

Big Data

데이터 분석 기획

류영표 강사

youngpyoryu@dongguk.edu

Copyright © “Youngpyo Ryu” All Rights Reserved.

This document was created for the exclusive use of “Youngpyo Ryu”.

It must not be passed on to third parties except with the explicit prior consent of “Youngpyo Ryu”.



류영표

Youngpyo Ryu

現 동국대학교 수학과/응용수학 석사수료

現 SD아카데미 국비과정 강사

現 Upstage AI X 네이버 부스트캠프 멘토

前 메가 IT아카데미(파이썬, 빅데이터) 강사

한국파스퇴르연구소 Image Mining 인턴(Deep learning)

前 (주)셈웨어(수학 콘텐츠, 데이터 분석 개발 및 연구인턴)

강의 경력

- 현대자동차 연구원 강의 (인공지능/머신러닝/딥러닝/강화학습)
- 딥러닝 집중 교육과정 강사
- (재)월튼블록체인 6일 과정 (파이썬기초, 크롤링, 머신러닝)
- 서울특별시 X AI 양재허브 X 모두의연구소 (중급 NLP과정) 보조강사
- SK아카데미_HLP(임원) 1차/2차 보조강사
- (주) 모두의연구소 Aiffel 1기 퍼실리테이터(인공지능 교육)
- LG전자 / LG 인화원 보조강사
- 인공지능 자연어처리(NLP) 기업데이터 분석 전문가 양성과정 멘토
- 고려대학교 선도대학 소속 30명 딥러닝 집중 강의

주요 프로젝트 및 기타사항

- 제1회 인공지능(AI)기반 데이터사이언티스트
전문가 양성과정 최우수상 수상(Q&A 챗봇)
- 인공지능(AI)기반 데이터사이언티스트 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 기획의 방향성 도출

- 분석 기획의 정의.

- 실제 분석을 수행하기에 앞서 분석을 수행할 과제의 정의 및 의도했던 결과를 도출할 수 있

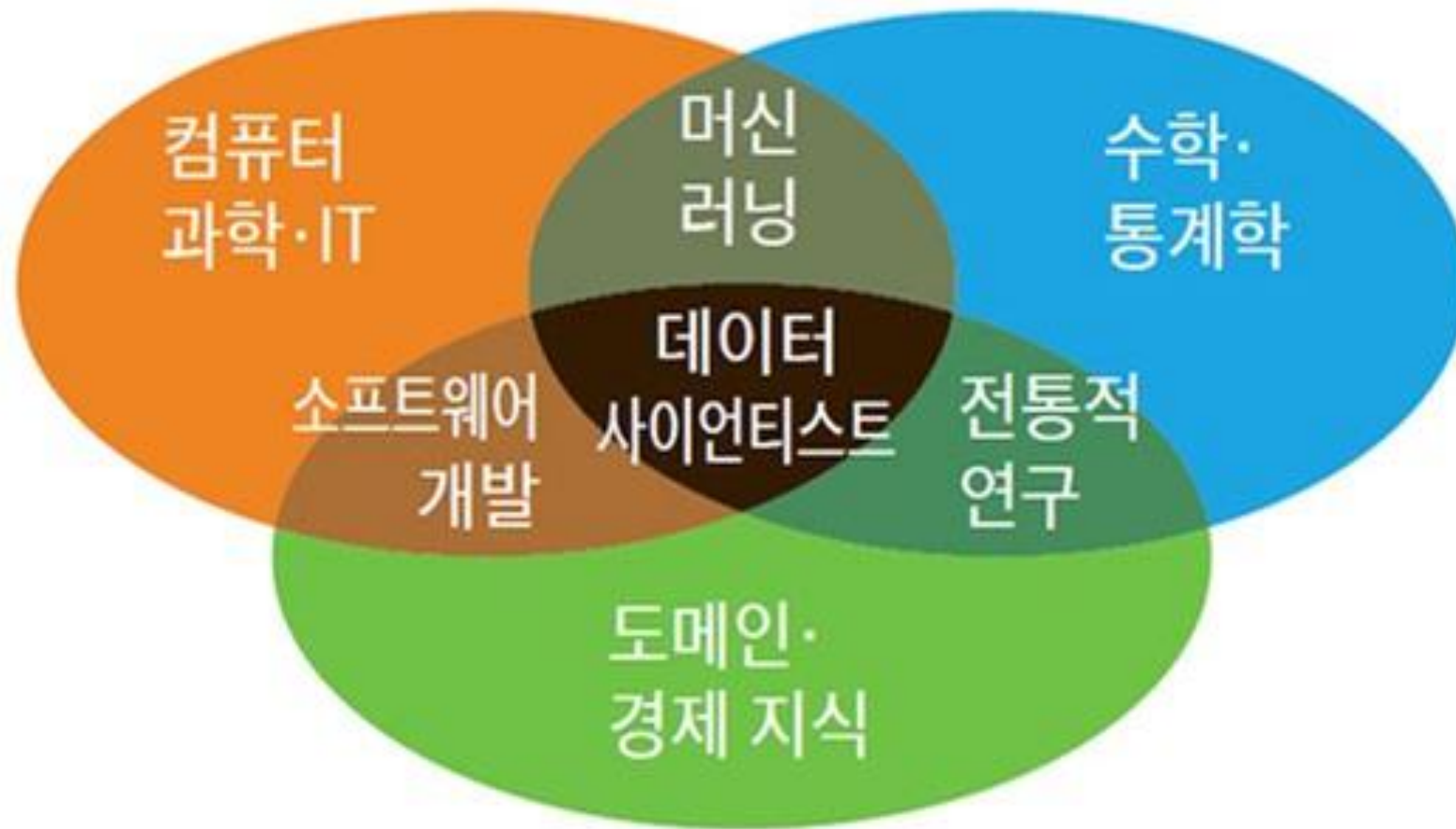
도록 하는 방안을 사전에 계획하는 일련의 작업

- 분석을 기획한다는 것은 해당 문제 영역에 대한 전문성 역량 및 수학/통계학적 지식을 활용

한 분석 역량과 분석의 도구인 데이터 및 프로그래밍 기술 역량에 대한 균형 잡힌 시각을 가

지고 방향성 및 계획을 수립해야 한다는 것을 의미

데이터 사이언티스트의 역량



분석의 대상 (What)

Known

Un-Known

분석의
방법
(How)

