

4) 타당성 평가

○ 경제적 타당성

- **비용대비 편익 분석 관점 접근(ROI)** [★] 같은 비용으로 더 이득 보는 방향으로

○ 데이터 및 기술적 타당성

- 비즈니스 지식과 기술적 지식 요구
- 우월한 대안 선택 / 솔루션 방안 포함 / 후속작업(분석과제 정의서 형태) 시행 / 프로젝트 계획의 입력물로 활용

2. 상향식 접근법 (Bottom up Approach)

- 문제의 정의가 어려운 경우, 데이터 기반으로 문제의 재정/해결방안을 탐색하고 지속적으로 개선하는 방식

1) 기존 하향식 접근법의 한계를 극복하기 위한 분석 방법론

- 기존 하향식 접근법은 새로운 문제 탐색에 한계

○ 디자인 사고 접근법 Design Thinking (스탠포드 대학의 d.shcool)

- 미리 답을 내는 것이 아니라 사물 있는 그대로 인식하는 What 관점
- 데이터 그 자체 관찰하고 실제적 행동으로 옮김 - 대상을 좀 더 잘 이해하는 방식으로 접근
- 감정이입(Empathize)을 특히 강조
- EMPATHIZE -> DEFINE -> IDEATE -> PROTOTYPE -> TEST

○ 비지도 학습 방법에 의해 수행 = 비교사학습

CLAP, 연관, 군집, SOM

- 데이터 자체의 결합, 연관성, 유연성 등을 중심으로 데이터 상태 표현
- 목표값을 사전에 정의안함, 해석은 용이하지 않지만 **새로운 유형의 인사이트** 도출에 유용
- 예 : 장바구니 분석, 군집 분석, 기술 통계 및 프로파일링 등 상관관계, 군집분석

○ 지도 학습 방법에 의해 수행 = 교사학습

기술 통계

- 명확한 목적 하에 데이터 분석
- 결과로 도출되는 값에 대하여 사전 인지, 어떠한 데이터를 넣었을 때 **어떠한 결과가 나올지 예측**하는 것
- 예 : 분류, 회귀분석 등

(지도) (비지도)
인과관계 → 상관관계

* 기존 통계적분석에서는 가설 설정, 검정으로 문제를 해결 (인과관계)

현대 빅데이터 환경에서 상관관계분석, 연관분석을 통해 다양한 문제해결 가능

2) 시행착오를 통한 문제 해결

○ 프로토타이핑 접근법

- 요구사항이나 데이터를 정확히 규정 어렵고 데이터 소스도 명확히 파악하기 어려운 상황에서 **일단 분석 시도**하고 그 결과를 확인하면서 **반복적으로 개선**해 나가는 방법
- 신속하게 해결책이나 **모형** 제시 - 문제를 명확하게 인식하고 필요한 데이터 식별하여 구체화하는 **상향식 접근 방법**
- **프로토타이핑 접근법의 기본 프로세스 구성**
 - 가설의 생성, 디자인에 대한 실험, 실제 환경 테스트 결과에서의 통찰도출 및 가설 확인
- 빅데이터 분석 환경에서의 프로토타이핑의 필요성 대두
 - **문제 정의가 불명확**하거나 이전에 접해보지 못한 **새로운 문제**일 경우
 - **필요 데이터 존재 여부의 불확실성**
 - **데이터 사용 목적의 가변성** : 기존 데이터 재검토로 데이터 사용목적과 범위 확대 가능
예) 이동통신사 위치추적 : 사용자의 호출의 효율적 처리 → 특정시간 인구 유동성 분석 정보

3. 분석과제 정의

○ 분석과제 정의서

- 분석별 필요한 소스 데이터, 분석방법, 데이터 입수 및 분석의 난이도 분석 수행주기, 분석결과에 대한 검증 오너십, 상세 분석 과정 등을 정의

