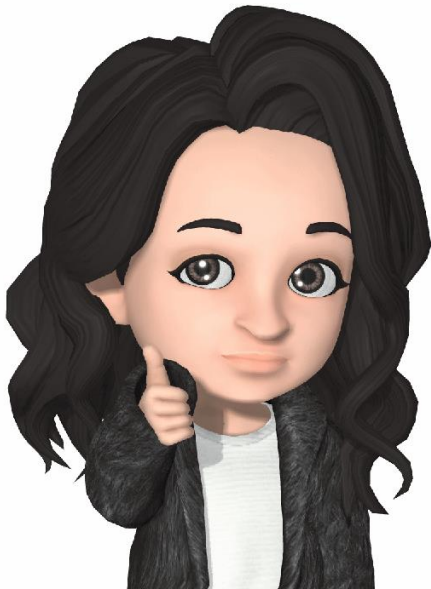


GD쌤

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

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

수업 마무리



GD쌤

지금까지 10회차 수업내용을 배워 보았습니다.

다음 시간에는 11회차 수업내용으로 시간 개념과 장소 특징을 연관해서
데이터 분석 및 시각화를 진행해 보겠습니다.

수고 많으셨어요. 다음 시간에 만나요.