

데이터분석준전문가(ADsP) Part 2

데이터 분석 기획

01

1장 데이터 분석 기획의 이해
2장 분석 마스터 플랜



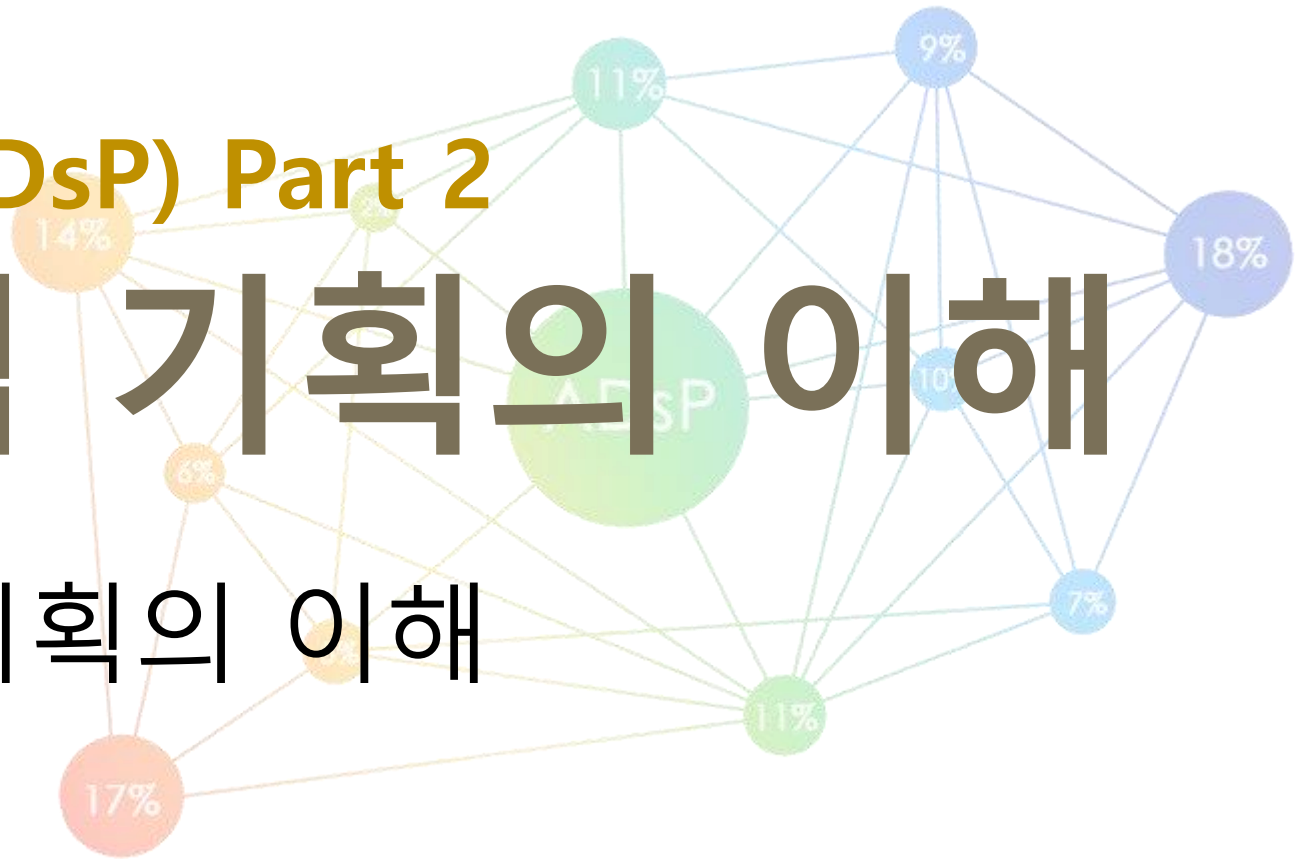
데이터분석전문가(ADsP) Part 2

데이터 분석 기획의 이해

01

1장 데이터분석 기획의 이해

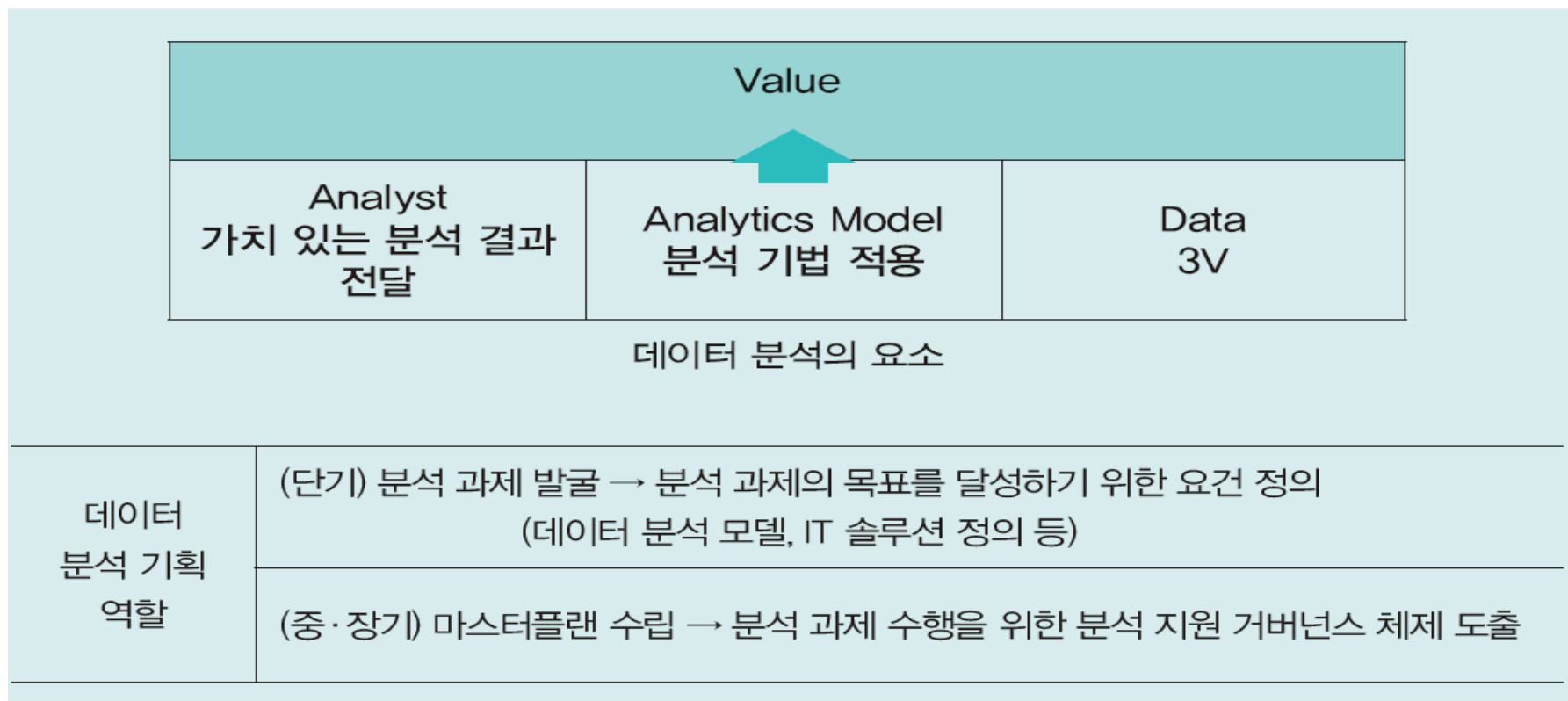
1. 분석 기획 방향성 도출
2. 분석방법론
3. 분석 과제 발굴
4. 분석 프로젝트 방안



1.데이터 분석 기획 방향성 도출



1) 데이터 분석 기획이란?



1. 데이터 분석 기획 방향성 도출



2) 분석기획의 방향성 도출

33

분석은 분석의 대상(what), 분석의 방법(how)에 따라서 4가지 구분

		분석의 대상(what)	
		Known	Un-Known
분석의방법(How)	Known	최적화(Optimization)	통찰(Insight)
	Un-Known	솔루션(Solution)	발견(Discovery)

- ① **Optimization** → 분석대상 및 분석방법을 이해하고 현 문제를 최적화의 형태로 수행
- ② **Solution** → 분석과제는 수행되고, 분석 방법을 알지 못하는 경우 솔루션을 찾는 방식으로 분석과제 수행
- ③ **Insight** → 분석대상이 불분명하고, 분석방법을 알고 있는 경우 인사이트 도출
- ④ **Discovery** → 분석대상, 방법을 모른다면 발견을 통하여 분석 대상 자체를 새롭게 도출

1. 데이터 분석 기획 방향성 도출



3) 목표 시점 별 분석 기획방안

34

목표시점별 분석기획방안 단기방안(과제중심적인 접근방식) 중장기(장기적인 마스터플랜)구분

당면한 분석 주제의 해결
(과제단위)

지속적 분석 문화 내재화
(마스터플랜 단위)

Speed & Test	〈1차 목표〉	Accuracy & Deploy
Quick-Win	〈과제의 유형〉	Long Term View
Problem Solving	〈접근방식〉	Problem Definition

1. 데이터 분석 기획 방향성 도출



4) 분석 기획 시 고려사항

35

분석 기획시 고려사항(3가지)

① 가용한 데이터(Available Data)

→ 분석을 위한 데이터의 확보가 필수적임. 데이터 유형에 따라서 적용 가능한 솔루션 및 분석방법이 다르기 때문에 유형에 대한 분석이 선행적으로 이루어져야 함.

예) 정형 데이터, 비정형 데이터, 반정형데이터

② 적절한 유스케이스(Proper Use-Case) 탐색

→ 유사분석 시나리오 및 솔루션이 있다면 이를 최대한 활용하는 것이 중요함

③ 장애요소들에 대한 사전 계획 수립이 필요(Low Barrier of Execution)

→ 정확도를 올리기 위해서는 기간과 투입 리소스가 늘어나게 되는데 이것은 비용 상승으로 이어질 수 있으므로 많은 사전 고려가 필요함. 일회성 분석으로 그치지 않고 조직의 역량을 내재화하기 위해서는 충분하고 지속적인 교육 및 활용방안 등의 변화관리(Change Management)가 고려되어야 함.

