꿈꾸는 라이언

데이터 분석 준전문가

핵심 요약 노트

꿈꾸는 라이언

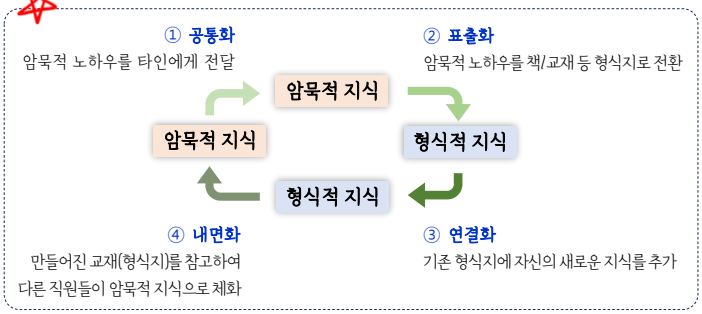
목 차

■ **데이터:** 추론과 추정의 근거를 이루는 사실 (사전적 정의)
상호관계 속에서 가치를 가질 수 있으며 추정, 예측을 위한 근거 자료

정성적 데이터 (Qualitative Data)	. 비정형 데이터, 저장/분석 등 자료 처리에 많은 비용/시간 소요 Ex. 주관식 응답 / 기상특보 / SNS 텍스트
정량적 데이터 (Quantitative Data)	. 수치 기반의 정형 데이터, 많은 자료 양에도 저장/분석 용이 Ex. 온도, 습도, 강우량, 풍속

■ **지식 경영:** 개인의 암묵지와 집단의 형식지 간에 상호 작용(생성/발전/전환)을 통해 지식의 발전을 기반으로 한 기업의 경영

암묵지	경험과 학습을 통해 개인에게 습득된 무형의 지식으로 공유 어려움 ex. 자동차 운전, 강연, 운동
형식지	구체적이거나 문서화된 지식으로 전달 및 공유 용이 ex. 작업 매뉴얼, 설계도, 데이터베이스, 프로그램



■ **DIKW 피라미드 :** 데이터, 정보, 지식을 통해 최종적으로 지혜를 얻어내는 일련의 과정을 나타내는 계층구조

. **Data (데이터) :** 가공 전의 객관적인 사실 및 순수한 수치/기호
→ *A마트는 100원에, B마트는 300원에 연필을 판매한다.*

. **Information (정보) :** 데이터 가공 및 상관관계 속에서 패턴을 인식하고 의미를 도출
→ *A마트의 연필이 B마트보다 싸다.*

. **Knowledge (지식) :** 상호 연결된 정보 패턴을 이해하여 이를 토대로 예측한 결과물
→ *상대적으로 저렴한 A마트에서 연필을 사야겠다.*

. **Wisdom (지혜) :** 근본 원리에 대한 깊은 이해를 바탕으로 도출되는 창의적 아이디어
→ *다른 상품들도 A마트가 B마트보다 저렴할 것이라 판단한다.*

■ **데이터베이스 4가지 특징**

. **통합** : 데이터베이스 내 중복된 데이터 없음 (중복 최소화)
. **저장** : 컴퓨터가 접근 가능한 저장 매체에 데이터 저장 가능
. **공용** : 여러 사용자가 동일 데이터를 공용으로 사용 (공동 접근)
. **변화** : 추가/삭제/갱신 등의 기존 데이터 변경 가능 (정확한 데이터 유지)

■ **데이터 종류**

정형 데이터 (Structured)	. 행/열로 구성된 스프레드시트 형태 (관계형 데이터베이스)
반정형 데이터 (Semi-structured)	. 데이터 구조에 대한 메타 정보 / Parsing 활용 구조 파악 (Parsing : 반정형 데이터 구조 해석하여 정보 추출) . HTML, XML, JSON 등 웹 기반 데이터
비정형 데이터 (Unstructured)	. 형태 및 구조가 정형화되지 않은 데이터 (사진, 영상, 소리, 텍스트 등)

