

이름	설명	결과	비고	사용 기술	유형
./data/subway.txt	- 서울시 주요 48 장소의 인근 지하철역 데이터 - 총 68 개 지하철역 - 크롤링 작업 중 검색용 데이터				데이터
./data/지하철역맛집정보링크.csv	- 역명 및 링크 정보 담긴 csv 데이터 - 총 6,398 건				데이터
./data/지하철역맛집정보최종.json	- 역명 + 각 음식점 정보 담은 json 데이터 - 총 6,398 건		- NoSQL 사용시 적재		데이터
./data/지하철역맛집정보최종.csv	- 위의 json 데이터 csv 로 가공한 데이터(메뉴 제외) - 총 6,398 건		- RDS 적재 필요		데이터
./data/지하철역맛집정보메뉴.csv	- 위의 json 데이터 중 메뉴만 따로 추출한 데이터 - 총 7,473 건		- DB 정규화 위해 따로 추출 - RDS 적재 필요 - 한 음식점당 5건		데이터
./data/지하철역별맛집정보최종_20230531.csv	- 지하철역맛집정보최종.csv 최종 가공 데이터	DB restaurant 테이블			데이터
./data/지하철역별맛집운영시간_20230531.csv	- 지하철역맛집정보최종.csv 최종 가공 데이터	DB restaurant_runtime 테이블			데이터
./data/지하철역별맛집휴식시간_20230531.csv	- 지하철역맛집정보최종.csv 최종 가공 데이터	DB restaurant_breaktime 테이블			데이터
MangoPlate_LinkExtract.py	- subway.txt 활용 맛집 검색 및 음식점 링크 추출	./data/지하철역맛집정보링크.csv		selenium	코드
MangoPlate_InfoExtract_json.py	- 지하철역맛집정보링크.csv 활용 음식점 데이터 추출	./data/지하철역맛집정보최종.json		bs4	코드
MangoPlate_InfoTransform_todf.py	- 지하철역맛집정보최종.json 활용 CSV로 데이터 변형 - 음식	./data/지하철역맛집정보최종.csv		pandas	코드
MangoPlate_InfoTransform_todf_menu.py	- 지하철역맛집정보최종.json 활용 메뉴 데이터 추출 - CSV로	./data/지하철역맛집정보메뉴.csv		pandas	코드
MangoPlate_ImageDownloader.py	- 지하철역맛집정보최종.csv 활용 이미지 다운로드	./images			코드
MangoPlate_ImageDownloader_Fail.py	- ImageDownloader 로그 정보 활용 - 다운로드 실패 이미지 재다운로드				코드
MangoPlate_Final_Transform.py	- DB 적재 과정에서 삭제 필요 컬럼 삭제(run_time, break_time) - 정규화 위해서, csv 분리(run_time, break_time) - S3 이미지 링크 가공(버킷 : standupseoul)	- ./data/지하철역별맛집정보최종_{date}.csv		pandas	코드
LoadMySQL.py	- Transform 결과 데이터 DB 적재	- ./data/지하철역별맛집운영시간_{date}.csv	- 보안 정보, git ignore	sqlalchemy	코드
LoadImageS3.py	- 이미지 데이터 S3 적재	- ./data/지하철역별맛집휴식시간_{date}.csv	- 보안 정보, git ignore	boto3	
./images	- 음식점 데이터 6,398건 중 6,215건의 이미지 - 183건은 원래 이미지 링크 없음		- S3 적재 필요		데이터, 이미지
test.ipynb	- 간이 이미지 데이터 검증 코드				코드
./logs	- 각 프로그램별 로그 정보				데이터