

파이썬으로 배우는 **따릉이** 데이터 분석과 시각화

13회차 수업 시각화 및 탐색

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

© Copyright Elixirr

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

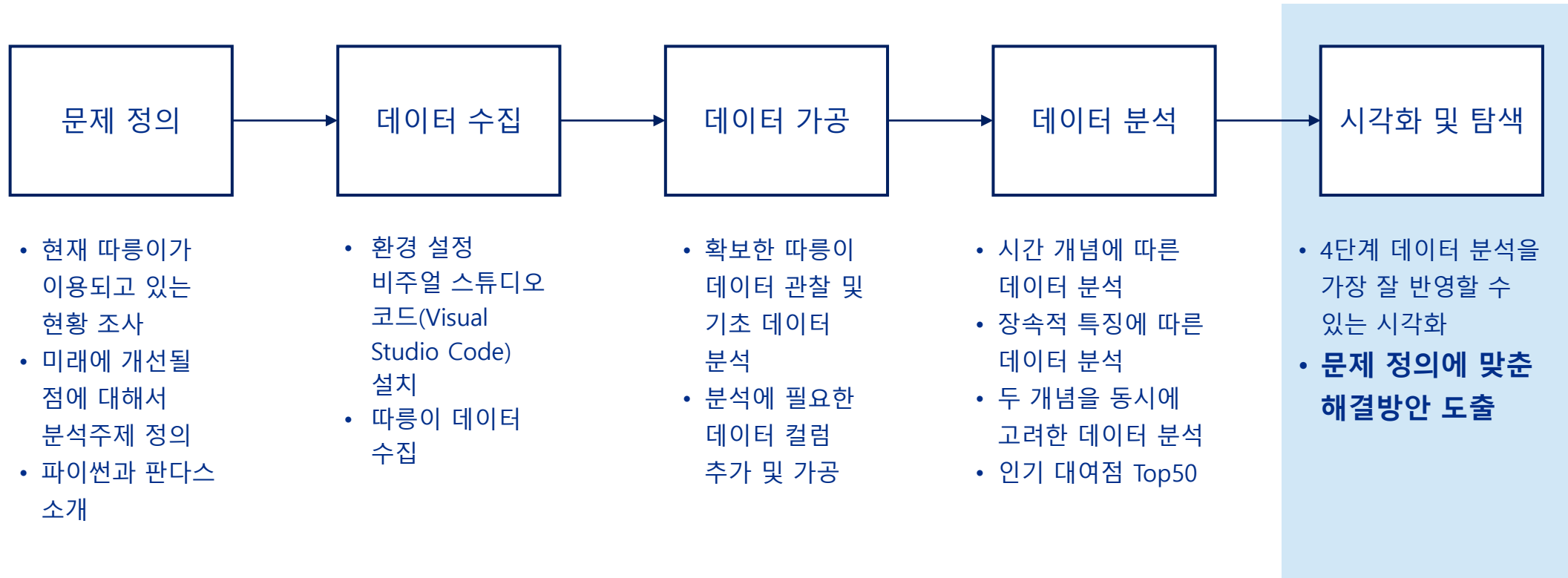
3. 데이터 가공

4. 데이터 분석

5. 시각화 및 탐색

데이터 분석 단계에 맞추어 따릉이 데이터 분석을 수행한다.

데이터 분석의 5단계





여기서 배울 내용은 ?

시각화 및 탐색

1.문제정의

2.데이터수집

3.데이터 가공

4.데이터 분석

5.시각화 및 탐색

문제 정의에 맞춘 해결방안 도출

지금까지 수행한 데이터 분석 총정리

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

인기 있는 대여소 TOP50

주말에 인기 있는 대여소 TOP50
평일에 인기 있는 대여소 TOP50

주제1. 시간 개념에 따른 이용 패턴

1.문제 정의

미디어에 알려진 따름이 이용 패턴에 대해서 궁금한 것이 생겼어요. 언제 많이 타는지(이용건수)와 한 번 대여 하면 어느 정도의 시간을 대여하는지(이용시간)에 대한 분석이 필요해 보인다.

출퇴근 목적으로 사용하는
것이 가장 많고 이용시간도
20분 이내가 가장 많네.



해리

그럼 출퇴근용이 아닌
경우에는 더 오래 탈까?