Known

Un-Known

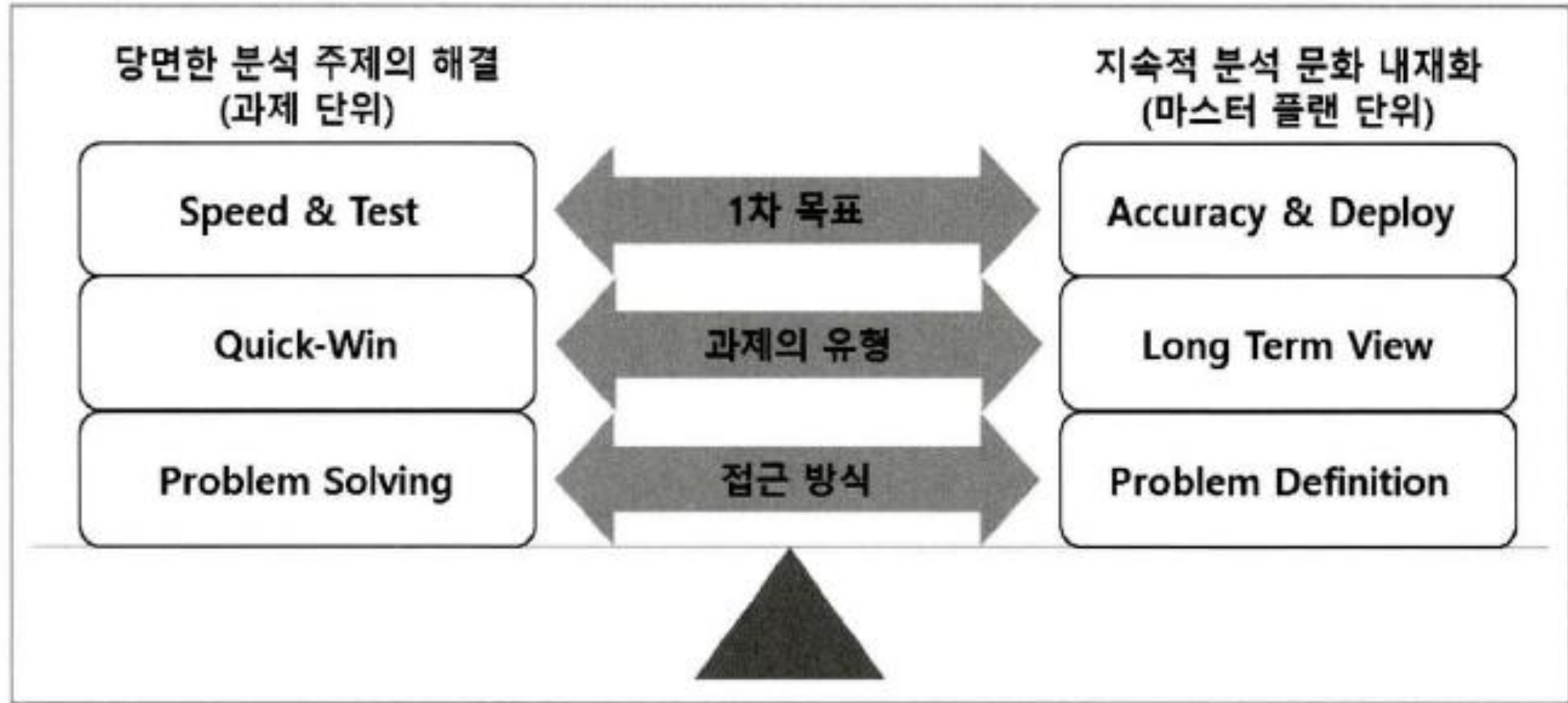
① Optimization

③ Insight

② Solution

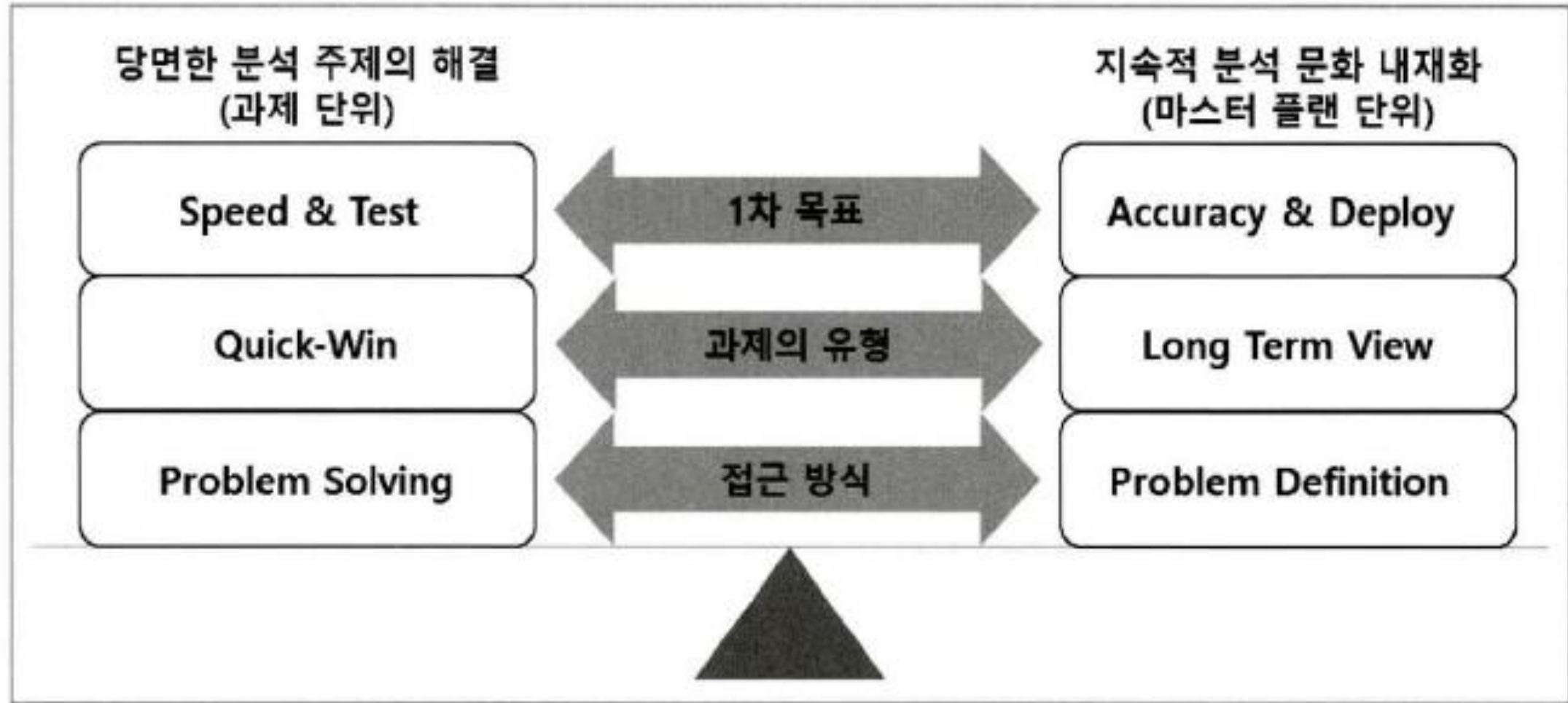
④ Discovery

목표시점 별 분석 기획 방안



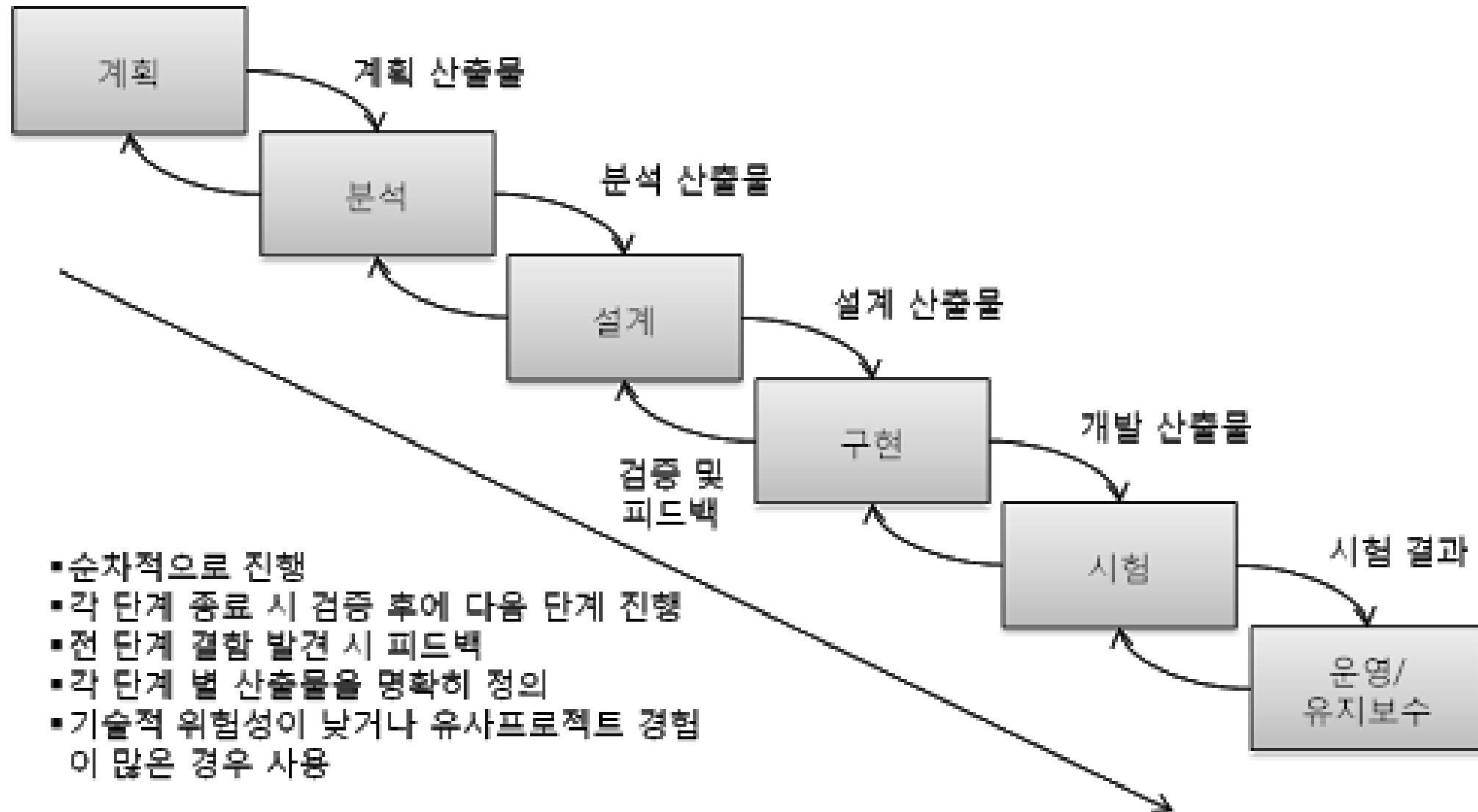
[그림 III-1-3] 목표 시점 별 분석 기획 방안

목표시점 별 분석 기획 방안

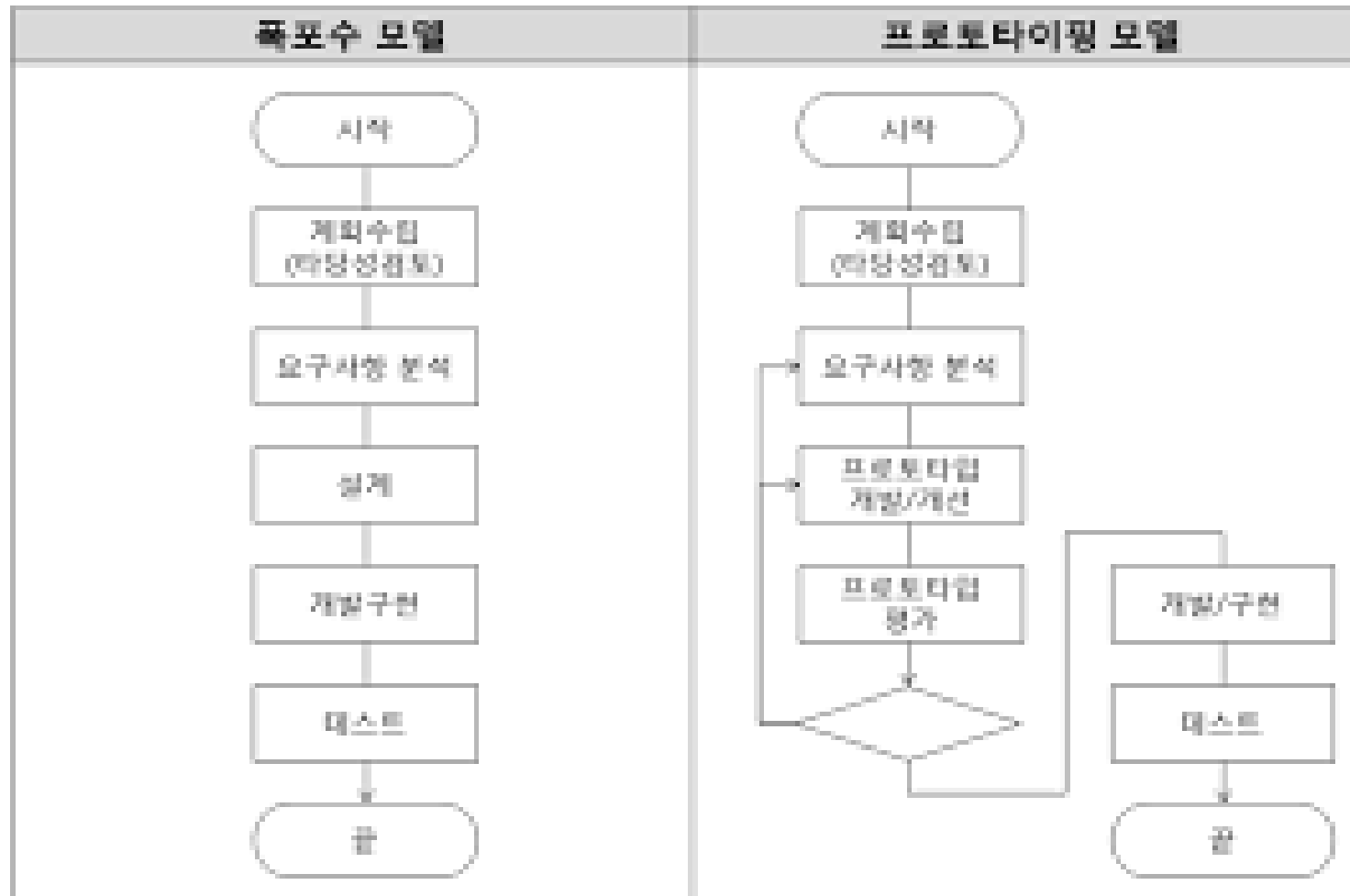