(소셜미디어, 오픈데이터)
광범위한
곳에서 수집가능
→ 상세하게 정의한다

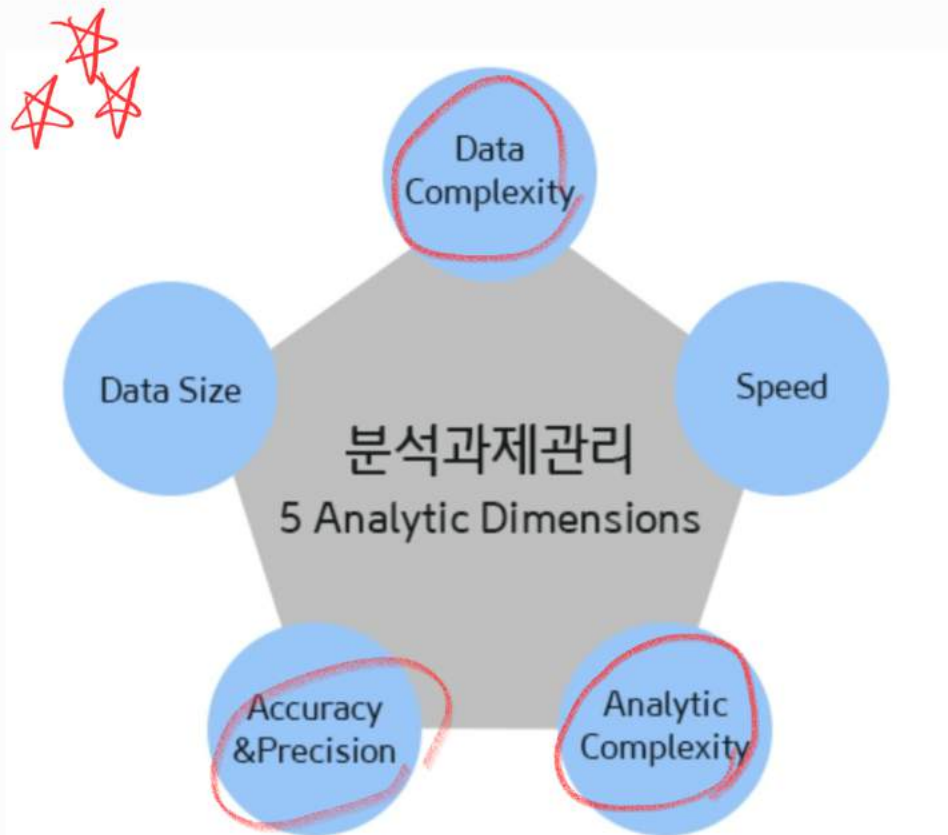
+ 난이도사유 작성

제4절 분석프로젝트 관리 방안

○ 범위, 일정, 품질, 리스크, 의사소통 영역별 관리 +

5가지 주요 속성

(data size, data complexity, speed, analytic complexity, Accuracy & Precision)



- 데이터 크기: 분석하고자 하는 데이터의 양을 고려한 관리 방안 수립

- 데이터 복잡성: 정형보다 복잡한 비정형데이터+원천데이터 고려한 분석 모델 선정

- 속도: 시나리오 측면에서 분석 모델의 성능 및 속도 고려한 개발 및 테스트

- 분석 복잡성: 모델 복잡도 - 정확도 향상 - 해석 어려움 (정확도 복잡도 Trade Off)

- 정확도와 정밀성: 모델과 실제 값 차이가 적다 (정확도, 활용성)

반복시 동일한 결과 제시 (일관성, 안정성) (정확도, 정밀성 Trade Off)

→ 정밀성

→ 과적합, 과소적합 관점에서 생각하면 이해 빠름

정확도 < 복잡도
정밀성

1. 분석 프로젝트의 특성

○ 분석가 목표: 분석의 정확도 + 개별 분석 업무 + 전반적인 프로젝트 관리 필요

○ 분석가 입장: 데이터 영역, 비즈니스 영역 간 분석모델을 통한 조율 수행하는 조정자
지속적 반복 및 개선을 통해 기간 내에 가능한 최선의 결과 도출

2. 분석 프로젝트 관리방안

프로젝트관리 지침의 프로젝트관리체계 (10개 주제그룹, KSA ISO 21500)