■ **데이터베이스 구성 요소**

. **메타 데이터 :** 데이터에 대한 데이터 (데이터 특성, 구조, 의미, 관리 정보 설명)
(MetaData) 데이터베이스 내 데이터에 대한 정보를 제공하거나 설명

. **인덱스 :** 데이터베이스에서 데이터 검색 및 정렬을 빠르게 수행하기 위한 자료 구조
(Index)

■ **DBMS** (Data-Base Management System)

- 사용자의 요구에 따라 정보를 처리하고 데이터베이스를 관리하는 소프트웨어
- 장점 : 데이터의 중복 최소화, 일관성 및 무결성 유지, 모든 사용자 동시 접근 가능
데이터 액세스 권한 제어를 통해 민감 정보 보호, SQL 기반 언어로 제어 용이
타 사용자가 트랜잭션 시에도 결과 즉시 확인 가능
(데이터 상태를 변환시키는 일련의 연산 작업)
- 단점 : 모든 데이터 문제 해결 불가, 유지/보수 비용 발생, 전문 지식/관리 필요
시스템 문제 발생 시 모든 데이터 영향, 새로운 버전 적용 시 호환성 문제

RDBMS (Relational DBMS)	. 관계형 데이터베이스 관리 시스템 . 정형화된 테이블로 구성된 데이터 항목들의 집합체 ex. MySQL (오픈소스 DBMS) / Oracle Database (상용 RDBMS) Microsoft SQL Server, IBM DB2, PostgreSQL, Sybase
ODBMS (Object-oriented DBMS)	. 객체 지향 데이터베이스 관리 시스템 . 복잡한 데이터 구조를 표현/관리 . 객체를 생성하여 계층에서 체계적으로 정리하고, 상위 계층으로부터 방법과 속성을 다시 물려받음(상속)

★ **NoSQL** (Non-SQL) : 관계형 데이터베이스보다 덜 제한적인 일관성 모델 이용
디자인 단순화, 수평적 확장성, 세세한 통제
ex. MongoDB, Apache Hbase, Redis, Apche Cassandra

■ **시대별 데이터베이스 솔루션**

1980년대	**OLTP**	. On-Line Transacton Processing / 온라인 거래 처리 . 주 컴퓨터와 연결된 사용자들의 실시간 트랜잭션 처리
	OLAP	. On-Line Analytical Processing / 온라인 분석 처리 . 다차원의 데이터를 대화식으로 분석 및 통계 요약 정보 제공
2000년대	**CRM**	. Customer Relationship Management . 고객 이해를 바탕으로 한 마케팅 전략을 통해 높은 이익 창출
	SCM	. Supply Chain Management (유통망 관리) . 정보기술을 활용하여 유통 재고 및 시간/비용 최적화

■ **분야별 데이터베이스 솔루션**

ERP	. Enterprise Resource Planning . 프로세스 관리를 돕는 여러 모듈로 구성된 통합 애플리케이션
RTE	. 비즈니스 프로세스를 투명하고 민첩하게 유지, 지연시간 제거, 대기업-중소기업 간 협업적 IT화
BI (Business Interlligence)	. 데이터 웨어하우스 내 데이터에 접근하여 경영 의사 결정에 필요한 정보 획득 및 활용 (데이터 통합 및 분석) . 하나의 특정 비즈니스 질문에 답변하도록 설계 ★ Ad Hoc Report : 특정 요구에 의해 즉각 생성된 보고서
BA (Business Analytics)	. 통계적이고 수학적인 분석에 초점 (BI보다 진보된 형태)
KMS	. 지식관리시스템 / 조직 내의 지식을 체계적으로 관리하는 시스템
EDW	. Enterprise Data Warehouse . 여러 애플리케이션의 비즈니스 정보를 중앙 집중화 → 조직 전체에서 분석/사용가능하도록 하는 데이터베이스

