PART 2. 데이터 분석 기획

데이터 준전문가

ADSP, Advanced Data Analytics semi-Professional

류영표 강사
ryp1662@gmail.com

강사 소개

류영표

Youngpyo Ryu

동국대학교 수학과/응용수학 석사수료

前 Upstage AI X 네이버 부스트 캠프 AI tech 1~3기 멘토

前 Innovation on Quantum & CT(IQCT) 이사

前 한국파스퇴르연구소 Image Mining 인턴(Deep learning)

前 ㈜셈웨어(수학컨텐츠, 데이터 분석 개발 및 연구인턴)

강의 경력

- 현대자동차 연구원 강의 (인공지능/머신러닝/딥러닝/강화학습)
- ㈜ 모두의연구소 Aiffel 1기 퍼실리테이터(인공지능 교육)
- 인공지능 자연어처리(NLP) 기업데이터 분석 전문가 양성과정 멘토
- 공공데이터 청년 인턴 / SW공개개발자대회 멘토
- 고려대학교 선도대학 소속 30명 딥러닝 집중 강의
- 이젠 종로 아카데미(파이썬, ADSP 강사)
- 최적화된 도구(R/파이썬)을 활용한 애널리스트 양성과정(국비과정) 강사
- 한화, 하나금융사 교육
- 인공지능 신뢰성 확보를 위한 실무 전문가 자문위원
- 보건 · 바이오 AI활용 S/W개발 및 응용전문가 양성과정 강사
- Upstage AI X KT 융합기술원 기업교육 모델최적화 담당 조교

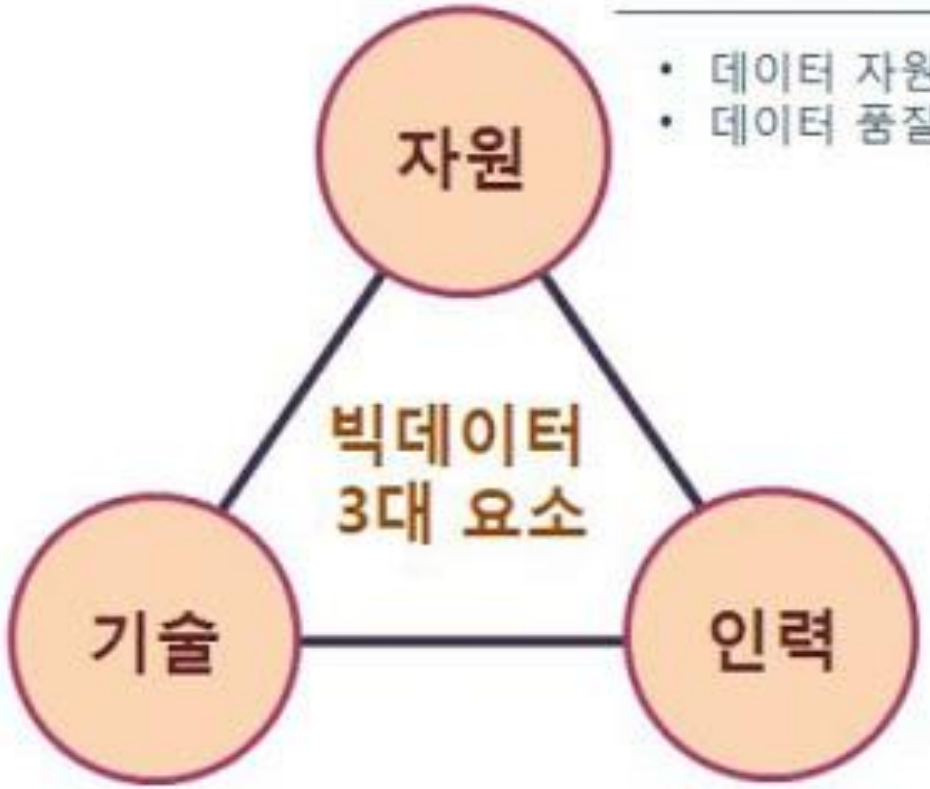
주요 프로젝트 및 기타사항

- 개인 맞춤형 당뇨병 예방·관리 인공지능 시스템 개발 및 고도화(안정화)
- 폐플라스틱 이미지 객체 검출 경진대회 3위
- 인공지능(AI)기반 데이터 사이언티스트 전문가 양성과정 1기 수료
- 제 1회 산업 수학 스터디 그룹 (질병에 영향을 미치는 유전자 정보 분석)
- 제 4,5회 산업 수학 스터디 그룹 (피부암, 유방암 분류)
- 빅데이터 여름학교 참석 (혼잡도를 최소화하는 새로운 노선 건설 위치의 최적화 문제)

분석 기획

- 실제 분석을 수행하기 앞서 분석을 수행할 과제의 정의 및 의도했던 결과를 도출할 수 있도록 이를 적절하게 관리 할 수 있는 방안을 사전에 계획하는 일련의 작업
- 분석 과제 및 프로젝트를 직접 수행하는 것은 아니지만, 어떠한 목표를 달성하기 위하여 어떤 데이터를 가지고 어떤 방식으로 수행할지에 대한 일련의 계획을 수립하는 작업이기 때문에 성공적인 분석 결과를 도출하기 위한 중요한 사전 작업.
- 분석을 기획한다는 것은 해당 문제 영역에 대한 전문성 역량 및 수학 / 통계학적 지식을 활용한 분석 역량과 분석의 도구인 데이터 및 프로그래밍 기술 역량에 대한 균형 잡힌 시각을 가지고 방향성 및 계획을 수립해야 한다는 것을 의미.

빅데이터의 주요 요소 3가지



빅 데이터(Big Data)

- 데이터 자원 확보
- 데이터 품질 관리

빅데이터 플랫폼 (Big Data Platform)

- 데이터 저장, 관리 기술(NoSQL, ETL..)
- 대용량 데이터 처리 (Hadoop, MapReduce..)
- 빅데이터 분석(자연어 처리, 의미분석, 데이터 마이닝..)
- 시각화(Visualization)

데이터 사이언티스트 (Data Scientist)

- 수학, 공학(IT기술과 엔지니어링) 능력
- 경제학, 통계학, 심리학 등 다문학적 이해
- 비판적 시각과 커뮤니케이션 능력
- 스토리텔링 등 시각화 능력

출처 : i-bada.blogspot.com/2012/04/blog-post_3023.html

데이터 사이언티스트의 역량

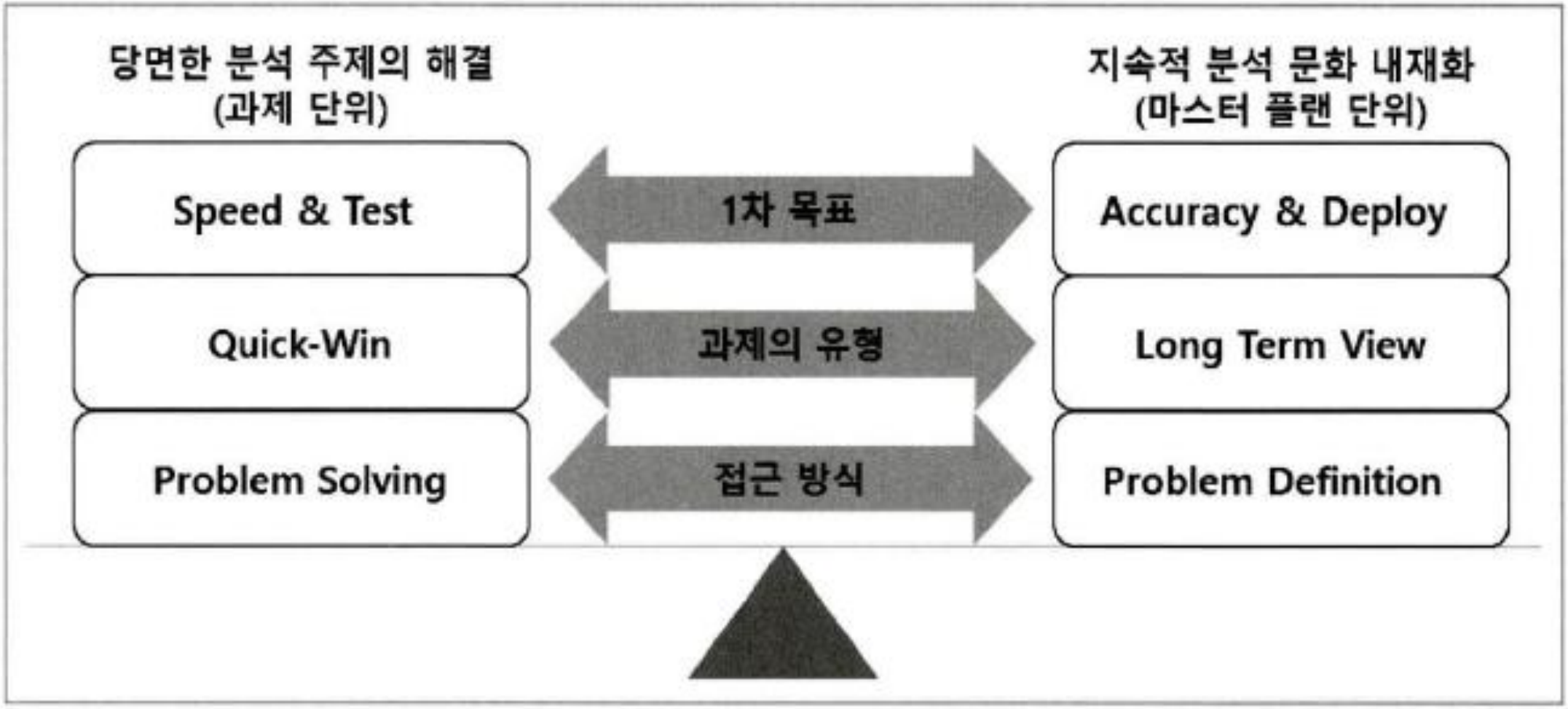
출처 : news.joins.com/article/23112392

분석주제 유형

분석의 방법 (How)	분석의 대상 (What)	*Known*	*Un-Known*
	Known	① Optimization	③ Insight
	Un-Known	② Solution	④ Discovery

출처 : https://m.blog.naver.com/PostView.naver?isHttpsRedirect=true&blogId=yjhead&logNo=222059021396

