

# 파이썬으로 배우는 **따릉이** 데이터 분석과 시각화

## 2회차 – (2) 데이터 수집



# 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

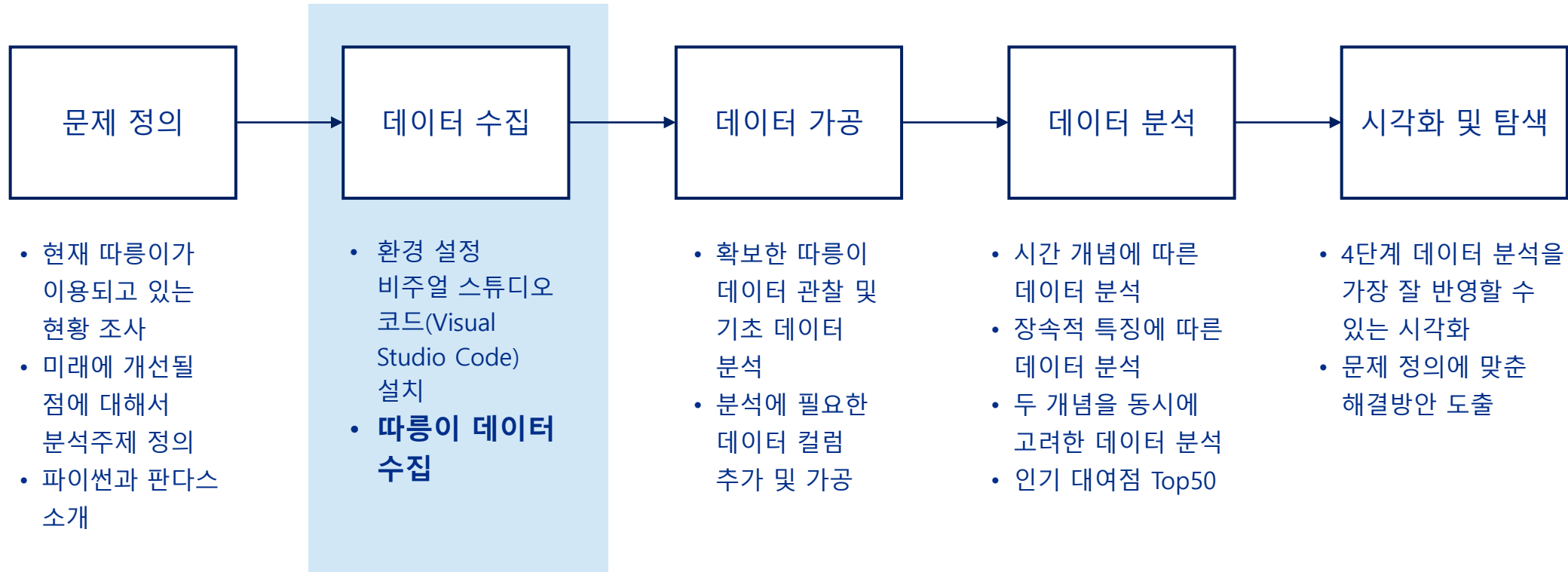
3. 데이터 가공

4. 데이터 분석

5. 시각화 및 탐색

데이터 분석 단계에 맞추어 따릉이 데이터 분석을 수행한다.

데이터 분석의 5단계



# 따릉이 데이터 : '서울 열린 데이터 광장'

## 2.데이터 수집

따릉이 데이터는 '서울 열린 데이터 광장' 사이트를 통해 공유 되어 있다.  
사이트에 접속하여 '따릉이' 로 검색한다.

data.seoul.go.kr

1 data.seoul.go.kr

로그 주피터 Home ppticon redmine preswot Google Google 계정 따릉이 머신러닝 OverTheMoon

서울특별시 시민감사추진위원회 출범 5주년 기념행사 온라인 참석 시민 모집 서울소식 응답소 정보공개 분야별정보

서울 열린데이터 광장 공공데이터 통계 소식&참여 이용안내 AI학습데이터 로그인 회원가입 사이트맵

모든 서울시민을 위한 공공데이터

열린데이터광장에서 서울시와 연계 기관이 공개한 공공데이터를 확인하실 수 있습니다.  
서울시와 관련된 다양한 공공데이터를 확인해 보세요.

2

Q 찾고 싶은 데이터를 입력해 주세요. 검색하기

데이터세 6,983 서비스 14,395 Open API 5,437

보건	일반행정	문화/관광	산업/경제	복지	환경
교통	도시관리	교육	안전	인구/가구	주택/건설

관심데이터 36종 생활이슈 자세히보기 >

뉴스레터 구독안내 Newsletter 열광뉴스레터를 신청하세요!

1 서울 열린데이터 광장  
data.seoul.go.kr 접속

2 검색창에 '따릉이'를  
입력하고 '검색하기' 클릭

# 따릉이 데이터 : 공유 데이터 종류

## 2.데이터 수집

'따릉이'의 검색 결과 14건을 찾았습니다. 조회순 조회

**서울시 공공자전거 이용현황**

서울시에서 운영하는 공공자전거(따릉이)에 대한 기간별, 대여소별 이용(대여, 반납)현황 데이터입니다.  
수정일자: 2019-12-23 제공기관: 서울특별시 제공부서: 도시교통실 보행친화기획관 자전거...

[교통] FILE

서울특별시 공공데이터

**서울특별시 공공자전거 대여이력 정보**

서울특별시 공공자전거 대여이력 정보입니다. 자전거 이동경로에 대한 데이터 분석이 가능하도록 년도...  
수정일자: 2020-10-22 제공기관: 서울특별시 제공부서: 도시교통실 보행친화기획관 자전거...

[교통] FILE

서울특별시 공공데이터

**서울특별시 공공자전거 대여소 정보**

서울특별시 공공자전거 대여소(따릉이) 현황정보입니다. 대여소의 이름, 관리번호, 위치정보와, 거치대...  
수정일자: 2020-10-22 제공기관: 서울특별시 제공부서: 도시교통실 보행친화기획관 자전거...

[교통] FILE

서울특별시 공공데이터

**서울특별시 공공자전거 실시간 대여정보**

서울특별시 공공자전거 실시간 대여정보입니다. 대여소별 실시간 자전거 대여가능 간수, 가치움, 대여...  
수정일자: 2021-05-07 제공기관: 서울특별시 제공부서: 도시교통실 보행친화기획관 자전거...

[교통] OpenAPI

서울특별시 공공데이터

**서울특별시 공공자전거 신규가입자 정보(월별)**

서울특별시 공공자전거 월별 신규가입자 정보입니다. 2017년 부터 2018년까지의 월별 신규가입자 ...  
수정일자: 2021-05-07 제공기관: 서울특별시 제공부서: 도시교통실 보행친화기획관 자전거...

[교통] SHEET OpenAPI FILE

공공데이터

**서울특별시 공공자전거 이용정보(시간대별)**

서울특별시 공공자전거 시간대별 이용정보입니다. 대여일자, 대여시간, 대여소번호, 대여소명, 정기간...  
수정일자: 2020-07-15 제공기관: 서울특별시 제공부서: 도시교통실 보행친화기획관 자전거...

[교통] FILE

공공데이터

**서울특별시 공공자전거 대여소별 이용정보(월별)**

서울특별시 공공자전거 대여소별 대여, 반납 정보입니다. 년월, 대여소번호, 대여소명, 대여간수, 반...  
수정일자: 2020-07-23 제공기관: 서울특별시 제공부서: 도시교통실 보행친화기획관 자전거...

[교통] FILE

공공데이터

**서울특별시 공공자전거 신규가입자 정보(일별)**

서울특별시 공공자전거 일별 신규가입자 정보입니다. 대여일자, 사용자구분, 성별, 연령대별 가입자 정...  
수정일자: 2020-07-07 제공기관: 서울특별시 제공부서: 도시교통실 보행친화기획관 자전거...

[교통] SHEET OpenAPI FILE

공공데이터

**서울특별시 공공자전거 이용정보(월별)**

서울특별시 공공자전거 이용정보입니다. 대여일자, 대여소번호, 대여소명, 대여구분(정기,비정기), 성...  
수정일자: 2020-07-23 제공기관: 서울특별시 제공부서: 도시교통실 보행친화기획관 자전거...

[교통] FILE

공공데이터

**서울특별시 공공자전거 이용정보(일별)**

서울특별시 공공자전거 일별 이용정보입니다. 대여일자, 대여소, 정기간유무, 성별, 연령대, 운동량...  
수정일자: 2020-07-06 제공기관: 서울특별시 제공부서: 도시교통실 보행친화기획관 자전거...

[교통] FILE

공공데이터

'따릉이'의 검색 결과 14건을 찾았습니다. 조회순 조회

**서울특별시 공공자전거 외국인 대여정보(월별)**

서울특별시 공공자전거 년도별 외국인 대여정보입니다. 2017년 부터 2018년까지의 외국인 대여정보...  
수정일자: 2020-07-23 제공기관: 서울특별시 제공부서: 도시교통실 보행친화기획관 자전거...

[교통] SHEET OpenAPI FILE

공공데이터

**서울특별시 공공자전거 외국인대여정보(일별)**

서울특별시 공공자전거 외국인 일별 이용정보입니다. 대여일자, 대여소번호, 대여소, 대여간수, 반납건...  
수정일자: 2020-07-23 제공기관: 서울특별시 제공부서: 도시교통실 보행친화기획관 자전거...

[교통] SHEET OpenAPI FILE

공공데이터

**서울시 공공자전거 고장신고 내역**

서울시 고장신고된 공공자전거의 자전거번호, 등록일시, 고장구분의 내용을 제공합니다.  
수정일자: 2019-12-23 제공기관: 서울특별시 제공부서: 도시교통실 보행친화기획관 자전거...

[교통] FILE

공공데이터

**서울특별시 공공자전거 대여 이력**

최근 7일 이내의 대여소 및 시간별 자전거 대여가능 간수, 가치움, 대여소 위치정보를 제공합니다. 호...  
수정일자: 2021-05-07 제공기관: 서울특별시 제공부서: 도시교통실 보행친화기획관 자전거...

