

🔥 Warm Up 🔥

🤔 왜 데이터 전처리가 중요해요?

😊 데이터 전처리는 분석의 정확성과 신뢰성을 높이기 위해 필수적입니다. 깨끗하고 구조화된 데이터를 사용해 의미 있는 통찰을 도출하고, 예측 모델의 성능을 향상시킬 수 있습니다.

좋은 재료를 통한 맛의 향상



목적에 위한 정확한 준비



👁 Definition

데이터 전처리란 수행하고자 하는 분석에 적합하게 데이터를 가공하는 작업을 의미

- 현업에서 발생하는 데이터를 포함하여 분석을 위해 수집한 데이터는 분석 모델에 적합하지 않은 경우가 많음
- 결측값 혹은 이상치가 있는 데이터

📌 데이터 전처리의 종류

구분	내용
데이터 클리닝	결측치 처리, 이상치 확인 및 제거 등
데이터 통합	다양한 데이터 피처의 결합 등
데이터 변환	스케일링, 유형 등
데이터 축소	변수 축소, 러닝레이 등
분류형 데이터 처리	언더 샘플링, 오버 샘플링 등
데이터 분할	train, test 데이터 분할 등