■ **데이터 웨어하우스**

데이터 웨어하우스 (Data Warehouse)	. 기업 내 의사결정 지원을 위한 하나의 통합된 데이터 저장 공간 **[4대 특징]** ① 통합 : 전사적 차원에서 일관된 형식으로 정의 ② 시계열성 : 시간의 흐름에 따라 변화값 저장 ③ 주제 지향적 : 특정 주제에 따라 분류/저장/관리 ④ 비소멸성(비휘발성) : Batch 작업에 의한 갱신 이외에 변화 X
데이터 마트 (Data Mart)	. 소규모 단일 주제의 데이터 웨어하우스 (특정 조직의 특정 업무)

■ **빅데이터 :** 대용량 데이터를 활용하여 새로운 통찰 및 가치를 생산
3V : Volume (데이터 양↑), Variety (정형 + 비정형 + 반정형), Velocity (빠른 속도)
4V : 3V + Value (빅데이터 4번째 V, ROI) ← 비즈니스 효과 관점
Return On Investment, 투자 대비 수익율
- 사전 처리 → 사후 처리 / 표본조사 → 전수 조사 / 질 → 양 / 인과관계 → 상관관계

★ **기억 용량 단위**
KB (2^{10}, 10^3) → MB (2^{20}, 10^6) → GB (2^{30}, 10^9) → TB (2^{40}, 10^{12})
→ PB (2^{50}, 10^{15}) → EB (2^{60}, 10^{18}) → ZB (2^{70}, 10^{21}) → YB (2^{80}, 10^{24})

★ **자료 구성 단위**
- Bit → Nibble (4Bit) → Byte (8Bit) → Word (명령 단위) → Field → Record → File → DB

※ 1Bit = 데이터 최소 단위 (이진수 하나)
1Byte = 숫자, 영어, 공백, 특수문자 (반각문자) / 2Byte = 한글, 한자(전각문자)

■ **빅데이터 출현 배경**
- 소셜 미디어(SNS), 영상 등 비정형 데이터 확산, 저장 처리 장치 가격 하락
- 클라우드 컴퓨팅(분산 병렬처리로 처리 비용 감소), 인터넷 보급, 디지털화
- 인간 게놈 프로젝트 / 양질 전환 법칙 / IoT 확산
데이터의 양적 누적이 질적 비약으로 변환

■ **빅데이터에 거는 기대 (비유적 표현)**
- **철/석탄**(1차 → 2차 산업혁명의 핵심 재료), **원유**(주요 에너지원)
렌즈(생물학에서 현미경 렌즈의 중요성), **플랫폼**(공동 활용 목적의 유/무형 구조물)

■ **빅데이터의 어려운 가치 산정**
- 데이터의 활용 방식이 다양해지면서 특정 데이터의 활용 시점/장소/대상 불분명
- 새로운 분석 기술의 적용으로 가치가 없던 데이터가 나중에는 가치 창출될 수 있음

■ **빅데이터 활용 방법**

연관 규칙 학습	. 데이터 변수 간 의미있는 상관관계(연관 규칙) 발견 ex. 기저귀 구매와 맥주 구매 간 상관성
유형 분석	. 사용자의 특성(유형)에 맞게 그룹 및 범주를 나누어 분류
유전 알고리즘	. 최적화 필요한 문제 해결을 자연선택, 돌연변이와 같은 메커니즘을 통해 점진적으로 진화
기계 학습	. 훈련 데이터로부터 패턴을 파악해 예측 (넷플릭스 추천 영상)
회귀 분석	. 선형 함수로 나타낼 수 있는 수치데이터 분석
감정 분석	. 글쓴이의 감정 분석 / 고객이 원하는 것을 찾아낼 때 활용 고객의 후기를 받아 서비스 개선에 활용
소셜 네트워크 분석	. 영향력 있는 사람을 찾아 사람들 간의 소셜 관계를 파악