제니

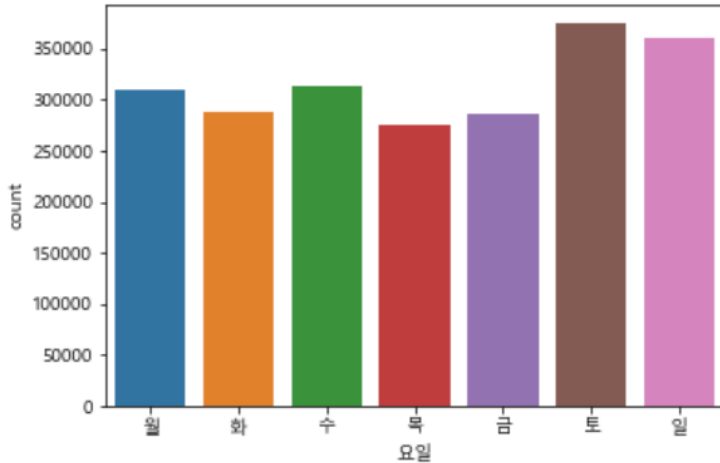
주말과 평일에 대여 건수의
차이가 있을까?



론

시간 개념에 따른 데이터 분석 - (1)

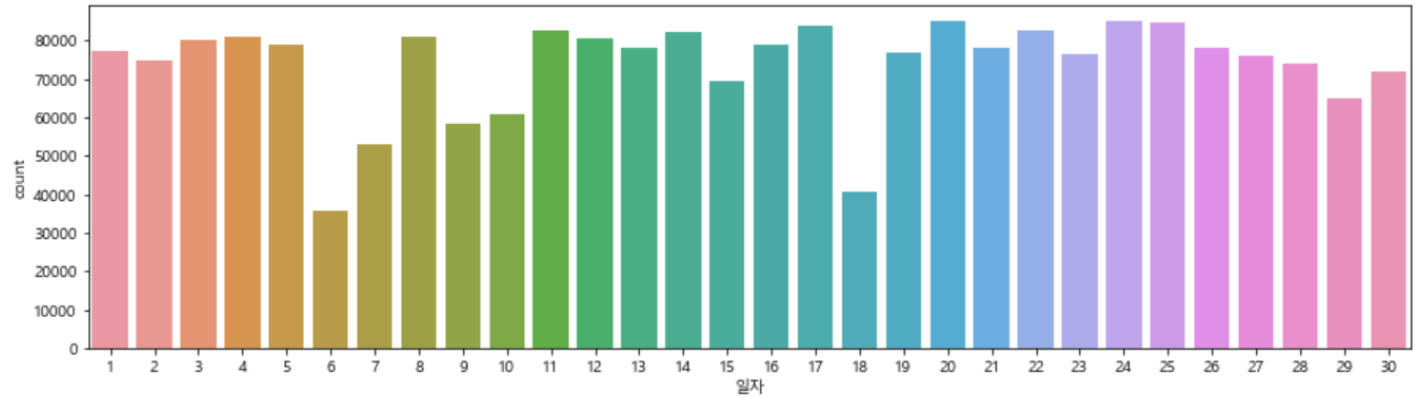
시각화 및 탐색



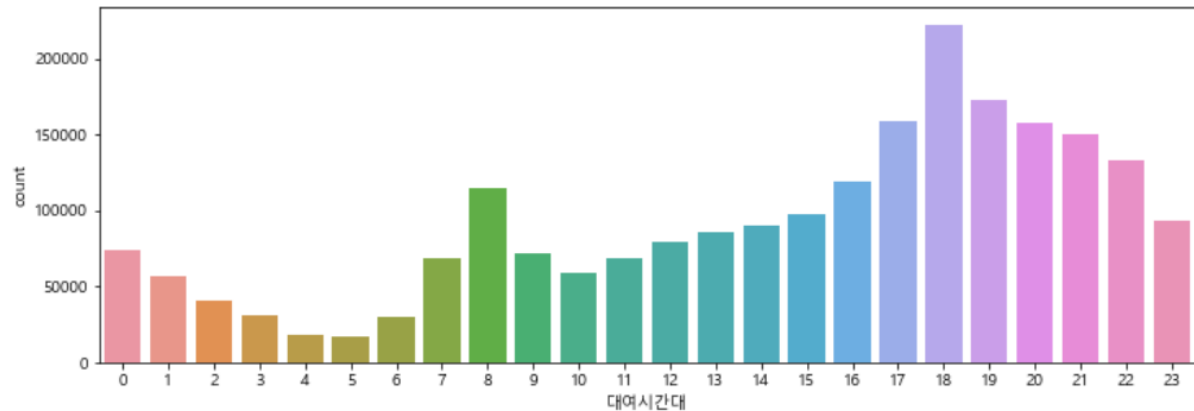
- 평일보다는 주말에 따릉이 이용건수가 많고 주말 중에서는 토요일에 가장 이용건수가 많다.
- 주말에 인기 있는 대여소 근처에 대여소를 추가로 설치하거나 따릉이를 추가 배치할 필요가 있다.