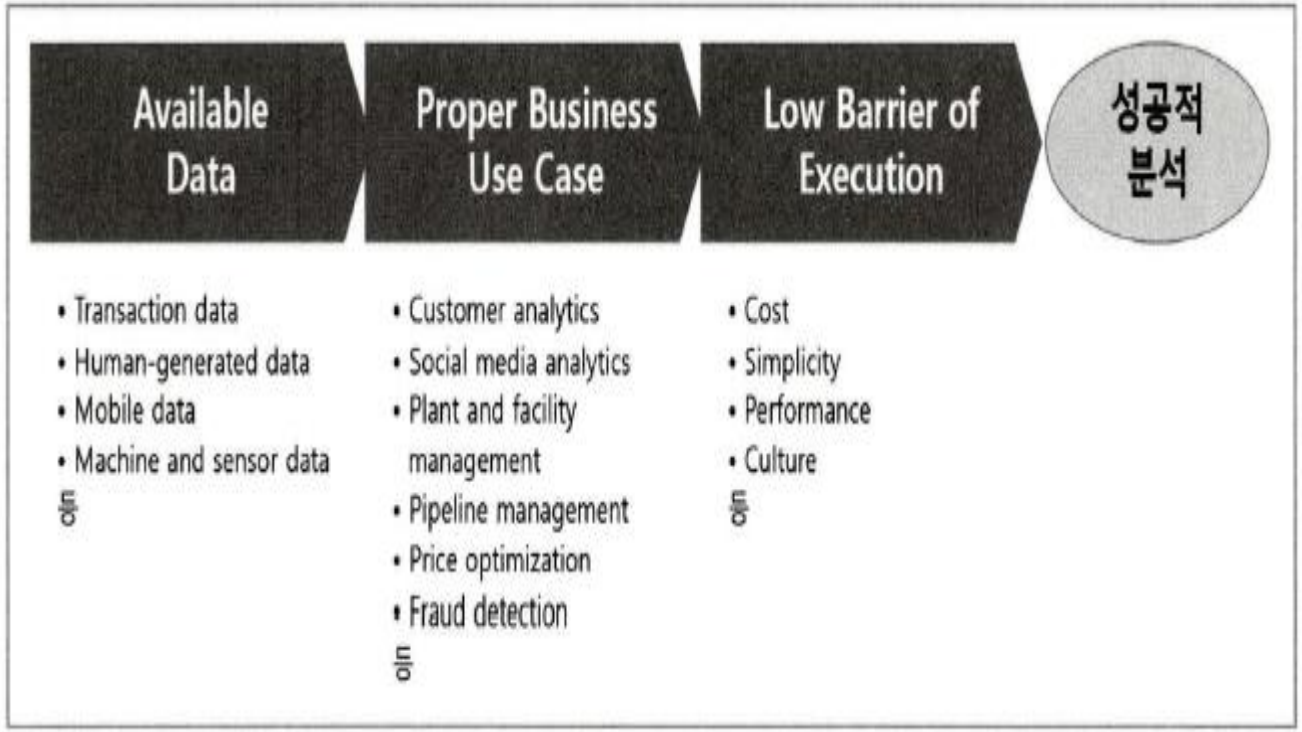
목표시점 별 분석 기획 방안



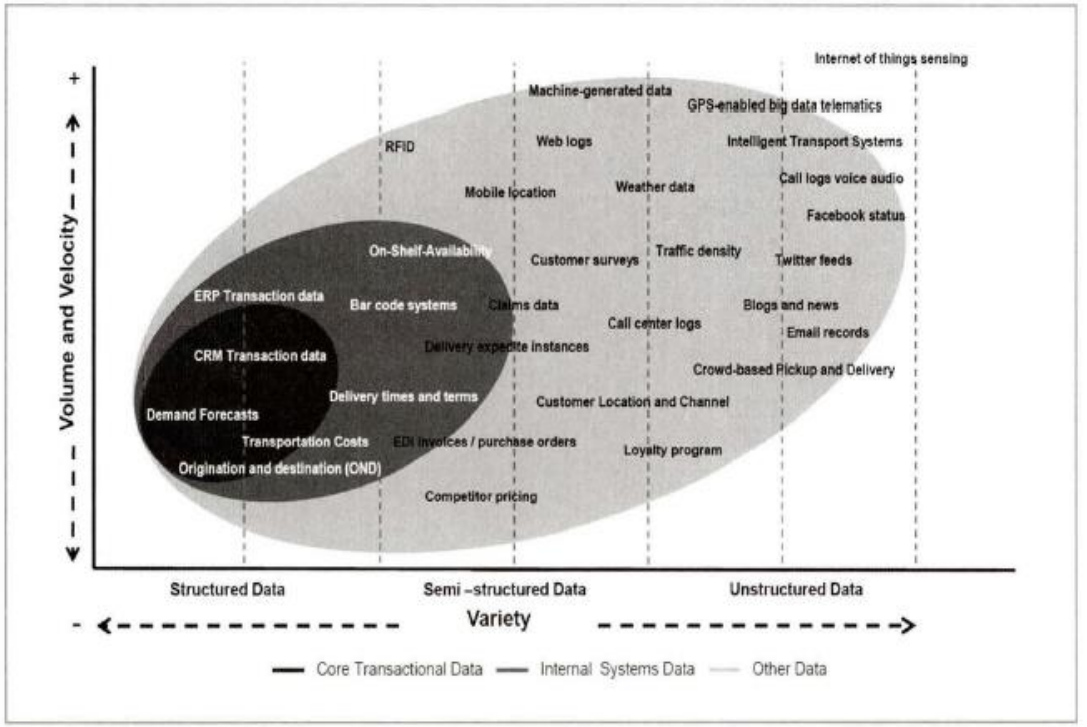
[그림 Ⅲ-1-3] 목표 시점 별 분석 기획 방안

출처 : https://jeonggi.tistory.com/

분석 기획시 고려사항



[그림 Ⅲ-1-6] 분석 기획 시 고려사항



[그림 Ⅲ-1-5] 다양한 데이터 유형

분석 방법론

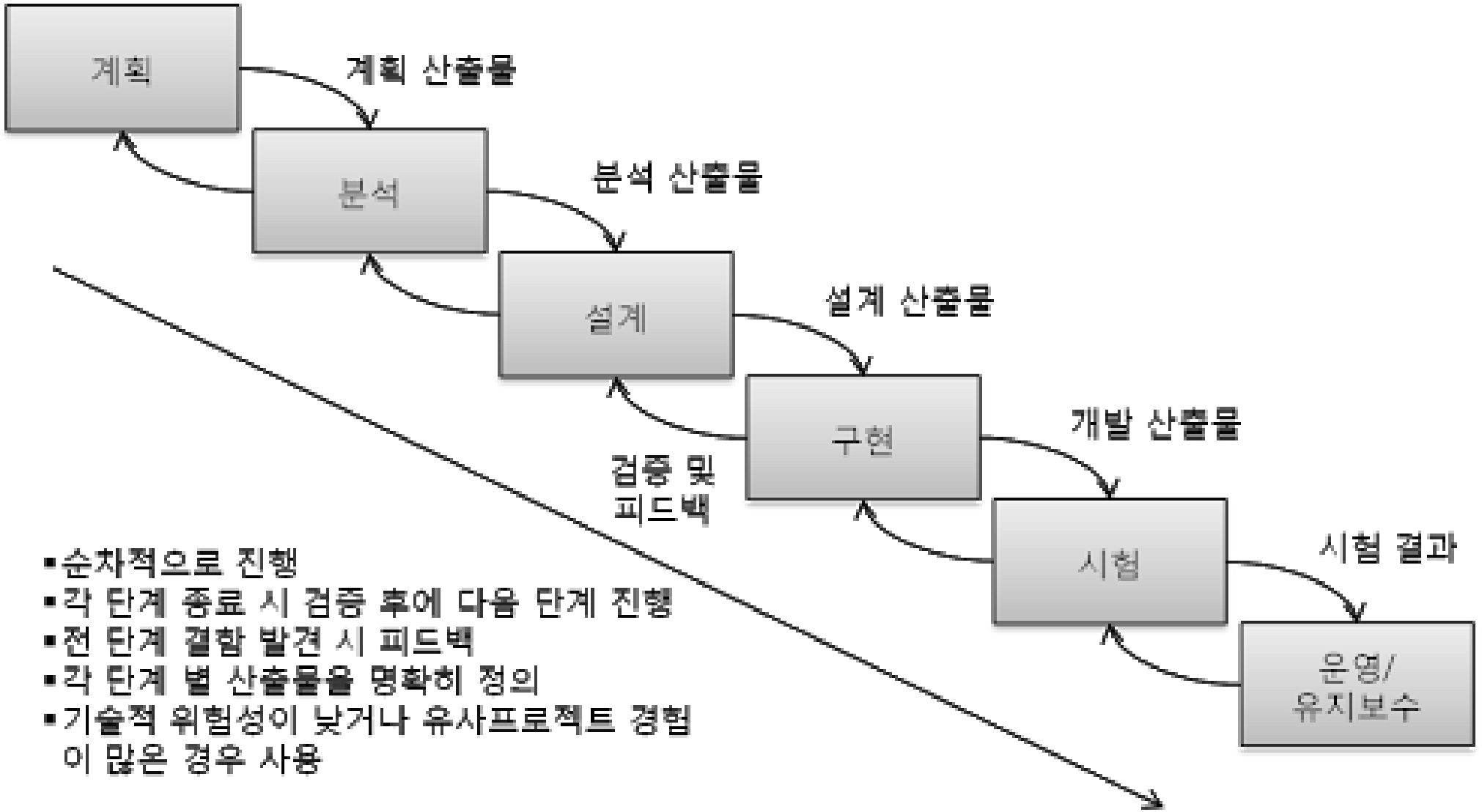
- 데이터 분석이 효과적으로 기업 내에 정착하기 위해서는 이를 체계화한 절차와 방법이 정리된 데이터 분석 방법론의 수립이 필수적.
- 방법론은 절차, 방법, 도구와 기법, 템플릿과 산출물로 구성되어 어느 정도 지식만 있으면 활용 가능.
- **데이터 기반 의사 결정 필요성**

  1. 경험과 감에 따른 의사결정 –〉 데이터 기반의 의사결정

  2. 기업의 합리적 의사결정을 가로막는 장애요소:

  고정관념, 편향된 생각, 프레이밍 효과(문제의 표현 방식에 따라 동일한 사건이나 상황임에도 불구하고 개인의 판단이나 선택이 달라질 수 있는 현상 등)

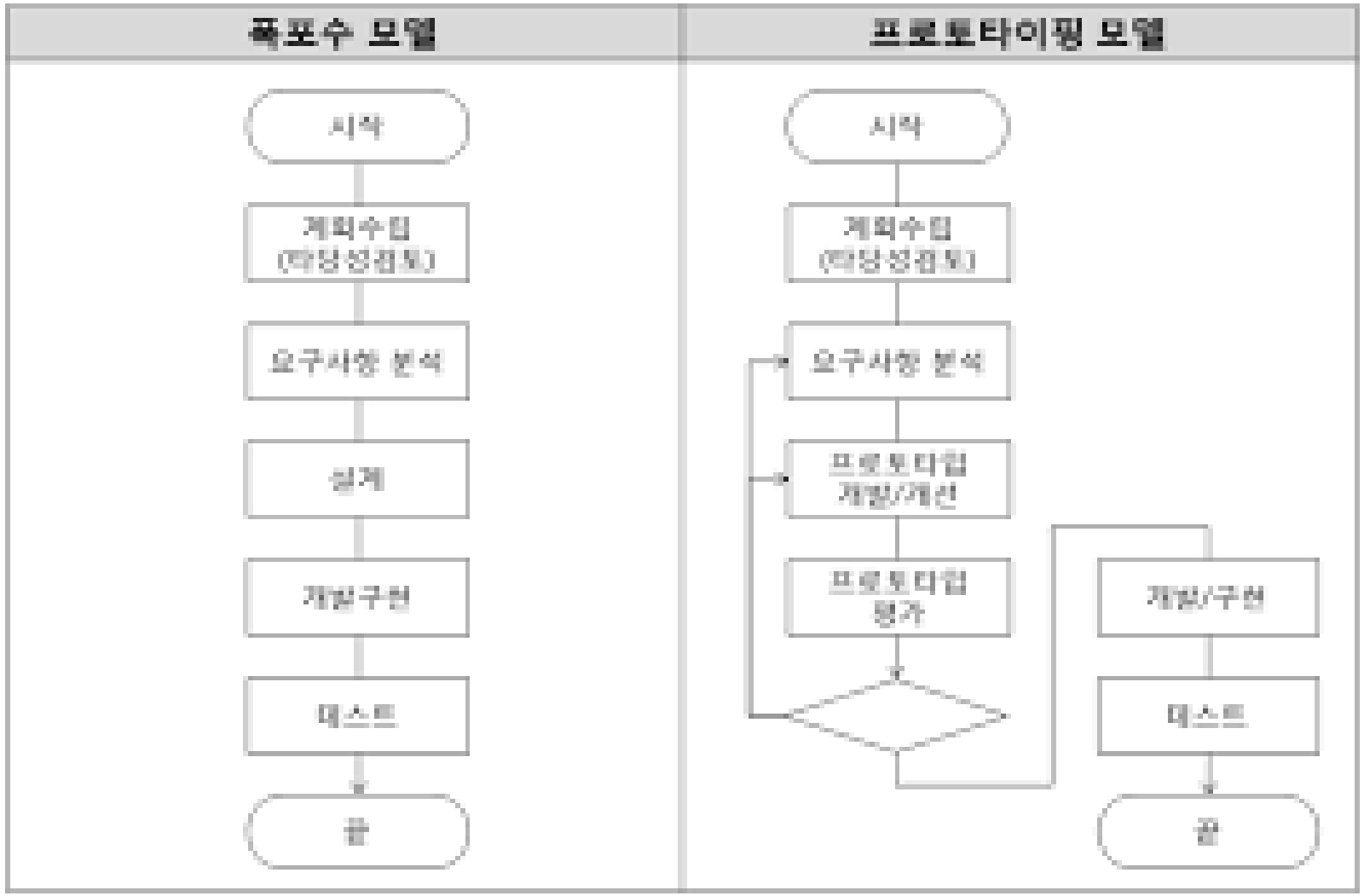
분석 방법론-폭포수 모델(Waterfall Model)



- 순차적으로 진행
- 각 단계 종료 시 검증 후에 다음 단계 진행
- 전 단계 결함 발견 시 피드백
- 각 단계 별 산출물을 명확히 정의
- 기술적 위험성이 낮거나 유사프로젝트 경험이 많은 경우 사용

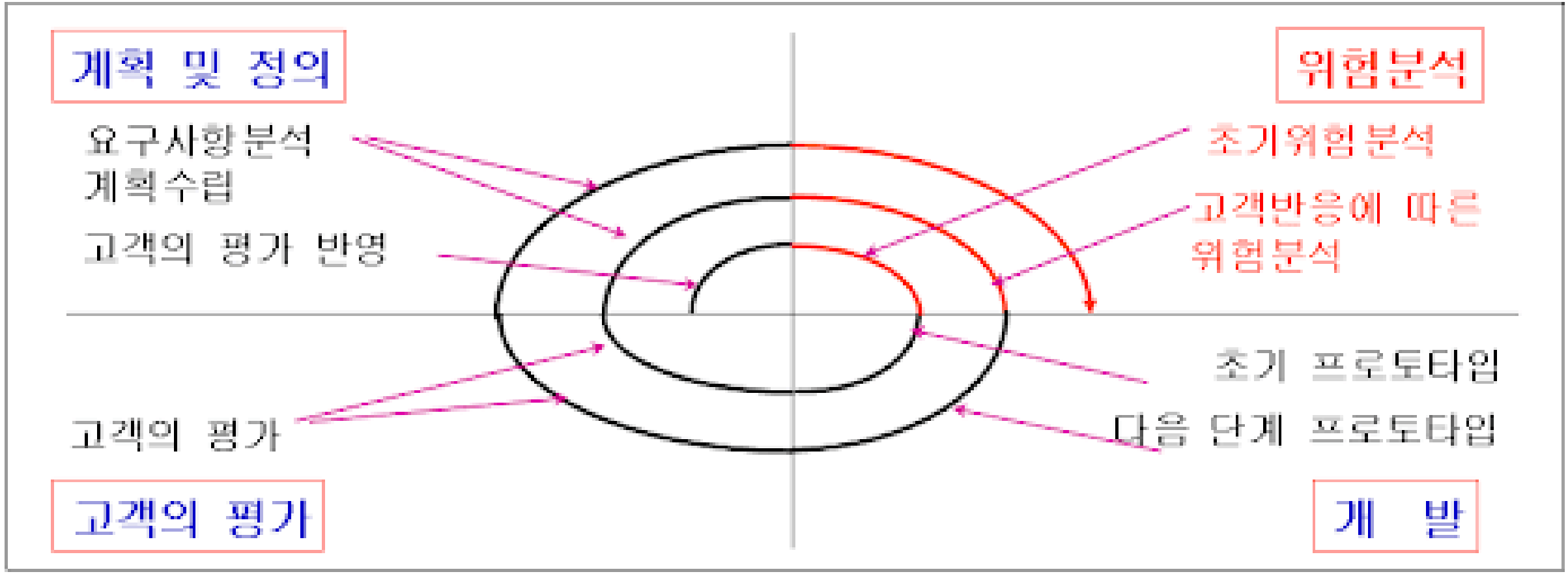
출처 : http://blog.skby.net/%ED%8F%AD%ED%8F%AC%EC%88%98-%EB%AA%A8%EB%8D%B8-waterfall-model/

분석 방법론-프로토타입 모델(Prototype model)



출처 : https://m.blog.naver.com/PostView.nhn?isHttpsRedirect=true&blogId=seilius&logNo=130185596031

분석 방법론-나선형 모델(Spiral model)



출처 : https://m.blog.naver.com/PostView.nhn?isHttpsRedirect=true&blogId=seilius&logNo=130185596031

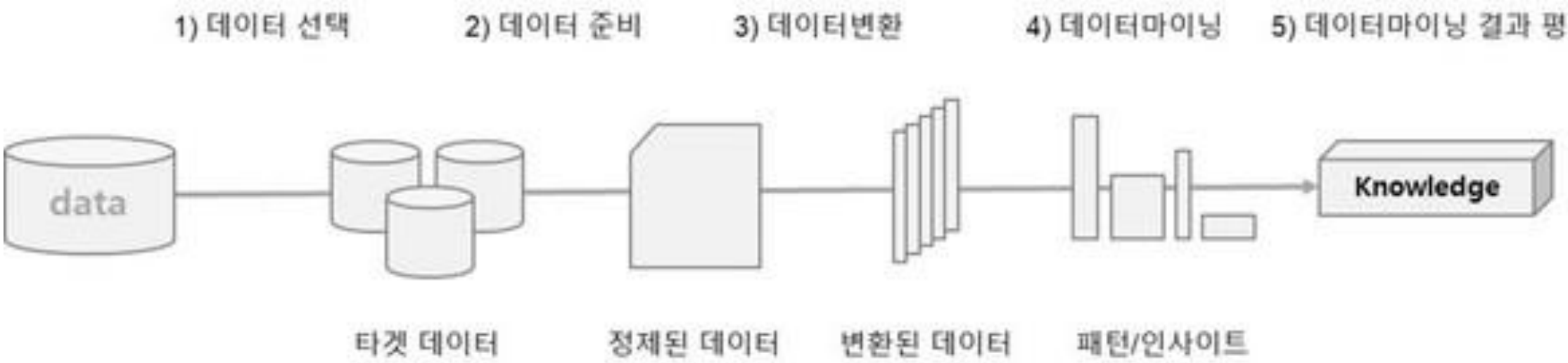
분석 방법론-나선형 모델(Spiral model)

구분	폭포수 모델	프로토타입 모델	나선형 모델
Concept			
특징	순차적 접근	프로로 타입 개발	위험분석~개발 반복수행
장점	이해가 쉽고 관리가 편함	요구분석 용이 개발 타당성 검증가능	위험성 감소와 변경에 유연한 대처
단점	초기 요구분석이 어려워 후반 문제발생 가능성 有	프로토 타입 평가 후 취소 시 폐기비용발생	단계 반복에 따른 공정관리가 어려움

출처 : https://m.blog.naver.com/PostView.naver?isHttpsRedirect=true&blogId=yjhead&logNo=222059021396