■ **빅데이터 위기 요인 :** 사생활 침해 / 책임 원칙 훼손 / 데이터 오용

사생활 침해	. 개인 정보가 n차 가공되어 공유 / 익명화 기술로는 해결 불충분 → 동의제에서 책임제로 전환 (개인 정보 사용자에게 책임 부여)
책임 원칙 훼손	. 빅데이터 예측 기반 잠재적 위험 감지 후 책임 추궁 (ex. 실제 범행 전 용의자 사전 체포 / 개인 신용도 무관 대출 거절) → 기존 책임 원칙 강화 (결과 기반 책임 원칙)
데이터 오용	. 빅데이터 기반 분석이 항상 맞는 결과를 제공하진 않음 → 데이터 알고리즘 접근권 허용 및 객관적 인증방안 도입 '알고리즈미스트(Algorithmist)' 도입

★ **개인 정보 비식별화 방법**
- 가명 처리 / 총계 처리 (개인 점수 → 그룹 평균) / 데이터 값 부분 및 전체 삭제 / 데이터 범주화 (27세 → 20대) / 데이터 마스킹 (주민번호 : 980402-2******)

■ **빅데이터 활용에 필요한 3요소**
① **데이터 :** 모든 것의 데이터화 (Datafication)
② **기술 :** 데이터 분석 및 처리기술의 진화 / 인공지능 AI 출현
③ **인력 :** 데이터 사이언티스트 및 알고리즈미스트의 중요성 증대

■ **데이터 사이언스**
- 정형/반정형/비정형 등 다양한 형태의 데이터에서 의미/가치를 추출
- 총체적(holistic) 접근법 사용 : 수학/통계학/컴퓨터 공학/데이터 공학 등
- 데이터 분석가 : 분석 보고서 및 시각화 자료 작성 → 데이터 기반 의사 결정 추구

■ **데이터 사이언티스트** : 머신 러닝 및 AI 활용을 위한 코딩 스킬, 통계적 지식 필요
+ **하드 스킬** : Machine Learning + Modeling + Data Technical Skill (이론 지식)
+ **소프트 스킬** : 창의적 사고, 스토리텔링, 시각화, 커뮤니케이션 (고객 공감 능력)
→ 통찰력 있는 분석 능력

★ **가트너(Gartner)가 주장한 데이터 사이언티스트의 역량**
- 데이터 관리, 분석 모델링, 비즈니스 분석, 소프트 스킬 (하드 스킬 제외)

★ **효과적인 분석모델 개발 위한 요구사항 (인문학적 요소)**
- 분석 모델이 예측할 수 없는 위험 요소를 지속적으로 판단
- 모델 능력에 항상 의구심을 가지며, 분석 모델의 한계에 대해 끊임없이 고찰
- 모델 범위 바깥의 외부 요인은 판단하지 않음 (가정과 현실의 불일치)

▶ 과학 기반의 정량 분석 + 인문학적 통찰 기반 합리적 추론
(과거: 모델링, 실험설계 / 현재 : 차선 행동 / 미래 : 예측, 최적화)
→ '인과관계'가 아닌 '상관관계' 분석 기반 인사이트 중요도 높아짐

▶ 사회경제적 변화 : 단순 세계 → 복잡 세계 (다양성, 연결성, 창조성)
(인문학 열풍) 비즈니스 중심의 제품 생산 → 서비스 중심
경제 및 산업 논리의 생산 → 시장 창조 (무형 자산)

■ **가치 패러다임의 변화 :** Digitalization ▶ Connection ▶ Agency
(아날로그 → 디지털) (디지털 정보 간 연결) (효율/효과적인 연결 지향)

- 일차적 분석: 큰 변화에 대응하거나 고객 환경의 변화 파악, 새로운 기회 포착 어려움
- 전략도출 가치기반 분석: 전략적 통찰력 창출을 통해 기회 발굴
주요 경영진의 지원 확보 / 넓은 활용 범위 / 전략적인 변화

