



2. 데이터 분석 개요

데이터 분석?

- 데이터를 정제 및 가공 해서 의미 있는 정보를 추출
- 추출된 정보를 중요한 의사 결정에 활용
 - ⊙ 본 강의에서는 데이터 분석에 대해 전반적인 이해와
 - ⊙ 실습을 통해 다양한 데이터 정제, 가공, 가시화 기술들을 배우며
 - ⊙ 새로운 가치를 찾아내는 경험을 축적



데이터 분석의 목적

- 데이터를 바탕으로 합리적인 사고 과정을 통해 올바른 의사결정을 지원
 - ⊙ 과거의 경험과 직감에만 의존하지 않고 객관적인 데이터 기반한 합리적인 결정
- 예시
 - ⊙ 신제품 개발을 위해
 - ⊙ 새로운 광고를 제작할 때
 - ⊙ 신규 매장 위치를 결정할 때
 - ⊙ 다이어트를 할 때 (섭취량, 운동량, 수면시간 등)



데이터-정보-지식-지혜



개념

- 개별 데이터 자체로는 특별한 의미부여가 안된 객관적 사실
- 타 데이터와 상관관계 없는, 가공하기 전 순수한 수치나 기호 자체

예시

- 월평균 로션판매량 조사
 - A대리점: 100개
 - B대리점: 200개

의미 부여가 되지 않은
있는 그대로의 사실

데이터-정보-지식-지혜



개념

- 데이터를 통해 도출된 다양한 정보를 구조화해 유의미한 정보로 분류함
- 데이터의 가공·처리, 데이터간 상관·연관관계 속에서 의미를 부여함
- 유용성은 상황에 따라 다름

예시

- B대리점의 판매실적이 A대리점보다 2배임
- 매월 총 300개 판매실적을 보이고 있음

판단

데이터-정보-지식-지혜



개념

- 정보를 바탕으로 의사결정에 활용하는 것
- 상호 연결된 정보패턴을 이용해 예측한 결과물
- 개인의 경험 결합해 고유의 지식으로 내재화

액션을 일으키는 근거

예시

- 각 대리점별 판매실적을 기준으로
 - 공급 부서장: 로션을 공급량, 공급량 및 주기 차별화
 - 성과 평가 부서장: 인센티브도 차등 지급함
 - 생산부서: 원료 수입, 제조라인 투입 인력 운영, 스케줄링, 재고량 등에 대한 의사결정

데이터-정보-지식-지혜



개념

- 로션이외에 다른 상품 판매실적도 비교해 봄
- 계절과 지역에 따라, 영업직원 판매방식에 따라 실적차이가 나는지 점검해 전략적 매장운영 관리체계를 수립함

예시

- 지식의 축적과 아이디어가 결합된 창의적인 산물
- 근본 원리에 대한 깊은 이해를 바탕으로 도출되는 창의적인 아이디어

데이터 분석 실무 예시

○ 제조

- ⊙ 제조의 공정을 개선하거나 문제점을 보완해 생산성 향상 및 품질 개선
- ⊙ 센서 데이터를 이용해 공장 전체 혹은 라인별 진행 상황을 실시간 분석하여
 - ⊙ 고장이 잦은 설비 파악, 원인 분석
 - ⊙ 문제점 조기 발견 및 예측

○ 유통

- ⊙ 수많은 주문 내역, 재고량 등의 정보를 기반으로 최적의 배송
- ⊙ 유동적으로 변하는 고객의 주문량에 대해 신속 대응
- ⊙ 배송 준비 및 진행 과정을 고객에게 공유

데이터 분석 실무 예시

○ 마케팅

- ⊙ 고객 구매 정보, 만족도 같은 기업 내부 데이터로 고객 취향 파악, 관리 및 잠재 고객을 유입
- ⊙ 포털 사이트 연관 검색어 같은 흔적 데이터를 사용하여 마켓 트렌드 및 경쟁사와의 우위 정보 사용
- ⊙ 이를 통해 마케팅 전략 수립