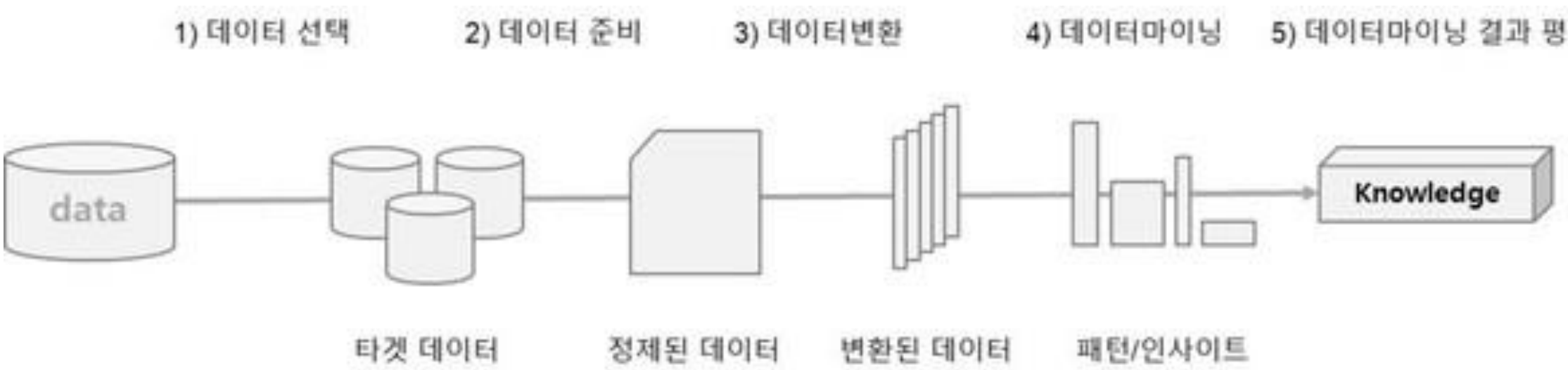
KDD 분석 방법론

- KDD(Knowledge Discovery in Database) 분석 방법론은 1996년 Fayyad가 소개한 방법론으로 데이터를 통해 통계적 패턴이나 지식을 찾을 수 있도록 정리한 데이터마이닝 프로세스이다. 데이터마이닝, 기계학습, 인공지능, 패턴인식, 데이터 시각화에서 응용 될 수 있는 구조를 갖고 있다. KDD 분석 방법론은 데이터셋 선택, 데이터 전처리, 데이터 변환, 데이터마이닝, 결과 평가로 이루어져 있다.



출처 : https://www.2e.co.kr/news/articleView.html?idxno=301010

KDD 분석 방법론



1) 데이터 선택(Selection)

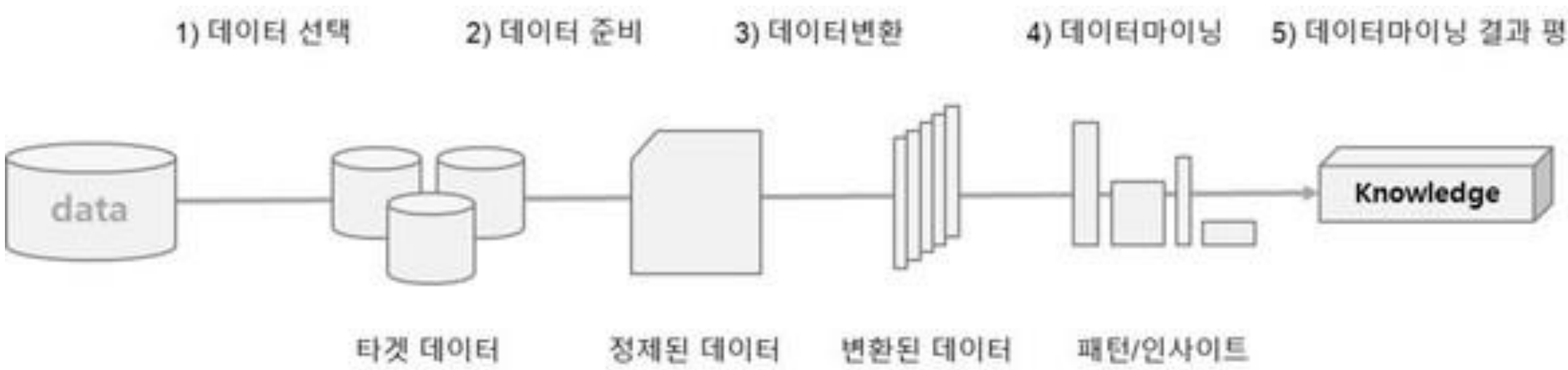
- 데이터셋 선택에 앞서 분석 대상의 비즈니스 도메인에 대한 이해와 프로젝트 목표 설정이 필수이며 데이터베이스 또는 원시 데이터에서 분석에 필요한 데이터를 선택하는 단계.

2) 데이터 전처리(preprocessing)

- 추출된 분석 대상용 데이터 셋에 포함되어 있는 잡음(Noise)과 이상치(Outlier), 결측치(Missing value)를 식별하고 필요시 제거하거나 의미있는 데이터로 재처리하여 데이터 셋을 정제하는 단계

출처 : https://www.2e.co.kr/news/articleView.html?idxno=301010

KDD 분석 방법론



3) 데이터 변환(Transformation)

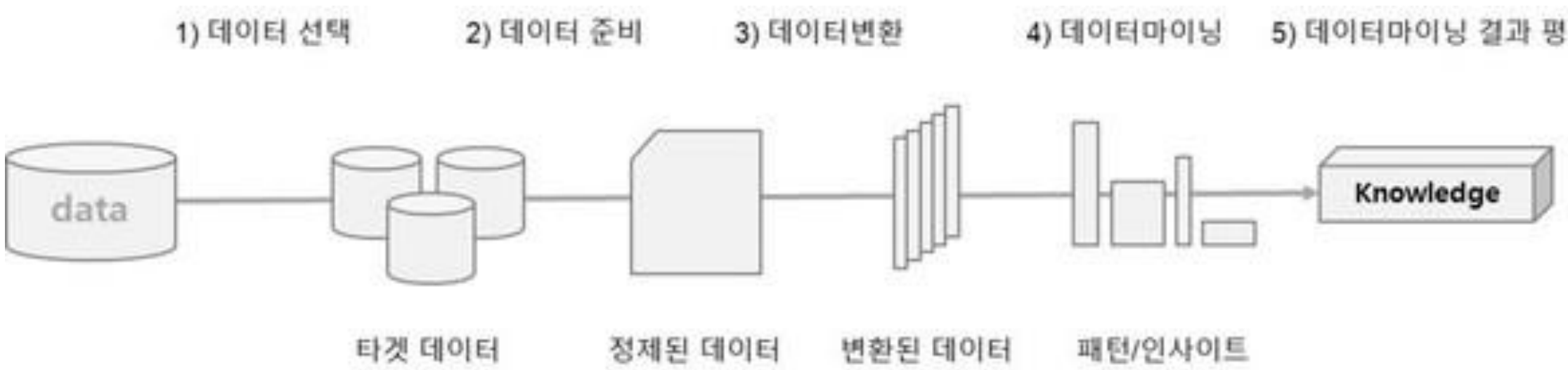
- 데이터 전처리 과정을 통해 정제된 데이터에 분석 목적에 맞게 변수를 생성, 선택하고 데이터의 차원을 축소하여 효율적으로 데이터 마이닝을 할 수 있도록 변경하는 단계 / 프로세스를 진행하기 위해 학습용 데이터와 검증용 데이터로 분리하는 단계

4) 데이터 마이닝(Data Mining)

- 학습용 데이터를 이용하여 분석목적에 맞는 데이터 마이닝 기법을 선택하고, 적절한 알고리즘을 적용하여 데이터 마이닝 작업을 수행 / 필요에 따라 전처리와 데이터 변환 프로세스를 추가로 실행하여 최적의 결과를 산출.

출처 : https://www.2e.co.kr/news/articleView.html?idxno=301010

KDD 분석 방법론



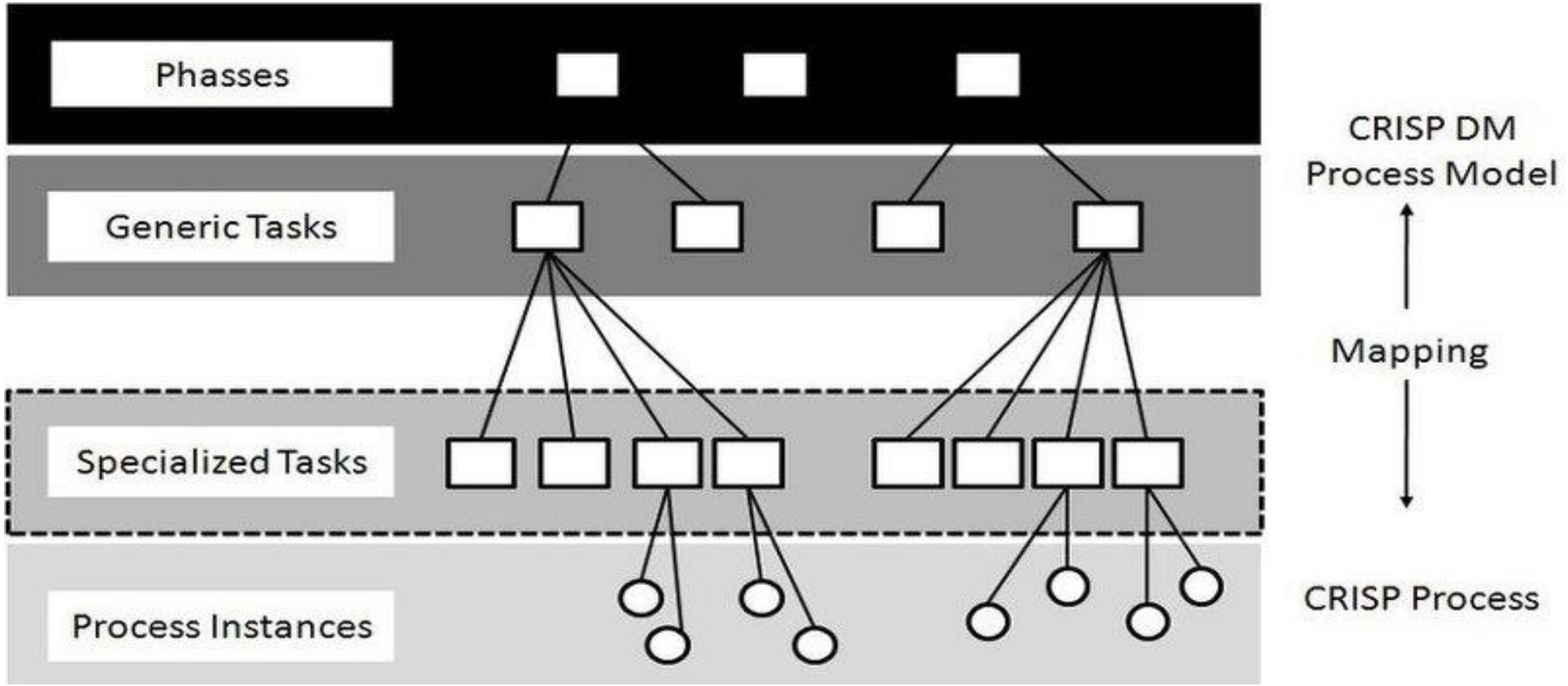
5) 데이터 마이닝 결과 평가(Interpretation / Evaluation)

- 데이터 마이닝 결과에 대한 해석과 평가, 그리고 분석목적과의 일치성을 확인.
- 데이터 마이닝을 통해 발견한 지식을 업무에 활용하기 위한 방안 마련의 단계
- 필요에 따라 데이터 선택 프로세스에서 데이터 마이닝 프로세스를 반복 수행.

출처 : https://www.2e.co.kr/news/articleView.html?idxno=301010

CRISP-DM 분석 방법론

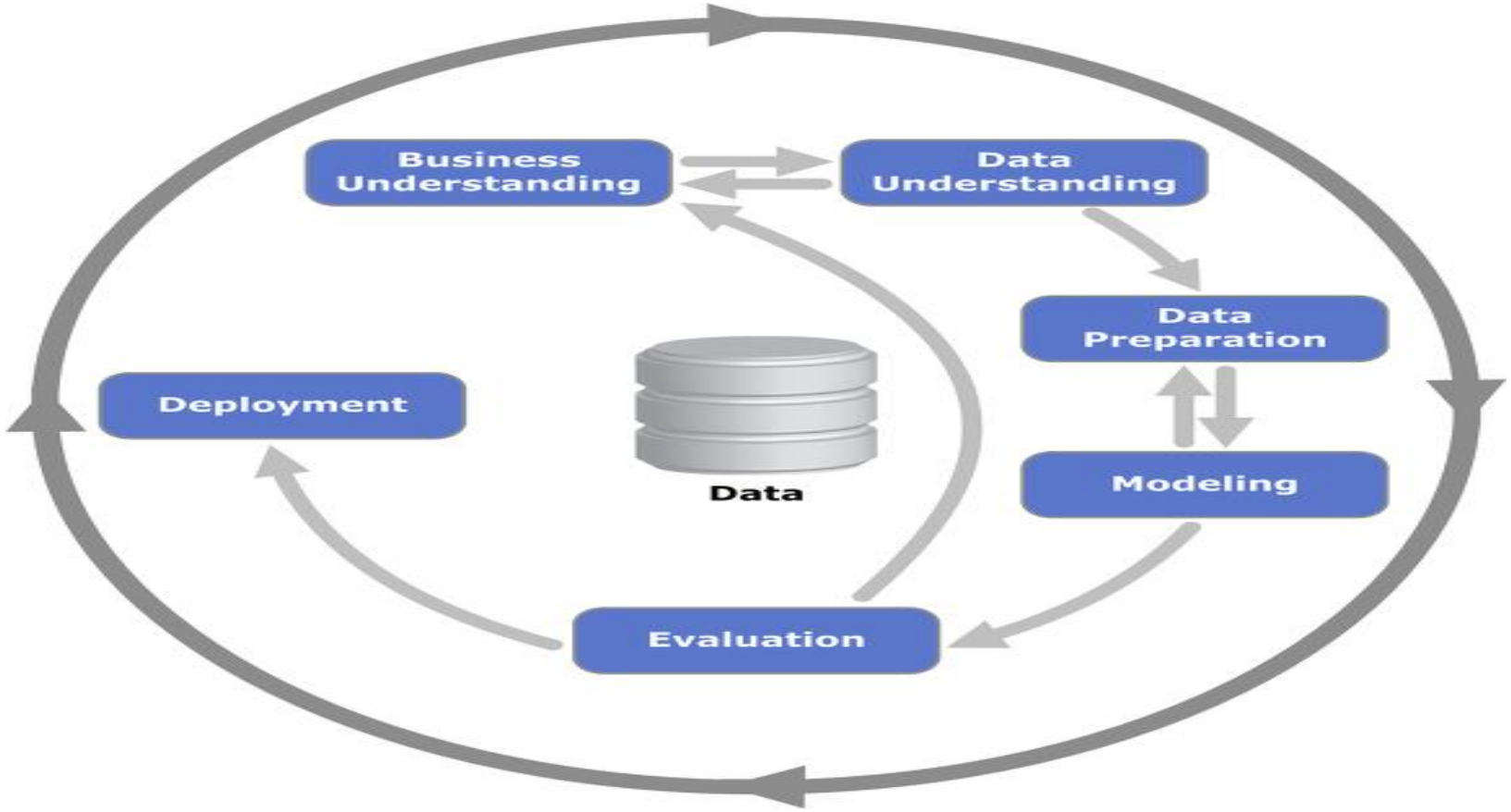
- CRISP-DM(Cross Industry Standard Process for Data Mining) 방법론은 전 세계에서 가장 많이 사용되는 데이터마이닝 표준 방법론으로 단계, 일반 과제, 세부과제, 프로세스 실행 등의 4가지 레벨로 구성된 계층적 프로세스 모델이기도 하다.



