



# PANDAS-CHEATSHEET

## # 데이터 불러오기

`pd.read_csv('파일명.csv')`  
: CSV 파일에서 데이터를 읽어옴

`pd.read_excel('파일명.xlsx')`  
: Excel 파일에서 데이터를 읽어옴

## # 데이터 정보 확인

`df.head()`  
: 데이터프레임의 처음 5개 행을 보여줌

`df.info()`  
: 데이터프레임의 각 열의 정보를 요약해서 보여줌

`df.describe()`  
: 데이터프레임의 기술 통계 정보를 보여줌

`df.shape`  
: 데이터프레임의 행과 열의 개수를 반환함

## # 데이터 조작

`df['열이름']`  
: 특정 열을 선택

`df[['열1', '열2']]`  
: 여러 열을 선택

`df.loc[행인덱스, '열이름']`  
: 특정 행과 열을 선택

`df.iloc[행번호, 열번호]`  
: 특정 행과 열을 숫자로 선택

`df.drop('열이름', axis=1)`  
: 열을 삭제

`df.rename(columns={'이전이름': '새이름'})`  
: 열의 이름을 변경합니다.

`df.groupby('열이름').agg({'다른열': '합수'})`  
: 그룹화하고 집계 함수를 적용

`df.merge(other_df, on='공통열')`  
: 두 데이터프레임을 병합

`pd.concat([df1, df2], axis=0)`  
: 두 데이터프레임을 연결

`df.T`  
: 데이터프레임의 행과 열을 바꿈

## # 데이터 열/인덱스 이름 변경

`df.columns = ['새로운열1', '새로운열2']`  
: 모든 열의 이름을 변경

`df.rename(columns={'이전이름': '새이름'}, inplace=True)`  
: 특정 열의 이름을 변경

`df.rename(index={'이전인덱스': '새인덱스'}, inplace=True)`  
: 특정 행의 인덱스를 변경

## # 데이터 전처리

`df.dropna()`  
: 결측값이 있는 행을 삭제

`df.fillna(값)`  
: 결측값을 지정된 값으로 채움

`df.isnull()`  
: 각 요소가 결측값인지 여부를 확인

`df.duplicated()`  
: 중복된 행을 확인

`df.drop_duplicates()`  
: 중복된 행을 삭제

`df.sort_values(by='열이름')`  
: 특정 열을 기준으로 데이터를 정렬

## # 기타

`df.apply(함수)`  
: 함수를 열 또는 행에 적용

`pd.to_datetime(df['열이름'])`  
: 날짜 열을 datetime 형식으로 변환

`df.plot(kind='그래프종류')`  
: 데이터프레임을 시각화

`pd.get_dummies(df['열이름'])`  
: 범주형 변수를 더미 변수로 변환

## # 날짜 및 시간 관련

`df['날짜열'].dt.year, df['날짜열'].dt.month, 등`  
: 날짜 열에서 연도, 월 등을 추출

`pd.to_timedelta(df['시간열'])`  
: 시간 열을 timedelta 형식으로 변환