5/30 강의노트

텍스트, 스크린샷, 폰트, 번호이(가) 표시된 사진

자동 생성된 설명

Read\_csv의 skiprows 매개변수 : 앞에서부터 몇줄을 스킵하고 데이터를 불러들임

텍스트, 스크린샷, 폰트, 번호이(가) 표시된 사진

자동 생성된 설명

시리즈의 .str 명령어 : 시리즈의 각 행에 함수를 각각 적용할 수 있도록 해줌

.strip() 함수 : 문자열 양 끝의 띄어쓰기, 엔터, 탭 등을 제거하는 함수

텍스트, 스크린샷, 폰트, 번호이(가) 표시된 사진

자동 생성된 설명

판다스 isna() 함수 : 데이터프레임 내에 각 요소가 결측치인지 확인하는 함수

텍스트, 스크린샷, 폰트, 번호이(가) 표시된 사진

자동 생성된 설명

sum 함수로 결측치의 개수 확인 (True = 1)

텍스트, 스크린샷, 번호, 폰트이(가) 표시된 사진

자동 생성된 설명

최저기온이 결측치인 행만 확인

텍스트, 스크린샷, 번호, 폰트이(가) 표시된 사진

자동 생성된 설명

최고기온이 결측치인 행만 확인

텍스트, 스크린샷, 폰트, 번호이(가) 표시된 사진

자동 생성된 설명

Idxmax : 가장 높은 값이 있는 인덱스를 반환

텍스트, 스크린샷, 폰트, 번호이(가) 표시된 사진

자동 생성된 설명

Loc으로 해당 열의 데이터를 추출

텍스트, 스크린샷, 번호, 폰트이(가) 표시된 사진

자동 생성된 설명

Endswith 함수를 이용하면 문자열의 마지막이 어떻게 끝나는지 조건을 걸 수 있음

텍스트, 스크린샷, 번호, 폰트이(가) 표시된 사진

자동 생성된 설명

To\_datetime 함수 : 문자열을 날짜 타입 데이터로 바꿔줌

텍스트, 스크린샷, 디스플레이, 번호이(가) 표시된 사진

자동 생성된 설명

날짜 타입 데이터는 날짜 관련 명령어 수행 가능 (dt = datetime)

텍스트, 스크린샷, 디스플레이, 소프트웨어이(가) 표시된 사진

자동 생성된 설명

Boxplot : 상자수염그림

노란선 : 중위값

박스의 상하 : 1, 3사분위값

위아래 가로선 : 최댓값, 최솟값

동그란 점 (위에는 없음) : 이상값 (상자로부터 1, 3사분위값 사이의 범위 = IQR의 1.5배를 벗어난 비정상적인 분포를 가지는 값) 해당 값은 최댓값, 최솟값 집계에서 제외됨

텍스트, 스크린샷, 폰트, 번호이(가) 표시된 사진

자동 생성된 설명

상관분석 – 상관계수

상관계수는 .corr() 함수를 사용하면 바로 출력됨

상관계수가 0.3 이하이면 거의 관계가 없다고 봄

0.3~0.7 : 상관이 있다

절댓값 0.7 이상 : 관계가 강하다

텍스트, 스크린샷, 번호, 폰트이(가) 표시된 사진

자동 생성된 설명

히트맵 시각화

Cbar : 참고 바 활성화

Annot : 숫자 표기

Square : 정사각형으로 표시 (안하면 직사각형)

fmt : 숫자 표기 포맷