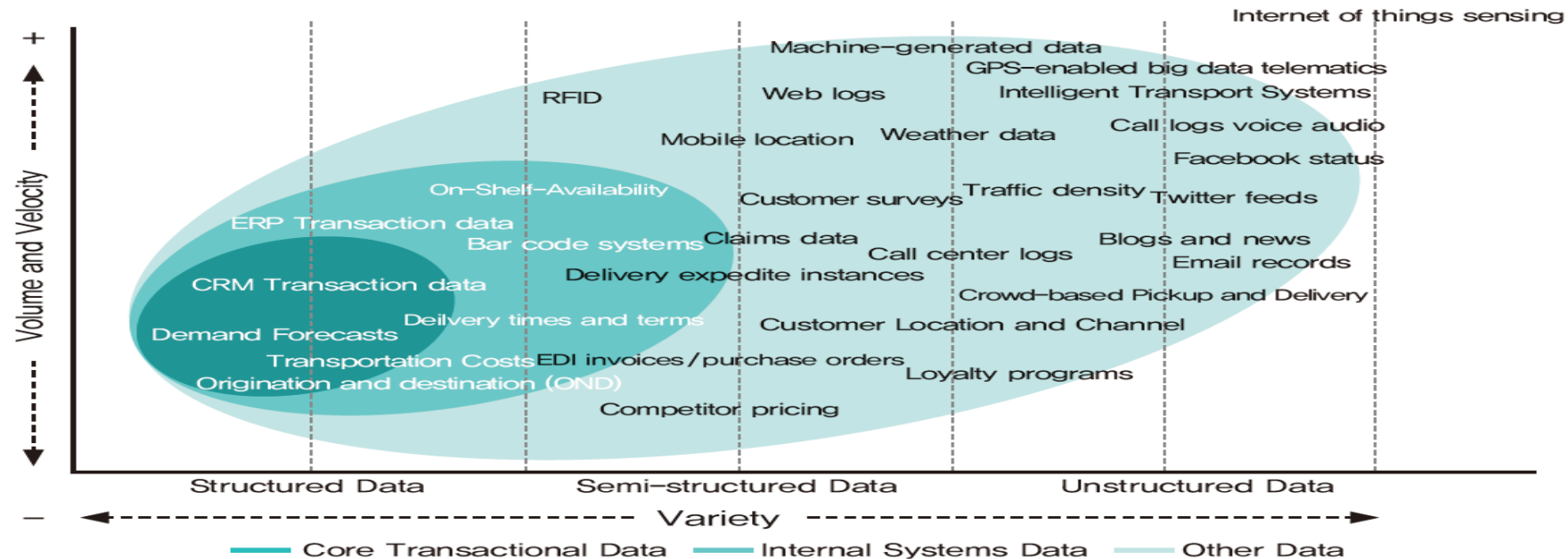
1. 데이터 분석 기획 방향성 도출



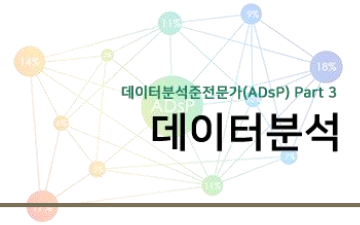
4) 분석 기획 시 고려사항

36 데이터의 유형 3가지

- ① 정형 데이터 ex) ERP, CRM Transaction data, Demand Forecast
- ② 비정형 데이터 ex) e-mail, 보고서, 소셜미디어, blogs and news
- ③ 반정형데이터 ex) machine data



2. 분석방법론



1) 분석방법론 개요

데이터 기반 의사결정 -> 데이터분석->기업문화의 변화, 업무프로세스 개선



분석 방법론 필요

분석방법론의 구성요소

- ① 절차(Procedures)
- ② 방법(Methods)
- ③ 도구와 기법(Tool & Techniques)
- ④ 템플릿과 산출물(Templates & Output)

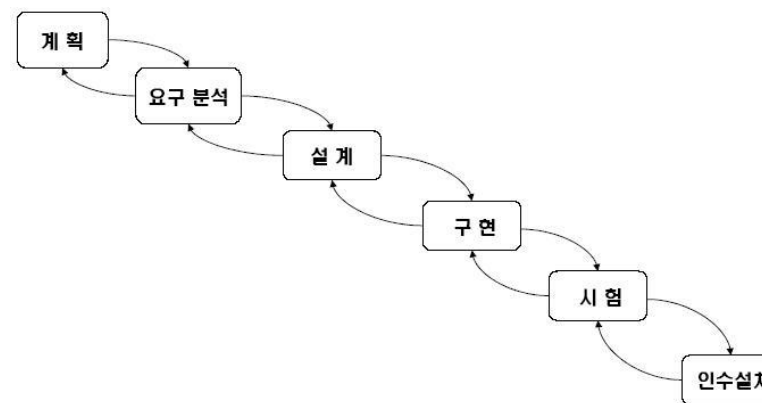
합리적 의사결정의 장애요소: 고정관념(Stereotype), 편향된 생각(Bias)
프레이밍 효과(Framing Effect)

2. 분석방법론



2.1) 폭포수 모형

- 1970년대 소개: 항공 방위 소프트웨어 개발 경험으로 습득
- 각 단계가 다음 단계 시작 전에 끝나야 함
 - 순서적 - 각 단계 사이에 중복이나 상호작용이 없음
 - 각 단계의 결과는 다음 단계가 시작 되기 전에 점검
 - 바로 이전 단계로의 피드백만이 가능함
- 단순하거나 응용 분야를 잘 알고 있는 경우 적합
- 장점
 - 개발자가 어떤 작업을 수행하고 있는지 그 단계가 명확하다.
 - 프로세스가 간단하여 일반인도 쉽게 이해할 수 있다.
 - 중간 산출물이 명확하게 지정되어 있다.
- 단점
 - 처음 단계를 지나치게 강조하면 코딩, 테스트가 지연된다.
 - 각 단계의 전환에 많은 노력을 기울여야 한다. (뒤로 돌아가기 어렵다.)



2. 분석방법론



2.2) 프로토타입 모형

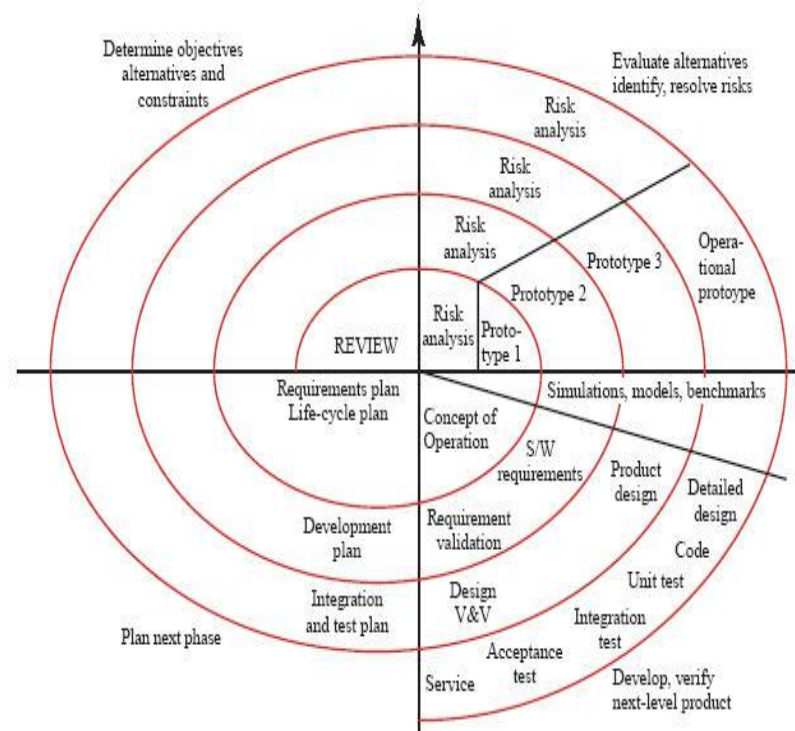
- 프로토타이핑: 시스템의 일부 혹은 모형을 만드는 과정
- 시뮬레이션 수행이나 데모 시스템 개발
 - 개발 단계에서 유지보수가 이루어짐
- 단점
 - (프로토타입을 보고, 전부 다 개발한 것 아니냐고 할 수 있음)

2. 분석방법론



2.3) 나선형 모형

- 소프트웨어의 기능을 나누어 점증적으로 개발
 - 실패의 위험을 줄임 → 위험 분석에 초점을 맞춘 개발 모형
 - 테스트 용이
 - 피드백이 용이
- 진화 단계
 - 계획 수립(planning): 목표, 기능 선택, 제약 조건의 결정
 - 위험 분석(risk analysis): 기능 선택의 우선순위, 위험요소의 분석
 - 개발(engineering): 선택된 기능의 개발
 - 평가(evaluation): 개발 결과의 평가
- 대규모 시스템 개발에 적합
 - Risk reduction mechanism
- 반복적인 개발 및 테스트
- 단점
 - 관리가 중요 (상당히 복잡하고,)
 - 상대적으로 새로운 모형 (적용이 쉽지 않음)



2. 분석방법론



38 KDD 분석이란?

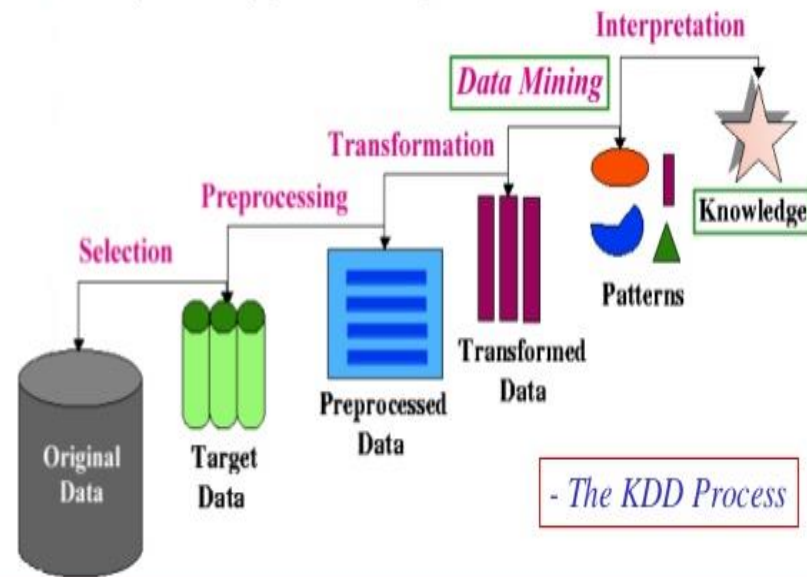
KDD 분석

- ① 분석대상의 비즈니스 도메인에 대한 이해와 프로젝트 목표를 정확하게 설정
- ② 데이터셋 선택 (Selection)
- ③ **데이터 전처리(Preprocessing)**: 분석 데이터셋에 포함되어 있는 잡음(Noise), 이상값(Outlier), 결측치 (Missing value)를 식별하고 필요시 제거
- ④ **데이터 변환(Transformation)**: 분석 목적에 맞는 변수를 선택하거나 데이터의 차원을 축소하여 데이터 마이닝을 효율적으로 적용할 수 있도록 데이터셋을 변경하는 작업
- ⑤ **데이터 마이닝(Data Mining)**: 변환된 데이터셋을 이용하여 분석 목적에 맞는 데이터 마이닝 기법을 선택하고 데이터 마이닝 알고리즘을 선택하여 데이터의 패턴을 찾거나 데이터를 분류 또는 예측 등의 마이닝 작업을 시행한다.
- ⑥ 데이터 마이닝 결과 평가 (Interpretation/Evaluation): 분석 결과에 대한 해석과 평가, 활용

The Knowledge Discovery Process

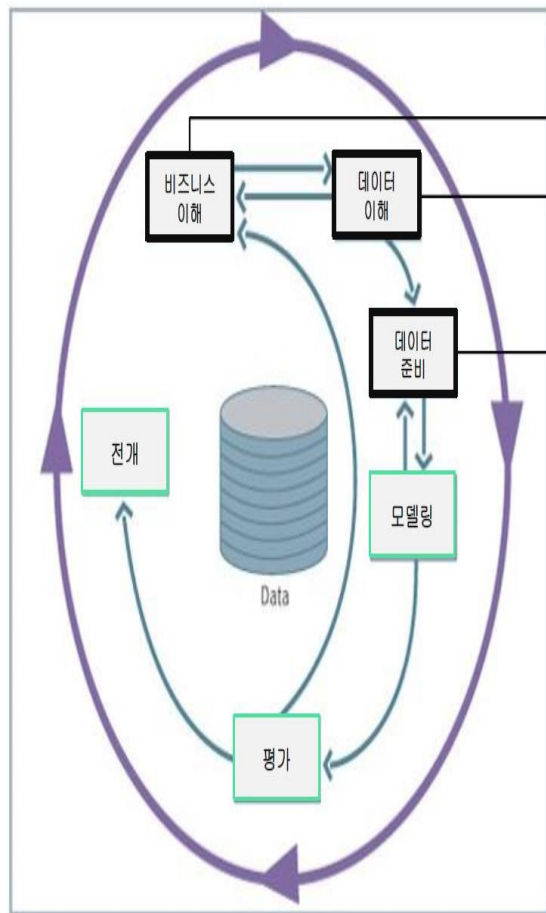
• Data Mining v. Knowledge Discovery in Databases (KDD)