- 범위 (Scope) : 분석 진행 과정 및 최종 결과물 형태에 따라 **범위** 빈번하게 변경
- 통합 (Integration) : **관리프로세스들이 통합적으로 운영**되도록 관리
- 자원 (Resource) : 인력, 시설, 장비, 자재, 기반 시설, 도구와 같은 **프로젝트 자원**
- 시간 (Time) : 품질보장된든 전제 하에 Time boxing 기법으로 **일정관리**
- 원가 (Cost) : 데이터, 도구 **비용** 고려 (외부데이터, 상용버전 도구 등 고가 고려)
- 리스크 (Risk) : **위험과 기회**를 식별하고 관리
 - 데이터 미확보 / 데이터 및 분석 알고리즘의 한계 등 **대응방안** 수립
- 품질 (Quality) : **품질목표** 사전 수립, **품질통제 / 품질보증** 나누어 수행
- 조달 (Procurement) : **외부 소싱** 적절하게 운영 (인프라 대신 클라우드 등..)
- 의사소통 (Communication) : 모든 프로젝트 관계자를 위한 **의사소통체계**
- 이해관계자 (Shareholder) : 스폰서, 고객사, 기타 **이해관계자**를 식별 관리

2장 분석마스터플랜

제1절 마스터 플랜 수립

○ 분석과제의 우선순위 평가 → 단계적 로드맵 구성 → 이행계획수립

1. 마스터 플랜 수립 개요 *기출*

★ **우선순위 고려요소** *이거 세개 기억!!!* *어떤것부터 적용할지 정한다*

- **전략적 중요도, 비즈니스 성과/ROI, 실행 용이성** → **적용 우선순위 설정**

○ **적용범위/방식 고려요소**

- **업무내재화 적용 수준, 분석 데이터 적용 수준(내·외부), 기술적용 수준**

→ **분석 구현 로드맵 수립**

* 특히 우선순위 고려요소를 묻는 문제

우선순위를 정한 과제를 어떻게 분석할지 정한다

주관식

* **ISP (Information Strategy Planning) = 방법론 / 정보전략계획**

정보기술,시스템의 전략적 활용을 위해 조직 내,외부 환경을 분석하여 기회나

문제점을 도출하고 사용자의 요구사항을 분석하여 시스템 구축 우선순위를 결정하는

중장기 마스터 플랜을 수립하는 절차 (기업, 정부기관에서 활용)

↪ **분석마스터플랜은 일반적인 ISP 방법론을 활용**

2. 우선순위 평가

○ **우선순위 평가**: 정의된 데이터 과제에 대한 실행 순서를 정하는 것

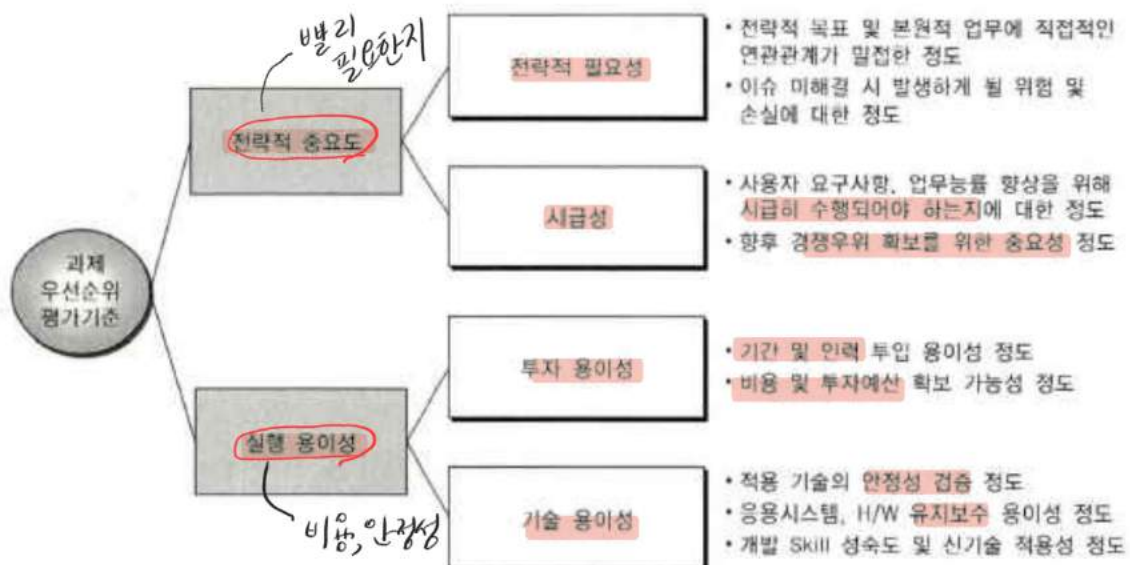
- 업무 영역별 도출된 분석과제를 **우선순위 평가 기준**에 따라 평가

