데이터 과학 길라잡이

교사용

교육부 한국과학창의재단

사이트 이용 안내

원활한 학습을 위한 학습 전 유의 사항 및 사이트 이용 안내 입니다.
아래 내용을 반드시 숙지하신 후 학습 참여 부탁드립니다.

학습 방법

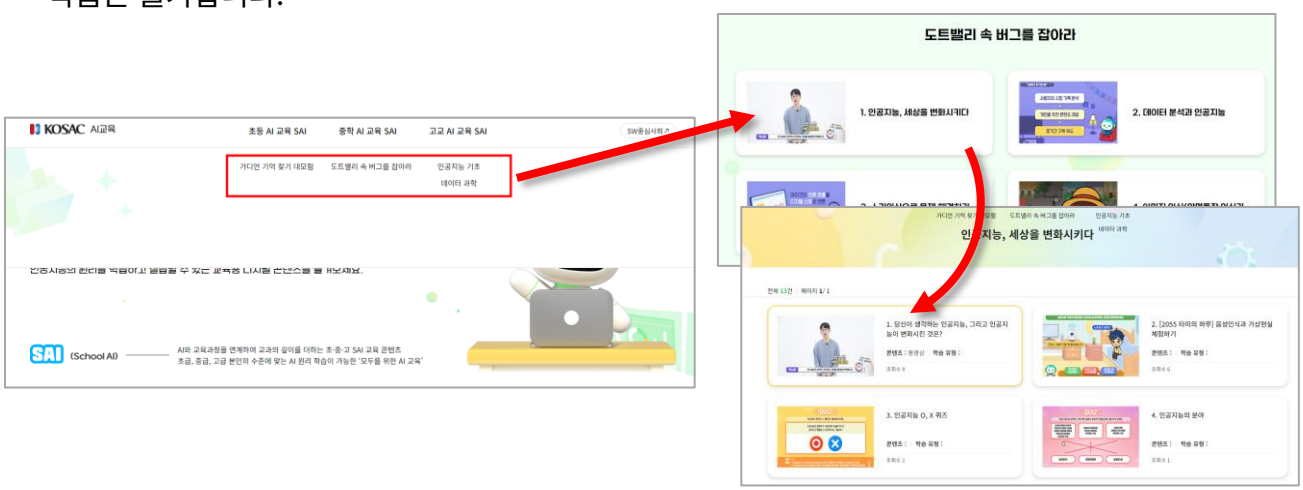
수강 완료 후 이수증 출력이 가능한 [학습 도장 모으기] 학습하기

[학습 도장 모으기] 버튼을 클릭하여 학습 시 수강 후 수강생의 이름이 적힌 이수증 출력이 가능합니다.
수강 방법은 홈페이지 메인 중앙의 학년별 배너를 클릭 후 [학습 도장 모으기] 버튼을 클릭하여 학습을 진행하셔야 이수증 출력이 가능합니다.

※ [학습 도장 모으기] 버튼 클릭 외 다른 경로로 콘텐츠 학습 시 이수증 출력이 불가할 수 있으니 이수증 출력이 필요한 경우 반드시 해당 경로로 학습을 진행해 주시기 바랍니다.
※ [학습 도장 모으기]으로 학습 중 학습 미완료 상태로 학습창을 종료할 경우 처음부터 재학습하셔야 하니 유의하셔서 학습에 참여해 주시기 바랍니다.
※ 이수증은 학습페이지 마지막 단계에서 진행되며 이수증 출력을 위해 수강생의 이름을 정확히 입력해주세요.
※ 학습창을 종료한 콘텐츠의 이수증 재발급은 불가하며 발급이 필요한 경우 처음부터 학습을 진행해야 하므로 이수증 출력 시 PDF 파일로도 저장하여 보관해주시기 바랍니다.
※ [학습 도장 모으기]으로 학습 시 학습창 하단의 영상 재생바 조작(재생바 이동)이 불가하므로 정배속으로 순차 학습하셔야 합니다.

학습 제한없이 자유롭게 학습하기

이수증 출력 없이 영상만 시청을 원할 경우 홈페이지 상단의 위치한 메뉴에서 각 학년별 과정명을 클릭하여 학습하시면 영상 재생바 이동 등 단순 시청이 가능합니다. 다만 해당 경로로 학습할 경우 학습을 모두 완료하였어도 이수증 출력은 불가하며 영상을 재학습하여도 이전 학습한 부분부터 이어서 학습은 불가합니다.



사이트 이용 안내

원활한 학습을 위한 학습 전 유의 사항 및 사이트 이용 안내 입니다.
아래 내용을 반드시 숙지하신 후 학습 참여 부탁드립니다.

학습 시작 전 유의 사항

본 사이트는 회원가입 및 로그인 없이 바로 콘텐츠 학습이 가능합니다.

해당 사이트는 별도의 회원가입 절차가 없고 바로 콘텐츠 학습이 가능합니다.
다만 로그인이 없기 때문에 학습에 대한 기록이 남지 않아 학습 중단 후 학습창 이탈 시 이전에 학습한 영상을 이어서 하거나 이수증 재출력이 불가하오니 종료 전 반드시 확인 후 종료해 주세요.

학습창을 종료하고 재접속 시 이어서 학습은 불가합니다.

개인의 학습 이력 관리가 없기 때문에 학습창 이탈 후 재접속 시 이전 학습에서 이어서 학습이 불가합니다. 이미 수강이 완료된 콘텐츠도 재접속 시 처음부터 새롭게 학습이 진행됩니다.
[학습 도장 모으기]의 콘텐츠를 학습할 경우 영상의 재생바 조작이 불가하며 이미 학습을 완료한 차시도 재생바 조작이 불가합니다.

이수증 출력을 원하시는 경우 반드시 [학습 도장 모으기] 버튼을 클릭하여 학습해주세요.

해당 사이트에서는 이수증을 출력할 수 있는 학습 방법과 단순 영상만 시청하는 학습 방법이 있습니다.
이수증 출력을 원하시는 경우 홈페이지 메인 화면에서 학년별 배너를 클릭 후 [학습 도장 모으기] 버튼을 클릭하여 학습하셔야 학습 완료 후 이수증 출력이 가능합니다.
다만 로그인 없이 학습하기 때문에 학습 이탈 시 이어서 학습하기가 불가하여 처음부터 학습을 진행해야 하니 [학습 도장 모으기]으로 학습하는 경우 반드시 끝까지 학습을 완료하여 이수증을 출력하고 종료해 주시기 바랍니다.

※ 이수증 재발급이 필요한 경우 처음부터 학습을 재시작하셔야 합니다.
※ 이수증 인쇄 전 이수증 내 입력한 이름 정보가 틀리지 않았는지 반드시 확인 후 출력해주세요.

학습은 데스크톱, 노트북, 태블릿PC 기기에서 학습해주세요.

본 콘텐츠는 화면을 클릭하여 진행해야 하는 다양한 상호작용이 있는 콘텐츠로 모바일 학습 시 원활한 학습이 불가합니다. 데스크톱, 노트북, 태블릿 PC를 통해 학습해 주시기 바랍니다.

무엇을 도와 드릴까요?

대표전화 1522-6841

문의메일 ai4school@kosac.re.kr

*평일 9시~18시 | 점심시간 12시~13시 (일요일/공휴일 휴무)

목차

1. 전체 콘텐츠 명

데이터과학

2. 콘텐츠 개요

◆ 과정 목차

모듈	순서	차시명	주제명	유형	소요시간
데이터 과학의 이해	1	데이터과학으로 여는 세계	데이터 기반 의사 결정	영상형	8분
	2	데이터의 속성	다양한 데이터의 속성	영상형	6분
			인공지능(AI) 학습을 위한 데이터 속성 분류하기	실습형	10분
			정형데이터와 비정형 데이터 구분하기	실습형	10분
	3	데이터셋과 데이터베이스	데이터셋과 데이터베이스 기초	영상형	6분
			데이터셋과 데이터베이스	실습형	10분
데이터 준비와 분석	4	데이터 수집의 중요성	데이터 수집방법 알아보기	영상형	7분
			데이터 수집의 중요성	실습형	10분
	5	데이터 전처리	결측치, 이상치, 정규화	영상형	6분
			데이터 전처리	실습형	10분
	6	데이터 시각화	데이터 시각화 특징과 그래프 종류	영상형	7분
			데이터 시각화	실습형	10분
	7	데이터 상관관계	데이터 시각화를 통한 상관관계 이해하기	영상형	4분
			데이터 상관관계	실습형	10분
	8	데이터 분석방법	데이터 분석방법 알아보기	영상형	4분
			건강데이터 살펴보기	실습형	15분
			건강데이터로 비만 분포 확인하기	실습형	15분
데이터 모델링과 평가	9	데이터 모델링 및 데이터 분석 도구 탐색하기	다양한 분석 도구 살펴보기	영상형	6분
	10	단순 선형 회귀, 다중 선형 회귀	선형 회귀의 기본 이해	영상형	6분
			회귀분석으로 예측하기	실습형	10분
	11	회귀 분석, 결정 계수	R^2 값 알아보기	영상형	6분
			회귀분석과 결정계수	실습형	10분
	12	군집화 개념 이해하기	보로노이 다이어그램과 센트로이드	영상형	5분
			SNS 사용 데이터를 활용한 군집화 알아보기	실습형	10분
	13	군집분석	포켓몬빵 사례로 알아보는 군집분석	영상형	7분
			군집분석	실습형	10분
	14	지지도, 신뢰도, 향상도	연관분석 기초	영상형	6분
			지지도, 신뢰도, 향상도	실습형	10분
	15	장바구니 분석	장바구니 분석 이해하기	영상형	4분
			장바구니 분석	실습형	5분

◆ 과정 구성

영상형 콘텐츠 15개, 실습형 콘텐츠 15개 (총 30개 콘텐츠)

1) 영상형

데이터과학의 이해 모듈(3개), 데이터 준비와 분석 모듈(5개),
데이터 모델링과 평가 모듈(7개)

❖ **구성:** 지식영상 & 캐릭터 애니메이션 & 인터뷰 영상

2) 실습형

데이터과학의 이해 모듈(3개), 데이터 준비와 분석 모듈(6개),
데이터 모델링과 평가 모듈(6개)

◆ 추천 학습 환경

실습형 콘텐츠 오류 발생 시 대처 방법

- 오류 발생 이유: 실습환경은, 사용하시는 기기의 메모리를 활용하여 진행할 수 있도록 구성되어 있습니다. 실습 진행 시, 기기의 메모리를 초과하는 실습을 진행하거나(다수의 버튼을 클릭하여 메모리 실행 초과 등) 기기의 메모리가 현저히 낮다면, 메모리 관련 오류가 발생할 수 있습니다.
- 오류 발생 해결 방법: 실습 진행 시, 기기의 메모리를 소모할 수 있는 다른 프로그램을 종료하여 최적의 실습 환경으로 구성해 주시기를 바랍니다. 이러한 조치에도 오류가 발생하였다면, 새로고침 혹은 이전/다음 페이지를 눌러 실습 환경을 초기화 후 다시 시도해 보시기 바랍니다.

실습형 실행 환경

- 해상도: 실습 콘텐츠는 해상도 1680 * 1050을 기준으로 제작되어 있습니다. 필요시 해상도를 조정해 주세요.
- 실행 환경: 실습 콘텐츠는 PC 환경 및 Chrome, Edge 브라우저 환경에 최적화되어 있습니다.

3. 세부 콘텐츠

◆ 개요

01차시. 데이터과학으로 여는 세계

성취기준, 학습목표, 학습내용, 개발유형

성취기준	[12데과01-01] 데이터과학의 개념을 이해하고, 문제 해결 사례를 데이터 기반 의사 결정 상황에 적용한다. [12데과01-04] 데이터로 인한 사회 변화를 인식하고, 진로 및 직업과 관련한 데이터 기반 문제 해결 사례를 분석한다.
학습 목표	• 데이터과학의 개념을 이해하고 데이터 기반 의사 결정의 사례를 설명할 수 있다. • 데이터로 인한 사회 변화를 이해하고 진로 및 직업과 관련한 데이터 기반 문제 해결 사례를 설명할 수 있다.
대상 학년	
연계교육과정	정보교과
세부콘텐츠	1. (강의실) 동영상 콘텐츠

◆ 구성

강의실

- ❖ **주제명:** 데이터 기반 의사 결정
- ❖ **세부 주제**
  1) 데이터과학의 개념
  2) 데이터 기반 의사 결정을 함으로써 문제를 해결하는 사례와 사회적 변화 안내
- ❖ **콘텐츠 개발 목적:** 현직자 인터뷰를 통해 데이터과학의 필요성을 전달한다.
- ❖ **선생님을 위한 팁!**
  추가적인 데이터 기반 의사 결정 사례 예시를 다루어 볼 수 있다.

3. 세부 콘텐츠

◆ 개요

02차시. 데이터의 속성

성취기준, 학습목표, 학습내용, 개발유형

성취기준	[12데과01-02] 정형 데이터와 비정형 데이터를 구분하고, 데이터 속성에서 데이터의 잠재적 가치를 파악한다.
학습 목표	• 정형 데이터와 비정형 데이터의 개념을 이해하고 특징을 설명할 수 있다. • 데이터 속성이 데이터과학 분야에서 갖는 잠재적 가치를 설명할 수 있다.
대상 학년	
연계교육과정	정보교과
세부콘텐츠	1. (강의실) 동영상 콘텐츠 2. (실험실) 실습형 콘텐츠

◆ 구성

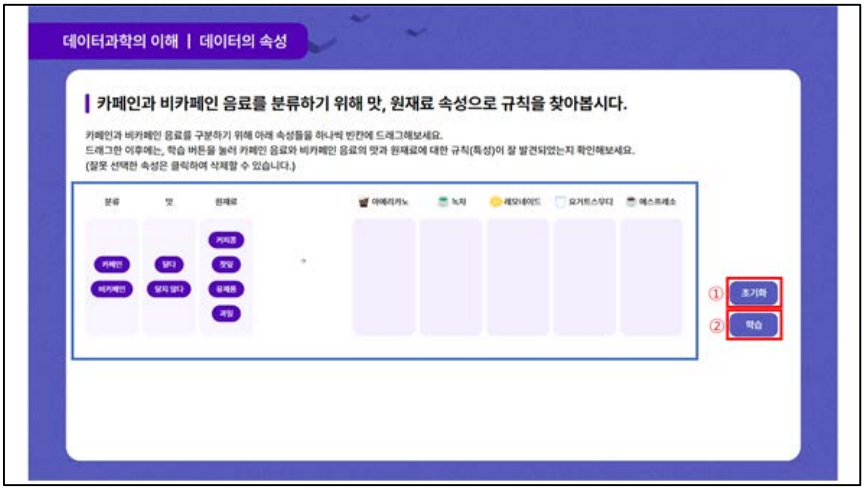
강의실

- ❖ **주제명:** 다양한 데이터의 속성
- ❖ **세부 주제**
  1) 정형 데이터와 비정형 데이터
  2) 범주형 데이터와 수치형 데이터
- ❖ **콘텐츠 개발 목적:** 같은 데이터라도 목적에 따라 속성들의 중요성이 달라질 수 있음을 이해할 수 있다.
- ❖ **선생님을 위한 팁!**
  정형 데이터와 비정형 데이터의 구조에 대해 추가 설명이 필요할 수 있다.