[그림 III-1-3] 목표 시점 별 분석 기획 방안

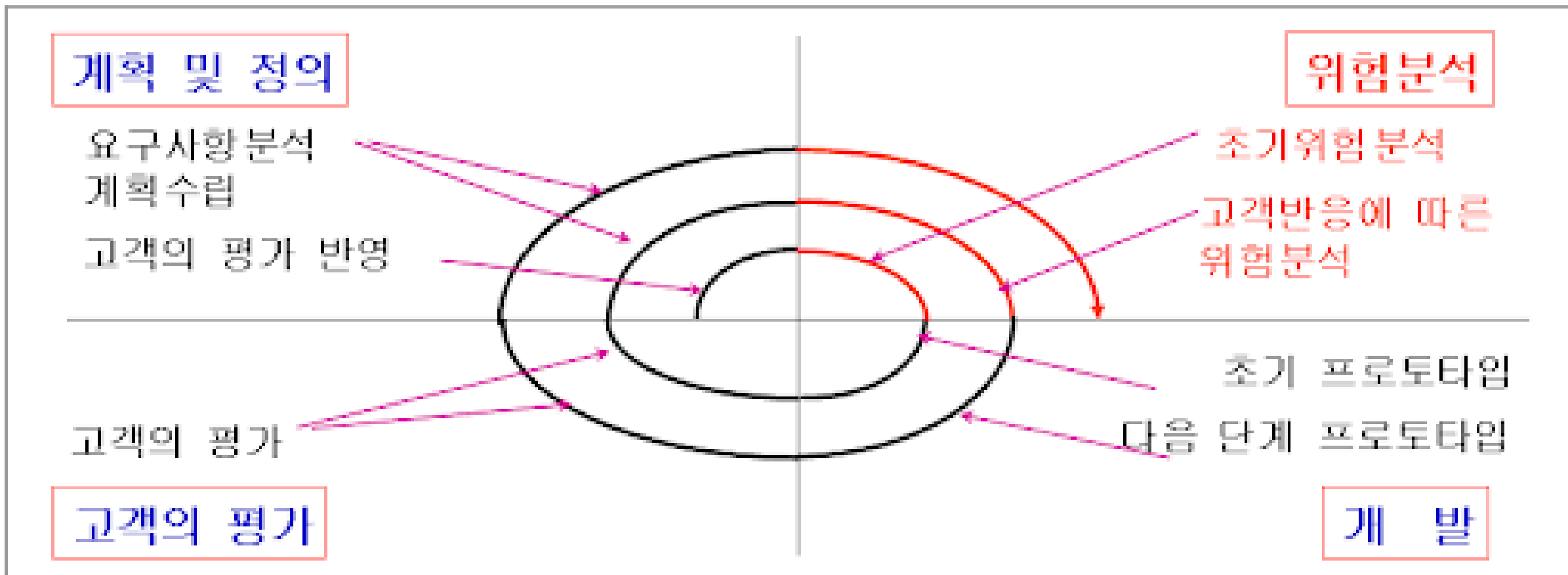
분석 방법론-폭포수 모델(Waterfall Model)



분석 방법론-프로토타입 모델(Prototype model)

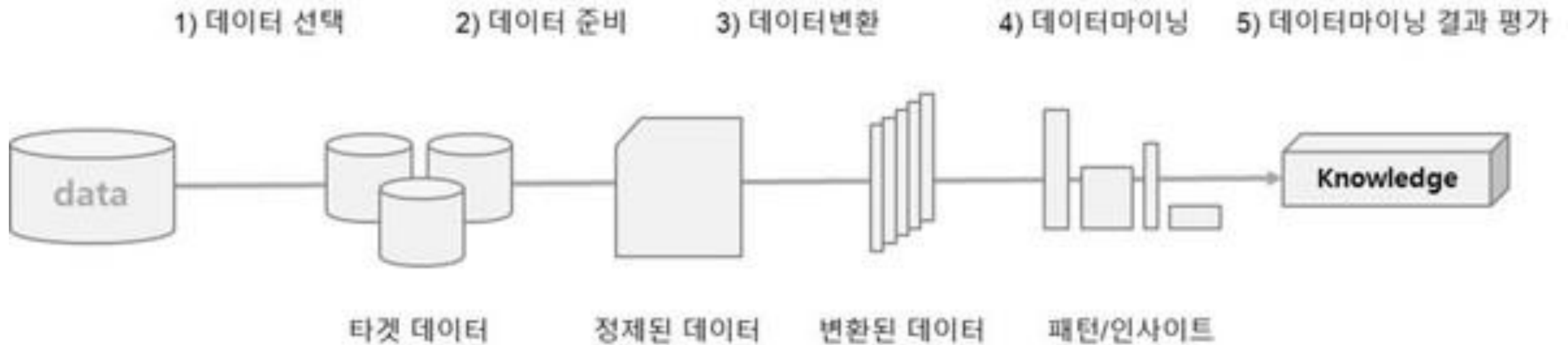


분석 방법론-나선형 모델(Spiral model)



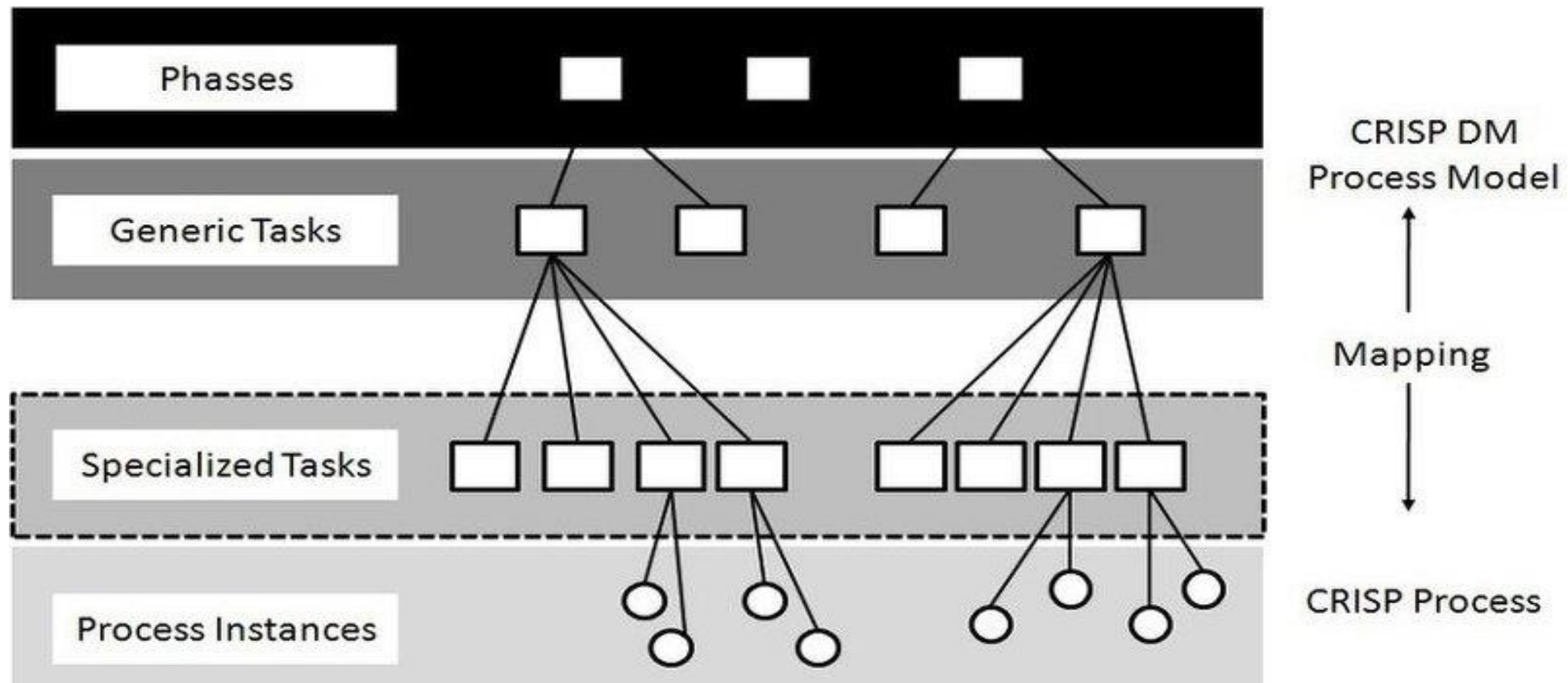
KDD 분석 방법론

- 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.



