

데이터가 쉬워지는

엑셀

데이터

분석

정성일 지음

길벗

엑셀 데이터 분석으로 이런 걸 할 수 있어요!

이 책은 실습을 통해 자연스럽게 데이터 분석 전문 지식과 기술을 습득할 수 있도록 구성했습니다. 특히, 컴퓨터 공학, 통계학, 기계 학습 알고리즘을 기반으로 한 실습은 데이터 분야 전문가에게도 충분히 유용한 정보를 제공합니다.

데이터 정리하는 방법과 핵심 단축키

관련 기술 ▶ 관계형 데이터베이스, 행 기반 데이터 구조, SQL

엑셀에는 데이터베이스와 SQL의 개념이 포함되어 있습니다. SQL은 데이터를 가공하거나 조회하는 데 사용되는 언어로 SQL의 개념과 단축키를 활용하면 워크시트를 더욱 빠르고 효율적으로 가공할 수 있습니다.

조건의 종류

[범례] 장의 '조건 종류'는 6개의 옵션을 선택할 수 있습니다. 대부분의 경우 [왼쪽 외곽]을 그대로 사용하지만, 각 옵션의 동작 방식을 알아보면 다음과 같습니다.

1. 왼쪽 외곽(첫 번째의 모두, 두 번째의 일치하는 행)

[범례] 장의 [조건 종류]에서 왼쪽은 첫 번째 테이블, 오른쪽은 두 번째 테이블을 의미합니다. 다음과 같이 두 개의 데이터 영역이 있고 키 값으로 '상품코드'를 사용할 경우 [범례] 장의 [조건 종류]에서 [왼쪽 외곽]을 선택하면 첫 번째 표는 전부 가져오지만, 두 번째 표에서는 [상품코드]가 일치하는 값만 가져옵니다.

상품코드	상품명
A	사과
B	배
C	수박
D	참외
E	멜론

상품코드	가격
A	1,000
C	5,000
D	25,000
F	2,000
G	8,000

상품코드	상품명	가격
A	사과	1,000
B	배	
C	수박	5,000
D	참외	25,000
E	멜론	

기초 통계와 피벗 테이블 이용 방법

관련 기술 ▶ BI(데이터 탐색 및 대시보드 툴), 기초 통계, 차원과 측정값

피벗 테이블은 데이터를 요약하고 탐색하기 위한 강력한 도구입니다. 이 책에서는 평균을 중심으로 기초 통계를 쉽게 설명하고 BI 도구의 개념을 바탕으로 피벗 테이블의 사용법을 체계적으로 안내합니다.

평균 (AVERAGE)

중심

- 중앙값 (MEDIAN)
- 최빈값 (MODE)

산포도

- 분산 (VAR)
- 표준편차 (STDEV)

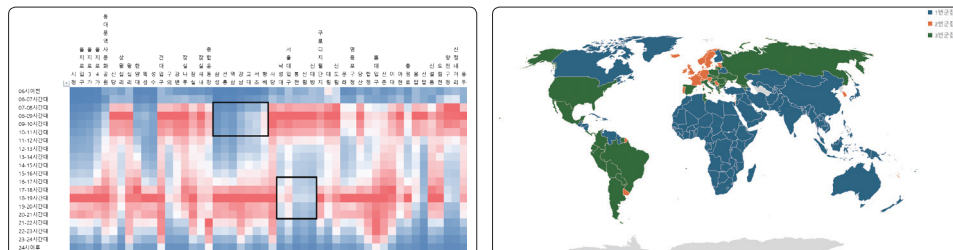
위치

- 사분위수 (QUARTILE)
- 백분위수 (PERCENTILE)

차트를 활용한 데이터 분석 방법

관련 기술 ▶ 생존 곡선, 히스토그램, 지도 차트, 상자 수염 차트, 팔레티형 차트

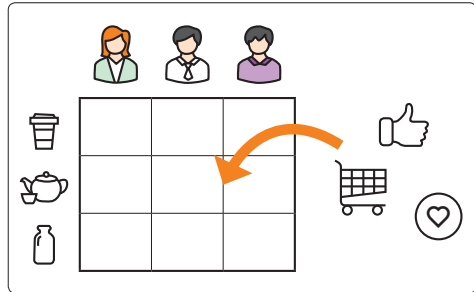
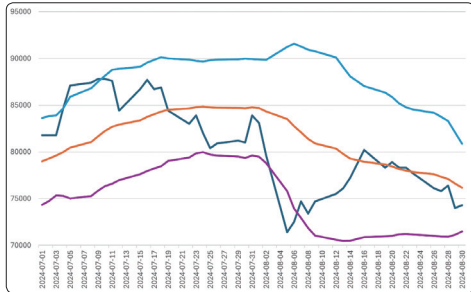
차트의 조립 순서가 있습니다. 무작정 따라하기만 하면 누구나 쉽게 차트를 통해 데이터를 분석할 수 있습니다. 공식적으로 제공하지 않는 차트부터 지도 차트까지 간단한 방법으로 고급 시각화를 구현하는 방법을 소개합니다.



통계 기반 분석 방법

관련 기술 ▶ 이동평균, 볼린저 밴드, 상관 분석, 코사인 유사도, 표준화 점수(Z-Score)

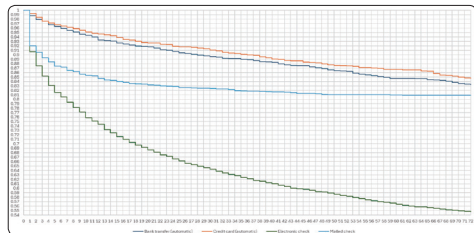
평균과 표준편차를 제대로 이해하고 활용하면, 지금까지 상상하지 못했던 다양한 방식으로 데이터를 분석할 수 있습니다. 이 책에서는 여러 CASE를 통해 통계치를 활용한 데이터 분석과 해석 방법을 알기 쉽게 설명합니다.



마케팅 및 비즈니스 분석 방법

관련 기술 ▶ RFM모델, 코호트 분석, 퍼널 분석

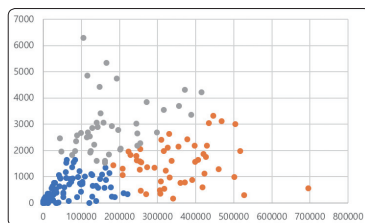
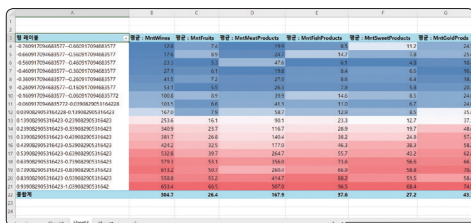
사회학, 경제학, 통계학을 기반으로 발전한 다양한 데이터 분석 방법을 통해, 전략 수립과 의사결정은 물론 제품 개발, 마케팅, 프로모션 기획 등 실무에 바로 적용할 수 있는 구체적인 방법을 소개합니다.

[illegible]

알고리즘 기반 분석 방법

관련 기술 ▶ 협업 필터링, 군집 분석

엑셀을 활용해 현재 데이터 분석가들 사이에서 인기 있는 고급 데이터 분석 알고리즘을 구현할 수 있습니다. 이 책에서는 엑셀을 보다 전문적으로 활용하는 방법을 소개합니다.



데이터를 빠르게 정리하는 방법

1장

EXCEL
DATA
LITERACY

엑셀이 쉬워지는 데이터 구조 만들기

(CASE 01) 엑셀이 추천하는 데이터 구조	019
엑셀 작업이 많아지는 이유	019
행 기반 데이터 구조의 이해	020
열 기반 데이터를 행 기반으로 바꾸는 방법	022
잘못 만들어진 행 기반 데이터 구조 바로잡기	024
(CASE 02) 여러 데이터 합치기	029
데이터가 분산되는 이유와 관리 방법	029
VLOOKUP 함수로 데이터 병합하기	031
파워 쿼리로 데이터 병합하기	035
(전문가의 조언) 조인의 종류	042
같은 양식의 여러 개 파일을 병합하기	044

2장

EXCEL
DATA
LITERACY

빠르게 찾고 한 번에 정리하기

(CASE 01) 데이터 정리에 유용한 단축키 모음	049
데이터 영역 확인 및 전체 선택	049
셀 선택한 후 행 단위로 삭제/삽입	051
근처에 있는 셀로 빈 셀 채우기	053
데이터 영역 안에서 빠르게 이동하기	056
(CASE 02) 필터와 정렬로 데이터 살펴보기	060
자동 필터 활용 방법	060
고급 필터 활용 방법	069
데이터 정렬 활용 방법	073
(CASE 03) 중복 데이터와 빈 셀 처리하기	076
중복 데이터와 빈 셀의 위험성	076
중복 행 삭제하기	077