+ **적용 순위 조정**(과제 수행 선후행 관계 고려)

→ 최종 확정

- 각 기업의 중요 가치 기준에 따라 다양한 관점에서의 우선순위 기준 수립하여 평가

* 예시



ROI 관점 빅데이터의 핵심 특성

ROI : 투자하는거에 비해서 얻을 수 있는 것의 비율

- 투자비용(Investment) 측면 요소(ROI관점) (3V)
Volume(규모/양), Variety(종류/유형), Velocity(속도)

자주출제

- 비즈니스 효과(Return) 측면 요소: Value(가치)

크기 (volume), 다양성 (variety), 속도 (velocity), 가치 (Value) = 4V
= 난이도 = 시급성

데이터 분석 과제 추진할 때 고려해야 하는 우선순위 평가 기준

- 난이도 : 데이터 생성, 저장, 가공, 분석하는 비용 & 기업 분석수준 고려
분석 적합성 여부, 해당기업의 상황에 따라 조율가능 → 크기, 다양성, 속도 조정해서
- 시급성 : 전략적 중요도(핵심), 분석과제 목표가치(KPI) 함께 고려하여 판단 난이도 낮음

포트폴리오 사분면 분석을 통한 과제 우선순위 선정 기법

D 난 이 도	I - 전략적 중요도가 높아 경영에 미치는 영향 큼 → 시급하게 추진 필요 - 난이도가 높아 현재 수준에서 과제를 바로 적용 어려움	II - 현재 전략적 중요도 높지 않지만 중장기적 관점에서 반드시 추진되어야 함 - 분석과제를 바로 적용하기엔 난이도 높음
	III - 전략적 중요도가 높아 현재 시점에 전략적 가치가 있음 - 과제 추진의 난이도가 어렵지 않아 우선적 바로 적용 가능할 필요성 있음	IV - 전략적 중요도가 높지 않아 중장기전 관점에서 관제추진 바람직 - 과제를 바로 적용하는 것은 어렵지 않음
현재 ----- 시급성 ----- 미래		

- 시급성 기준 : 3 - 4 - 2 순 의사결정 / 난이도 기준 : 3 - 1 - 2 순 의사결정
- 1사분면 : 경영진 또는 실무 담당자의 의사결정에 따라 적용 우선순위 조정 가능
- 기술적 요소, 분석 범위에 따라서도 분석과제 적용 우선순위 조정 가능

3. 이행계획 수립

○ 단계적 구현 로드맵 수립

- 포트폴리오 사분면 분석으로 1차적 우선순위 결정

- 분석 과제별 적용 범위·방식 고려해 최종 우선순위 결정, 단계적 구현 로드맵 수립
- 단계별 추진 목표 정의 → 추진 과제별 선·후행 관계 고려하여 단계별 추진내용 정렬
= 추진과제

- 단계적 구현 로드맵 수립된 것 예시

- setp 1 : 데이터 분석체계 도입 → 추진단계
↓ (분석기획 발굴 및 분석과제 정의, 마스터 플랜 수립) → 추진내용
- setp 2 : 데이터 분석 유효성 검증
↓ (분석알고리즘 및 아키텍처 설계, 분석과제 pilot 수행)
- setp 3 : 데이터 분석 확산 및 고도화
(업무 프로세스 내재화율을 위한 process Innovation, 변화관리
빅데이터 분석 활용 시스템 구축, 유관 시스템 고도화)

○ 세부 이행계획 수립

→ 반복적인 정련(refine)으로 완성도 ↑

- 혼합형 분석 모델 (순차적 + 반복형 / Refine Analytics Model)