■ **분석 기획 :** 분석을 수행할 과제 정의 및 관리할 수 있는 방안을 사전에 계획하는 작업
→ 균형 잡힌 시각 / 데이터 프로그래밍 기술 역량 필요

☆ 분석 주제 유형 4가지

분석 방법 \ 분석 대상	Known	Un-Known
Known	최적화 (Optimization)	통찰 (Insight)
Un-known	솔루션 (Solution)	발견 (Discovery)

☆ 목표 시점별 분석 기획 방안

구 분	과제 중심적 (단기)	마스터 플랜 (장기)
1차 목표	Speed & Test	Accuracy & Deploy
과제 유형	Quick-Win*	Long Term View
접근 방식	Problem Solving (문제 해결)	Problem Definition (문제 정의)

* Quick-Win : 단기 해결을 위한 즉각적인 실행 및 추진 방식

■ **분석 기획 시 고려사항**

가용한 데이터	. 데이터 확보 및 데이터 유형 분석 사전 시행 (정형, 반정형, 비정형)
적절한 유스케이스	. 기존 수립된 유사 분석 시나리오 및 솔루션 최대한 활용
장애 요소 확인	. 분석 방해하는 장애 요소들에 대한 사전 계획 수립 필요 → 지속적인 교육 및 활용방안을 통한 조직 역량 내재화

■ **분석 방법론**

폭포수	. 선형 순차적 개발(하향식) / 고전적, 전통적 개발 모형 . Step-by-Step 단계마다 진행 / 문제 시 전 단계로 피드백 수행
나선형 (Spiral)	. 점진적 개발 과정 반복 / 정밀하며 유지보수 과정 필요없으나 효과적인 관리 체계 부재 시 복잡도 상승으로 프로젝트 진행 불가 ※ 계획/목표 수립 → 위험 분석 → 개발 및 검증 → 고객 평가
프로토타입	. 고객 need 파악 위해 견본/시제품을 통해 최종 결과 예측 (상향식) . 불분명한 사용자 요구사항 시 일단 프로토타입 먼저 개발 후 그 결과를 토대로 반복적인 개선 작업 진행

■ **KDD 분석 방법론** (Knowledge Discovery in Database)

데이터 셋 선택	. 분석에 필요한 데이터(raw data) 선택 → 분석 대상 산업 도메인의 이해와 해당 과제 목표 설정 필수
데이터 전처리	. 데이터의 노이즈, 이상값, 결측치 확인 후 필요 시 제거(정제)
데이터 변환	. 분석 목적에 맞는 데이터 선별 및 데이터 차원 축소
데이터 마이닝	. 분석 목적에 맞는 데이터 마이닝 알고리즘 선택 . 데이터 간 패턴 및 경향/예측 등 데이터 마이닝 작업 실시
분석 결과 평가	. 데이터 마이닝 결과에 대한 해석과 평가 진행

■ **CRISP-DM 분석 방법론**

업무 이해	. 분석 과제의 목적과 요구사항 이해 → 업무 목적 파악 / 상황 파악 / 데이터 마이닝 목표 설정 프로젝트 계획 수립
데이터 이해	. 분석을 위한 데이터 수집 및 이해 → 초기 데이터 수집, 데이터 기술 분석, 데이터 품질 확인
데이터 준비	. 수집된 데이터에서 분석 용이한 데이터만 추출 및 편성 → 데이터 셋 선택, 데이터 정제, 데이터 통합, 데이터 포맷팅
모델링	. 다양한 알고리즘 적용으로 모델링 수행 및 파라미터 최적화 → 데이터 분할, 모델링 기법 선택, 모델링 작성, 모델 평가
평가	. 수행된 모델링이 프로젝트 목적에 적절한지 평가 → 분석결과 평가, 모델링 과정 평가, 모델 적용성 평가
전개	. 완성된 모델을 실제 현업에 적용하기 위한 이행 계획 수립 → 전개 계획 수립, 모니터링/유지보수 계획 수립, 프로젝트 리뷰

