

Data Science

데이터 분석

데이터 분석 기획

강사 양석환

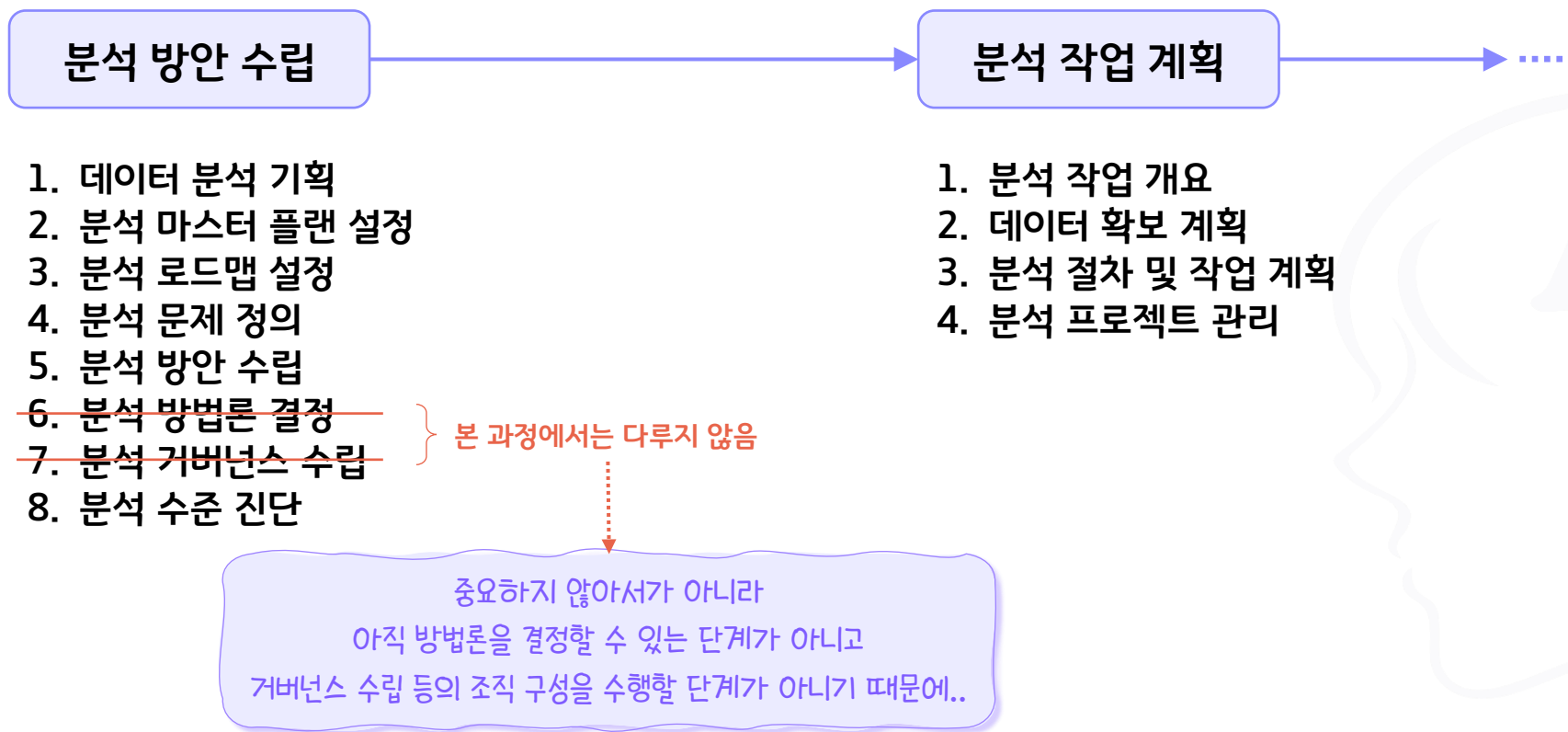


데이터 분석 계획



• 데이터 분석 계획

- 데이터의 분석을 위한 방안을 수립하고 세부적인 작업을 계획하는 과정



• 데이터 분석의 지향점

• 전략적 통찰이 없는 데이터 분석 배제

- 단순히 자주, 많이 분석한다고 좋은 것이 아니다.
- 본질을 놓친 분석은 불필요한 결과를 초래한다. 오히려 역효과인 경우도 있다.

데이터 분석이란?