출처 : https://needjarvis.tistory.com/512

CRISP-DM의 프로세스

- CRISP-DM의 절차는 6단계로 구성되어 있는데 각 단계들은 순차적으로 진행되는 것이 아니라, 필요에 따라 단계 간의 반복 수행을 통해 분석의 품질을 향상시킨다.



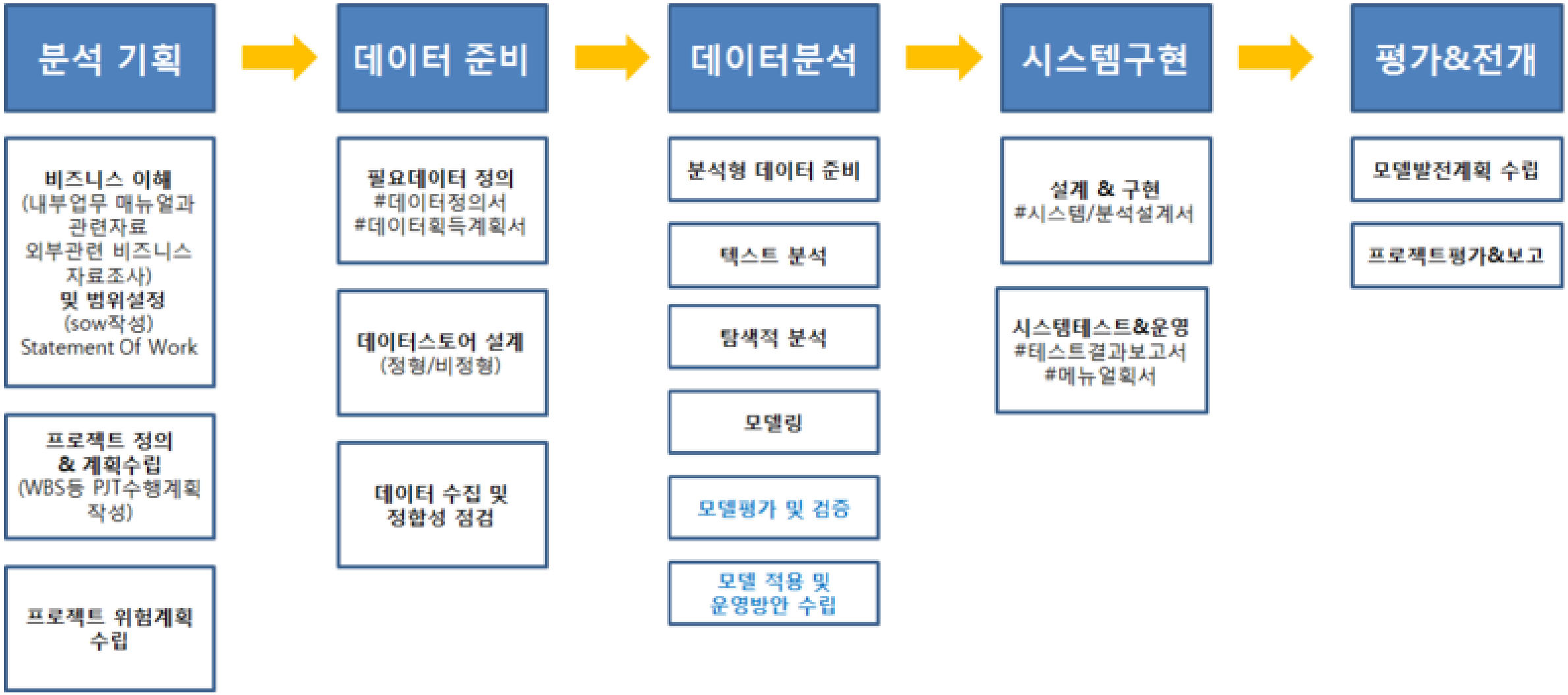
출처 : https://needjarvis.tistory.com/512

KDD와 CRISP-DM의 비교

KDD	CRISP-DM
분석대상 비즈니스 이해	업무 이해
데이터셋 선택	데이터의 이해
데이터 전처리	
데이터 변환	데이터 준비
데이터 마이닝	모델링
데이터 마이닝 결과 평가	평가
데이터 마이닝 활용	전개

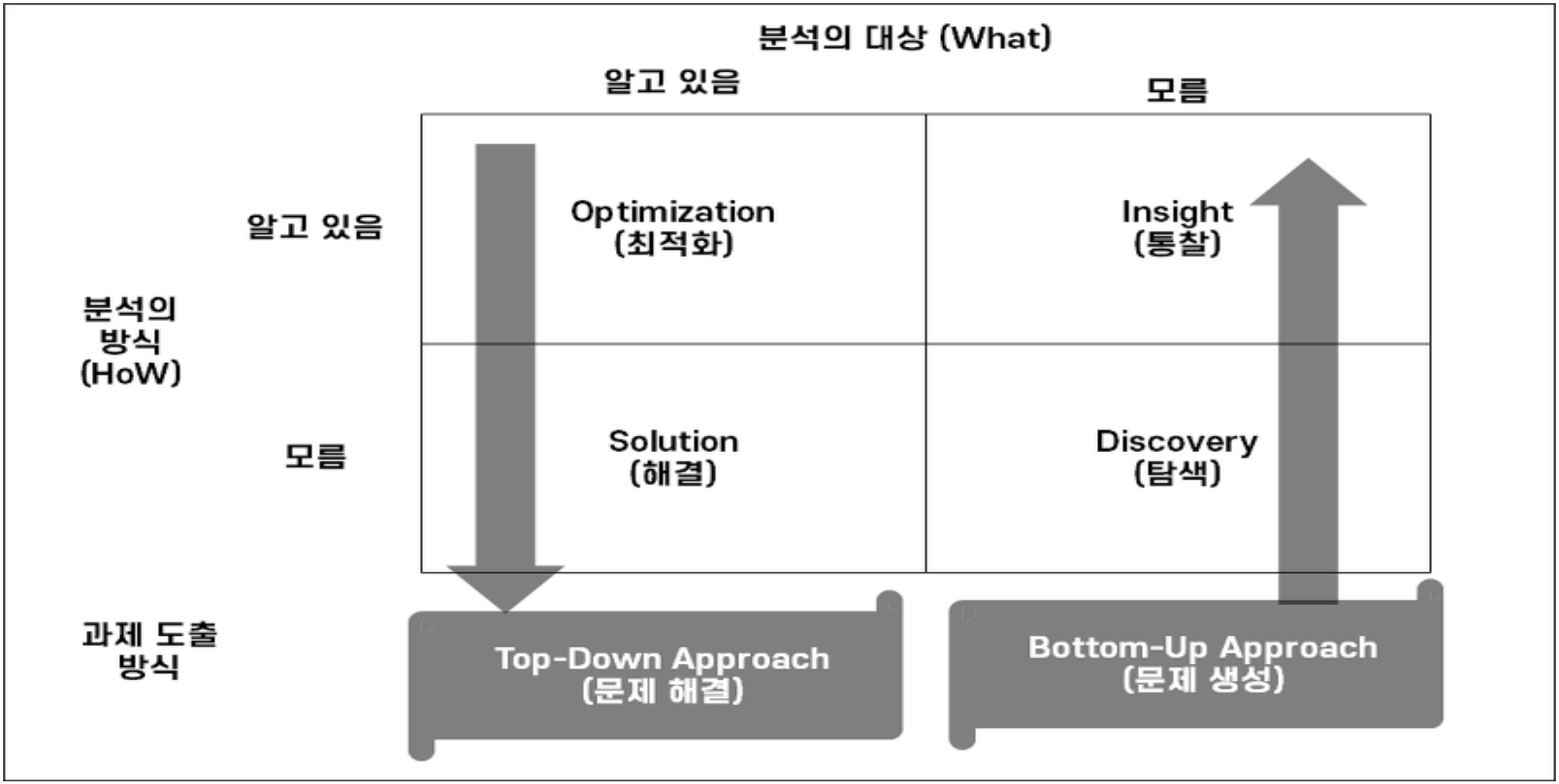
출처 : https://velog.io/@eunpro/2-1-2-%EB%B6%84%EC%84%9D-%EB%B0%A9%EB%B2%95%EB%A1%A0

빅데이터 분석 5단계



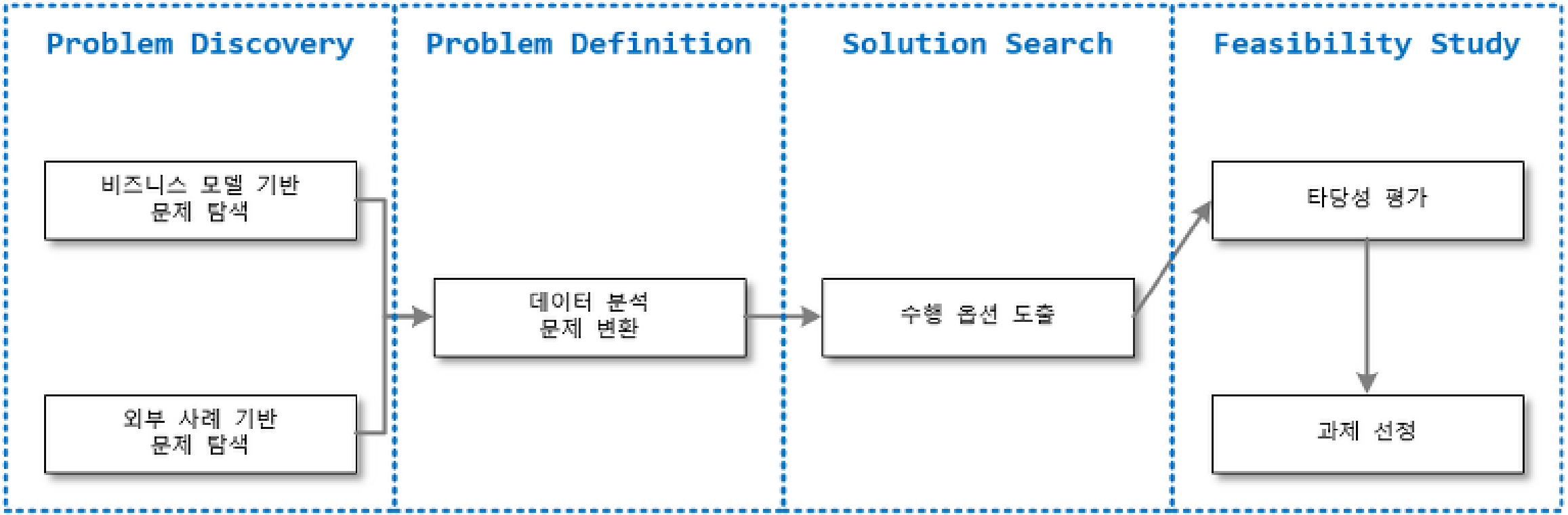
출처 : https://m.blog.naver.com/PostView.naver?isHttpsRedirect=true&blogId=pouing_v&logNo=221478992326

분석 과제 발굴



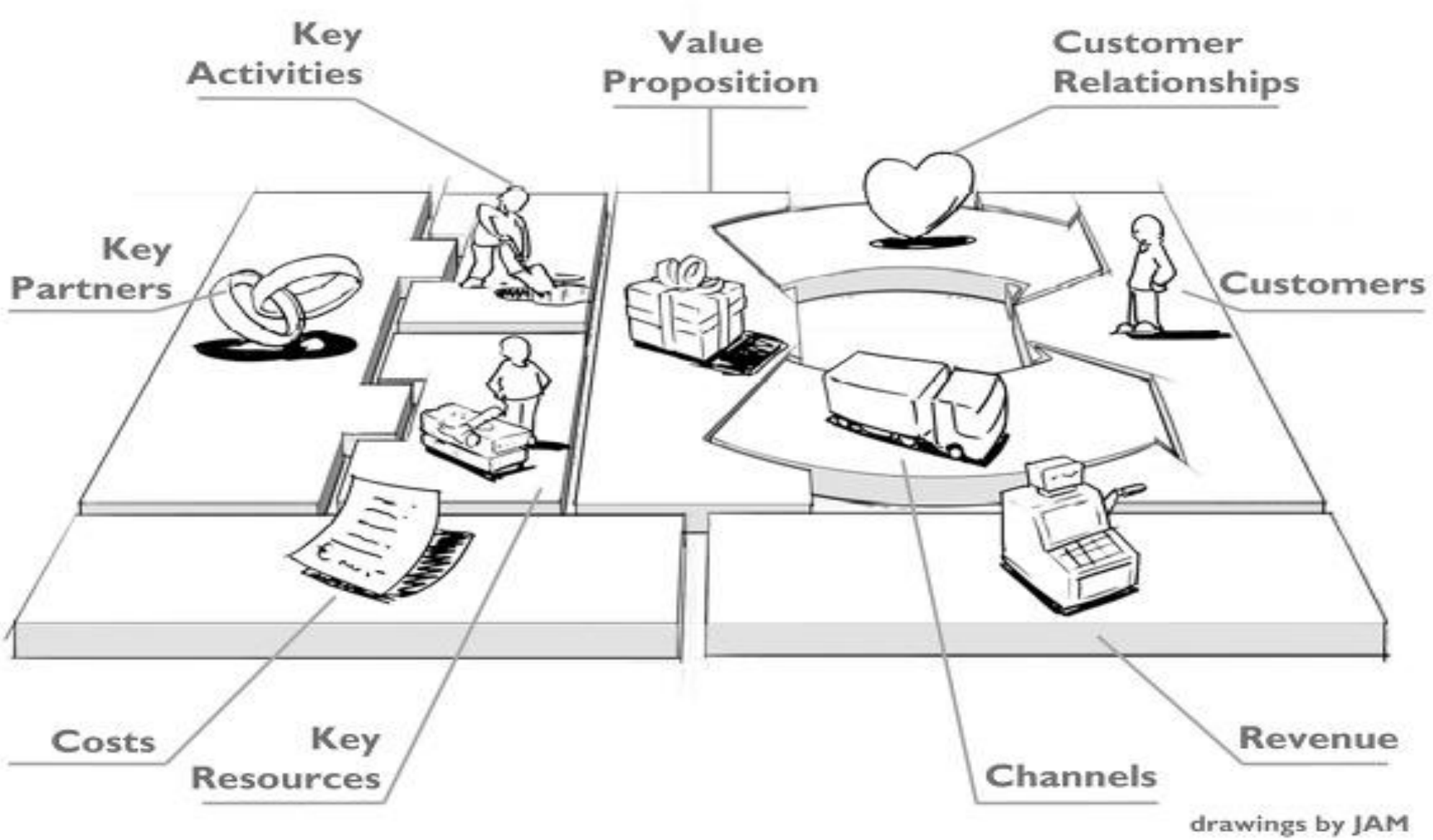
출처 : https://needjarvis.tistory.com/527

하향식 접근 방식(Top Down Approach)