○ 공공

- ⊙ 정부, 행정 지자체 별로 공공 데이터를 축적하고 행정 업무의 효율성을 높임
- ⊙ 교통 카드 내역을 기반으로 승하차 데이터 취득, 유동인구의 지역별 밀도를 분석
- ⊙ 합리적인 버스 노선도 설계 및 심야 버스 노선 개발

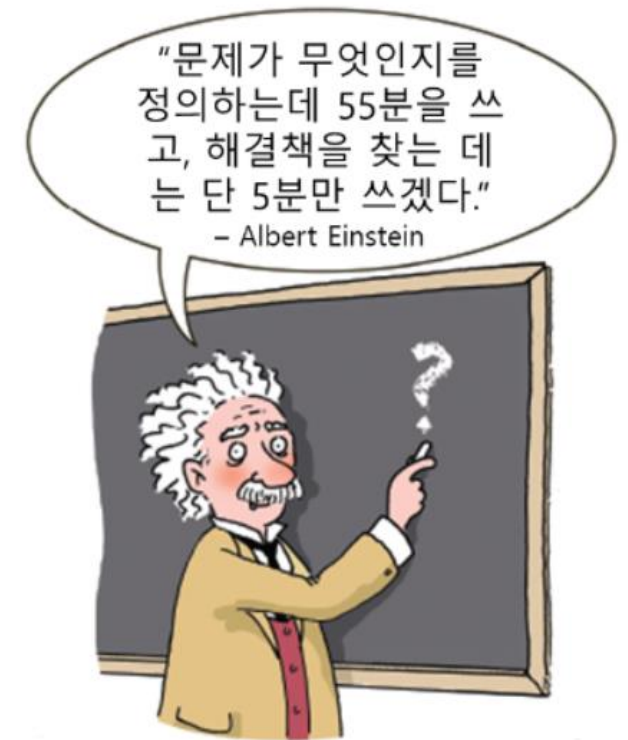
데이터 분석 절차

1. 문제 정의
2. 데이터 수집
3. 데이터 전처리
4. 데이터 분석
5. 보고 및 피드백 반영

1) 문제 정의

○ 가장 중요하기도 하고 가장 어려운 단계

- ⊙ 많은 사람들이 공감할 만한 가치가 있는 문제
- ⊙ 문제가 제대로 설정 되지 않으면 분석 목표가 불분명
- ⊙ 분석의 목적, 이유, 및 예상 산출물 등에 대해 협의
- ⊙ 분석을 위한 전문가와 분석 기간을 확보



아인슈타인은 이렇게 말했다.

2) 데이터 수집

- 본인/회사 에서 보유하고 있는 데이터 수집
- 공공 포털 또는 크롤링 등을 이용하여 데이터 수집

공공

공공데이터포털 <https://www.data.go.kr/>
서울시 열린데이터 광장 <https://data.seoul.go.kr/>

행정

주민등록 인구통계 <https://jumin.mois.go.kr/>
지방행정 데이터 <https://www.localdata.go.kr/>

지도

국가 공간정보 포털 <http://www.nsdi.go.kr/>

건축

건축데이터 민간개방 시스템 <https://open.eais.go.kr/>
국가공간정보포털 <http://www.nsdi.go.kr/>
등기정보광장 <https://data.iros.go.kr/>

치안

경찰청 공공 데이터 개방 <https://www.police.go.kr/www/open/public/public01.jsp>

관광

한국관광공사 TourAPI <https://api.visitkorea.or.kr/>

기상

기상 자료 개방 포털 <https://data.kma.go.kr/>

농림

농림축산부 <https://www.mafra.go.kr/mafra/322/subview.do>

금융

금융 빅데이터 개방시스템 <https://credb.kcredit.or.kr/>
금융데이터 거래소 <https://www.findatamall.or.kr/>

