02 - 데이터마트

01. - ②

여러 곳에 흩어진 data 수집한 뒤
이용자 목적에 맞게 구성된 data의 집합
⇒ data mart

---

02. - ①

dataframe은

. 하나 이상의 특정 변수를 기준으로 나누는 함수
⇒ melt()

· 나누어진 data를 원하는 구성으로 재결합하는 함수
⇒ cast()

∴ ①

---

03 - 데이터 탐색

01. - EDA

data의 기초통계량 값 확인
다양한 관점에서
data 바라보며
data 이해하기 위한 목적으로 수행하는 작업

⇒ 탐색적 데이터 분석
(EDA: Exploratory Data Analysis)

02. - ①

단순대치법으로 결측값 처리

⇒ complete.cases 이용

college라는 dataframe을

사전에 copy_college ← college 로 복사.

∴ ①

03. - ④

이상값

: 값이 존재하는 결측값과 달리

다른 data와 비교했을 때

극단적으로 (크거나 / 작은) 값

입력자의 실수로 입력 or 응답과 악의적 의도

로도 가능.

ESD (Extreme Studentized Deviation)

: mean으로부터 sd 3 만큼 떨어진 값들을 이상값 판단.

사분위수

⇒ 사분범위에서 1.5분위수 벗어나는 경우 이상치.

⇔ (Q1 - 1.5 × IQR / Q3 + 1.5 × IQR)

∴ ④