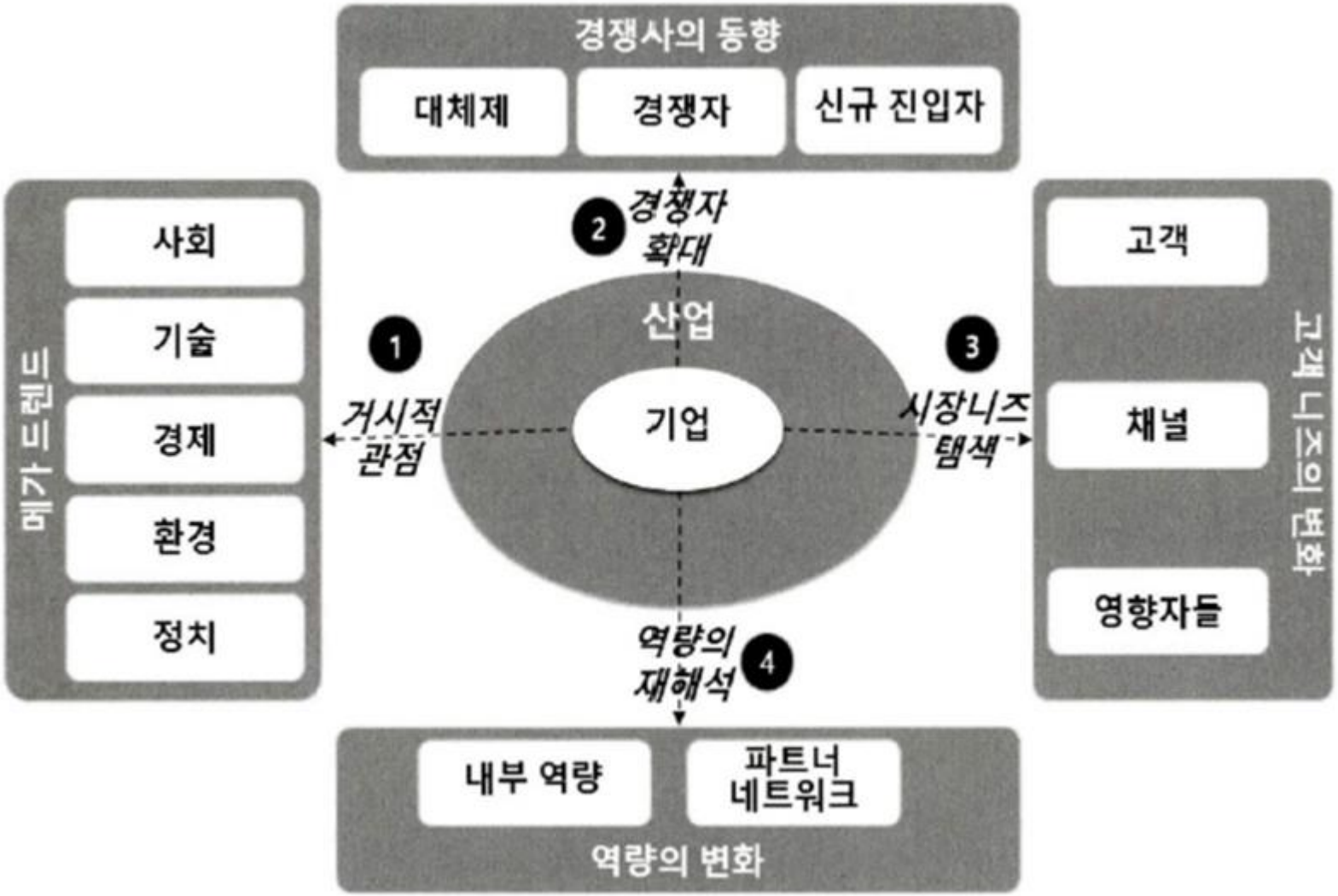
출처 : https://codedragon.tistory.com/6909

비즈니스 모델 기반 문제 탐색- 하향식 접근법 1단계



출처 : https://needjarvis.tistory.com/528?category=850659

분석 기회 발굴의 범위 확장-하향식 접근법 1단계

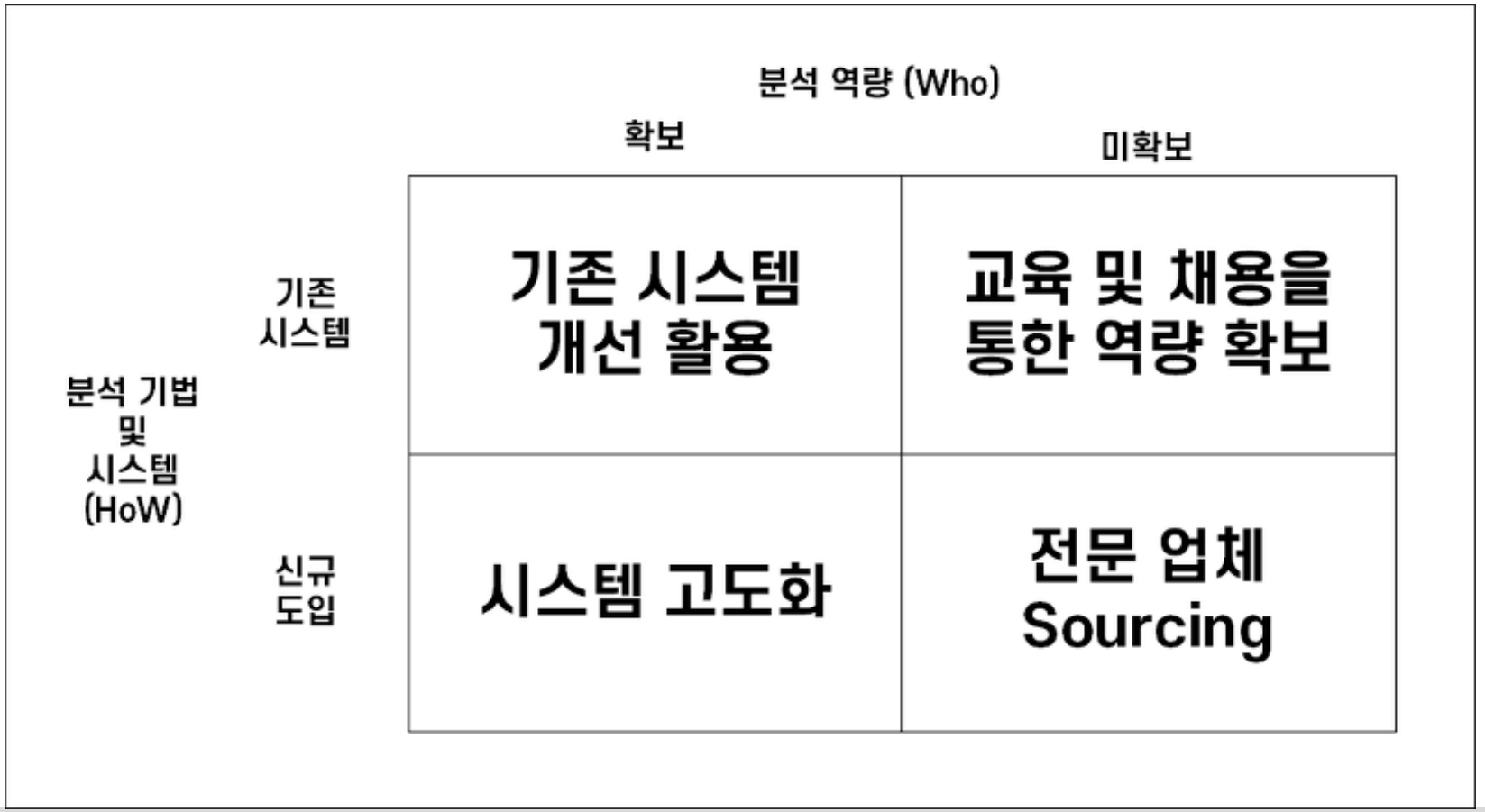


데이터 분석 기반 문제 탐색-하향식 접근법 2단계

- 데이터 분석 문제의 정의 및 요구사항 : 분석을 수행하는 당사자뿐만 아니라 해당 문제가 해결되었을 때 효용을 얻을 수 있는 최종 사용자(End User) 관점에서 이루어져야 한다.
- 데이터 분석 문제가 잘 정의 되었을 때 필요한 데이터의 정의 및 기법 발굴이 용이하기 때문에 가능한 정확하게 분석의 관점으로 문제를 재정의할 필요가 있다.

데이터 분석 기반 문제 탐색- 하향식 접근 3단계

- 데이터 분석 문제를 해결하기 위한 다양한 방안이 모색된다.
- 분석역량을 기존에 가지고 있는 지의 여부를 파악하여 보유하고 있지 않은 경우에는 교육이나 전문인력 채용을 통한 역량을 확보하거나 분석 전문업체를 활용하여 과제를 해결하는 방안에 대해 사전 검토를 수행한다.



데이터 분석 기반 문제 탐색- 하향식 접근 4단계

- 도출된 분석 문제나 가설에 대한 대안을 과제화하기 위해서는 다각적인 타당성 분석이 수행되어야 한다.
- 경제적 타당성 : 비용대비 편익 분석 관점의 접근이 필요.
- 데이터 및 기술적 타당성 : 데이터 분석에는 데이터 존재 여부, 분석 시스템 환경 그리고 분석 역량이 필요

분석 역량의 경우 실제 프로젝트 수행시, 걸림돌이 되는 경우가 많기 때문에 기술적 타당성 분석시 역량 확보 방안을 사전에 수립하고 이를 효과적으로 평가하기 위해서는 비즈니스 지식과 기술적 지식이 요구됨.

1) 평가 과정을 거쳐 가장 우월한 대안을 선택.
2) 도출한 데이터 분석 문제 및 선정된 솔루션 방안을 포함
3) 분석과제 정의서의 형태로 명시하는 후속작업을 시행
4) 프로젝트 계획의 입력물로 활용됨.

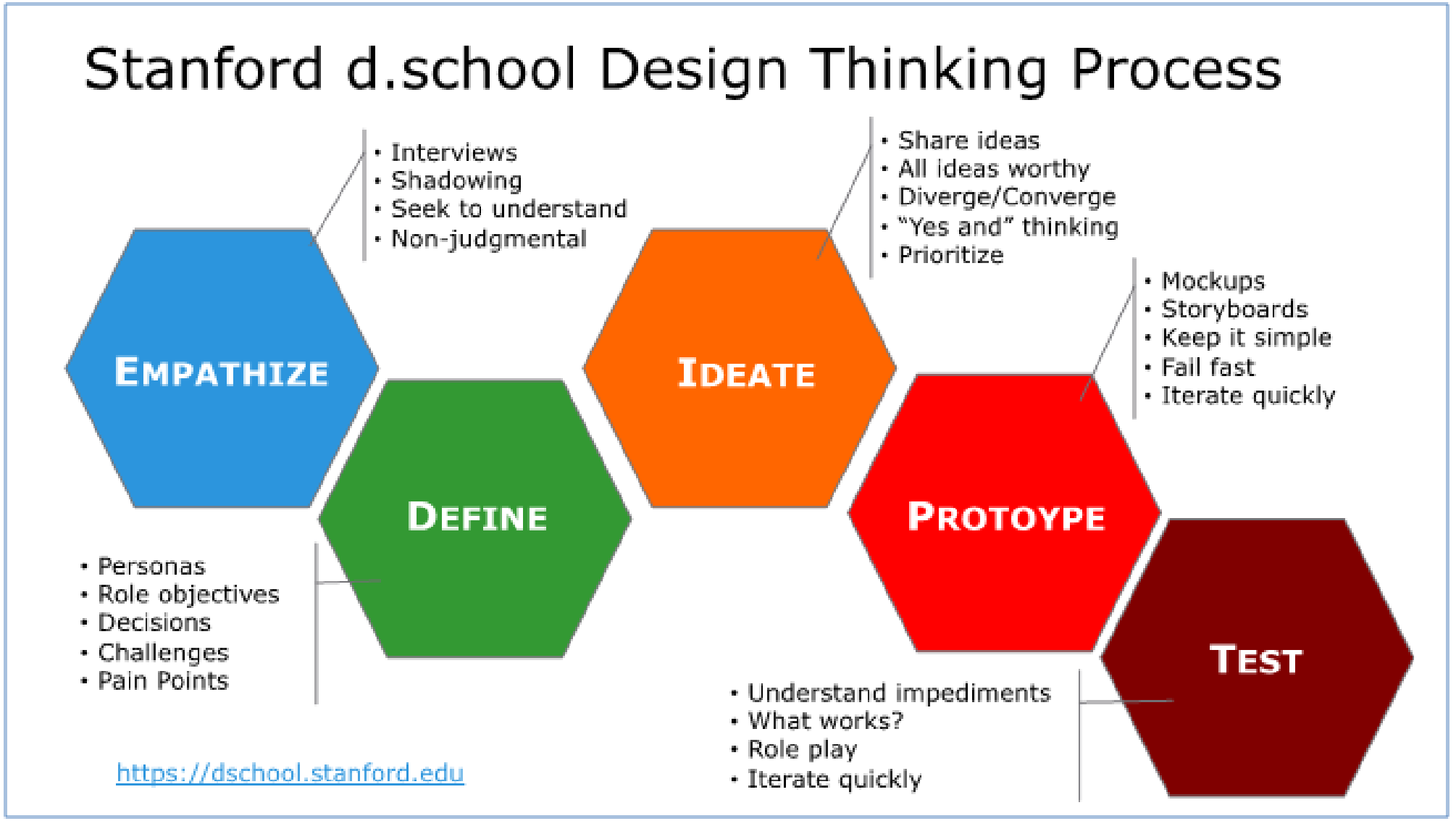
상향식 접근법(Bottom up Approach)

