

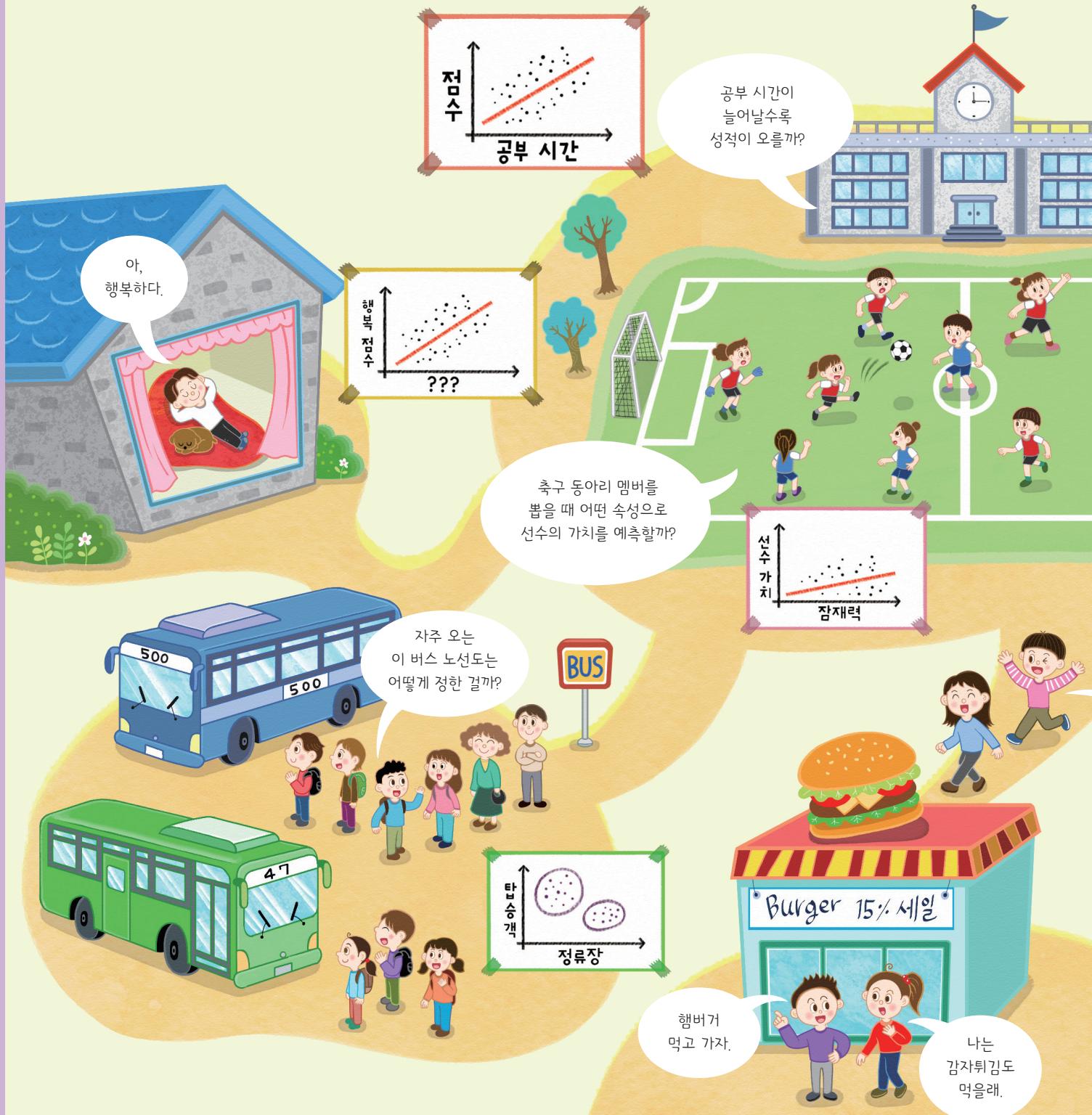


생각에
날개 달기

데이터 모델링은 어디에, 어떻게 사용될까?



영상으로 확인해요!
[https://dtgr.kr/
VwVHzA](https://dtgr.kr/VwVHzA)



데이터 모델링을 통한 분석은 일상 생활에서도 쉽게 찾아볼 수 있다. 문제 상황에 따라 회귀 분석, 군집 분석, 연관 분석 등 다양한 데이터 모델링으로 데이터를 분석하는 과정을 배워 보자.



01

데이터 모델의 개념

학습 목표

- 데이터 모델의 개념을 이해하고 사례를 알 수 있다.
- 데이터 분석 과정을 이해하고, 데이터 분석 도구의 종류를 알 수 있다.

학습 요소

- 데이터 모델, 데이터 분석

1. 데이터 모델의 이해

데이터 모델은 데이터를 이해하는 방법을 제공한다.

