📊 데이터 정규화 실생활 예시 (채용 데이터 기준)

1. 연봉 **vs.** 기술 스택 수 비교

- 문제:
  - 연봉 범위: `3,000만원 ~ 1억원` (단위: 1,000만)
  - 기술 스택 수: `1개 ~ 20개` (단위: 1)
    → 연봉이 분석에 지배적 영향을 미침!
- 해결: **Min-Max** 정규화 적용

원본 데이터	정규화 후 **(0~1)**
연봉 5,000만원	`(5000-3000)/(10000-3000) = 0.29`
기술 10개	`(10-1)/(20-1) = 0.47`

→ 이제 기술 스택 영향력이 공정하게 반영됨!

---

2. 회사 규모**(**직원 수**)** 비교

- 원본 데이터:
  - 스타트업: `10명`
  - 대기업: `10,000명`
    → 차이가 **1,000**배라 분석 왜곡 가능성 ↑
- 해결: **Log** 변환

회사 규모	**Log10** 적용 후
10명	`log10(10) = 1`
10,000명	`log10(10000) = 4`

→ 차이를 **4**배로 축소해 더 합리적인 비교 가능!

---

3. 지역별 평균 연봉 표준화

- 문제:
  - 서울 평균: `6,000만원` (표준편차 1,000만)
  - 부산 평균: `4,500만원` (표준편차 500만)
    → 단순 비교 시 표준편차 차이 무시됨
- 해결: **Z-Score** 정규화

지역	원본 연봉	**Z-Score ((값-평균)/표준편차)**
서울	7,000 만 원	`(7000-6000)/1000 = +1.0`
부산	5,000 만 원	`(5000-4500)/500 = +1.0`

→ 둘 다 평균보다 **1**표준편차 높음으로 동등한
"상대적 위치" 파악 가능!

.