대용량의 데이터 집합으로부터 유용한 정보를 찾고, 결과를 예측하기 위해 목적에 따라 분석 기술과 방법론을 기반으로
정형, 비정형 대용량 데이터를 구축, 탐색, 분석하고 시각화를 수행하는 업무

• 전략 도출을 위한 가치 기반의 데이터 분석 지향

- 사업 성과를 견인하는 요소들과 차별화를 꾀할 기회에 대해 전략적 인사이트를 주는 가치 기반 분석단계로 진출
- 사업과 관련 트렌드에 대한 청사진을 그리고, 인구통계학적 변화나 사회경제적 트렌드 및 고객 니즈의 변화 등을 고려하여 분석 수행

???

- 각 분야 별로 많이 수행되는 데이터 분석 내용

산업군	데이터 분석 내용
금융	신용점수 산정, 사기 방지, 프로그램 트레이딩, 고객 수익성 분석 등
소매	프로모션, 판매 관리, 수요 예측, 재고 관리 등
제조	공급사슬 최적화, 수요 예측, 재고 관리, 상품 개발 등
운송	일정 관리, 노선 배정, 수익 관리 등
병원(의료)	약품 거래, 질병 관리 등
에너지	트레이딩, 공급, 수요 예측 등
정부	사기 방지, 범죄 방지 등
온라인	고객 추천, 사이트 설계 등
기타	성과 관리 등

• 데이터 분석 기획

• 정의

- 실제 분석을 수행하기에 앞서
- 분석을 수행할 과제의 정의 및 의도했던 결과를 도출할 수 있도록
- 이를 적절하게 관리할 수 있는 방안을 사전에 계획하는 작업

• 작업 내용

- 어떠한 목표(What)를
- 달성하기 위해(Why)
- 어떠한 데이터를 가지고 어떤 방식으로(How) 수행할 것인가
- 에 대한 일련의 계획을 수립



- 어떻게 기획할 것인가?

- 분석 대상과 방법에 따른 분류를 따라서

- 분석 주제와 방법에 대한 특성상 4가지 유형으로 구분
 - 각 유형을 넘나들며 분석을 수행하고 결과를 도출하는 과정을 반복함

- 데이터 분석의 주제 유형

		분석의 대상(What)	
		Known	Un-Known
분석의 방식 (How)	Known	Optimization	Insight
	Un-Known	Solution	Discovery

• 어떻게 기획할 것인가?

• 목표 시점에 따른 분류를 따라서

단기적 접근 방식 (과제 중심적 접근)	<ul style="list-style-type: none">• 당면한 과제를 빠르게 해결하기 위한 목적을 가짐• 명확한 해결을 위해 Quick-Win(즉각적 성과 도출) 방식으로 접근
중장기적 접근 방식 (마스터 플랜 접근)	<ul style="list-style-type: none">• 지속적인 분석 문화를 내재화하기 위한 목적을 가짐• 전사적으로 장기적 관점에서 과제를 도출하여 수행
혼합 방식 (분석 기획 시 적합)	<ul style="list-style-type: none">• 마스터 플랜을 수립하고 장기적인 관점에서 접근하는 것이 바람직함• 분석의 가치를 증명하고 이해관계자들의 동의를 얻기 위해 과제를 빠르게 해결하여 그 가치를 조기에 체험 시킴