- ▶ DM and KDD are often used interchangeably
- ▶ actually, DM is only part of the KDD process



2. 분석방법론

CRISP-DM



단계 1: 비즈니스 이해(Business Understanding)

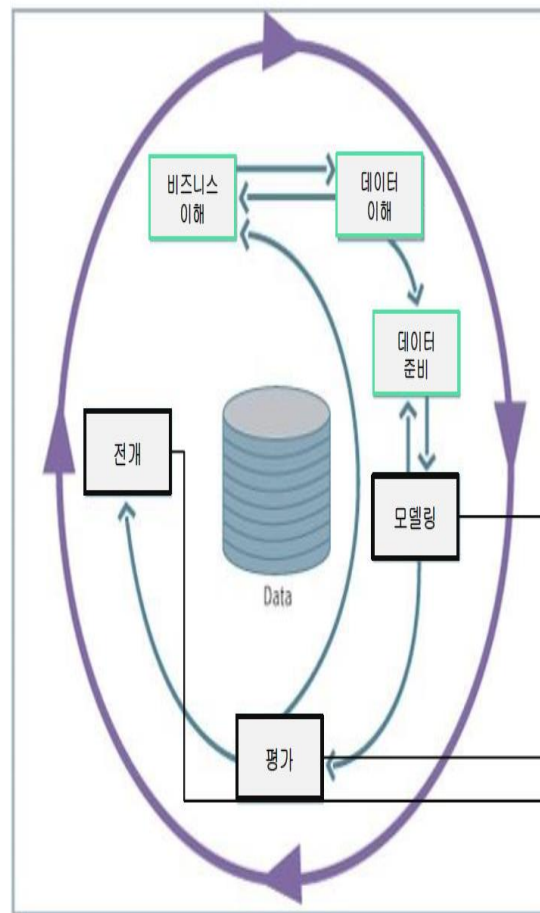
- 각종 참고 자료와 현업 책임자와의 커뮤니케이션을 통해 해당 비즈니스를 이해하는 단계

단계 2: 데이터 이해(Data Understanding)

- 레코드 수, 변수종류, 자료의 질 등, 현업이 보유하고 있는 데이터를 이해하는 단계

단계 3: 데이터 준비(Data Preparation)

- 데이터의 정제, 새로운 데이터 생성 등, 자료를 분석 가능한 상태로 만드는 단계



단계 4: 모델링 (Modeling)

- 자료 기술 및 탐색을 포함하여 필요한 각종 모델링을 하는 단계

단계 5: 평가 (Evaluation)

- 모형의 해석 가능 여부, 독립적인 새 자료에 적용되는 경우에도 재현가능한지를 검토하는 단계

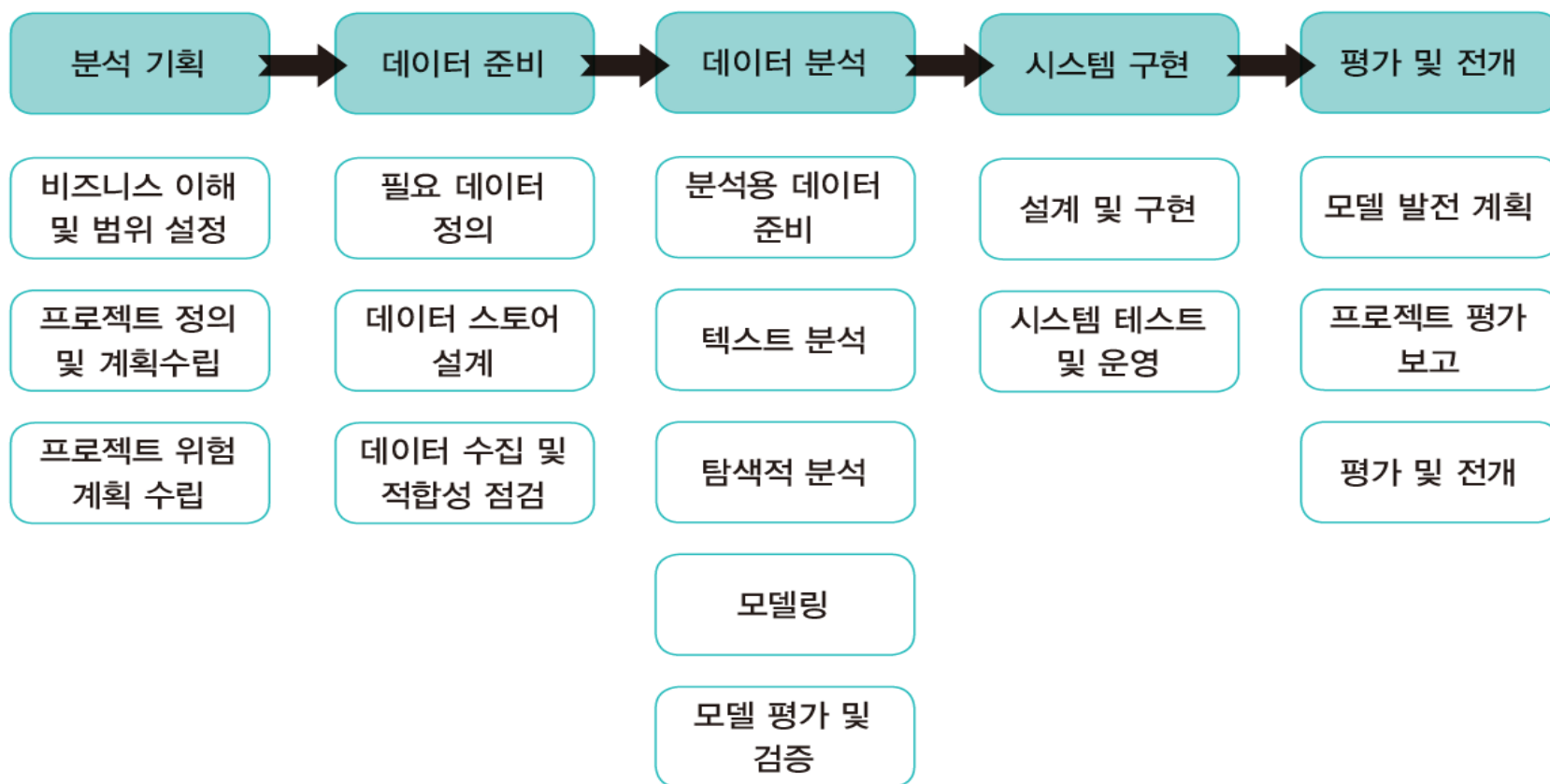
단계 6: 전개 (Deployment)

- 각 관리자에게 전달하여 필요한 조치를 취하는 등 검토가 끝난 모형을 실제 현업에 적용하는 단계

2. 분석방법론



빅 데이터 분석 방법론



2. 분석방법론



WBS의 Working Package

WBS(Working Breakdown Structure)

프로젝트 전체 범위를 단계적으로 분할하여 계층적 구조로 설명

계층도 형식의 WBS



테이블 형식의 WBS

마음 가득 선물 세트	1.0 통조림	1.1 돼지고기
		1.2 참치
	2.0 식용유	2.1 카놀라유
		2.2 포도씨유
	3.0 양념	3.1 된장
		3.2 고추장
		3.3 참기름

WP: WBS의 최하위 계층 구성하는 요소

2. 분석방법론



41

분석기획 단계

① 비즈니스 이해 및 범위 설정:

– 비즈니스 이해

빅데이터 분석 대상인 업무 도메인을 이해하기 위해서는 내부 업무 매뉴얼과 관련자료, 외부의 관련 비즈니스 자료를 조사하고 향후 프로젝트 진행을 위한 방향을 설정한다.

– 프로젝트 범위 설정

빅데이터 분석 프로젝트의 대상인 비즈니스에 대한 이해와 프로젝트 목적에 부합하는 범위를 명확히 설정하고 프로젝트에 참여하는 관계자들(Stakeholders)의 이해를 일치시키기 위하여 구조화된 프로젝트 범위 정의서인 SOW(Statement of Work)를 작성한다.

2. 분석방법론



② 프로젝트 정의 및 계획 수립

- 데이터 분석 프로젝트 정의

프로젝트의 목표 및 KPI(핵심성과지표), 목표수준 등을 구체화하여 상세 프로젝트 정의서를 작성하고 프로젝트의 목표를 명확화하기 위하여 모델 이미지 및 평가 기준을 설정한다.

- 프로젝트 수행 계획 수립

프로젝트 수행 계획서를 작성하는 단계로써 프로젝트의 목적 및 배경, 기대효과, 수행방법, 일정 및 추진조직, WBS을 작성한다.

③ 프로젝트 위험 계획 수립

-데이터 분석 위험 식별, 계획수립 단계에서 빅데이터 분석 프로젝트를 진행하면서 발생 가능한 모든 위험을 식별한다. 식별된 위험은 위험의 영향도와 빈도, 발생가능성 등을 평가하여 위험의 우선순위를 설정한다.

2. 분석방법론



위험 대응 방법

접근방법	내용
회피(Avoid)	모든 프로젝트 위험 제거 할 수 없으나 회피 할 수 있음 예) 기간연장, 범위축소
전이(Transfer)	위험의 결과를 제3자에게 이전 노력 예) 이행보증, 사후보증
완화(Mitigate)	용인 가능한 임계치까지 절감하고자 노력
수용(Accept)	위험을 취급하는 프로젝트 계획을 변경하지 않음

2. 분석방법론



42

데이터준비 단계

① 필요 데이터 정의

– 데이터 정의: 정형·비정형·반정형 등의 모든 내·외부 데이터를 포함하고 데이터의 속성, 데이터 오너, 데이터 관련 시스템 담당자 등을 포함하는 데이터 정의서를 작성한다

예) 메타데이터 정의서, ERD 포함.

– 데이터 획득방안 수립.

② 데이터 스토어 설계

– 정형 데이터 스토어 설계: 일반적으로 관계형데이터베이스(RDBMS)를 사용하고 데이터의 효율적인 저장과 활용을 위하여 데이터 스토어의 논리적, 물리적 설계를 구분하여 설계한다.

예) 데이터 매핑 정의성

– 비정형 데이터 스토어 설계: 하둡, NoSQL 등을 이용하여 비정형 또는 반정형 데이터를 저장하기 위한 논리적, 물리적 데이터 스토어를 설계한다.

③ 데이터 수집 및 정합성 점검

– 데이터 수집 및 저장

크롤링 등의 데이터 수집을 위한 ETL 등의 다양한 도구와 API, 스크립트 프로그램 등을 이용하여 데이터를 수집하고, 수집된 데이터를 설계된 데이터를 스토어에 저장한다.

– 데이터 정합성(무결성) 점검

데이터 스토어의 품질 점검을 통하여 데이터의 정합성을 확보하고 데이터 품질개선이 필요한 부분에 대하여 보완 작업을 한다.