실험실

- ❖ **주제명:** 데이터의 속성
- ❖ **세부 주제**
  1) 데이터 속성 분류의 중요성
  2) 인공지능(AI)을 학습시키기 위한 데이터 속성 분류 체험
- ❖ **콘텐츠 개발 목적:** 인공지능(AI)을 만들 때 데이터 속성을 분류하는 과정을 단계별로 실습하며 중요성을 인식

❖ 실습 콘텐츠 안내:



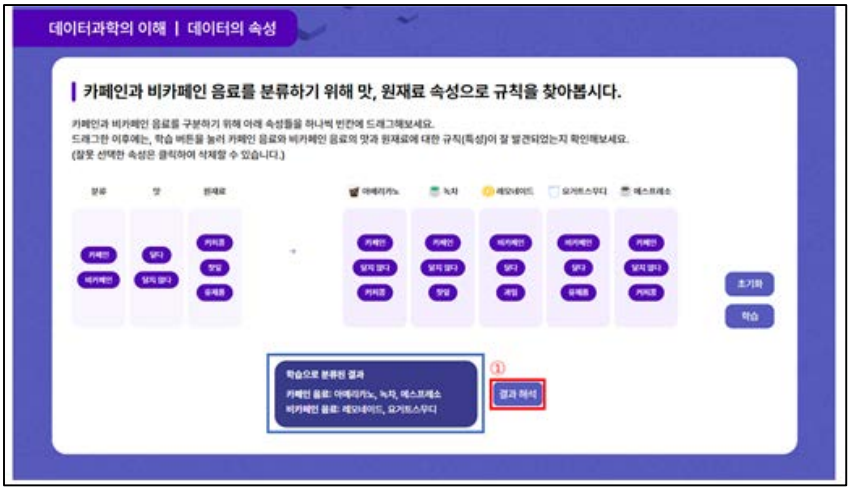
Step 1

분류, 맛, 원재료 속성을 각 음료 영역으로 드래그하여 카페인/비카페인 음료 구분을 위한 속성을 인공지능에게 학습시키기

기능 사용법

① 초기화 버튼: 각 속성을 드래그 앤 드롭으로 분류하는 과정에서 초기화 버튼을 누르면 분류 전 처음 상태로 돌아갈 수 있습니다. (전체 초기화를 하지 않고 특정 속성만 지우고 싶을 경우, 클릭하여 지울 수 있습니다.)

② 학습 버튼: 모든 분류가 끝나면, 분류한 결과로 인공지능(AI)을 학습시킬 수 있는 버튼입니다.



Step 2

학습된 인공지능(AI)의 음료 분류 결과를 확인

기능 사용법

① 결과 해석 버튼: 인공지능(AI)이 학습한 결과를 토대로 분류된 결과에 대한 설명을 볼 수 있습니다.

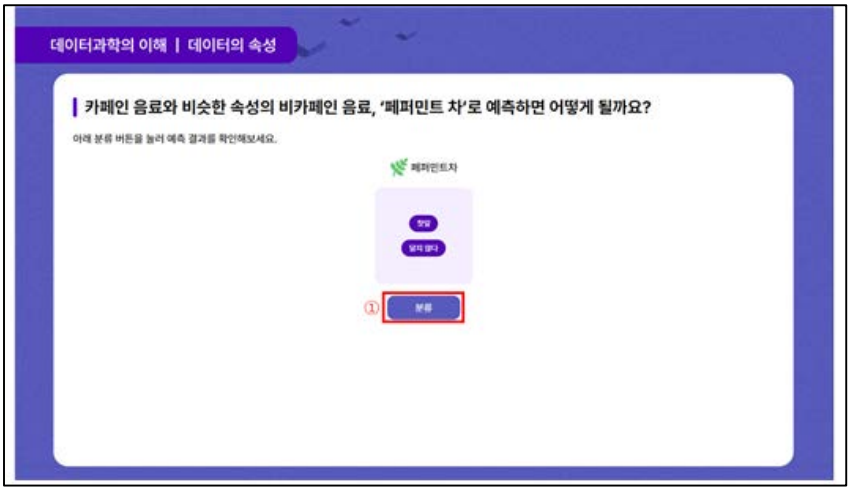
❖ 실습 콘텐츠 안내:

Step 3

기존에 분류해 둔 속성들을 기준으로, 다른 음료들의 분류 결과가 어떻게 나올지 예측해 본 뒤 분류 버튼을 눌러 확인하기

기능 사용법

① 분류 버튼: 인공지능(AI)을 학습시킨 분류 규칙으로 음료를 분류하는 버튼입니다. 버튼을 누르면 분류 결과와 설명을 확인할 수 있습니다.



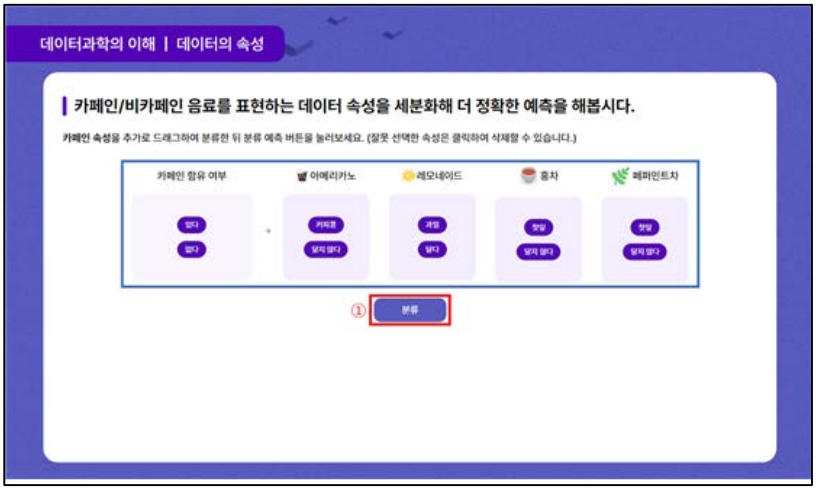
Step 4

기존에 학습된 결과로 분류할 수 없는 예외 케이스를 분류 버튼을 눌러 확인

기능 사용법

① 분류 버튼: 인공지능(AI)을 학습시킨 분류 규칙으로 음료를 분류하는 버튼입니다. 버튼을 누르면 분류 결과와 설명을 확인할 수 있습니다.

❖ 실습 콘텐츠 안내:



Step 5

예외 케이스까지 포함할 수 있는 속성을 드래그 앤 드롭으로 추가 분류 하고, 분류 버튼을 눌러 결과 확인하기

① 분류 버튼: 인공지능(AI)을 학습시킨 분류 규칙으로 음료를 분류하는 버튼입니다. 버튼을 누르면 분류 결과와 설명을 확인할 수 있습니다.

주의 분류 실습 시, 아이콘을 너무 빠르게 움직이면 아이콘이 없어지는 경우가 있으므로 천천히 움직여주세요.

❖ **선생님을 위한 팁!**

인공지능에 학습시킬만한 주제, 그에 따른 속성들을 추가적으로 탐구해보는 시간을 가지면 원리 학습에 더욱 도움이 됩니다.

3. 세부 콘텐츠

◆ 개요

02차시. 데이터의 속성

성취기준, 학습목표, 학습내용, 개발유형

성취기준	[12데과01-02] 정형 데이터와 비정형 데이터를 구분하고, 데이터 속성에서 데이터의 잠재적 가치를 파악한다.
학습 목표	• 정형 데이터와 비정형 데이터의 개념을 이해하고 특징을 설명할 수 있다. • 데이터 속성이 데이터과학 분야에서 갖는 잠재적 가치를 설명할 수 있다.
대상 학년	
연계교육과정	정보교과
세부콘텐츠	1. (강의실) 동영상 콘텐츠 2. (실험실) 실습형 콘텐츠

◆ 구성

강의실

- **주제명:** 다양한 데이터의 속성
- **세부 주제**
  1) 정형 데이터와 비정형 데이터
  2) 범주형 데이터와 수치형 데이터
- **콘텐츠 개발 목적:** 같은 데이터라도 목적에 따라 속성들의 중요성이 달라질 수 있음을 이해할 수 있다.
- **선생님을 위한 팁!**
  정형 데이터와 비정형 데이터의 구조에 대해 추가 설명이 필요할 수 있다.

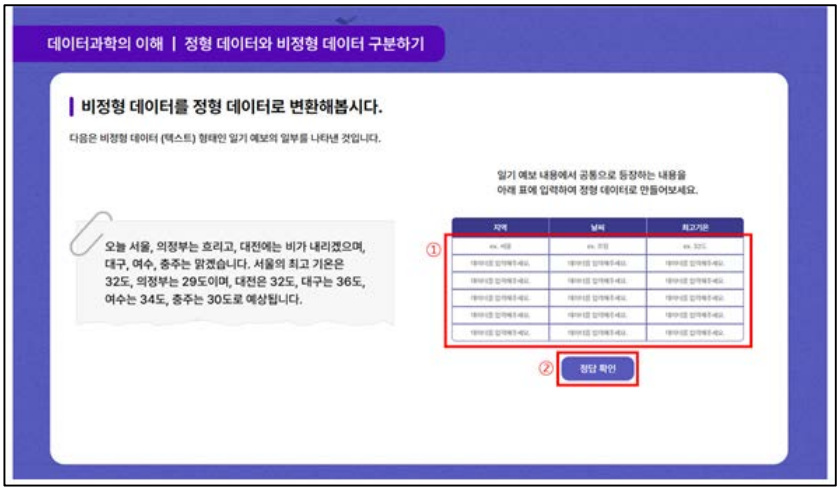
실험실

- **주제명:** 정형 데이터와 비정형 데이터 구분하기
- **세부 주제**
  1) 정형/비정형 데이터를 구분
  2) 비정형 데이터를 정형 데이터로 바꾸어 표현
  3) 정형 데이터 내에서 문제를 해결하기 위해 어떤 데이터 속성이 필요한지 결정
- **콘텐츠 개발 목적:** 분석을 위한 데이터의 유형 및 특성 이해

❖ 실습 콘텐츠 안내:

Step 1

선택지에 있는 데이터 유형을 정형/비정형 데이터로 각각 분류



Step 2

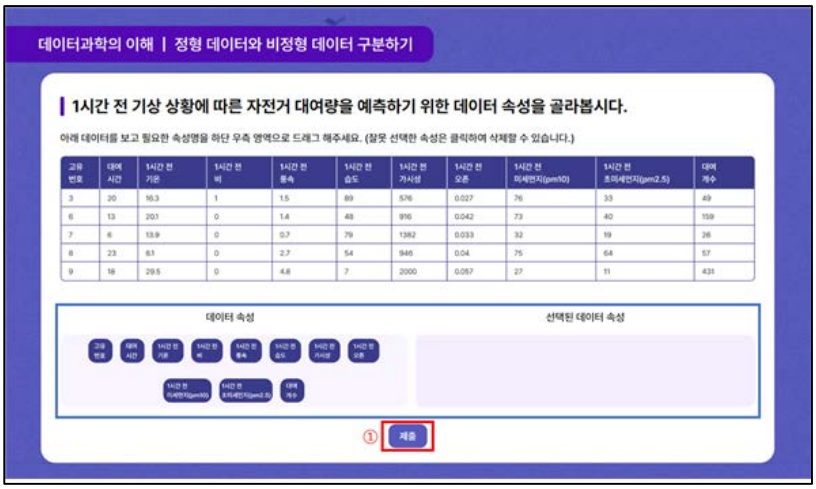
비정형 데이터 (텍스트) 형태인 왼쪽의 일기예보를 오른쪽 정형 데이터 틀에 옮겨 적어 정형 데이터로 만들고,정답 확인 버튼을 눌러 올바르게 변환했는지 확인하기

기능 사용법

① 데이터 입력란: 왼쪽 일기예보를 보고 첫번째 행에 ex로 표시된 정형데이터 형식과 같이 작성합니다.

② 정답 확인 버튼: ①번 입력란을 전부 기입한 뒤, 정답 확인 버튼을 누르면 일기예보를 정형 데이터로 변환했을 때의 정답과 입력한 내용을 비교해 확인할 수 있습니다.

❖ 실습 콘텐츠 안내:



Step 3

예측 조건이 주어졌을 때, 예측에 필요한 데이터 속성을 '선택된 데이터 속성' 영역에 드래그하고 제출 버튼을 눌러 제대로 분류 했는지 확인하기 기능 사용법

① 제출 버튼: 데이터 속성 영역에서, 조건에 맞는 속성을 선택된 데이터 속성으로 드래그 앤 드롭한 후 제출 버튼을 누르면 정/오답 여부를 판별할 수 있습니다.

주의 일기 예보를 정형 데이터로 변환하는 실습에서, 날씨를 글로 작성할 때는 예시와 동일한 형태로 작성해 주세요. 예를 들어, 날씨는 흐림/비/맑음 세 가지로만 작성할 수 있으며, 흐리다/비가온다/맑다 등으로 작성하면 오답 처리 될 수 있습니다.

❖ **선생님을 위한 팁!**

Step 3에 해당하는 실습에서, 미세먼지 수치에 따른 대여량은 어떤 속성을 가지고 예측할 수 있을지 학생들에게 추가 질문을 해보세요.

3. 세부 콘텐츠

◆ 개요

03차시. 데이터셋과 데이터베이스

성취기준, 학습목표, 학습내용, 개발유형

성취기준	[12데과01-03] 데이터셋의 집합인 데이터베이스를 이해하고, 서로 다른 데이터셋의 데이터를 분석이 가능한 형태로 통합하는 것의 의미를 파악한다.
학습 목표	• 데이터셋과 데이터베이스를 이해하고 설명할 수 있다. • 서로 다른 데이터셋의 데이터를 분석 가능한 형태로 통합하고 의미를 설명할 수 있다.
대상 학년	
연계교육과정	정보교과
세부콘텐츠	1. (강의실) 동영상 콘텐츠 2. (실험실) 실습형 콘텐츠

◆ 구성

강의실

- ❖ **주제명:** 데이터셋과 데이터베이스 기초
- ❖ **세부 주제**
  1) 데이터셋과 데이터베이스의 개념
  2) 데이터베이스를 사용하는 것의 장점
  3) 데이터베이스를 사용하는 사례
- ❖ **콘텐츠 개발 목적:** 데이터셋과 데이터베이스의 개념을 이해하고 실생활에 적용할 수 있다.
- ❖ **선생님을 위한 팁!**
  데이터베이스 실생활 예시를 추가 제공할 수 있다.

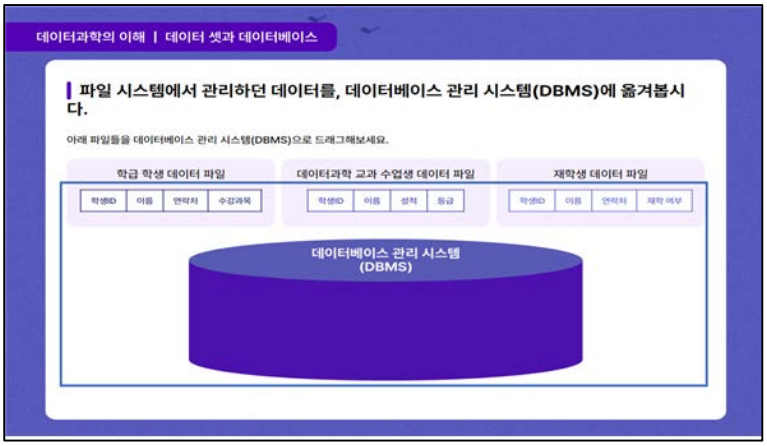
실험실

- ❖ **주제명:** 데이터셋과 데이터베이스
- ❖ **세부 주제**
  1) 파일 시스템 관리 방법 이해
  2) 데이터셋과 DBMS 이해
  3) DBMS로 데이터 관리하는 방법, 장점 이해
- ❖ **콘텐츠 개발 목적:** 데이터셋을 관리하는 방법과, DBMS의 장점을 실습을 통해 이해

❖ 실습 콘텐츠 안내:

Step 1

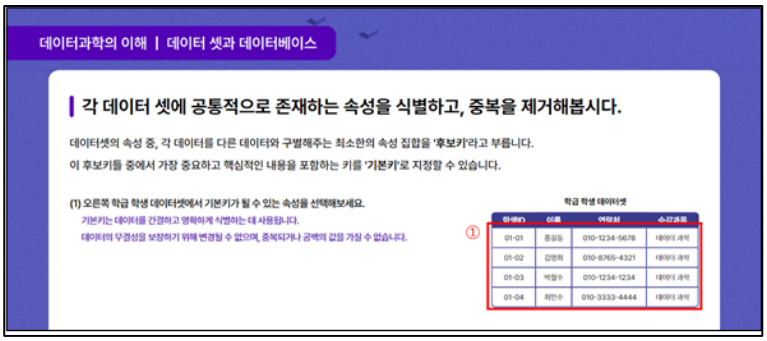
파일 시스템 환경에서, 데이터 변경/추가로 인한 관리 방법을 실습
(초록색으로 표시된 부분에 필요한 내용을 입력하여 수정, 추가)
기능 사용법
① 데이터 입력란: 지시문을 따라 데이터를 추가, 수정하여 입력합니다.
② 제출 버튼: 데이터를 수정/추가한 후 해당 버튼을 누르면 정/오답 여부를 확인할 수 있습니다.