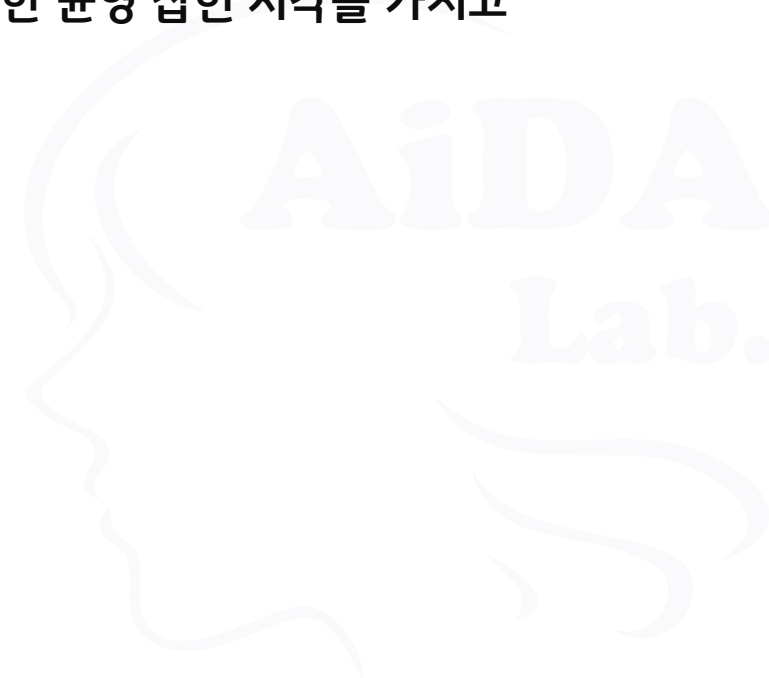
- **분석 기획 시 필요 역량**

- **분석 기획을 위한 기본적인 소양**

- 분석 기획은
 - 도메인 지식과 정보기술, 수학 및 통계학적 지식이라는 3가지 역량에 대한 균형 잡힌 시각을 가지고
 - 분석의 방향성과 계획을 수립하는 것

- **프로젝트 관리 역량과 리더십**

- 분석 기획 시 기본적인 3가지 소양과 함께
 - 프로젝트 관리 역량과
 - 분석 프로젝트를 잘 이끌어 갈 리더십이 중요함



• 분석 기획 시 고려사항

• 사용 가능한 데이터 확인

- 데이터 확보 가능 여부, 데이터의 유형 등을 미리 확인하여야 함
- 데이터의 유형에 따라 적용 가능한 솔루션이나 분석 방법론이 달라짐

• 적합한 사례 탐색

- 기존에 잘 구현되어 활용되고 있는 유사 분석 시나리오나 솔루션이 있다면 이를 최대한 활용하는 것이 유리함
- 분석 결과를 활용할 사용자의 측면에서 공감대를 얻을 수 있으며, 분석 수행이 원활하게 될 수 있도록 도와줌

• 분석 수행 시 발생 가능한 요소 고려

- 분석 결과의 정확도를 높이기 위하여 기간과 투입 자원 증가가 불가피하며, 이로 인한 비용 상승을 고려해야 함
- 분석 결과를 실제 환경에서도 성능에 문제없이 적용할 수 있도록 충분히 고려해야 함

• 분석 마스터 플랜 수립

• 분석 마스터 플랜이란?

- 분석 과제를 수행함에 있어 그 과제의 목적이나 목표에 따라 전체적인 방향성을 제시하는 기본 계획

• 분석 마스터플랜 수립

- 분석 마스터 플랜 수립 시 일반적인 **정보 전략 계획 방법론**을 활용할 수 있음
- 데이터 **분석 기획의 특성**을 고려하여 수행하여야 함
- 과제 도출 방법을 활용하여 데이터 **분석 과제들을 빠짐없이 정의**함
- 분석 과제의 **중요도와 난이도** 등을 고려하여 **우선 순위 결정**
- 단기와 중장기로 나누어 **분석 로드맵** 수립

정보 전략 계획(Information Strategy Planning, ISP)

- 정보기술 및 시스템을 전략적으로 활용하기 위한 중장기 마스터 플랜을 수립하는 절차
- 조직 내·외부의 환경을 충분히 분석하여 새로운 기회나 문제점을 도출함
- 사용자의 요구사항을 확인하여 시스템 구축 우선순위를 결정함

- **분석 로드맵 설정**

- **분석 로드맵이란?**

- 마스터 플랜에서 정의한 목표를 기반으로
 - 분석 과제를 수행하기 위해 필요한 기준 등을 담아 만든 종합적인 계획

- **분석 로드맵 수립 절차**

- 최종적인 실행 우선순위를 결정하여 단계적 구현 로드맵을 수립
 - 각 단계별로 추진하고자 하는 목표를 명확하게 정의
 - 추진 과제별 선행 관계를 고려하여 단계별 추진 내용을 정렬