[교통] OpenAPI

공공데이터

서울시 열린 데이터에는 따릉이 관련 데이터가 있다.

번호	제목	설명	주요 데이터 항목
1	서울특별시 공공자전거 대여소 정보	대여소 위치정보	대여소명, 위도, 경도, 구분, 거치대수
2	서울특별시 공공자전거 실시간 대여정보	실시간 대여정보	Open API
3	서울특별시 공공자전거 신규가입자 정보(월별)	월별 신규가입자 정보	가입일자, 성별, 연령대코드, 신규가입자수
4	서울특별시 공공자전거 대여이력 정보	자전거별 대여이력	자전거번호, 대여일시, 반납일시, 이용시간
5	서울특별시 공공자전거 이용정보(월별)	월별 사용자별 대여정보	성별, 연령대코드, 대여구분코드, 이용건수
6	서울특별시 공공자전거 외국인 대여정보(월별)	월별 외국인 사용이력	대여일자, 대여소명, 대여건수, 반납건수
7	서울특별시 공공자전거 대여소별 이용정보(월별)	월별 대여소별 대여건수	대여소명, 대여일자, 대여건수
8	서울특별시 공공자전거 외국인 대여정보(일별)	일별 외국인 사용이력	대여일자, 대여소명, 대여건수, 반납건수
9	서울특별시 공공자전거 이용정보(시간대별)	일별 사용자별 대여정보	성별, 연령대코드, 대여구분코드, 이용건수
10	서울특별시 공공자전거 신규가입자 정보(일별)	일별 신규가입자 정보	가입일자, 성별, 연령대코드, 신규가입자수
11	서울특별시 공공자전거 이용정보(일별)	일별 사용자별 대여정보	성별, 연령대코드, 대여구분코드, 이용건수
12	서울특별시 공공자전거 고장신고 내역	고장신고 내역	자전거번호, 등록일시, 고장구분
13	서울특별시 공공자전거 이용현황	일별 이용현황	대여일시, 대여건수



이번 분석에는 자전거 대여이력 정보를 사용하므로 ① 를 클릭한다.

The screenshot shows the Seoul Open Data Portal interface. At the top, there's a navigation bar with '서울특별시' (Seoul City) and various menu items. Below that, a search bar contains the text '찾고 싶은 데이터를 입력해 주세요.' (Enter the data you want to find). The main content area displays search results for '따릉이' (Ttarengi). On the left, a sidebar lists categories like '공공데이터' (Public Data), '통계' (Statistics), and '카테고리' (Category). Under '카테고리', '교통' (Transportation) is selected, showing 14 items. The main results area shows three items, with the second item, '서울특별시 공공자전거 대여이력 정보' (Seoul City Public Bicycle Rental History Information), highlighted with a red circle and the number 1. This item is categorized under '교통' (Transportation) and has a 'FILE' button next to it. The other items are '서울시 공공자전거 이용현황' (Seoul City Public Bicycle Usage Status) and '서울특별시 공공자전거 대여소 정보' (Seoul City Public Bicycle Rental Station Information).

'서울특별시 공공자전거 대여이력 정보'로 들어간다.

서울특별시 공공자전거 대여이력 정보

서울특별시 공공자전거 대여이력 정보입니다. 자전거 이동경로에 대한 데이터 분석이 가능하도록 년도별, 대여소별, 자전거가별 대여이력 원천 데이터를 제공합니다.(2015년 최초 공공자전거 서비스부터 2019년 11월까지 제공중)

파일내려받기 \* 파일에 이상이 있는 경우 오류신고를 통해 운영자에게 알려주세요. 오류신고

NO	항목	파일명	용량 (MB)	수정일	내려받기
1	데이터	공공자전거(대여이력 정보_2020.05.zip	69.7	2020.10.22	
2	데이터	공공자전거(대여이력 정보_2020.04.zip	68.2	2020.10.22	
3	데이터	공공자전거(대여이력 정보_2020.03.zip	49.5	2020.10.22	
4	데이터	공공자전거(대여이력 정보_2020.02.zip	25.0	2020.10.22	
5	데이터	공공자전거(대여이력 정보_2020.01.zip		2020.10.22	

1 [전체 파일보기](#) 2

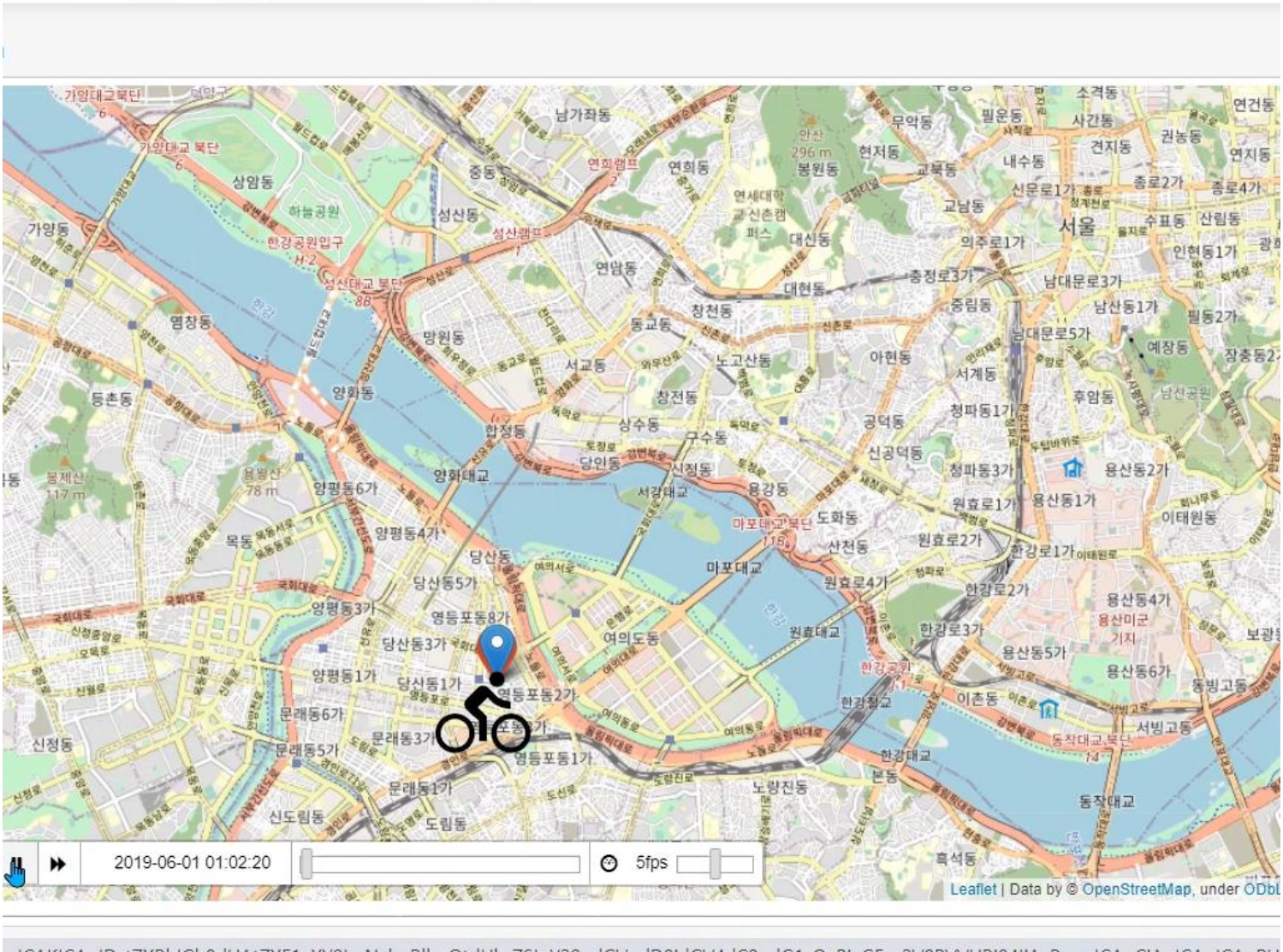
1 전체파일보기 를 클릭한다.

2 2019년 6월 데이터를 찾는다.

3 클릭해서 3개의 파일을 다운받는다.

'서울특별시공공자전거대여정보\_201906\_1.csv'  
'서울특별시공공자전거대여정보\_201906\_2.csv'  
'서울특별시공공자전거대여정보\_201906\_3.csv'

28	데이터	서울특별시 공공자전거대여정보_201907_1.csv	106.3	2019.12.10	
29	데이터	서울특별시 공공자전거대여정보_201906_3.csv	82.8	2019.12.10	
30	데이터	서울특별시 공공자전거대여정보_201906_2.csv	107.1	2019.12.10	
31	데이터	서울특별시 공공자전거대여정보_201906_1.csv	106.3	2019.12.10	
32	데이터	서울특별시 공공자전거대여정보_201810_01.xlsx	78.9	2019.06.20	



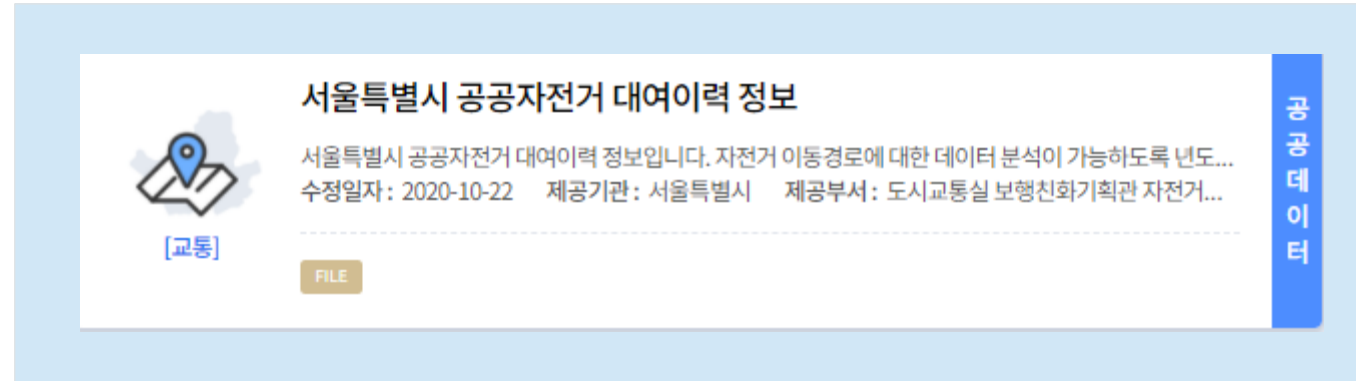
따릉이가 움직일 때마다 수집되는 데이터

- 1. 자전거번호 : 따릉이 고유식별번호
- 2. 대여일자, 대여 대여소번호, 대여대여소명, 대여거치대
- 3. 반납일자, 반납 대여소번호, 반납대여소명, 반납거치대
- 4. 이용시간, 이용거리



대여소에 자전거가 부족할 때 트럭으로  
자전거가 많이 있는 대여소에서 이동시  
킨다.