Step 2

파일 시스템으로 관리하던 파일들을 데이터 베이스 관리 시스템 영역으로 드래그하여 옮기기

❖ 실습 콘텐츠 안내:

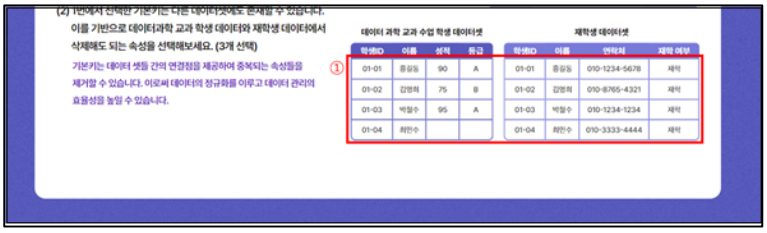


Step 3

DBMS로 옮긴 데이터셋에 공통적으로 존재하는 속성을 식별하고, 중복을 제거하기

기능 사용법

① 속성 선택: 지시문에 해당하는 속성을 클릭하여 선택합니다. 열(학생ID, 이름, 연락처, 수강과목)을 기준으로 데이터를 선택할 수 있습니다. 선택 시 해당 열이 초록색으로 표시되며 정오답 여부를 확인할 수 있습니다. 올바른 속성을 선택할 경우 (2)번 문제가 화면에 표시됩니다.



Step 4

기본키로 연결된 데이터셋에서, 중복되는 속성들을 제거하기

기능 사용법

① 속성 선택: 지시문에 해당하는 속성을 클릭하여 선택합니다. 열(학생ID, 이름, 연락처, 수강과목)을 기준으로 데이터를 선택할 수 있으며, 선택 시 해당 열이 초록색으로 표시되며 정오답 여부를 확인할 수 있습니다.

주의 3, 4 페이지에서 초록색으로 표시된 영역 외의 부분을 수정할 경우 정답 처리에 영향을 줄 수 있습니다.

❖ **선생님을 위한 팁!**

DBMS로 데이터를 관리하려면, 관리할 데이터가 정형 데이터여야 합니다. 이전 차시의 실습에서 배웠던 비정형 데이터를 정형데이터로 변환하는 실습을 다시 언급하며, DBMS로 관리하고 싶은 데이터가 정형데이터가 아닐 경우, 정형화 과정이 필요하다는 점을 알려주시면 더욱 좋습니다.

❖ 실습 콘텐츠 안내:

철수가 수강과목을 변경했을때 데이터를 수정해봅시다.

1. 철수가 수강과목을 인공지능 기초에서 데이터 과학으로 바꾸었습니다.
2. 철수의 데이터 과학 과목 성적은 95, 등급은 A입니다.

학급 학생 데이터

학생ID	이름	연락처	수강과목
01-01	홍길동	010-1234-5678	데이터 과학
01-02	김영희	010-8765-4321	데이터 과학
01-03	박철수	010-1234-1234	데이터 과학

데이터 과학 교과 수업 학생 데이터

학생ID	이름	성적	등급
01-01	홍길동	90	A
01-02	김영희	75	B
01-03	박철수	95	A

재학생 데이터

학생ID	이름	연락처	재학 여부
01-01	홍길동	010-1234-5678	재학
01-02	김영희	010-8765-4321	재학
01-03	박철수	010-1234-1234	재학

최민수 학생이 전학을 왔을때 데이터를 관리해봅시다.

1. 최민수라는 학생이 "전학"을 왔습니다.
2. 최민수 학생의 학생 ID는 01-04, 연락처는 010-3333-4444, 수강과목은 데이터 과학입니다.

학급 학생 데이터

학생ID	이름	연락처	수강과목
01-01	홍길동	010-1234-5678	데이터 과학
01-02	김영희	010-8765-4321	데이터 과학
01-03	박철수	010-1234-1234	데이터 과학
01-04	최민수	010-3333-4444	데이터 과학

데이터 과학 교과 수업 학생 데이터

학생ID	이름	성적	등급
01-01	홍길동	90	A
01-02	김영희	75	B
01-03	박철수	95	A
01-04	최민수		

재학생 데이터

학생ID	이름	연락처	재학 여부
01-01	홍길동	010-1234-5678	재학
01-02	김영희	010-8765-4321	재학
01-03	박철수	010-1234-1234	재학
01-04	최민수	010-3333-4444	재학

(1) 오른쪽 학급 학생 데이터셋에서 기본키가 될 수 있는 속성을 선택해보세요.

기본키는 데이터를 간결하고 명확하게 식별하는 데 사용됩니다.
데이터의 무결성을 보장하기 위해 변경될 수 없으며, 중복되거나 공백의 값을 가질 수 없습니다.

학급 학생 데이터셋

학생ID	이름	연락처	수강과목
01-01	홍길동	010-1234-5678	데이터 과학
01-02	김영희	010-8765-4321	데이터 과학
01-03	박철수	010-1234-1234	데이터 과학
01-04	최민수	010-3333-4444	데이터 과학

(2) 1번에서 선택한 기본키는 다른 데이터셋에도 존재할 수 있습니다.
이를 기반으로 데이터과학 교과 학생 데이터와 재학생 데이터에서 삭제해도 되는 속성을 선택해보세요. (3개 선택)

기본키는 데이터 셋들 간의 연결점을 제공하여 중복되는 속성들을 제거할 수 있습니다. 이로써 데이터의 정규화를 이루고 데이터 관리의 효율성을 높일 수 있습니다.

데이터 과학 교과 수업 학생 데이터셋

학생ID	이름	성적	등급
01-01	홍길동	90	A
01-02	김영희	75	B
01-03	박철수	95	A
01-04	최민수		

재학생 데이터셋

학생ID	이름	연락처	재학 여부
01-01	홍길동	010-1234-5678	재학
01-02	김영희	010-8765-4321	재학
01-03	박철수	010-1234-1234	재학
01-04	최민수	010-3333-4444	재학

3. 세부 콘텐츠

◆ 개요

04차시. 데이터 수집의 중요성

성취기준, 학습목표, 학습내용, 개발유형

성취기준	[12데과02-01] 데이터를 편향되지 않도록 수집하고, 수집된 데이터의 특성을 분석한다.
학습 목표	• 데이터를 수집하는 방법을 이해하고 목적에 맞게 데이터를 수집할 수 있다. • 데이터의 편향성에 유의하며 데이터를 수집할 수 있다.
대상 학년	
연계교육과정	정보교과
세부콘텐츠	1. (강의실) 동영상 콘텐츠 2. (실험실) 실습형 콘텐츠

◆ 구성

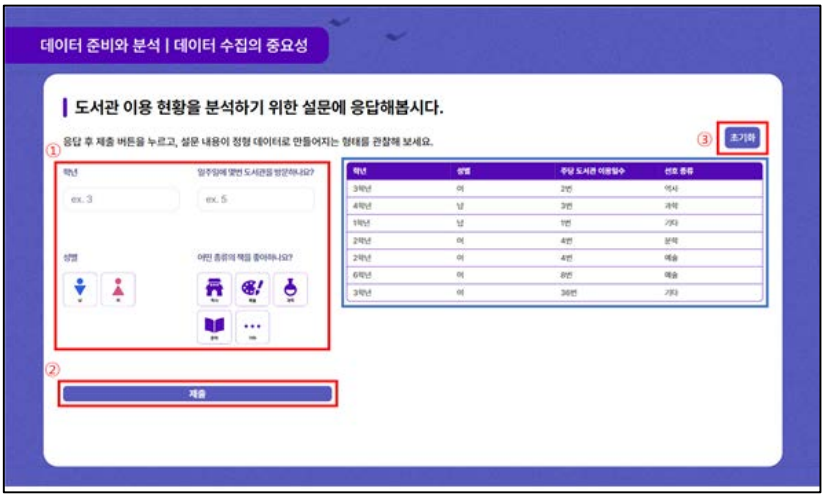
강의실

- ❖ **주제명:** 데이터 수집방법 알아보기
- ❖ **세부 주제**
  1) 데이터셋 직접 구성
  2) 공유 데이터셋 사용
- ❖ **콘텐츠 개발 목적:** 데이터 분석의 기초가 되는 데이터 수집에 대해 이해할 수 있다.
- ❖ **선생님을 위한 팁!**
  좋은 데이터를 선별하는 활동을 추가 구성할 수 있다.

실험실

- ❖ **주제명:** 데이터 수집의 중요성
- ❖ **세부 주제**
  1) 데이터 수집 방법 학습
  2) 수집한 데이터 특성 파악
- ❖ **콘텐츠 개발 목적:** 데이터를 수집하고, 수집한 데이터의 특성을 파악

❖ 실습 콘텐츠 안내:



Step 1

설문지 형태로 데이터를 기입하고, 정형 데이터로 변환되는 과정 관찰

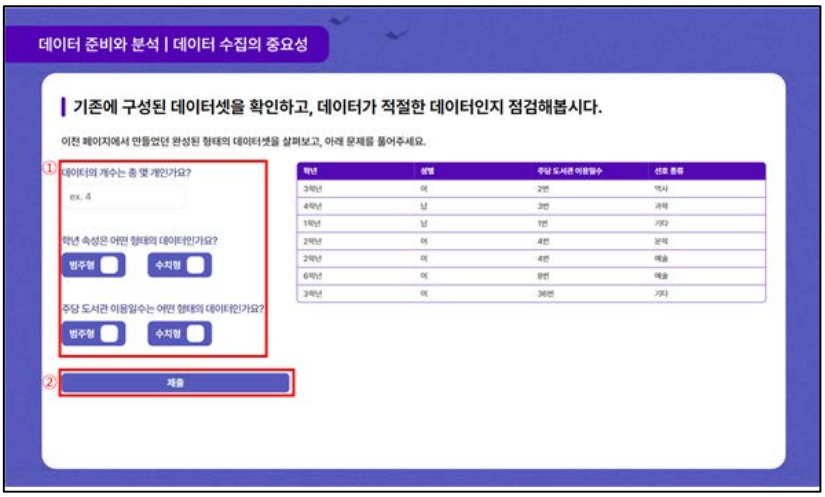
기능 사용법

① 데이터 입력란: 학년 / 도서관 방문 횟수는 숫자를 입력하고, 성별과 좋아하는 책 종류는 아이콘을 선택하여 데이터를 기입합니다.

② 제출 버튼: 해당 버튼을 누르면 ① 데이터 입력란에 입력된 내용을 기반으로, 오른쪽 정형 데이터 형식의 표에 데이터 행이 추가됩니다. (제출은 여러번 할 수 있으며, 응답을 자유롭게 추가하고 제출하여 나만의 도서관 이용 현황 데이터를 만들어보세요.)

③ 초기화 버튼: 초기 데이터 외에, 추가한 데이터를 삭제하고 싶을 경우 해당 버튼을 눌러 한번에 삭제할 수 있습니다.

❖ 실습 콘텐츠 안내:



Step 2

Step 1 (3 페이지)에서 생성한 데이터의 특성을 파악하여 답안을 작성하고, 제출 버튼을 눌러 정답 확인

기능 사용법

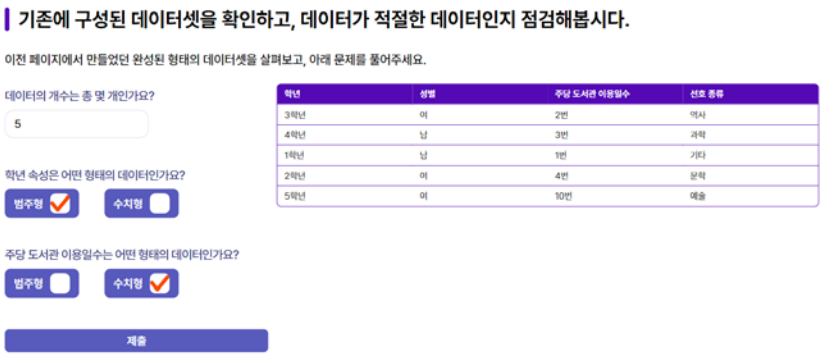
① 데이터 입력란: 데이터 개수에는 숫자 형식, 학년 속성 및 주당 도서관 이용 일수의 형태는 체크 형식으로 답안을 작성해 주세요.

② 정답 확인 버튼: ①번 입력란을 전부 기입한 뒤, 제출 버튼을 누르면 해당 답안에 대한 정/오답 여부를 확인할 수 있습니다.

주의 3,4 페이지 실습에서 사용되는 데이터에서, 처음에 주어지는 4개의 데이터는 초기데이터로 고정되어 있으며, 초기화 버튼을 눌러도 수정할 수 없습니다.

❖ **선생님을 위한 팁!**

데이터를 추가하는 3페이지 실습에서, 성별은 여성만 혹은 남성만 선택하는 등의 행위는 데이터 편향이 발생할 수 있음을 함께 안내해주시면 더욱 좋습니다.



3. 세부 콘텐츠

◆ 개요

05차시. 데이터 전처리

성취기준, 학습목표, 학습내용, 개발유형

성취기준	[12데과02-02] 이상치와 결측치 탐색 및 정규화를 통해 전처리하여 오류 가능성을 최소화하고, 데이터 분석을 위해 시각화한다.
학습 목표	• 이상치와 결측치의 개념과 탐색 방법을 이해한다. • 이상치와 결측치의 처리 방법을 이해하고, 적절한 처리 방법을 선택할 수 있다. • 정규화의 개념과 방법을 이해하고, 데이터의 분포를 일정하게 조절할 수 있다.
대상 학년	
연계교육과정	정보교과
세부콘텐츠	1. (강의실) 동영상 콘텐츠 2. (실험실) 실습형 콘텐츠

◆ 구성

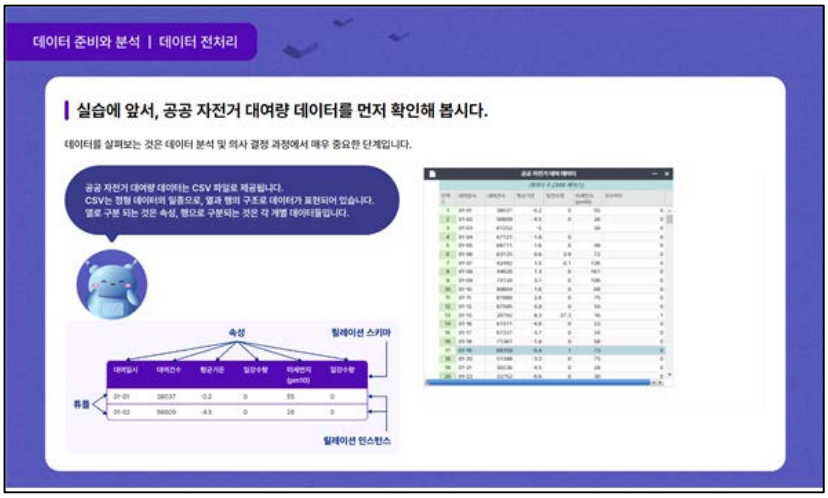
강의실

- ❖ **주제명:** 결측치, 이상치, 정규화
- ❖ **세부 주제**
  1) 결측치 처리방법
  2) 이상치 탐색방법
  3) 정규화 방법
- ❖ **콘텐츠 개발 목적:** 데이터 전처리의 중요성을 이해하고, 데이터 분석의 전 단계임을 인지한다.
- ❖ **선생님을 위한 팁!**
  데이터 전처리의 기본 개념에 대해 이해하지 못하는 학생이 있다면 보충 설명을 먼저 진행해야 한다.

