# <오전>

## [실습 환경]

* GitHub의 Codespace 사용
* Python 3.10.13 사용
* Pandas 라이브러리 사용 / import pandas as pd
* Rocksamples.csv 데이터 사용

## [실습 시작]

- notebooks 폴더에 Daltamsa\_excersize.ipynb 생성



- Python에 대한 설명

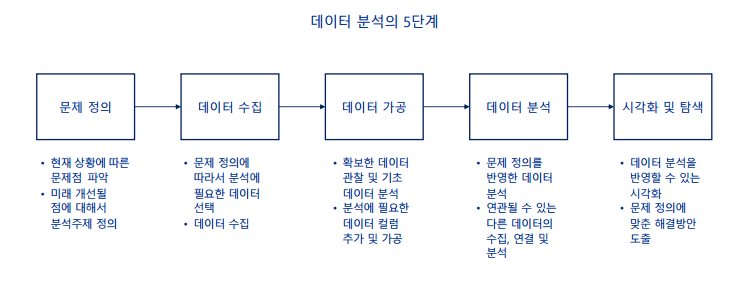
\* Python은 인터프리터 프로그래밍 언어이다. 때문에 문법이 간결하고 표현 구조가 인간의 사고 체계와 유사하지만 느리다.

\* Python은 풍부한 라이브러리를 가지고 있다. (최대의 장점)

\* 파이썬은 웹 개발, 데이터 분석, 머신 러닝, 딥 러닝 등 다양한 분야에서 활용된다.

- 수업 환경은 판다스(Pandas)로 활용

- 데이터 분석은 아래와 같이 5단계로 수행된다. 시각화 과정은 생략될 수도 있다.



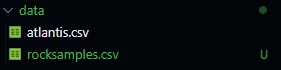
## [문제 정의 단계]

- PPT 참고

## [데이터 수집 단계]

-<https://github.com/drguthals/learnwithdrg/blob/main/OverTheMoon/sample-return/data/rocksamples.csv> 에서 rocksamples.csv 다운로드

- 다운로드 받은 rocksamples.csv를 data 폴더에 드래그하여 삽입



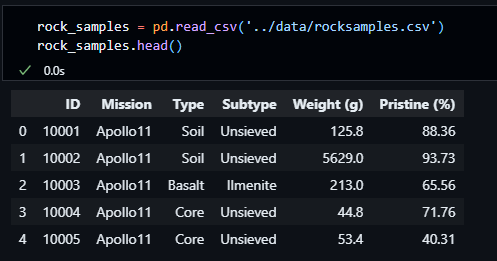
## [데이터 가공 단계]

- pandas 라이브러리 import 문을 활용하여 읽기.

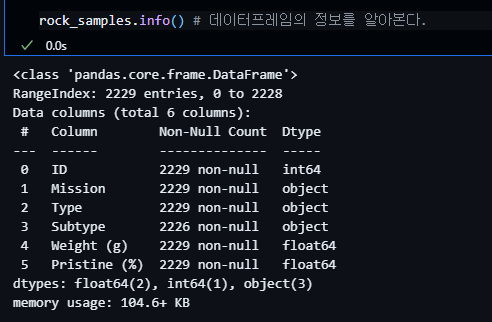
- rock\_samples : 아폴로 임무에서 수집된 암석 샘플 데이터프레임

\* pd.read\_csv() : 아폴로 임무에서 수집된 암석 샘플 데이터를 읽어 들인다.

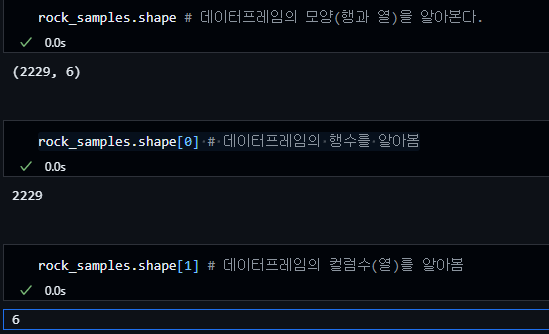
\* df.head(), df.tail() : 읽어 들인 데이터프레임을 둘러본다.