- **분석 로드맵 설정**

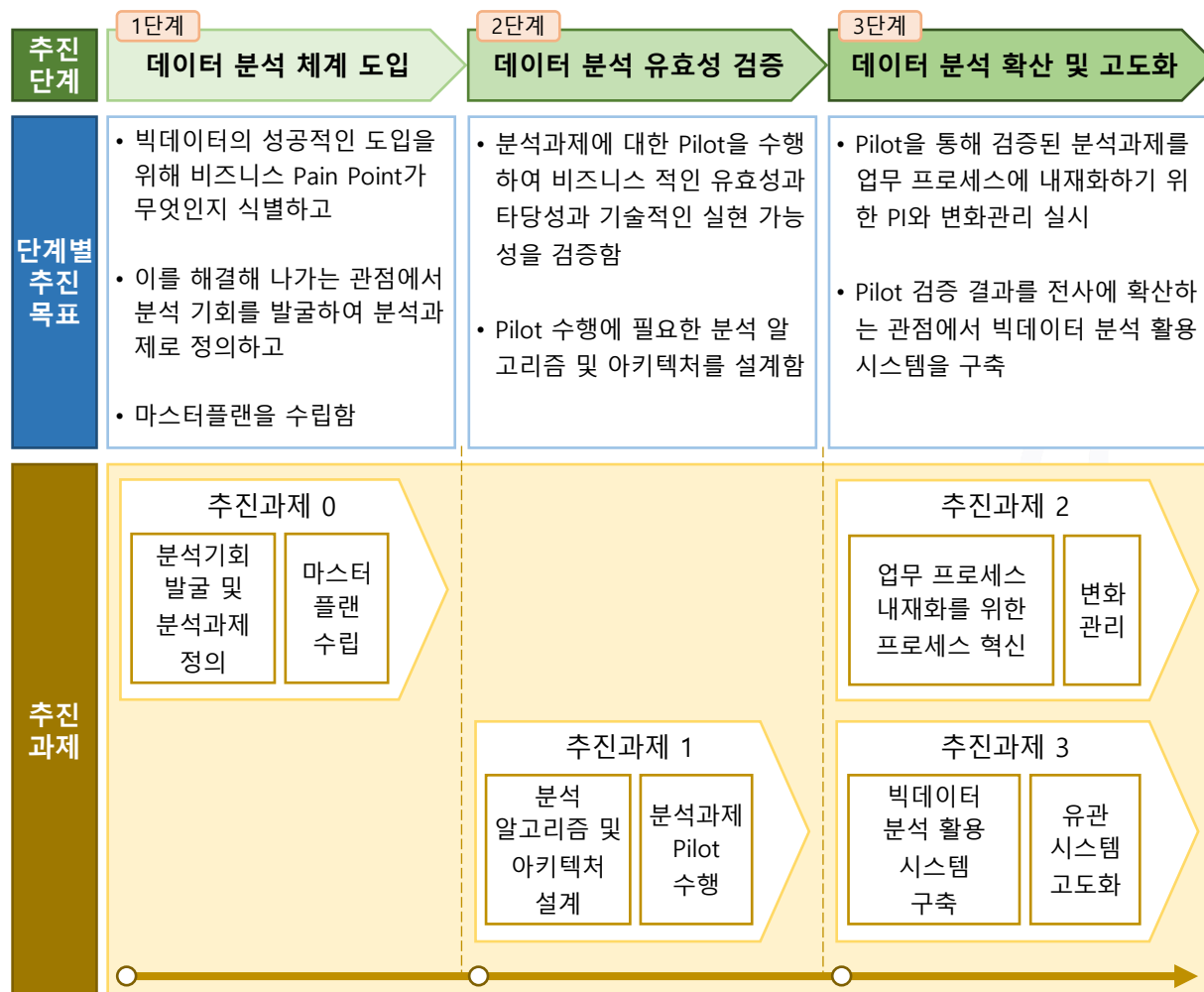
- **세부적인 일정계획 수립 방법**

- 반복적인 정련 과정을 통해 프로젝트의 완성도를 높여 나감
 - 데이터 수집 및 확보와 분석 데이터 준비 단계는 순차적으로 진행하고 모델링 단계는 반복적으로 수행함
 - 주로 순차형과 반복형을 혼합하여 사용함



• 분석 로드맵 설정

• 예시



• 분석 문제 정의

• 분석 과제 도출

- 해결해야 할 다양한 문제들을 데이터 분석 문제로 변환
- 분석 프로젝트로 수행할 수 있는 과제 정의서 형태로 도출

• 대표적인 분석 과제 도출 방법

• 하향식 접근

- 문제가 먼저 주어지고 이에 대한 해법을 찾아가는 방식
- 전통적인 분석과제 발굴 방식 → 근래의 문제들은 변화가 심하고 문제를 사전에 정확하게 정의하기 어려움

• 상향식 접근

- 데이터를 기반으로 문제 재정의 및 해결방안을 탐색하고 이를 지속적으로 개선하는 방식
- 문제의 정의 자체가 어려운 경우에 많이 사용됨

근래의 데이터 분석에는 상향식 접근이 좀 더 잘 어울림

• 분석 문제 정의

• 최적의 의사결정을 위한 혼합 방식

- 동적인 환경에서 발산과 수렴 단계를 반복적으로 수행
→ 상호 보완을 통해 분석의 가치를 극대화할 수 있음

신상품 개발이나 전략 수립 등
중요한 의사결정이 필요할 때 주로 사용

- 상향식 접근 방식의 발산(Diverge) 단계: 가능한 옵션을 도출
- 하향식 접근 방식의 수렴(Converge) 단계: 도출된 옵션을 분석하고 검증함