데이터 수집/확보 > 데이터 준비 > (분석모델 설계 > 적용 > 평가 > 설계) 반복 > 개발

↳ 순차적

↳ 모델링단계 반복적

- 반복 정련과정을 고려해 수립하고 최종적으로 세부 일정계획 수립

제2절 분석 거버넌스 체계

* 거버넌스 체계 개요

○ 데이터 분석과 활용에 대한 체계적인 관리의 중요성

○ 분석의 지속적 개선/개발, 확산 및 서비스관리를 위한 거버넌스 체계 구성

- 1) 조직 : 분석 기획 및 관리 수행 조직
- 2) 프로세스 : 과제 기획 및 운영
- 3) 시스템 : 분석 관련 tool
- 4) 데이터
- 5) Human resource(인력자원) : 분석 관련 교육 및 마인드 육성 체계

1. 데이터 분석 수준 진단

○ 데이터 분석 수준 진단 궁극적 목표 2가지

- 현재 분석 수준을 이해하고 수준진단 결과를 토대로 미래 목표수준 정의
- 경쟁력 확보를 위해 선택과 집중, 보완 등 개선상안 도출 가능

○ 진단방법 2가지 : 분석 준비도, 분석 성숙도

1) 분석준비도(readiness) : 데이터 분석 도입의 수준을 파악을 위한 진단방법

- * 6가지 종류를 묻거나 6가지 중 1가지에 내용에 대해 물어보는 문제 6가지
- 영역 별 진단결과 일정수준 이상 충족 시 분석업무 도입 / 미달 시 분석환경 조성

분석 업무 파악	인력 및 조직	분석기법
<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"> · 분석전문가 직무 존재 · 전문가 교육훈련 프로그램 · 관리자 기본분석능력 · 전사총괄조직 · 경영진 분석 업무 이해 	<ul style="list-style-type: none"> · 업무별 적합한 분석기법 · 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 <u>충분성/신뢰성/적시성</u> · 비구조적 데이터 관리 · 외부데이터 활용 체계 · 기준데이터 관리(MDM) 	<ul style="list-style-type: none"> · 사실에 근거한 의사결정 · 관리자의 데이터 중심 · 회의 등에서 데이터 활용 · 직관보다 데이터활용 · 데이터 공유 및 협업 문화 	<ul style="list-style-type: none"> · 운영시스템 데이터 통합 · EAI,ETL등 데이터 유통체계 · 분석 전용 서버 및 스토리지 · 분석 환경 (빅데이터/통계/비주얼)

→ 비정형데이터

⊗ CMMI

역량숙성도를 측정하기 위한 모델로 소프트웨어 품질보증과 시스템엔지니어링 분야의 품질보증 기술을 통합하여 개발된 평가모델로 1~5단계로 구성된 성숙도모델

2) 분석성숙도 (maturity) : 조직의 성숙도 수준 평가 (평가도구 : CMMI 모델)

- 분석수준은 성숙 단계에 따라 점차 진화

주관식!!!

도입, 활용, 확산, 최적화

내용이해~!

단계	도입단계	활용단계	확산단계	최적화단계
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 데이터 기반 	<ul style="list-style-type: none"> 전사 성과 실시간 분석 프로세스 혁신 3.0 분석규칙 관리 이벤트 관리 	<ul style="list-style-type: none"> 외부 환경분석 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 분석기법 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 통계 분석 환경 	<ul style="list-style-type: none"> 빅데이터 관리 환경 시뮬레이션·최적화 비주얼 분석 분석 전용 서버 	<ul style="list-style-type: none"> 분석 협업환경 분석 Sandbox 프로세스 내재화 빅데이터 분석