실험실

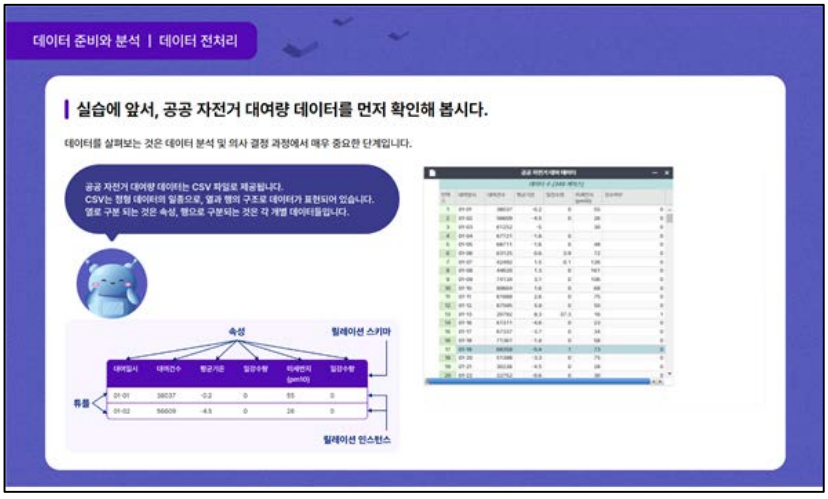
- ❖ **주제명:** 데이터 전처리
- ❖ **세부 주제**
  1) 데이터의 결측치를 확인하고, 제거
  2) 데이터 결측치와 이상치를 박스 플롯으로 확인
- ❖ **콘텐츠 개발 목적:** 데이터의 결측치, 이상치를 제거하거나 시각화하는 과정으로 데이터 전처리의 중요성 인식

❖ 실습 콘텐츠 안내:



Step 1

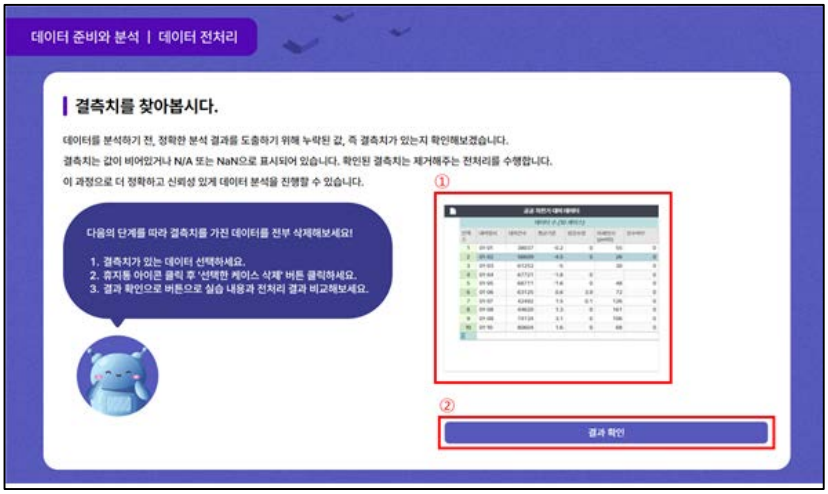
특정 데이터를 살펴보고, 구조를 이해



Step 2

Step 1 (3 페이지)의 데이터를 요약한 정보를 통해 전반적인 데이터 분포를 확인

❖ 실습 콘텐츠 안내:



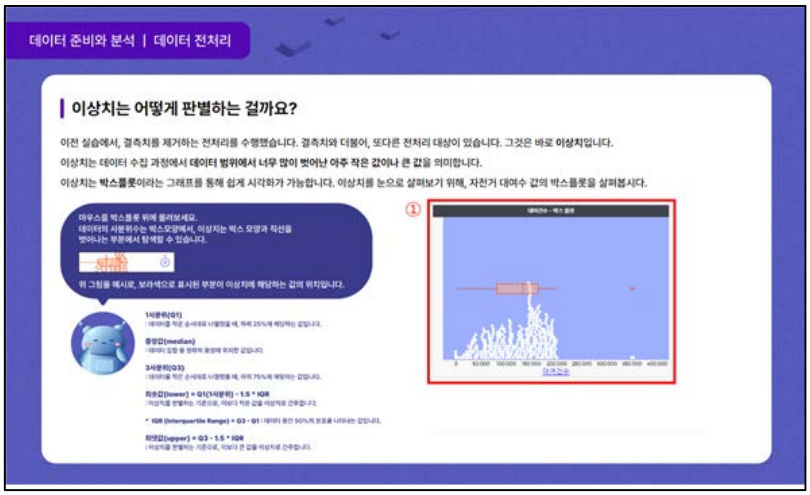
Step 3

결측치(비어있는 칸)가 있는 데이터 행을 삭제하여 전처리 과정을 실습해 보고, '결과 확인' 버튼을 눌러 제대로 수행했는지 확인

기능 사용법

① 데이터 조작 영역: 데이터를 조작할 수 있는 영역입니다. 말풍선을 따라 해당 영역을 조작하여 실습을 수행해 보세요.

② 결과 확인 버튼: 말풍선의 지시문에 따라 실습을 수행한 후, 해당 버튼을 클릭해 정답을 확인하고 작성한 답안과 일치하는지 비교해 보세요.



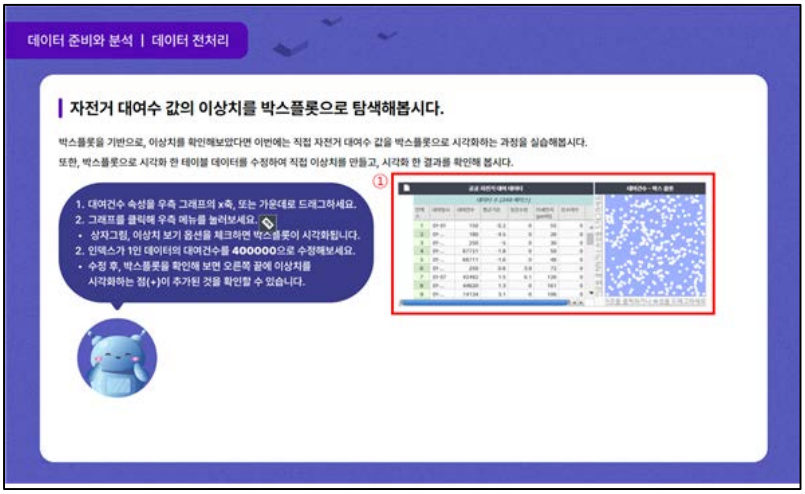
Step 4

설명을 참고하여 박스플롯 그래프에 표시된 지표 확인하면서, 이상치 판별 기준을 확인

기능 사용법

① 그래프 조작 영역: 말풍선과 말풍선 밑 설명을 토대로 그래프에서 지표를 탐색하는 영역입니다.

❖ 실습 콘텐츠 안내:



Step 5

데이터를 시각화하고, 특정 행의 대여건수 값을 매우 큰 값으로 수정하고 박스플롯의 변화를 관찰

기능 사용법

① 데이터 조작 영역: 데이터를 조작할 수 있는 영역입니다. 말풍선을 따라 해당 영역을 조작하여 실습을 수행해 보세요.

주의 실습 공간이 한정적이므로, 테이블의 크기, 위치 등을 임의로 조정할 경우 실습 진행에 방해가 될 수 있습니다. 가급적으로 주어진 실습 공간의 배치를 움직이지 말아주세요!

❖ **선생님을 위한 팁!**

결측치, 이상치를 삭제하는 방법 외에 다른 전처리 방법 (결측, 이상치 전치, 데이터 정규화, 구간화 등)을 추가 설명하면 더욱 좋습니다.

3. 세부 콘텐츠

◆ 개요

06차시. 데이터 시각화

성취기준, 학습목표, 학습내용, 개발유형

성취기준	[12데과02-02] 이상치와 결측치 탐색 및 정규화를 통해 전처리하여 오류 가능성을 최소화하고, 데이터 분석을 위해 시각화한다.
학습 목표	• 데이터 시각화의 개념과 필요성을 이해한다. • 시각화의 종류와 특징을 알고, 적절한 시각화를 선택할 수 있다.
대상 학년	
연계교육과정	정보교과
세부콘텐츠	1. (강의실) 동영상 콘텐츠 2. (실험실) 실습형 콘텐츠

◆ 구성

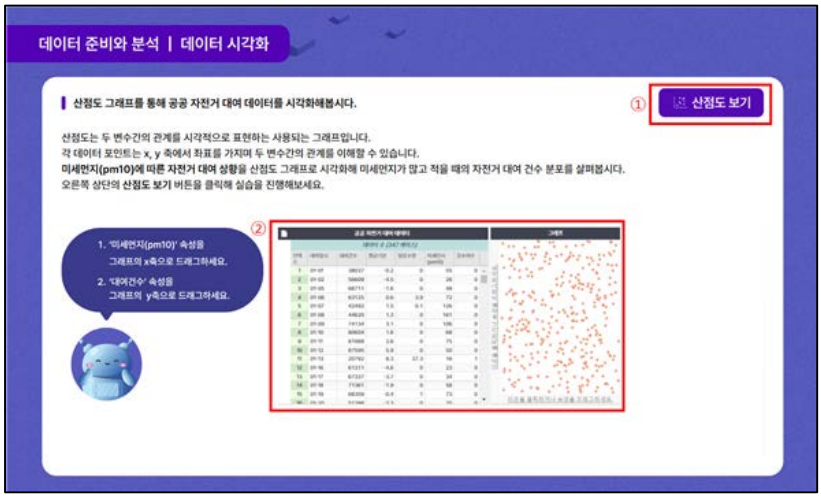
강의실

- ❖ **주제명:** 데이터 시각화 특징과 그래프 종류
- ❖ **세부 주제**
  1) 데이터 시각화 특징
  2) 시각화 그래프 종류와 활용 예시
- ❖ **콘텐츠 개발 목적:** 그래프 예시를 통해 데이터 시각화를 알기 쉽게 설명한다.
- ❖ **선생님을 위한 팁!**
  필요시, 서울시 공공 자전거 데이터를 활용할 수 있습니다.

실험실

- ❖ **주제명:** 데이터 시각화
- ❖ **세부 주제**
  1) 데이터 시각화의 개념과 필요성 인식
  2) 데이터 시각화 실습
- ❖ **콘텐츠 개발 목적:** 데이터 분석에서 중요한 단계인 시각화의 개념과 필요성을 인식하고, 배운 이론을 활용해 직접 실습 진행

❖ 실습 콘텐츠 안내:



Step 1

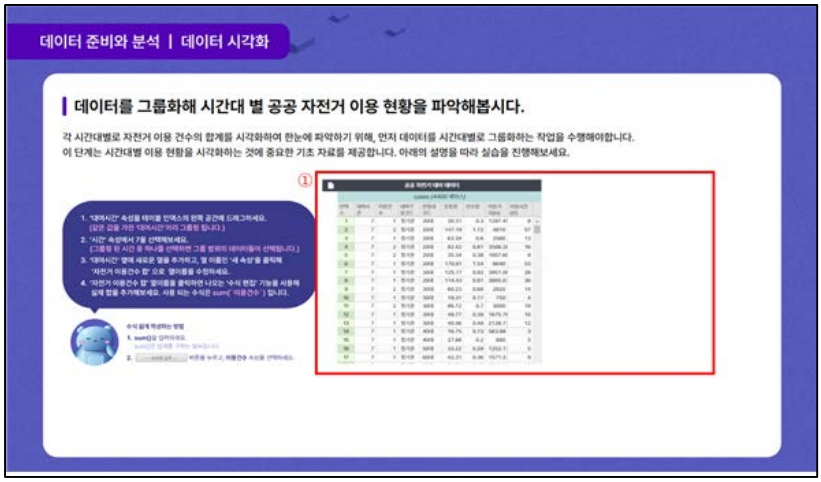
미세먼지, 대여 건수 속성들로 구성된 산점도를 시각화

기능 사용법

① 산점도 보기 버튼: 상단의 설명을 토대로 진행할 실습 공간을 여는 버튼입니다.

② 데이터, 그래프 조작 영역: 데이터, 그래프 조작이 가능한 영역입니다. 말풍선을 따라 해당 영역을 조작하여 실습을 수행해 보세요.

❖ 실습 콘텐츠 안내:

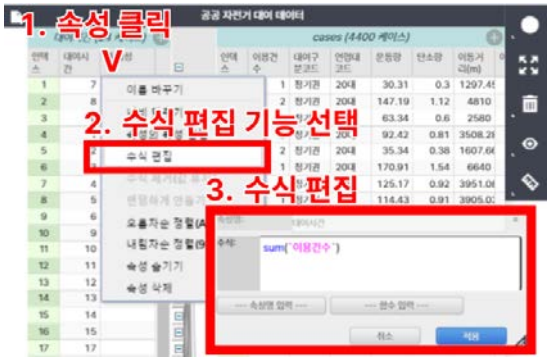


Step 2

시각화를 위한 데이터 그룹화 작업을 실습

기능 사용법

① 데이터 조작 영역: 데이터 조작이 가능한 영역입니다. 말풍선을 따라 해당 영역을 조작하여 실습을 수행해 보세요.



❖ 실습 콘텐츠 안내:

Step 3

그룹화 한 데이터를 선 그래프 버튼을 눌러 시각화하기

기능 사용법

① 선 그래프 버튼: 버튼을 누르면 그룹화한 데이터를 선그래프로 시각화한 결과와, 시각화에 대한 설명을 확인할 수 있습니다.

주의 실습 공간이 한정적이므로, 테이블의 크기, 위치 등을 임의로 조정할 경우 실습 진행에 방해가 될 수 있습니다. 가급적으로 주어진 실습 공간의 배치를 움직이지 말아주세요!

❖ **선생님을 위한 팁!**

말풍선은 한정적인 가이드라인을 표시하지만, 다른 조건으로 시각화 하는 것은 자유입니다. 실습 시 콘텐츠에서 제공한 조건 외, 여러 조건을 제시해서 시각화하는 수업을 다양하게 구성하시면 더욱 좋습니다.