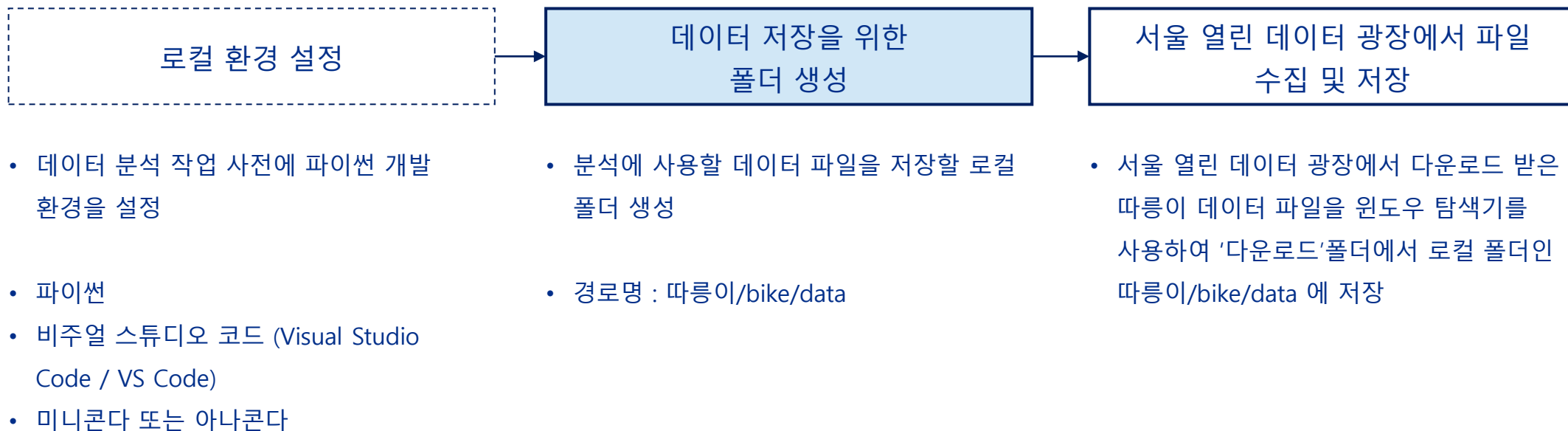
다운받은 데이터를 엑셀로 열어서 데이터를 확인한다.



	A	B	C	D	E	F	G	H	I	J	K
1	자전거번호	대여일시	대여 대여소번호	대여 대여소명	대여거치대	반납일시	반납대여소번호	반납대여소명	반납거치대	이용시간	이용거리
2	SPB-22040	2019-06-03 8:49	646	장한평역 1번출구 (국민은행앞)	1	2019-06-03 9:17	3	중랑센터	14	27	1330
3	SPB-07446	2019-06-03 8:33	526	용답토속공원 앞	8	2019-06-03 9:27	3	중랑센터	14	54	1180
4	SPB-20387	2019-06-05 8:27	646	장한평역 1번출구 (국민은행앞)	1	2019-06-05 8:41	3	중랑센터	2	12	1930
5	SPB-16794	2019-06-05 8:46	646	장한평역 1번출구 (국민은행앞)	6	2019-06-05 8:53	3	중랑센터	14	6	1340
6	SPB-18266	2019-06-10 8:27	529	장한평역 8번 출구 앞	10	2019-06-10 8:33	3	중랑센터	2	5	1230
7	SPB-13926	2019-06-11 8:29	646	장한평역 1번출구 (국민은행앞)	4	2019-06-11 8:37	3	중랑센터	2	7	1360
8	SPB-14638	2019-06-12 8:29	646	장한평역 1번출구 (국민은행앞)	4	2019-06-12 8:35	3	중랑센터	2	5	1340
9	SPB-18588	2019-06-17 8:34	646	장한평역 1번출구 (국민은행앞)	5	2019-06-17 8:44	3	중랑센터	2	8	1360
10	SPB-21148	2019-06-17 8:47	646	장한평역 1번출구 (국민은행앞)	6	2019-06-17 9:10	3	중랑센터	14	22	1330



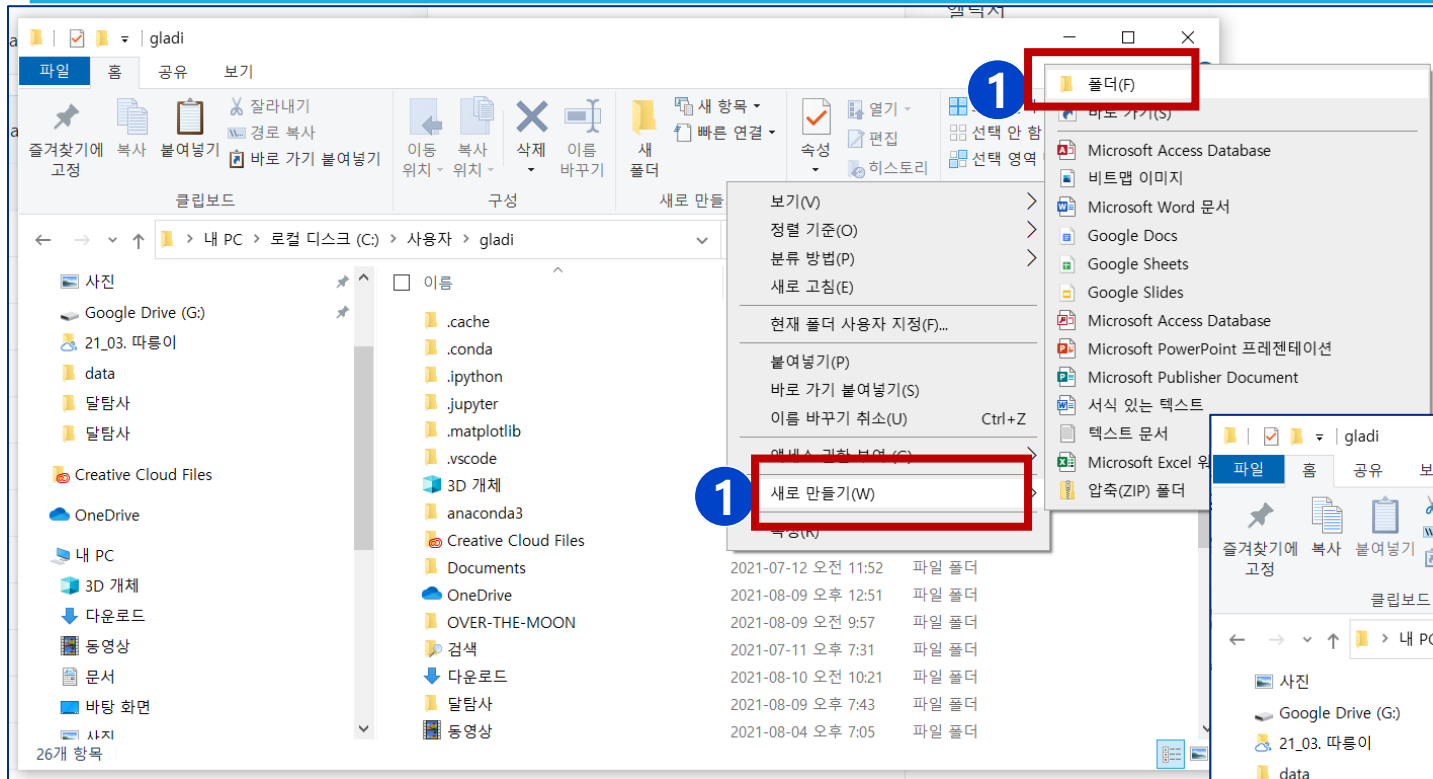
분석에 사용할 데이터 파일을 깃허브에서 수집하여 로컬 폴더에 저장한다.