출처 : https://www.slideshare.net/StartupAlliance/171026-82508966

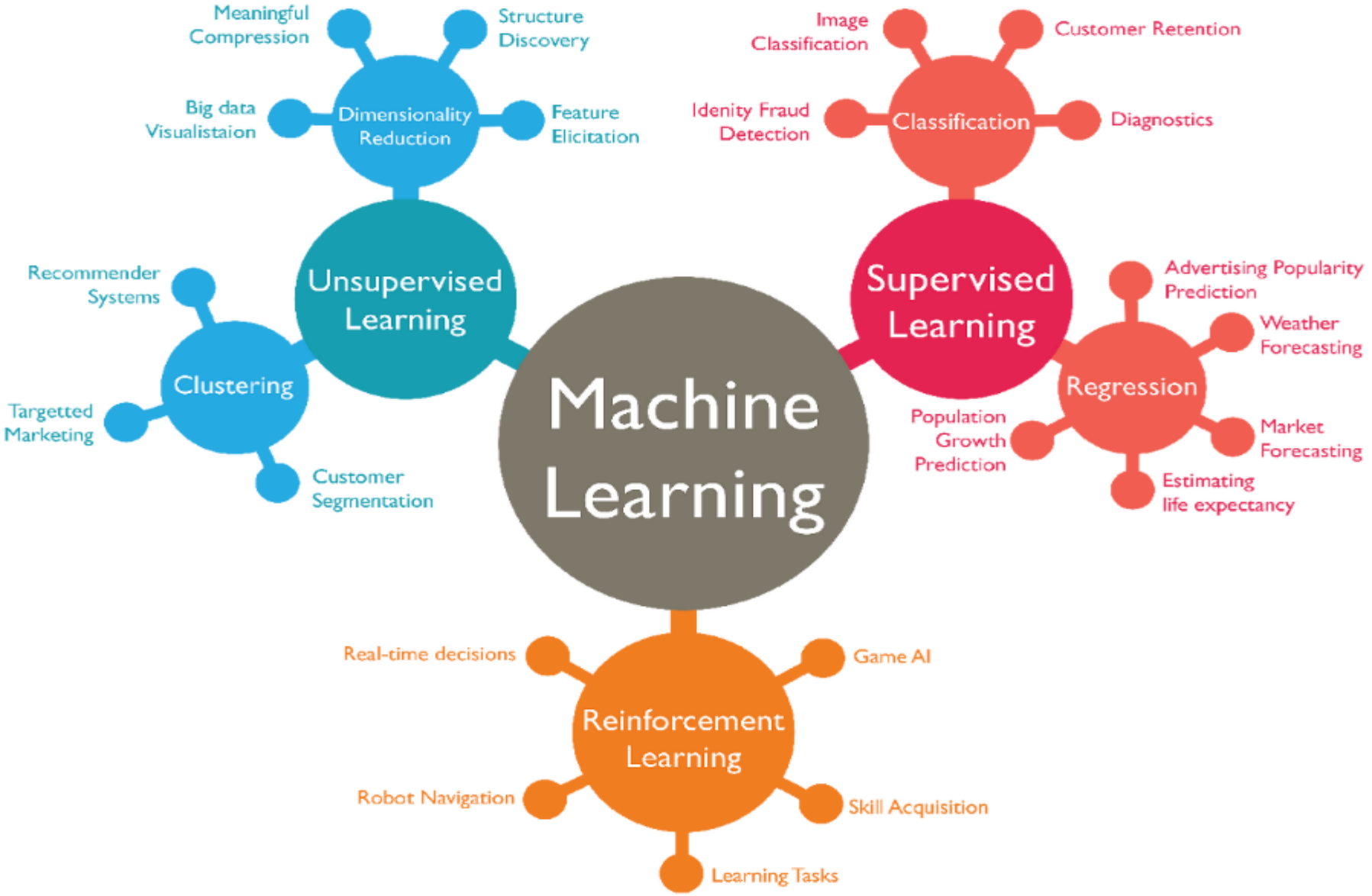
하향식 접근 방법의 한계

- 솔루션 도출에는 유효하지만 새로운 문제 탐색의 한계
- 논리적인 단계별 접근법 기반의 문제해결 방식은 복잡하고 다양한 환경에서 발생하는 문제에는 비적합
- 이를 해결하기 위해 스탠포드 대학의 d.school(Institute of Design at Stanford)은 디자인 사고(Design Thinking) 접근법을 통해 전통적인 분석적 사고를 극복하려고 함
- 통상적인 관점에서는 'Why'를 강조하지만, 있는 그대로 인식하는 'What' 관점 필요
- 이와 같은 점을 고려하여 d.school에서는 첫단계로 Empathize(감정이입)을 강조

하향식 접근 방법의 한계


Stanford d.school Design Thinking Process
• Interviews
• Shadowing
• Seek to understand
• Non-judgmental
EMPATHIZE
DEFINE
• Personas
• Role objectives
• Decisions
• Challenges
• Pain Points
IDEATE
• Share ideas
• All ideas worthy
• Diverge/Converge
• "Yes and" thinking
• Prioritize
PROTOYPE
• Mockups
• Storyboards
• Keep it simple
• Fail fast
• Iterate quickly
TEST
• Understand impediments
• What works?
• Role play
• Iterate quickly
https://dschool.stanford.edu

Machine learning

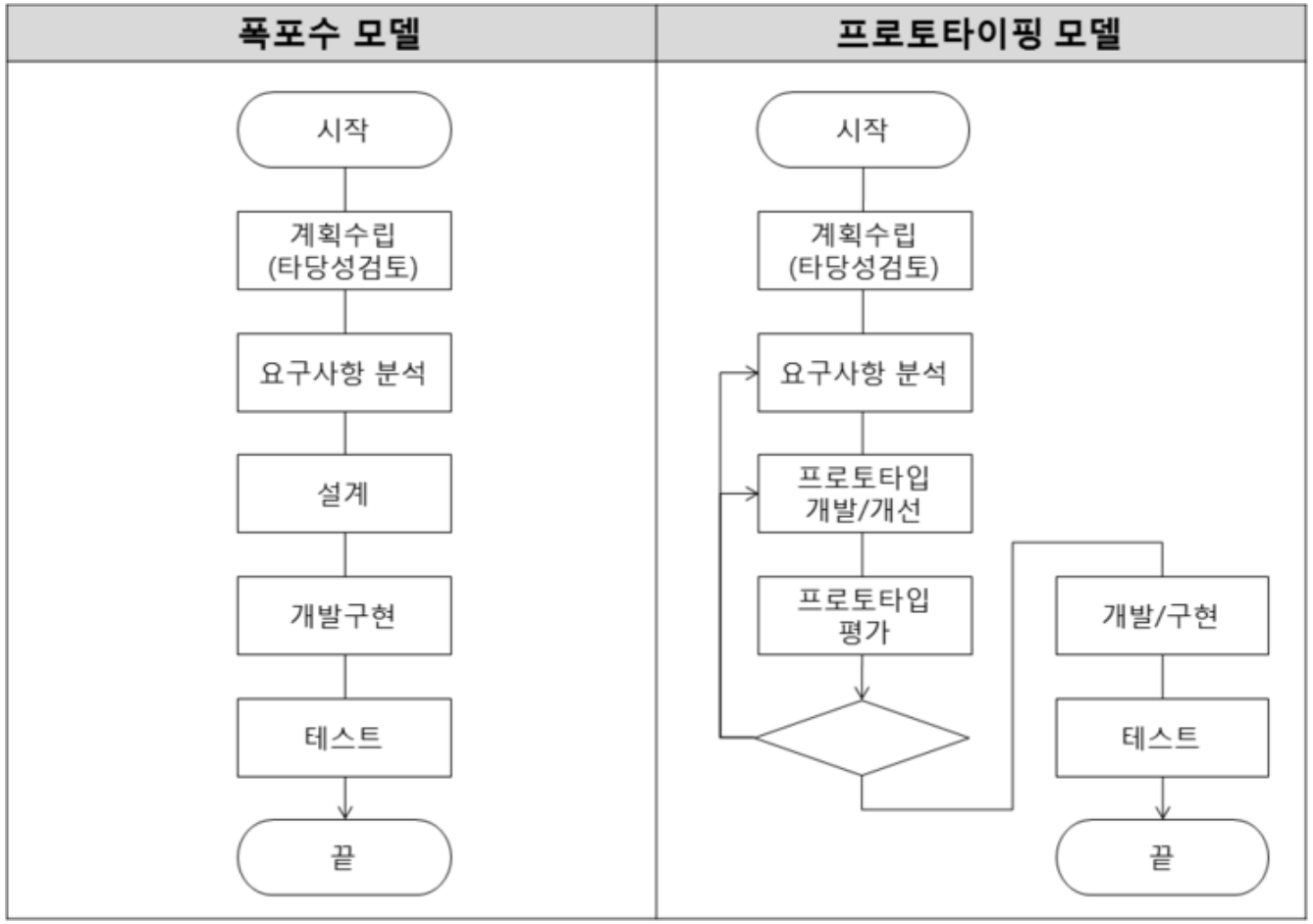


통계적 분석

- 인과관계 분석을 위해 가설을 설정하고 이를 검정하기 위해 모집단으로 부터 표본을 추출하고 그 표본을 이용한 가설 검정을 실시하는 방식으로 문제 해결
- 빅데이터 환경에서는 이와 같은 논리적인 인과관계 분석뿐만 아니라 상관관계 분석 또는 연관 분석을 통하여 다양한 문제 해결에 도움을 받을 수 있음.
- 인과관계에서 상관관계 분석으로의 이동이 빅데이터 분석에서의 주요 변화
- 다량의 데이터 분석을 통해서 “왜” 그러한 일이 발생하는지 역으로 추적하면서 문제를 도출하거나 재정의 할 수 있는 것이 상향식 접근법.

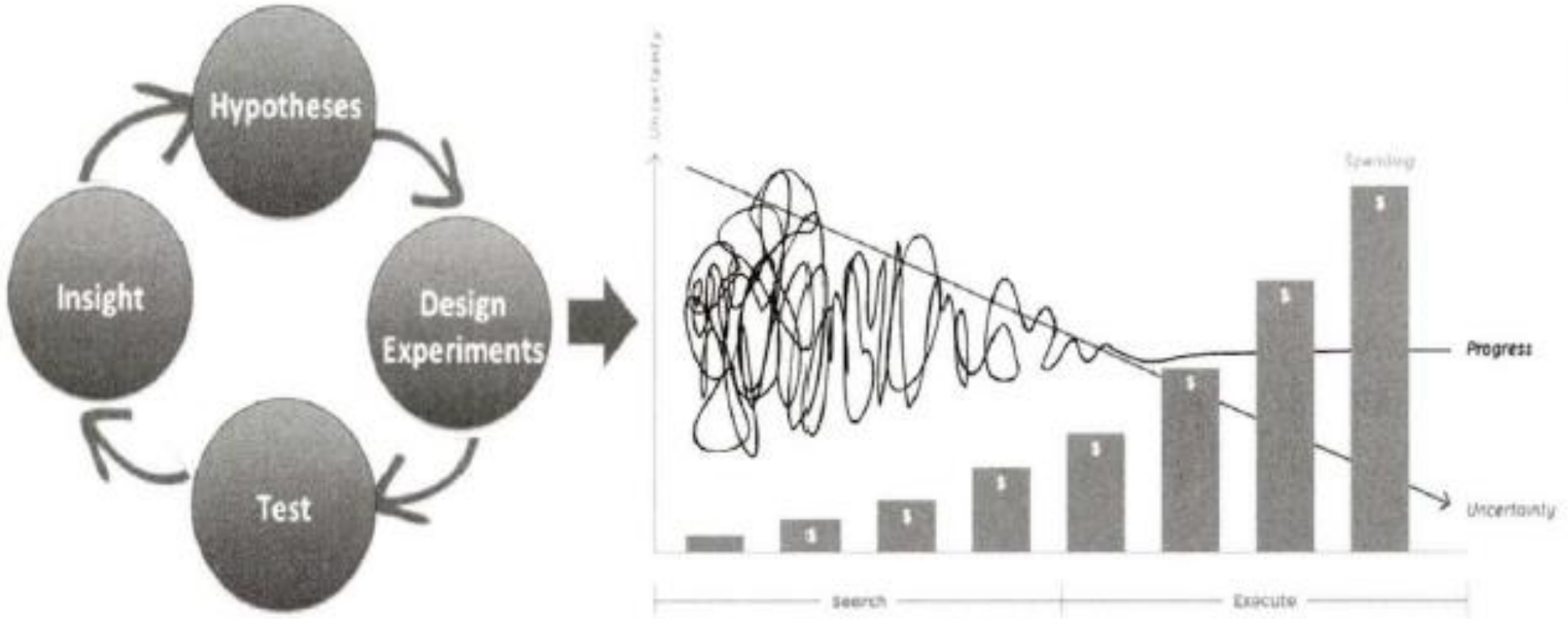
프로토타이핑 모델

- 사용자가 요구사항이나 데이터를 정확히 규정하기 어렵고 데이터 소스도 명확히 파악하기 어려운 상황에서 일단 분석을 시도해 보고 그 결과를 확인해 가면서 반복적으로 개선해 나가는 방법



프로토타이핑 모델

- 사용자가 요구사항이나 데이터를 정확히 규정하기 어렵고 데이터 소스도 명확히 파악하기 어려운 상황에서 일단 분석을 시도해 보고 그 결과를 확인해 가면서 반복적으로 개선해 나가는 방법



빅데이터 분석 환경에서 프로토타이핑의 필요성

1. **문제에 대한 인식 수준**

   : 문제 정의가 불명확하거나 이전에 접해보지 못한 새로운 문제일 경우 사용자 및 이해관계자는 프로토타입을 이용하여 문제를 이해하고, 이를 바탕으로 구체화함.

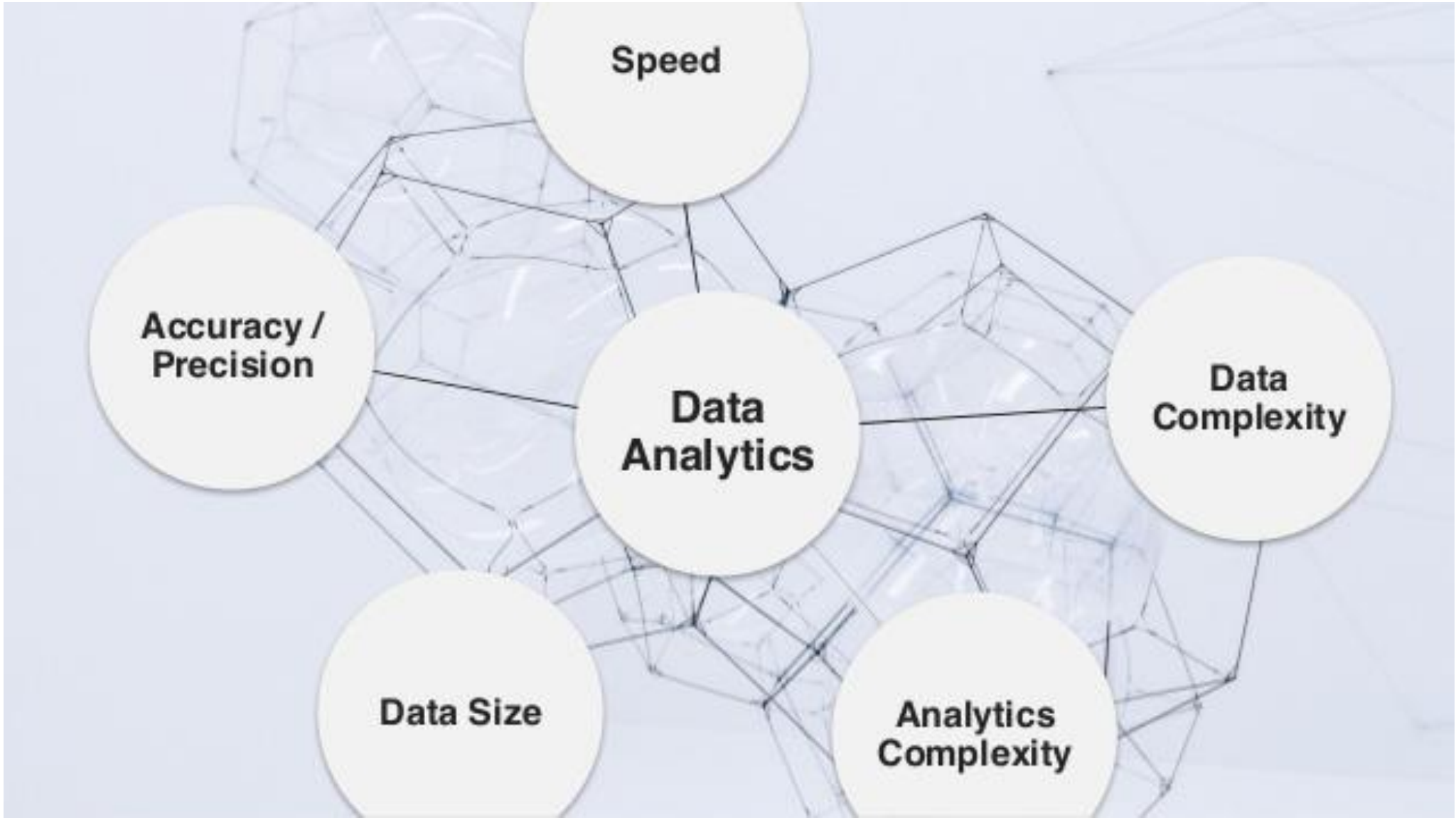
2. **필요 데이터 존재 여부의 불확실성**

   :문제 해결을 위해 필요한 데이터의 집합이 모두 존재하지 않을 경우, 그 데이터의 수집을 어떻게 할 것인지 또는 그 데이터를 다른 데이터로 대체할 것인지 등에 대한 사용자와 분석자간의 반복적이고 순환적인 협의 과정이 필요.