빈 셀이 있는 행을 한 번에 삭제하는 방법	080
빈 셀을 한 번에 채우는 방법	081

둘째
마당

제대로 배우는 데이터 탐색과 시각화

1장

EXCEL
DATA
LITERACY

데이터가 쉬워지는 탐색과 요약

(CASE 01) 피벗 테이블로 데이터 요약하기	087
차원과 측정값의 개념	087
피벗 테이블로 데이터 요약하기	088
(전문가의 시선) 피벗 테이블의 차원과 측정값	091
(전문가의 시선) 피벗 테이블을 활용한 데이터 분석	096
(전문가의 조언) 피벗 테이블의 값 표시 형식	097
(CASE 02) 통계 함수의 이해와 표 활용 방법	099
평균과 기초 통계 함수	099
표를 활용해 빠르게 통계값 탐색하기	103
(전문가의 조언) 표의 기능적 제약	109

2장

EXCEL
DATA
LITERACY

차트만 그려도 알 수 있는 인사이트 도출 방법

(CASE 01) 히스토그램을 활용한 분포 분석	111
히스토그램의 구조와 활용	111
히스토그램을 활용한 분포 분석	112
(전문가의 시선) 평균을 중심으로 밀집된 히스토그램	116
(전문가의 시선) 밀집된 구간이 여러 개인 히스토그램	117
(전문가의 시선) 한쪽으로 치우친 히스토그램	120
(CASE 02) 분산형 차트로 상관관계 시각화하기	121
분산형 차트의 구조와 활용	121

분산형 차트로 열별 상관관계 알아보기	122
(전문가의 시선) 분산형 차트로 상관관계를 분석하는 방법	125
(전문가의 시선) 분산형 차트와 상관계수를 함께 살펴보는 이유	127
(CASE 03) 히트맵으로 패턴 발견하기	128
히트맵의 구조와 활용	128
히트맵 차트로 혼잡도 시각화하기 ①	129
(전문가의 시선) 히트맵 차트로 패턴을 분석하는 방법	130
히트맵 차트로 혼잡도 시각화하기 ②	131
(전문가의 시선) 히트맵 차트로 데이터를 탐색하며 패턴 찾기	138

셋째
마당

사례별 고객 데이터 분석 방법

1장

EXCEL
DATA
LITERACY

고객 관리의 시작, 우수 고객과 속성 분석

(CASE 01) RFM 모델로 우수 고객 분류하기	143
우수 고객의 실질적인 의미	143
RFM으로 고객 분류하기	144
(전문가의 조언) LOOKLUP 함수의 동작 방식	154
(전문가의 시선) RFM 점수별 고객 현황 살펴보기	156
RFM점수별 소비 패턴 분석	157
(전문가의 시선) RFM 점수별 고객 분석 및 활용 방법	163
(CASE 02) 평행 좌표 차트로 고객 속성 분석하기	164
고객 속성별 행동 패턴 분석	164
출생 연도별 평행 좌표 차트 그리기	166
(전문가의 시선) 출생 연도별 평행 좌표 차트 분석	170
(전문가의 시선) 소득별 평행 좌표 차트 분석	173

2장

EXCEL
DATA
LITERACY

그로스 해킹의 핵심, 이탈 고객 분석

(CASE 01) 코호트 분석으로 고객 유지율 살펴보기	177
코호트 분석이란	177
코호트 분석하기	178
(전문가의 시선) 코호트 분석 결과를 해석하는 방법	184
(CASE 02) 생존 곡선으로 이탈 고객 분석하기	186
생존 곡선이란	186
생존 곡선 그리기	187
(전문가의 조언) 계단식 차트를 생성하는 방법	196
(전문가의 시선) 생존 곡선을 활용해 이탈 고객을 분석하는 방법	197
(전문가의 시선) 속성별 생존 곡선 비교 분석 방법	200
(전문가의 조언) 속성별 생존 곡선 만드는 방법	204
(CASE 03) 깔때기형 차트로 이탈 행동과 원인 분석하기	205
퍼널 분석과 그로스 해킹	205
깔때기형 차트로 퍼널 분석하기	206
(전문가의 시선) 월별 깔때기형 차트를 활용한 이탈 원인 분석	214
(전문가의 시선) 상품별 깔때기형 차트를 활용한 이탈 원인 분석	220

3장

EXCEL
DATA
LITERACY

[디지털 플랫폼의 성장 비법, 페르소나와 상품 추천

(CASE 01) 코사인 유사도로 페르소나 고객 찾기	223
페르소나 마케팅이란	223
코사인 유사도를 이용한 실제 고객 찾기	224
페르소나 기반 고객 추출하기	225
(전문가의 시선) 코사인 유사도별 고객 행동 살펴보기	238
(전문가의 시선) 페르소나 유사도를 통한 고객 속성 비교	240
(CASE 02) 상관계수로 추천 상품 찾기	241
상품 기반 추천과 활용 방법	241
상품 기반 추천을 위한 유사도 측정표	242
상관계수로 추천 상품 찾기	244
(전문가의 시선) 상품 기반 추천 활용 방법	252

1장

EXCEL
DATA
LITERACY

주식으로 알아보는 추세와 상관관계

(CASE 01) 스파크라인으로 한 눈에 추세 분석 하기	257
국내 주식 시세 데이터 가져오기	257
여러 종목 시세 데이터 한 번에 가져오기	265
스파크라인으로 종목별 추세 한 눈에 보기	274
(전문가의 시선) 시계열 데이터의 추세	278
(CASE 02) 상관계수로 시장을 분석하고 대응하는 방법	279
시계열 데이터의 상관관계 분석	279
지수와 종목별 상관관계 분석하기	280
(전문가의 시선) 지수와 종목별 상관관계 분석	282
(전문가의 시선) 양의 상관관계를 가진 분산형 차트	283
(전문가의 시선) 음의 상관관계를 가진 분산형 차트	286
시계열 데이터의 평균과 표준편차	286
이동평균과 볼린저 밴드로 주식 매매하기	287
(전문가의 시선) 볼린저 밴드 차트를 읽는 방법	292
(전문가의 시선) 주식 가상 매매 결과 분석	296

2장

EXCEL
DATA
LITERACY

이상 징후를 탐지하고 모니터링하는 방법

(CASE 01) 상자 수염 차트로 이상치 찾기	299
실무에서 만나는 이상치	299
상자 수염 차트의 모양과 의미	300
상자 수염 차트 그리기	302
(전문가의 시선) 이상치 탐색과 상자 수염 차트 활용 방법	305
(CASE 02) 평균과 표준편차로 이상 징후 탐지하기	308
평균과 표준편차를 활용한 이상치 탐지	308
표준화 점수로 이상치 추출하기	309
(전문가의 시선) 이상 징후 탐지와 사후 분석	316

3장

EXCEL
DATA
LITERACY

군집 분석을 활용한 고급 인사이트 분석

(CASE 01) 해 찾기로 군집 만들기	321
군집 분석의 활용 방법	321
K-Mean 군집 분석의 동작 방식 알아보기	322
(전문가의 조언) 해 찾기 기능 설치하기	323
해 찾기로 군집 만들기	324
(전문가의 시선) 분산형 차트를 활용하여 군집 개수 결정하기	328
(전문가의 조언) 해 찾기 해법의 종류	342
(CASE 02) 군집에서 인사이트 발굴하기	343
패턴의 발견과 활용	343
군집 분석 결과 해석하기	344
(전문가의 시선) 군집의 기본 속성을 활용한 유형 정의	350
(전문가의 시선) 군집의 확장 속성을 활용한 인사이트 발굴	352
(전문가의 시선) 지도차트를 활용한 인사이트 발굴	357
(전문가의 조언) 잘못 연결된 지역 정보 수정하기	358
찾아보기	359



실습예제 다운로드

이 책에 사용된 예제는 길벗출판사 홈페이지(www.gilbut.co.kr)에서 다운로드할 수 있습니다.
홈페이지 회원으로 가입하지 않아도 누구나 실습 예제 파일을 다운로드 할 수 있습니다.



1 길벗 홈페이지에서 '엑셀 데이터 분석'을 검색하세요.



자료실



2 해당 도서의 페이지에서 [자료실]을 클릭해 실습 예제 파일을 다운로드하세요.



3 압축을 해제하여 실습에 활용해 보세요.

2

들
재
마
당

제대로 배우는 데이터 탐색과 시각화

1장 데이터가 쉬워지는 탐색과 요약

2장 차트만 그려도 알 수 있는 인사이트 도출 방법

1

DATA

LITERACY

장

데이터가 쉬워지는 탐색과 요약

엑셀을 활용하면 데이터를 쉽게 탐색할 수 있습니다. 방향키로 셀을 이동하며 값을 하나씩 살펴보는 건 엑셀만의 장점이죠. 하지만, 피벗 테이블과 표를 활용하면 업무가 더욱 효율적으로 바뀝니다. 이번 장에서는 피벗 테이블과 표, 그리고 요약에 사용되는 통계값을 효과적으로 사용하는 방법을 알아보겠습니다.



