

2-2

# 심화로 향하는 인공지능 개념완성

: 데이터





# 데이터에 대해 알아봅시다

인공지능에서 필수적이고 기계가 학습하기 위해 필요한 것



데이터 속성(특성, 변수)

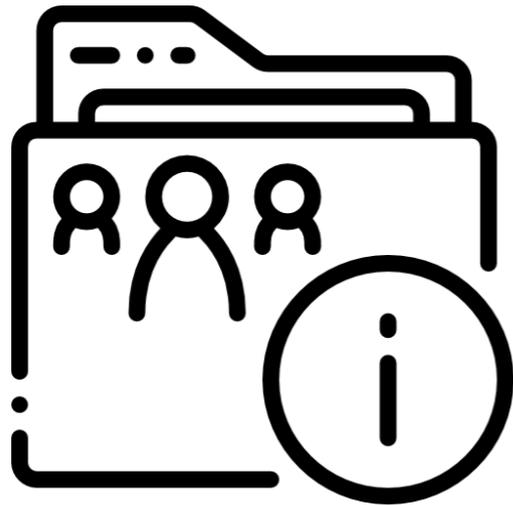
독립 변수/종속 변수

연속형/범주형 데이터

정형/비정형 데이터

데이터 시각화

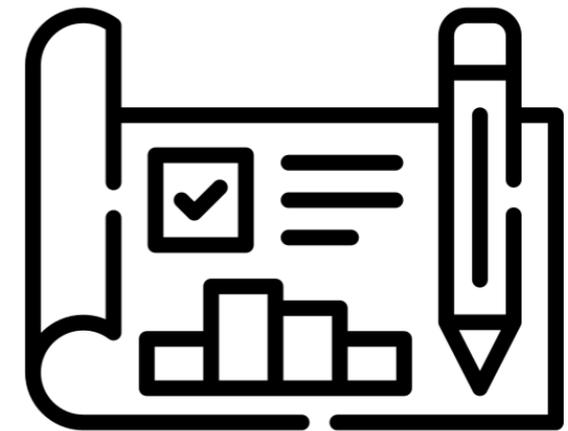
# 어떤 데이터를 수집해야 좋은가요?



많은 양의 데이터



좋은 질의 데이터  
- 최신 데이터 -  
- 문제에 맞는 데이터-



사용할 수 있는 형태의 데이터

# 데이터의 의미를 알아보자

## : 데이터의 특성(Feature), 속성, 변수

데이터의 의미를 가지고 있는 개별 요소 및 최소 단위

### 데이터 속성

종류	이름	품종	성별	귀여움	귀 길이	눈 색깔	먹는 양	개월
강아지	정남이	진돗개	여	4	5.1cm	갈색	3kg	33
강아지	뽀삐	시바견	여	3	4.2cm	검은색	2.5kg	24
고양이	냐옹이	페르시안	남	5	3.6cm	파란색	0.5kg	8

# 데이터 속성의 종류

## : 연속형 vs 범주형 vs 순서형

종류	이름	품종	성별	귀여움	귀 길이	눈 색깔	먹는 양	개월
강아지	정남이	진돗개	여	4	5.1cm	갈색	3kg	33
강아지	뽀삐	시바견	여	3	4.2cm	검은색	2.5kg	24
고양이	냐옹이	페르시안	남	5	3.6cm	파란색	0.5kg	8

연속형 데이터 => 실수 데이터

범주형 데이터 => 카테고리 데이터

순서형 데이터 => 척도 데이터

# 학습을 위한 데이터간의 관계

## : 독립 변수 vs 종속 변수

$$y = ax$$

입력 값, 원인

독립 변수

종속 변수

결과 값, 효과



# 독립 변수와 종속 변수 어떻게 정할까?

풀려는 문제에 따라 독립 변수와 종속 변수는 바뀜

- 문제 1. 강아지와 고양이에게 적절한 양의 음식 주기
- 문제 2. 강아지와 고양이이 분류하기

종류	이름	품종	성별	귀여움	귀 길이	눈 색깔	먹는 양	개월
강아지	정남이	진돗개	여	4	5.1cm	갈색	3kg	33
강아지	뽀삐	시바견	여	3	4.2cm	검은색	2.5kg	24
고양이	냐옹이	페르시안	남	5	3.6cm	파란색	0.5kg	8