• 데이터 분석 방안 수립

• 분석 방법론

- 데이터 분석을 효과적으로 수행하기 위하여 분석 절차를 체계적으로 정리한 방법
- 데이터 분석 시 품질 확보를 위하여 단계별로 수행해야 하는 활동, 작업, 산출물을 정의함

• 분석 방법론의 구성 요건

- 상세한 절차(Procedures)
- 방법(Methods)
- 도구와 기법(Tools & Techniques)
- 템플릿과 산출물(Templates & Outputs)
- 어느 정도의 지식만 있으면 활용 가능한 수준의 난이도

분석 방법론은 왜 필요한가?

프로젝트는 한 개인의 역량이나 조직의 우연한 성공에 의해서는 안되고
일정 품질 수준 이상의 산출물과 프로젝트의 성공 가능성을 제시해야 하므로!

• 데이터 분석 방안 수립

• 분석 방법론

• 분석 방법론의 생성 과정(선순환 과정)

• 형식화

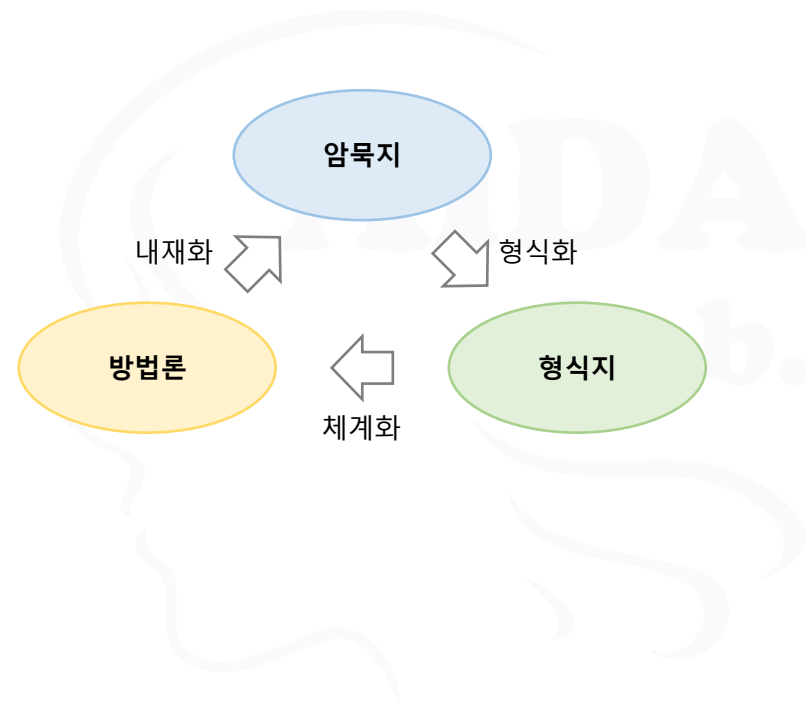
- 개인의 암묵지가 조직의 형식지로 발전되었음
- 분석가의 경험을 바탕으로 정리하여 문서화

• 체계화

- 문서화한 최적화된 형식지로 전개됨으로써 방법론이 생성됨
- 문서에는 절차나 활동 및 작업, 산출물, 도구 등을 정의함

• 내재화

- 개인에게 전파되고 활용되어 암묵지로 발전됨
- 전파된 방법론을 학습하고 활용하여 내재화함



• 데이터 분석 방안 수립

• 분석 방법론

• 계층적 프로세스 모델 구성

- 분석 방법론은 일반적으로 계층적 프로세스 모델 형태로 구성이 가능하며 단계, 태스크, 스텝의 3계층으로 구성됨
- 최상위 계층 – 단계(Phase)
 - 프로세스 그룹을 통해 완성된 단계별 산출물 생성. 각 단계는 기준선으로 설정되어 관리. 버전 관리 등을 이용하여 통제
- 중간 계층 – 태스크(Task)
 - 각 태스크는 단계를 구성하는 단위 활동임. 물리적 또는 논리적 단위로 품질검토가 가능함
- 최하위 계층 – 스텝(Step)
 - WBS(Work Breakdown Structure)의 워크 패키지(Work Package)임.
 - 입력자료, 처리 및 도구, 출력자료로 구성된 단위 프로세스