피벗 테이블로 데이터 요약하기

기업의 데이터 활용이 강조되면서 BI 도구의 도입이 활발해졌습니다. Tableau, 파워 BI 등의 BI 도구는 데이터를 요약하고 시각화하는 데 특화되어 있죠. 엑셀의 피벗 테이블은 BI 도구만큼 유연하고 다양한 분석이 가능하면서도, 엑셀 안에서 작동하여 접근성과 활용도가 높다는 장점이 있습니다. 피벗 테이블의 핵심 개념과 기본 사용법을 익히면 엑셀에서도 고도화된 데이터 요약 수행할 수 있습니다.

✓ 차원과 측정값의 개념

‘BI(Business Intelligence)’ 도구나 데이터 시각화 솔루션에서는 ‘차원(Dimension)’, ‘측정값(Measure)’이라는 용어가 자주 등장합니다.

차원은 데이터를 그룹화하여 요약하고 비교 분석할 수 있게 해줍니다. 주로 범주형 데이터로 구성되어 있으며, 데이터의 성격에 따라 계층 구조로 이루어져 있습니다. 예를 들어, 매출 데이터의 ‘지역’ 차원이 [국가]-[지역]-[도시] 등의 계층으로 구성된 것을 생각하면 쉽게 이해할 수 있습니다.

측정값은 차원에 따라 구분된 데이터에 대해 실제 계산을 수행하는 것을 말합니다. 대표적인 측정값에는 ‘합계’, ‘평균’, ‘개수’, ‘최대’, ‘최소’ 등이 있으며, 매출 데이터의 상품 카테고리별 매출액을 합산하거나, 지역별 평균 판매량을 구하는 것이 측정값 계산에 해당합니다. 측정값은 주로 숫자형 데이터를 대상으로 하며, 하나의 차원을 여러 측정값으로 동시에 계산할 수 있습니다.

원본 데이터: 매출 기록

지역	분류	채널	매출액	판매량
서울	의류	온라인	1,000,000	100
서울	의류	오프라인	2,000,000	150
서울	가전	온라인	1,500,000	50
서울	가전	오프라인	3,000,000	80
부산	의류	온라인	800,000	120
부산	의류	오프라인	1,200,000	100
부산	가전	온라인	2,000,000	60
부산	가전	오프라인	1,800,000	70

범주형 값
(주로 차원에 사용)

숫자형 값
(주로 측정값에 사용)

요약: 분류/채널별 평균 판매량

구분	온라인	오프라인
의류	110	125
가전	55	75

차원

요약: 지역/분류별 최대/최소 매출

지역	분류	최대매출	최소매출
서울	의류	2,000,000	1,000,000
서울	가전	3,000,000	1,500,000
부산	의류	1,200,000	800,000
부산	가전	2,000,000	1,800,000

측정값

차원과 측정값은 BI 도구를 비롯한 다양한 시각화 솔루션에서 공통적으로 사용되는 만큼, 그 원리를 이해하는 것만으로도 데이터를 자유자재로 요약하여 분석할 있습니다. 엑셀의 피벗 테이블에서는 ‘행’과 ‘열’에 차원을 배치하고 ‘값’에 측정값을 배치합니다.

실습

피벗 테이블로 데이터 요약하기

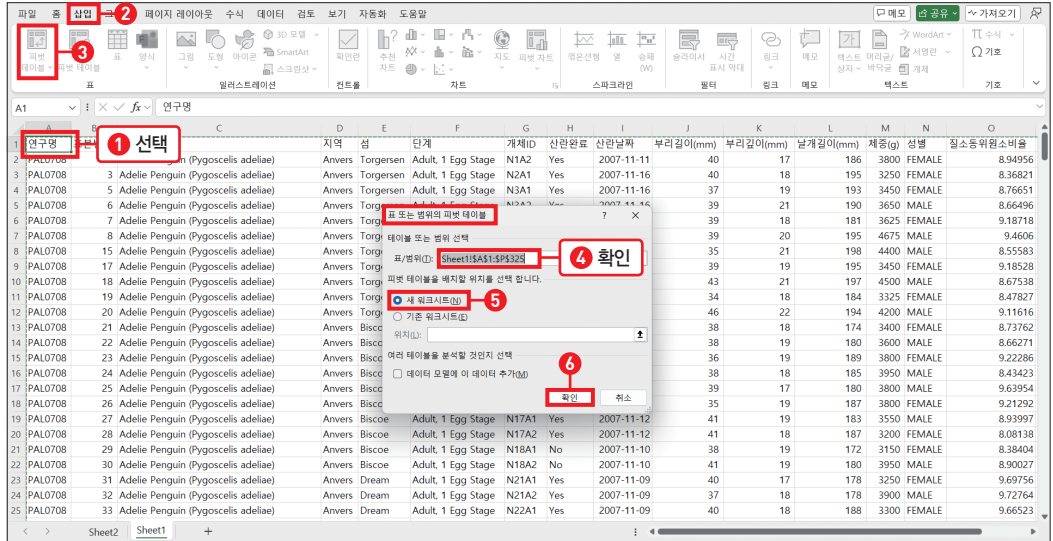
🔗 CASE_01

실습 예제는 ‘크리스틴 고먼’ 박사와 연구팀이 수집한 ‘팔머 펭귄 데이터’입니다. 이 데이터는 남극의 팔머 군도에 서식하는 ‘아델리(Adélie)’, ‘턱끈(Chinstrap)’, ‘젠투(Gentoo)’ 펭귄의 데이터로 생태 및 해양 환경 변화를 이해하기 위한 목적으로 수집되었습니다. 데이터에는 펭귄의 종과 서식지(섬)을 비롯하여, 부리와 날개 등의 신체 측정치, 그리고 깃털과 혈액 등에서 채취한 동위원소 비율이 포함되어 있습니다.

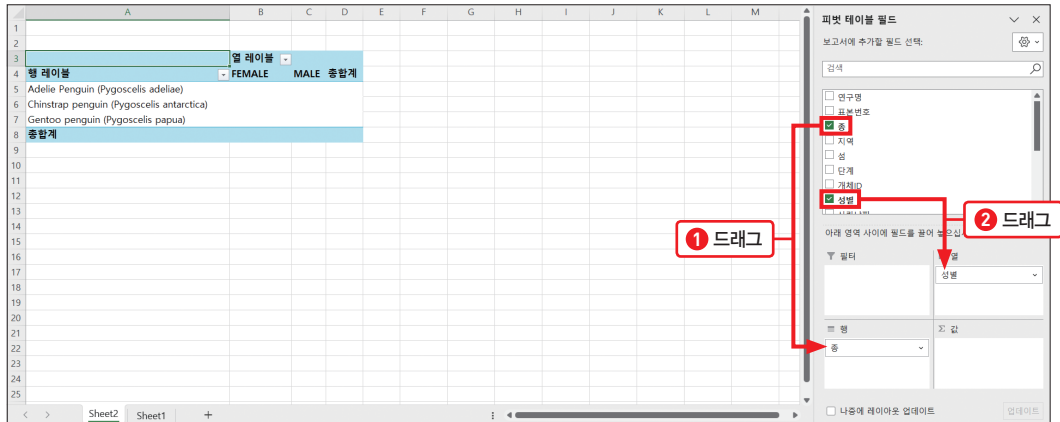
머리글	데이터 설명
연구명	연구 프로젝트의 이름 및 식별자
표본번호	개별 표본에 부여된 고유 번호
종	펭귄의 종(예: 아델리 펭귄, 턱끈 펭귄, 젠투 펭귄)
지역	연구가 수행된 지역(예: Anvers 등)
섬	펭귄이 관찰된 섬(예: Torgersen, Biscoe, Dream 등)
단계	펭귄의 생물학적 상태(예: 성체, 알 1개 단계 등)
개체ID	각 개체에 부여된 고유 식별자
산란완료	산란 완료 여부(예: Yes/No)
산란날짜	알을 낳은 날짜
부리길이(mm)	펭귄의 부리 길이(mm 단위)
부리깊이(mm)	펭귄의 부리 깊이(mm 단위)
날개길이(mm)	펭귄의 날개 길이(mm 단위)
체중(g)	펭귄의 체중(g 단위)
성별	펭귄의 성별(예: 수컷, 암컷)
질소동위원소비율	질소 동위원소 비율(생태적 위치 및 식성 연구에 사용)
탄소동위원소비율	탄소 동위원소 비율(생태적 위치 및 식성 연구에 사용)

1 | 피벗 테이블 기본 사용법

- ① 데이터 영역 중 임의의 셀을 선택한 다음, 메뉴의 [삽입]-[표]-[피벗 테이블]을 클릭합니다.
- ② [표 또는 범위의 피벗 테이블] 창에 데이터 범위가 제대로 입력되었는지 확인한 후, [새 워크시트]를 선택하고 [확인]을 클릭합니다.



- ③ '피벗 테이블 필드'에서 [중]을 [행], [성별]을 [열]로 드래그합니다.



- ④ [체중(g)]을 [값]으로 드래그한 후, [값]의 [합계 : 체중(g)]을 클릭하고 [값 필드 설정]를 선택합니다.