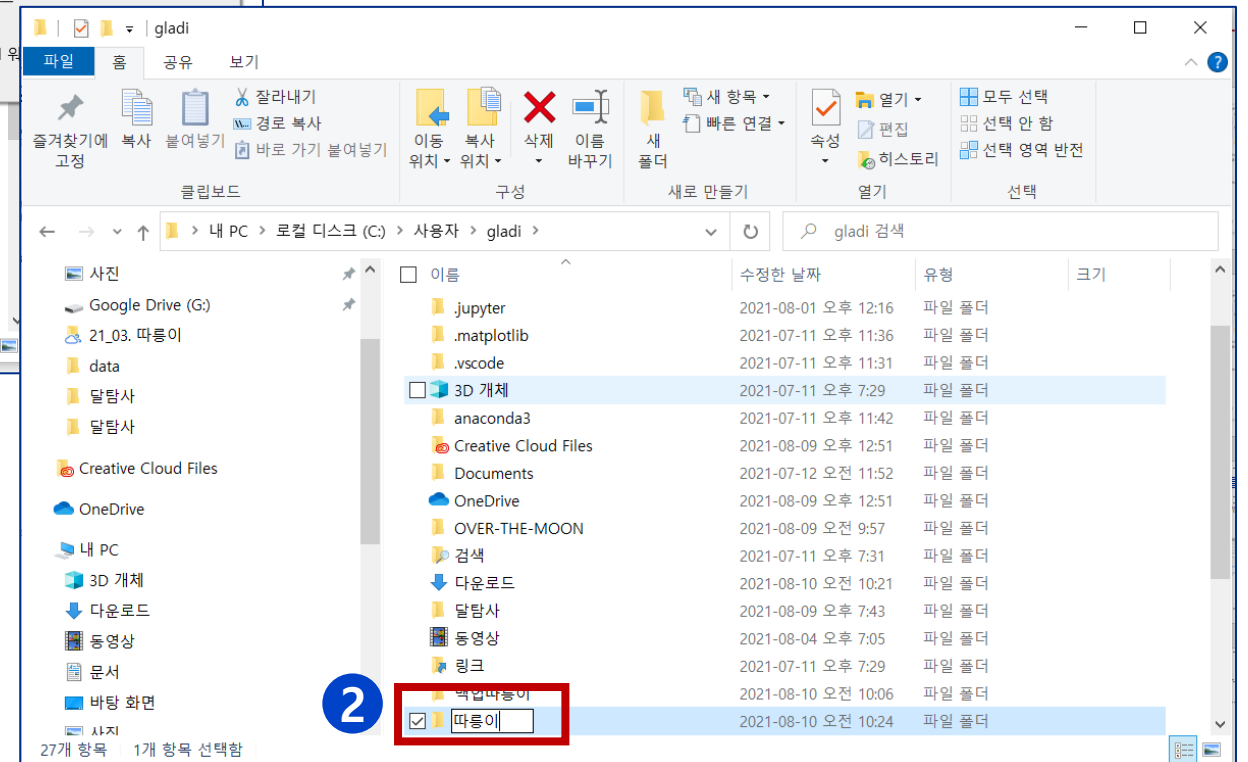
# 로컬 폴더 생성

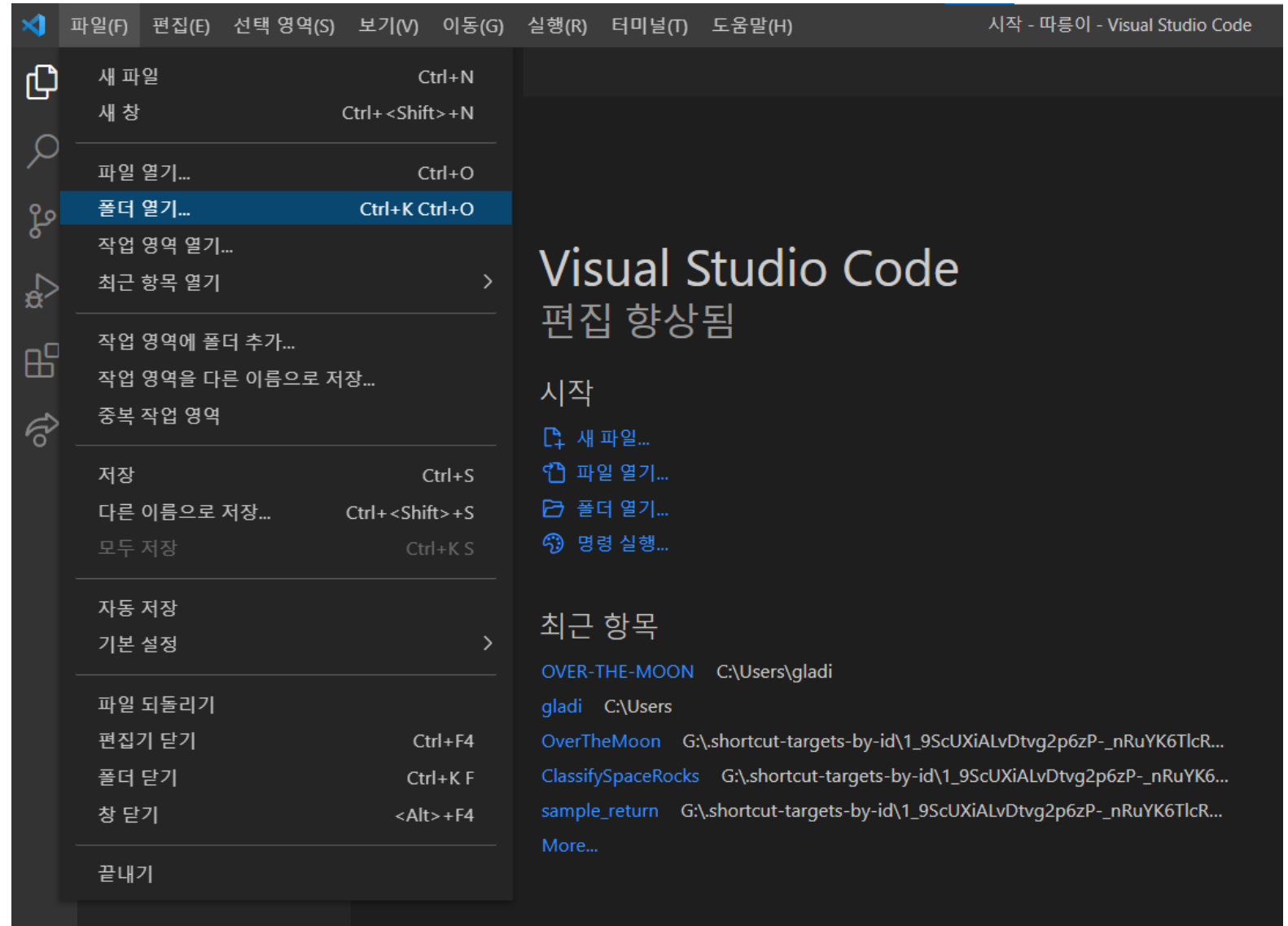
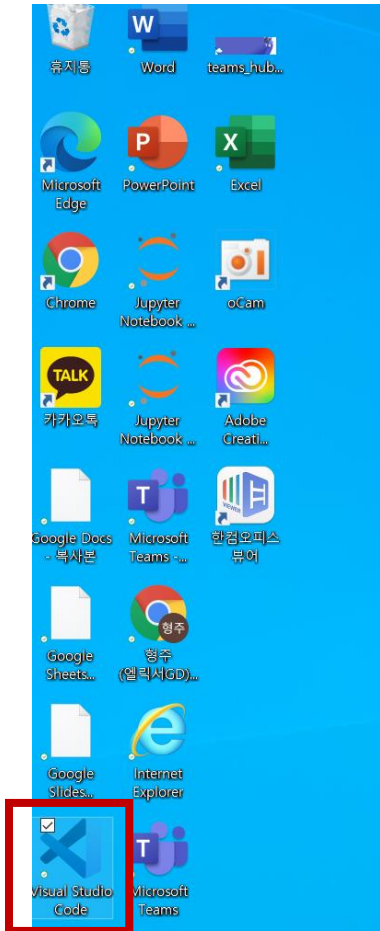
## 2.데이터 수집



1 탐색기를 사용하여 컴퓨터 내에서 원하는 위치에 새 폴더를 만든다.

2 따릉이 폴더를 만든다.



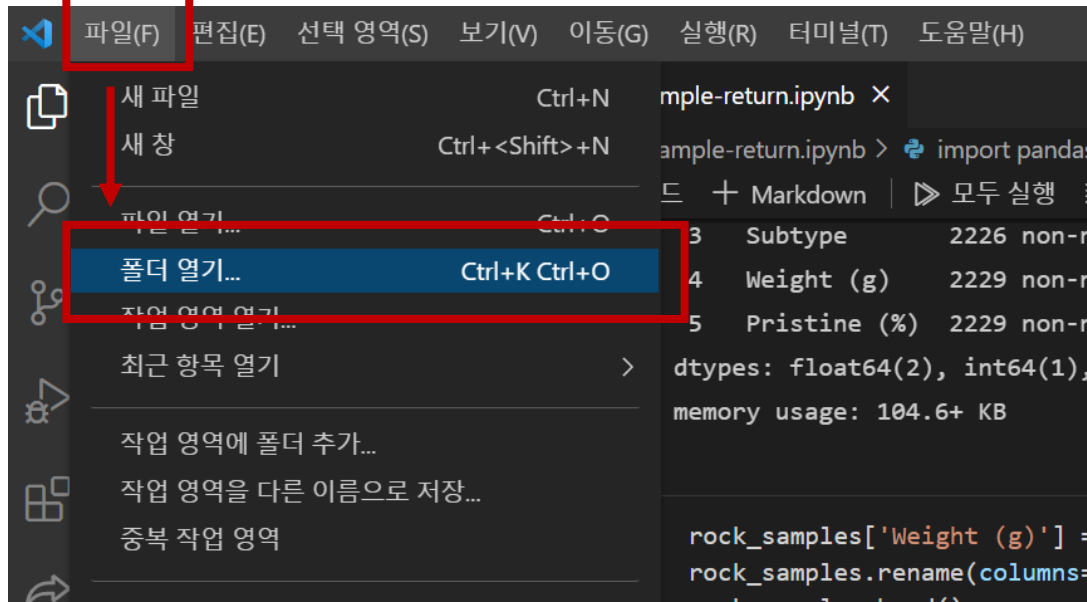


비주얼 스튜디오 코드 아이콘을  
클릭하면 실행되어 왼쪽 화면이 나온다.