데이터 분석을 하고 시각화
하니까 해결방안을 생각하기
편하네



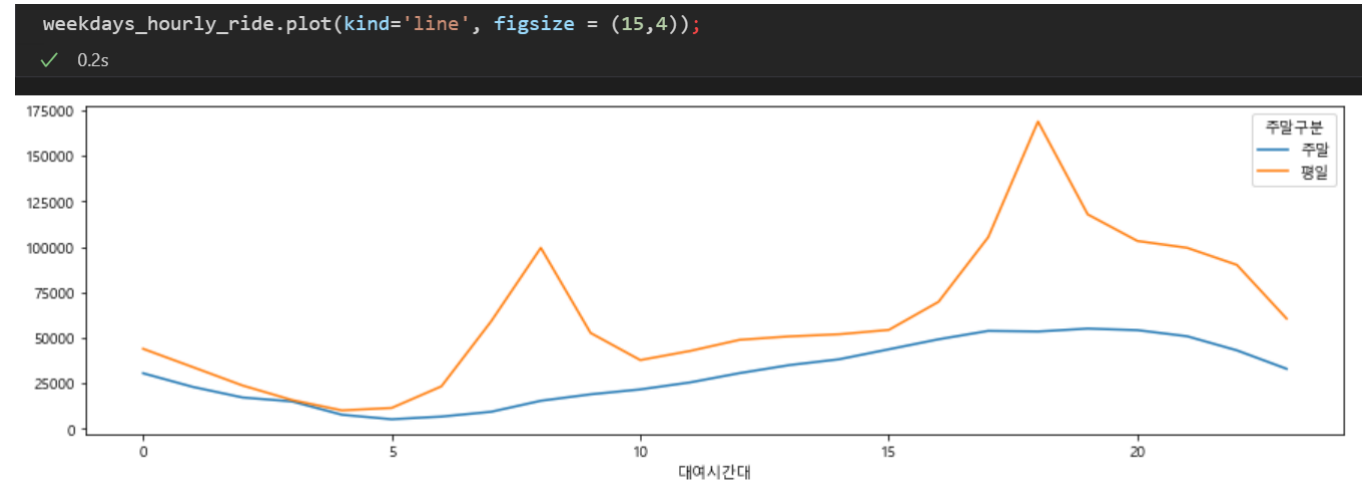
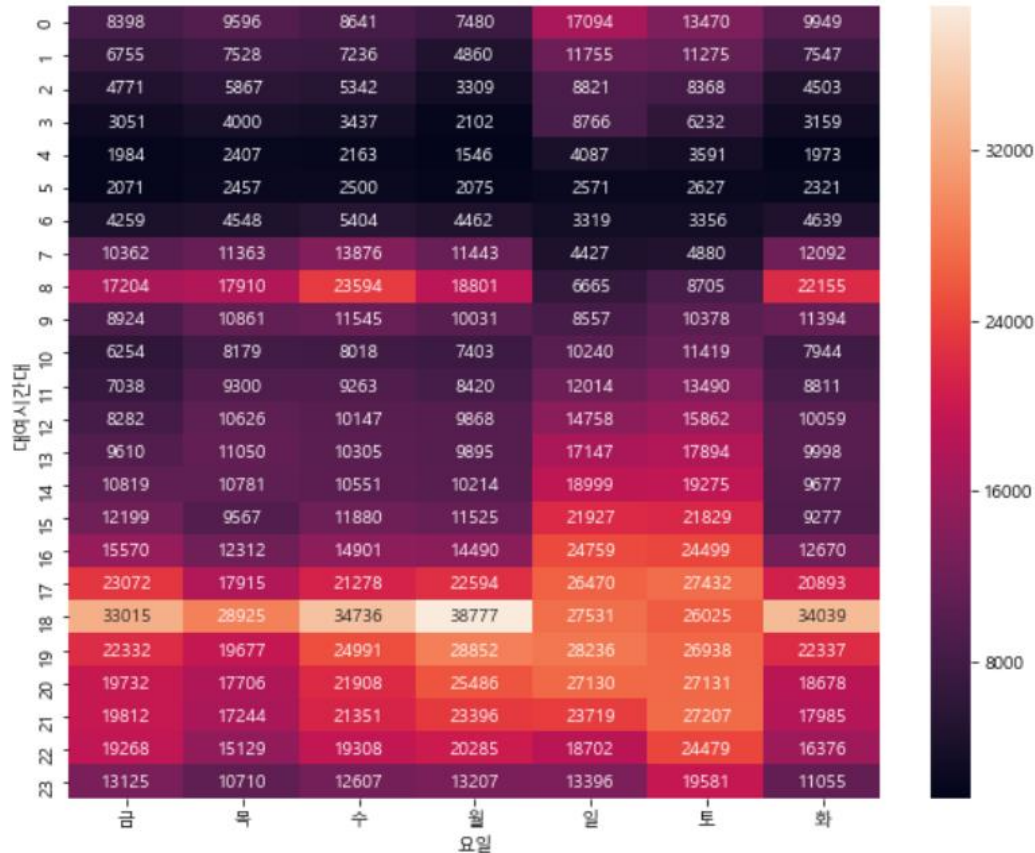
- 6일이 현충일이지만 비가 와서 이용건수가 적다. 강수량에 영향을 많이 받는다.
- 일회용 우비 등을 비치해서 비 오는 날도 따릉이 이용에 불편이 없도록 한다.