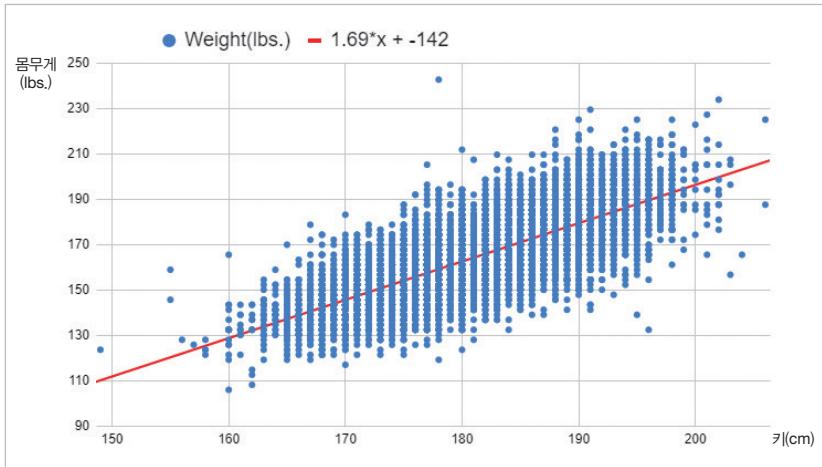
① 데이터 모델

데이터 모델은 사람이 세상을 이해하는 간소화된 표현 방식으로, 사용되는 분야에 따라서 수식이나 그림 등 다양한 형태로 표현할 수 있다.

데이터 모델은 주어진 데이터에서 속성들의 관계나 패턴을 파악하여 모델로 표현하므로 간소화된 형태로 데이터를 파악할 수 있다. 데이터 모델을 사용하면 주어진 문제를 합리적으로 해결하는 데 도움을 준다.



데이터 모델링은 기계학습 및 통계 분석을 포함한 다양한 기법을 사용하여 데이터를 이해하고 예측하기 위한 모델을 만드는 과정을 의미한다. 주로 예측, 분류, 군집화 등의 작업에 중점을 둔다. 문제에 따라서 속성을 독립 변수와 종속 변수로 나누어 분석할 수 있다. [☞ Link 23쪽](#)



축구 선수는 키와 몸무게는 상관관계가 있을까요?



[그림 III-2] '피파 23 게임 선수'들의 키와 몸무게의 상관관계를 나타낸 산점도

피파 23 게임 선수들은 키와 몸무게의 상관관계를 분석하기 위한 데이터 모델을 [그림 III-2]의 그래프에서 빨간색 선을 $y = ax + b$ 형태의 식으로 표현 할 수 있다. 독립 변수인 키가 증가함에 따라 종속 변수인 몸무게가 늘어나는 점을 보아, 키와 몸무게 간의 상관관계가 있음을 확인할 수 있다.



2 데이터 모델의 구분

데이터 모델은 목적에 따라 회귀 모델, 분류 모델, 군집 모델, 연관 모델 등으로 구분할 수 있다.

[표 III-1] 데이터 모델의 구분

구분	설명	시각화 예시
회귀 모델	<ul style="list-style-type: none"> 독립 변수와 종속 변수의 관계를 나타내는 방법을 제공해주는 모델이다. 과거의 데이터를 이용하여 모델을 구축하고, 이를 토대로 새로운 데이터에 대한 미래의 값을 예측할 수 있다. 통계적 혹은 기계학습 등의 분석 방법을 사용하여 만들 수 있다. 	
분류 모델	<ul style="list-style-type: none"> 입력 데이터를 기준으로 서로 다른 범주 또는 클래스로 분류하는 모델이다. 구축된 모델로 새로운 데이터에 대해 범주를 예측할 수 있다. 로지스틱 회귀, 의사결정트리, 인공신경망 등의 분석 방법을 사용하여 만들 수 있다. 	
군집 모델	<ul style="list-style-type: none"> 데이터들을 유사한 속성을 갖는 군집으로 나누는 모델이다. 이전에는 알지 못했던 새로운 사실을 발견할 수 있다. k-평균 군집 분석을 사용하여 만들 수 있다. 	
연관 모델	<ul style="list-style-type: none"> 속성 간에 존재하는 연관 규칙을 찾는 모델이다. 연관 분석 기법을 활용하여 만들 수 있다. 	

