
데이터산업 현황 및 데이터 비즈니스 모델

임종호

연세대학교 통계데이터사이언스학과

2022 한국경제학회/한국경영학회 공동학술대회

2022.01.13

개요

1. 데이터산업
2. Think Data
3. 데이터 비즈니스 모델
4. 결론

1. 데이터산업 - 요약

(데이터산업) 데이터의 생산, 수집, 처리, 분석, 유통, 활용 등을 통해 가치를 창출하는 상품과 서비스를 생산·제공하는 산업

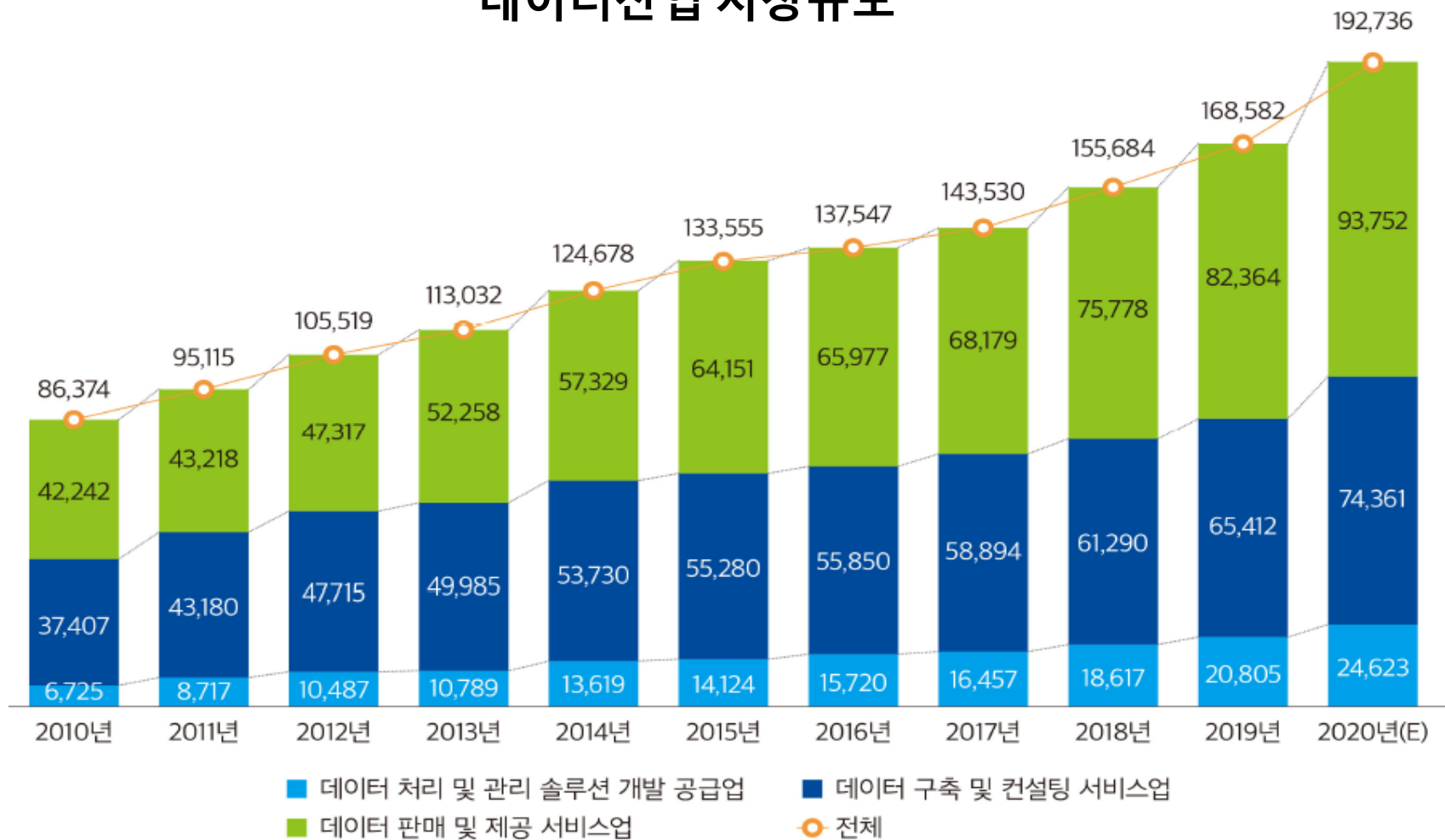
데이터산업 분류 및 내용

구분	내용
데이터 처리 및 관리 솔루션 개발·공급업	• DBMS, DBMS관리 • 데이터 모델링, 분석·시각화, 검색엔진 관련 솔루션
데이터 구축 및 컨설팅 서비스업	• DB설계, 데이터 이행 등을 포함한 DB시스템 구축 • 비정형데이터 DB로 변환, 정비 • 데이터 관련 컨설팅 비즈니스
데이터 판매 및 제공 서비스업	• 데이터 혹은 DB 판매 및 중개 • 데이터를 가공·활용·분석하여 제공
데이터 인프라 서비스업	• 서버, 스토리지, 네트워크 데이터 인프라 서비스 제공