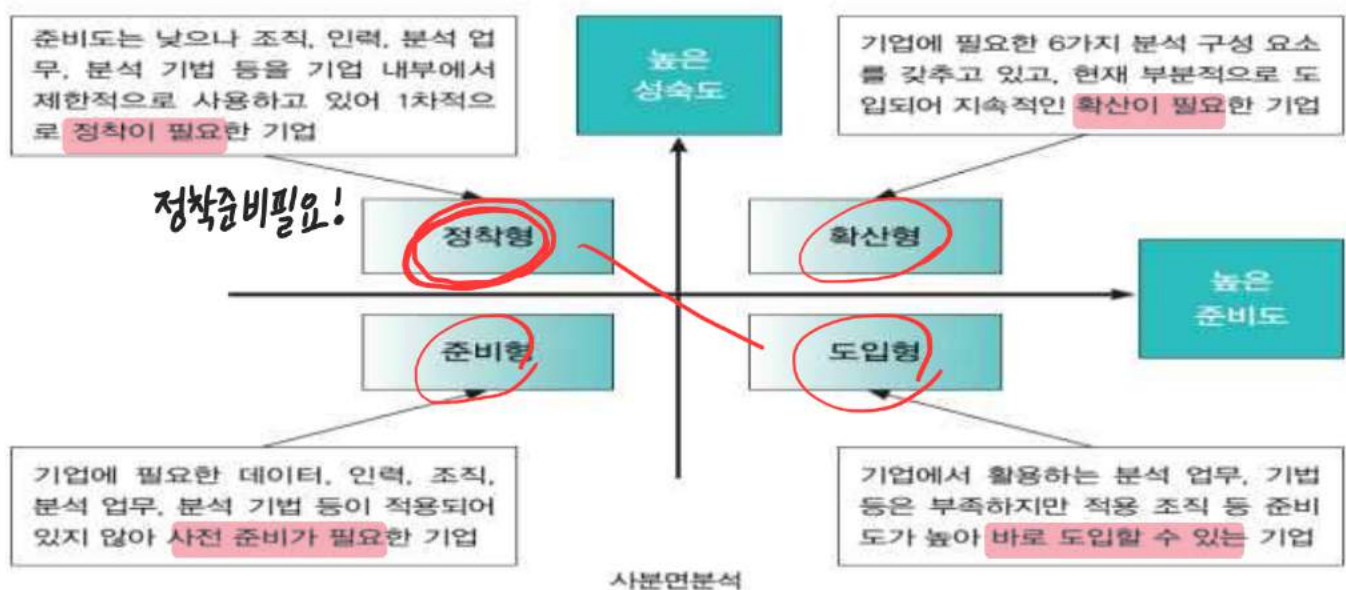
3) 분석 수준 진단 결과

○ 기업의 현재 분석 수준 객관적 파악

○ 경쟁사와 비교하여 분석 경쟁력 확보·강화를 위한 목표수준 설정

네가지 꼭 확인

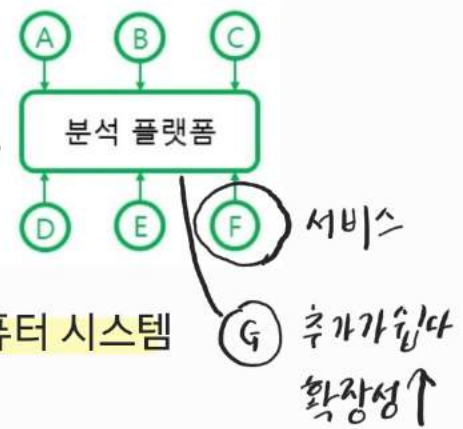
○ 분석관점에서의 사분면 분석 : 준비도, 성숙도 기준 목표방향, 개선방안 수립