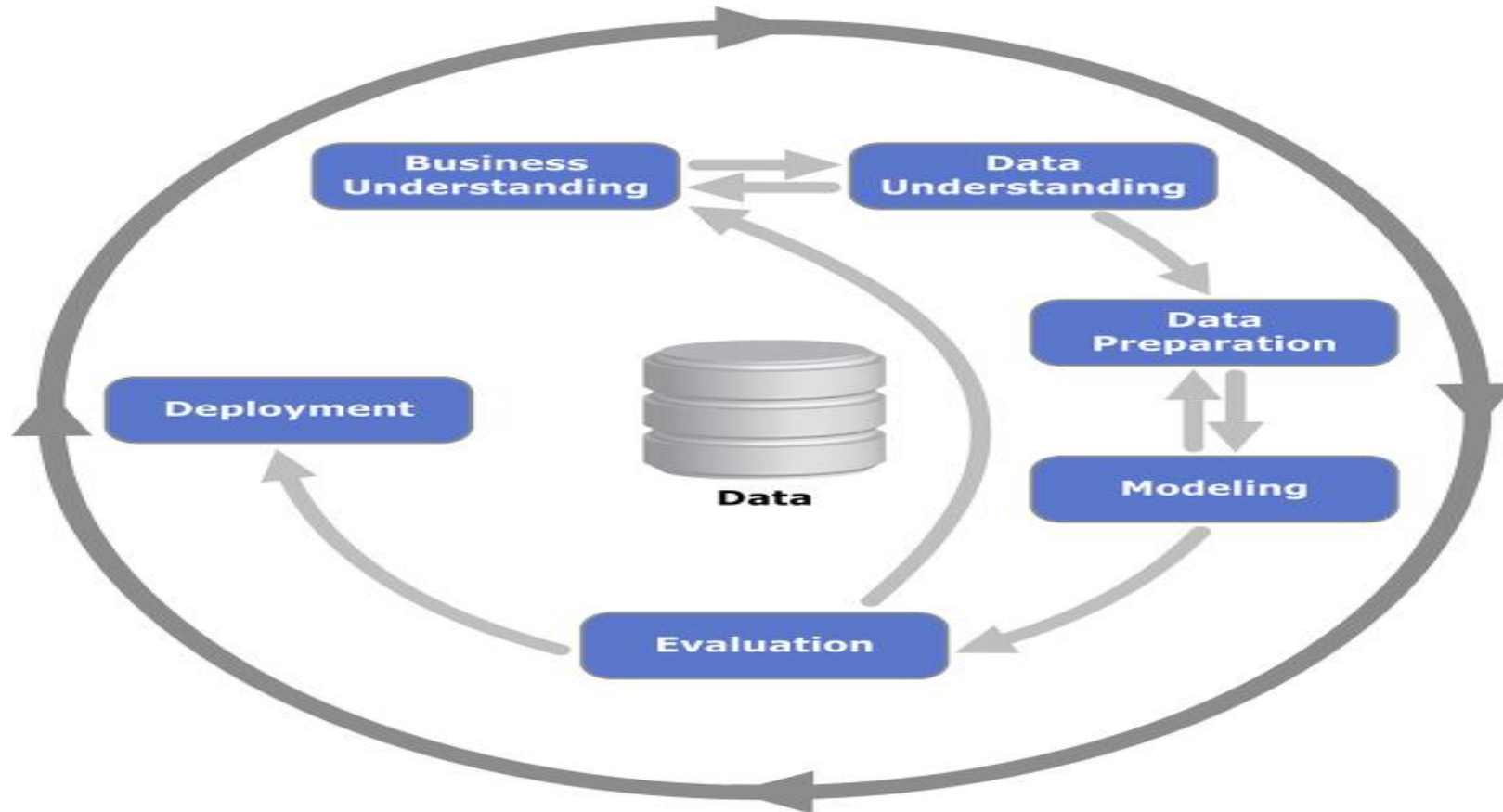
CRISP-DM 분석 방법론

- 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.

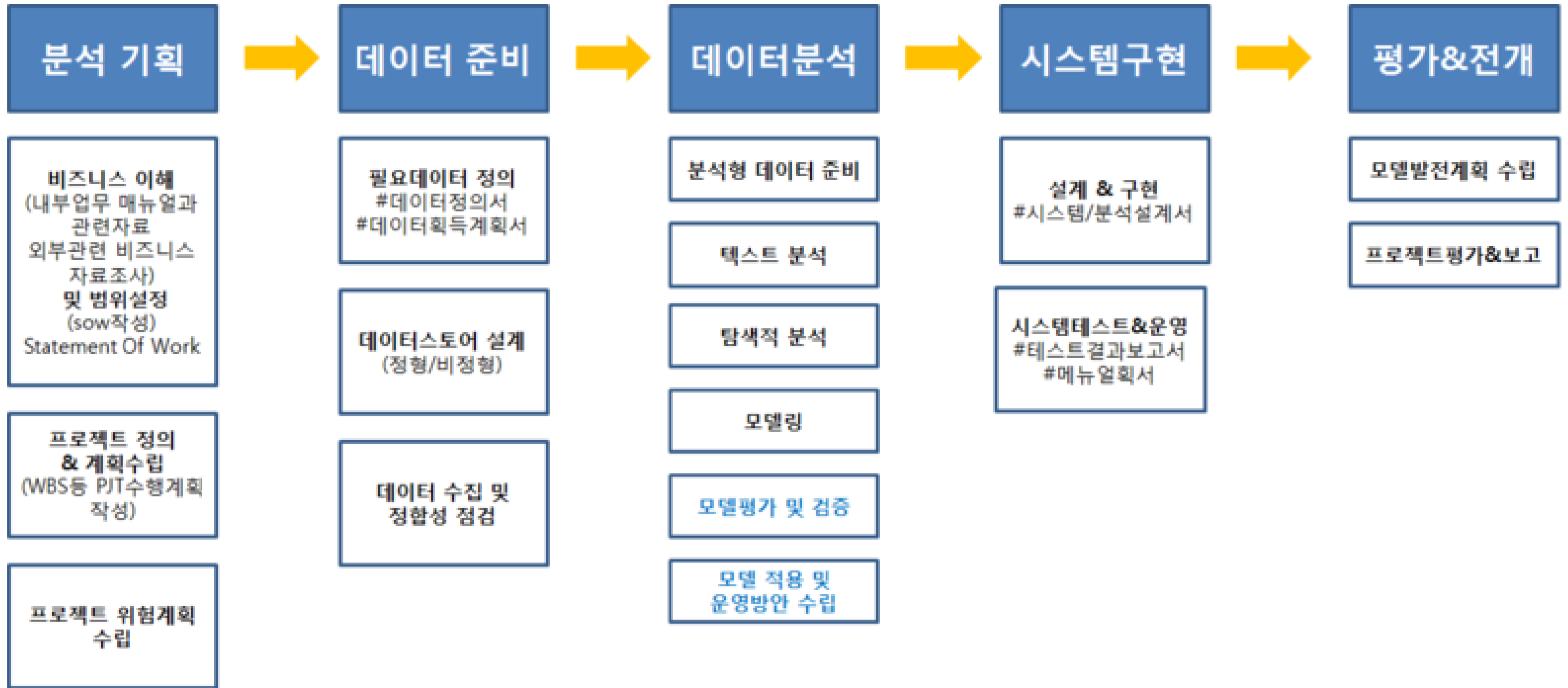


CRISP-DM의 프로세스

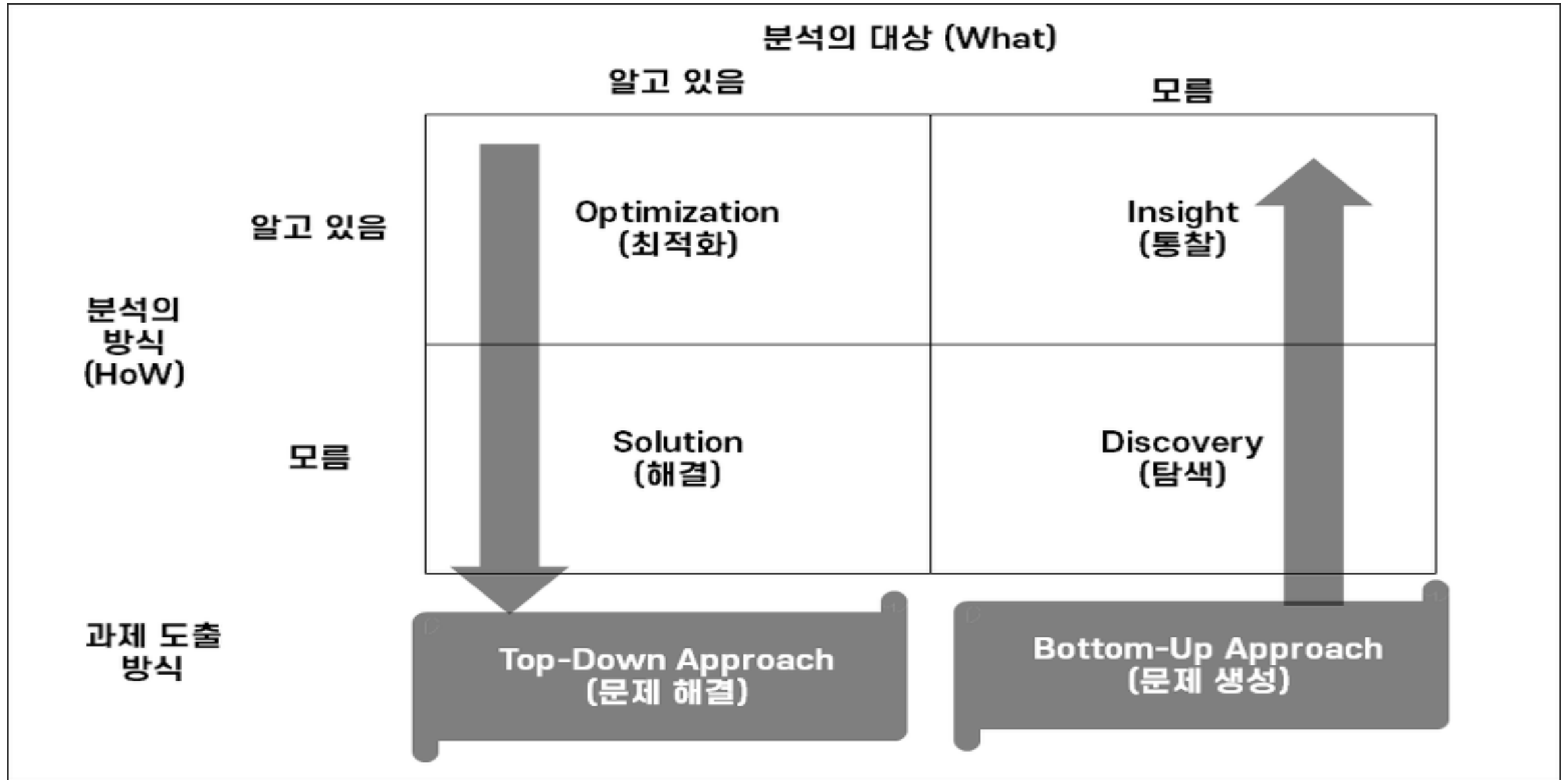
- CRISP-DM의 절차는 6단계로 구성되어 있는데 각 단계들은 순차적으로 진행되는 것이 아니라, 필요에 따라 단계 간의 반복 수행을 통해 분석의 품질을 향상시킨다.