- 출퇴근 시간대를 중심으로 이용건수가 많고 오전<오후<저녁 순으로 이용건수가 증가한다.
- 출퇴근 시간대에 많이 이용되는 대여소에 따릉이를 추가 배치하자.

시간 개념에 따른 데이터 분석 - (2)

시각화 및 탐색



주말과 평일에 이용패턴이 확연히 차이가 난다. 주말과 평일에 인기 있는 대여소를 알아볼 필요가 있다.

두 그래프 모두 주말과 평일에 이용패턴이 다르네. 하지만 히트맵은 시간대에 따른 정확한 수치까지 제공해 주네.



주말과 평일에 이용패턴이 다르고 평일중에서는 월요일 퇴근 시간에 따릉이 이용이 많고 목요일이 상대적으로 적다.
-> 평일 출퇴근 시간대에 따릉이를 이용하는 사람들에게 포인트 적립 등의 인센티브를 제공해서 출퇴근시의 교통적체를 줄여보자.

주제2. 장소적 특징에 따른 이용패턴

1.문제 정의

대여소가 위치한 **장소적 특징**에 따라서 사용자의 이용패턴이 달라진다.

우리 집 앞은 언덕이 심해서 자전거를 타기가 힘들어.



해리

지역구區 별로 자전거를 이용하기 쉬운 장소가 인기가 많지 않을까?