3. 세부 콘텐츠

◆ 개요

07차시. 데이터 상관관계

성취기준, 학습목표, 학습내용, 개발유형

성취기준	[12데과02-03] 데이터를 분석하기 위해 데이터 속성 간의 관계를 파악하고 통합한다.
학습 목표	• 상관관계의 개념을 설명할 수 있다. • 상관관계를 파악하기 위한 데이터의 시각화와 통계적 분석 방법을 이해할 수 있다.
대상 학년	
연계교육과정	정보교과
세부콘텐츠	1. (강의실) 동영상 콘텐츠 2. (실험실) 실습형 콘텐츠

◆ 구성

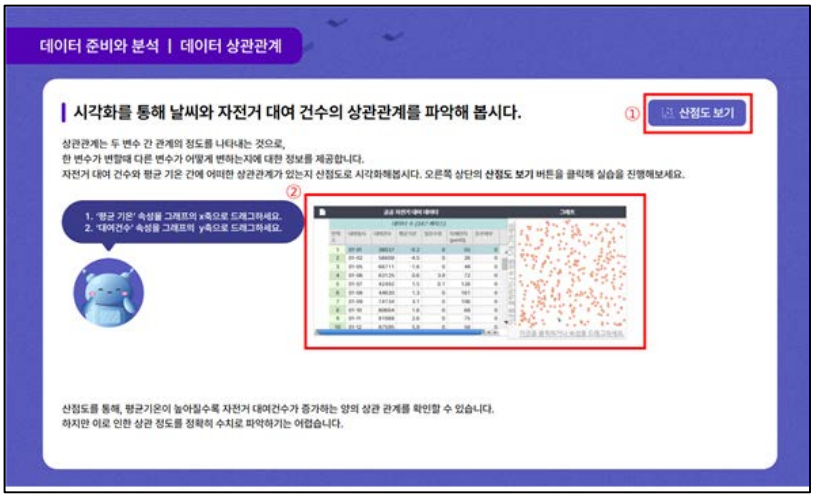
강의실

- ❖ **주제명:** 데이터 시각화를 통한 상관관계 이해하기
- ❖ **세부 주제**
  1) 데이터 시각화
  2) 통계적 분석
- ❖ **콘텐츠 개발 목적:** 상관관계의 개념을 이해하고 해석할 수 있다.
- ❖ **선생님을 위한 팁!**
  학습 전, 학생들이 데이터 시각화 차시 내용을 복습할 필요가 있습니다.

실험실

- ❖ **주제명:** 데이터 상관관계
- ❖ **세부 주제**
  1) 상관관계 이해
  2) 데이터 속성 간의 상관관계 시각화
- ❖ **콘텐츠 개발 목적:** 데이터 상관관계를 시각화하는 실습으로 분석에서 상관관계의 역할을 확인

❖ 실습 콘텐츠 안내:



Step 1

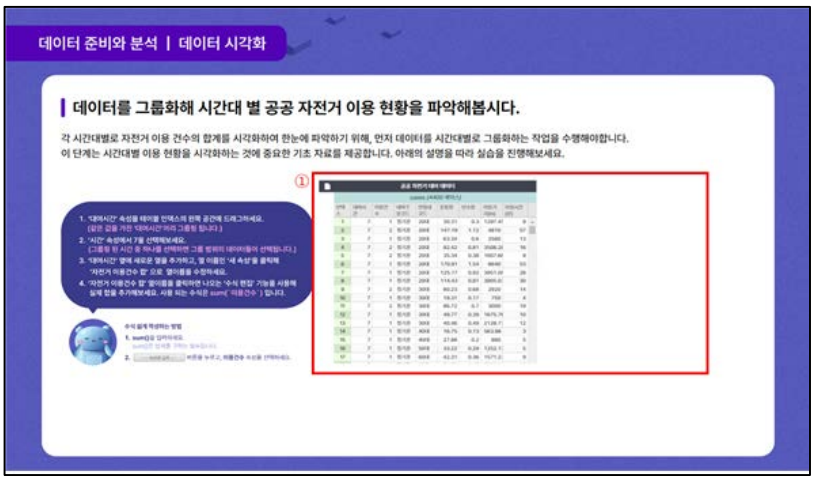
날씨와 자전거 대여건수의 상관관계를 그래프로 시각화해 확인

기능 사용법

① 산점도 보기 버튼: 상단의 설명을 토대로 진행할 실습 공간을 여는 버튼입니다.

② 데이터, 그래프 조작 영역: 데이터, 그래프 조작이 가능한 영역입니다. 말풍선을 따라 해당 영역을 조작하여 실습을 수행해 보세요.

❖ 실습 콘텐츠 안내:

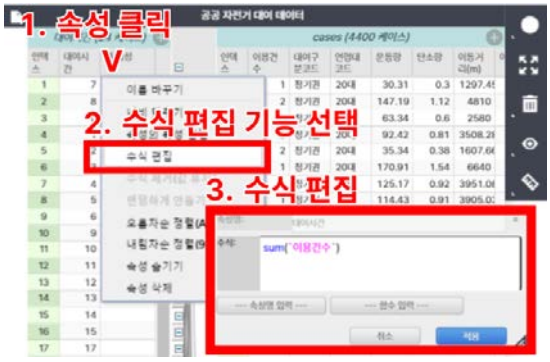


Step 2

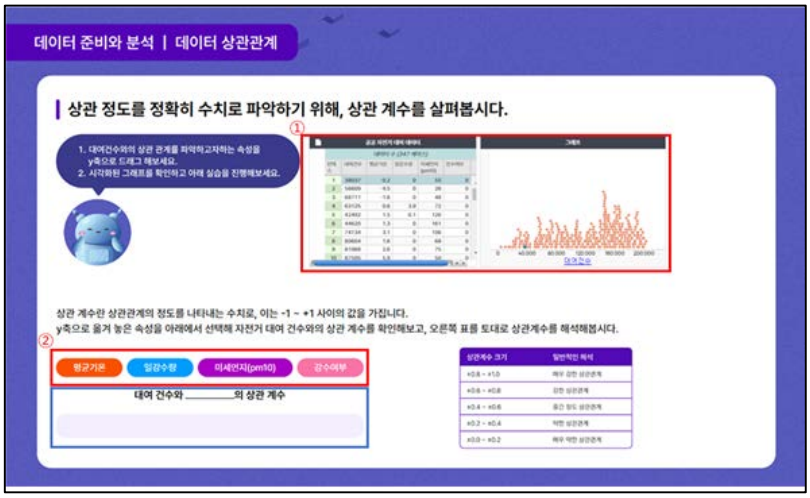
산점도로 속성 간 상관관계를 시각화해 보고, 상관관계에 대한 지표인 상관계수를 확인한 뒤 지표를 토대로 상관관계 해석

기능 사용법

① 데이터 조작 영역: 데이터 조작이 가능한 영역입니다. 말풍선을 따라 해당 영역을 조작하여 실습을 수행해보세요.



❖ 실습 콘텐츠 안내:



Step 3

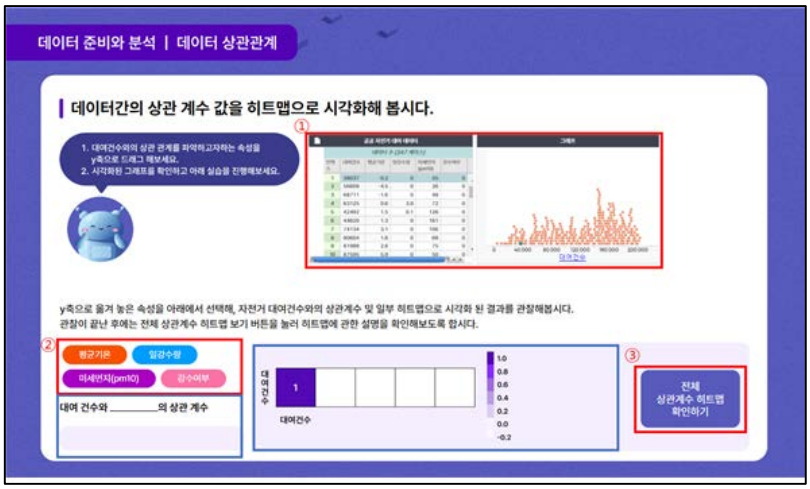
그룹화한 데이터를 선 그래프 버튼을 눌러 시각화하기

기능 사용법

① 데이터, 그래프 조작 영역: 데이터 및 그래프 조작이 가능한 영역입니다. 말풍선을 따라 해당 영역을 조작하여 실습을 수행해 보세요.

② 속성 선택 버튼: 각 속성들을 선택하면 아래 영역에 상관계수를 표시하는 버튼입니다.

❖ 실습 콘텐츠 안내:



Step 4

상관관계를 산점도, 히트맵을 시각화하여 상관관계 시각화의 다각적 측면 이해하기

기능 사용법

① 데이터, 그래프 조작 영역: 데이터 및 그래프 조작이 가능한 영역입니다. 말풍선을 따라 해당 영역을 조작하여 실습을 수행해 보세요.

② 속성 선택 버튼: 각 속성들을 선택하면 아래에 상관계수를 표시하고, 오른쪽 영역에 상관계수를 기반으로 시각화한 히트맵의 일부를 표시하는 버튼입니다.

③ 전체 상관계수 히트맵 확인하기 버튼: ②번 버튼은 히트맵으로 시각화된 일부만 확인할 수 있던 반면, 해당 버튼은 데이터 속성들에 대한 전체적인 히트맵을 확인하고 그에 대한 설명을 확인할 수 있습니다.

주의 실습 공간이 한정적이므로, 테이블의 크기, 위치 등을 임의로 조정할 경우 실습 진행에 방해가 될 수 있습니다. 가급적으로 주어진 실습 공간의 배치를 움직이지 말아주세요!

❖ **선생님을 위한 팁!**

상관관계를 이용해 분석할 만한 데이터, 속성이 무엇이 있을지 탐구해 보는 시간을 가지면 상관관계 원리 이해에 더욱 도움이 됩니다.

3. 세부 콘텐츠

◆ 개요

08차시. 데이터 분석방법

성취기준, 학습목표, 학습내용, 개발유형

성취기준	[12데과02-04] 동일한 데이터를 서로 다른 분석방법을 적용하여 분석 결과를 비교한다.
학습 목표	• 탐색적 데이터 분석과 확증적 데이터 분석을 이해한다. • 동일한 데이터로 서로 다른 분석방법을 적용해 분석결과를 비교한다.
대상 학년	
연계교육과정	정보교과
세부콘텐츠	1. (강의실) 동영상 콘텐츠 2. (실험실) 실습형 콘텐츠

◆ 구성

강의실

- ❖ **주제명:** 데이터 분석방법 알아보기
- ❖ **세부 주제**
  1) 탐색적 데이터 분석
  2) 확증적 데이터 분석
- ❖ **콘텐츠 개발 목적:** 분석방법에 따른 분석결과의 차이를 이해할 수 있다.
- ❖ **선생님을 위한 팁!**
  다양한 응용활동이 가능합니다.

실험실

- ❖ **주제명:** 데이터 분석(1) - 건강데이터 살펴보기
- ❖ **세부 주제**
  1) 학생 건강 검진 데이터를 이해하고 해석
  2) 특정 데이터로 데이터 분석을 하는 과정 실습
- ❖ **콘텐츠 개발 목적:** 건강 검진 데이터를 사례로, 데이터를 이해하고, 시각화하면서 해석하는 과정을 통해 데이터 분석의 기초를 실습할 수 있도록 구성

❖ 실습 콘텐츠 안내:

Step 1

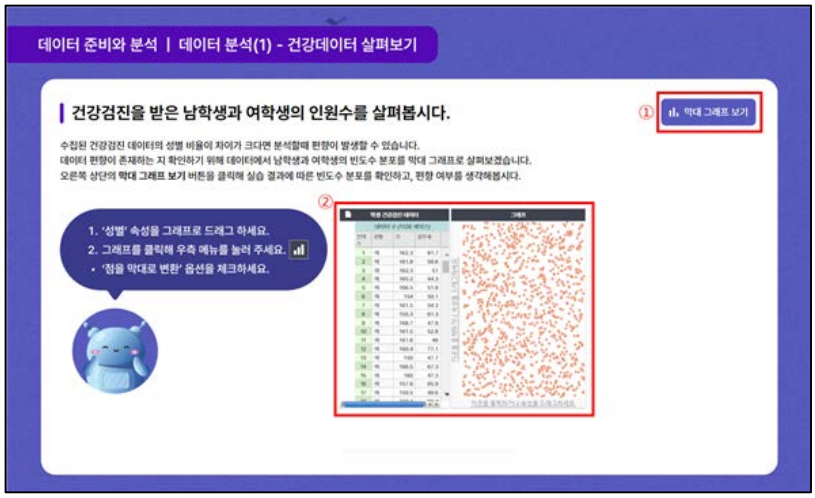
학생 건강 검진 데이터를 확인 후 해당 데이터의 특성 파악

기능 사용법

① 데이터 입력란: 우측 데이터를 확인한 뒤 숫자 입력 및 체크로 답안을 작성해 주세요.

② 제출 버튼: 입력란에 기입한 답안을 토대로 정/오답 여부를 판별하는 버튼입니다.

❖ 실습 콘텐츠 안내:



Step 2

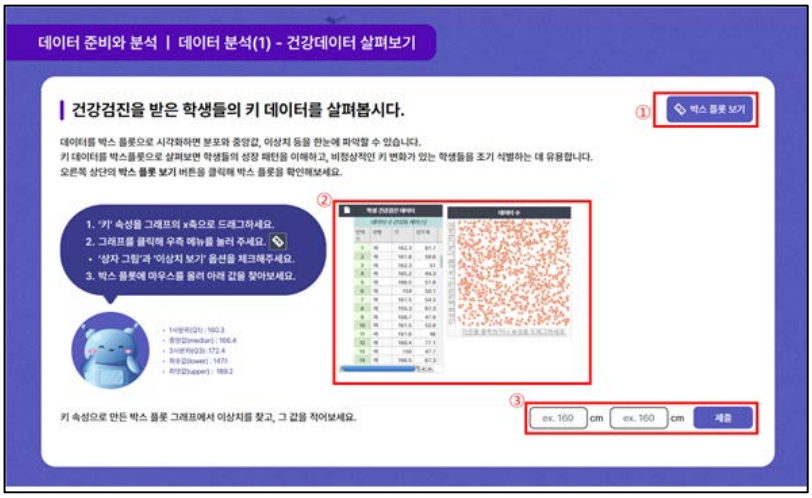
'성별' 속성을 막대그래프로 시각화하여 분포를 살펴보고, 데이터 편향 여부 검토하기

기능 사용법

① 막대그래프 보기 버튼: 상단의 설명을 토대로 진행할 실습 공간을 여는 버튼입니다.

② 데이터, 그래프 조작 영역: 데이터, 그래프 조작이 가능한 영역입니다. 말풍선을 따라 해당 영역을 조작하여 실습을 수행해 보세요.

❖ 실습 콘텐츠 안내:



Step 3

데이터의 '키' 속성을 박스플롯으로 시각화한 뒤 분포, 중앙값, 이상치 확인하기

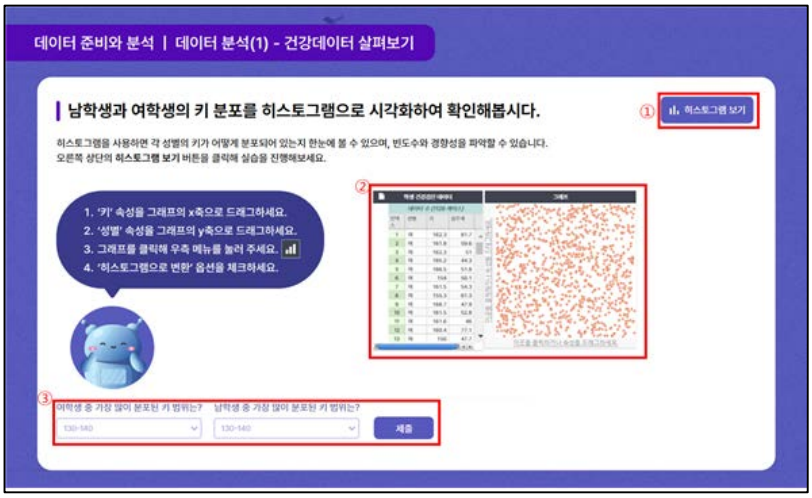
기능 사용법

① 박스플롯 보기 버튼: 상단의 설명을 토대로 진행할 실습 공간을 여는 버튼입니다.