※ 2020 데이터산업 현황조사

1. 데이터산업 - 시장현황

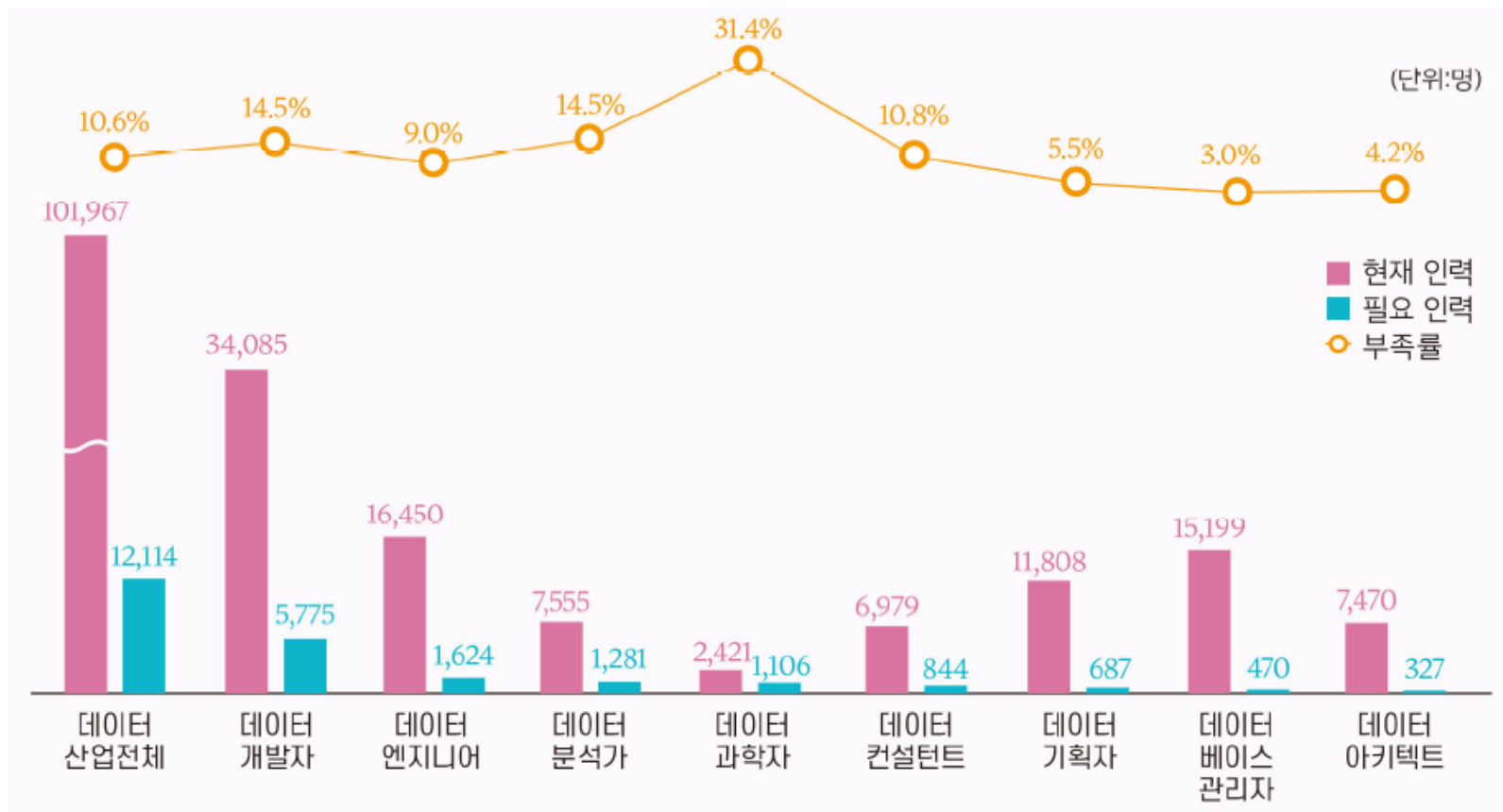
데이터산업 시장규모



※ 2020 데이터산업 현황조사

1. 데이터산업 - 인력 현황