\* df.info(), 판다스의 자료형 : rock\_samples 데이터프레임의 정보를 알아본다.



\* df.shape : 데이터프레임의 모양을 알아본다.

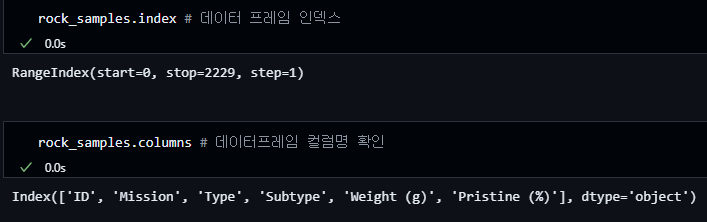


\* df.describe() : rock\_samples 데이터프레임의 요약통계를 알아본다.

\* df.apply(), lambda함수, df.rename() : rock\_samples안의 중량 단위를 ‘g’에서 ‘kg’으로 변경한다.

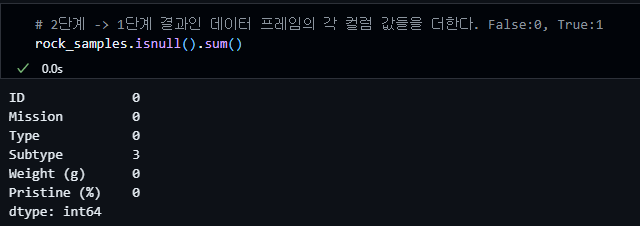
\* df.isnull().sum() : 데이터프레임의 각 컬럼의 갖는 누락값들의 합을 보여준다.

- codespace 작성









- 인덱스 연산자 : []라면 한 컬럼만 입력하는 형식의 시리즈, [[]]이라면 이중으로 선택한 데이터프레임이 된다.

- df,apply() : 데이터프레임의 컬럼, 시리즈 또는 데이터프레임 전체에 대해 함수르 적용하게 해주는 명령어.

- lambda 함수

\* 함수(Function) : 입력 값을 받아서 어떤 일을 수행한 뒤 그 결과값을 돌려 주는 구문들의 모음

\* 함수를 사용하는 이유 : 동일한 일을 반복적으로 수행해야 할 때 구문들을 매번 작성할 필요가 없다

\* 람다(lambda) 함수 : 함수를 재사용하지 않고 즉시 실행이 필용한 경우 익명 함수(일회성 함수)를 사용한다.

특히 판다스의 데이터프레임에 새로운 컬럼을 추가하거나 특정 컬럼 값을 변형시킬 때 apply 명령어와 결합하여 사용하면 유용하다.

\* 람다 함수의 표현 : lambda 입력값 : 결과값

예) rock\_samples[‘Weight (g)’].apply( lambda x : x \* 0.001 )

- 딕셔너리(dictionary, dic)

\* 정의 : 일상생활에서 사용하는 사전처럼 이름과 그 이름에 대응되는 내용이 하나의 항목으로 연결되어 있는 파이썬 자료구조이다. 이 때 이름을 키(key), 대응되는 내용을 값(value)라고 한다.

\* 만들기 : 딕셔너리 이름 = { key1 : value1, key2 : value2, … } 예) dict\_ex = {‘Weight (g)’ : ‘Weight (kg)’}

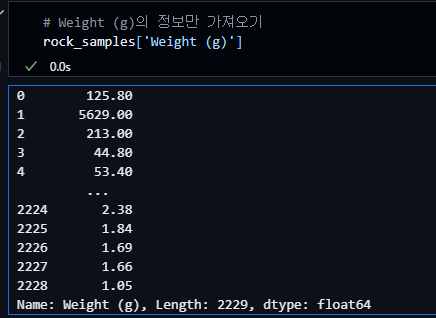
\* 항목 추출하기 : 딕셔너리 이름[ 항목의 키]