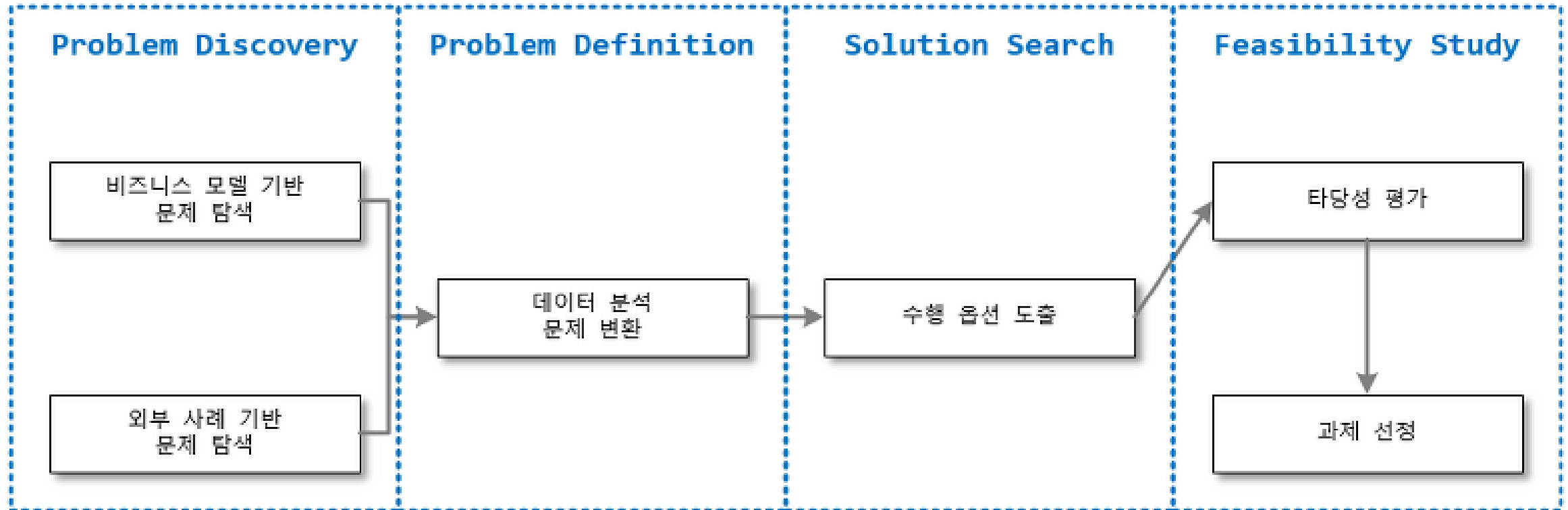
빅데이터 분석 5단계



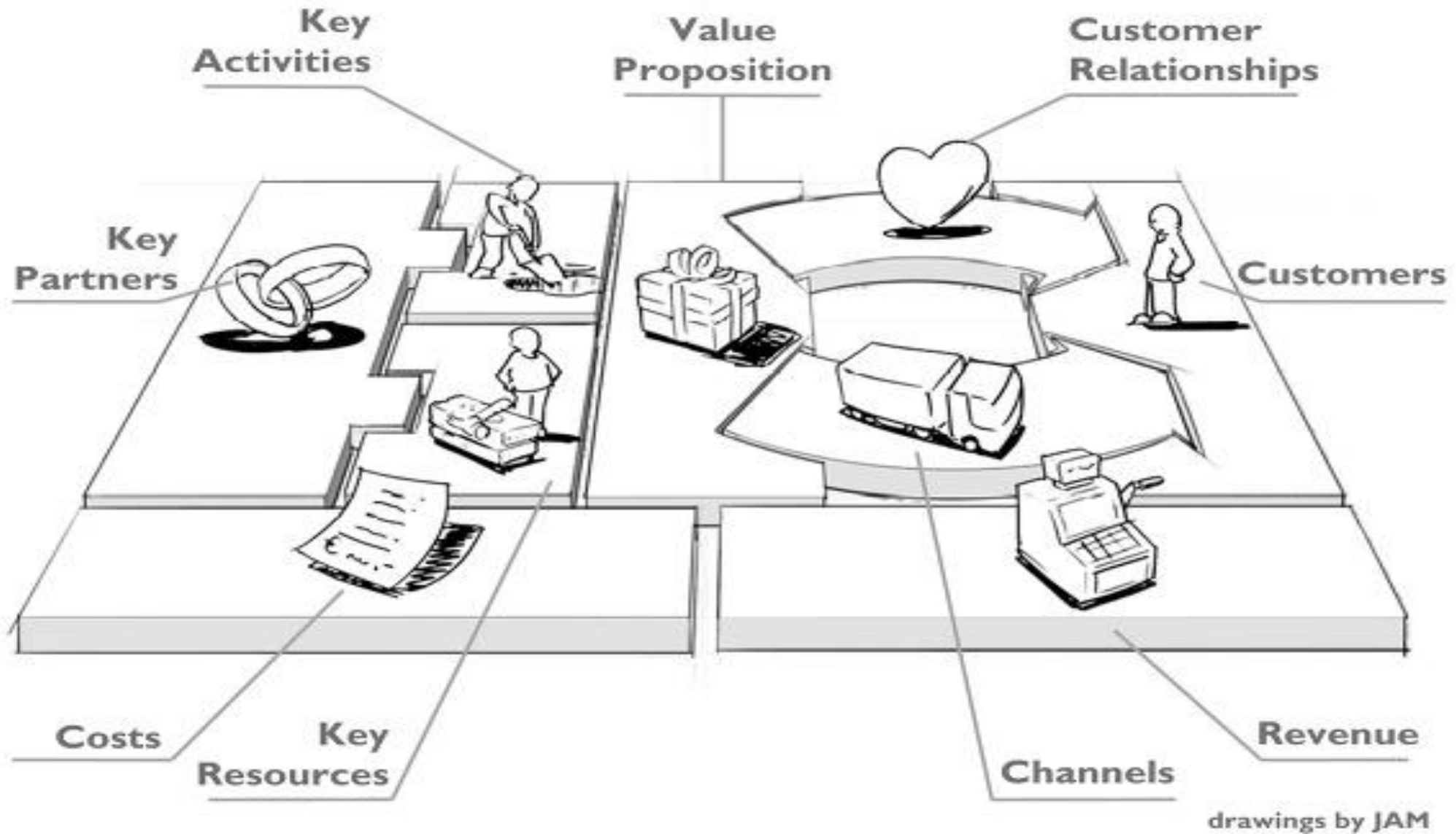
분석 과제 발굴



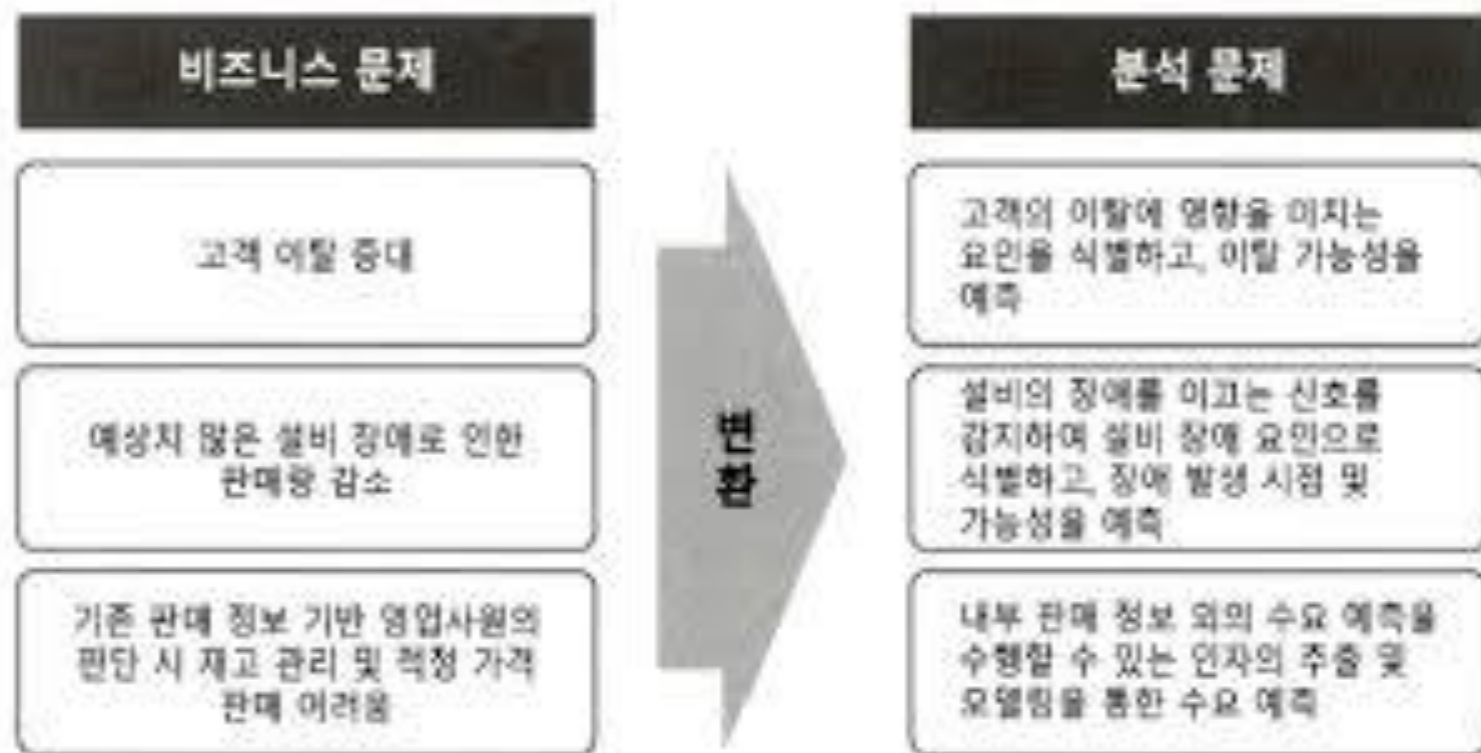
하향식 접근 방식(Top Down Approach)



비즈니스 모델 기반 문제 탐색



데이터 분석 기반 문제 탐색



데이터 분석 기반 문제 탐색

		분석 역량 (Who)	
		확보	미확보
분석 기법 및 시스템 (HoW)	기존 시스템	기존 시스템 개선 활용	교육 및 채용을 통한 역량 확보
	신규 도입	시스템 고도화	전문 업체 Sourcing

상향식 접근법(Bottom up Approach)

기업 고객 분류



상향식 접근 (Bottom-up Approach)

즉시 도입 가능
(구매 프로세스 한 단계)

거래 액수 작음
(개인, 팀 단위)

서비스 적극적 수용

구전 효과 높음
확산 속도 빠름

슬랙, 드롭박스, 트렐로



하향식 접근 (Top-down Approach)

일정 기간 소요