- ⑤ 여기서는 펭귄의 종과 성별에 따른 평균 체중을 비교해보겠습니다. [값 필드 설정] 창의 [값 요약 기준] 탭에서 '값 필드 요약 기준'을 [평균]으로 변경하고 [확인]을 클릭합니다.

- ⑥ 피벗 테이블에 [종]과 [성별]을 기준으로 [평균 체중]이 요약됩니다.

	A	B	C	D	E	F
1						
2						
3	평균 : 체중(g)	열 레이블				
4	행 레이블	FEMALE	MALE	총합계		
5	Adelie Penguin (Pygoscelis adeliae)	3367	4054	3703		
6	Chinstrap penguin (Pygoscelis antarctica)	3527	3939	3730		
7	Gentoo penguin (Pygoscelis papua)	4680	5489	5091		
8	총합계	3867	4565	4214		
9						
10						



피벗 테이블의 차원과 측정값

피벗 테이블은 요약할 '차원'을 '피벗 테이블 필드'의 [행], [열]에 추가하고, '측정값'은 [값]에 추가하는 방식으로 동작합니다. 그리고 측정값의 계산 방식은 '값 필드 설정'에서 변경할 수 있습니다.

차원

- ☐ 연구명
- ☐ 표본번호
- ☒ **중**
- ☐ 지역
- ☐ 섬

- ☐ 날개길이(mm)
- ☐ 체중(g)
- ☒ **성별**
- ☐ 질소동위원소비율
- ☐ 탄소동위원소비율

측정값

- ☐ 부리깊이(mm)
- ☐ 날개길이(mm)
- ☒ **체중(g)**
- ☐ 성별
- ☐ 질소동위원소비율

	A	B	C	D	E
1					
2					
3	평균 : 체중(g)	열 레이블			
4	행 레이블	FEMALE	MALE	총합계	
5	Adelle Penguin (Pygoscelis adeliae)	3367	4054	3703	
6	Chinstrap penguin (Pygoscelis antarctica)	3527	3939	3730	
7	Gentoo penguin (Pygoscelis papua)	4680	5489	5091	
8	총합계	3867	4565	4214	
9					
10					

요약된 피벗 테이블을 살펴보면 차원으로 추가한 [종]과 [성별]은 [행]과 [열]의 머릿글에 각각 배치되었습니다. 그리고 행과 열이 교차되는 지점에 측정값이 평균값으로 표시됩니다. 작업량으로 살펴보면 차원으로 선택한 세 개의 종과 두 개의 성별이 조합되어 여섯 개의 칸을 만들어지고, 각 조합별 평균을 계산하는 작업이 똑딱 처리된 셈입니다.

이러한 형태의 표는 기업에서 흔히 사용하는 재무장표나 대시보드(현황판)에서 자주 볼 수 있습니다. 복잡한 데이터에서 핵심을 추출하여 목적에 따라 요약한 후, 현황을 파악하고 중요한 의사결정을 하는데 사용할 수 있는 피벗 테이블은 데이터를 요약할 때 값을 채우는 작업을 쉽게 할 수 있습니다.

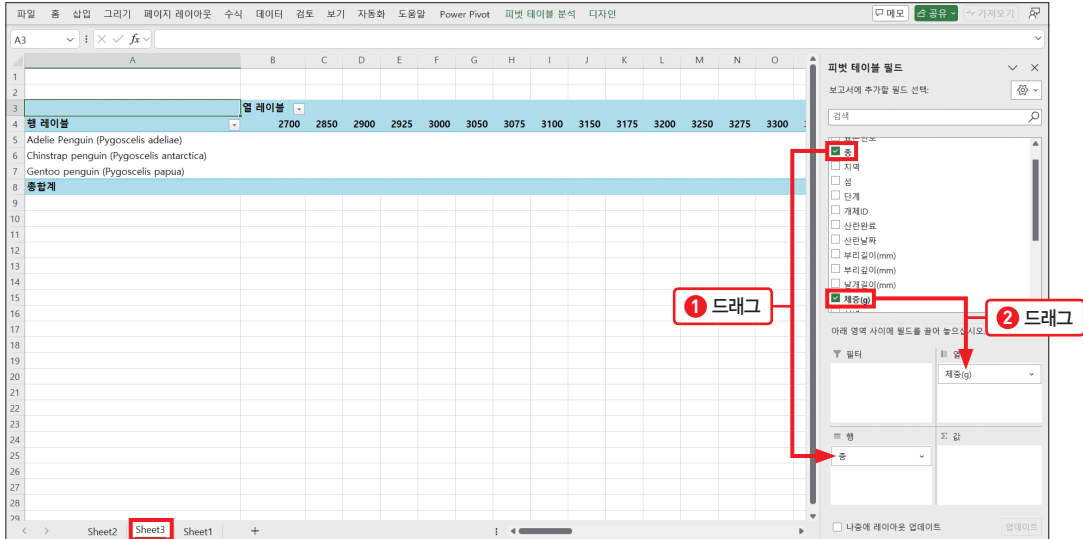
피벗 테이블은 데이터를 필터링하고 함수를 적용하는 것보다 작업이 간단하고, 항목을 자유자재로 바꾸가며 데이터를 탐색할 수 있기 때문에 데이터 분석가들에게도 각광받는 도구입니다. '엑셀 5.0'에 처음으로 추가된 피벗 테이블은 꾸준히 발전하면서 별도의 분석 도구가 필요 없을 만큼 강력한 기능을 갖추게 되었죠.

실습에서도 간단한 작업만으로 종과 성별을 교차 계산하여 젠투 펭귄의 수컷이 가장 체중이 많이 나가고, 아델리 펭귄 암컷이 가장 작은 것을 확인할 수 있습니다. 행과 열의 끝에 있는 총합계에서는 각 항목의 요약 값도 쉽게 확인할 수 있어, 행의 총합계를 통해 젠투 펭귄이 다른 두 종에 비해 체중이 큰 걸 볼 수 있으며, 열의 총합계에서 모든 종의 수컷이 암컷보다 많이 나가는 것을 쉽게 확인할 수 있습니다.

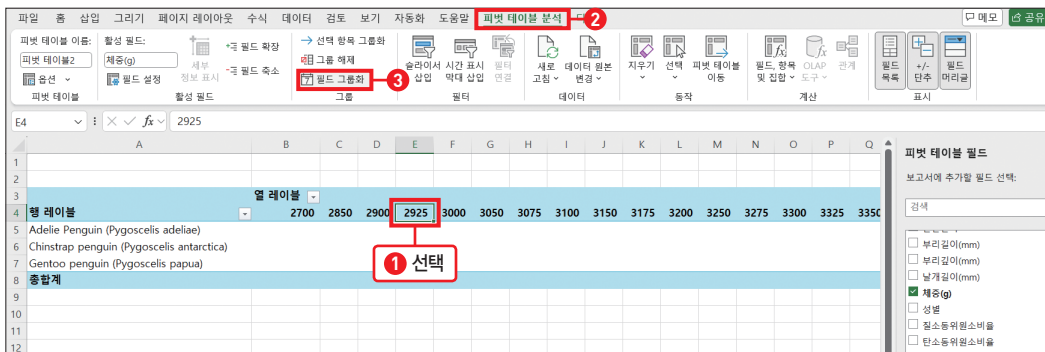
2 | 그룹화 기능을 활용한 분포 분석

① [Sheet1] 시트의 임의의 데이터 영역을 선택하고 [삽입]-[표]-[피벗 테이블]을 클릭하여 새 워크시트([Sheet3])에 피벗 테이블을 추가합니다.

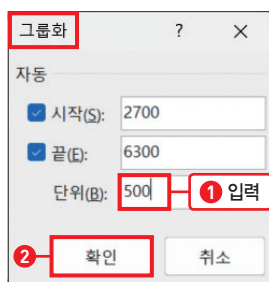
② [Sheet3] 시트의 '피벗 테이블 필드'에서 [중]을 [행], [체중(g)]을 [열]로 드래그합니다.



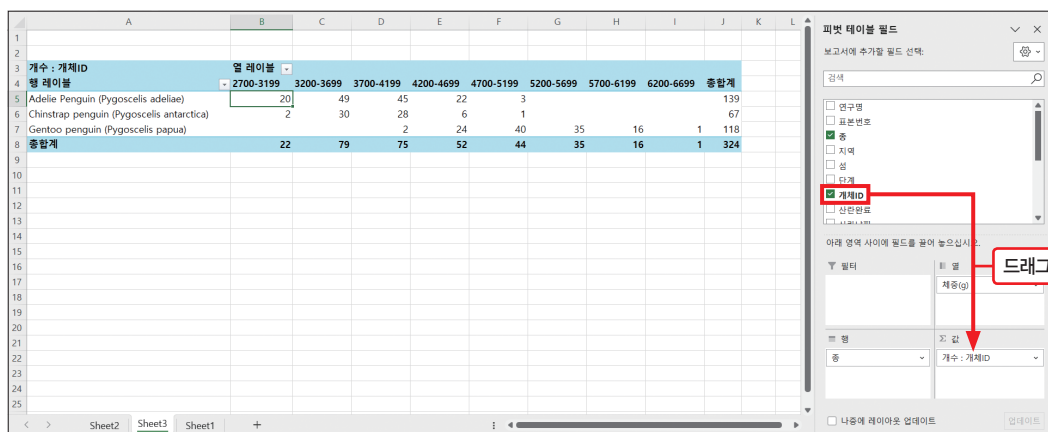
③ 체중과 같은 숫자형 값을 열에 추가하면 그림처럼 표가 길어져 데이터를 확인하기 어려워집니다. 이럴 때는 숫자형 값을 일정한 간격으로 묶어주는 [그룹화]를 이용하면 편리합니다. 열 레이블이 있는 [B4:CP4] 영역에서 임의의 셀을 선택한 후, 메뉴의 [피벗 테이블 분석]-[그룹]-[필드 그룹화]를 선택합니다.