3. **데이터 사용 목적의 가변성**

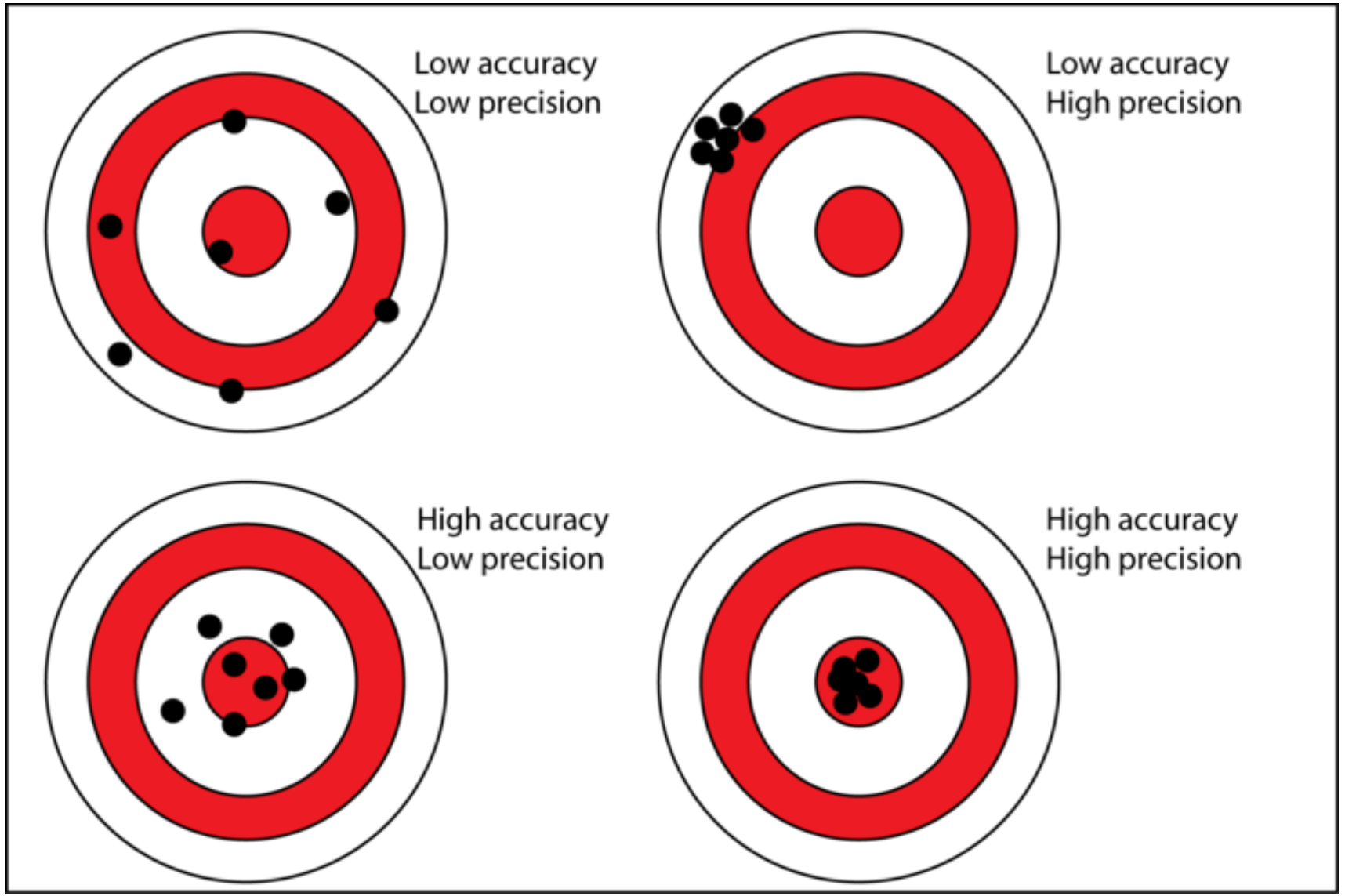
   : 조직에서 보유 중인 데이터라 하더라고 기존의 데이터 정의를 재검토하여 데이터의 사용 목적과 범위를 확대

분석 프로젝트 관리 방안



출처 : https://needjarvis.tistory.com/533

분석 프로젝트 관리 방안



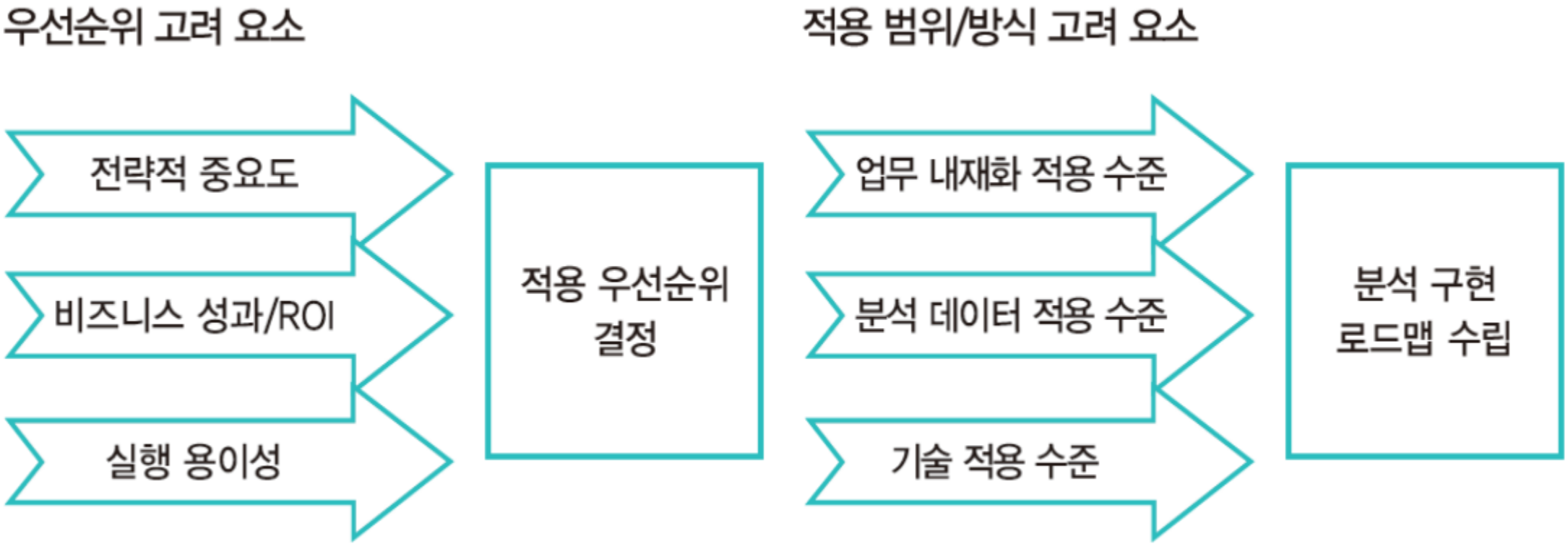
출처 : https://needjarvis.tistory.com/533

분석 프로젝트 관리 방안

관리 영역	주요 프로세스
통합관리 (Integration Management)	프로젝트 헌장 개발, 프로젝트 관리 계획 수립, 프로젝트 실행 지시 및 관리, 프로젝트 작업 감시 및 통제, 통합 변경 통제, 프로젝트 종료 관리 등
범위관리 (Scope Management)	프로젝트 범위 계획, 범위 정의, 작업분류체계(WBS) 작성, 범위 검증, 범위 통제 관리 등
일정관리 (Time Management)	작업 정의, 작업 순서 배열, 작업별 자원 산정, 작업 기간 산정, 일정 개발, 일정 통제 등
비용관리 (Cost Management)	자원 계획, 비용 산정, 비용 예산 및 비용 통제 등
품질관리 (Quality Management)	품질 계획, 품질 보증, 품질 관리 등
인적자원관리 (Human Resource Management)	조직 계획, 인적 자원 획득, 프로젝트 팀 확보, 프로젝트 팀 개발, 프로젝트 팀 관리 등
위험관리 (Risk Management)	위험 관리 계획, 위험 식별, 정성적 위험 분석, 정량적 위험 분석, 위험 대응 계획, 위험 감시 통제 등
의사소통관리 (Communication Management)	의사소통 계획, 정보 배포, 진척 관리, 종료 절차 등
조달관리 (Procurement Management)	획득계획, 공급자 유치 계획, 공급자 선정, 계약 관리, 계약 종료 등

분석 마스터 플랜

- 데이터 기반 구축을 위해서 분석 과제를 대상으로 전략적 중요도, 비즈니스 성과 및 ROI, 분석과제의 실용 용이성 등 다양한 기준을 고려해 적용 우선순위를 설정



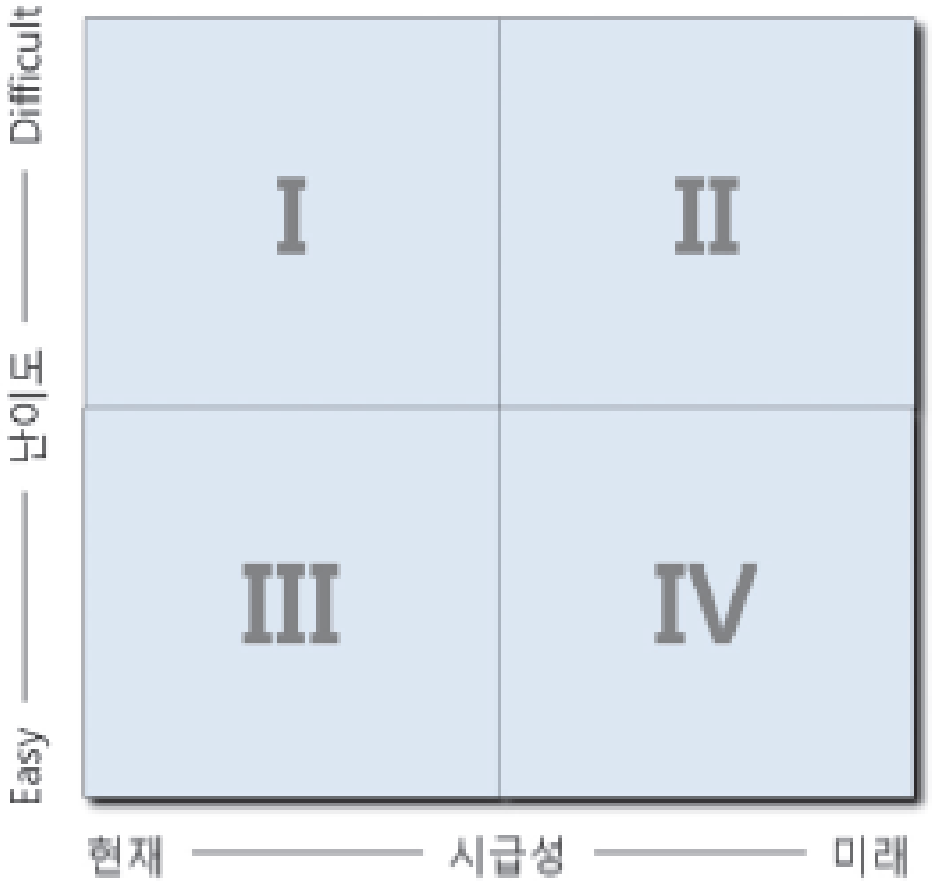
마스터플랜 수립 개요 (★기출)

출처 : https://acdongpgm.tistory.com/80

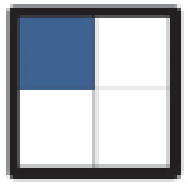
분석 마스터 플랜

출처 : https://acdongpgm.tistory.com/80

분석 마스터 플랜

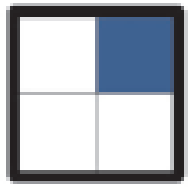


I



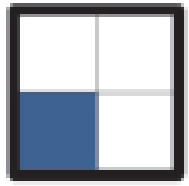
•전략적 중요도가 높아 경영에 미치는 영향이 크므로 현재 시급하게 추진이 필요함
•난이도가 높아 현재 수준에서 과제를 바로 적용하기에 어려움