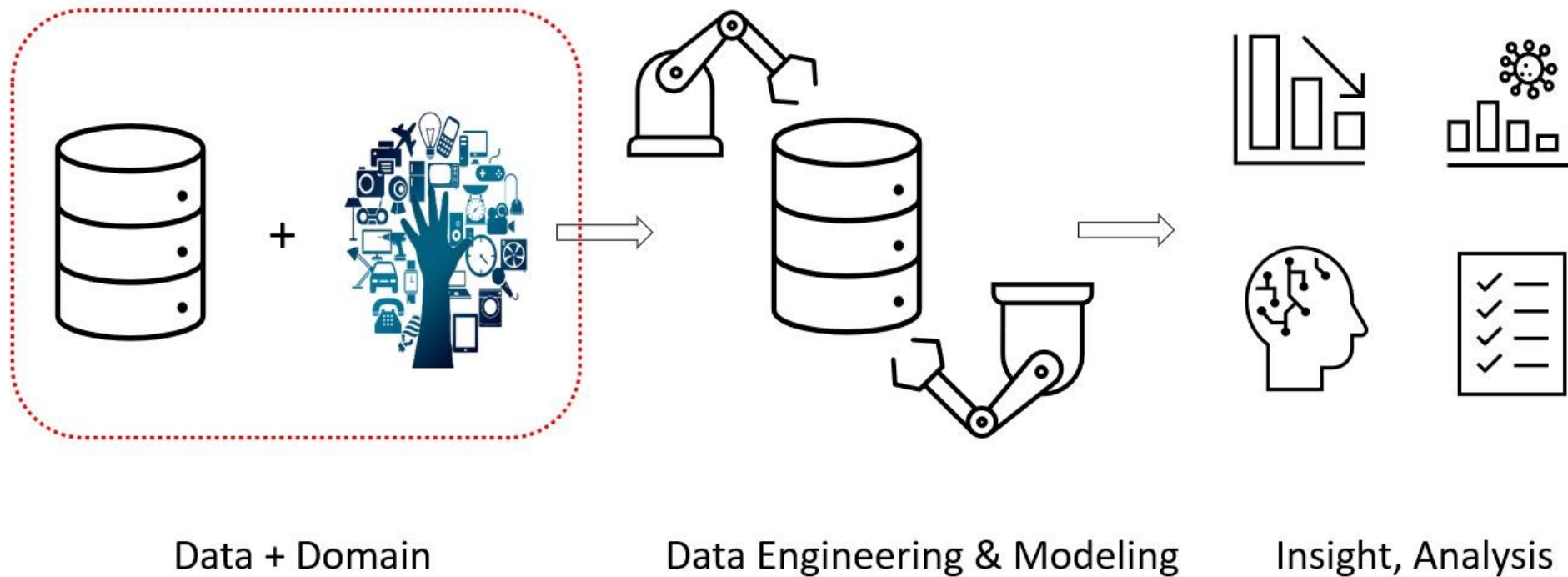
데이터산업 직무별 인력 현황 및 수요



※ 2020 데이터산업 현황조사

1. 데이터산업 - 연관 기술 흐름도

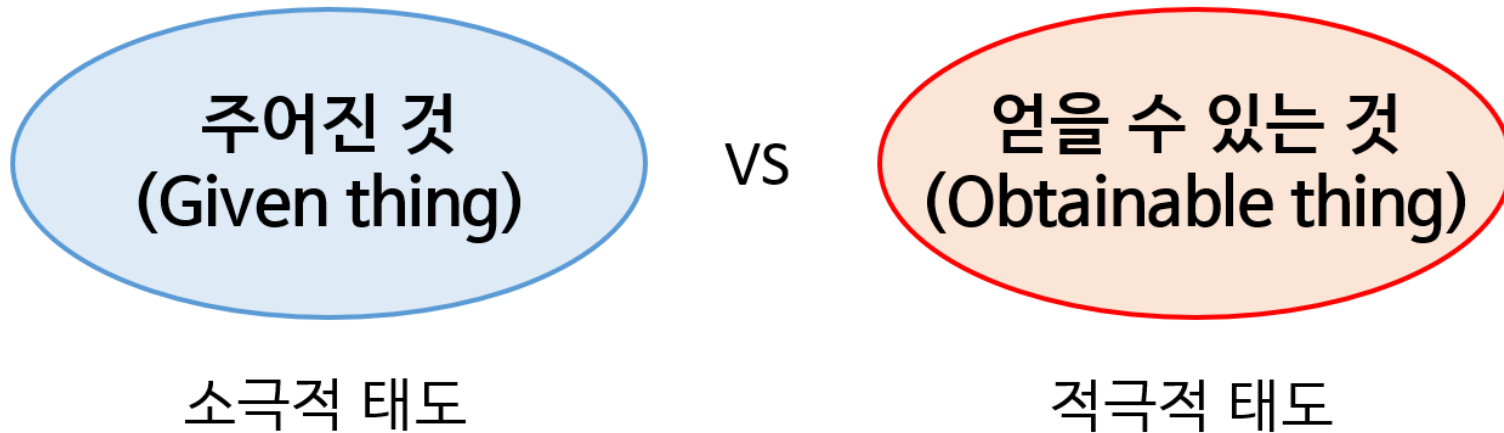
데이터 수집·처리·분석 흐름도