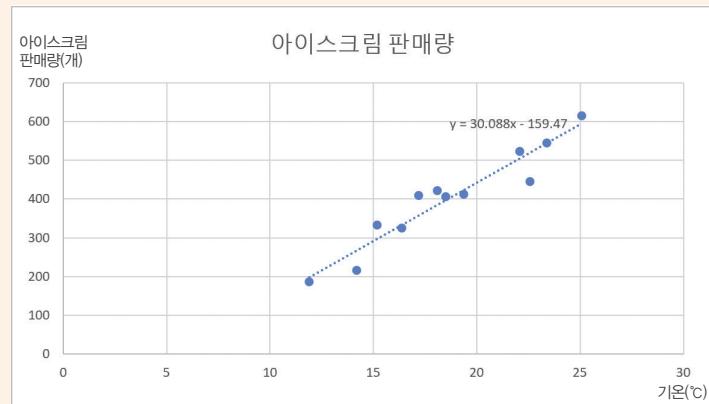
해 보기

아이스크림 판매량 데이터로 기온에 따른 판매량 예측하기



기온에 따른 아이스크림 판매량에 대해 다음과 같은 데이터 모델을 생성했다. 데이터 모델로 아래의 문제에 답해 보자.

기온(°C)	아이스크림 판매량(개)
14.2	215
16.4	325
11.9	185
15.2	332
18.5	406
22.1	522
19.4	412
25.1	614
23.4	544
18.1	421
22.6	445
17.2	408



1. 아이스크림 판매량을 예측하기 위해 선택한 독립 변수와 종속 변수는 각각 무엇인가?

독립 변수	종속 변수

2. 기온이 10°C일 때, 아이스크림의 판매량을 산점도에 표시된 다음 수식을 이용하여 예측해 보자.

$$y = 30.088x - 159.47$$

3. 기온이 30°C일 때, 아이스크림의 판매량을 산점도에 표시된 다음 수식을 이용하여 예측해 보자.

$$y = 30.088x - 159.47$$

위와 같이 데이터 모델을 이용하면 데이터에 있지 않은 새로운 독립 변수에 해당하는 종속 변수의 값을 예측하는 것이 가능하다.

2. 데이터 분석

상관관계

- 두 속성이 얼마나 관련되어 있는지를 말한다.
- 상관관계의 강약을 나타내는 상관계수는 $-1 \sim 1$ 의 사이의 값을 가지며, 양의 상관관계는 두 속성이 같은 방향으로 움직이고, 음의 상관관계는 그 반대이다.

인과 관계

한 속성이 다른 속성에 영향을 미치는 것을 말한다.

연관 관계

두 변수 또는 두 속성 간의 연결이나 관계를 말한다.

데이터 분석은 통계적 또는 기계학습 방법으로 데이터로부터 유용한 정보를 얻는 방법이다.

1 데이터 분석의 개념

데이터 분석은 데이터 모델링 과정에서 다양한 데이터 분석 방법을 사용하여 데이터로부터 의미 있는 정보를 추출한다. 속성 간의 상관관계, 인과 관계, 연관 관계 등을 찾아 데이터 기반의 합리적인 의사 결정을 한다.



[그림 III-3] 상관관계, 인과 관계, 연관 관계의 구분

데이터 분석은 데이터의 특성과 목적에 따라 다양하며, 목적에 맞는 데이터 분석 방법을 선택할 수 있어야 한다. 또한 데이터 분석으로 만들어진 데이터 모델의 성능을 평가하고 새로운 분석 방법을 선택하는 과정을 반복해야 한다.

2 데이터 분석의 종류

데이터 모델을 이용한 데이터 분석 방법은 회귀 분석, 군집 분석, 연관 분석 등이 있다.

[표 III-2] 데이터 분석의 종류

종류	특징							
회귀 분석	<ul style="list-style-type: none"> 주어진 데이터에서 독립 변수와 종속 변수 사이의 관계를 가정하고, 이것을 가장 잘 표현하는 수식을 구하는 분석 방법이다. 새로운 독립 변수가 주어질 때 종속 변수의 값을 예측할 수 있다. <p>예 주택 가격 예측, 주가 예측, 판매 예측 등</p>							
군집 분석	<p>데이터의 유사성을 측정하여 n개의 군집으로 나누는 분석 방법이다.</p> <p>예 세분화, 이상 탐지, 분리 등</p>							
연관 분석	<p>데이터 내부에 존재하는 속성 간의 관계를 찾아내는 분석 방법으로, '장바구니 분석'이라고도 한다.</p> <p>예 상품 추천, 상품 진열 등</p> <table border="1"> <tr> <td>우유를 산 고객이</td> </tr> <tr> <td>식빵을 함께 구매한 확률</td> <td>80%</td> </tr> <tr> <td>달걀을 함께 구매한 확률</td> <td>60%</td> </tr> <tr> <td>휴지를 함께 구매한 확률</td> <td>45%</td> </tr> </table>	우유를 산 고객이	식빵을 함께 구매한 확률	80%	달걀을 함께 구매한 확률	60%	휴지를 함께 구매한 확률	45%
우유를 산 고객이								
식빵을 함께 구매한 확률	80%							
달걀을 함께 구매한 확률	60%							
휴지를 함께 구매한 확률	45%							