- ④ [그룹화] 창에 데이터의 [단위]에는 기본값으로 '1000'이 입력되어 있습니다. 이대로 그룹화를 실행하면 '1,000' 단위로 네 개의 그룹이 생성됩니다. 여기서는 데이터를 좀 더 자세히 살펴보기 위해 '단위'를 '500'으로 변경 후, [확인] 클릭합니다.



- ⑤ '피벗 테이블 필드'에서 [개체ID]를 [값]으로 드래그합니다.



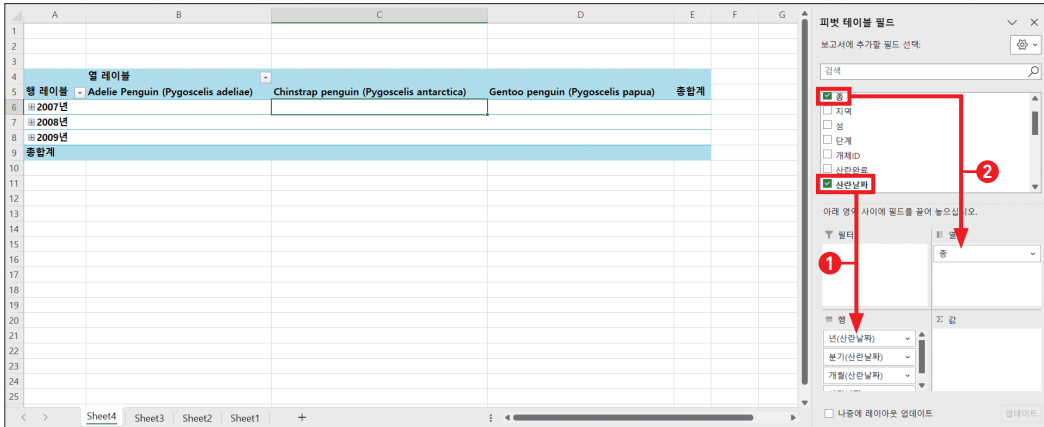
피벗 테이블에서는 보통 문자 등의 형태를 가진 범주형 값을 차원으로 활용하지만, 필요에 따라 숫자형 값도 차원에 추가할 수 있습니다. 단, 숫자형 값은 데이터에 따라 행과 열의 개수가 달라져 표가 길어질 수 있습니다. 이런 때에는 데이터를 일정한 간격으로 그룹화하여 사용합니다. 기업의 활용사례에서는 고객을 나이에 따라 10살 단위로 구분하거나, 가격대에 따라 상품을 구분하는 것을 예시로 들 수 있습니다.

그룹화 기능을 이용하면 각 그룹의 분포를 쉽게 확인할 수 있습니다. 실습 데이터를 예시로 부리와 날개 길이를 그룹화하거나 체중이 큰 개체와 작은 개체 등을 나눠 신체적 차이에 따라 개체를 분류한 후 분포를 파악할 수 있을 것입니다. 여기에서는 체중을 500g 단위로 그룹화함으로써 아델리 펭귄과 턱큰 펭귄의 크기별 분포를 확인함으로써 대부분의 아델리 펭귄이 젠투 펭귄보다 체중이 적게 나가지만, 일부 개체는 젠투 펭귄만큼 큰 것을 확인할 수 있습니다. 이러한 작업은 마치 실

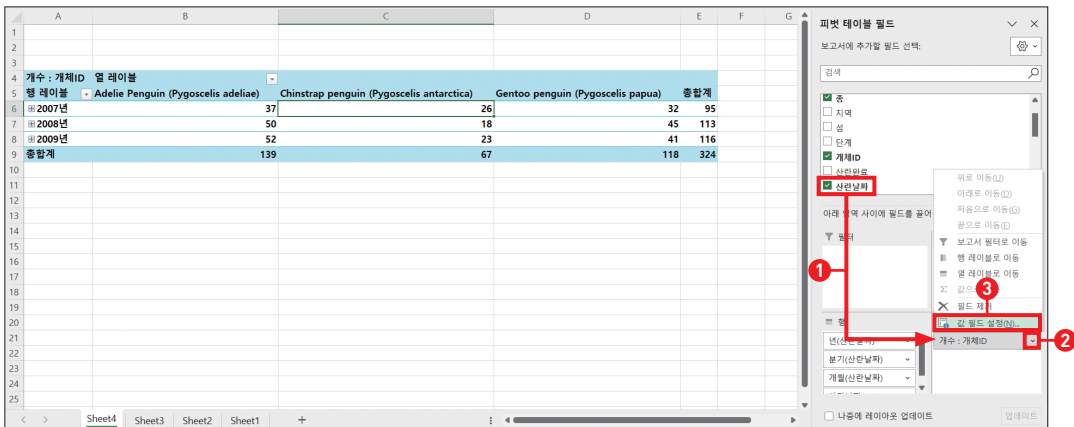
무에서 매출 규모에 따른 매장의 분포를 확인하는 것과 유사합니다.

3 | 값 표시 형식을 활용한 종별 비율 변화 분석

- ① [Sheet1] 시트의 데이터 영역에서 임의의 셀을 선택하고 [삽입]-[표]-[피벗 테이블]을 클릭하여 새 워크시트([Sheet4])에 피벗 테이블을 추가합니다.
- ② [Sheet4] 시트의 '피벗 테이블 필드'에서 [산란날짜]를 [행], [종]을 [열]로 드래그합니다.



- ③ '피벗 테이블 필드'에서 [개체ID]를 [값]으로 드래그 후, [값]의 [개수 : 개체ID]를 클릭하고 [값 필드 설정]을 선택합니다



- ④ [값 필드 설정] 창에서 [값 표시 형식] 탭을 선택한 다음, '값 표시 형식'에서 [행 합계 비율]을 선택하고 [확인]을 클릭합니다.

- ⑤ 행 방향 총합계가 100%로 요약되고 년도별/종별 비율이 계산됩니다. 여기서는 산란날짜를 기준으로 했기 때문에 년도별 태어난 종별 펭귄 비율로 해석하면 됩니다.

개수 : 개체ID	종	행 합계 비율
2007년	Adelie Penguin (Pygoscelis adeliae)	38.95%
2007년	Chinstrap penguin (Pygoscelis antarctica)	27.37%
2007년	Gentoo penguin (Pygoscelis papua)	33.68%
2008년	Adelie Penguin (Pygoscelis adeliae)	44.25%
2008년	Chinstrap penguin (Pygoscelis antarctica)	15.93%
2008년	Gentoo penguin (Pygoscelis papua)	39.82%
2009년	Adelie Penguin (Pygoscelis adeliae)	44.83%
2009년	Chinstrap penguin (Pygoscelis antarctica)	19.83%
2009년	Gentoo penguin (Pygoscelis papua)	35.34%
총합계	Adelie Penguin (Pygoscelis adeliae)	42.90%
총합계	Chinstrap penguin (Pygoscelis antarctica)	20.68%
총합계	Gentoo penguin (Pygoscelis papua)	36.42%
총합계	종 합계	100.00%



개체: 개체 ID	열 레이블			
행 레이블	Adelie Penguin (Pygoscelis adeliae)	Chinstrap Penguin (Pygoscelis antarctica)	Gentoo Penguin (Pygoscelis papua)	총합계
2007년	37	26	32	95
2008년	50	18	45	113
2009년	52	23	41	116
총합계	139	67	118	324

▲ 년도/종별 출생한 개체 수 계산

실습을 통해 팔머 군도의 연도별 펭귄 종 비율을 쉽게 확인할 수 있습니다. 3년간 출생한 총 개체 수를 합산하면 아델리 펭귄이 139마리로 가장 많고, 젠투 펭귄이 118마리로 그 뒤를 이었으며, 턱끈 펭귄은 67마리로 다른 두 종의 절반 수준에 그칩니다.

2008년에는 턱끈 펭귄의 출생 비율이 전년 대비 감소한 반면, 아델리와 젠투 펭귄의 비율은 상대적으로 증가했습니다. 이후 턱끈 펭귄의 출생 비율이 다소 회복되었지만, 여전히 2007년보다는 낮은 수준을 유지하고 있습니다.