2. 분석방법론

• 메타 데이터의 정의

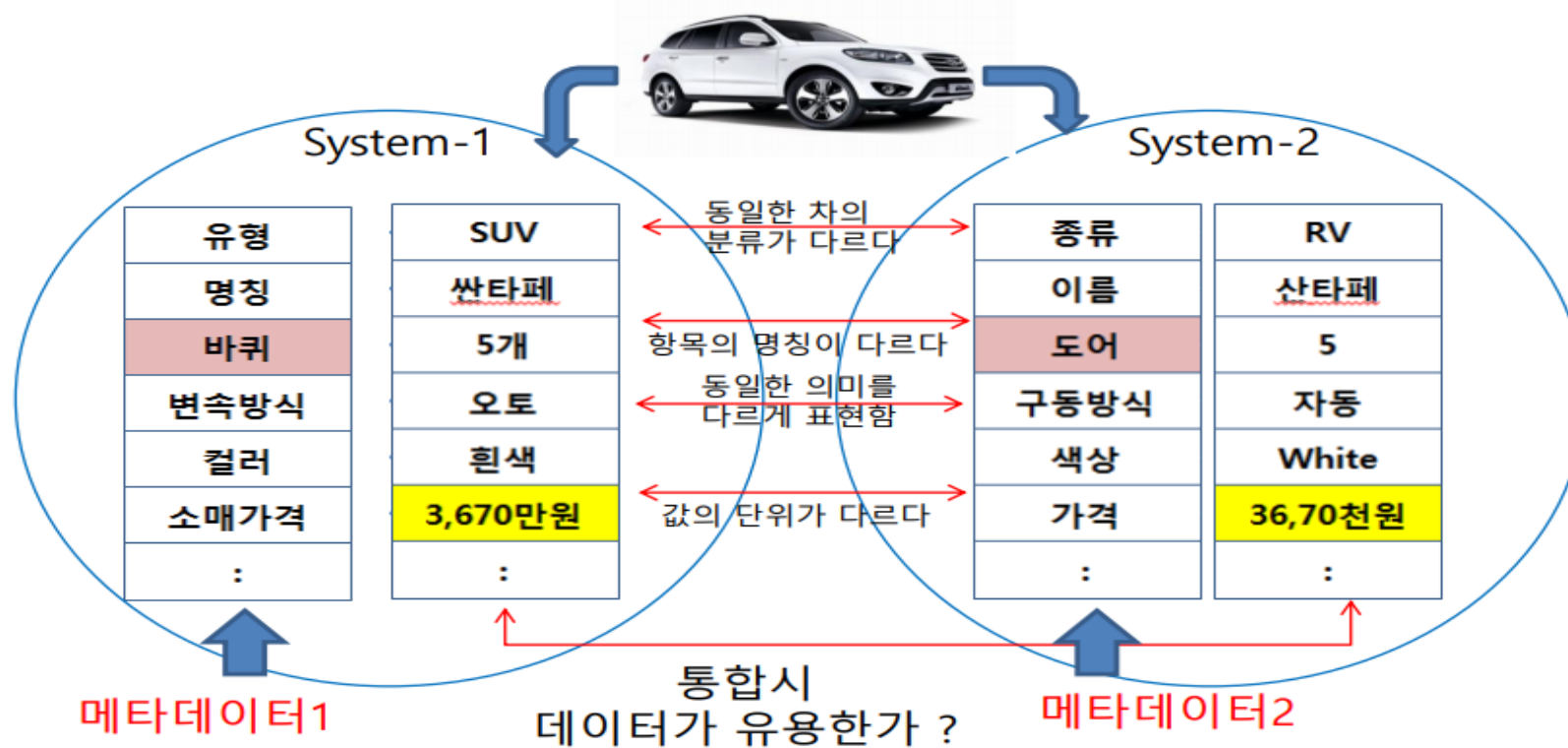
- 조직에서 관심을 갖는 데이터의 특성을 요약한 데이터
- 데이터 항목들의 이름, 생성 및 저장의 위치, 추출 및 통합 과정과 방법, 품질, 조직에서의 가치, 다른 데이터와의 연관성, 프라이버시 등의 정보를 포함



Meta data for Big Data ?
Knowing what we have ?
False discoveries 방지

2. 분석방법론

메타 데이터(metadata)의 불일치 => 정보 통합, 분석, 유통의 걸림돌



2. 분석방법론



43

데이터분석 단계

① 분석용 데이터 준비

- 비즈니스 룰 확인: 분석 계획 단계에서 비즈니스 이해, 도메인 문제점인식, 프로젝트 정의 등을 이용하여 프로젝트의 목표를 정확하게 인식
- 분석용 데이터셋 준비: 데이터 스토어로부터 분석에 필요한 정형·비정형 데이터를 추출한다.

2. 분석방법론



② 텍스트 분석

웹 페이지 데이터, 로그 데이터, 텍스트 자료 등 비정형·반정형 텍스트 데이터를 이용하여 어휘/구문 분석(Word Analysis), 감성 분석(Sentimental Analysis), 토픽 분석(Topic Analysis), 오피니언 분석(Opinion Analysis), 소셜 네트워크분석(SNA) 등을 실시하여 텍스트로부터 분석 목적에 맞는 적절한 모델을 구축한다.

③ 탐색적 분석(EDA)

다양한 관점 별로 기초 통계량을 산출하고 데이터의 분포와 변수 간의 관계 등 데이터 특성 및 데이터의 통계적 특성을 이해하고 모델링을 위한 기초 자료로 활용한다. 데이터 시각화는 탐색적 데이터 분석을 위한 도구로 활용한다.

④ 모델링(Modeling)

모델링이란 분석용 데이터를 이용한 가설 설정을 통하여 통계 모델을 만들거나 기계학습을 이용한 데이터의 분류, 예측, 군집 등의 기능을 수행하는 모델을 만드는 과정이다.

— 데이터 분할

모델링을 효과적으로 진행하기 위해서는 모델링 전에 데이터셋을 훈련용과 테스트용으로 분할함으로써 모델의 과적합을 방지하거나 모델을 일반화에 이용된다.

— 데이터 모델링

기계학습 등을 이용한 데이터 모델링은 훈련용 데이터를 활용하여 분류, 예측, 군집 등의 모델을 만들어 가동중인 운영 시스템에 적용한다. 또한 필요시 비정형 데이터 분석결과를 통합적으로 활용하여 프로젝트 목적에 맞는 통합 모델링을 수행한다.

—모델 적용 및 운영 방안

모델을 가동중인 운영시스템에 적용하기 위해서는 모델에 대한 상세한 알고리즘 작성이 필요하다. 필요시 의사 코드 수준의 작성이 필요할 수도 있다.

⑤ 모델 평가 및 검증

프로젝트 정의서의 모델 평가 기준에 따라 모델을 객관적으로 평가하고 품질관리 차원에서 모델 평가 프로세스를 진행한다. 모델의 실제 적용성을 검증하기 위하여 검증용 데이터(실제 활용되는 데이터)를 이용해 모델 검증 작업을 실시하고 모델링 검증 보고서를 작성한다.

3. 분석과제 발굴

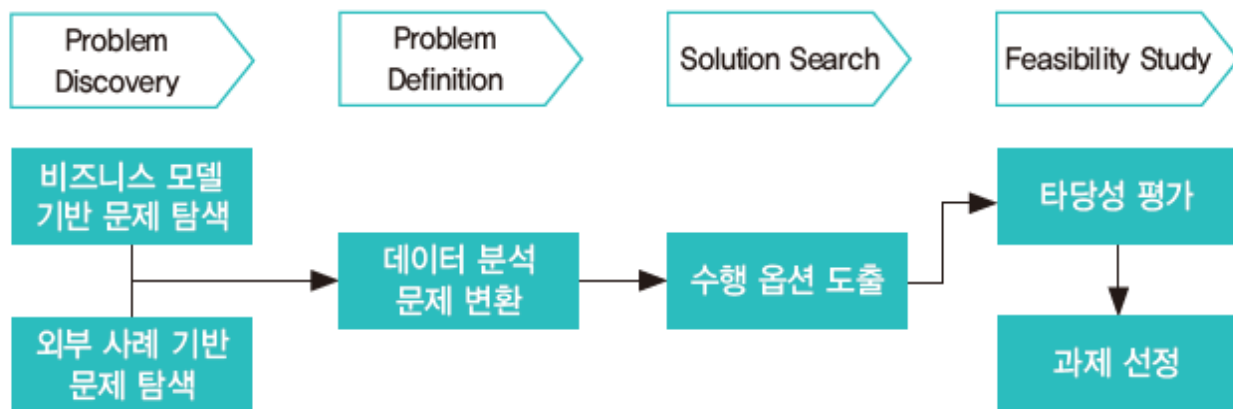


44

하향식 접근 방식 vs 상향식 접근 방식

– 하향식 접근 방식(Top Down Approach)

현황분석 또는 인식된 문제점, 전략으로부터 기회나 문제를 탐색(Problem Discovery), 해당 문제를 데이터 문제로 정의(Problem Definition) 한 후 해결방안 탐색(Solution Search) 그리고 데이터 분석의 타당성 평가(Feasibility Study)를 거쳐 분석 과제를 도출하는 과정으로 이루어 진다.



하향식 접근 방법

3. 분석과제 발굴



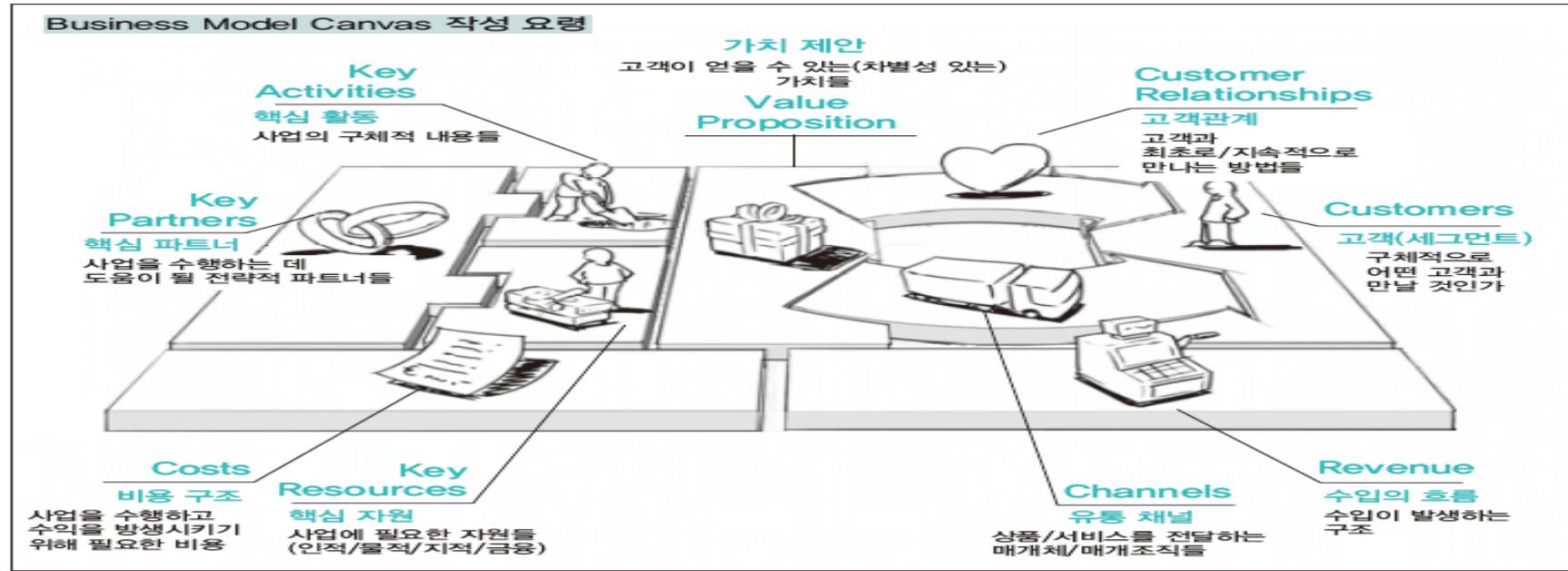
문제탐색(Problem Discovery) 단계
전체적인 관점의 기준 모델을 활용하여 문제를 도출하고 식별하는 단계

기업 내.외부 환경을 포괄하는 비즈니스 모델과 외부참조 모델이 있다.