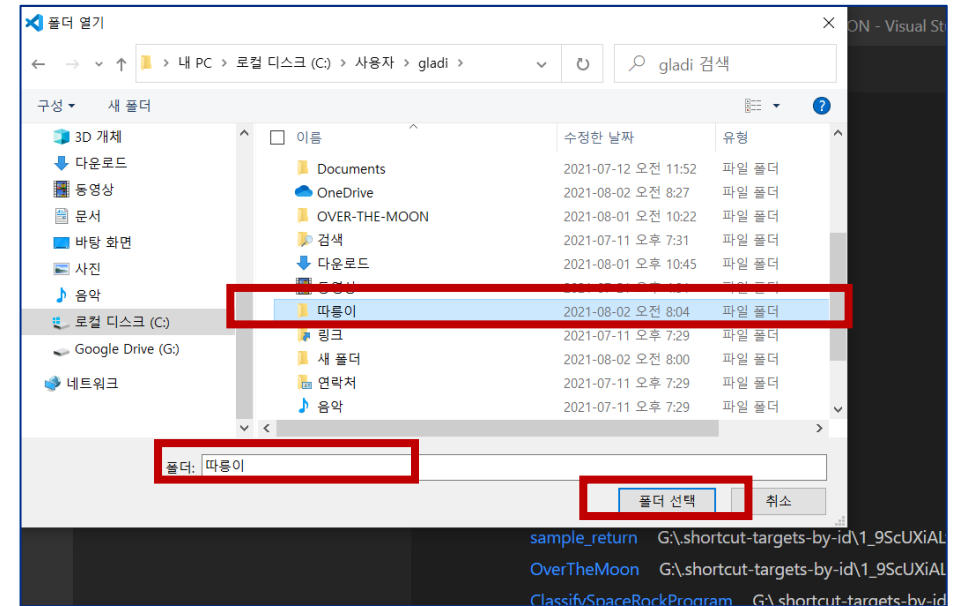
# 비주얼 스튜디오 코드에서 폴더 열기

## 2.데이터 수집

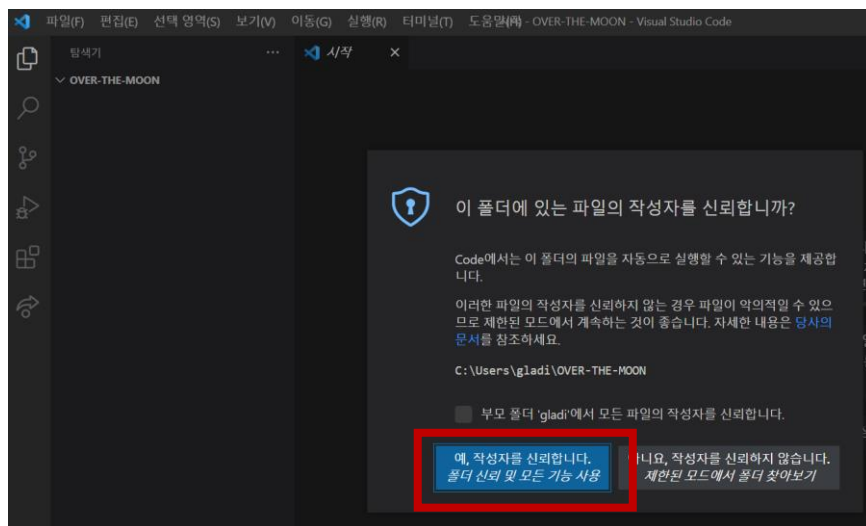
1



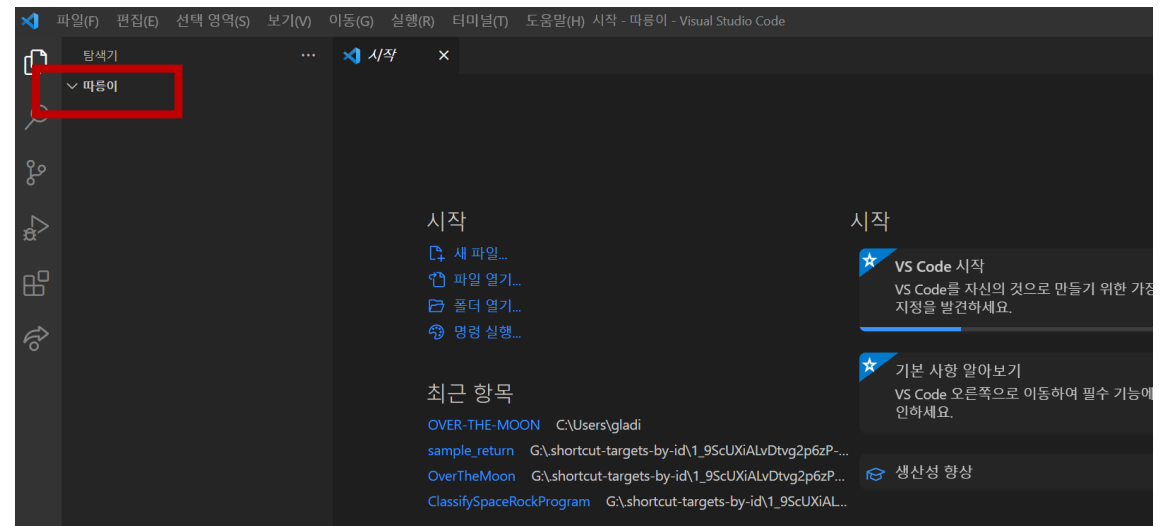
2



3



4





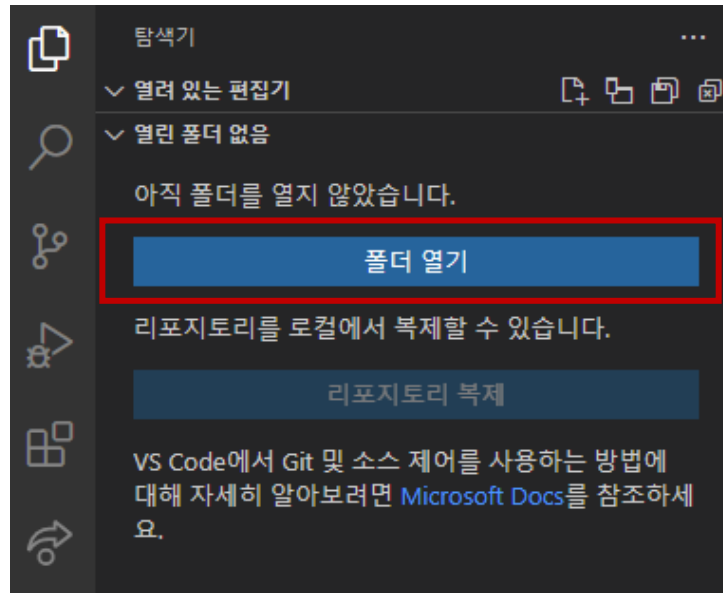
윈도우 탐색기에서 따릉이 폴더를 생성한 후 해당 폴더를 열고 신뢰 모드로 변경합니다.

### 폴더 생성

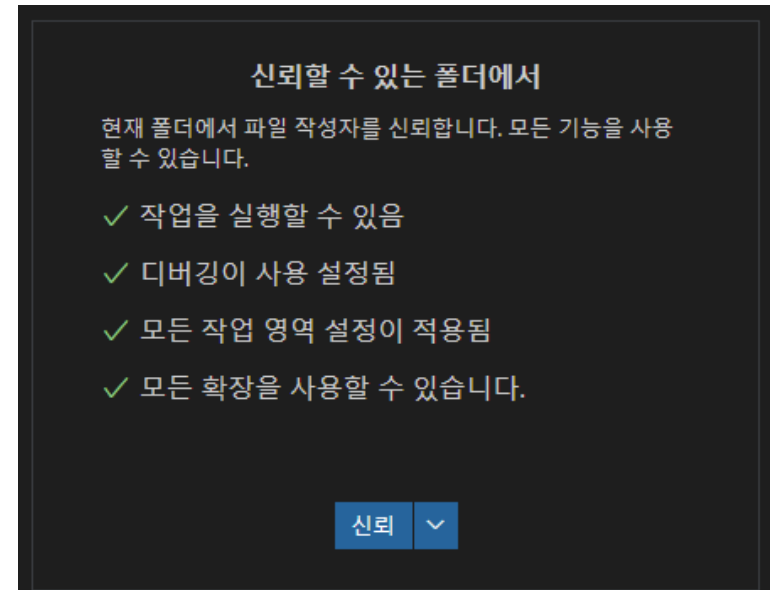
#### 따릉이

- 윈도우 탐색기에서 따릉이 폴더 생성

### 폴더 열기



### 폴더 신뢰

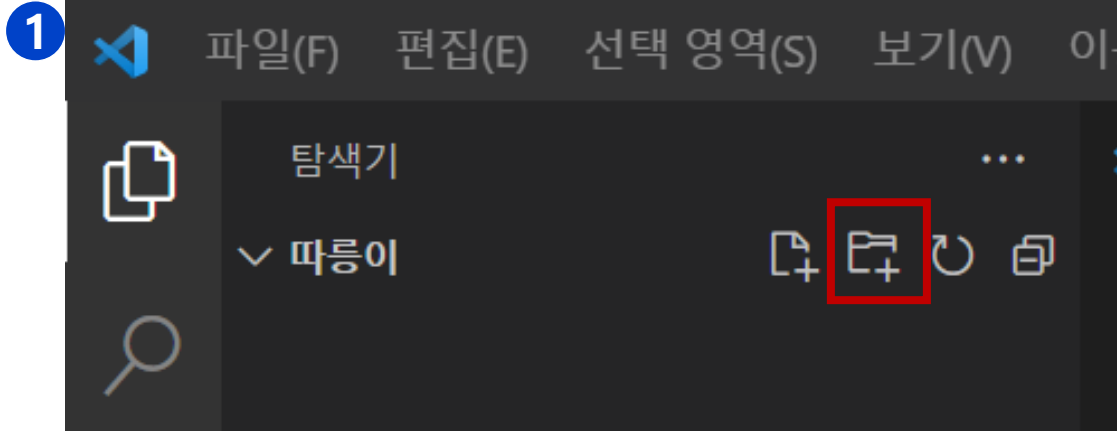


- 비주얼 스튜디오 코드에서 따릉이 폴더 열기

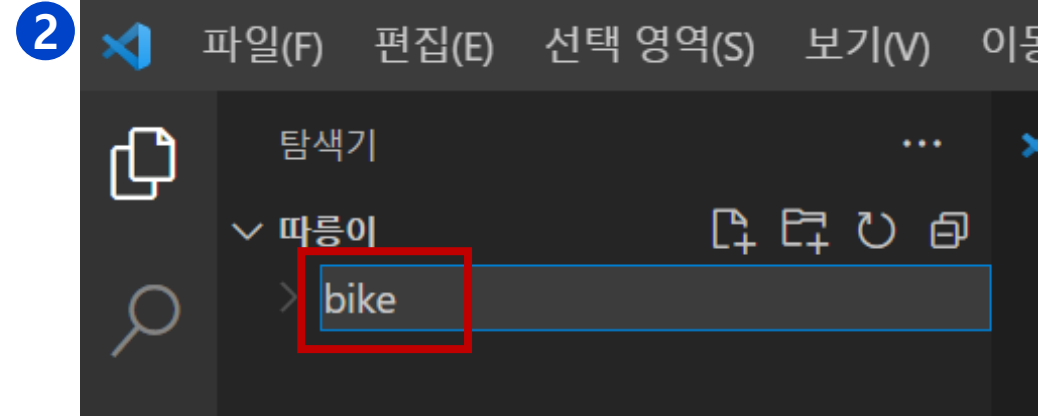
- 처음 여는 폴더는 신뢰 모드 적용이 필요
- 비주얼 스튜디오 코드는 폴더 (워크스페이스) 기준으로 신뢰 모드 적용

