Sigmadream jtablesaw

R

깃 프로젝트 3종 : 라이선스 이그노어 리드미(설명)

깃이그노어 : 윈도우 mac os 리눅스 vscode, r,

data폴더와 output 폴더 구분

eda – 탐색적 분석. 데이터의 경향성을 보는 목적

1. 데이터분석의 기초

프로젝트구조를 잘 결정하세요(사람들에게 내가 만든 방법 알려줘야함)

즉 일관된 흐름, 구조를 가져야함

1. 데이터가 가장 중요

데이터에 대한 설명 리드미파일에 적는다

데이터의 컬럼명,자료구조, 자료형 파악하고 있어야함.

데이터는 숫자로 구성되는 게 원칙. 숫자로 안 돼있으면 숫자로 바꿔야함

1. 결측치 및 극단치 일단 ‘삭제’. 보관은 나중에.

Db에 넣으려면 na를 제거해야한다.

삭제한 데이터 따로 보관. 나중에 필요하면 db없데이트해서 추가

Nosql , 자스.

1. 그래프는 반드시 확인하고 표현(google seaborn)
2. 모든 그래프는 X,y, data가 반드시 필요(옵션은 기억x)

X,y 연관관계 파악

(y~x | 조건, data)

1. 산점도(xyplot) 히스토그램(histo), 라인그래프(line)

기술면접 때 코드하나도 못치면 너무 슬픔…

하루에 8시간 배우면, 오늘 배운게 목차처럼 떠올라야하고 말할 수 있어야함

khen-roe.github.io 만들어서 관리하라

보여주기 싫은건 보여주지 마라

구글에 코드가 있어도 내가 짠것과 안 짜본건 다름. 안해보면 머리가 하얘짐

프로그래밍에서는 행보다는 컬럼에 주목해야함. 데이터분석의 기본적 접근은

컬럼으로 접근한다. Db도 동일함.

코테 통과 못하면 프리렌서도 못한다

R쓰는 이유? 패키지 때문에

<https://seaborn.pydata.org/examples/index.html>

계산가능한 건 생략 가능.

산포도 Xyplot(y, ~ x | )

산포도 히스토그램 라인그래프 - 3개는 필수

정리

Csv 파일 – data 폴더 – 읽어서 str로 구조확인. Na값 확인

결측치 제거.

데이터분포 – 산점도로 확인.

히스토그램으로 분포확인하고 추세선 그림.

보고서 작성