“데이터”에 더 주목하고 시각을 넓힐 필요가 있음

2. Think Data – 데이터란?

(데이터 어원) 주어진 것

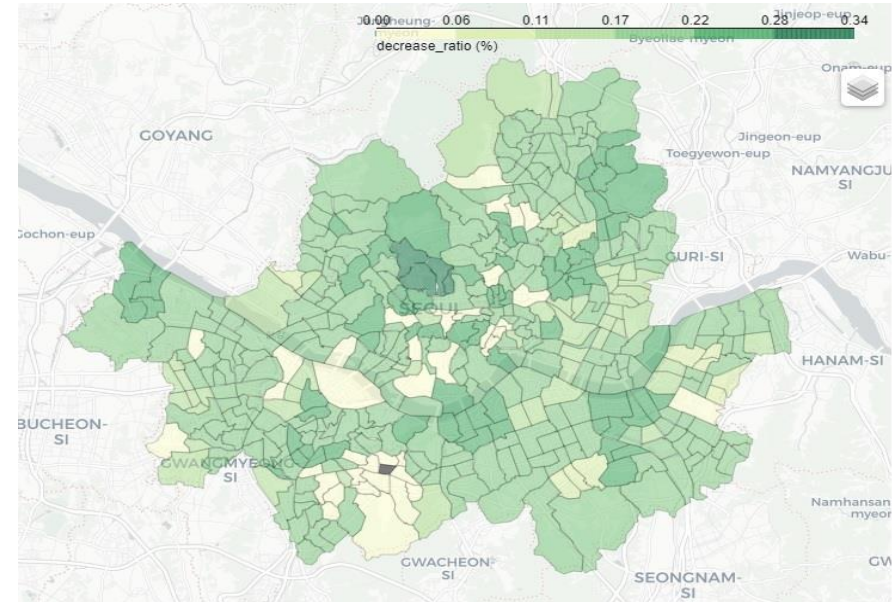
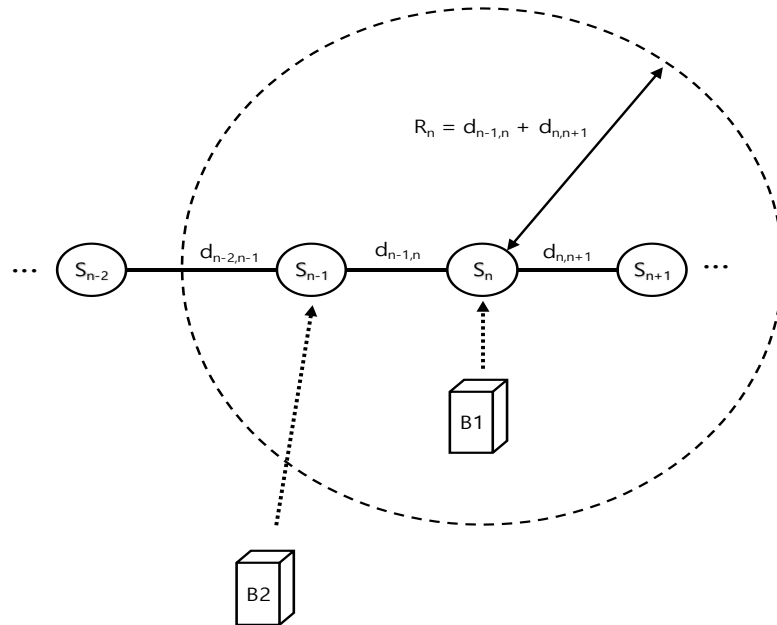


- (A사) 생산관리를 위하여 노트북 수요 및 생산 예측을 하는 경우를 가정
- 경쟁사 혹은 시장에 대한 데이터는 없거나 부족한 경우가 대부분
- 온라인 정보 수집, 시장 조사 등 다양한 형태로 연관 데이터를 수집해야 함