# 따릉이 하위 폴더 생성

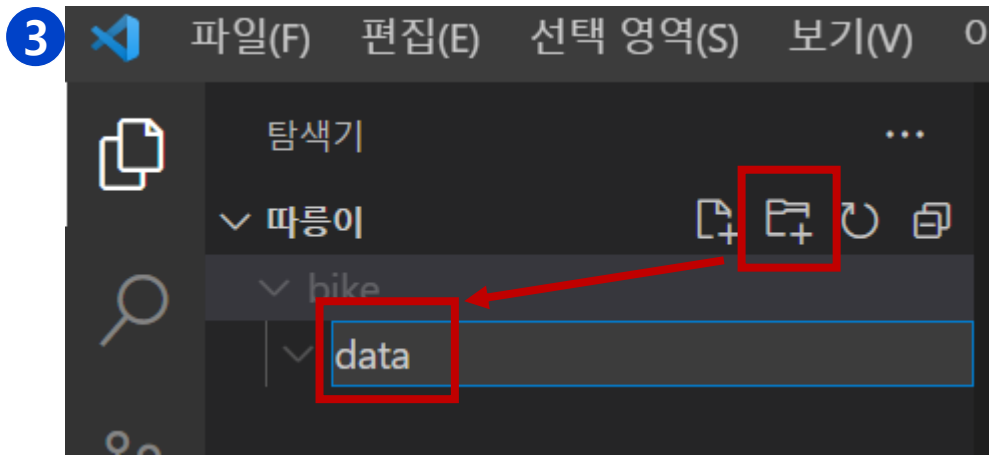
## 2.데이터 수집



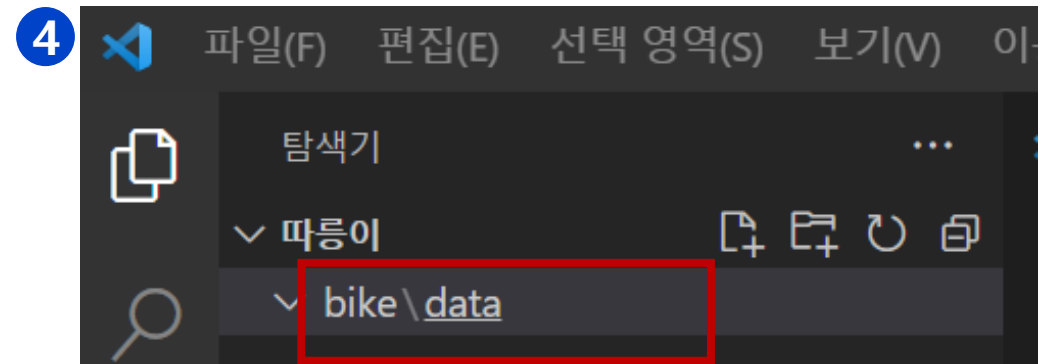
폴더생성 을 클릭한다.



bike 폴더를 만든다.



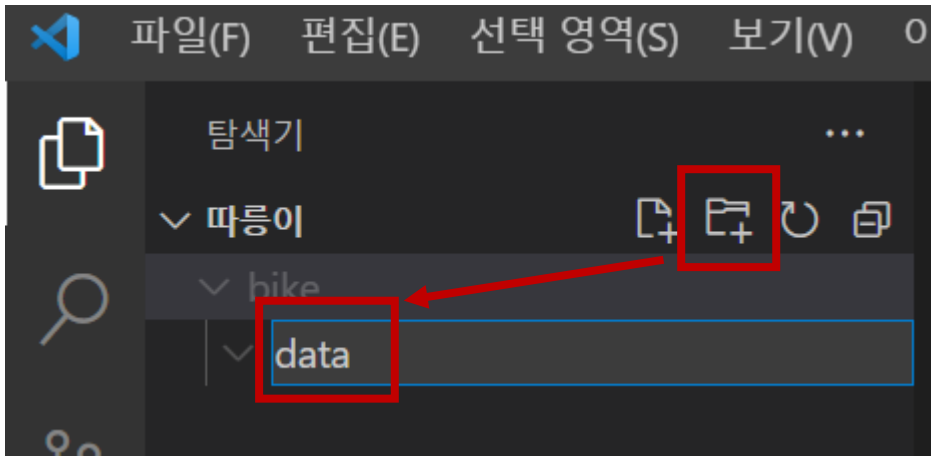
bike 폴더의 하위폴더인 data폴더를 만든다.



bike\data폴더에 데이터를 저장한다.

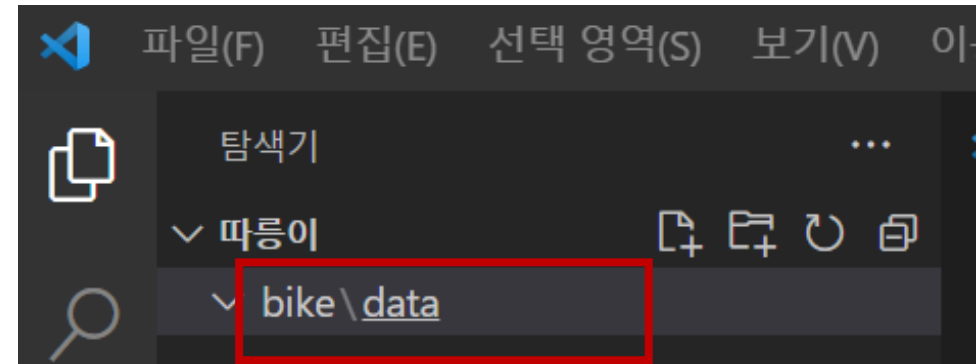
따릉이 폴더 아래에 bike 폴더 및 data 폴더를 생성합니다.

폴더 생성



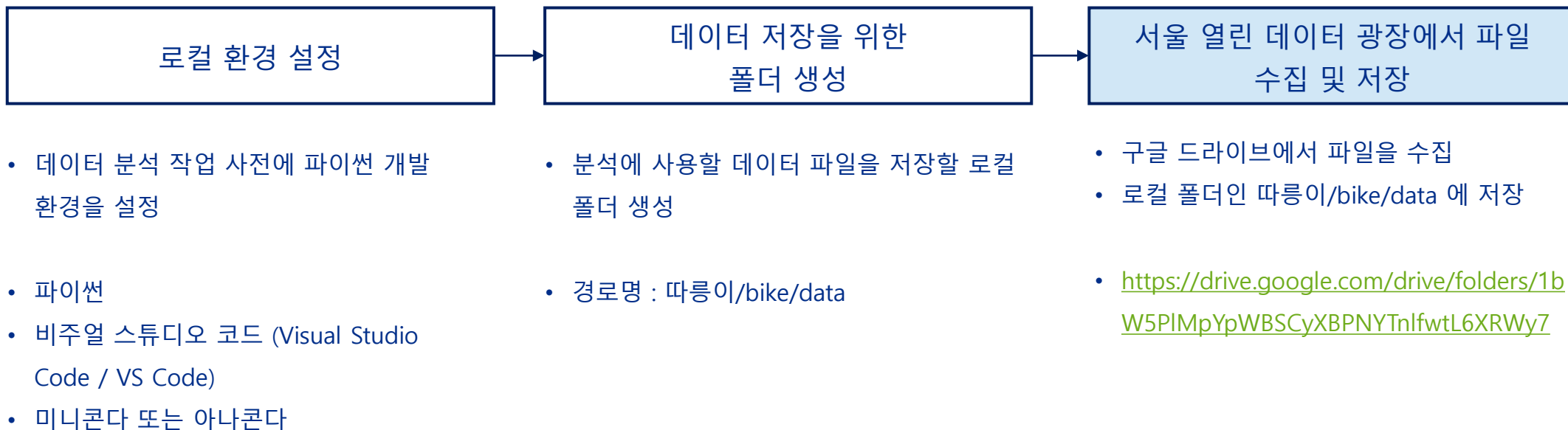
- 탐색기에서 따릉이 옆에 폴더 생성 아이콘을 클릭
- bike 폴더 생성
- bike 폴더 아래에 data 폴더를 생성