이러한 내용을 종합해 보면, 팔머 군도에서 턱끈 펭귄은 다른 펭귄에 비해 개체 수가 적을 뿐만 아니라 출생률도 낮아지고 있는 추세입니다. 물론 이 데이터는 연구원들이 관찰한 개체에 한정된 것이므로, 팔머 군도 전체의 종 비율을 대변한다고 보기는 어렵습니다. 그럼에도 불구하고, 먹이와 서식지가 제한된 환경에서 이러한 추세가 지속된다면 턱끈 펭귄의 개체 수는 점차 감소할 것으로 예측할 수 있습니다.

피벗 테이블은 설정 방식에 따라 다양하게 활용할 수 있으며, 행, 열, 값을 지정하고 요약 방식을 선택하는 순서로 작업하면 대부분 쉽게 사용할 수 있습니다. **피벗 테이블은 작업 시간을 줄여줄 뿐만 아니라 원본 데이터에도 영향을 주지 않아 데이터 분석에 매우 유용합니다.** 이후 실습에서도 자주 활용될 예정이니, 데이터에 따라 상황별 활용 방법을 익혀보시기 바랍니다.

피벗 테이블의 값 표시 형식

피벗 테이블의 '값 표시 형식'은 표를 두어 번 가공한 것과 같은 작업 효과를 낼 수 있습니다. 실습에서는 다음과 같이 년도/종별 출생 개체 수를 계산한 후, 종별 개체 수를 년도별 합계로 나눠 비율을 계산되었습니다.

개체: 개체 ID	열 레이블			
행 레이블	Adelie Penguin (Pygoscelis adeliae)	Chinstrap Penguin (Pygoscelis antarctica)	Gentoo Penguin (Pygoscelis papua)	총합계
2007년	37	26	32	95
2008년	50	18	45	113
2009년	52	23	41	116
총합계	139	67	118	324

▲ 년도/종별 출생한 개체 수 계산

개체: 개체 ID	열 레이블			
행 레이블	Adelie Penguin (Pygoscelis adeliae)	Chinstrap Penguin (Pygoscelis antarctica)	Gentoo Penguin (Pygoscelis papua)	총합계
2007년	38.95%	27.37%	33.68%	100.00%
2008년	44.25%	15.93%	39.82%	100.00%
2009년	44.83%	19.83%	35.34%	100.00%
총합계	42.90%	20.68%	36.42%	100.00%

▲ 종별 개체수를 년도별 합계로 나눠 비율 계산

값 표시 형식은 표를 복사하거나 수식을 적용하지 않고도 다양한 계산을 할 수 있어, 다소 복잡해 보여도 익혀두면 유용하게 활용할 수 있습니다. 사용할 수 있는 옵션은 다음과 같습니다.

구분	값 표시 형식	설명
기본 값	계산 없음	피벗 테이블 필드에 지정된 계산 방식(합계, 평균 등)을 그대로 표시합니다.
	총합계 비율	피벗 테이블 전체 총합에 대한 백분율로 각 셀의 값을 표시합니다.
비율	열 합계 비율	전체 열 합계에서 해당 셀이 차지하는 비율을 백분율로 표시합니다.
	행 합계 비율	전체 행 합계에서 해당 셀이 차지하는 비율을 백분율로 표시합니다.
	[기준값]에 대한 비율	기준 셀을 설정하여 해당 셀과 비교한 비율을 계산하여 표시합니다.
	상위 행 합계 비율	피벗 테이블에서 행을 기준으로 상위 레벨의 값이 전체에서 차지하는 비율을 표시합니다.
	상위 열 합계 비율	피벗 테이블에서 열을 기준으로 상위 레벨의 값이 전체에서 차지하는 비율을 표시합니다.



히트맵으로 패턴 발견하기

히트맵 차트는 색상의 차이로 데이터의 크기나 빈도를 직관적으로 표현하는 시각화 기법으로 데이터의 전체적인 형태를 빠르게 파악하고 패턴을 발견하거나 이상치를 찾아내는 데 유용하게 활용합니다. 히트맵 차트는 엑셀이 제공하는 차트에는 포함되어 있지 않지만, 시트를 활용하면 손쉽게 생성할 수 있습니다. 이번 CASE에서는 엑셀에서 히트맵을 생성하는 방법과 데이터 분석에 활용하는 방법을 알아보겠습니다.

✓ 히트맵의 구조와 활용

히트맵은 1991년 ‘코맥 키니(Cormac Kinney)’가 복잡한 금융 시장 데이터를 시각화하기 위해 고안한 기법입니다. 크기가 작은 값은 차가운 색(파란색)으로, 크기가 큰 값은 따뜻한 색(빨간색)으로 구분하여 트레이더가 시장의 움직임과 패턴을 빠르게 파악할 수 있도록 고안했습니다.

히트맵의 기본 구조는 2차원 격자 형태로, 각 블록의 값 크기에 따라 색상을 채워서 생성합니다. 이를 응용하면 지도를 배경으로 기온, 트래픽, 인구 밀집도 등을 표현하거나, 웹사이트의 레이아웃에 고객 이용 빈도를 나타내는 등 다양한 형태로도 활용할 수 있습니다.

히트맵의 가장 큰 장점은 방대한 데이터를 한눈에 파악할 수 있는 점입니다. 수치 데이터를 하나씩 비교하는 것보다 색상의 변화로 데이터를 파악할 수 있으므로 방대한 데이터의 전체적인 패턴을 빠르게 인지할 수 있어 비슷한 색상으로 표현된 데이터의 집합을 찾아 유사한 크기나 빈도의 데이터를 한눈에 알아챌 수 있고, 주변 데이터의 색상과 큰 차이가 나는 데이터를 색상으로 구분할 수 있어 이상치를 빠르게 탐색할 수 있습니다.

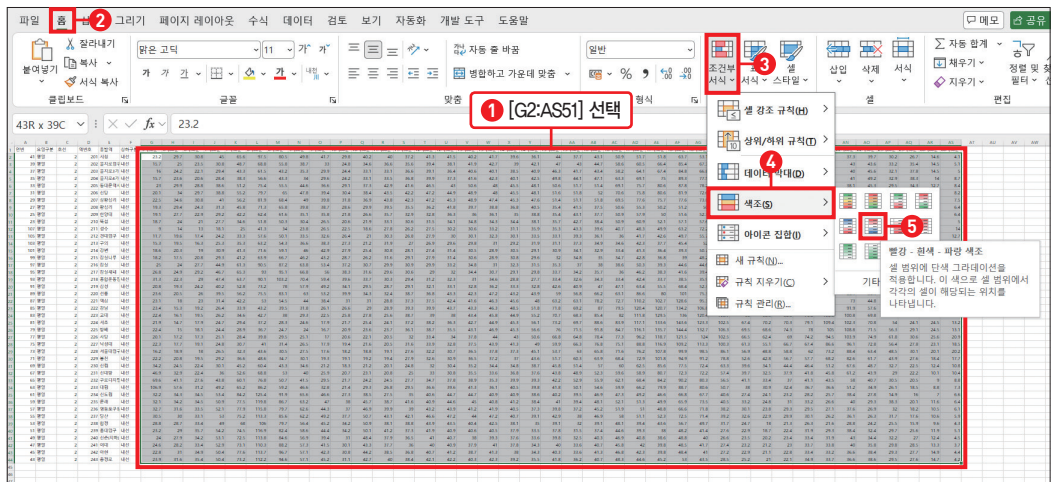
엑셀에서 제공하는 차트에 히트맵이 포함되어 있지 않지만, 격자무늬 형태의 셀 구조를 활용하면 조건부 서식을 활용하여 손쉽게 생성할 수 있습니다. 워크시트 자체를 차트로 표현할 수 있기 때문에 차트 생성을 위해 데이터를 따로 준비할 필요도 없으며 데이터를 가공하면서 수시로 히트맵 차트를 생성하며 심도 있는 분석을 이어 나갈 수 있습니다.

실습 예제는 ‘서울교통공사’에서 제공하는 ‘지하철 혼잡도’입니다. ‘혼잡도’란 정원대비 승차인원
으로 승차인원과 좌석수가 일치할 경우 혼잡도를 34%로 산정합니다. 즉, 좌석수보다 많은 사람이
승차하면 34%를 초과한 수치가 기록되죠. 실습 예제는 평일 2호선 내선 순환열차의 혼잡도로 오
전 5:30부터 다음날 0시 30분까지 30분간격으로 평균 혼잡도와 역번호, 역명, 상하선구분 등이
포함되어 있습니다.

① 혼잡도 데이터가 있는 [G2:AS51] 영역을 선택합니다.

TIP [G2] 셀을 선택한 상태에서 [Ctrl]+[Shift]+[→], [Ctrl]+[Shift]+[↓]을 누르면 데이터 영역 한번에 선택할 수 있습니다. 데이터 영역을 간편하게
선택하는 자세한 내용은 49쪽을 참고하세요.

② 메뉴에서 [홈]-[스타일] 그룹-[조건부 서식]-[색조]-[빨강-흰색-파랑 색조]를 차례대로 선택
합니다.