② 데이터, 그래프 조작 영역: 데이터, 그래프 조작이 가능한 영역입니다. 말풍선을 따라 해당 영역을 조작하여 실습을 수행해 보세요.

③ 데이터 입력란 및 제출 버튼: ②번 기능에서 수행한 내용을 토대로 답안을 숫자로 기입한 뒤, 해당 버튼을 누르면 정/오답 여부를 확인할 수 있습니다.

❖ 실습 콘텐츠 안내:



Step 4

'키' 분포를 '성별에 따라 히스토그램으로 시각화하여 빈도수와 경향성 파악하기

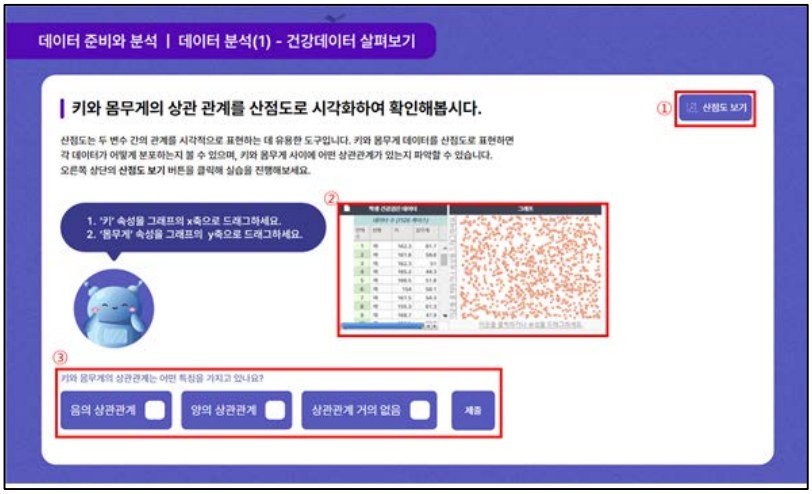
기능 사용법

① 히스토그램 보기 버튼: 상단의 설명을 토대로 진행할 실습 공간을 여는 버튼입니다.

② 데이터, 그래프 조작 영역: 데이터, 그래프 조작이 가능한 영역입니다. 말풍선을 따라 해당 영역을 조작하여 실습을 수행해 보세요.

③ 데이터 입력란 및 제출 버튼: ②번 기능에서 수행한 내용을 토대로 답안을 선택한 뒤, 해당 버튼을 누르면 정/오답 여부를 확인할 수 있습니다.

❖ 실습 콘텐츠 안내:



Step 5

'키'와 '몸무게' 속성을 산점도로 시각화하여 상관관계 파악하기

기능 사용법

① 산점도 보기 버튼: 상단의 설명을 토대로 진행할 실습 공간을 여는 버튼입니다.

② 데이터, 그래프 조작 영역: 데이터, 그래프 조작이 가능한 영역입니다. 말풍선을 따라 해당 영역을 조작하여 실습을 수행해보세요.

③ 데이터 입력란 및 제출 버튼: ②번 기능에서 수행한 내용을 토대로 답안을 체크한 뒤, 해당 버튼을 누르면 정/오답 여부를 확인할 수 있습니다.

주의 답안은 ex로 주어진 형식으로 작성해 주세요. 만약 'ex. 160'으로 표시된 입력란에 '160cm'라고 적을 경우 오답 처리될 수 있습니다.

❖ **선생님을 위한 팁!**

실습 콘텐츠에서 언급하지 않은 다른 시각화 그래프 혹은 이전 차시에 배웠던 시각화의 중요성 등에 대해 추가 부연 설명을 하면서 실습을 진행하면 더욱 좋습니다.

3. 세부 콘텐츠

◆ 개요

08차시. 데이터 분석방법

성취기준, 학습목표, 학습내용, 개발유형

성취기준	[12데과02-04] 동일한 데이터를 서로 다른 분석방법을 적용하여 분석 결과를 비교한다.
학습 목표	• 탐색적 데이터 분석과 확증적 데이터 분석을 이해한다. • 동일한 데이터로 서로 다른 분석방법을 적용해 분석결과를 비교한다.
대상 학년	
연계교육과정	정보교과
세부콘텐츠	1. (강의실) 동영상 콘텐츠 2. (실험실) 실습형 콘텐츠

◆ 구성

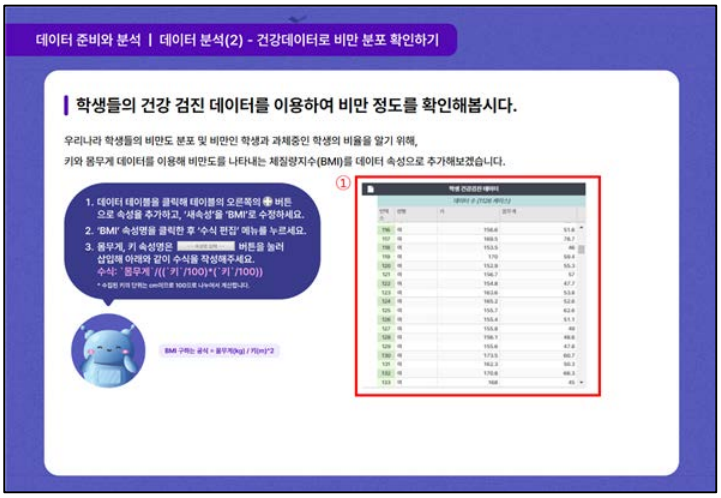
강의실

- ❖ **주제명:** 데이터 분석방법 알아보기
- ❖ **세부 주제**
  1) 탐색적 데이터 분석
  2) 확증적 데이터 분석
- ❖ **콘텐츠 개발 목적:** 분석방법에 따른 분석결과의 차이를 이해할 수 있다.
- ❖ **선생님을 위한 팁!**
  다양한 응용활동이 가능합니다.

실험실

- ❖ **주제명:** 데이터 분석(2) - 건강데이터로 비만 분포 확인하기
- ❖ **세부 주제**
  1) 데이터를 조합해 분석에 필요한 새로운 지표를 생성하는 방법 습득
  2) 분석에 필요한 데이터를 조합하여 시각화
- ❖ **콘텐츠 개발 목적:** 분석에 필요한 데이터를 조합하여 새로운 지표를 생성해 보고, 시각화하는 역량을 키우기 위한 실습

❖ 실습 콘텐츠 안내:

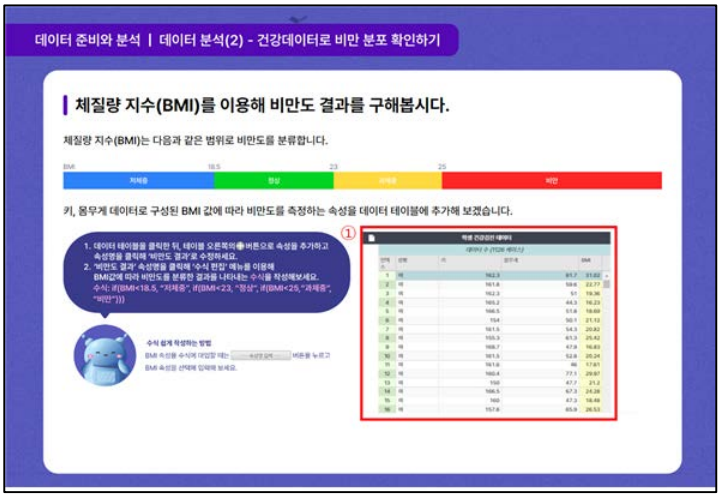


Step 1

키와 몸무게 속성값으로 체질량 지수 속성 추가하기

기능 사용법

① 데이터, 조작 영역: 데이터 조작이 가능한 영역입니다. 말풍선을 따라 해당 영역을 조작하여 실습을 수행해 보세요.



Step 2

산점도로 속성 간 상관관계를 시각화해 보고, 상관관계에 대한 지표인 상관계수를 확인한 뒤 지표를 토대로 상관관계 해석

기능 사용법

① 데이터 조작 영역: 데이터 조작이 가능한 영역입니다. 말풍선을 따라 해당 영역을 조작하여 실습을 수행해 보세요.

❖ 실습 콘텐츠 안내:

Step 3

그룹화한 데이터를 선 그래프 버튼을 눌러 시각화하기

기능 사용법

① 막대 그래프 보기 버튼: 상단의 설명을 토대로 진행할 실습 공간을 여는 버튼입니다.

② 데이터, 그래프 조작 영역: 데이터, 그래프 조작이 가능한 영역입니다. 말풍선을 따라 해당 영역을 조작하여 실습을 수행해 보세요.

③ 데이터 입력란 및 제출 버튼: ②번 기능에서 수행한 내용을 토대로 답안을 체크한 뒤, 해당 버튼을 누르면 정/오답 여부를 확인할 수 있습니다.

주의 답안은 ex로 주어진 형식으로 작성해 주세요. 만약 'ex. 10'으로 표시된 입력란에 '10%'라고 적을 경우 오답 처리될 수 있습니다.

❖ **선생님을 위한 팁!**

새롭게 생성한 지표를 가지고 다른 분석 주제를 선정할 수 있을지, 가능하다면 어떤 주제를 선정하여 어떤 분석 목표를 이룰 수 있을 지에 대한 토론을 실시하면 더욱 좋습니다.

3. 세부 콘텐츠

◆ 개요

09차시. 데이터 모델링 및 데이터 분석 도구 탐색하기

성취기준, 학습목표, 학습내용, 개발유형

성취기준	[12데과03-01] 데이터 모델 개념을 이해하고 데이터 분석에 활용할 수 있는 도구를 탐색한다.
학습 목표	• 데이터 모델 개념 이해 및 데이터간 관계 분석 및 상호 연관성을 파악한다. • 데이터 속성에 대한 유사성 측정, 분석의 원리 이해할 수 있다.
대상 학년	
연계교육과정	정보교과
세부콘텐츠	1. (강의실) 동영상 콘텐츠

◆ 구성

강의실

❖ **주제명:** 다양한 분석 도구 살펴보기

❖ **세부 주제**

1) 데이터 모델링의 특징

2) 데이터 분석 도구

❖ **콘텐츠 개발 목적:** 다양한 데이터 분석 도구를 소개한다.

❖ **선생님을 위한 팁!**

데이터 분석 도구 예시로 설명된 사이트들을 추가적으로 실습해 볼 수 있다.

3. 세부 콘텐츠

◆ 개요

10차시. 단순 선형 회귀, 다중 선형 회귀

성취기준, 학습목표, 학습내용, 개발유형

성취기준	[12데과03-02] 동일한 데이터를 통계적 회귀모델과 기계학습을 통한 회귀모델로 분석하여 결과 해석 내용을 비교한다.
학습 목표	• 회귀 분석의 개념에 대해 설명할 수 있다. • 단순 성형 회귀, 다중 선형 회귀 분석의 개념을 설명할 수 있다. • 예에 따른 회귀 분석 방법을 선택할 수 있다.
대상 학년	
연계교육과정	정보교과
세부콘텐츠	1. (강의실) 동영상 콘텐츠 2. (실험실) 실습형 콘텐츠

◆ 구성

강의실

❖ **주제명:** 선형 회귀의 기본 이해

❖ **세부 주제**

1) 단순 선형 회귀

2) 다중 선형 회귀

❖ **콘텐츠 개발 목적:** 회귀 분석의 개념을 알고 실행할 수 있다.

❖ **선생님을 위한 팁!**

상황에 따라 단순 선형 회귀와 다중 선형 회귀 중 어떤 것을 적용해야 하는지 설명할 수 있습니다.

실험실

❖ **주제명:** 회귀분석으로 예측하기

❖ **세부 주제**

1) 회귀분석에 사용되는 독립변수, 종속변수 이해

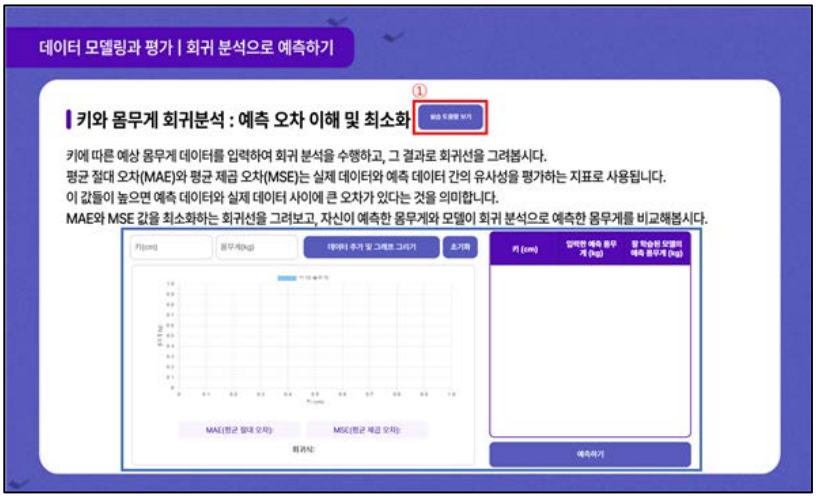
2) 회귀 모델 생성 과정 이해 및 예측 오차 최소화 해보기

❖ **콘텐츠 개발 목적:** 단순 선형 회귀 방식으로 키에 기반한 몸무게를 예측하며 회귀 분석의 원리 이해

❖ 실습 콘텐츠 안내:

Step 1

드래그 앤 드롭으로 회귀 분석을 위한 종속 변수/독립변수 분류



Step 2

회귀 분석으로 키에 따른 몸무게 예측해보기

기능 사용법

① 실습 도움말 보기 버튼: 실습을 진행하기 위한 가이드를 차례대로 제공하는 버튼입니다. 실습을 진행하기 전 해당 버튼을 눌러 실습 과정을 이해해보세요.

주의 4 페이지에서, 키와 예측 몸무게를 입력할 때 키는 100(cm) 이상, 몸무게는 10(kg) 이상 입력해야 합니다.

❖ **선생님을 위한 팁!**

회귀 분석에 사용되는 회귀식에 대한 추가 설명을 해주신다면 학생들의 회귀분석 이해에 더욱 도움이 됩니다.

3. 세부 콘텐츠

◆ 개요

11차시. 회귀 분석, 결정 계수

성취기준, 학습목표, 학습내용, 개발유형

성취기준	[12데과03-02] 동일한 데이터를 통계적 회귀모델과 기계학습을 통한 회귀모델로 분석하여 결과 해석 내용을 비교한다.
학습 목표	• 성능 지표(평균제곱오차, 결정계수)의 개념을 알고 값을 구할 수 있다. • AI 모델의 적합도를 확인할 수 있다.
대상 학년	
연계교육과정	정보교과
세부콘텐츠	1. (강의실) 동영상 콘텐츠 2. (실험실) 실습형 콘텐츠

◆ 구성

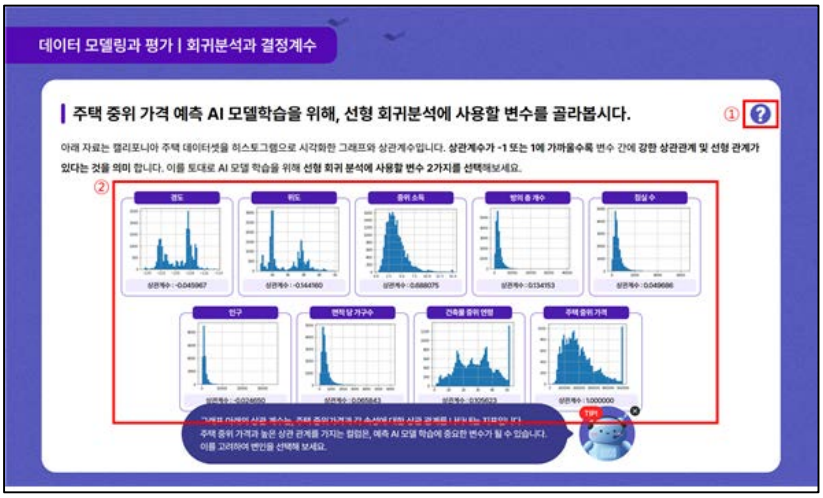
강의실

- ❖ **주제명:** R^2 값 알아보기
- ❖ **세부 주제**
  1) 회귀 분석
  2) 결정 계수
- ❖ **콘텐츠 개발 목적:** 동일한 데이터셋을 활용하더라도 데이터 전처리 과정, 데이터 선택, 분석방법에 따라 모델링 결과가 다를 수 있음을 인지한다.
- ❖ **선생님을 위한 팁!**
  R^2 값에 대한 부연설명이 필요할 수 있다.