(구매 프로세스 다단계, 약 3-6개월 소요)

거래 액수 큼
(기업 단위)

서비스 소극적 수용

기업별 맞춤 전략 및 대응 요구
지속적 관계 유지 필요
(세일즈 역할 중요)

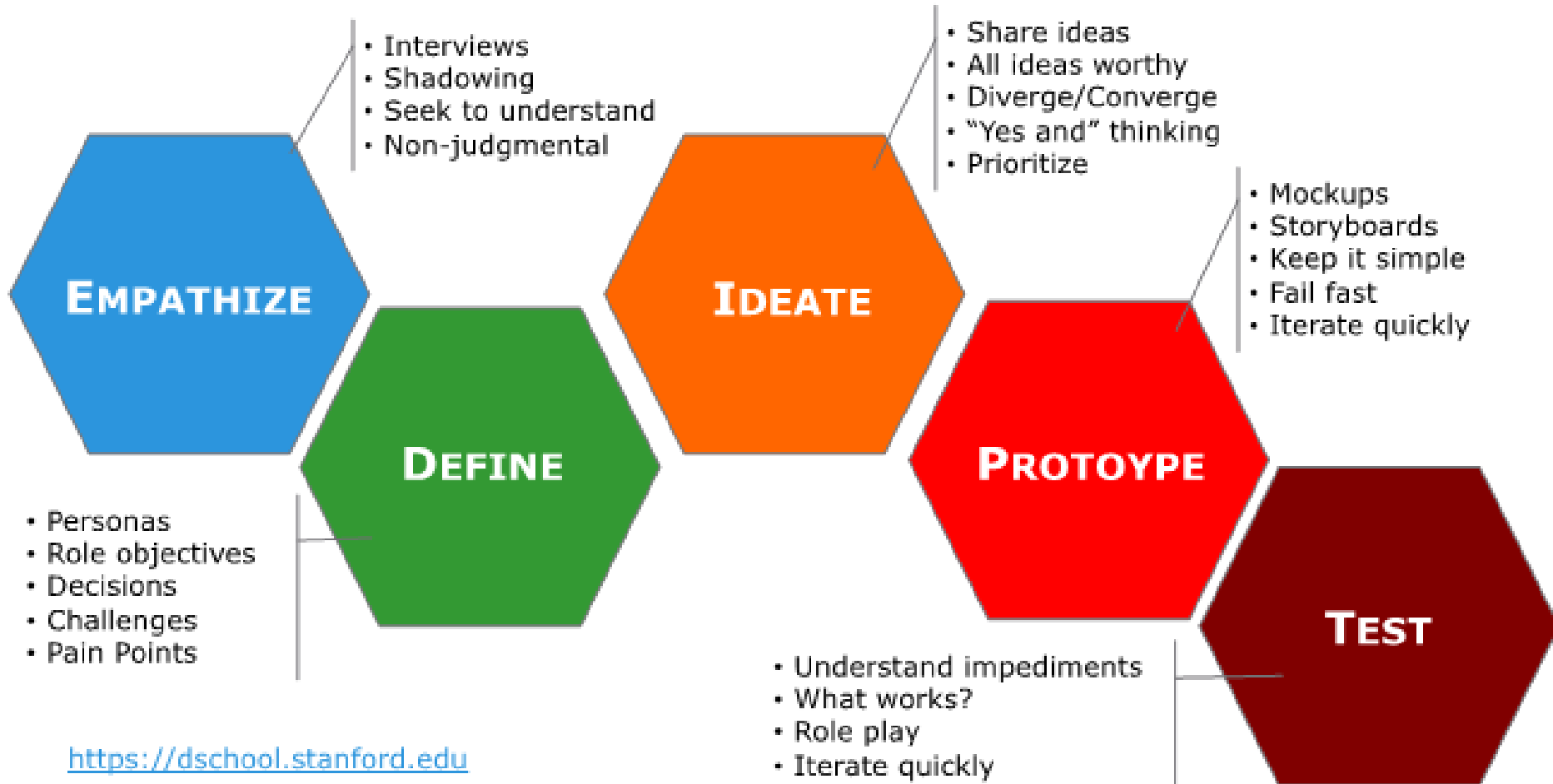
그룹웨어, ERP 시스템

하향식 접근 방법의 한계

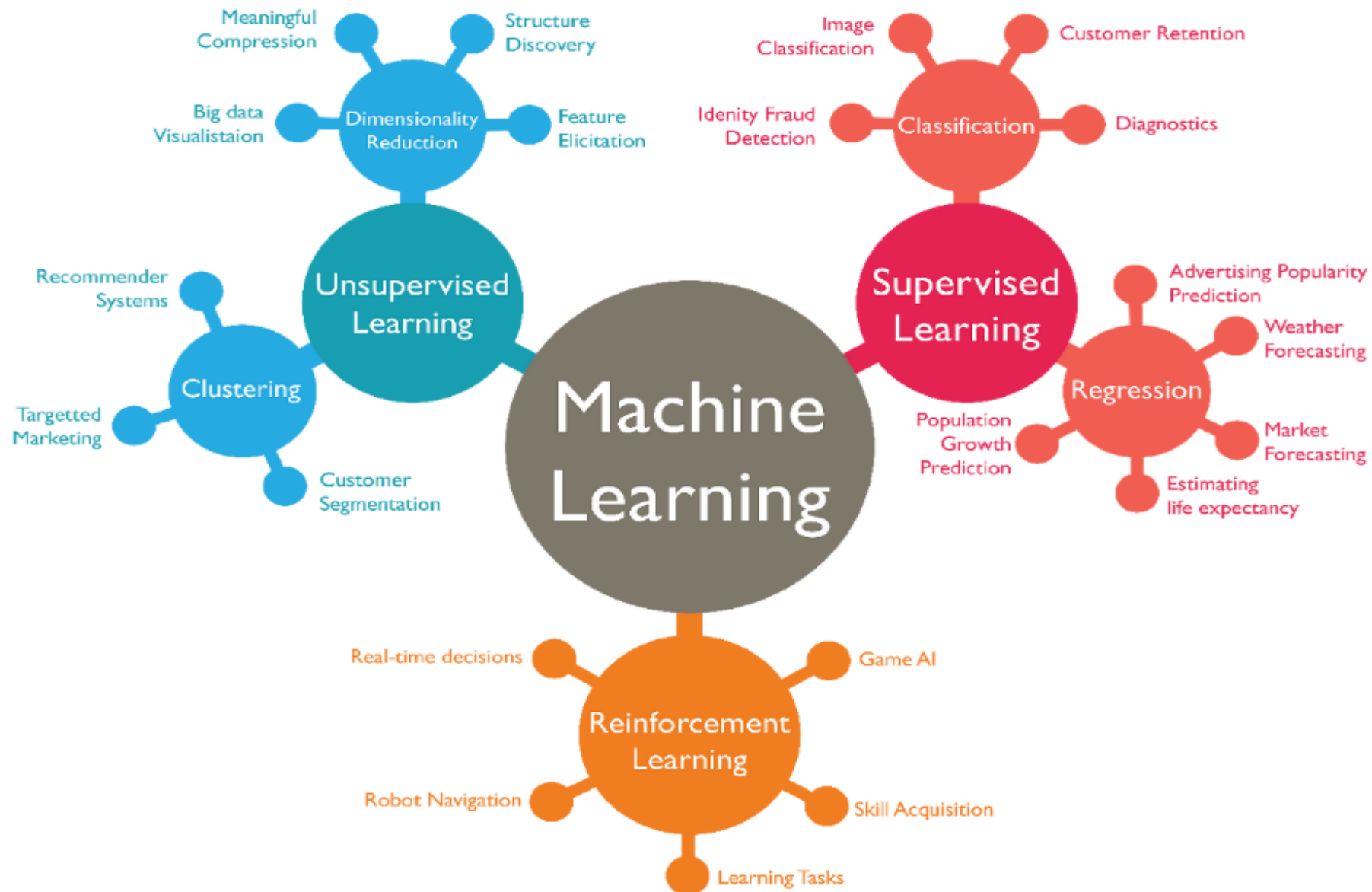
- 솔루션 도출에는 유효하지만 새로운 문제 탐색의 한계
- 논리적인 단계별 접근법 기반의 문제해결 방식은 복잡하고 다양한 환경에서 발생하는 문제에는 비적합
- 이를 해결하기 위해 스탠포드 대학의 d.school(Institute of Design at Stanford)은 디자인 사고(Design Thinking) 접근법을 통해 전통적인 분석적 사고를 극복하려고 함
- 통상적인 관점에서는 'Why'를 강조하지만, 있는 그대로 인식하는 'What' 관점 필요
- 이와 같은 점을 고려하여 d.school에서는 첫단계로 Empathize(감정이입)을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