③ 데이터가 큰 셀은 ‘빨간색’, 작은 셀은 ‘파란색’으로 채워집니다.

[illegible]

환



전문가의 시선

히트맵 차트로 패턴을 분석하는 방법

히트맵 차트는 행과 열로 구성된 시트에서 각 셀의 데이터를 색상으로 표현합니다. 막대 차트나 선 차트와 달리, 가로/세로축에서 수치가 아닌 범주형 값을 표시하며, 데이터의 크기는 오직 색상으로만 표현하는 특성으로 인해 범주형 데이터의 분포와 패턴을 파악하는 데 매우 효과적입니다. 이번 실습에서도 별다른 가공 없이 데이터가 있는 영역에 조건부 서식만 적용했음에도 불구하고 혼잡도가 높은 구간과 시간대를 한눈에 파악할 수 있습니다.

[illegible]

히트맵 차트 왼쪽의 빨간색이 모여 있는 구간은 오전 7시 30분에서 9시로 출근 시간에 해당합니다. 대부분의 역이 90에 달하는 높은 혼잡도를 기록하고 있으나, 강남부터 서울대입구에 이르는 역은 오히려 한가한 것을 볼 수 있습니다. 이 부근을 통과하여 위치한 강남, 역삼, 선릉, 서초, 교대 등은 서울의 대표적인 업무 중심 지역으로, 출근 시간대에는 이 지역으로 유입되는 인구가 많습니

다. 하지만 이 시간대에 해당 지역에서 다른 지역으로 이동하는 인구는 상대적으로 적기 때문에 지하철 혼잡도가 낮게 나타나는 것으로 해석할 수 있습니다.

히트맵 차트의 오른쪽은 퇴근 시간이 지난 오후 9시부터 10시 30분에 빨간색이 다시 모이는 것을 볼 수 있는데, 이는 이 지역이 업무 지구인 동시에 유흥 및 엔터테인먼트 문화의 중심지이기 때문입니다. 강남역 주변에는 각종 맛집, 술집, 클럽 등이 밀집해 있어 퇴근 후에도 많은 사람들이 모이는 곳입니다. 또한 코엑스, 멀티플렉스 영화관, 대형 서점 등 다양한 문화 시설도 있어 퇴근 후 여가 시간을 보내는 인구가 많은 것으로 보입니다.

이처럼 히트맵 차트는 지역과 시간대라는 범주형 데이터를 기반으로 실제 값인 혼잡도를 색상으로 표현함으로써, 데이터에 내재된 패턴을 시각적으로 드러내 줍니다. 이를 통해 우리는 복잡한 데이터 속에서 의미 있는 특징과 경향성을 직관적으로 파악할 수 있게 됩니다.

실습 히트맵 차트로 혼잡도 시각화하기 ②

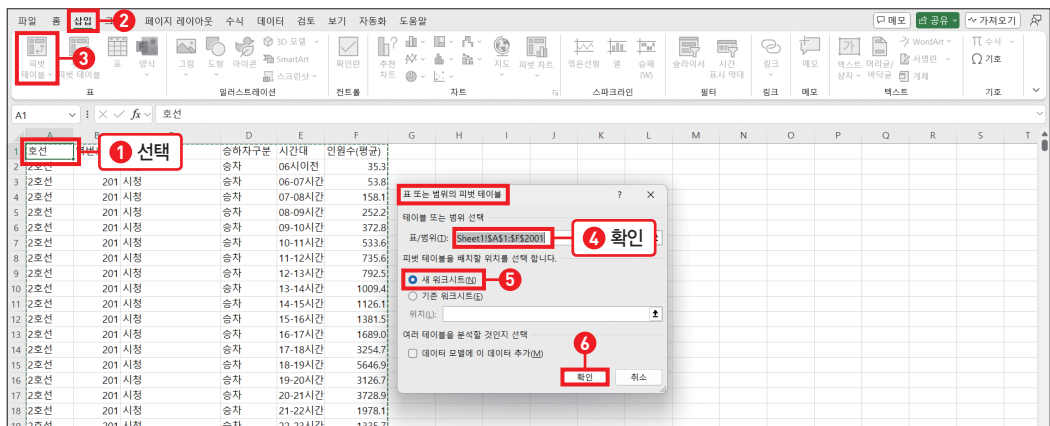
📎 CASE_03_02

1 | 데이터를 탐색하며 히트맵 활용하기 ①

이번에는 시간대별 승하차인원 정보를 활용하여 실습을 진행해 보겠습니다. 실습 예제는 2호선만 추출한 후 시간대별 인원 수를 [열 피벗 해제]하여 가공하였습니다.

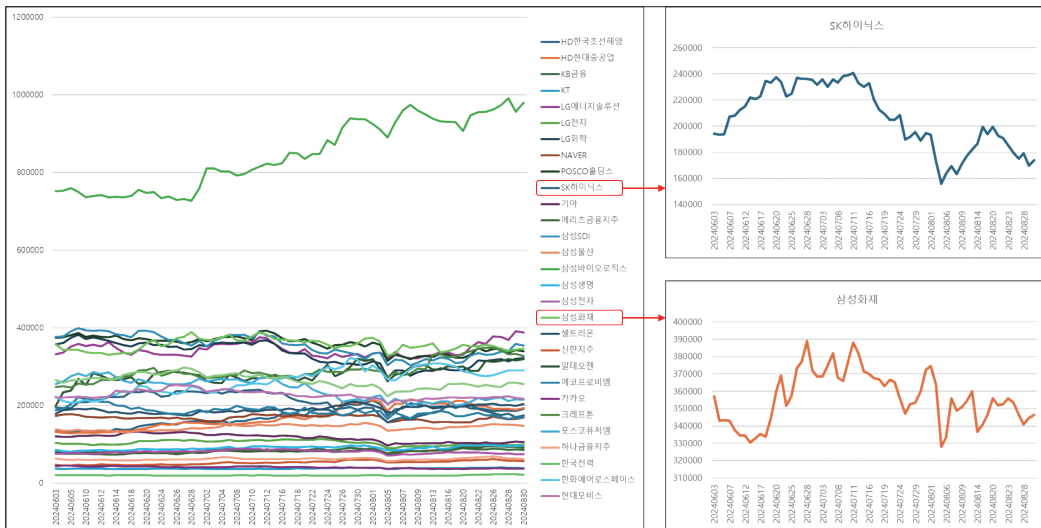
TIP 열 피벗 해제에 대한 자세한 내용은 ###페이지를 참고하세요.

① 데이터가 있는 [A1:F2001] 영역에서 임의의 셀을 선택하고, 메뉴에서 [삽입]-[피벗 테이블]을 선택하여 새 워크시트에 피벗 테이블을 생성합니다.



날짜나 시간의 흐름에 따라 순차적으로 기록된 데이터를 '시계열 데이터'라고 합니다. '시계열(時系列)'이라는 단어 자체가 '시간의 흐름에 따라 연관된 데이터를 일정한 순서로 나열한 것'이라는 의미를 담고 있죠. 일반적으로 데이터는 특정 이벤트가 발생한 시점에 기록됩니다. 예를 들어, 고객이 회원 가입을 하거나 제품을 구매했을 때 해당 정보가 저장됩니다. 그러나, 시계열 데이터는 정해진 시간 간격마다 일정한 규칙에 따라 데이터를 기록합니다.

시계열 데이터는 시간의 흐름에 따라 연속적으로 데이터를 기록하므로 데이터의 변화 양상과 추세를 파악하는 데 유용하게 활용됩니다. 주식 시세 데이터는 대표적인 시계열 데이터로, 정해진 시간마다 거래 내역을 기록하여 가격과 거래량의 변동 추이를 파악할 수 있습니다. 시계열 데이터는 주로 선형 차트로 추세를 확인하지만, 주식 시세 데이터의 경우 종목별 시세 범위가 달라 한 차트에 모두 표시하면 추세를 파악하기 어렵습니다. 또한, 각 종목마다 차트를 따로 그리려면 같은 작업을 여러 번 반복해야 하는 번거로움이 있습니다.