①비즈니스 모델 기반 문제 탐색

기업 내.외부 환경을 포괄하는 비즈니스 모델이라는 Frame을 활용하여 가치가 창출될 문제를 누락 없이 도출
기업의 사업 모델을 도식화한 비즈니스모델 캔버스의 9가지 블록을 단순화
업무(Operation), 제품(Product), 고객(Product) 단위로 문제 발굴
이를 관리하는 두 가지의 영역인 규제와 감사(Audit & Regulation)영역
지원 인프라(IT & Human Resource)영역 추가도출

3. 분석과제 발굴

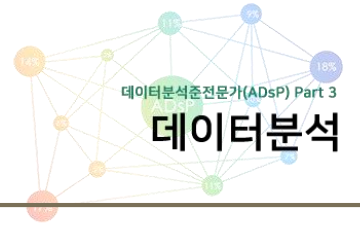


비즈니스 모델 캔버스 작성 요령

출처: Business Model Generation

- ① 업무(Operation); 예) 생산공정 최적화, 재고량 최소화 등
- ② 제품(Product); 예) 제품의 주요 기능 개선
- ③ 고객(Customer); 예) 고객 전화 대기 시간 최소화
- ④ 규제와 감사(Regulation & Audit)
- ⑤ 지원 인프라(IT & Human Resource)

3. 분석과제 발굴



② 분석 기회 발굴의 범위 확장

- 거시적 관심의 요인: STEEP로 요약되는 사회(Social), 기술(Technological), 경제(Economic), 환경(Environmental), 정치(Political) 영역으로 나누어서 좀 더 폭넓게 기획 탐색을 수행한다.
- 경쟁자 확대 관점: 사업영역의 직접 경쟁사영역 및 제품·서비스의 대체재영역과 신규 진입자 영역 등으로 관점을 확대하여 위협이 될 수 있는 상황에 대한 분석 기회 발굴의 폭을 넓혀서 탐색한다.
- 시장의 니즈 탐색: 고객 영역과 고객과 접촉하는 역할을 수행하는 채널 영역 및 고객의 구매와 의사결정에 영향을 미치는 영향자들 영역에 대한 관점을 바탕으로 분석 기회를 탐색한다.
- 역량의 재해석 관점: 내부역량 영역 뿐만 아니라 해당 조직의 비즈니스에 영향을 미치는 파트너 네트워크 영역을 포함한 활용 가능한 역량을 토대로 폭넓은 분석 기회를 탐색한다.

3. 분석과제 발굴



(3) 외부 참조 모델 기반 문제 탐색

유사·동종 사례 벤치마킹을 통한 분석 기회 발굴은 제공되는 산업별, 업무 서비스별 분석 테마 후보 그룹을 통해 “Quick & Easy” 방식으로 필요한 분석 기회가 무엇인지에 대한 아이디어를 얻고 기업에 적용할 분석 테마 후보 목록을 워크숍 형태의 브레인스토밍을 통해 빠르게 도출하는 방법이다. 특히 현재 환경에서는 데이터를 활용하지 않은 업종 및 업무 서비스가 사실상 존재하지 않기 때문에 업무에 활용되는 사례들을 발굴하여 자사의 업종 및 업무 서비스에 적용할 수 있다.

(4) 분석 유즈 케이스(Analytics Use Case) (★기출)

분석 유즈 케이스는 풀어야 할 문제에 대한 상세한 설명 및 해당 문제를 해결했을 때 발생하는 효과를 명시함으로써 향후 데이터 분석 문제로의 전환 및 적합성 평가에 활용하도록 한다.

업무	분석 유즈 케이스	설명	효과
재무	자금 시재 예측	일별로 예정된 자금 지출과 입금을 추정	자금 과부족 현상 예방, 자금 운용 효율화
	구매 최적화	구매 유형과 구매자별로 과거 실적과 구매 조건을 비교·분석하여 구매 방안 도출	구매 비용 절감

분석 유즈 케이스 예시

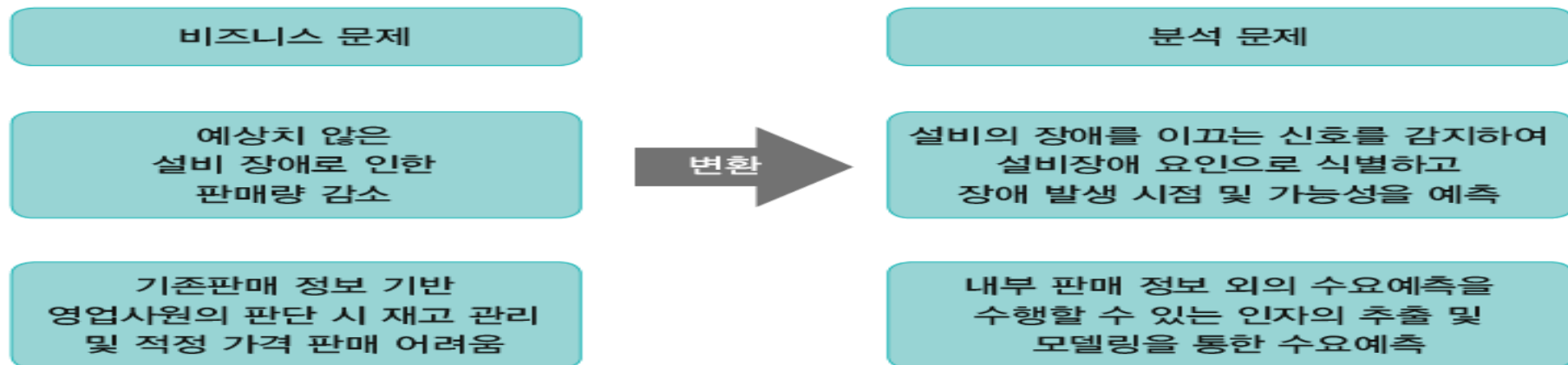
3. 분석과제 발굴



2) 문제 정의(Problem Definition) (★★가출)

식별된 비즈니스 문제를 데이터의 문제로 변환하여 정의하는 단계이다. 필요한 데이터 및 기법을 정의하기 위한 데이터 분석의 문제로의 변환을 수행하게 된다. 앞서 수행한 문제 탐색의 단계가 무엇을(What) 어떤 목적으로(Why) 수행해야 하는지에 대한 관점이었다면 문제 정의(Problem Definition) 단계에서는 이를 달성하기 위해서 필요한 데이터 및 기법(How)을 정의하기 위한 데이터 분석의 문제로의 변환을 수행하게 된다.

예) ‘고객 이탈의 증대’라는 비즈니스 문제는 ‘고객의 이탈에 영향을 미치는 요인을 식별하고 이탈 가능성을 예측’하는 데이터 분석 문제로 전환



비즈니스 문제의 분석 문제 변환 예시

3. 분석과제 발굴



3) 해결 방안 탐색(Solution Search) 단계

동일한 데이터 분석 문제라 해도 어떤 데이터 또는 분석 시스템을 사용할 것인지에 따라서 소요되는 예산 및 활용 가능한 도구가 다르기 때문에 다각도로 고려할 필요가 있다.

예) 엑셀 등의 간단한 도구로 분석이 가능한지, 또는 하둡 등 분산병렬처리를 활용한 빅데이터 분석 도구를 활용할지를 결정하거나, 전문 인력 채용을 통한 역량을 확보해야 하는지에 대한 사전 검토를 수행한다.

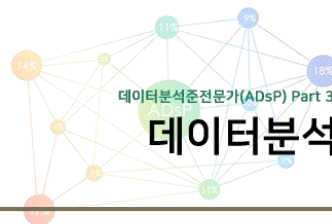
3. 분석과제 발굴



4) 타당성 검토(Feasibility Study) 단계

- ① 경제적 타당도: 비용 대비 편익 분석 관점의 접근
- ② 데이터 및 기술적 타당도: 데이터 분석에는 데이터 존재 여부, 분석 시스템 환경, 그리고 분석 역량이 필요하다.

3. 분석과제 발굴



2. 상향식 접근 방식(Bottom Up Approach) (★★★기출)

문제의 정의 자체가 어려운 경우 데이터를 기반으로 문제의 재정의 및 해결 방안을 탐색하고 이를 지속적으로 개선하는 방식이며, 일반적으로 상향식 접근 방식의 데이터 분석은 비지도학습(Unsupervised Learning) 방법에 의해 수행된다.

통계적 분석에서는 인과관계(원인과 결과) 분석을 위해 가설을 설정하고 이를 검정하기 위해 모집단으로 표본을 추출하고, 그 표본을 이용한 가설검정을 실시하는 방식으로 문제를 해결한다. 그러나 빅데이터 환경에서는 이와 같은 논리적인 인과관계 분석뿐만 아니라 상관관계 분석 또는 연관분석을 통하여 다양한 문제 해결에 도움을 받을 수 있다. 즉 인과관계로부터 상관관계 분석으로의 이동이 빅데이터 분석에서의 주요 변화라고 할 수 있다.

3. 분석과제 발굴



3) 시행착오를 통한 문제 해결

- 하향식 접근방식은 문제가 정형화되어 있고 문제해결을 위한 데이터가 완벽하게 조직에 존재할 경우에 효과
 - 프로토타이핑 접근법은 사용자가 요구사항이나 데이터를 정확히 규정하기 어렵고 데이터 소스도 파악하기 어려운 상황에서 반복적으로 개선해 나가는 방법
- 대표적 상향식 접근방법

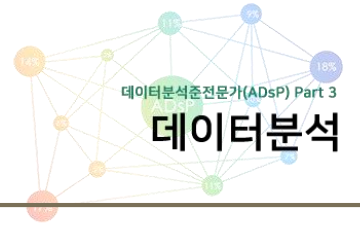
3. 분석과제 발굴



프로토타이핑에 대한 학자들의 다양한 정의

뷰라키 홀그림슨 Bjarki Hallgrímsson	제품 디자인과정에서 신제품이 어떻게 다루어 질지 연구실험하기 위해 실물을 제작하는 것
존 에드슨 John Edson	제품이나 서비스가 어떤 모습일지를 가장 신속하고 간단하게 만들어보는 것
타드 자키 워플 Tod Zaki Warfel	최종적인 시스템의 시뮬레이션, 혹은 대표 모델
칼 울리히 Karl T. Ulrich	제품 전체에서 하나 이상의 관심 있는 영역을 유사하게 만드는 것
툼 캘리 Tom Kelley	이노베이션 과정 의 한 단계이면서, 전망이 불투명한 상황 에서도 앞으로 나아가게 해주는 실질적인 과정
조너던 리트맨 Jonathan Littman	아이디어를 가능한 한 빨리 구체적이고 가시적인 형태로 만들어내는 것
캔 우드 Ken Wood	진화유도
벨라마틴 Bella Martin	아이디어를 실험하고 발전 시켜나가기 위해 다양한 완성도의 물건을 실제로 제작하는 방법
존 콜코 Jon Kolko	해결책 영역을 탐구하고 아이디어에 대해 소통하기 위한 방법

3.분석과제 발굴



하향식 접근법 : 해법을 찾아라

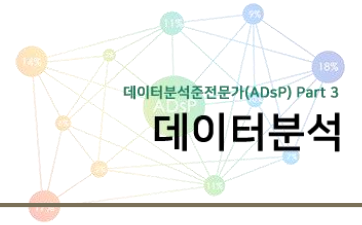
상향식 접근법 : 데이터가 말하게 하라

프로토타이핑 접근법 : 시행착오로 찾아라

빅 데이터 분석 환경에서 프로토타이핑 필요성

- ①문제에 대한 인식 수준:문제의 정의가 불명확, 이전에 접해 보지못한 새로운 문제
- ②필요데이터 존재 여부의 불확실성: 문제해결을 위해 필요한 데이터의 집합이 모두 존재하지 않을 경우
- ③데이터 사용 목적의 가변성:데이터의 가치는 사전에 정해진 수집목적에 따라 확정되는 것이 아니라 지속적으로 변화 할 수 있다.