■ **데이터 분석 방법론**

- 단계(Phase), 태스크(Task), 스텝(Step) 3개의 층으로 구성

분석 기획	. 비즈니스 이해 및 범위 설정 / 프로젝트 정의 및 위험(Risk) 계획 수립 * SOW (Statement of Work, 작업 기술서/명세서) : 고객의 요구사항 및 프로젝트 예상 결과 등을 세부적으로 기술 * WBS (Work Breakdown Structure, 업무 분업 구조도) : 전체 업무를 분류 및 계획하여 각 업무 담당자에게 업무 분할 → 위험 대응 방안 : 회피, 전이, 완화, 수용
↓	
데이터 준비	. 필요한 데이터 정의 / 데이터 스토어(Data store) 수립 . 데이터 수집 및 정합성 점검 * 웹 크롤링 : 웹 데이터의 구조 분석 후 데이터 자동 수집 * API : 프로그램 라이브러리 접근을 위해 정의한 규칙 * ETL : 데이터 추출/변환 및 데이터베이스에 적재하는 작업
↓	
데이터 분석 ↓↑	. 분석용 데이터 준비 / 텍스트 분석 / 탐색적 분석 . 데이터 분할 - 훈련용 학습데이터와 검증용 실험데이터로 분류 . 모델링 / 모델링 평가 및 검증 → 추가적 데이터 확보 필요 시 데이터 준비 단계로 피드백 * EDA (Exploratory Data Analysis, 탐색적 데이터 분석) : 데이터셋 분석/조사하여 주요 특성을 파악하는 데에 사용
시스템 구현	설계 및 구현 / 시스템 테스트 및 운영 단순 데이터 분석으로 프로젝트 종료 시 [시스템 구현] 단계 미진행
↓	
평가 및 전개	모델 발전 계획 수립 / 프로젝트 평가 및 보고

■ **분석 과제 도출 방법**

. **하향식 접근** : 제시된 문제(분석 대상)가 확실한 경우 체계적인 단계화로 문제 해결
→ 분석 주제 유형이 '최적화' 및 '솔루션'일 때

①문제 탐색	. 비즈니스 모델 기반 문제 탐색 → 비즈니스 모델 캔버스 활용 (9가지 블록을 5가지로 단순화) 업무, 제품, 고객, 지원인프라, 규제/감사 ※ 거시적 관점 : STEEP (사회, 기술, 경제, 환경, 정치) 경쟁자 확대 관점 : 대체자, 경쟁자, 신규 진입자 시장니즈 탐색 관점 : 고객, 채널, 영향자 . 외부 참조 모델 기반 문제 탐색 (외부 사례 벤치마킹) . 분석 유즈 케이스 (데이터 분석 문제로 전환)
②문제 정의	. 비즈니스 문제를 데이터 분석 문제로 변환 → 해결하기 위한 데이터 및 방법(How)을 정의하기 위한 목적
③해결방안 탐색	. 수행 옵션 도출 분석 기법 \ 분석 역량 / 확보 / 미확보 기존 시스템 / 기존 시스템 개선 / 교육/채용 통한 역량 확보 신규 도입 / 시스템 고도화 / 전문 업체
④타당성 검토	. 타당성 평가 및 과제 선정 (비용 대비 편익 / 데이터 존재 여부 / 분석 시스템 환경)

. **상향식 접근** : 문제 정의 자체가 어려운 경우 데이터 기반으로 문제 탐색 / 지속 개선
→ 분석 주제 유형이 '통찰' 및 '발견'일 때 (What 관점)
→ 인사이트 도출 후 반복적 시행착오를 통한 문제 도출
→ 비지도 학습 방법에 의해 수행