하향식 접근 방법의 한계

Stanford d.school Design Thinking Process



Machine learning



분석 프로젝트 관리 방안



분석 프로젝트 관리 방안

**Accurate
Precise**



**Not Accurate
Precise**



**Accurate
Not Precise**



**Not Accurate
Not Precise**



분석 프로젝트 관리 방안

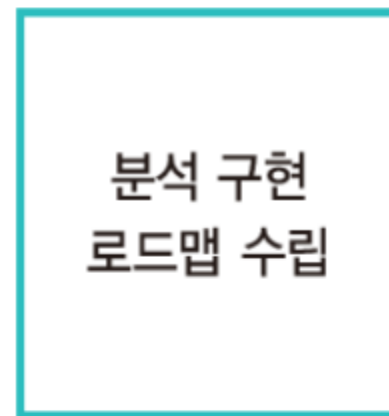
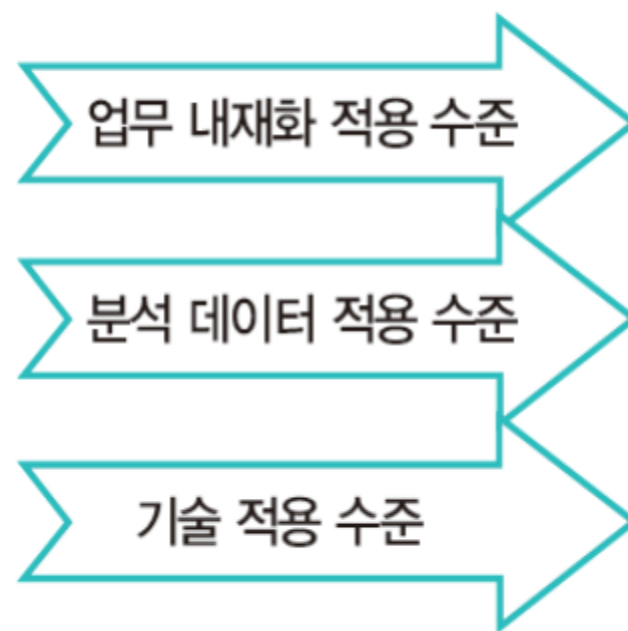
관리 영역	주요 프로세스
통합관리 (Integration Management)	프로젝트 헌장 개발, 프로젝트 관리 계획 수립, 프로젝트 실행 지시 및 관리, 프로젝트 작업 감시 및 통제, 통합 변경 통제, 프로젝트 종료 관리 등
범위관리 (Scope Management)	프로젝트 범위 계획, 범위 정의, 작업분류체계(WBS) 작성, 범위 검증, 범위 통제 관리 등
일정관리 (Time Management)	작업 정의, 작업 순서 배열, 작업별 자원 산정, 작업 기간 산정, 일정 개발, 일정 통제 등
비용관리 (Cost Management)	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

우선순위 고려 요소



적용 범위/방식 고려 요소



마스터플랜 수립 개요 (★기출)