문화

문화 데이터 광장 <https://www.culture.go.kr/data/>

복지

보건복지 데이터 포털 <https://data.kihasa.re.kr/>

교통

국가 교통 DB <https://www.ktdb.go.kr/>
교통사고 분석 시스템 <http://taas.koroad.or.kr/>

전기

전력데이터 개방 포털시스템 <https://bigdata.kepco.co.kr/>

기타

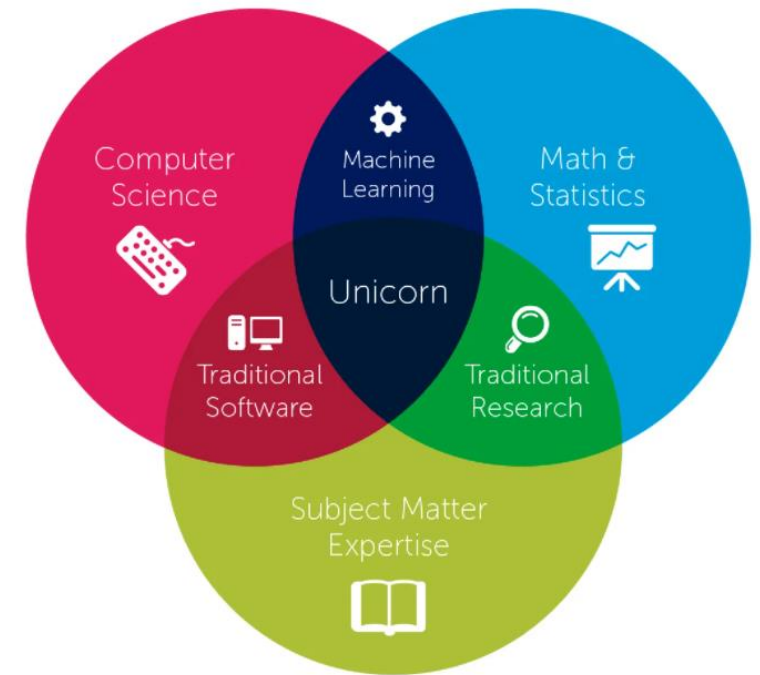
AI 데이터 허브 <https://aihub.or.kr/>
데이터 스토어 <https://www.datastore.or.kr/>

3) 데이터 전처리

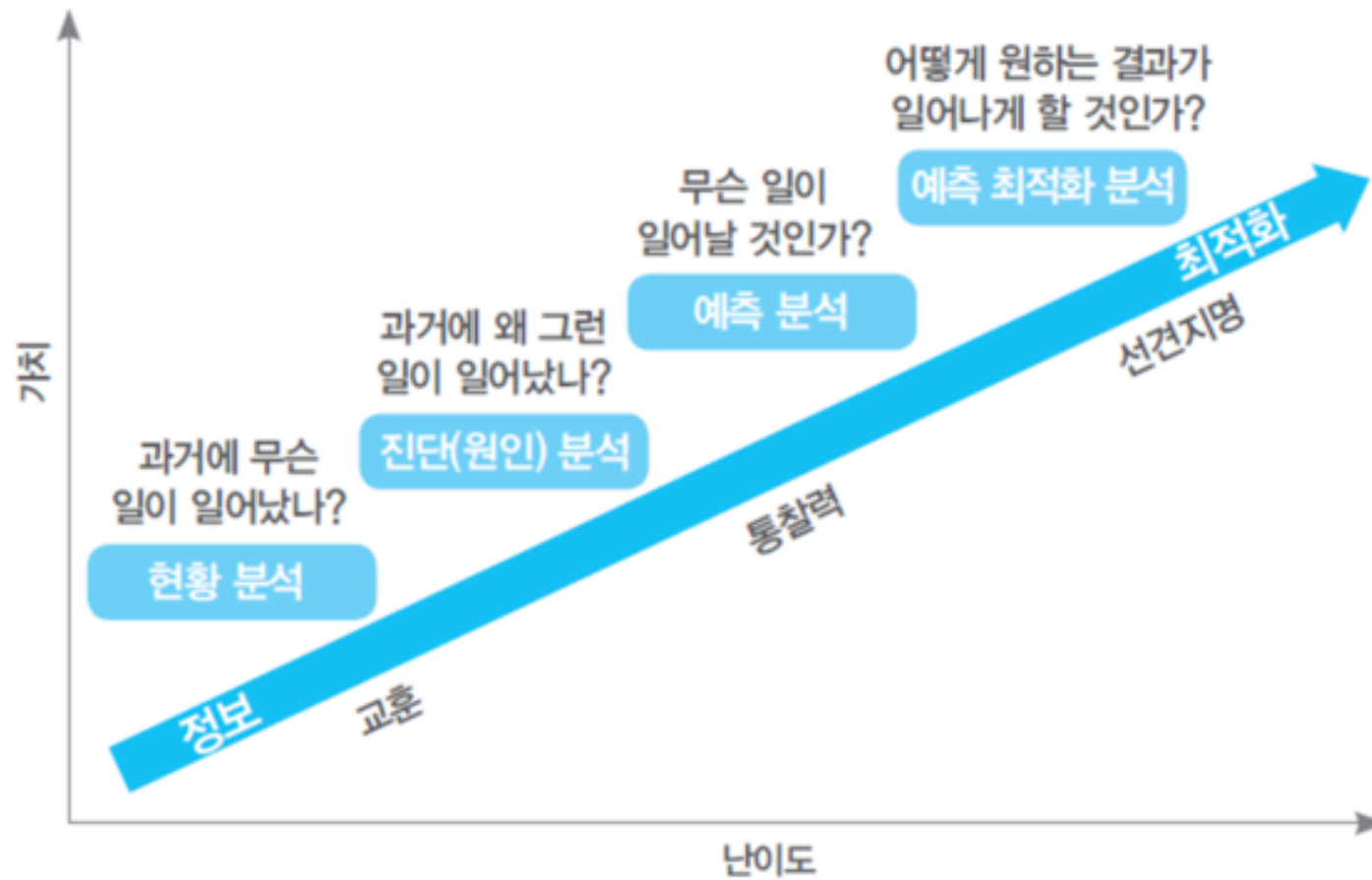
- 데이터 분석 단계 중 가장 많은 시간이 소요되는 단계
- 수집한 데이터가 분석에 바로 사용되기에는 어려운 경우 대부분
- 결측 값, 이상 값, 중복 값 등을 적절히 처리해주어야 함
- Feature Engineering
 - ⊙ scaling (예: 로그함수 적용)
 - ⊙ binning (예: 연속형 데이터를 범주형으로)
 - ⊙ transforming (예: 개별 일자 데이터를 주중 & 주말 데이터로)