2. Think Data – 데이터의 방향성

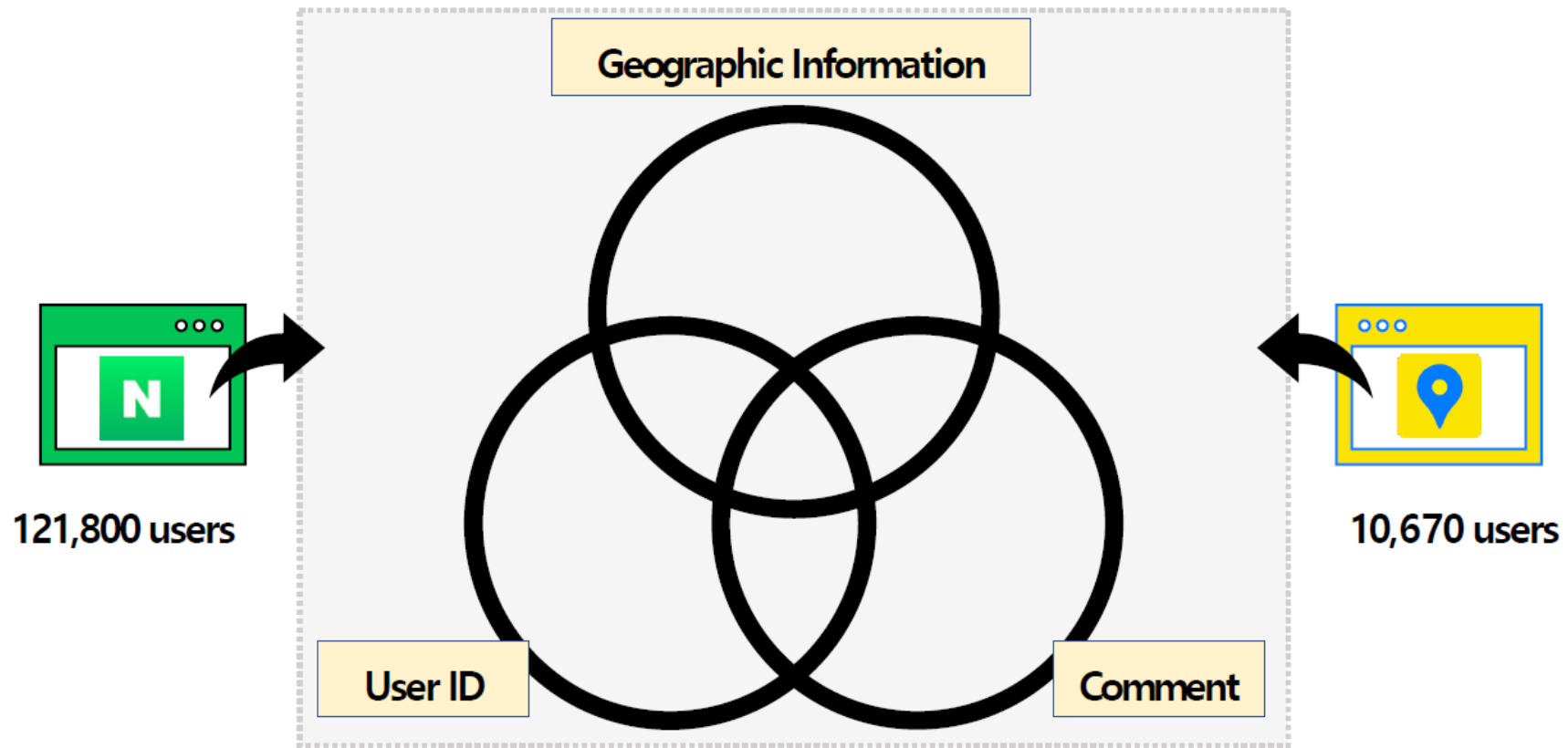
Bigger Data 와 데이터 결합

“사회적 거리 두기” 효과? 방역 전략? (2020년 서울시 2월 & 3월 버스이용자)



- 대체적으로 버스 이용객 수는 감소하였음
- 거의 감소하지 않았거나 증가한 곳이 있음 → **물류창고 시설 부근**

2. Think Data - 정보보호



- 개인정보보호법 (2020.08) – 가명정보 제도화
- 식별 및 결합 가능성을 개인정보로 규정
- 약 5% 정도의 User를 식별하여 개인 정보를 만들어 낼 수 있음

2. Think Data – 데이터 품질

Relevance: 우리가 원하는 정보가 있는가? 우리가 원하는 지식을 발견할 수 있는가?