2. 분석지원 인프라 방안 수립

○ 개별로 분석시스템 구축시 관리의 복잡도·비용 증가 발생

○ 장기·안정적으로 활용할 수 있는 **플랫폼 구조 도입**



○ **플랫폼** = 인프라 비슷한 것

- 분석 서비스를 위한 응용프로그램이 실행될 수 있는 기초 컴퓨터 시스템

- 분석에 필요한 프로그래밍 환경과 실행·서비스 환경 제공

3. 데이터 거버넌스 체계 수립

○ 데이터 거버넌스

- 데이터에 대한 표준화된 관리 체계를 수립하고 운영을 위한 프레임워크 및 저장소를 구축하는 것

- **중요 관리대상** : 마스터 데이터, 메타 데이터, 데이터 사전

- 데이터 **가용성·유용성·통합성·보안성·안전성** 확보

- 구성요소 : 원칙, 조직, 프로세스

{ 원칙: 데이터 유지 관리 지침, 가이드
조직: 데이터 관리 조직
프로세스: 데이터 관리 활동 및 체계

○ 데이터 거버넌스 체계

- **데이터 표준화**

표준 용어 설정, 명명 규칙 수립, 메타데이터 구축, 데이터 사전 구축 업무

- **데이터 관리체계**

데이터 정합성 및 활용의 효율성을 위해 메타데이터와 데이터사전 관리 원칙 수립
빅데이터의 경우, 데이터의 생명 주기 관리방안 수립

- **데이터 저장소 관리**

전사차원의 저장소 구성

워크플로우 관리용 응용 소프트웨어를 지원하고 관리 대상 시스템과의 인터페이스를 통한 통제 필요, 데이터 구조 변경에 따른 **사전 영향 평가**도 수행

- **표준화 활동**

표준 준수 여부를 주기적으로 점검, 모니터링 실시

계속적인 변화관리 및 주기적인 교육, 지속적인 데이터 표준화 개선활동

4. 데이터 분석 조직 및 인력방안 수립

○ 전문 분석조직 필요(기업의 차별화된 경쟁력 확보 수단)

○ 분석 조직 개요

= DSCoE = 분석조직

- (목표) 기업 경쟁력 확보를 위해 비즈니스 질문과 이에 부합하는 가치를 찾고 비즈니스 최적화하는 것
- (역할) 전사/부서의 분석업무 발굴 & 전문적 기법과 분석도구 활용하여 기업 내 빅데이터 속 insight 발견, 전파, action화
- (구성) 지식(기초통계학, 분석방법)과 분석경험을 가진 인력 또는 부서 내 조직으로 구성

○ 조직 인력 구성 시 고려사항

- 조직 구조 : 분석 업무 및 지원에 효율적인 구조인가
- 인력 구성 : 어떤 역량의 인력을 어떤 규모로 구성할 것인가

- 분석 조직 구조 3가지 유형 (집중, 기능, 분산)

무조건출제

* 3가지 유형 구별 및 내용 이해(출제 빈도 굉장히 높음)



※ DSCoE : Data Science Center of Excellence

집중형 조직구조	기능 중심의 조직구조	분산된 조직구조 Best
<ul style="list-style-type: none"> · 분석업무를 별도 분석 조직이 담당 · 전략적 중요도에 따라 우선순위 정해 진행 · 현업 부서와 <u>이원화/이중화</u>가 <u>가능</u> 높음 	<ul style="list-style-type: none"> · 일반적 분석수행구조 · 별도 조직없고 해당 부서에서 분석 수행 · 전사적 핵심 분석 어려움 · 과거실적에 국한된 분석 수행 	<ul style="list-style-type: none"> · 분석조직인력을 현업부서로 배치 (<u>분산</u>) · 전사차원의 <u>우선순위</u> 수행 · 분석결과에 따른 신속한 Action · 베스트프랙티스 공유 가능 · 분석업무와 역할분담 명확화 필요 → <u>업무과다 이원화</u> 가능성

○ 분석 조직(DSCoE) 인력 구성

비즈니스, IT기술, 분석전문, 변화관리(검직 가능), 교육담당 인력(검직 가능)

5. 분석 과제 관리 프로세스 수립

주관식



1) 과제 **발굴**: 분석 아이디어 발굴, 아이디어 과제화, **분석과제Pool로 관리**

2) 과제 **수행**: 팀 구성, 과제 실행, 지속적인 모니터링, **결과 공유 및 개선**

결과물 Pool로 관리하여

시행착오 최소로, 효율적 프로젝트 진행

6. 분석교육 및 변화관리

○ **분석교육의 목표**: 분석역량을 확보하고 강화하는 것

- 분석기획자: 분석 큐레이션 교육
- 분석실무자: 데이터 분석기법 및 툴
- 업무수행자: 분석기회 발굴, 구체화, 시나리오 작성법

○ 분석 내재화 단계

많은 조직이 분석과제 성공
못하고 포기하는 단계