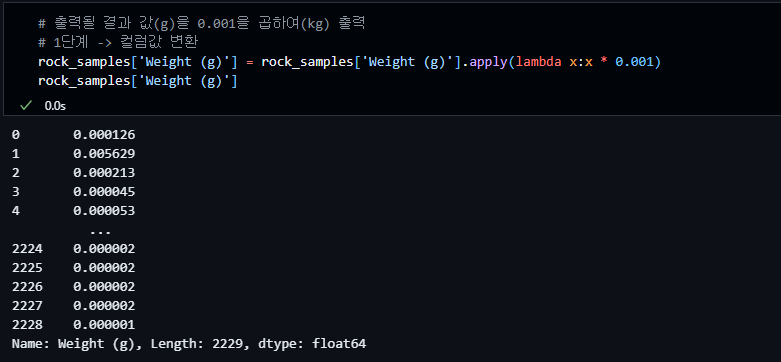
예) val = dict\_ex[‘Weight (g)’] / val에 할당되는 값은 딕셔너리 dict\_ex 에서 키인 Weight (g)에 대응되는 값인 Weight(kg)이 할당된다.

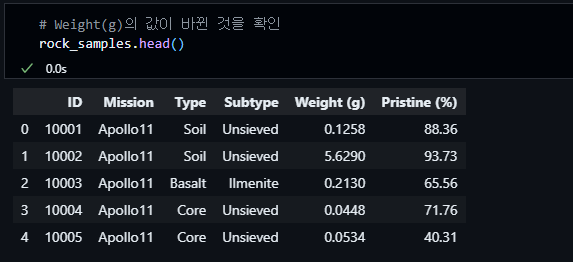
- df.rename() : 컬럼명 변경하기

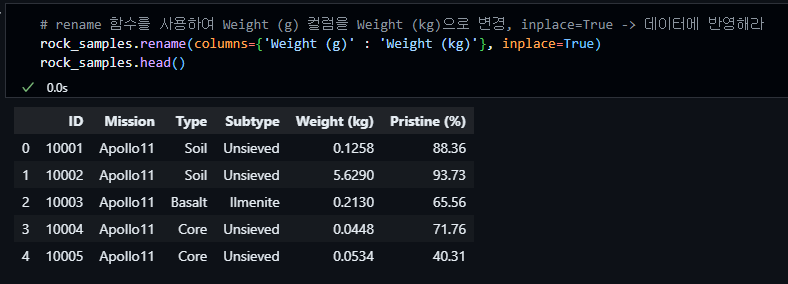
예) rock\_samples.rename( columns={ 변경전 컬러명 : 변경후 컬럼명 }, inplace=True )