3.분석과제 발굴



3. 분석 과제 정의

분석 과제 정의서는 향후 프로젝트 수행 계획의 입력물로 사용되기 때문에 수행하는 이해관계자가 프로젝트의 방향을 설정하고, 성공 여부를 판별할 수 있는 주요한 자료를 명확하게 작성하여야 한다.

분석명		분석 정의	
해지 상담 접촉 패턴 분석		기 해지 계약 건 발생 고객의 해지 시점 상담 정보 분석을 통해 해지 고객의 상담 특성을 발굴하는 분석	
소소 데이터	데이터 획득 난이도	분석 방법	
접속 채널, 건수, 접촉 평균시간 이후 해지까지 시간 상담인력 업무 능숙도	하	해지로 이어지는 해지 상담의 유의미한 속성을 요인분석을 통해 발굴하고, 클러스터링 분석을 통해 영향 요인을 그룹핑하고, 그룹핑된 요인 그룹이 해지에 미치는 영향도를 회귀분석함	
분석 적용 난이도	분석 적용 난이 사유	분석 주기	분석 결과 검증 소유
중	접촉 로그 등의 비구조적 데이터 분석 필요	월별	전략기획팀

분석 과제 정의서 예시

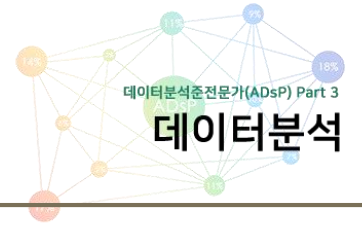
4. 분석 프로젝트 관리 방안



■ 분석 과제의 주요 5가지 특성 관리 영역 (★★기출)

Data Size	분석하고자 하는 데이터양을 고려한 관리 방안 수립이 필요
Data Complexity	다양한 유형의 데이터가 통합해서 분석 프로젝트를 진행할 때는 해당 데이터에 잘 적용될 수 있는 분석 모델의 선정을 고려해야 한다.
Speed	배치 형태인지 실시간 분석의 형태인지를 고려해야 한다.
Analytic Complexity	분석 모델이 복잡할수록 정확도는 올라가지만 해석이 어려워지는 단점이 존재하므로 이에 대한 기준점을 사전에 정의해두어야 한다.
Accuracy & Precision	Accuracy는 모델과 실제값 사이의 차이가 적다는 정확도를 의미한다. Precision은 모델을 지속적으로 반복했을 때의 편차의 수준으로서 일관적으로 동일한 결과를 제시한다는 것을 의미한다.

4. 분석 프로젝트 관리 방안



프로젝트 영역별 주요 관리 항목

주제 그룹	개념 및 관련 프로세스
통합	프로젝트와 관련된 다양한 활동과 프로세스를 도출, 정의, 결합, 단일화, 조정, 통제, 종료에 필요한 프로세스
이해관계자	프로젝트 스폰서, 고객사, 기타 이해관계자를 식별하고 관리하는 데 필요한 프로세스
범위	작업과 인도물을 식별하고 정의하는 데 요구되는 프로세스
자원	인력, 시설, 장비, 자재, 기반 시설, 도구와 같은 적절한 프로젝트 자원을 식별하고 확보하는 데 필요한 프로세스
시간	프로젝트 활동의 일정을 수립하고 일정 통제의 진척상황을 관찰하는 데 요구되는 프로세스
원가	예산개발과 원가통제의 진척상황을 관찰하는 데 요구되는 프로세스를 포함
리스크	위험과 기회를 식별하고 관리하는 프로세스
품질	품질보증과 품질통제를 계획하고 확립하는 데 요구되는 프로세스
조달	계획에 요구된 프로세스를 포함하며 제품 및 서비스 또는 인도물을 인수하고 공급자와의 관계를 관리하는 데 요구되는 프로세스
의사소통	프로젝트와 관련된 정보를 계획, 관리, 배포하는 데 요구되는 프로세스

시간: Time boxing 기법-> 반복형 개발에 쓰이는 프로젝트관리기법 반복할당된 작업을 완수하지 못했을 경우 다음 과정으로 넘긴다.

5. 기출문제



Q1

빅데이터 분석 방법론 중 추가적인 데이터 확보가 필요한
경우 반복적인 피드백을 수행하는 구간은 어디인가?

- ① 분석 기획~데이터 준비
- ② 데이터 준비~데이터 분석
- ③ 데이터 분석~데이터 구현
- ④ 시스템 구현~평가 및 전개

5. 기출문제



Q2

다음 중 비즈니스 모델 관점 이외에 경쟁자 관점의 영역에 해당되지 않는 것은?

- ① 대체재(Substitute) 영역
- ② 경쟁자(Competitor) 영역
- ③ 신규진입자(New Entrant) 영역
- ④ 소비자(Consumer) 영역

5. 기출문제



Q3

다음 중 분석 과제 접근 방법이 틀린 것은?

- ① 문제가 확실할 때 상향식 접근법을 사용한다.
- ② 문제가 주어지고 해법을 찾기 위해서 하향식 접근 방식을 사용한다
- ③ 문제의 정의 자체가 어려운 경우 상향식 접근 방식을 사용한다.
- ④ 디자인 사고(Design Thinking)의 경우 상향식과 하향식을 반복적으로 사용하기 쉽다

5. 기출문제



Q4

데이터 분석 방법론 중 KDD는 데이터 전처리(Preprocessing)에 분석 대상용 데이터셋에 포함되어 있는 잡음(Noise)과 이상치(Outlier), 결측치(Missing Value)를 식별하고 필요시 제거하거나 의미 있는 데이터로 처리하는 데이터셋 정제작업을 시행한다.

또 다른 분석 방법인 CRISP-DM 분석 방법론 중 이와 유사한 프로세스 단계는 무엇인가? ?

5. 기출문제



Q5

**비즈니스 모델 관점에서 해당 기업의 사업 모델을 도식화한
비즈니스 모델 캔버스 구성 단위는?**

- ① 업무, 제품, 고객, 지원 인프라, 규제&감사
- ② 범위, 의사소통, 이해관계자, 자원, 조달
- ③ 범위, 이해관계자, 자원, 시간, 원가
- ④ 리스크, 품질, 조달, 의사소통, 통합

5. 기출문제



Q6

프로토타이핑 접근법에 대한 설명 중 적절하지 않은 것은?

- ① '사용자들이 이렇게 만들면 편하게 사용할 거야' 라는 가설을 생성하게 된다.
- ② 특정 가설을 갖고 서비스를 설계하고 디자인에 대한 실험도 실행한다.
- ③ 프로토타입을 보고 완성시킨 결과물을 통해 가설을 확인할 수 있다.
- ④ 프로토타이핑은 순환적 문제탐색으로 인해 개발시간이 많이 필요로한다.

5. 기출문제



Q7

동일한 사안이라고 해도 제시되는 방법에 따라 그에 관한 해석이나 의사결정이 달라지는 인식의 왜곡현상을 무엇이라 하는가?

5. 기출문제



Q8

프로토타이핑 기법의 내용으로 적절한 것은?

- ① 신속하게 해결책 모형 제시, 상향식 접근방법에 활용
- ② 빠른 결과보다 모델의 정확성에 중점을 둔 기법
- ③ 워터폴 방식처럼 전체적인 플랜을 짜고 문서를 통해 개발한다.
- ④ 대표적인 하향식 접근방법 기법이다.

5. 기출문제



Q9

정형-반정형-비정형 데이터 순서로 된 것은?

- ① Demand Forecast-Competitor- pricing-Email record
- ② RFID-sensor-loyalty
- ③ SNS media-sensor-ERP
- ④ Bar code system-internet of things sensing-Email

데이터분석준전문가(ADsP) Part 2

데이터 분석 기획

01

2장 분석 마스터 플랜

1. 마스터 플랜 수립
2. 분석 거버넌스 체계 수립



1.마스터 플랜 수립



1. 분석 마스터 플랜 수립 프레임워크

중·장기적 관점의 마스터플랜 수립을 위해서는 분석 과제를 대상으로 **전략적 중요도**, **비즈니스 성과 및 ROI(투자회수율)**, **분석 과제의 실행 용이성** 등 다양한 기준을 고려해 **적용할 우선순위를 설정**

분석을 업무에 내재화할 것인지, 별도의 분석 화면으로 일단 적용할 것인지, **분석 데이터**를 내부의 데이터로 한정할 것인지, 외부의 데이터까지 포함할 것인지, **분석 기술은 어느 기술요소까지** 적용할 것인지 등 **분석의 적용범위 및 방식에** 대해서도 종합적으로 고려하여 데이터 분석을 실행하기 위한 로드맵을 수립한다.

1.마스터 플랜 수립



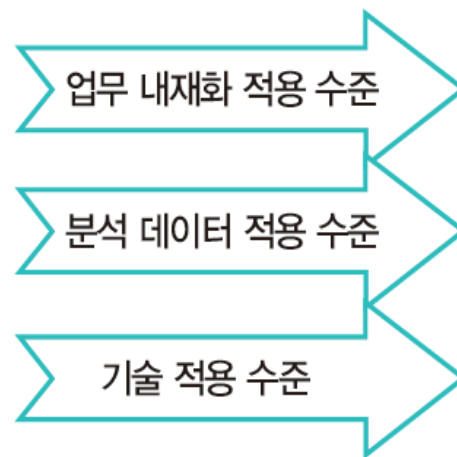
1)분석 마스터 플랜 수립 프레임워크

우선순위 고려 요소



적용 우선순위
결정

적용 범위/방식 고려 요소



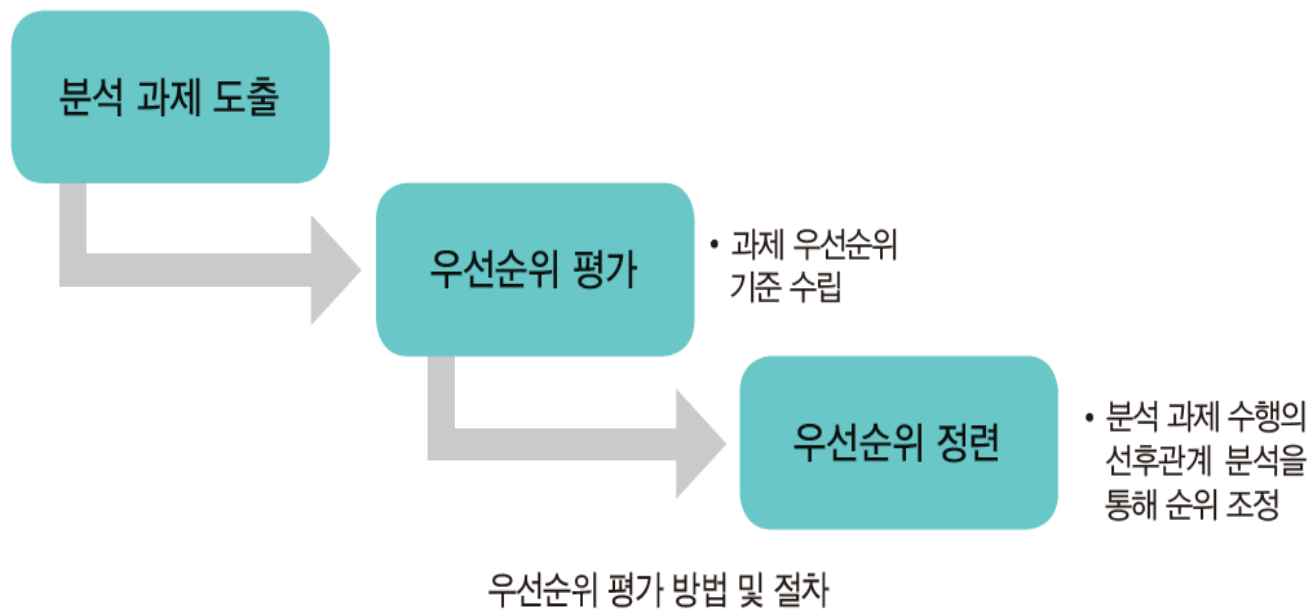
분석 구현
로드맵 수립

마스터플랜 수립 개요 (★기출)

1.마스터 플랜 수립



2) 수행과제 도출 및 우선순위 평가



1.마스터 플랜 수립



3) ROI 관점에서의 분석과제 우선순위 평가 기준

시급성: 판단 기준은 전략적 중요도가 핵심이며, 이는 전략적 중요도가 시점에 따라 시급성 여부를 고려할 수 있다는 뜻이다. 예) 현재는 미래보다 시급성이 높다.