제니

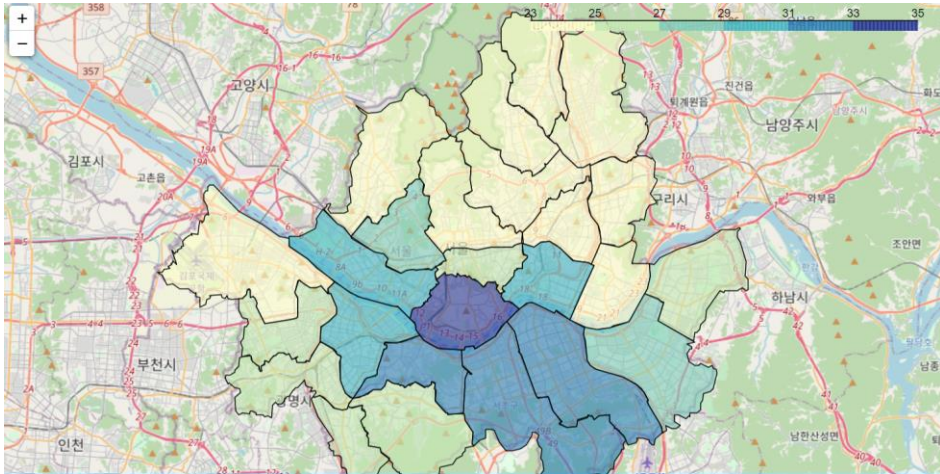
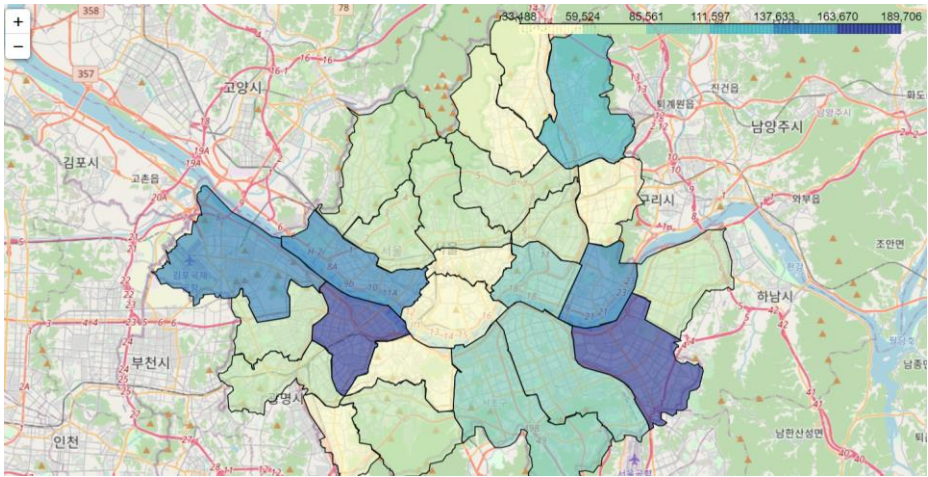
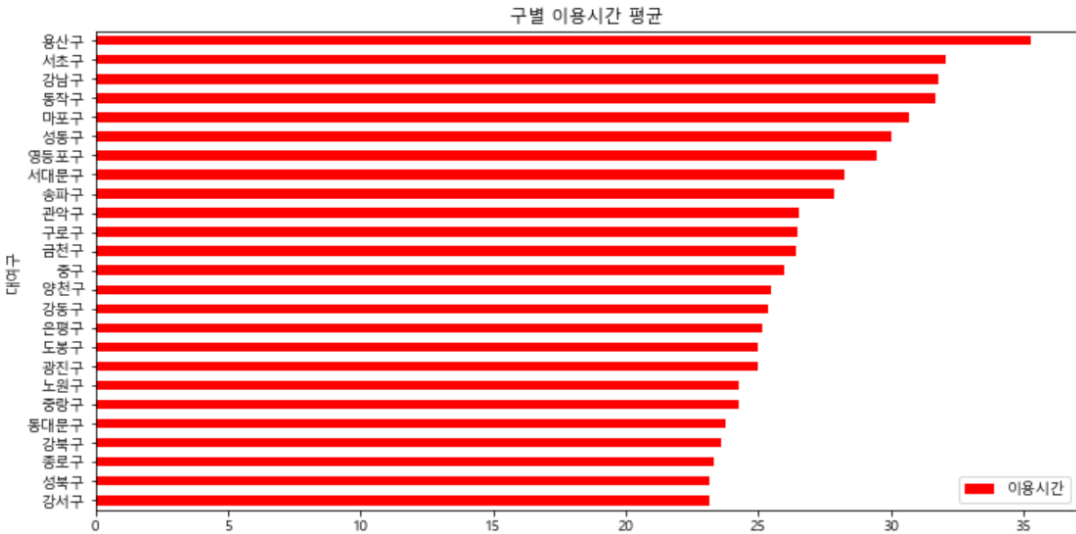
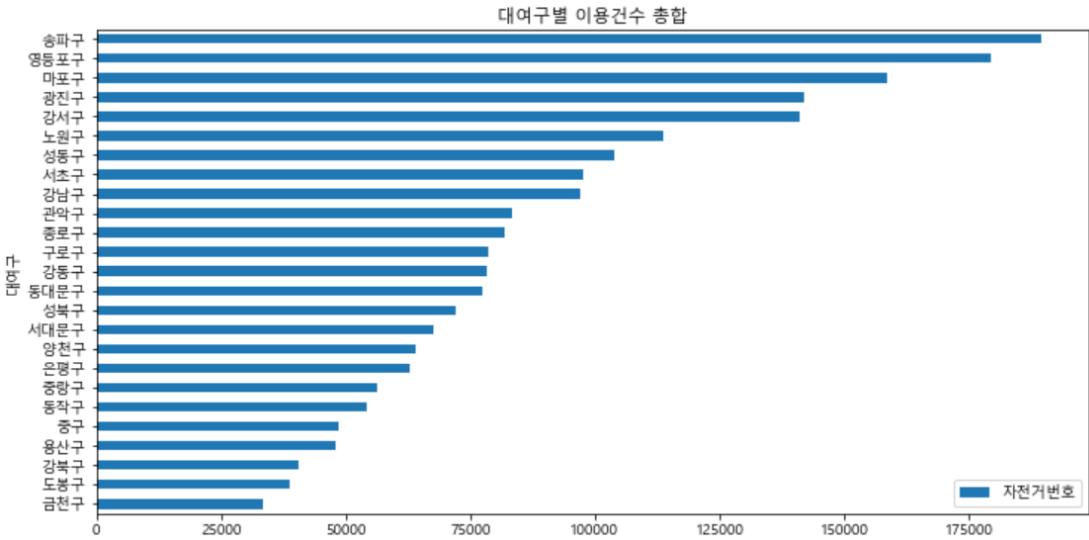
자전거 전용도로가 있는 곳이 사람들이 많이 타지 않을까?



론

장소적 특징에 따른 데이터 분석

시각화 및 탐색



- 위의 그래프는 이용건수를 나타내고
아래 그래프는 이용시간을 나타낸다.
- 이용건수가 많은 지역은 인구밀도가 높은 지역으로 교통정체가 심한 지역이다.
- 이용시간이 많은 지역은 한강이나 공원을 끼고 있거나 자전거 도로가 잘 되어 있는 지역구이다.
- **이용이 저조한 지역구에 자전거 도로를 확충하고 자전거를 타기 좋은 공원이나 강변을 개발해서 따릉이 이용을 유도해보자.**

주제3. 시간 개념 X 장소적 특징에 따른 이용패턴

1.문제 정의

시간 개념과 대여소가 위치한 장소적 특징을 모두 고려해서 사용자의 이용패턴을 분석해 본다.

서울시의 모든 지역구별로
따릉이 이용시간의 평균을
계산해서 가장 많이 사용하는
지역구를 알아볼까?



해리

6월 한달 일자에 따라 또는
하루 24시간 시간대에 따라
지역구별로 따릉이
이용건수를 세어볼까?