# <오후>

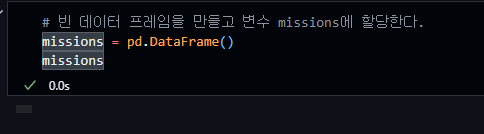


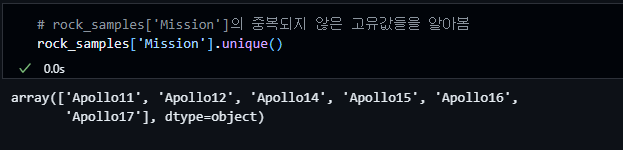


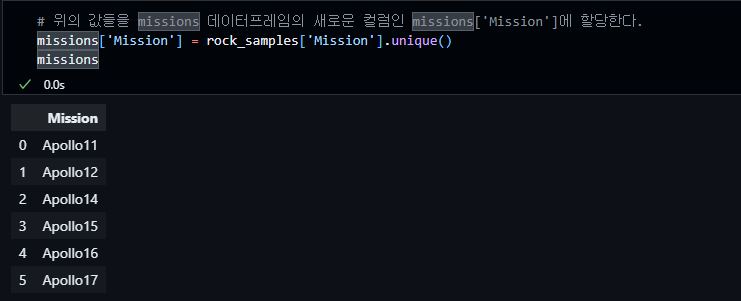




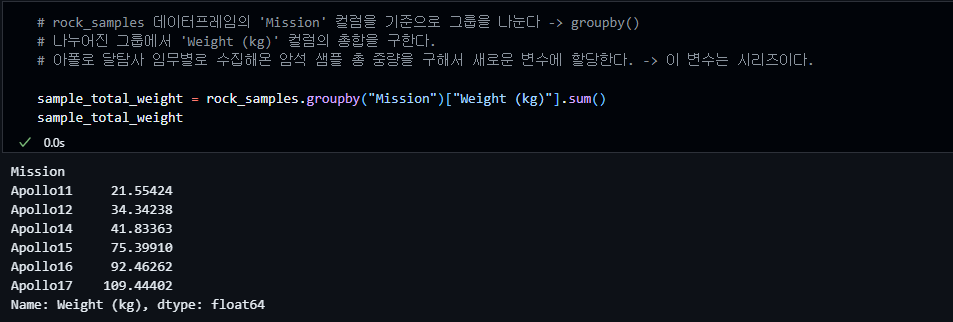
[아폴로 임무별 데이터 프레임 만들기 : missions]



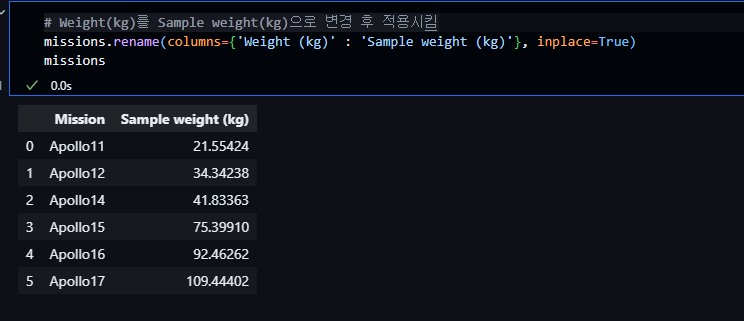


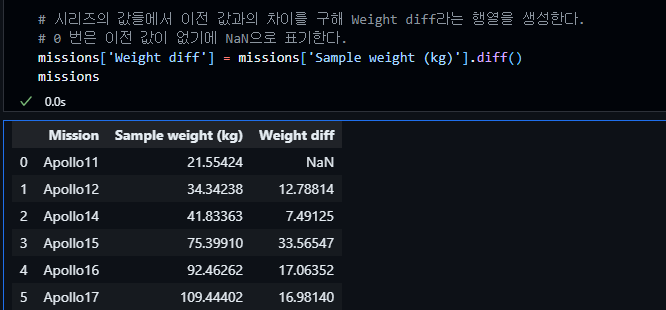


[groupby를 사용하여 임무별 샘플 중량의 총합 구하기]