따릉이 / bike / data



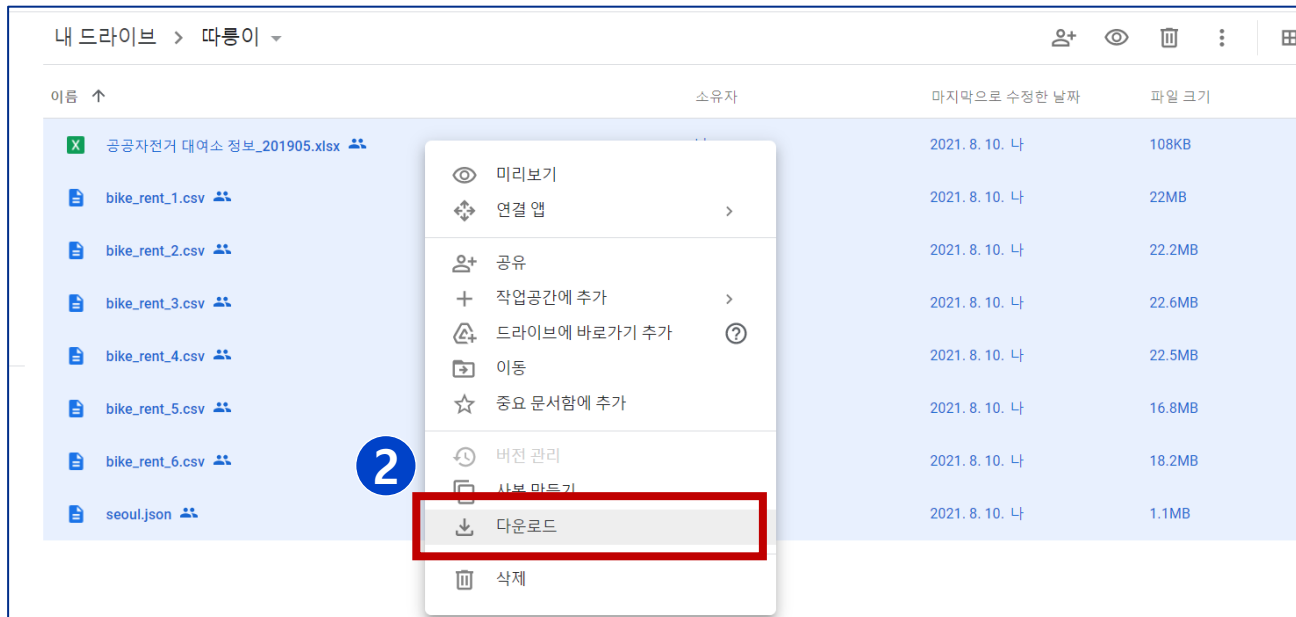
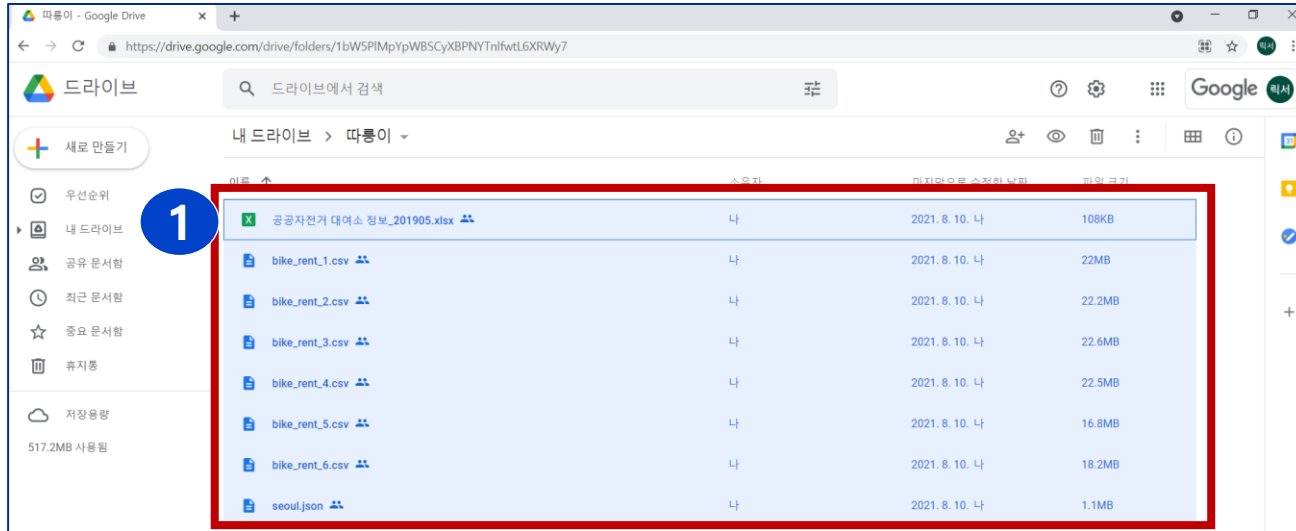
- bike 폴더 아래에 data 폴더만 있어서  
두 폴더가 한 줄에 겹쳐 보임
- 밑줄이 그어진 폴더가 선택된 폴더임

분석에 사용할 데이터 파일을 깃허브에서 수집하여 로컬 폴더에 저장한다.



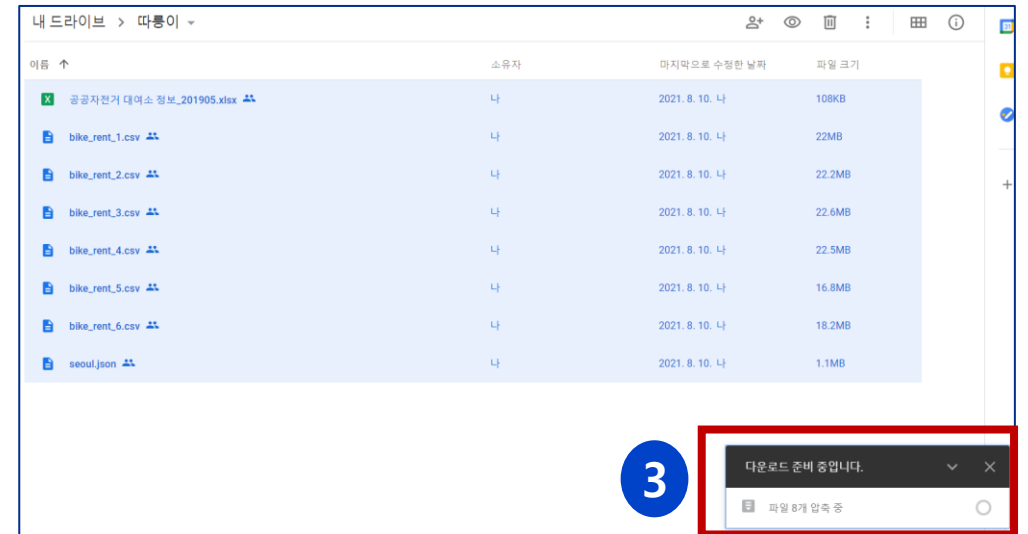
# 구글 드라이브에서 데이터 파일 다운로드

## 2.데이터 수집



1 Ctrl + 'A'를 눌러서 모든 파일을 선택한다.

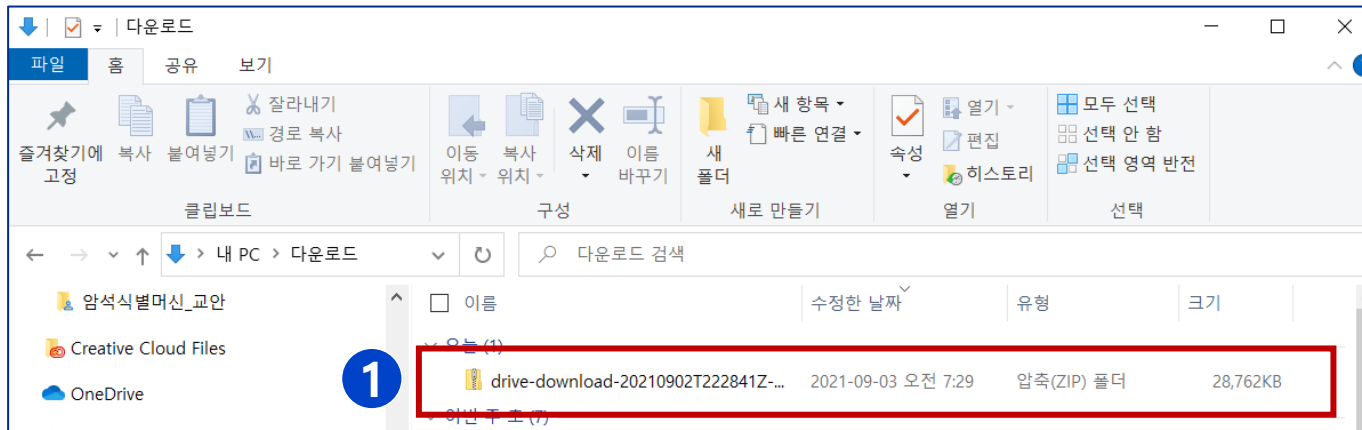
2 다운로드를 선택한다.



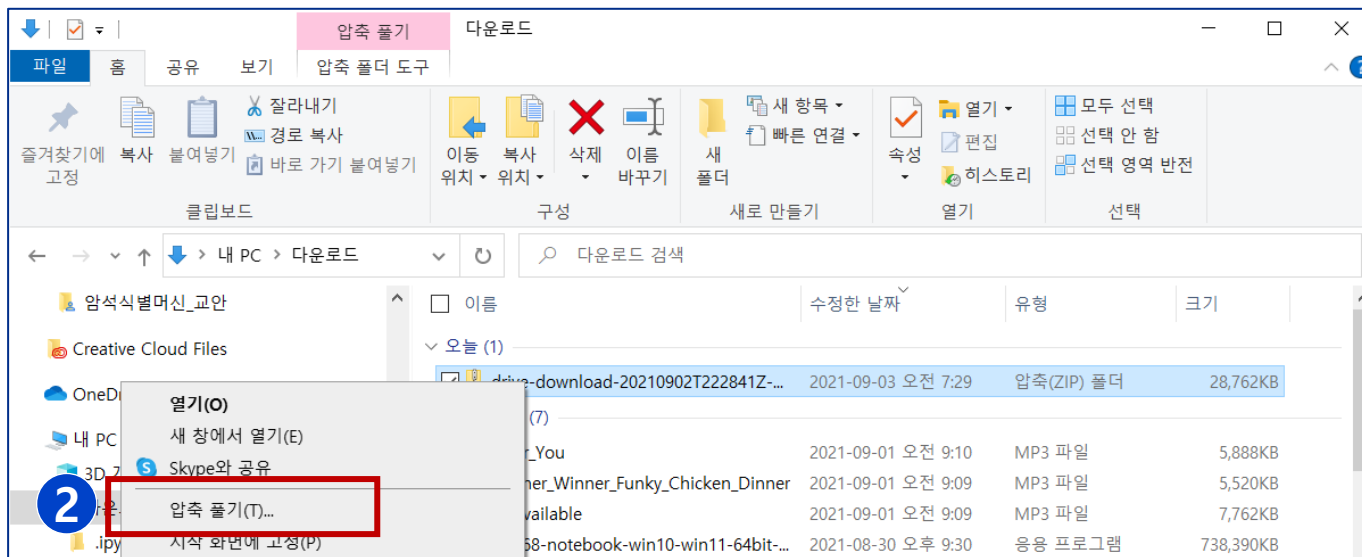
3 다운로드될 파일이 압축되어 내 컴퓨터의 다운로드 폴더에 저장된다.

