

공통기초스터디 6주차

DONGDUK AI LEARNING CREW TEAM SCALAR

TIME: PM 4:00~6:00



CSV

Comma-separated values

각 열을 쉼표나 공백으로 구분하며, 모든 행을 나열한 파일 형식
단순하면서도 쉽게 접근이 가능해 파이썬이나 빅데이터 분야에서 널리 사용되고 있음.



pandas에서 제공하는 csv관련 함수

- read_csv
: csv를 읽어서 데이터 프레임으로 만들어준다.
- to_csv
: csv 파일로 쓰기 위한 함수 .



데이터의 요약정보 확인

- info

:클래스 유형, 행 인덱스의 구성, 열 이름의 종류와 개수, 각 열의 데이터형과 개수 등에 대한 정보

- dtypes

:각 열에 대한 데이터형 정보

- describe

:숫자 데이터를 갖는 열에 대해서 평균, 표준편차 등 주요 통계정보를 요약해서 출력

- value_counts

:시리즈 객체의 고유값의 개수를 반환



데이터의 요약정보 확인

- mean,max,stdemd 등
:각각의 결과를 시리즈 객체로 반환
- corr
:두 열 사이의 상관계수



Pandas에서 그래프 그리기

plot 내장함수를 적용하면 바로 그래프 그리기가 가능하다.
아무것도 설정하지 않으면 기본적으로 라인그래프를 그린다.

kind 값	설명	kind 값	설명
'line'	선 그래프	'box'	박스 플롯
'bar'	수직 막대 그래프	'scatter'	산점도 그래프
'barh'	수평 막대 그래프	'area'	면적 그래프
'hist'	히스토그램	'pie'	파이 그래프



중복 데이터 판별

- duplicated

:이전에 나온 행들과 비교하여 중복된 행이라고 판별되면 True, 아니면 False를 반환

- drop_duplicates

:중복된 행들을 제거하고 고유한 값을 가진 행들을 남긴다.

subset 옵션을 적용해 열을 지정하여 일부 데이터의 중복 항목을 제거할 수도 있다.





THANK YOU