지도 학습 (Supervised Learning)	비지도 학습 (Unsupervised Learning)
명확한 input / output	불명확한 input / output
예측 (Regression)	연관 (Association)
분류 (Classification)	군집화 (Clustering)

. **디자인 사고**(Thinking) : 상향식 접근 '발산' 단계 + 하향식 접근 '수렴' 단계 통합 적용
공감 → 문제정의 → 아이디어 도출 → 프로토타입 → 테스트

. **Self Service Analytics**
- BI도구, Ad hoc report, OLA, Visual Discovery, Machine Learning 기능 포함
- R, Python 등의 데이터 분석 언어와 많은 통계적 지식 필요 (분산처리 지원 없음)

■ **분석 프로젝트 :** 비즈니스 영역과 데이터 영역의 현황 파악 / 구성원과 협업 중요
애자일(Agile) 프로젝트 관리 방식 / 분석 과제 정의서 기반 프로젝트 수행

☆ 분석 과제 정의서
- 분석에 필요한 소스 데이터, 분석 방법, 데이터 입수 및 분석 난이도, 분석 수행주기, 분석 결과에 대한 검증 등을 정의
- 이해관계자들의 프로젝트 방향 설정 및 성공 여부 판단의 명확한 기준 작성

■ **분석 과제 주요 특징 5가지**
- 데이터 크기 / 데이터 복잡성 / 속도 / 분석 복잡성 / 정확도 및 정밀도
Data size / Data Complexity / Speed / Analytic Complexity / Accuracy & Precision

* 정확도(Accuracy): 높을수록 모델과 실제 값 사이의 차이가 적음
* 정밀도(Precision): 높을수록 모델을 지속 반복했을 때의 결과 편차가 적음
→ '정확도'와 '정밀도' 간에 Trade-off 관계 존재 (정확도 ↔ 복잡도 마찬가지)
→ 분석의 활용 측면에서는 Accuracy 중요, 안정성 측면에선 Precision 중요

■ **분석 프로젝트 관리 방안 10가지**
- 시간 / 범위 / 품질 / 통합 / 이해관계자 / 자원 / 원가 / 리스크 / 조달 / 의사소통

■ **분석 마스터 플랜** (전략적 중요도 / 실행용이성 / ROI ↑)
- 데이터 분석 과제 도출 → 과제 우선순위 결정 → 적용 우선순위 결정 → 분석 구현 로드맵 수립
(적용 범위/방식 고려)
업무 내재화 / 분석 데이터 / 기술 적용 수준

* 모델링 단계만 반복적으로 수행하는 혼합형 주로 적용

■ **ISP** (Information Strategy Planning / 정보 전략 계획)
- 전략적 주요 정보를 대상으로 전사적인 종합추진 계획 / 중장기 마스터 플랜 수립

■ **분석 우선순위 평가 기준**
. **시급성** : 전략적 중요도 / 목표 가치 → 비즈니스 효과 (Return) : Value
. **난이도** : 데이터 획득, 저장, 가공 비용 / 분석 적용 비용 / 분석 수준
→ 투자비용 요소 (Investment) : Volume, Variety, Velocity

. 시급성 우선 시 : 3 → 4 → 2
. 난이도 우선 시 : 3 → 1 → 2
※ 시급성 높으나 난이도가 높은 과제(1 영역)도 경영진 의사결정에 따라 우선순위 조정 가능

★ **로드맵 :** 과제 계획서로 목표 달성을 위한 방향 및 일정을 시각적으로 표현한 문서

■ **분석 거버넌스 체계 :** 의사결정을 위한 데이터 분석과 활용을 위한 관리체계
- 구성 요소 : 프로세스 (Process) / 조직 (Organization) / 시스템 (System)
인적 자원 (Human resource) / 데이터 (Data)
분석 관련 교육 / 마인드 육성 체계 데이터 거버넌스