• 데이터 분석 방안 수립

• 분석 방법론

• 소프트웨어 개발 생명주기 활용

- 분석 방법론은 소프트웨어 공학의 소프트웨어 개발 생명주기를 활용하여 구성할 수도 있음
- 소프트웨어 개발 생명주기(Software Development Life Cycle, SDLC)란?
 - SW에 대해 요구분석과 설계, 구현과정을 거쳐 설치, 운영과 유지보수, 폐기까지의 전 과정을 가시적으로 표현한 것

• SDLC 구성요소

구성요소	내용
계획(요구명세)	<ul style="list-style-type: none">• 고객의 요구사항을 명세화.• 타당성 조사 및 소프트웨어의 기능과 제약조건을 정의하는 명세서를 작성함• 요구사항은 일반적으로 모호하고 불완전하며 모순되기도 함
요구분석	<ul style="list-style-type: none">• 대상이 되는 문제 영역과 사용자가 원하는 태스크를 이해함
설계	<ul style="list-style-type: none">• 분석 모형을 가지고 이를 세분화함으로써 구현될 수 있는 형태로 전환시킴
구현	<ul style="list-style-type: none">• 실행 가능한 코드를 생성함
시험(테스트)	<ul style="list-style-type: none">• 발생 가능한 실행 프로그램의 오류를 발견하고 수정함
유지보수	<ul style="list-style-type: none">• 인수가 완료된 후 일어나는 모든 개발 활동을 가리킴

• 데이터 분석 방안 수립

• 분석 방법론

- KDD (Knowledge Discovery in Database) 분석 방법론
 - 통계적인 패턴, 지식을 탐색하는데 활용할 수 있도록 체계적으로 정리한 프로파일링 기술 기반의 데이터 마이닝 프로세스
 - 데이터 마이닝, 기계학습, 인공지능, 패턴인식, 시각화 등에 응용 가능
- CRISP-DM (Cross Industry Standard Process for Data Mining) 분석 방법론
 - 계층적 프로세스 모델로써 4계층으로 구성된 데이터 마이닝 프로세스
 - 데이터 탐색과 이해 및 데이터를 통한 문제 인식과 해결 가능
 - 비즈니스 요구사항에 맞게 데이터 마이닝을 반복적으로 수행할 수 있음
- SEMMA (Sample, Explore, Modify, Model and Assess) 분석 방법론
 - SAS Institute의 주도로 만들어진 기술과 통계 중심의 데이터 마이닝 프로세스

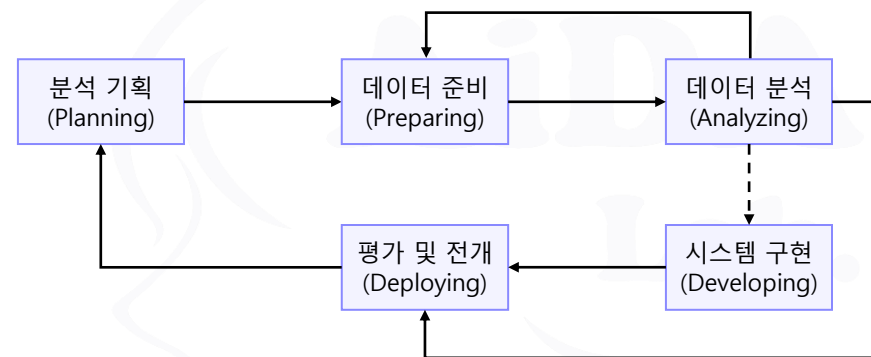
• 데이터 분석 방안 수립

• 분석 방법론

• 빅데이터 분석 방법론

- 응용 서비스 개발을 위한 3계층으로 구성

구분	내용
단계 (Phase)	<ul style="list-style-type: none">• 데이터 분석을 수행하기 위한 절차• 기준선(Baseline)을 설정하고 버전 관리를 통해 통제되어야 함
태스크 (Task)	<ul style="list-style-type: none">• 각 단계별로 수행되어야 하는 세부 업무• 각 태스크가 완료되면 그에 대한 성과를 얻을 수 있음
스텝 (Step)	<ul style="list-style-type: none">• 단기간 내에 수행 가능한 워크패키지(Work Package)• 입력자료, 처리 및 도구, 출력자료로 구성된 단위 프로세스