제니

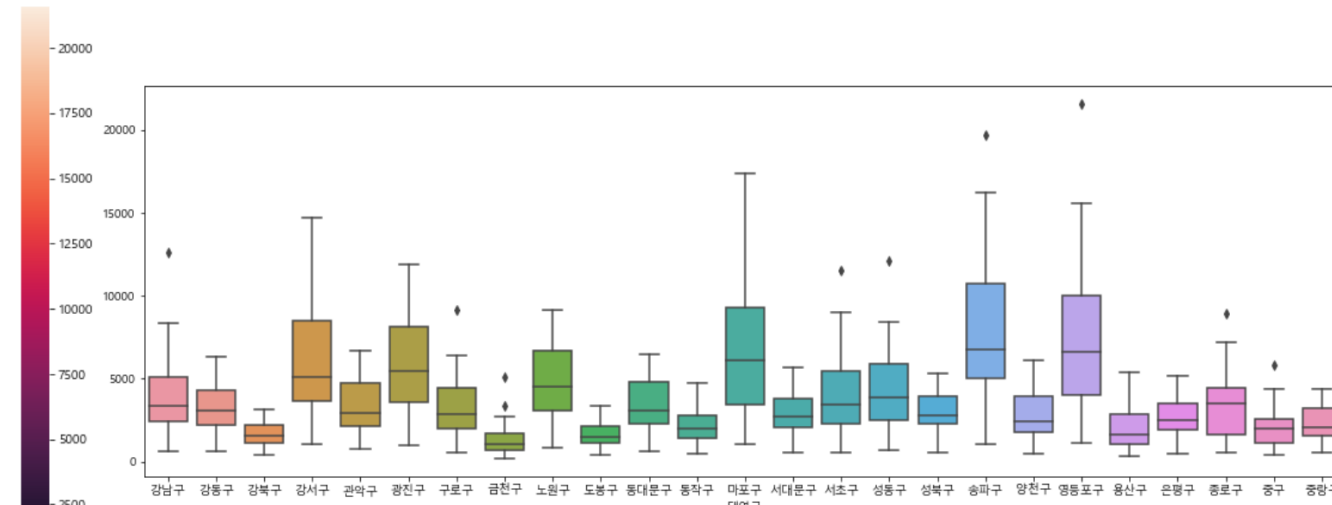
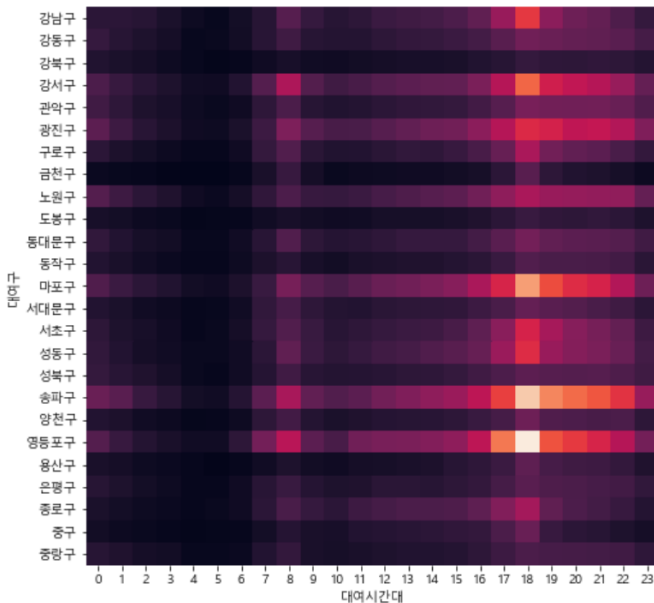
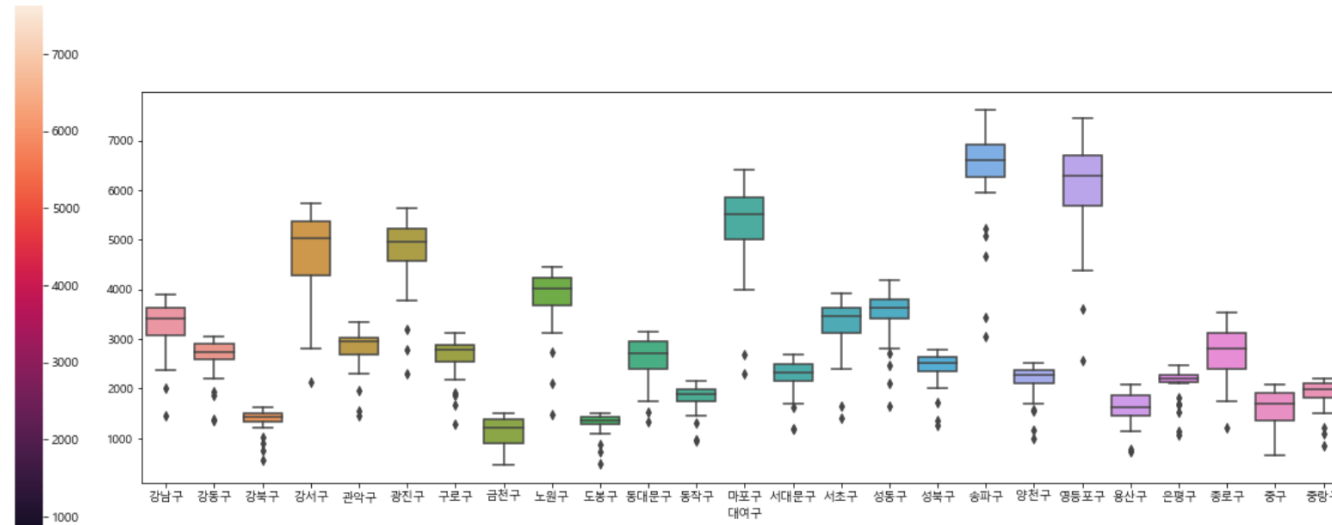
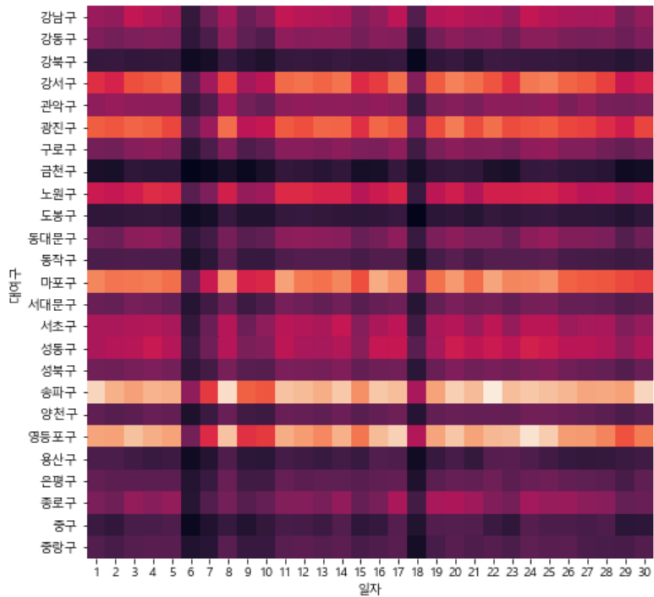
주말과 평일에 각각 인기있는
대여점 50개를 한번 알아보자.