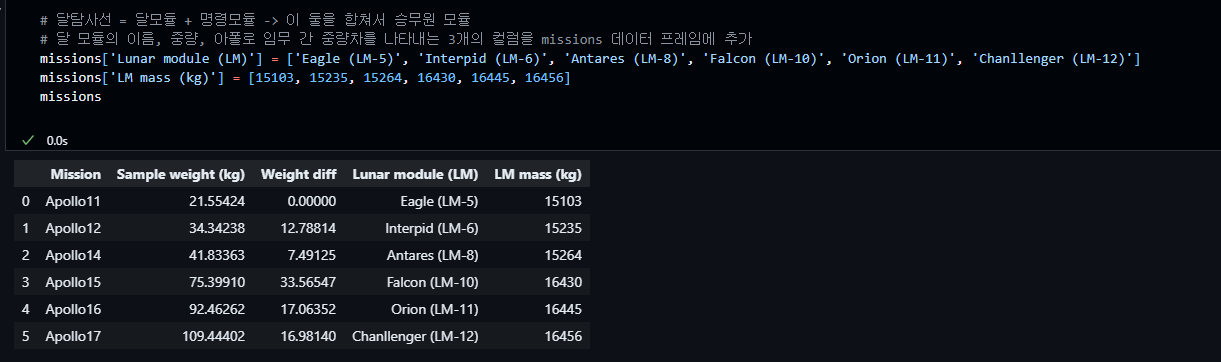








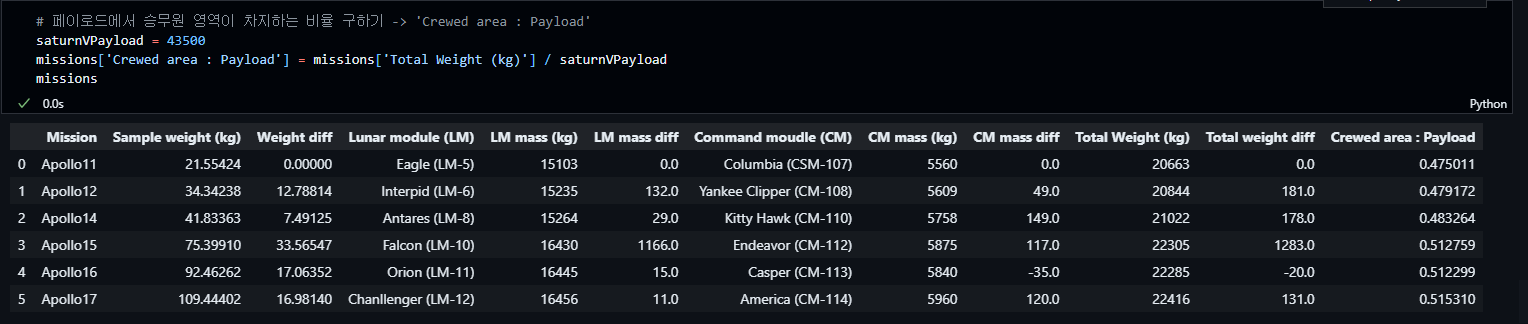


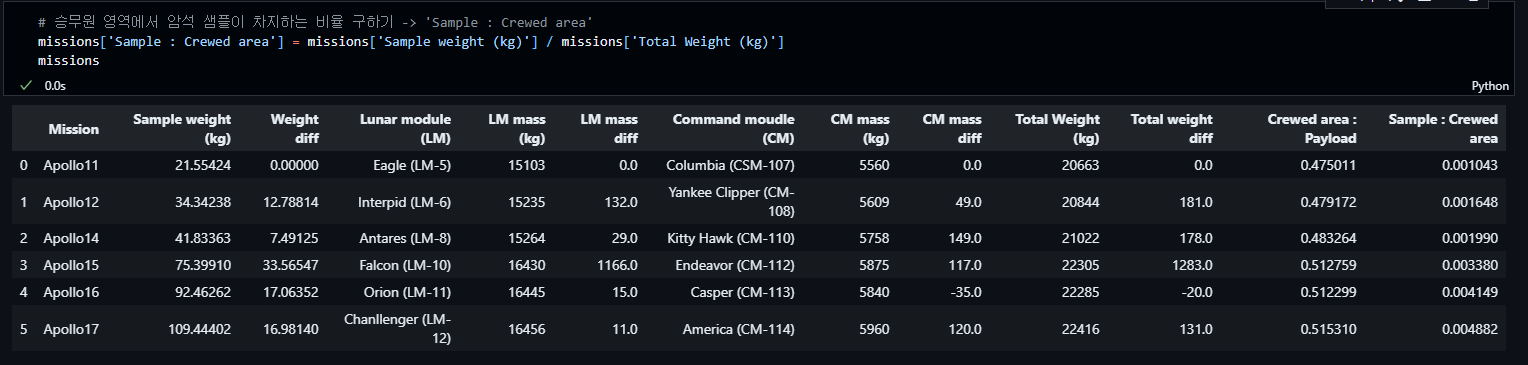














## [데이터 분석 단계]