물 품	가 격
	30,000
	55,000

- (옵션1) 물건의 판매와 관련된 내용 (물품, 가격, 수량) 기록
- (옵션2) 판매된 물건과 같은 윈도우(진열장)에 노출된 상품들의 정보도 같이 기록
- 고객들의 Primary Demand와 Primary Demand에 대한 행태를 분석하는 것이 목표라면 (옵션1)의 데이터는 같은 판매기록 데이터라도 도움이 되지 않을 수 있음

2. Think Data – 데이터 품질

Selection Bias: 관측 데이터의 경우 대부분 선택 편의에 노출되어 있음

레스토랑 A



레스토랑 평점: 4.5

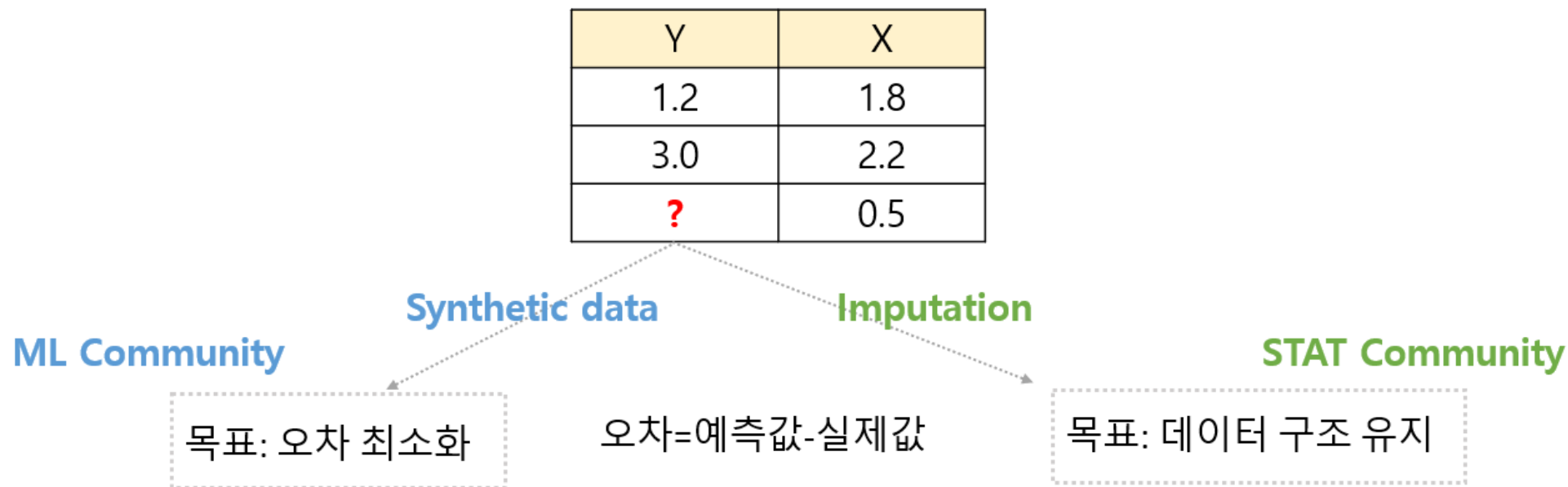
레스토랑 B



레스토랑 평점: 4.0

- 낯선 여행지에서 레스토랑을 방문하는 상황을 가정
- 동일한 메뉴를 판매하는 레스토랑 A, B가 있음
- 두 레스토랑 가격대는 비슷하나 평점은 상이함
- 어느 레스토랑을 가는 것이 현명할까?

2. Think Data – 불확실성에 대한 이해



- 불확실성을 대하는 방법에 따라서 두 가지 접근 방법이 있음
불확실성(오차)를 최소화하는 방법과 불확실성(오차)를 제어하는 방법임
- 일반적으로 ML community는 전자를 선호하고 Stat community는 후자를 선호함