# 데이터 파일 압축 풀기 - (1)

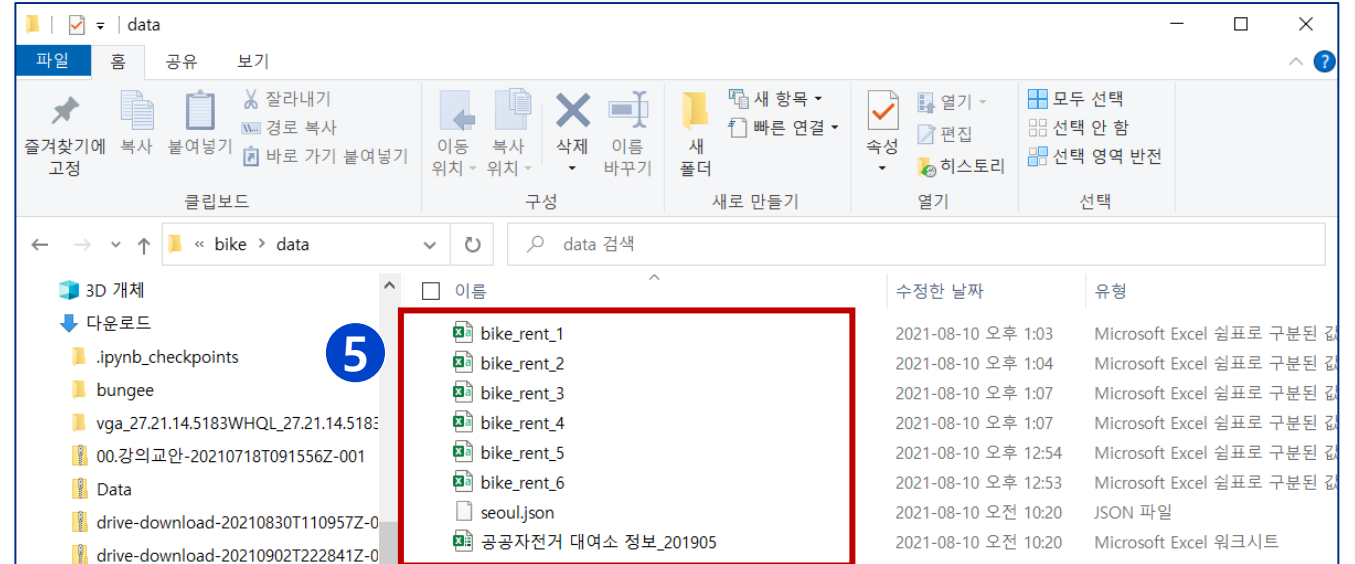
데이터 준비



1 다운로드 폴더에 가서 압축 파일을 클릭한다.

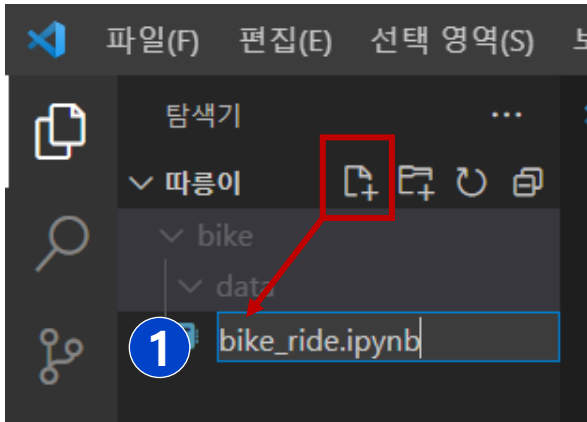


2 압축 파일일 클릭한 상태에서 마우스 오른쪽 버튼을 클릭하면 새로운 메뉴가 나오고 '압출 풀기'를 선택한다..

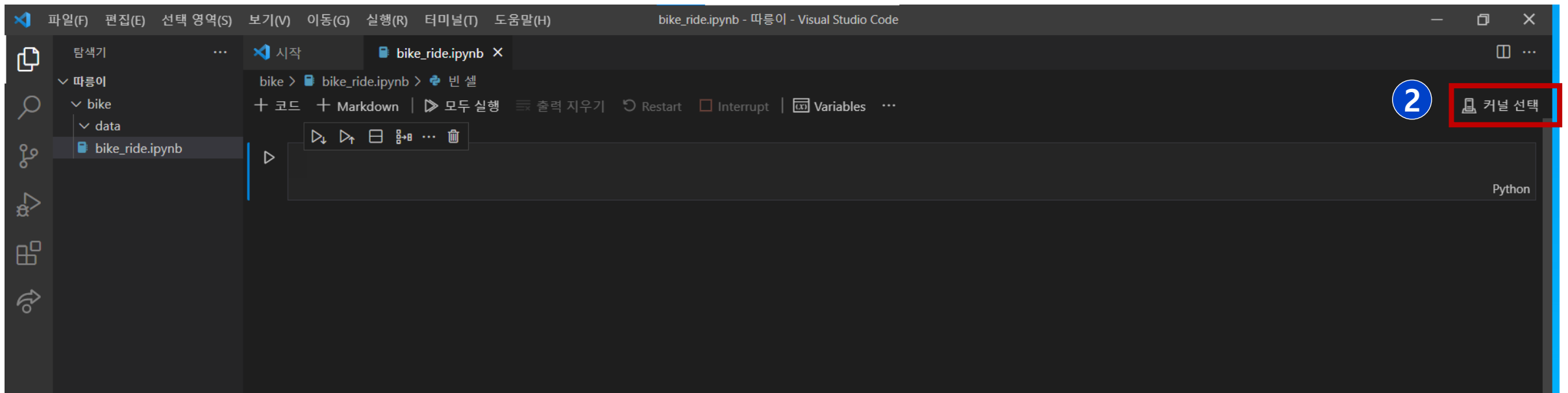


5 따릉이/bike/data폴더에 데이터가 있는 것을 확인한다.

- 3 찾아보기를 클릭해서 따릉이 폴더를 선택한다.
- 4 압축 풀기를 선택한다.



- 1 bike 폴더에서 '새파일'을 클릭하면 빈 칸이 나오고 bike Ride.ipynb 라는 따릉이 데이터 분석 파일을 만든다.
- 2 주피터 노트북이 열리면 오른쪽에 '커널 선택'을 클릭한다.





The image consists of two screenshots of the Visual Studio Code interface. The top screenshot shows the 'Python' kernel selection menu for the file 'bike Ride.ipynb'. The menu lists three options: 'Python 3.5 64-bit' (selected), 'Python 3.8.10 64-bit (myenv: conda)', and 'Python 3.8.8 64-bit (base: conda)'. A red box highlights the 'Python 3.8.8 64-bit (base: conda)' option, and a blue circle with the number '1' is placed next to it. The bottom screenshot shows the same interface after the selection. The 'Python 3.8.10 64-bit (myenv: conda)' option is now selected and highlighted with a red box. The status bar at the bottom of the window shows 'Python 3.8.5 64-bit' and 'Live Share'.

- 1 원하는 커널 파이썬 커널을 선택한다. (base : conda)
- 2 선택한 파이썬 커널로 파일이 실행되며 선택된 파이썬 커널이 표시된다.



# 나 지금 어느 단계를 공부하는 거지?

## 2.데이터 수집

1.문제정의

2.데이터수집

3.데이터 가공

4.데이터 분석

5.시각화 및 탐색

'서울시 열린 데이터'에 들어가서 따릉이 데이터를 검색

따릉이 관련 14개의 데이터를 살펴보고 분석할 데이터를 선택한다.

선택한 데이터를 엑셀로 열어서 데이터를 관찰하고 필요 데이터를 만든다.

아나콘다와 folium 라이브러리를 설치하고 따릉이 폴더와 하위 폴더를 만든다.

구글 드라이브에서 따릉이 데이터를 다운로드 받아 data 폴더에 저장한다.

비주얼 스튜디오 코드에서 따릉이 분석 파일을 만들고 커널을 선택한다.



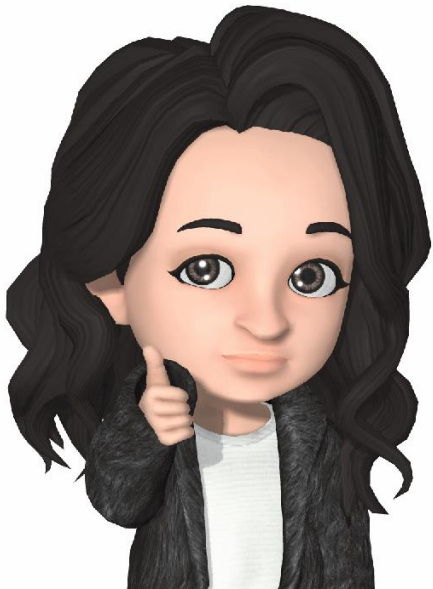
퀴즈를  
풀어봅시다

1. 따릉이 공공 데이터를 얻을 수 있는 인터넷 사이트는 ?

2. 이번 수업에서 분석할 따릉이 데이터는 몇 년도 몇 월의 데이터인가요?

3. 따릉이 데이터 파일의 확장자는 무엇인가요 ??

4. 소스 코드 에디터로서 코드 자동 완성 등 편리한 기능을 포함하고 파이썬, 주피터 노트북 등 다양한 확장 팩 (Extension)을 가진 것은 ?



GD쌤

지금까지 2회차의 두번째 수업내용을 배워 보았습니다.

다음 시간에는 3회차 수업내용으로 파이썬 라이브러리를 알아보고 따릉이 데이터를 관찰해 보겠습니다. 다음 시간부터는 파이썬 코딩 실습을 해봅니다.

수고 많으셨어요. 다음 시간에 만나요.