• 데이터 분석 방안 수립

• 분석 방법론

• 빅데이터 분석 방법론

• 빅데이터 분석 방법론의 개발 절차



• 분석 수준 진단

• 분석 수준진단의 필요성

- 조직 경쟁력 강화를 위한 데이터 분석의 도입 여부와 활용을 위해 현 상태에 대한 명확한 점검이 필요함
- 데이터 분석의 수준 진단을 통해 데이터 분석 기반을 만들기 위해 무엇을 준비하고 더 보완해야 하는지 확인 가능
- 데이터 분석의 유형이나 방향을 결정할 수 있음

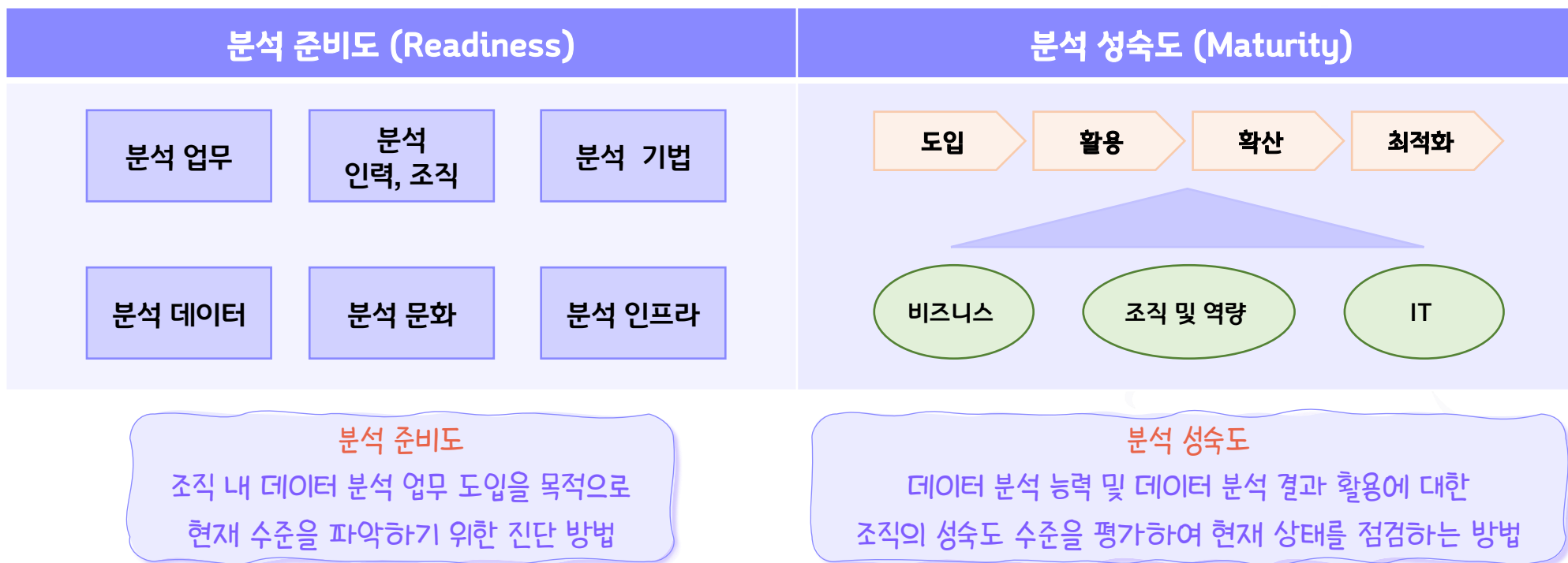
• 분석 수준진단의 목표

- 각 조직이 현재 수행하고 있는 데이터 분석 수준을 명확하게 이해하고 수준 진단 결과를 바탕으로 미래 목표 수준을 정의함
- 데이터 분석을 위한 기반이나 환경이 타사 대비 어느 정도 수준이고, 어느 영역에 선택과 집중을 해야 하는지, 무엇을 보완해야 하는지 등의 개선 방안을 도출함