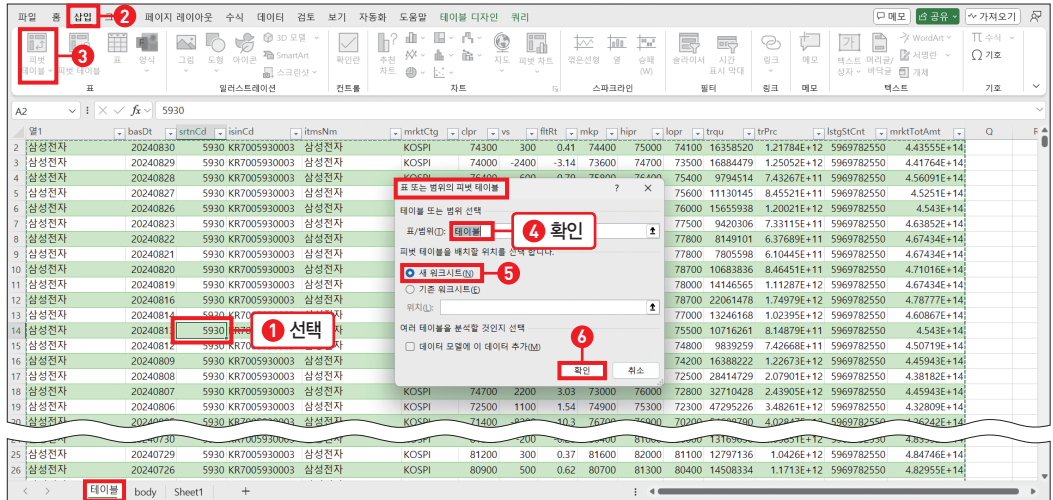


‘스파크라인’은 셀 안에 그려지는 작은 차트로 ‘선’, ‘열’, ‘승패’의 세 종류가 있으며, 이중 ‘선’ 스파크라인을 이용하면 여러 셀에 걸친 주식의 시세 데이터를 하나의 셀 안에 간단한 선으로 표현할 수 있습니다. 또한, 행 또는 열 단위로 차트를 생성할 수 있어 종목이 많고 시세가 다양한 주식 추세를 분석할 때 유용하게 활용할 수 있습니다.

1 | 피벗테이블로 종목별 시세 데이터 추출하기

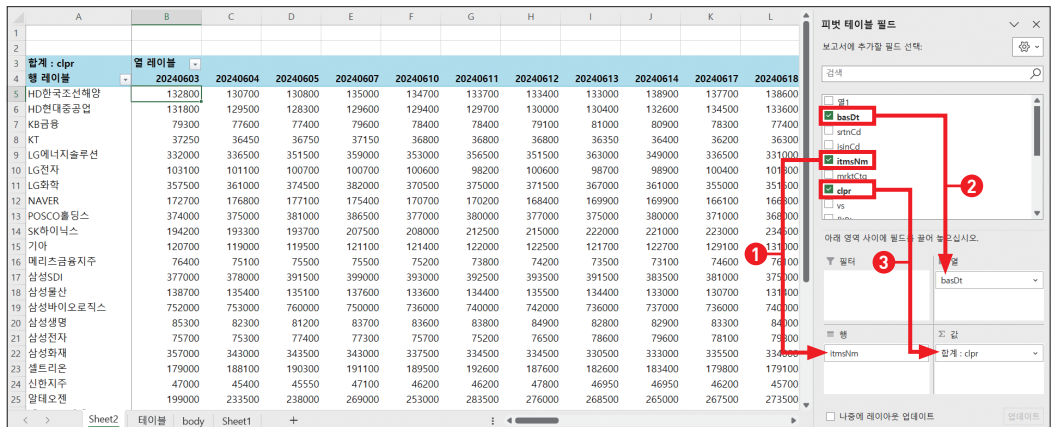
① [테이블] 시트의 데이터 영역 중 임의의 셀을 선택하고 메뉴에서 [삽입]-[피벗 테이블]을 선택합니다.

② [표 또는 범위의 피벗 테이블] 창에서 '표/범위'에 '테이블'이 입력되어 있는 것을 확인하고 [새 워크시트]에 피벗 테이블을 생성합니다.



TIP 피벗 테이블은 표를 지정해서 데이터를 가져올 수 있습니다. 이처럼 표가 선택된 상태에서 피벗 테이블을 추가하면 표 이름이 자동으로 입력됩니다.

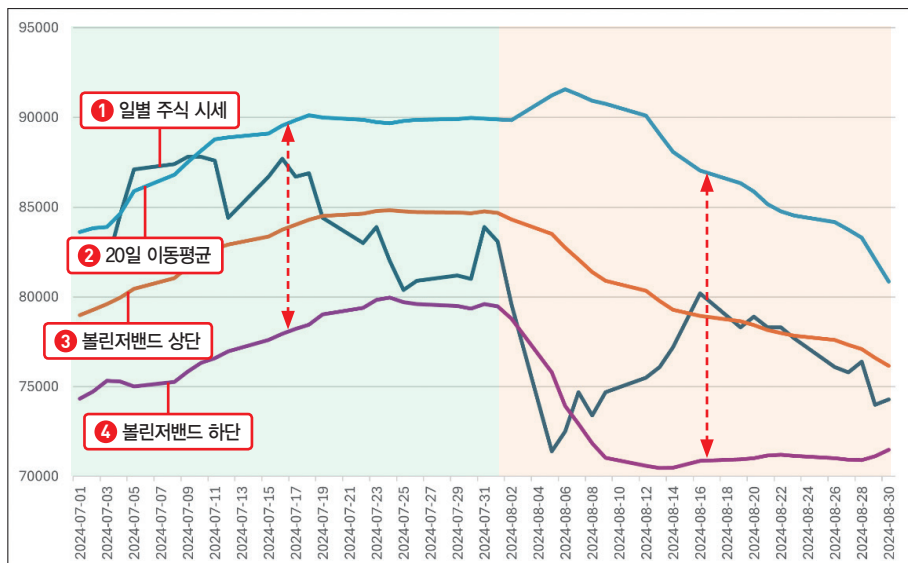
③ '피벗 테이블 필드'에서 [itemsNm](종목명)을 [행], [basDt](기준일자)를 [열], [clpr](종가)을 [값]으로 드래그합니다.



TIP [값]으로 드래그한 [clpr](종가)는 피벗 테이블의 행과 열에 하나씩만 집계되므로 요약 기준을 변경할 필요가 없습니다.



실습으로 완성한 선형 차트는 ① 일별 주식 시세, ② 20일 이동평균, ③ 볼린저밴드 상단, ④ 볼린저밴드 하단 데이터로 구성되어 있습니다.



볼린저 밴드 차트를 보면, 왼쪽과 오른쪽 구간에서 밴드의 너비에 큰 차이가 있음을 알 수 있습니다. 왼쪽 구간에서는 볼린저 밴드의 상단과 하단 사이의 간격이 좁은 반면, 오른쪽 구간에서는 상단과 하단 사이의 간격이 훨씬 넓습니다.

이러한 차이는 볼린저 밴드가 표준편차를 사용하여 계산되기 때문에 발생합니다. 일별 시세의 변동폭이 클 경우, 표준편차도 커지게 되고, 이에 따라 볼린저 밴드의 상단과 하단 사이의 간격도 넓어집니다. 이렇게 볼린저 밴드를 활용하면 시세 데이터의 변동성을 한눈에 파악할 수 있습니다. **밴드의 너비가 좁을 때는 시장이 안정적이고 변동성이 낮은 상태이며, 밴드의 너비가 넓을 때는 시장의 변동성이 높은 상태임을 쉽게 알 수 있습니다.** 이러한 정보는 투자자들이 시장의 동향을 파악하고 적절한 투자 전략을 수립하는 데 도움이 됩니다.

3 | 볼린저 밴드를 활용한 주식 매매방법

주식의 시세는 시장에 참여하는 개인, 기관, 외국인 등의 매매에 의해 결정됩니다. 정치, 경제, 또는 다른 금융 상품의 영향을 받을 수 있지만, 궁극적으로는 매매에 참여하는 주체들의 행동에 따

라 형성됩니다.

볼린저 밴드를 활용한 매매 전략은 시장 참여자들의 심리를 이용한 기법입니다. 이 전략은 주가가 일별 변동성을 나타내는 볼린저 밴드의 상한선과 하한선을 넘어설 때, 과도한 매수 또는 매도 심리가 발생한 것으로 해석하고 매매를 실행합니다. 구체적으로, 주가가 볼린저 밴드의 하한선 아래로 내려가면 '과매도' 상태로 판단하여 '매수'하고, 반대로 주가가 상한선을 넘어서면 '과매수' 상태로 보고 '매도'합니다.

이제 앞서 계산한 볼린저 밴드를 활용하여 주식을 매매하는 방법을 알아보겠습니다. 단, 차트에만 의존하여 거래하는 방식이므로 매도 시에는 기술적 매매에 주로 사용되는 '트레일링 스탱 (Trailing Stop)' 방식을 적용해 보겠습니다. 트레일링 스탱은 매수가 또는 고점 대비 일정 비율로 가격이 하락하면 매도하는 거래 방법으로 실습에서는 '2%' 하락 시 매도합니다.

① [H20] 셀에 머리글을 입력합니다. 여기에서는 '매수'를 입력했습니다.

② [H21] 셀에 다음의 수식을 입력하고 **[Enter]**를 누릅니다.

[H21] f_x **=C21<G21**
이 수식은 [C21] 셀에 있는 '가격'이 [G21] 셀의 볼린저 밴드 하단보다 작은지 확인하는 조건문입니다. 함수를 사용하지 않아도 사칙연산처럼 동작합니다.

③ [H21] 셀의 수식을 복사하여 [H21:H64] 영역에 붙여 넣으면, 셀의 값이 볼린저 밴드 하단보다 클 경우 'TRUE'(참), 작을 경우 'FALSE'(거짓)를 반환합니다.