4) 데이터 분석

- 매우 큰 영역을 아우르며, 도메인과 여러 상황에 따라 다양한 분석 유형이 존재
- 지표 정의 및 추적
 - ⊙ 비즈니스와 관련한 주요 지표를 개발하고 지속적으로 추적
- 탐색적 데이터 분석
 - ⊙ 다양한 각도에서 관찰하고 이해하는 과정
 - ⊙ 데이터 분포확인, 시각화, 패턴 발견
- 통계적 분석
 - ⊙ 변수간 상관 관계
 - ⊙ 가설 검정, 선형 모형
- 기계학습
 - ⊙ 분류 및 회귀 (지도학습)
 - ⊙ 클러스터링 (비지도학습)



4) 데이터 분석



자료: 가트너

5) 보고 및 피드백 반영

- 분석 결과 및 인사이트를 설득력 있게 전달
- 내용의 초점은 분석가 아닌 상대방
 - ⊙ 상대의 입장과 니즈에 맞춰 정리
 - ⊙ 분석 과정 보다는 '결과' 중심으로 전달
- 간결하고 명확한 메시지로 전달
 - ⊙ 짧은 문장 (불필요한 수식어 제거)
 - ⊙ 핵심 내용 먼저 전달 후 근거를 나열
- 적절한 시각화 방법 활용
 - ⊙ 차트, 그래프와 같은 시각적인 요소 활용
 - ⊙ 분석 결과의 전달력을 극대화
- 투명하고 공정한 반영
 - ⊙ 사실 기반 공유 & 보고
 - ⊙ 반복적이고 지속적인 개선

데이터 분석의 중요성 요약

○ 합리적 의사결정

- ⊙ 데이터 기반 흐름을 이해하고 미래를 예측하는 합리적 의사결정

○ 설득과 협상의 도구

- ⊙ 설득과 협상 과정에서 탄탄한 근거와 뒷받침할 데이터

○ 생산성, 효율성, 수익성 성장

- ⊙ 오류를 일으키는 원인 확인, 낭비되는 자원 식별, 시간과 비용을 절감