론

시간적 개념 x 장소적 특징 연관 데이터 분석

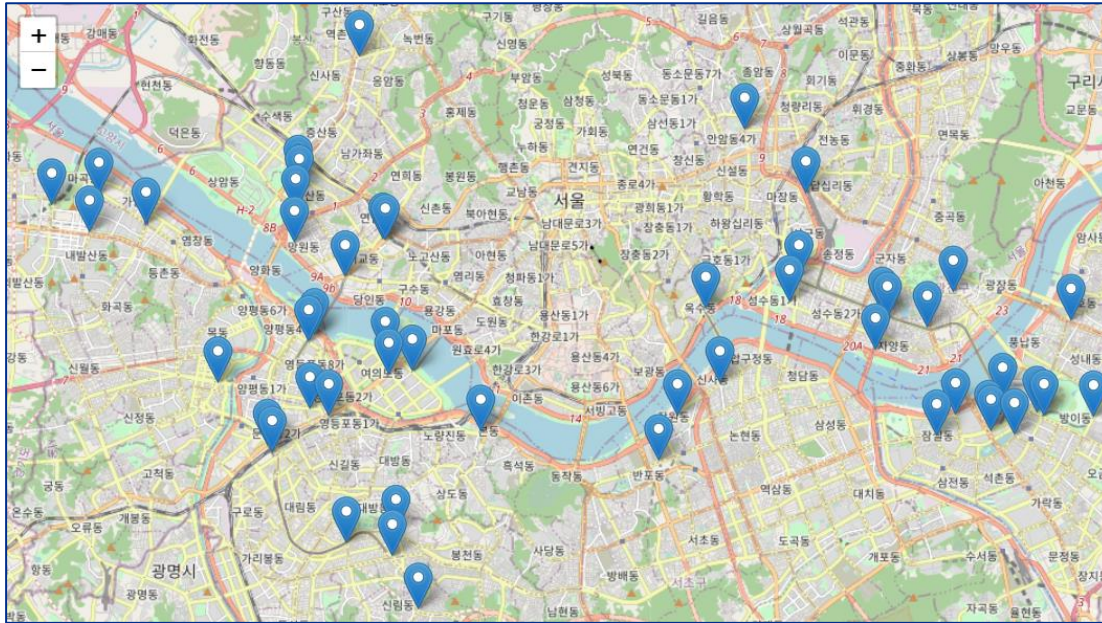
시각화 및 탐색



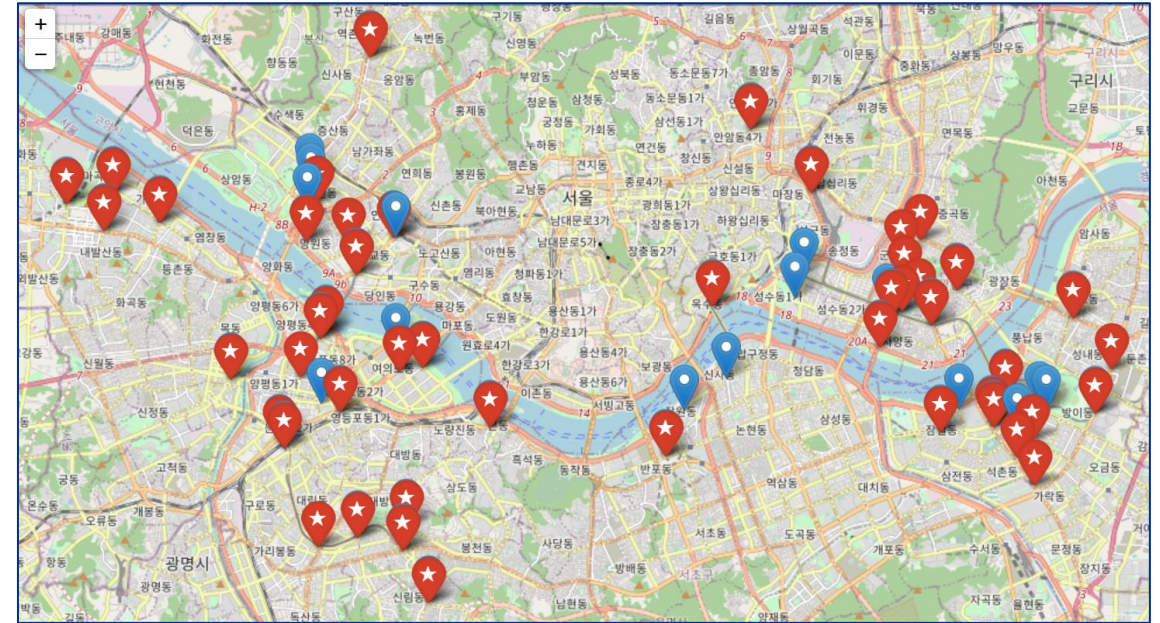
- 위의 그래프는 일자 x 대여구별 이용건수를 나타낸 것으로 송파구, 영등포구, 마포구 순으로 높다.
- 아래 그래프는 시간대 x 대여구별 이용건수를 나타낸 것으로 많은 구가 오후 6시 이후에 이용건수가 높으나 강북구, 금천구, 도봉구 등은 저녁시간 이용이 저조하다.
- 산을 끼고 있는 지역이 저녁시간 이용이 저조한데 산 초입을 강변처럼 조명을 밝게 하고 자전거를 탈 수 있는 도로 조성을 해서 따릉이 이용을 높여보자.

인기 있는 대여소 TOP50 지도 표시

시각화 및 탐색



- 주말에 인기 있는 대여소 TOP50을 표시한 것으로 주로 한강변, 호수나 공원 근처이다.
- 가족 동반 나들이로 이용하는 경우에 대비하여 아동용 따릉이나 안전용 헬멧 등을 비치하자.
- 따릉이를 이용하는 사람들이 많이 모이게 되니 안전 수칙을 알리는 표지판, 물을 마실 수 있는 개수대, 쓰레기통 비치 등에 신경 쓰자.



- 평일에 인기 있는 대여소 TOP50을 표시한 것으로 많은 대여소가 주말에 인기 있는 대여소와 일치한다. 예외적인 곳은 한강변 자전거 도로를 이용해서 출퇴근하는 사람들이 강변 안쪽 회사 밀집 지역을 이용하는 경우이다.
- 한강변에서 회사 밀집 지역으로 진입하는 도로에 자전거 도로를 편하게 만들어서 좀더 많은 사람들이 자전거로 출퇴근 할 수 있도록 유도하자.



나 지금 어느 단계를 공부하는 거지?

시각화 및 탐색

1.문제정의

2.데이터수집

3.데이터 가공

4.데이터 모델링

5.시각화 및 탐색

문제 정의에 맞춘 해결방안 도출

지금까지 수행한 데이터 분석 총정리

수업 마무리



GD쌤

지금까지 13회차 수업내용을 배워 보았습니다.

여러분은 지금까지 데이터 분석 5단계에 따라서 따릉이 데이터를 분석하고 해결방안을 도출해 보았습니다. 이런 방법으로 여러분 주위에 관심있는 분야의 데이터를 수집해서 분석하실 수 있습니다.

그동안 수고 많으셨어요. 감사합니다!