실험실

- ❖ **주제명:** 회귀분석으로 예측하기
- ❖ **세부 주제**
  1) 회귀분석에 사용되는 변수 선정하기
  2) 선정한 변수에 따라 달라지는 결정계수 이해하기
- ❖ **콘텐츠 개발 목적:** 더 정확하고 정교한 회귀 분석 인공지능(AI) 모델을 만들기 위해 속성을 선택하는 기준들을 학습

❖ 실습 콘텐츠 안내:



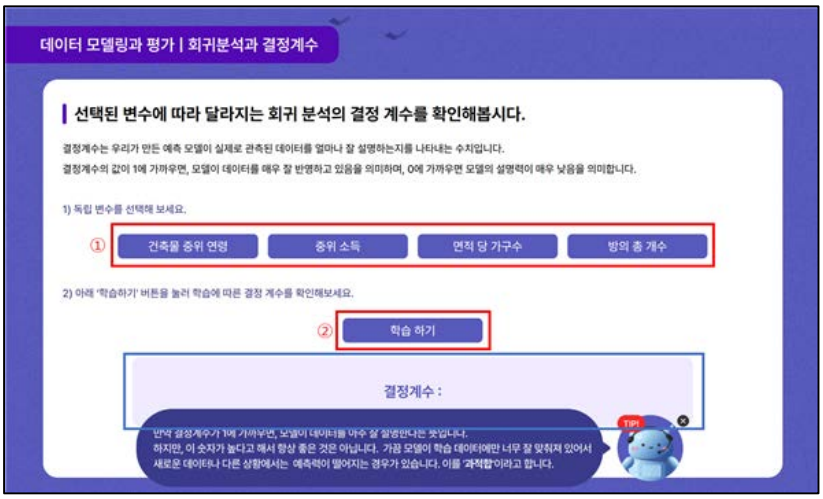
Step 1

드래그 앤 드롭으로 회귀 분석을 위한 종속 변수/독립변수 분류 기능 사용법

① 도움말 보기 버튼: 변수들에 대한 설명을 확인할 수 있는 버튼입니다.

② 아이콘 선택 기능: 설명을 따라 올바른 변수들을 두 가지 선택합니다. 선택한 변수 아이콘은 밝은색으로 표시되며, 두 가지를 모두 선택할 경우 정/오답 여부가 판별됩니다.

❖ 실습 콘텐츠 안내:



Step 2

변수를 선택하고 달라지는 결정계수 확인하기

기능 사용법

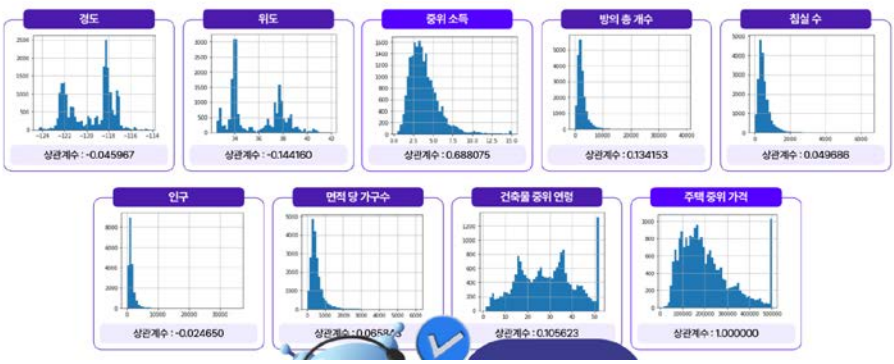
① 아이콘 선택 기능: 회귀 분석에 사용할 독립 변수를 선택합니다.최소 한 가지 ~ 최대 네 가지의 변수를 선택할 수 있습니다. (다중 선택이 가능)

② 학습하기 버튼: 해당 버튼을 누르면 ①번 기능으로 선택한 변수들에 대한 결정계수를 확인할 수 있습니다.

주의 3페이지에서, 요소를 선택할 때 TIP이 꺼지지 않은 상태일 경우 선택이 잘 되지 않을 수 있습니다. 선택이 잘 되지 않는다면 TIP을 닫아주세요.

❖ **선생님을 위한 팁!**

이전 실습에서 학습한 MAE, MSE 지표와 결정계수에 관계를 추가 설명해주면 회귀분석 이해에 더욱 도움을 줄 수 있습니다.



3. 세부 콘텐츠

◆ 개요

12차시. 군집화 개념 이해하기

성취기준, 학습목표, 학습내용, 개발유형

성취기준	[12데과03-03] 데이터의 속성에 대한 유사성을 측정하고 분석하여 군집을 형성하고, 군집 분석 결과의 의미를 해석한다. [12데과03-04] 데이터 간의 관계를 분석하고 상호 연관성을 파악하여 결과의 의미를 해석한다.
학습 목표	• 데이터 속성에 따라 유사도가 높은 데이터를 묶을 수 있다. • 데이터의 속성에 따라 다수의 군집으로 나누고 군집 내 유사성과 군집 간 상이성을 이해한다.
대상 학년	
연계교육과정	정보교과
세부콘텐츠	1. (강의실) 동영상 콘텐츠 2. (실험실) 실습형 콘텐츠

◆ 구성

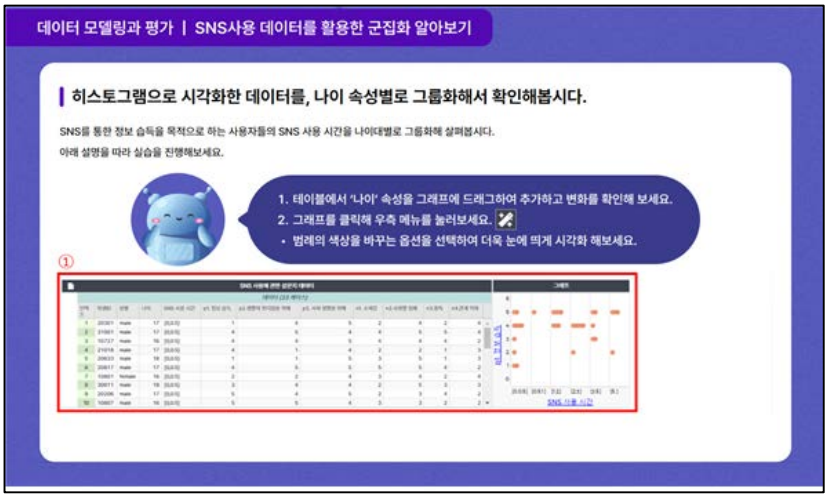
강의실

- **주제명:** 보로노이 다이어그램과 센트로이드
- **세부 주제**
  1) 보로노이 다이어그램 활용 예시
  2) 센트로이드 설정 과정
- **콘텐츠 개발 목적:** 센트로이드 개념을 설명하기 위해 보로노이 다이어그램으로 흥미를 유발한다.
- **선생님을 위한 팁!**
  보로노이 다이어그램 실습을 위해서는 하얀 그릇, 색상별 초콜릿 물을 준비하셔야 합니다.

실험실

- **주제명:** SNS 사용 데이터를 활용한 군집화 알아보기
- **세부 주제**
  1) 데이터 그룹화 방법 이해
  2) 데이터 그룹별 시각화 방법 이해
- **콘텐츠 개발 목적:** 데이터를 원하는 목적에 따라 그룹으로 분류하고, 시각화하여 군집 유형별 의미 유추해 보기

❖ 실습 콘텐츠 안내:

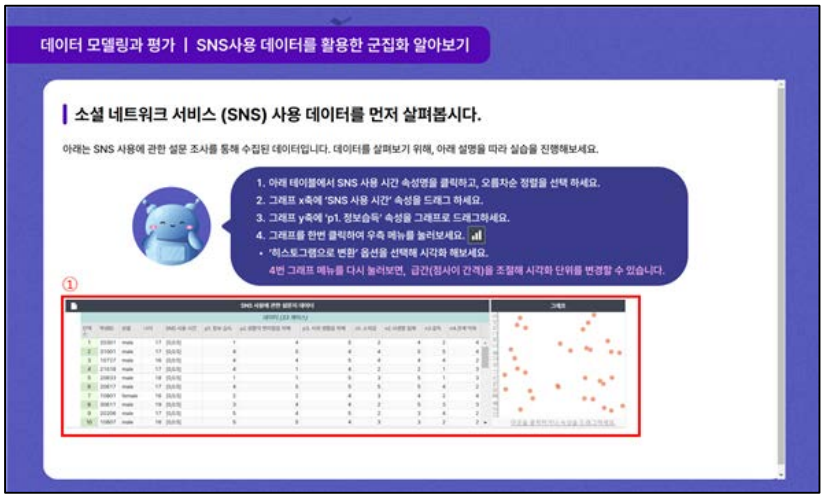


Step 1

히스토그램으로 시각화한 데이터를 나이 속성별로 그룹화
기능 사용법

① 데이터, 그래프 조작 영역: 데이터, 그래프 조작이 가능한 영역입니다.

말풍선을 따라 해당 영역을 조작하여 실습을 수행해 보세요.



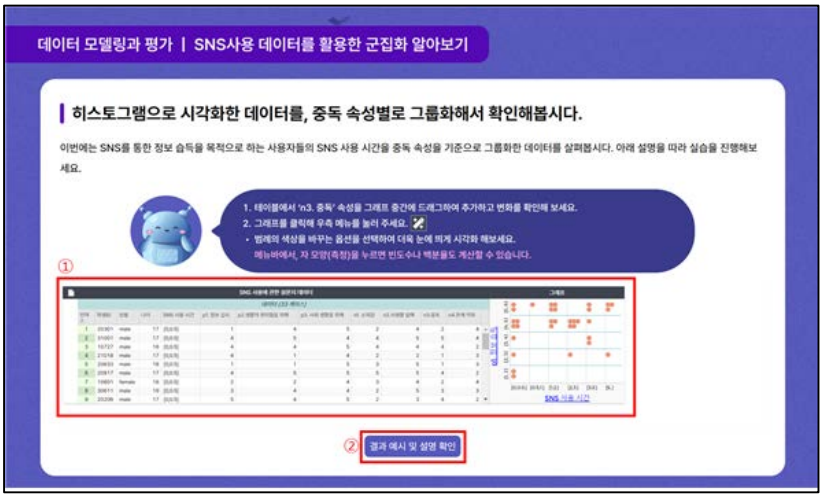
Step 2

정보 습득을 위해 sns를 사용하는 사용자들의 사용 시간을 그룹화
기능 사용법

① 데이터, 그래프 조작 영역: 데이터, 그래프 조작이 가능한 영역입니다.

말풍선을 따라 해당 영역을 조작하여 실습을 수행해 보세요.

❖ 실습 콘텐츠 안내:



Step 3

Step 2에서 그룹화 했던 데이터를 중독 속성별로 시각화

기능 사용법

① 데이터, 그래프 조작 영역: 데이터, 그래프 조작이 가능한 영역입니다. 말풍선을 따라 해당 영역을 조작하여 실습을 수행해보세요.

② 결과 예시 및 설명 확인 버튼: 지금까지 단계별로 시각화 해왔던 실습에 대한 결과 예시 및 그에 대한 설명을 확인할 수 있는 버튼입니다.

주의 실습 공간이 한정적이므로, 테이블의 크기, 위치 등을 임의로 조정할 경우 실습 진행에 방해가 될 수 있습니다. 가급적으로 주어진 실습 공간의 배치를 움직이지 말아주세요!

❖ **선생님을 위한 팁!**

해당 데이터를 그룹화 하는 이유와, 그룹화 해서 확인하면 알 수 있는 점을 군집 분석과 연계하여 설명하면 더욱 좋습니다.

3. 세부 콘텐츠

◆ 개요

13차시. 군집분석

성취기준, 학습목표, 학습내용, 개발유형

성취기준	[12데과03-03] 데이터의 속성에 대한 유사성을 측정하고 분석하여 군집을 형성하고, 군집 분석 결과의 의미를 해석한다. [12데과03-04] 데이터 간의 관계를 분석하고 상호 연관성을 파악하여 결과의 의미를 해석한다.
학습 목표	• 데이터 간 관계를 분석하고 데이터 분석에 활용할 수 있는 도구를 탐색한다.
대상 학년	
연계교육과정	정보교과
세부콘텐츠	1. (강의실) 동영상 콘텐츠 2. (실험실) 실습형 콘텐츠

◆ 구성

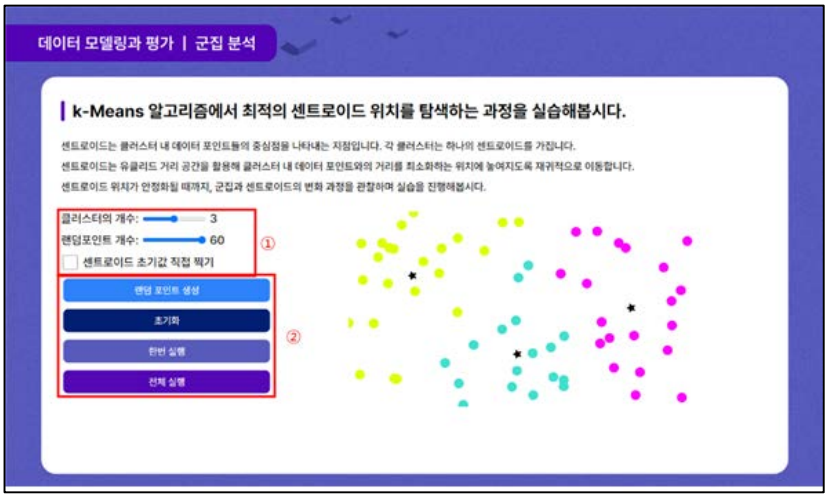
강의실

- **주제명:** 포켓몬빵 사례로 알아보는 군집분석
- **세부 주제**
  1) 군집분석, 군집간 유사성
  2) 군집간 상이성 결과 해석 및 설명
- **콘텐츠 개발 목적:** 군집분석 사례를 통해 데이터 분석에 대한 이해를 돕는다.
- **선생님을 위한 팁!**
  포켓몬빵 데이터를 활용하여 활동을 구성할 수 있습니다.