• 분석 수준 진단

• 분석 수준 진단 프레임워크

- 6개의 분석 준비도와 3개 영역의 분석 성숙도를 동시에 평가할 수 있음



- **분석 수준 진단**

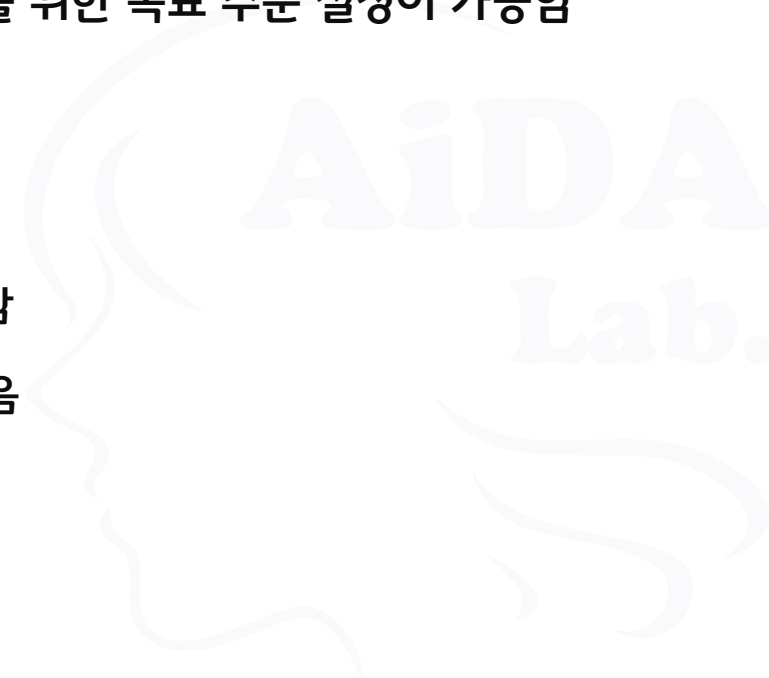
- **분석 수준 진단 결과**

- 조직의 현재 데이터 분석 수준을 객관적으로 파악할 수 있음
 - 타사의 데이터 분석 수준과 비교하여 데이터 분석 경쟁력 확보 및 강화를 위한 목표 수준 설정이 가능함

- **대표적인 방법**

- **사분면 분석 (Analytics Quadrant)**

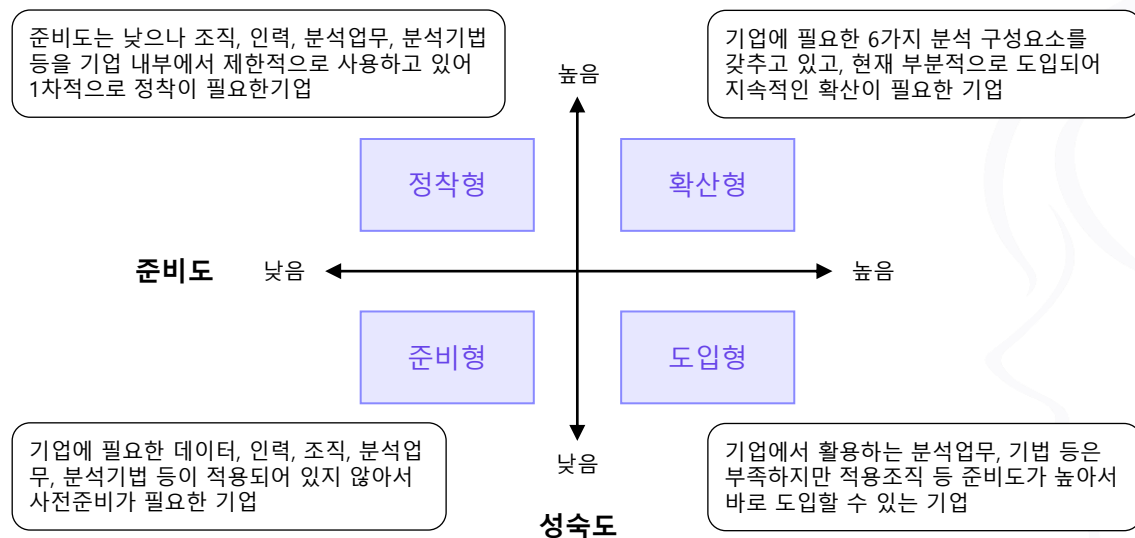
- 데이터 분석 관점에서 4가지 유형으로 데이터 분석 수준진단 결과를 구분함
 - 향후 고려해야 하는 데이터 분석 수준에 대한 목표나 방향을 정의할 수 있음
 - 유형별 특성에 따라 개선방안을 수립할 수 있음



- 분석 수준 진단

- 대표적인 방법

- 사분면 분석 (Analytics Quadrant)



• 분석 작업 계획

• 분석 작업 개요

- 분석 작업 계획을 수립하기 위해 데이터 처리 프로세스 전체에 대한 이해가 필요하며, 데이터 처리 영역과 데이터 분석 영역으로 나누어 살펴볼 수 있음

• 데이터 처리 프로세스



**THANK
YOU**