II



•현재 시점에서는 전략적 중요도가 높지 않지만 중장기적 관점에서는 반드시 추진되어야 함
•분석과제를 바로 적용하기에는 난이도가 높음

III

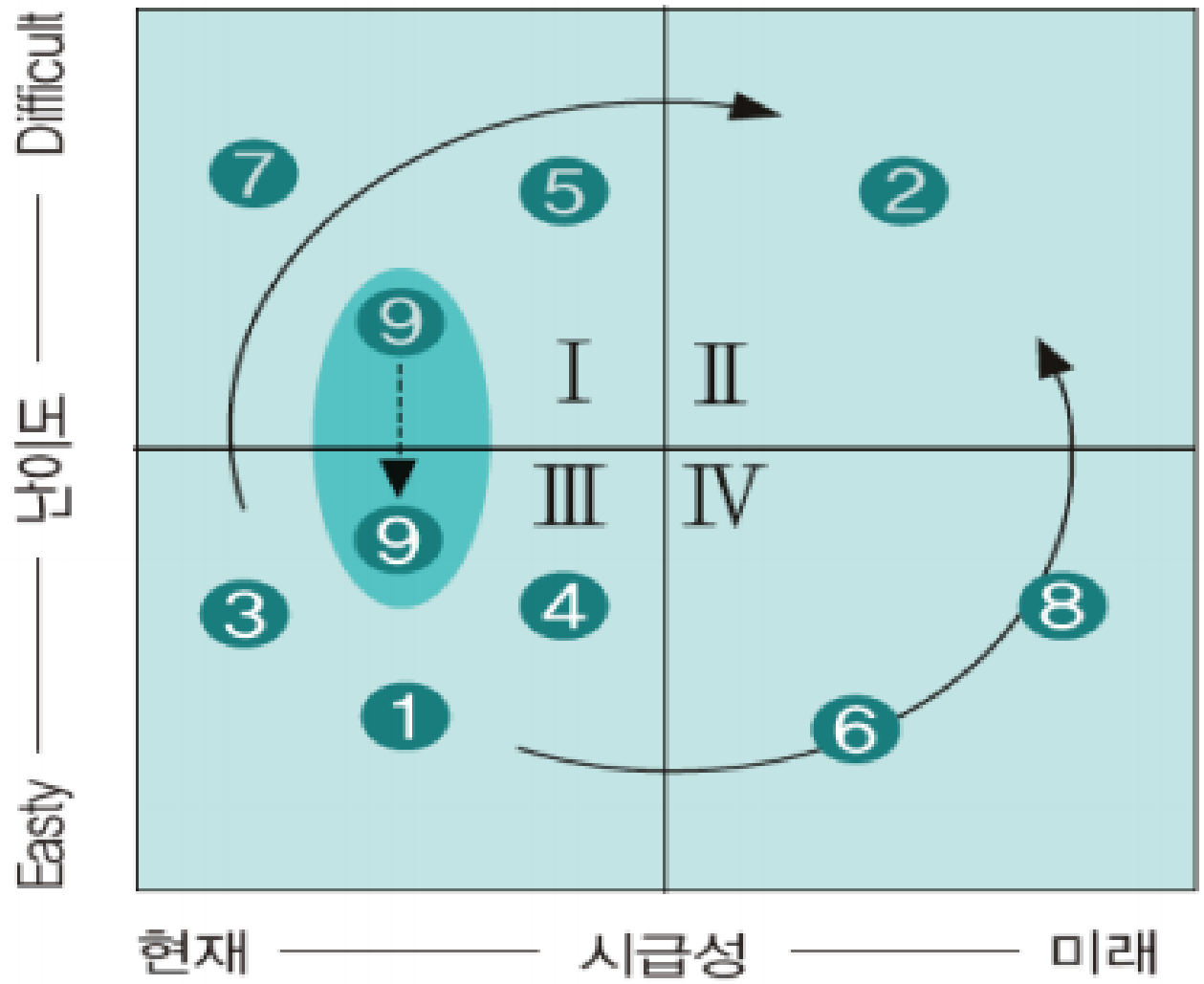


•전략적 중요도가 높아 현재 시점에 전략적 가치를 두고 있음
•과제 추진의 난이도가 어렵지 않아 우선적으로 바로 적용 가능할 필요성이 있음

IV

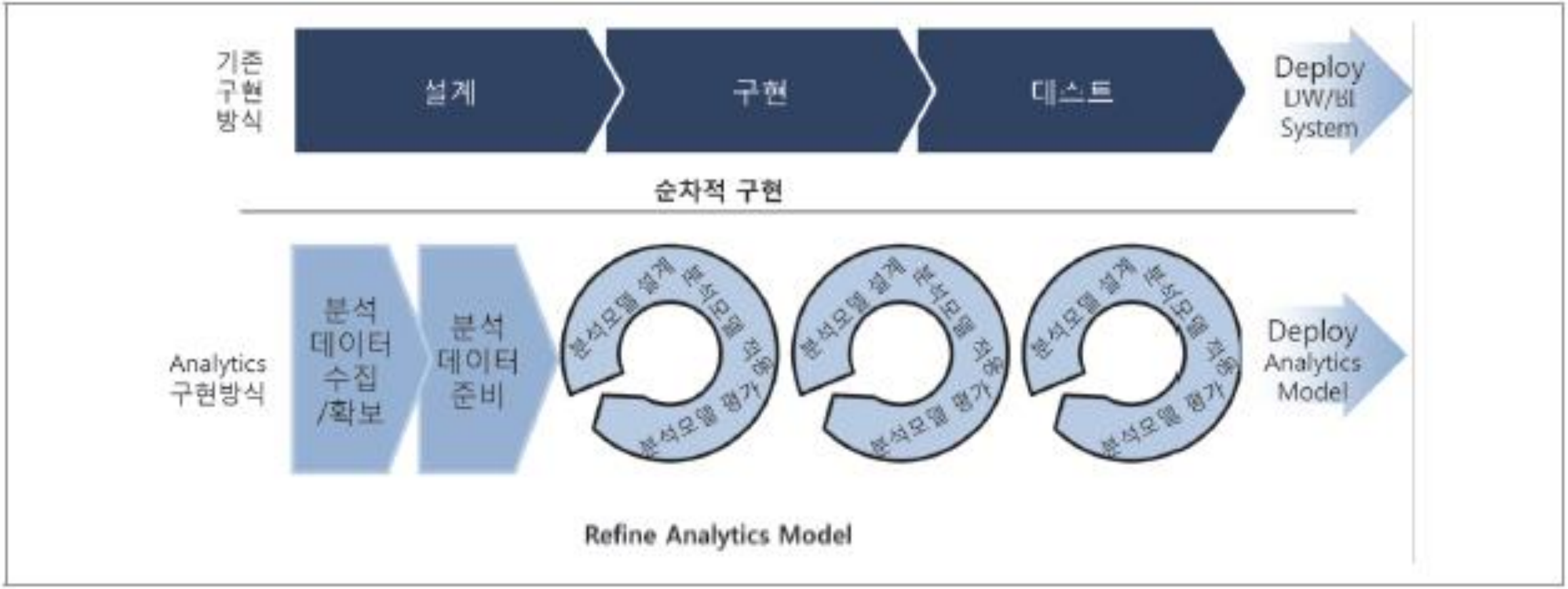
•전략적 중요도가 높지 않아 중장기적 관점에서 과제 추진이 바람직함
•과제를 바로 적용하는 것은 어렵지 않음

분석 과제 우선순위 선정 조정



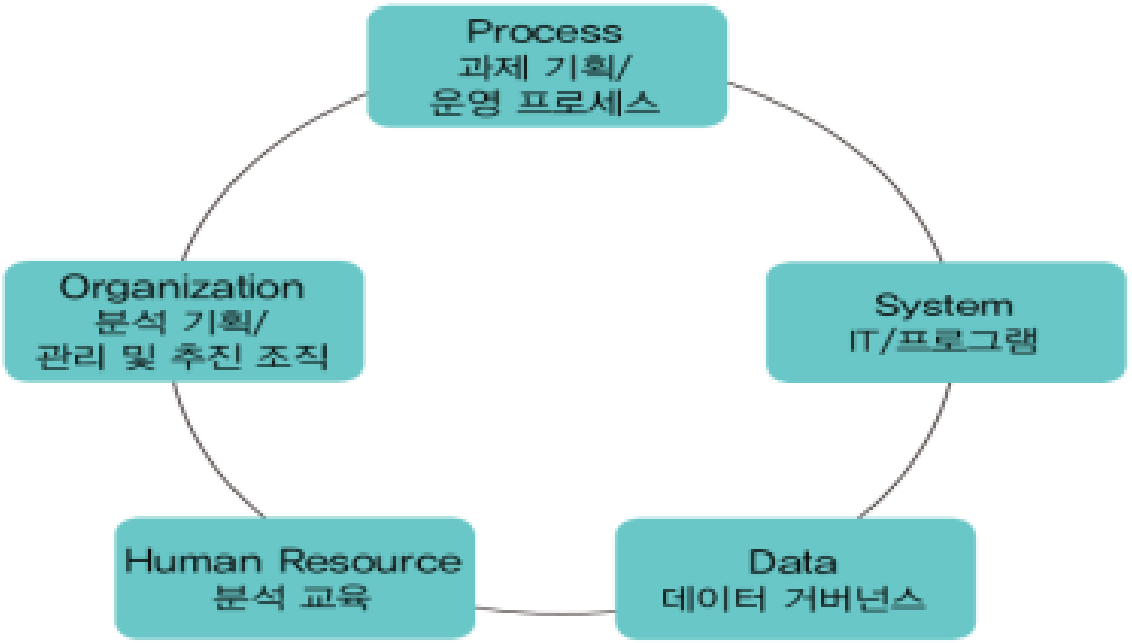
출처 : https://acdongpgm.tistory.com/80

이행계획 수립


기존 구현 방식
설계
구현
테스트
Deploy DW/BI System
순차적 구현
Analytics 구현방식
분석 데이터 수집 /확보
분석 데이터 준비
분석모델 설계
분석모델 적용
분석모델 평가
Deploy Analytics Model
Refine Analytics Model

분석 거버넌스 체계 수립

- 기업에서 데이터를 이용한 의사결정이 강조될수록 데이터 분석과 활용을 위한 체계적인 관리가 중요해 짐.



분석 거버넌스 체계 구성 요소

출처 : https://acdongpgm.tistory.com/80

분석 준비도

분석 업무 파악	인력 및 조직	분석 기법
• 발생한 사실 분석 업무 • 예측 분석 업무 • 시뮬레이션 분석 업무 • 최적화 분석 업무 • 분석 업무 정기적 개선	• 분석 전문가 직무 존재 • 분석 전문가 교육 훈련 프로그램 • 관리자들의 기본적 분석 능력 • 전사 분석업무 총괄 조직 존재 • 경영진 분석 업무 이해 능력	• 업무별 적합한 분석기법 사용 • 분석 업무 도입 방법론 • 분석기법 라이브러리 • 분석기법 효과성 평가 • 분석기법 정기적 개선

분석 데이터	분석 문화	IT 인프라
• 분석업무를 위한 데이터 충분성 • 분석업무를 위한 데이터 신뢰성 • 분석업무를 위한 데이터 적시성 • 비구조적 데이터 관리 • 외부 데이터 활용 체계 • 기준데이터 관리(MDM)	• 사실에 근거한 의사결정 • 관리자의 데이터 중시 • 회의 등에서 데이터 활용 • 경영진의 직관보다 데이터 • 데이터 공유 및 협업 문화	• 운영시스템 데이터 통합 • EAI, ETL 등 데이터유통체계 • 분석 전용 서버 및 스토리지 • 빅데이터 분석 환경 • 통계 분석 환경 • 비쥬얼 분석 환경

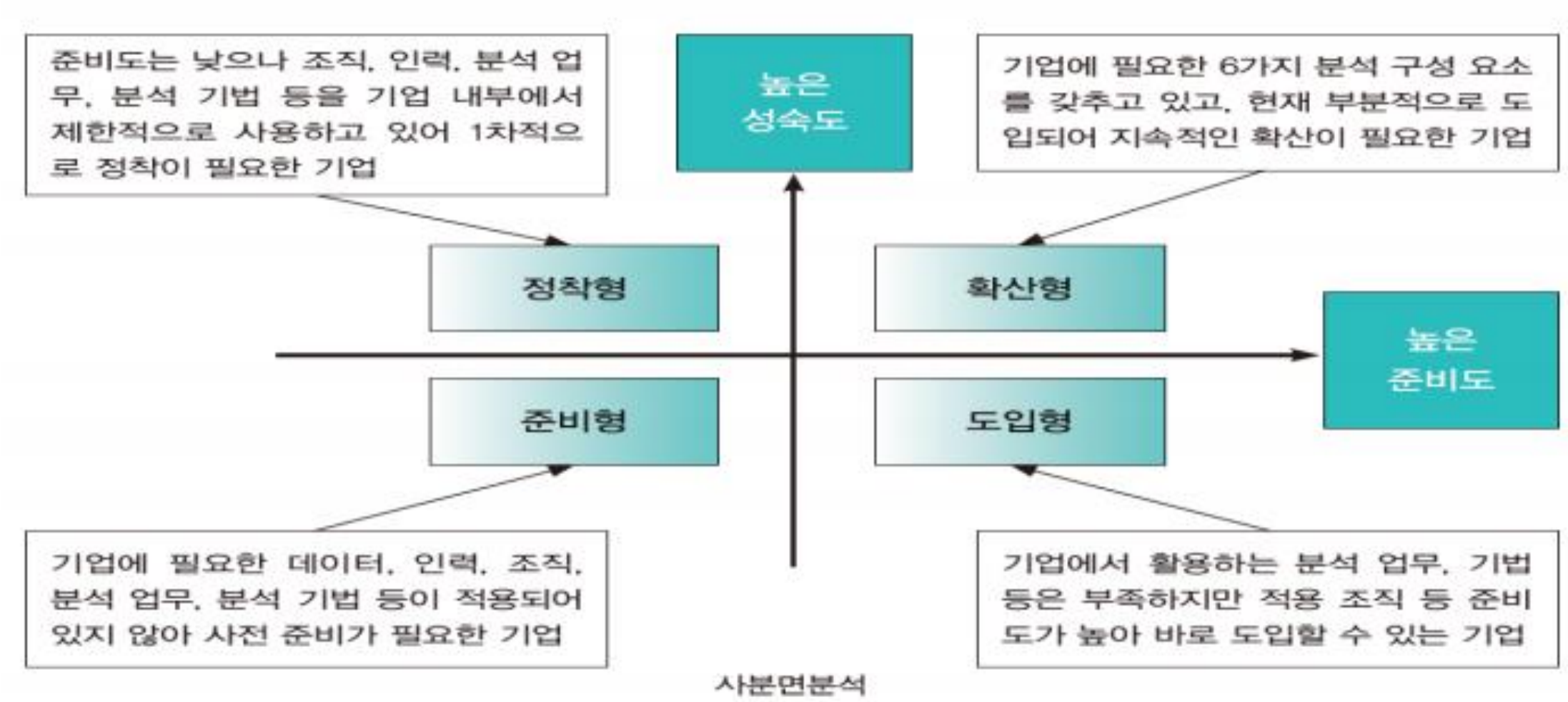
출처 : https://acdongpgm.tistory.com/80

분석 성숙도

단계	도입 단계	활용 단계	확산 단계	최적화 단계
설명	분석을 시작하여 환경과 시스템을 구축	분석 결과를 실제 업무에 적용	전사 차원에서 분석을 관리하고 공유	분석을 진화시켜서 혁신 및 성과 향상에 기여
비즈니스 부문	√ 실적분석 및 통계 √ 정기보고 수행 √ 운영 데이터 기반	√ 미래 결과 예측 √ 시뮬레이션 √ 운영 데이터기반	√ 전사 성과 실시간 분석 √ 프로세스혁신 3.0 √ 분석규칙 관리 √ 이벤트 관리	√ 외부환경 분석 활용 √ 최적화 업무 적용 √ 실시간 분석 √ 비즈니스 모델 진화
조직·역량 부문	√ 일부 부서에서 수행 √ 담당자 역량에 의존	√ 전문 담당부서에서 수행 √ 분석 기법 도입 √ 관리자가 분석 수행	√ 전사 모든 부서 수행 √ 분석 COE 조직 운영 √ 데이터 사이언티스트 확보	√ 데이터 사이언스 그룹 √ 경영진 분석 활용 √ 전략 연계
IT 부문	√ 데이터 웨어하우스 √ 데이터 마트 √ ETL/EAI √ OLAP	√ 실시간 대시보드 √ 통계분석 환경	√ 빅데이터 관리 환경 √ 시뮬레이션·최적화 √ 비주얼 분석 √ 분석 전용 서버	√ 분석 협업 환경 √ 분석 Sandbox √ 프로세스 내재화 √ 빅데이터 분석

출처 : https://acdongpgm.tistory.com/80

분석 진단 결과



사분면분석

출처 : https://acdongpgm.tistory.com/80

연습문제

1. 데이터에 대한 설명으로 가장 부적절한 것은 무엇인가?

① 데이터를 단순한 객체로서 가치 뿐만 아니라, 다른 객체와의 상호관계 속에서 가치를 갖는 것으로 설명할 수 있다.

② 데이터는 그 형태에 따라 언어,문자 등으로 기술되는 정량적 데이터와 수치, 기호, 도형으로 표시되는 정성적 데이터로 구분된다.

③ 설문조사와 주관식 응답, 트위터나 페이스북, 블로그 등에 올린 글 등과 같은 정성 데이터의 경우 그 형태와 형식이 정해져 있지 않아, 비정형 데이터라고 한다.

④ 지역별 온도, 풍속, 강수량과 같이 수치로 명확하게 표현되는 데이터를 정량 데이터라고 한다.


Thank you.
ADSP / 류영표 강사
ryp1662@gmail.com