# 두가지 종류의 데이터!

## : 정형 데이터 vs 비정형 데이터

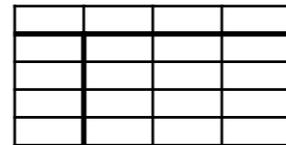
### 정형 데이터

형태가 정해져서 **행렬로 표현 가능**

숫자, 날짜, 문자 등의 데이터

상대적으로 작은 저장 공간

데이터 관리와 분석이 쉬움




데이터 형태



데이터 관리

### 비정형 데이터

형태가 불규칙하여 **행렬로 표현 불가능**

이미지, 음성, 비디오, 문서 등의 데이터

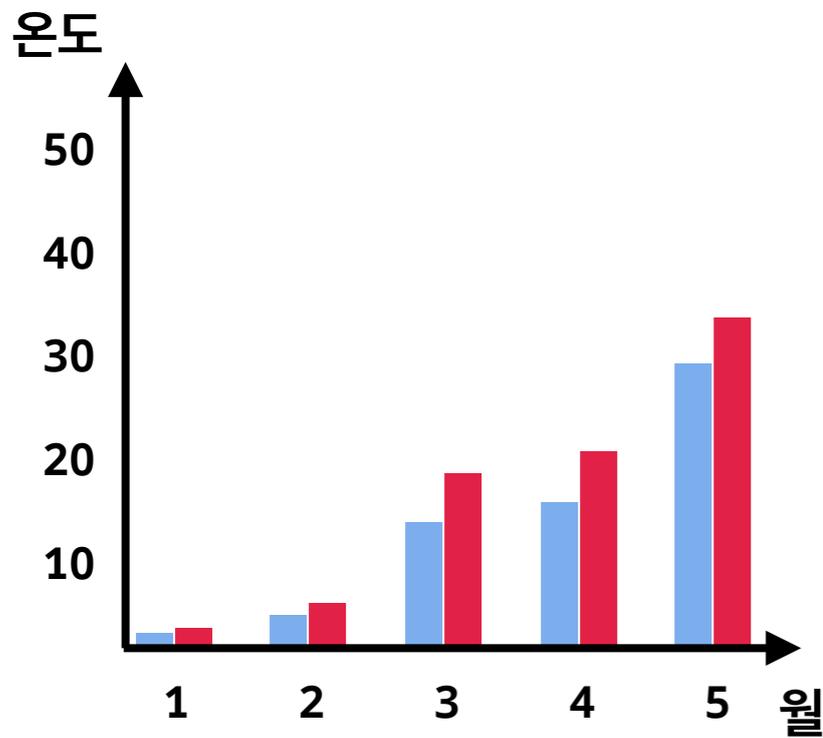
상대적으로 큰 저장 공간

데이터 관리와 분석이 어려움

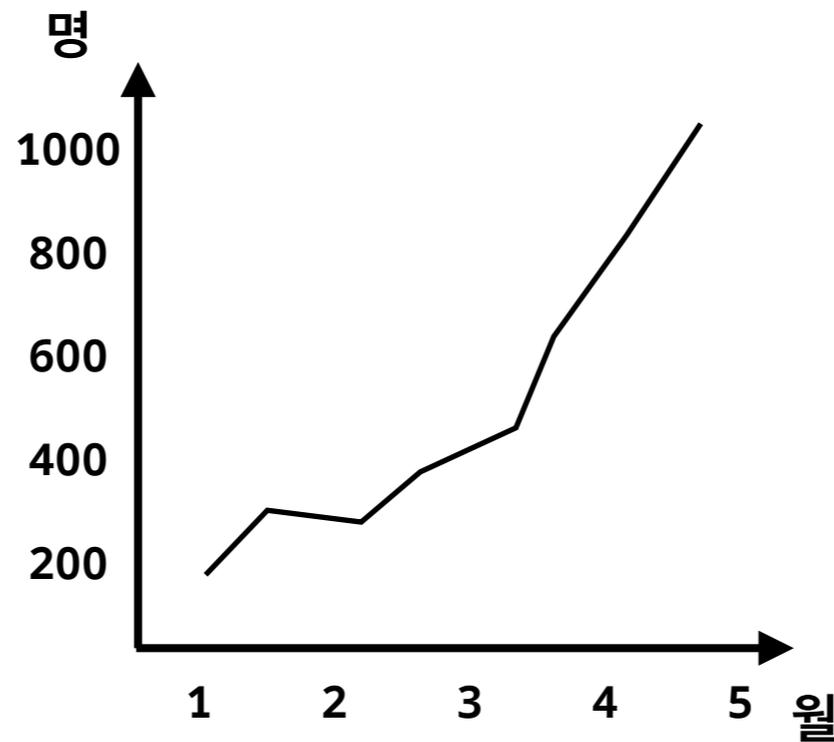
# 데이터를 이해하기 위한 시각화 방법

## : 막대 vs 선 vs 산점도

지역별 온도 변화



A게임 동시접속자



공부시간별 성적

