

print()  
type()  
len()  
sum()  
median()  
max()  
min()  
round()  
bool (True:1, False:0)

파이썬

퇴근후딴짓  
<https://inf.run/HYmN>  
(v.1.0)

판다스

라이브러리 및 데이터 불러오기  
import pandas as pd  
pd.read\_csv()  
데이터프레임과 시리즈  
DataFrame()  
df[['컬럼', '컬럼']]  
df[['컬럼']]  
Series()  
df['컬럼']  
데이터프레임 합치기  
concat([df1, df2], axis={0,1})

EDA  
head()  
shape  
info()  
상관관계  
corr() numeric\_only=True (only 코랩)  
isnull().sum()  
기초 통계  
describe() describe(include='O')  
mode()[0]  
카테고리  
df['컬럼명'].value\_counts() ascending=False  
df['컬럼명'].nunique()  
표준편차  
std()  
분산  
var()  
분위수  
df['컬럼'].quantile(.25)  
df['컬럼'].quantile(.75)  
idxmax()

인덱싱과 슬라이싱  
loc  
loc[인덱스 명]  
loc[행 시작: 행 끝, 열 시작: 열 끝] 끝 포함  
iloc  
iloc[인덱스 번호]  
iloc[행 시작 번호: 끝 번호, 열 시작 번호: 끝 번호] 끝 번호 포함 X

전처리  
dropna()  
dropna(subset=['컬럼명'])  
dropna(axis={0,1})  
fillna()  
drop()  
drop("컬럼명, axis=0) 행 삭제  
drop("컬럼명, axis=1) 컬럼 삭제  
축(axis)는 drop과 반대로 외울 것  
sum()  
axis=0 컬럼의 합  
axis=1 행의 합  
drop\_duplicates()  
get\_dummies()  
pop()

정렬  
sort\_values() ascending={True, False}  
sort\_index()  
reset\_index() drop={True, False}  
자료형 변경  
df['컬럼명'].astype('float')  
df['컬럼명'].astype('int')  
df['컬럼명'].astype('category')

문자열  
df['컬럼'].str.replace(변경 전, 변경 후)  
df['컬럼'].str.contains('키워드') 각 행별 True, False 반환  
df['컬럼'].str.len() 글자수  
df['컬럼'].str.lower() 소문자

시계열

pd.to\_datetime(df['날짜'])  
format = '%Y-%m-%d %H:%M:%S'  
%Y 연도 4자리  
%y 연도 2자리  
%m 월 2자리  
%d 일 2자리  
%H 시간 2자리  
%M 분 2자리  
%S 초 2자리  
df['날짜'].dt.  
year  
month  
day  
hour  
minute  
second  
dayofweek  
pd.Timedelta()  
diff = df[날짜1] - df[날짜2]  
분 diff.dt.total\_seconds()/60  
시간 diff.dt.total\_seconds()/60/60  
일 diff.dt.total\_seconds()/60/60/24