난이도: 현시점에서 과제를 추진하는 것이 비용과 범위 측면을 고려했을 때 바로 적용하기 쉬운 것인지 또는 어려운 것인지를 판단하는 것이다.



1.마스터 플랜 수립



4) 포트폴리오 사분면 분석을 통해 과제 우선순위 선정 기법

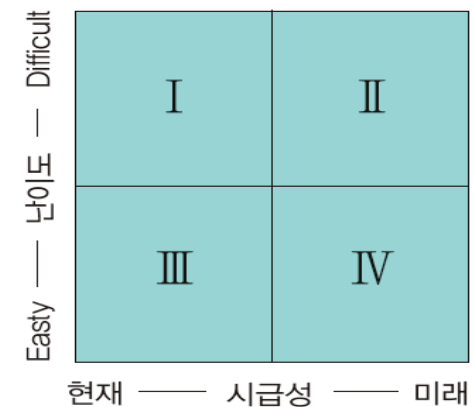
우선순위 기준을 난이도와 시급성을 고려하여 우선 추진해야 하는 분석 과제와 제한된 자원을 고려하여 단기적 또는 중장기적으로 추진해야 하는 분석 과제 등을 4가지 유형

분석 과제의 적용 우선순위를 '시급성'에 둔다면 Ⅲ→Ⅳ→Ⅱ 영역순이며, 난이도를 기준으로 둔다면 Ⅲ→Ⅰ→Ⅱ 영역순으로 의사결정을 할 수 있다

1.마스터 플랜 수립



5) 포트폴리오 사분면 분석을 통해 과제 우선순위 선정 기법



분석 과제 우선순위 선정 매트릭스

I 사분면: 전략적 중요도가 높아 경영에 미치는 영향이 크므로 현재 시급하게 추진이 필요함. 난이도가 높아 현재 수준에서 과제를 바로 적용하기에 어려움.

II 사분면: 현재 시점에서는 전략적 중요도가 높지 않지만 중장기적 관점에서는 반드시 추진되어야 함. 분석 과제를 바로 적용하기에는 난이도가 높음.

III 사분면: 전략적 중요도가 높아 현재 시점에 전략적 가치를 두고 있음. 과제 추진의 난이도가 어렵지 않아 우선적으로 바로 적용 가능할 필요성이 있음.

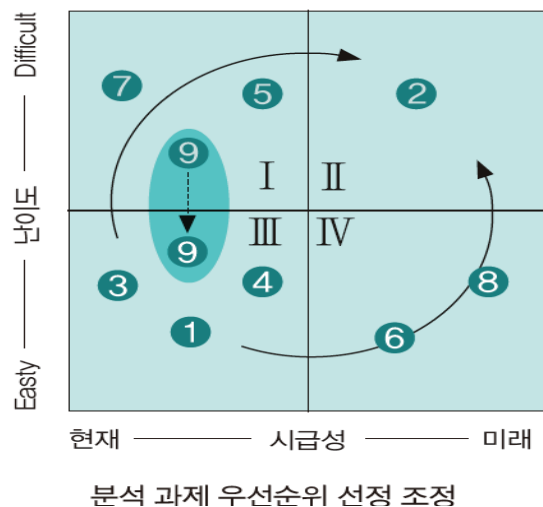
IV 사분면: 전략적 중요도가 높지 않아 중장기적 관점에서 과제 추진이 바람직함. 과제를 바로 적용하는 것은 어렵지 않음.

1.마스터 플랜 수립



6) 포트폴리오 사분면 분석을 통해 과제 우선순위 선정 기법

- ① 시급성과 난이도가 높은 1사분면은 경영진 또는 실무 담당자의 의사결정에 따라 적용 우선순위를 조정할 수 있다.
- ② 기술적 요소에 따라서도 적용 우선순위를 조정할 수 있다.
- ③ 분석 범위에 따라서도 적용 우선순위를 조정할 수 있다.



분석 과제 우선순위 선정 조정

2.분석 거버넌스 체계 수립



1) 거버넌스 체계 개요

거버넌스(Governance)는 정부라는 의미의 Government와 같은 어원을 가지고 있는 단어 인데, Government보다는 더 폭넓은 의미로 진화하여 기업, 비영리 기관 등에서 규칙, 규범 및 행동이 구조화되고, 유지되고, 규제되고 책임을 지는 방식 및 프로세스를 지칭한다.

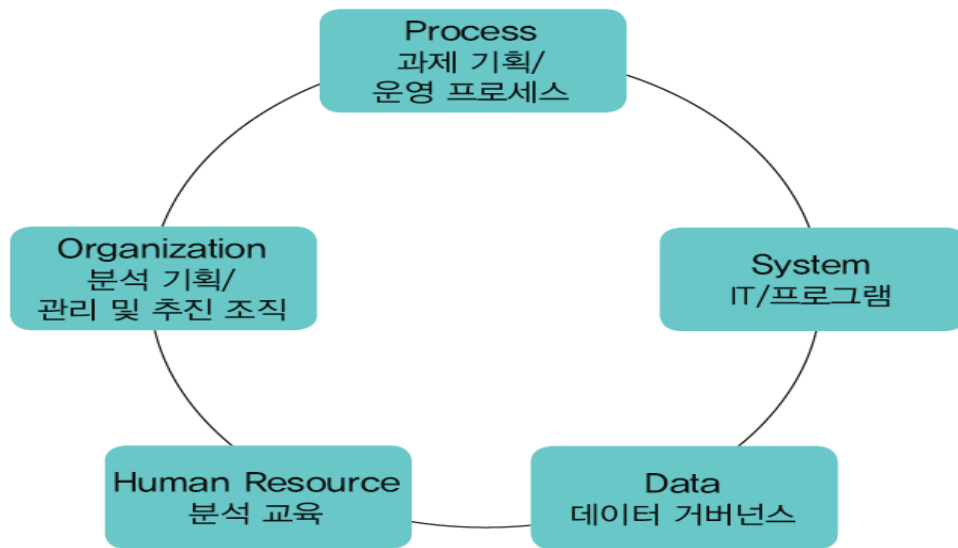
분석 거버넌스는 기업에서 데이터가 어떻게 관리되고, 유지되고, 규제되는지에 대한 내부적인 관리 방식이나 프로세스를 의미한다.

데이터 거버넌스는 데이터의 품질보장, 프라이버시 보호, 데이터 수명관리, 전담조직과 규정정립, 데이터 소유권과 관리권 명확화 등을 통하여 데이터가 적시에 필요한 사람에게 제공되도록 체계를 확립하는 것이다

2.분석 거버넌스 체계 수립



2) 분석 거버넌스 체계 구성 요소



분석 거버넌스 체계 구성 요소

2.분석 거버넌스 체계 수립

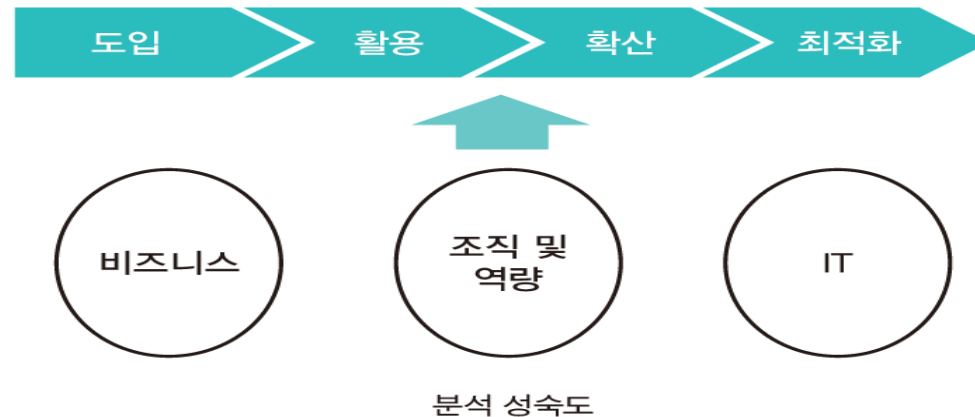


3) 데이터분석 성숙도 모델 및 수준 진단

데이터 분석의 수준 진단을 통해 데이터 분석 기법을 구현하기 위해 무엇을 준비하고 보완해야 하는지 등 분석의 유형 및 분석의 방향성을 결정할 수 있다.



분석 준비도



2.분석 거버넌스 체계 수립



4) 분석준비도 진단

분석 준비도는 기업의 데이터 분석 도입의 수준을 파악하기 위한 진단방법으로서 총 6가지 영역을 대상으로 현 수준을 파악
진단 결과 전체 요건 중 일정 수준 이상 충족하면 분석 업무를 도입하고, 충족하지 못하면 먼저 분석 환경을 조성.