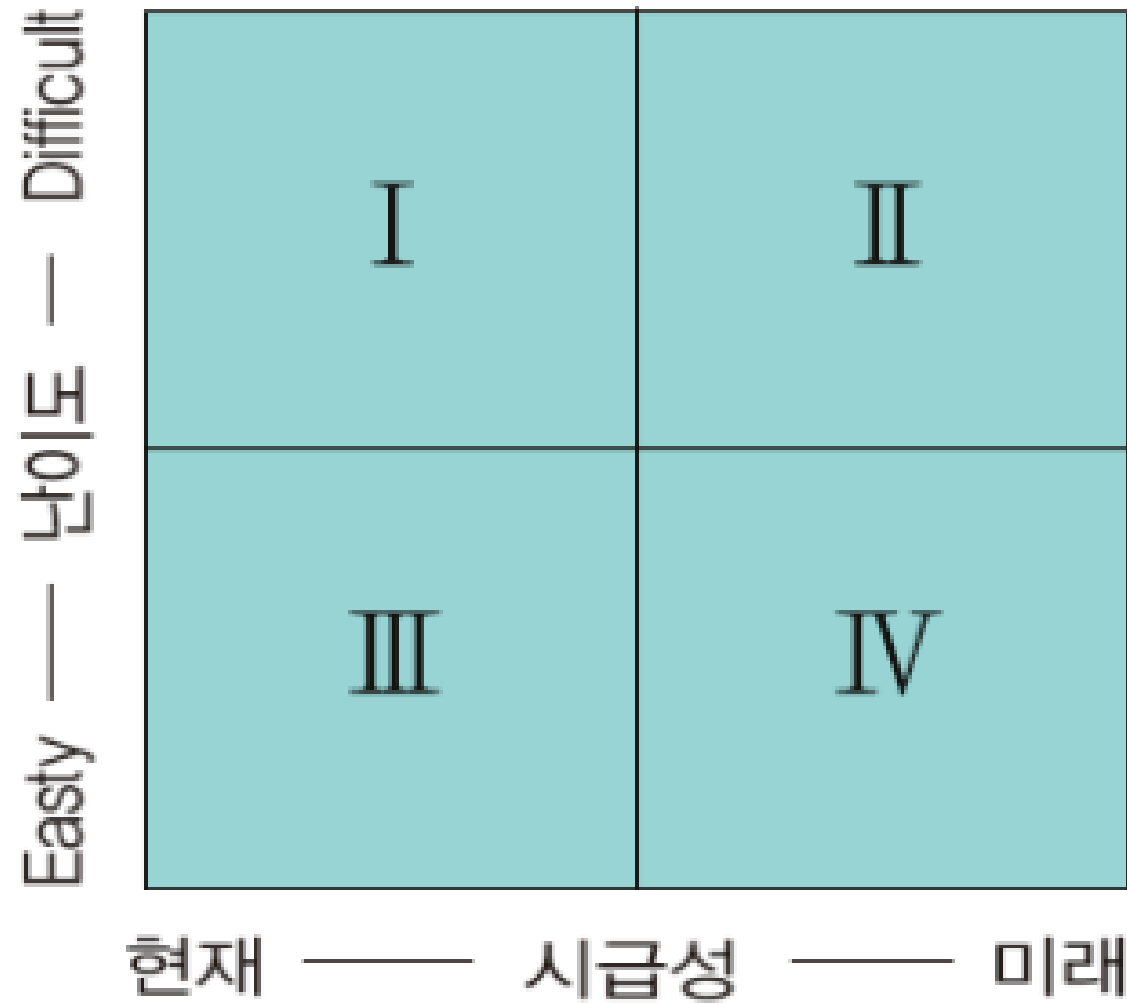
분석 마스터 플랜



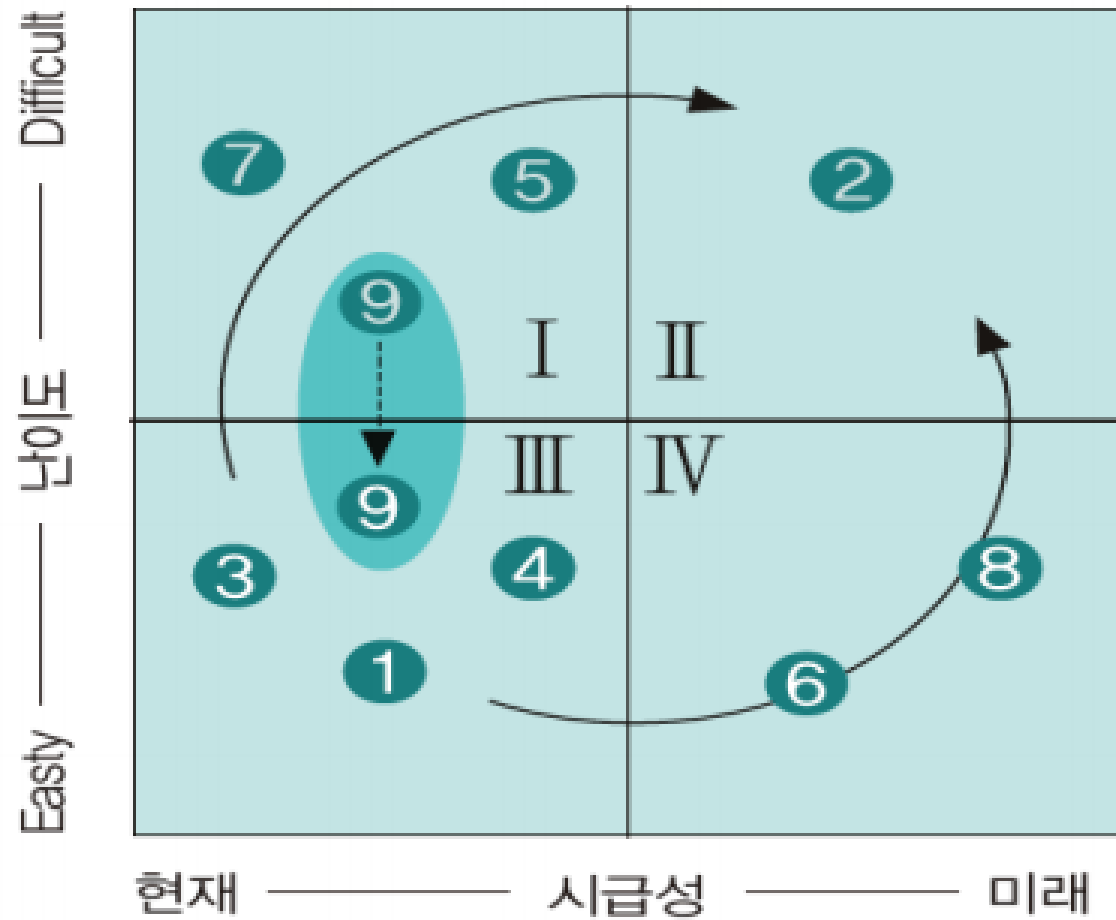
분석 마스터 플랜



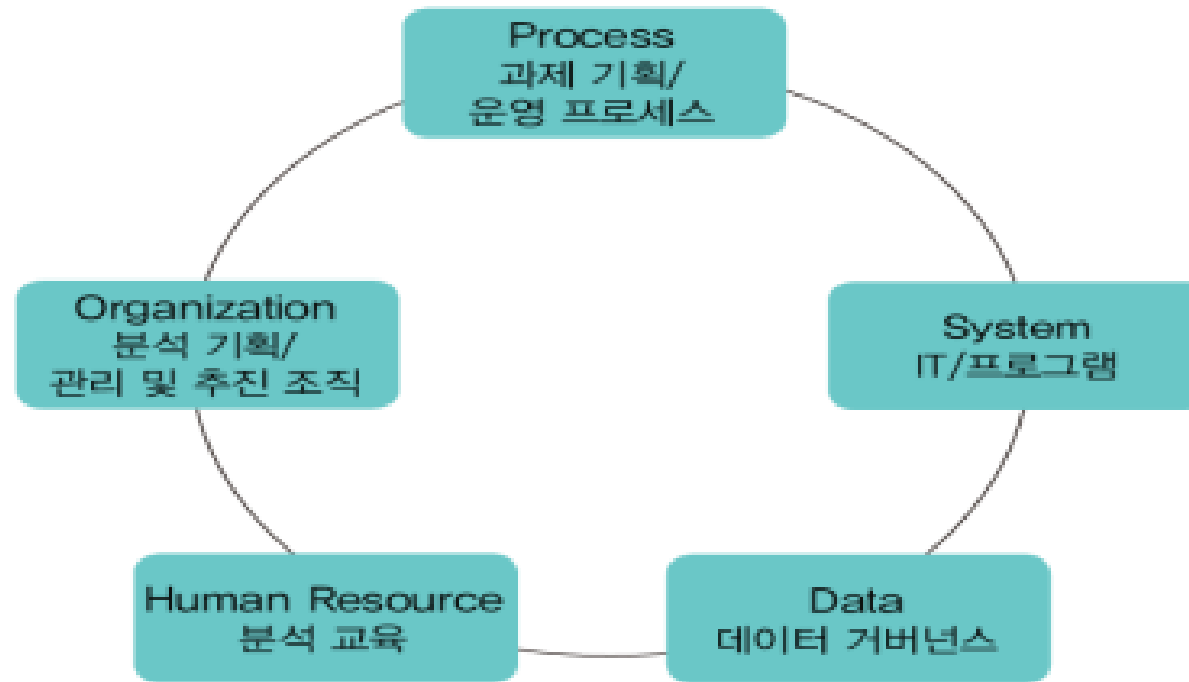
분석 마스터 플랜



분석 과제 우선순위 선정 조정



분석 거버넌스 체계 수립



분석 거버넌스 체계 구성 요소

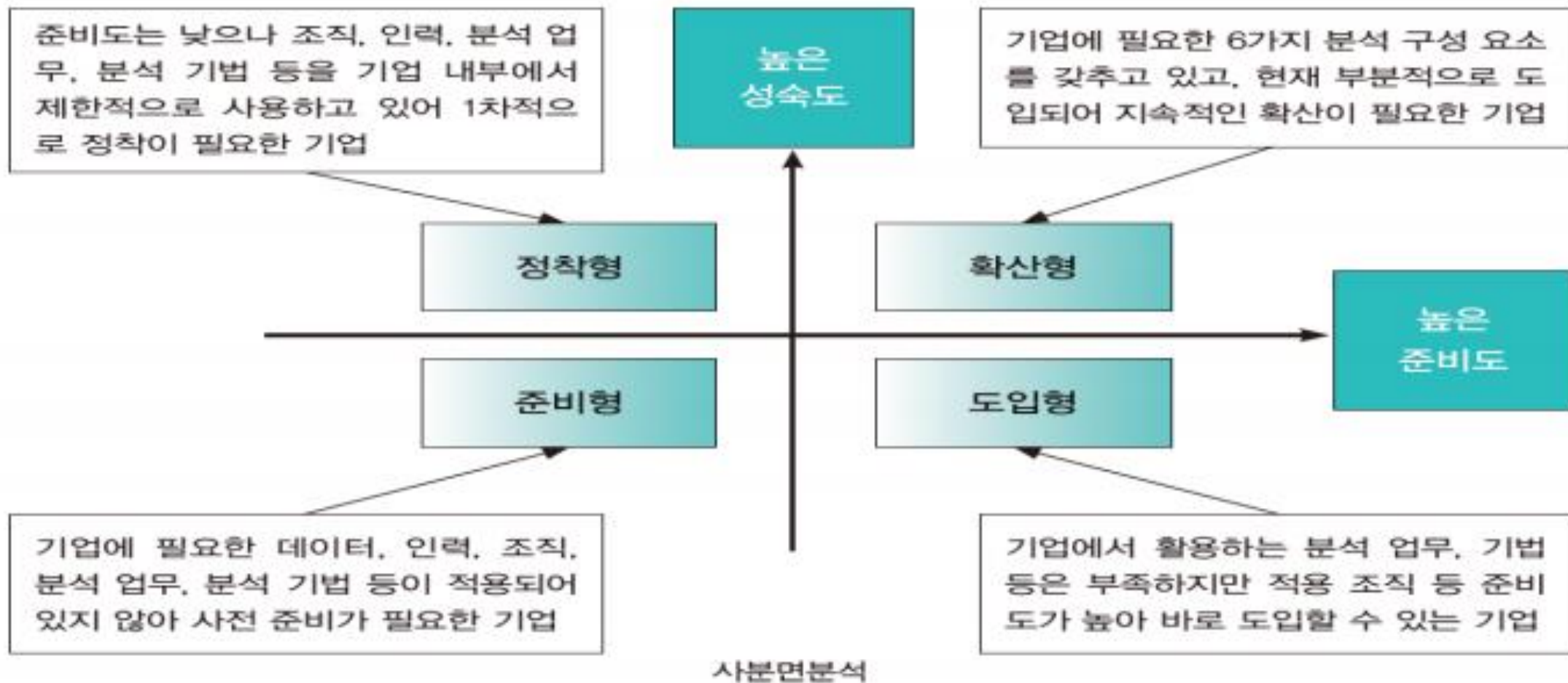
분석 준비도

분석 업무 파악	인력 및 조직	분석 기법
<ul style="list-style-type: none">• 발생한 사실 분석 업무• 예측 분석 업무• 시뮬레이션 분석 업무• 최적화 분석 업무• 분석 업무 정기적 개선	<ul style="list-style-type: none">• 분석 전문가 직무 존재• 분석 전문가 교육 훈련 프로그램• 관리자층의 기본적인 분석 능력• 전사 분석업무 총괄 조직 존재• 경영진 분석 업무 이해 능력	<ul style="list-style-type: none">• 업무별 적합한 분석기법 사용• 분석 업무 도입 방법론• 분석기법 라이브러리• 분석기법 효과성 평가• 분석기법 정기적 개선
분석 데이터	분석 문화	IT 인프라
<ul style="list-style-type: none">• 분석업무를 위한 데이터 충분성• 분석업무를 위한 데이터 신뢰성• 분석업무를 위한 데이터 적시성• 비구조적 데이터 관리• 외부 데이터 활용 체계• 기준데이터 관리(MDM)	<ul style="list-style-type: none">• 사실에 근거한 의사결정• 관리자층의 데이터 중시• 회의 등에서 데이터 활용• 경영진의 직관보다 데이터• 데이터 공유 및 협업 문화	<ul style="list-style-type: none">• 운영시스템 데이터 통합• EAI, ETL 등 데이터유통체계• 분석 전용 서버 및 스토리지• 빅데이터 분석 환경• 통계 분석 환경• 비주얼 분석 환경

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

분석 진단 결과





Thank you.

빅데이터 기초 / 류영표 강사
youngpyoryu@dongguk.edu

Copyright © “Youngpyo Ryu” All Rights Reserved.
This document was created for the exclusive use of “Youngpyo Ryu”.
It must not be passed on to third parties except with the explicit prior consent of “Youngpyo Ryu”.