■ **데이터 분석 조직 구조**

집중형 조직	. 별도 독립적 분석 조직 구성 (조직 내 분석 업무 일괄 담당) → 전사 차원에서 전략적 중요도에 따라 우선순위 결정 → 협업 부서와 중복 업무 가능성
기능 중심 조직	. 별도 독립 조직 없이 각 해당 부서에서 직접 분석 → 전사적 관점 분석 관리 어려움
분산 조직	. 독립적 분석 조직의 인력을 협업부서에 배치시켜 업무 수행 → 분석 결과의 신속한 실무적용 가능

■ **분석 과제 관리 프로세스**
. 과제 발굴 : 분석 아이디어 발굴 / 과제 후보 제안 / 분석 과제 확장
. 과제 수행 : 팀 구성 / 분석 과제 실행 / 분석 과제 진행 관리 / 결과 공유 및 개선
→ 분석 과제 진행 간 시사점과 분석 결과물들은 풀(Pool)에 축적하여 관리

■ **데이터 분석 수준 진단**

1. 데이터 분석 준비도 (Readiness)

분석 업무 파악	사실 분석 / 예측 / 시뮬레이션 / 최적화 / 분석 업무 정기적 개선
인력 및 조직	분석 전문가 / 관리자 분석 능력 / 분석 업무 조직 / 경영진 이해
분석 기법	적절한 분석 기법 / 분석 기법 라이브러리, 평가, 정기적 개선
분석 데이터	데이터 관리 / 외부 데이터 활용 / 기준데이터 관리 (MDM)
분석 문화	사실 기반 의사결정 / 회의에서 데이터 활용 / 데이터 공유 및 협업 / 관리자의 데이터 중요성 인지
IT 인프라	운영 시스템 데이터 통합 / 분석 환경

2. 데이터 분석 성숙도 (Maturity)
- CMMI 모델 기반 평가 / 비즈니스, 조직 및 역량, IT 부문 관점으로 구분

도입	. 환경, 시스템 구축 / 데이터 웨어하우스, 데이터 마트, ETL, OLAP
활용	. 분석 결과 업무에 적용 / 실시간 대시보드, 통계분석 환경
확산	. 전사 차원 분석 관리 및 공유 / 비주얼 분석, 분석 전용 서버
최적화	. 혁신 및 성과 향상에 기여 / 분석 Sandbox, 프로세스 내재화, 빅데이터

■ **데이터 분석 성숙도 모델**

준비형	데이터, 조직, 분석업무, 분석기법 등 미적용되어 사전 준비 필요
정착형	데이터, 조직, 분석업무, 분석기법 등을 내부에서 제한적으로 사용
도입형	분석업무 및 기법이 부족하나 조직 및 인력 등 준비도는 높은 상황
확산형	6가지 분석 구성요소 갖춘 상태 / 지속적 확산 가능

▶ 분석 지원 인프라 방안 수립 : 확장성을 고려한 플랫폼 시스템 적용 (중앙집중적)

■ **데이터 거버넌스 :** 전사 차원의 데이터 관리 체계 구축
- 구성 요소 : 원칙 / 조직 / 프로세스
- 관리 대상 : 마스터 데이터, 메타데이터, 데이터 사전

. 데이터 거버넌스 체계 요소

데이터 표준화	데이터 표준용어 설정, 명명 규칙 수립, 메타데이터 및 데이터 사전 구축
데이터 관리체계	효율적 데이터 관리를 위한 관리 원칙 수립
데이터 저장소관리	전사 차원의 저장소 구성
표준화 활동	거버넌스 체계 구축 후 표준 준수 여부를 주기적 점검

■ **빅데이터 거버넌스**
- 기존 데이터 거버넌스 체계에 빅데이터의 효율적 관리, 다양한 데이터 관리 체계, 데이터 최적화, 정보 보호, 데이터 카테고리 별 관리자 책임자 지정 등을 포함