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• 관리자등의 기본적 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• 관리자의 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

2.분석 거버넌스 체계 수립



5) 분석 성숙도 모델

분석 성숙도 진단은 비즈니스 부문, 조직·역량 부문, IT 부문 등 3개 부문을 대상으로 성숙도 수준에 따라 도입 단계, 활용 단계, 확산 단계, 최적화 단계로 구분해 살펴볼 수 있다

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	<ul style="list-style-type: none"> ✓ 실적분석 및 통계 ✓ 정기보고 수행 ✓ 운영 데이터 기반 	<ul style="list-style-type: none"> ✓ 미래 결과 예측 ✓ 시뮬레이션 ✓ 운영 데이터기반 	<ul style="list-style-type: none"> ✓ 전사 성과 실시간 분석 ✓ 프로세스혁신 3.0 ✓ 분석규칙 관리 ✓ 이벤트 관리 	<ul style="list-style-type: none"> ✓ 외부환경 분석 활용 ✓ 최적화 업무 적용 ✓ 실시간 분석 ✓ 비즈니스 모델 진화
조직·역량 부문	<ul style="list-style-type: none"> ✓ 일부 부서에서 수행 ✓ 담당자 역량에 의존 	<ul style="list-style-type: none"> ✓ 전문 담당부서에서 수행 ✓ 분석 기법 도입 ✓ 관리자가 분석 수행 	<ul style="list-style-type: none"> ✓ 전사 모든 부서 수행 ✓ 분석 COE 조직 운영 ✓ 데이터 사이언티스트 확보 	<ul style="list-style-type: none"> ✓ 데이터 사이언스 그룹 ✓ 경영진 분석 활용 ✓ 전략 연계
IT 부문	<ul style="list-style-type: none"> ✓ 데이터 웨어하우스 ✓ 데이터 마트 ✓ ETL/EAI ✓ OLAP 	<ul style="list-style-type: none"> ✓ 실시간 대시보드 ✓ 통계분석 환경 	<ul style="list-style-type: none"> ✓ 빅데이터 관리 환경 ✓ 시뮬레이션·최적화 ✓ 비주얼 분석 ✓ 분석 전용 서버 	<ul style="list-style-type: none"> ✓ 분석 협업 환경 ✓ 분석 Sandbox ✓ 프로세스 내재화 ✓ 빅데이터 분석

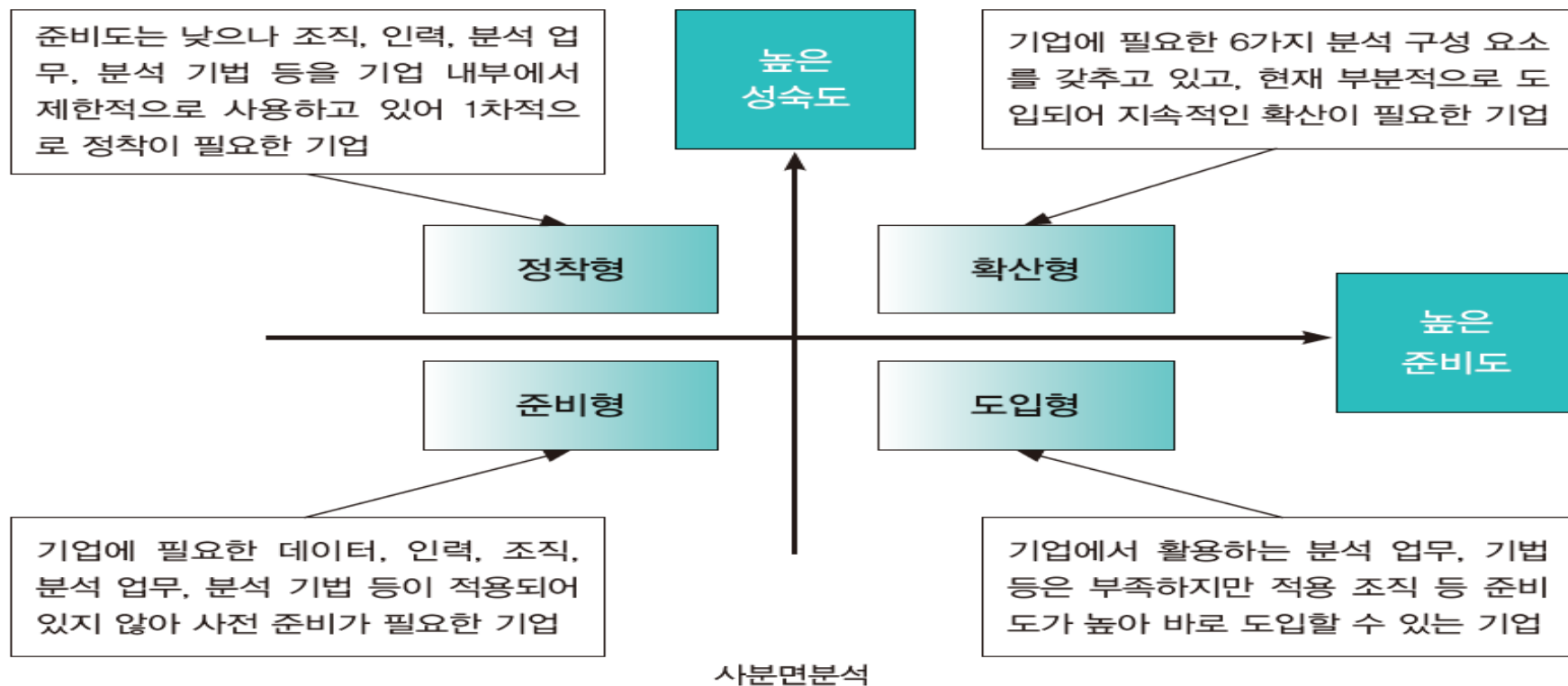
분석 성숙도 모델 (★★★)

2.분석 거버넌스 체계 수립



6) 분석 수준 진단 결과

분석 준비도와 성숙도 진단 결과를 토대로 해당 기업의 분석 수준을 객관적으로 파악



2.분석 거버넌스 체계 수립



7) 데이터 거버넌스 체계 수립

데이터 거버넌스란 전사 차원의 모든 데이터에 대하여 정책 및 지침, 표준화, 운영조직 및 책임 등의 표준화된 관리 체계를 수립하고 운영을 위한 프레임워크 및 저장소를 구축하는 것을 말한다

1) 데이터 거버넌스 구성 요소 (★기출)

- ① 원칙: 데이터를 유지 관리하기 위한 지침과 가이드 및 보안, 품질 기준, 변경 관리
- ② 조직: 데이터를 관리할 조직의 역할과 책임 및 데이터 관리자, 데이터 아키텍트
- ③ 프로세스: 데이터 관리를 위한 활동과 체계 및 작업 절차, 모니터링 활동

2) 데이터 거버넌스 체계 요소 (★★★기출)

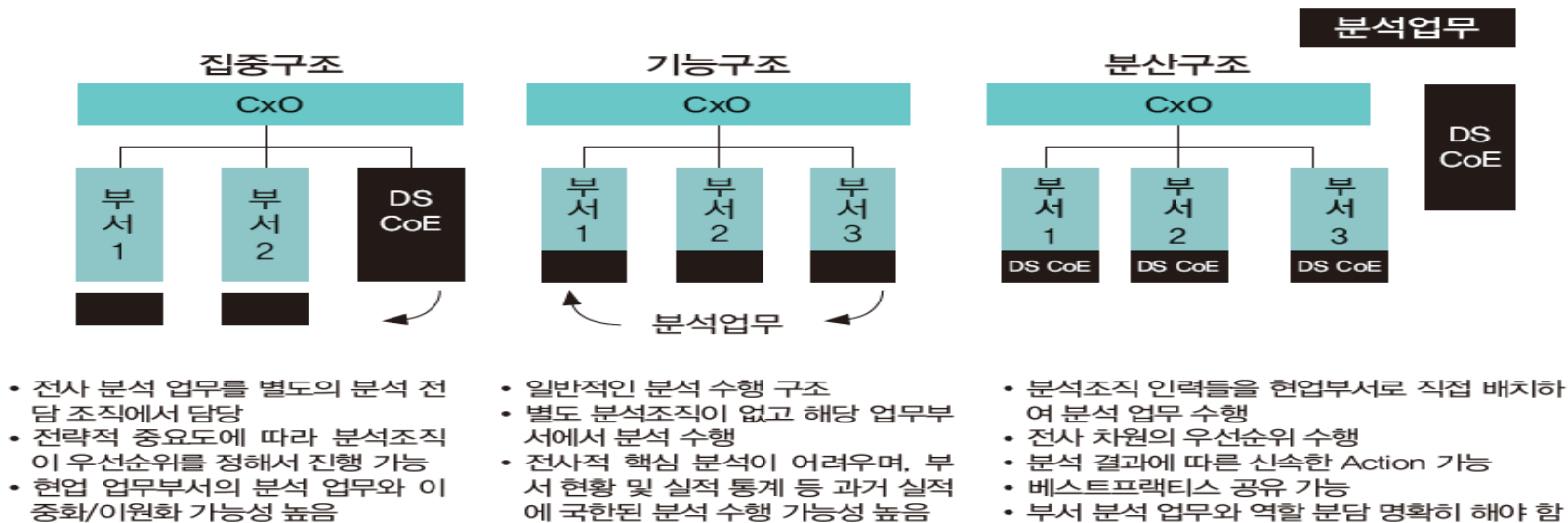
- ① 데이터 표준화 → 데이터 표준 용어 설명, 명명 규칙, 메타 데이터 구축, 데이터 사전 구축
- ② 데이터 관리 체계 → 메타 데이터와 데이터 사전의 관리 원칙을 수립
- ③ 데이터 저장서 관리 → 메타 데이터 및 표준 데이터를 관리하기 위한 전사 차원의 저장소를 구성
- ④ 표준화 활동 → 데이터 거버넌스 체계를 구축한 후 표준 준수 여부를 주기적으로 점검하고 모니터링을 실시

2.분석 거버넌스 체계 수립



8) 데이터 조직 및 인력 방안 수립

비즈니스변화를 인식하고 차별된 경쟁력을 확보하는 수단으로 데이터 과제 발굴, 기술 검토 및 전사 업무 적용 계획 수립 등 데이터를 효과적으로 분석 활용하기 위해 기획 및 운영을 관리할 수 있는 전문 분석조직의 필요성이 제기
특히 분석 업무 수행 주체에 따라 그림과 같이 세 가지 유형의 조직구조



분석조직 구조

3. 기출문제



Q1

분석 과제 평가 및 선정에 대해서 적절하지 않은 것은?

- ① 시급성의 판단 기준은 전략적 중요도가 핵심이다.
- ② 난이도는 현시점에서 과제를 추진하는 것이 비용과 범위 측면에서 쉬운(Easy) 것인 지 또는 어려운(Difficulty) 것인지에 대한 판단 기준이다.
- ③ 시급성이 높고 난이도가 높은 영역은 경영진 또는 실무 담당자의 의사결정에 따라 적용 우선순위를 조정할 수 있다.
- ④ 시급성은 분석 비용 및 데이터 적용 범위에 따라 달라진다.

3. 기출문제



Q2

데이터 분석 준비도 프레임워크 중 분석업무 파악에 관한 항목으로 적절하지 않은 것은? ?

- ① 최적화 분석 업무
- ② 시뮬레이션 분석 업무
- ③ 예측 분석 업무
- ④ 분석 기법 라이브러리

3. 기출문제



Q3 다음은 데이터 분석을 위한 조직 구조에 관한 설명 중 옳바르지 않은 것은?

- ① 집중형 조직 구조는 조직내에 별도의 독립적인 분석 전담조직이 구성하고, 회사의 모든 분석 업무를 전담 조직에서 담당한다.
- ② 집중형 조직 구조는 일부 협업 부서와 분석 업무가 이원화될 가능성이 있다.
- ③ 기능중심의 조직 구조는 별도로 분석 조직을 구성하지 않고 각 해당 업무부서에서 직접 분석하는 형태다.
- ④ 분산된 조직구조는 조직의 인력들을 협업 부서에 배치가 되어 신속한 업무에는 적합하지 않다

3. 기출문제



Q4

분석 프로젝트 관리의 영역이 아닌 것은?

- ① 프로세스 관리
- ② 시간
- ③ 범위
- ④ 품질

5. 기출문제



Q5

분석 거버넌스 체계 고려 사항이 아닌 것은?

- ① 분석 기획 및 조직
- ② 운영 프로세스
- ③ 분석 비용 및 예산
- ④ IT 기술과 마인드 육성체계

5. 기출문제



Q6

다음 중 데이터 분석 구현을 위한 적용 우선순위 평가 시 주요 고려 요소가 아닌 것은?

- ① 전략적 중요도
- ② ROI
- ③ 실행 용이성
- ④ 업무 내재화 적용 수준