실험실

- **주제명:** 군집분석
- **세부 주제**
  1) k-Means 알고리즘으로 군집 분석 이해하기
  2) k-Means 알고리즘과 보로노이 다이어그램의 유사성 확인하기
- **콘텐츠 개발 목적:** k-Means 알고리즘을 기반으로 군집 분석 원리 및 과정 이해하기

❖ 실습 콘텐츠 안내:



Step 1

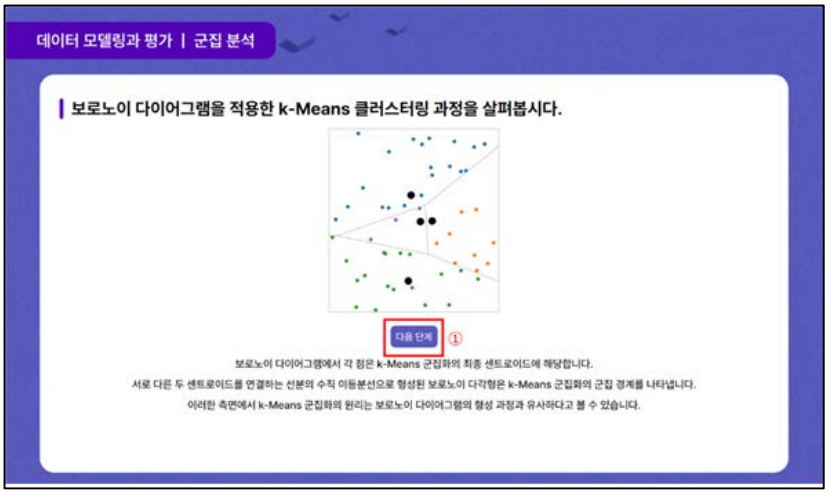
데이터 생성 및 센트로이드 탐색 관련 설정 후 알고리즘 원리 파악

기능 사용법

① 센트로이드 관련 초기 설정: 클러스터의 개수 및 랜덤포인트의 개수, 센트로이드 초깃값 설정 여부 등을 확인해주세요. 센트로이드 초기값 직접 찍기 모드를 활성화할 경우 알고리즘이 랜덤하게 센트로이드 위치를 설정하는 것이 아닌, 클릭한 위치를 기반으로 k-means 알고리즘이 실행됩니다.

② 센트로이드 탐색 관련 버튼: 랜덤 포인트 생성 버튼으로 분류할 데이터를 생성할 수 있습니다. 초기화 버튼으로 보드에 생성된 데이터를 초기화할 수 있습니다. 한 번 실행 버튼으로 k-means 알고리즘을 단계별로, 전체 실행 버튼으로 센트로이드 값이 결정될 때까지 k-means 알고리즘을 실행할 수 있습니다.

❖ 실습 콘텐츠 안내:



Step 2

보로노이 다이어그램과 k-means 클러스터링 비교

기능 사용법

① 다음 단계 버튼: 해당 버튼을 클릭하면 보로노이 다이어그램이 형성되는 과정을 단계별로 확인할 수 있습니다.

주의 센트로이드 초기값을 직접 찍을 경우, 찍을 수 있는 센트로이드의 개수는 '클러스터의 개수' 슬라이더를 통해 설정한 값과 동일합니다.

❖ **선생님을 위한 팁!**

상기 실습을 반복하며 센트로이드의 초기 위치를 다양한 지점에 설정한 뒤 k-means 알고리즘을 적용하여 적절한 초기값의 조건을 고려할 수 있도록 유도합니다.

3. 세부 콘텐츠

◆ 개요

14차시. 지지도, 신뢰도, 향상도

성취기준, 학습목표, 학습내용, 개발유형

성취기준	[12데과03-04] 데이터 간의 관계를 분석하고 상호 연관성을 파악하여 결과의 의미를 해석한다.
학습 목표	• 연관 분석의 개념을 알고 설명할 수 있다. • 지지도, 신뢰도, 향상도를 알고 설명할 수 있다.
대상 학년	
연계교육과정	정보교과
세부콘텐츠	1. (강의실) 동영상 콘텐츠 2. (실험실) 실습형 콘텐츠

◆ 구성

강의실

❖ **주제명:** 연관분석 기초

❖ **세부 주제**

1) 지지도

2) 신뢰도

3) 향상도

❖ **콘텐츠 개발 목적:** 기저귀와 맥주 판매 추이 예시를 통해 연관분석을 설명한다.

❖ **선생님을 위한 팁!**

공유데이터를 활용하여 연관분석을 실시해보는 활동을 할 수 있습니다.

실험실

❖ **주제명:** 지지도, 신뢰도, 향상도

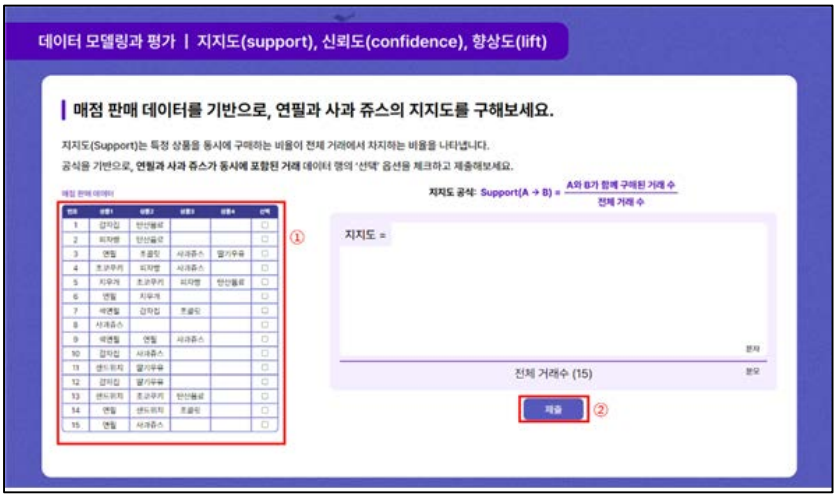
❖ **세부 주제**

1) 지지도, 신뢰도, 향상도의 의미 및 구하는 방법 습득

2) 연관 분석 지표 해석

❖ **콘텐츠 개발 목적:** 지지도, 신뢰도, 향상도를 실습을 기반으로 연관 분석 수행 방법 습득

❖ 실습 콘텐츠 안내:



Step 1

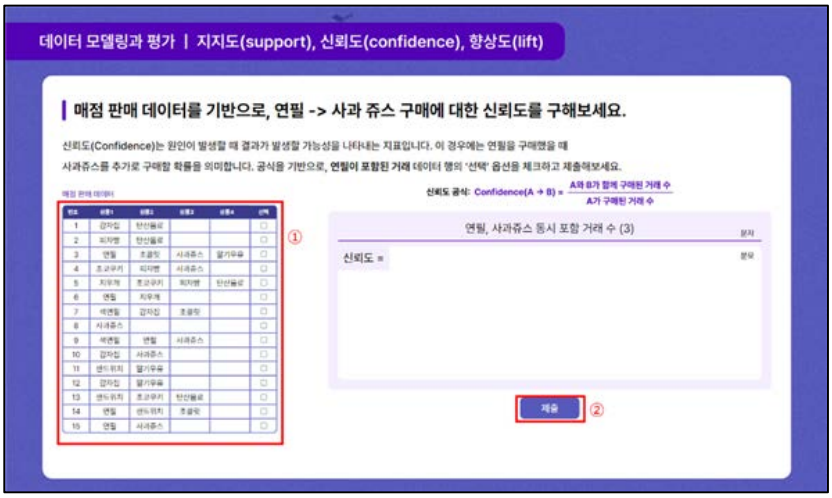
연필과 사과 주스의 지지도 계산하기

기능 사용법

① 행 선택 기능: 조건에 해당하는 행을 선택하는 기능입니다. 표의 맨 오른쪽부분의 박스를 체크하면 선택이 이루어지고, 오른쪽 분자 영역에 데이터 행이 추가 됩니다.

② 제출 버튼: 분자 영역에 추가된 데이터에 대한, 정/오답 여부를 확인하는 버튼입니다. 정답일 경우, 해당 조건에 대한 지지도를 함께 안내합니다.

❖ 실습 콘텐츠 안내:



Step 2

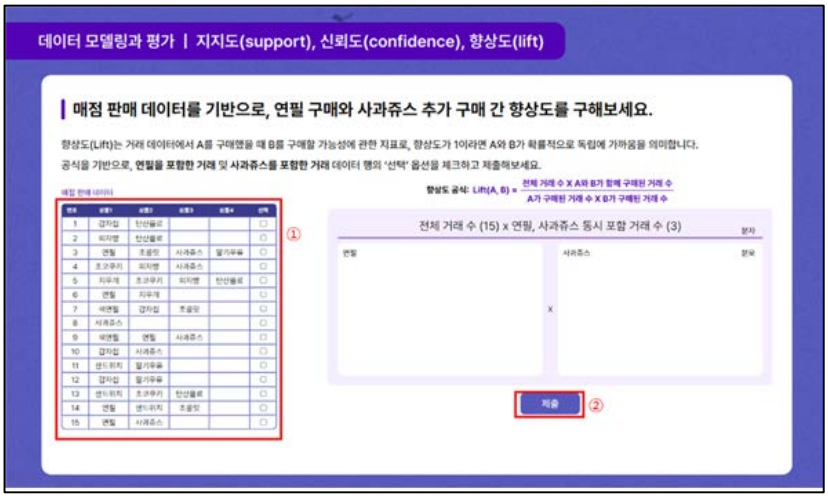
연필 -> 사과 주스 구매에 대한 신뢰도 계산하기

기능 사용법

① 행 선택 기능: 조건에 해당하는 행을 선택하는 기능입니다. 표의 맨 오른쪽부분의 박스를 체크하면 선택이 이루어지고, 오른쪽 분모 영역에 데이터 행이 추가 됩니다.

② 제출 버튼: 분모 영역에 추가된 데이터에 대한, 정/오답 여부를 확인하는 버튼입니다. 정답일 경우, 해당 조건에 대한 신뢰도를 함께 안내합니다.

❖ 실습 콘텐츠 안내:



Step 3

연필 구매와 사과주스 추가 구매 간 향상도 계산하기

기능 사용법

① 행 선택 기능: 조건에 해당하는 행을 선택하는 기능입니다. 표의 맨 오른쪽부분의 박스를 체크하면 선택이 이루어지고, 오른쪽 분모 영역에 각각 데이터 행이 추가 됩니다.

② 제출 버튼: 분모 영역에 추가된 데이터에 대한, 정/오답 여부를 확인하는 버튼입니다. 정답일 경우, 해당 조건에 대한 향상도를 함께 안내합니다.

주의 향상도를 구하는 실습에서 두가지 조건이 있으나, 선택은 나누어서 하지 않습니다. 두 조건에 전부 해당하는 것은 양쪽에 추가가 되고, 조건 하나에만 해당하는 것은 각 조건의 영역에 개별로 추가됩니다.

❖ **선생님을 위한 팁!**

다른 조건을 가정하여 지지도, 신뢰도, 향상도 식을 기반으로 값을 도출하는 실습을 진행해도 좋습니다.

3. 세부 콘텐츠

◆ 개요

15차시. 장바구니 분석

성취기준, 학습목표, 학습내용, 개발유형

성취기준	[12데과03-04] 데이터 간의 관계를 분석하고 상호 연관성을 파악하여 결과의 의미를 해석한다.
학습 목표	• 장바구니 분석을 해볼 수 있다. • 연관 분석의 지표들을 해석하여 적용할 수 있다.
대상 학년	
연계교육과정	정보교과
세부콘텐츠	1. (강의실) 동영상 콘텐츠 2. (실험실) 실습형 콘텐츠

◆ 구성

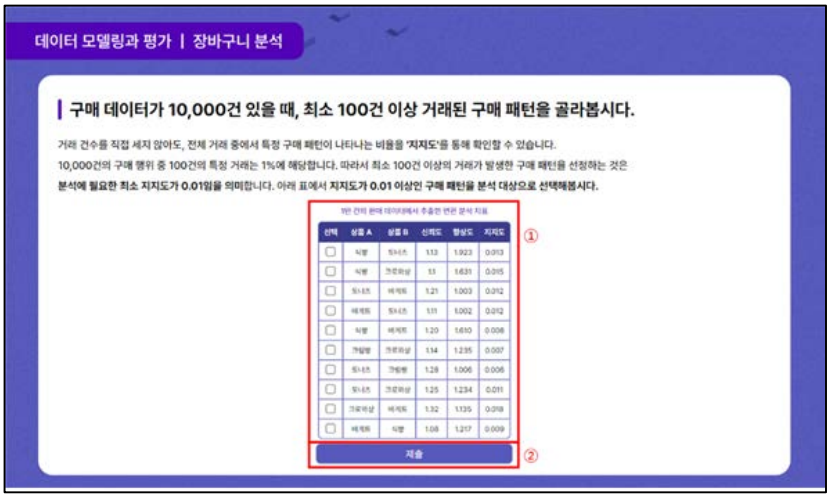
강의실

- ❖ **주제명:** 장바구니 분석 이해하기
- ❖ **세부 주제**
  1) 아프리오리(Apriori) 알고리즘
  2) 데이터 분석 - 최소 지지도 선정
- ❖ **콘텐츠 개발 목적:** 연관 분석 지표들을 바탕으로 상품을 진열할 수 있다.
- ❖ **선생님을 위한 팁!**
  데이터 수집의 중요성 차시와 연계하여 활용가능한 데이터를 찾을 수 있는 사이트를 추가적으로 안내할 수 있다.

실험실

- ❖ **주제명:** 장바구니 분석
- ❖ **세부 주제**
  1) 신뢰도 및 향상도에 대한 해석 실습
  2) 아프리오리 알고리즘 적용 데이터 분석
- ❖ **콘텐츠 개발 목적:** 데이터를 기반으로 실생활에 유의미한 인사이트를 도출하는 과정 실습

❖ 실습 콘텐츠 안내:



Step 1

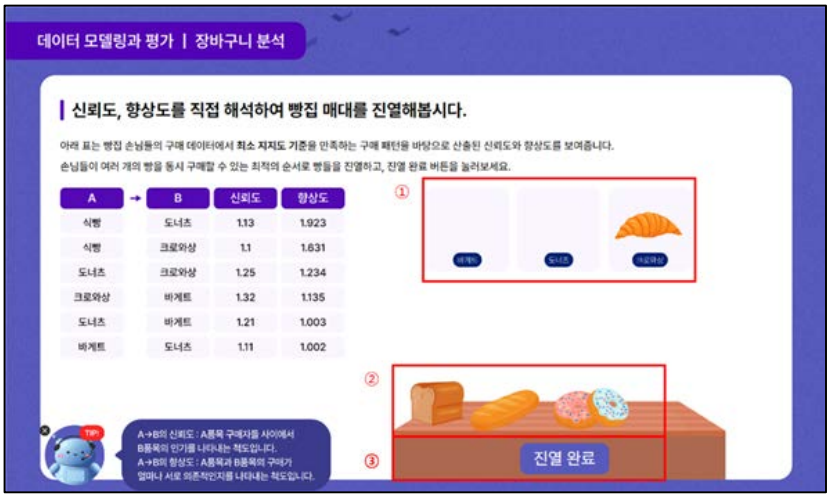
지지도 기반 분석 대상 구매 패턴 추출

기능 사용법

① 데이터 선택: '선택' 열에 있는 체크박스를 클릭하여, 발문의 조건에 해당하는 구매 패턴을 분석 대상으로 선택할 수 있습니다.

② 제출 버튼: 체크박스에 선택된 항목을 토대로 정/오답 여부를 판별하는 버튼입니다.

❖ 실습 콘텐츠 안내:



Step 2

신뢰도와 향상도를 고려하여 빵 진열 순서 정하기

기능 사용법

① 빵 기본 위치: 진열해야 할 빵이 있는 기본 위치입니다.

② 빵 진열 위치: ①에서 드래그하여 빵을 순서대로 진열해야 할 위치입니다.진열을 잘못했을 경우, 진열대에서 각 빵의 위치를 교체할 수 있습니다.

③ 진열 완료 버튼: 진열 완료 시 해당 버튼을 클릭하여 적절한 순서로 진열이 이루어졌는지 확인할 수 있습니다.

주의 빵을 진열하는 순서를 정할 때는 신뢰도와 향상도 두 가지를 모두 고려해야 합니다.

❖ **선생님을 위한 팁!**

실습 콘텐츠에서 제시된 경우의 수 외에 학생들이 다른 경우의 수를 스스로 생각해 보고, 해당 경우의 수에 대해서는 어떤 식으로 진열해야 할지 자유롭게 토론할 수 있도록 유도합니다.