

[과제] API를 활용한 로켓 발사 데이터 ETL 프로세스 구현

1. 과제 개요

외부 URL로부터 로켓 발사 관련 데이터를 수집(Extract), 변환(Transform), 저장(Load)하는 ETL 프로세스를 이해하고 구현하는 것을 목표

- 데이터를 다루는 다양한 파이썬 라이브러리(requests, pandas, os, json)의 실습
 - 기본 설정 10개의 이미지 데이터외에 50개를 추출하여 시도를 추천(Optional)
-

2. 학습 목표

- requests 모듈을 사용하여 외부 API 데이터를 호출(Client-Server Architecture)
 - JSON 형식의 데이터를 파일로 저장하거나 pandas 데이터프레임으로 변환
 - os 모듈을 활용하여 로컬 디렉토리를 생성하고 파일을 관리
 - 반복문과 조건문을 활용하여 데이터 내의 특정 정보(예: 이미지 URL)를 추출하고 처리
-

3. 과제 수행 방법 (Step-by-Step)

단계 1: 데이터 수집 및 JSON 저장 (load_data_from_api)

- <https://ll.thespacedevs.com/2.0.0/launch/upcoming> URL을 통해 다가올 로켓 발사 정보를 호출
- DB대신 저장 폴더를 생성하고, API 응답 결과를 launches.json 파일로 저장
- 불러온 데이터를 pandas.DataFrame 형태로 반환하여 구조를 확인

단계 2: 이미지 데이터 추출 및 다운로드 (get_pictures)

- 단계 1에서 저장한 launches.json 파일을 다시 읽기
- 데이터 리스트 내의 각 항목에서 이미지 URL(image)을 추출하며 이때 이미지가 없는 경우를 대비한 예외 처리를 포함
- 별도 폴더를 생성하고, 각 URL에서 이미지 파일을 다운로드하여 저장
- 네트워크 연결 오류나 유효하지 않은 URL에 대한 try-except 예외 처리를 확인

단계 3: 이미지 데이터 읽어와 시각화



단계 4: 심화 과제(아래 2개에서 선택)

제공된 코드 하단의 연습문제를 참고하여 다음 데이터 중 하나를 선택해 **ETL** 과정을 직접 설계해 봅니다.

1. **Titanic** 데이터: 지정된 **URL**에서 **CSV** 데이터를 가져와 데이터프레임으로 분석하기.
2. **Scikit-learn** 데이터: **fetch_california_housing** 또는 **iris** 데이터를 불러와 탐색하기.

4. 제출 가이드

- 제출물: 수정 및 실행이 완료된 **.ipynb** 파일을 개인 깃허브에 **upload**
- 주의사항:
 - 모든 코드 셀은 오류 없이 순차적으로 실행
 - 데이터 다운로드 후 생성된 **data** 폴더와 **image** 폴더에 실제 파일이 존재하는지 확인
 - **API** 호출 시 발생할 수 있는 네트워크 오류에 주의