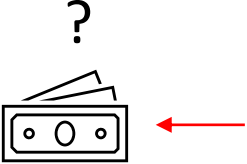
3. 데이터 비즈니스 모델



경제적 가치가 높은 데이터

- 비공개 혹은 휘발성이 있는 데이터
- 지속적으로 업데이트 되는 데이터

3. 데이터 비즈니스 모델 – New 모델



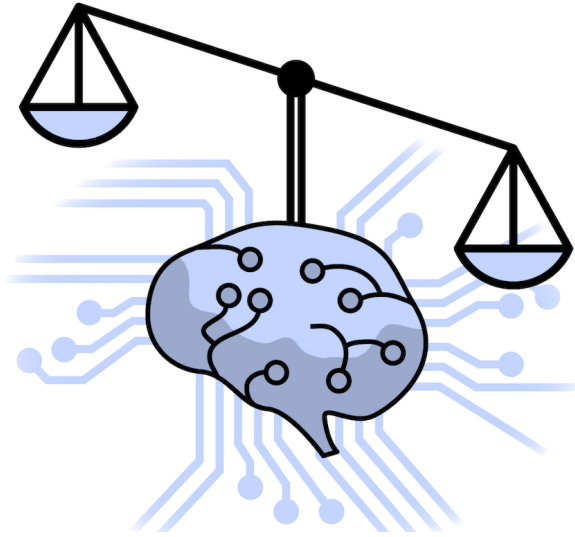
이름	생년월일	월 소득	금융부채
김연세	81.01.01	450만원	1천만원
김데이터	88.01.01	400만원	500만원
김통계	89.01.01	380만원	0원

- New 데이터 비즈니스 모델

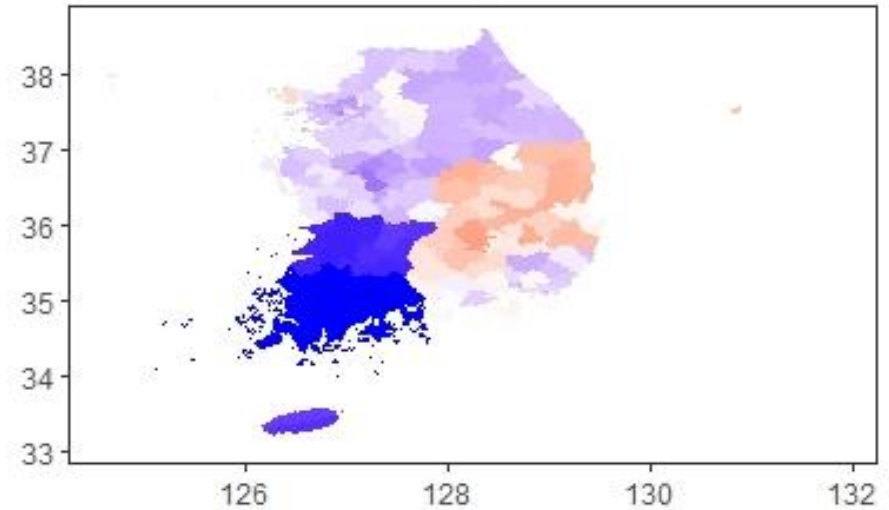
데이터 유닛 하나의 경제적 가치 (economic value)를 계산할 수 있는 모델

- Which data/unit is Informative ?

3. 데이터 비즈니스 모델 – Desirable 모델



※ Figures by *Nicholas Lue*



New 데이터 비즈니스 모델 + 데이터 윤리 + 이익공유

Fairness AI, 안심번호 기반 여론조사

4. 결론

- * 데이터산업은 빠르게 성장하고 있으나 Skilled 인력은 매우 부족한 상황
- * 데이터에 내재된 가치보다는 주어진 데이터에 알고리즘/모형을 적용하는 비즈니스 모형 중심
- * Bigger 데이터의 관점에서 데이터를 얻을 수 있는 것으로 이해하려는 태도와 전략이 필요
- * 데이터 윤리, 데이터 정보보호, 데이터 품질 이슈에 더 관심을 가져야 할 것으로 사료됨
- * 데이터의 경제적 가치를 다양한 형태로 측정하려는 노력이 필요