

ADsP 2과목 : 데이터의 분석 기획

▼ 01 분석의 4가지 유형

분석 대상을 알면 OS 모르면 ID 로 암기하자

- **최적화(Optimization)**: 분석 대상 및 분석 방법을 이해하고 현 문제를 최적화의 형태로 수행함
- **솔루션(Solution)**: 분석 대상 알고, 분석의 방법을 모르는 경우 솔루션을 찾는 방식으로 과제 수행
- **통찰(Insight)**
 - 분석의 대상(WHAT)을 모르고 하는 분석 방법(HOW)의 경우 분석 방향 (16회)
 - 분석 대상을 모르나 기존 분석 방식을 활용해 분석 (18회)
- **발견(Discovery)**: 분석 대상을 모르고, 분석 방법도 모르는 경우 발견을 통해 대상을 새로 선정하는 것

분석 대상 (what)

		Known	Un-Known
분석 방법 (how)	Known	최적화(Optimization)	통찰(Insight)
	Un-Known	솔루션(Solution)	발견(Discovery)

Insight를 찾는 객관식 (17회)
 통찰, 발견을 찾는 객관식 (18회)
 Solution찾는 객관식 (21회)

▼ 02 목표 시점별 분석기획 방안

- **과제 중심적인 접근 방식**의 특징
 - 목표 시점별로 당면한 과제를 빠르게 해결하는 방식
 - Speed & Test
 - Quick - Win
 - Problem Solving
- **마스터플랜 단위**
 - 지속적 분석 문화 내재화
 - Accuracy & Deploy

- Long Term View
- Problem Definition

과제 중심적 접근 방식의 특징 아닌 것 (19회)

- Accuracy * Deploy는 분석 마스터플랜

▼ 03 분석 기획 시 고려사항

- **가용한 데이터** (Available Data)
 - 데이터 유형에 따라 적용 가능한 솔루션 및 분석 방법이 다르기 때문에 유형에 대한 분석이 선행적으로 이루어져야 함
 - 예) 정형, 비정형, 반정형 데이터
- **적절한 유스케이스** (Proper Use-Case) 탐색
 - 유사 분석 시나리오 및 솔루션이 있다면 이를 최대한 활용해야 함
- **장애요소들에 대한 사전 계획 수립** (Low Barrier of Execution)
 - 장애요소들에 대한 사전 계획 수립 필요
 - 일회성 분석으로 그치지 않고 조직의 역량을 내재화하기 위해서는 **충분하고 지속적인 교육 및 활용방안 등의 변화관리** (Change Management)가 고려되어야 함

장애 요소에 대한 부적절 설명 찾기 (18회)

- 복잡하고 정교한 모형이 효과적이라해서 틀림

분석 과제 기획 시 고려요소 아닌 것 고르기 (19회)

- 데이터 분석을 위해 데이터 정형화 필수라고 해서 틀림

▼ 04 데이터의 종류

- **정형** : CRM Transaction data, ERP, Demand Forecast
- **반정형** : sensor, machine data, Competitor pricing
- **비정형** : Email, SNS, voice, IoT, 보고서, news

정형, 반정형, 비정형의 예 고르기 (18회)

다른 종류 고르기 (21회)

- ERP, CRM, Demand Forecast는 정형

▼ 05 분석방법론의 구성요소 (절방도 템)

- **절차**
- **방법**
- **도구와 기법**
- **템플릿**

구성요소가 아닌 것 고르기 (20회, 22회)

▼ 06 분석 방법론의 모델 3가지

- **폭포수모델**: 단계를 순차적으로 진행, 이전단계가 완료되어야 다음 단계로 순차 진행하는 **하향식 진행**, 문제점이 발견되면 전단계로 돌아가는 피드백 수행
- **나선형모델**: 반복을 통하여 점증적으로 개발, 처음 시도하는 프로젝트에 적용이 용이하지만, 반복에 대한 관리체계를 효과적으로 갖추지 못한 경우 복잡도가 상승하여 프로젝트 진행이 어려울 수 있음
- **프로토타입모델**
 - 사용자가 **요구사항이나 데이터를 정확히 규정하기 어렵고 데이터 소스도 명확히 파악하기 어려운 상황에서 일단 분석을 시도**해보고 그 결과를 확인해가면서 반복적으로 개선해 나가는 방법
 - 사용자들이 이렇게 만들면 편하게 사용할거야 라는 가설을 생성하게 된다
 - 특정 가설을 갖고 서비스를 설계하고 디자인에 대한 실험도 실행한다
 - 프로토타입을 보고 완성시킨 결과물을 통해 가설을 확인할 수 있다
 - 신속하게 해결책 모형제시, **상향식 접근방법**이다
 - 시제품이 나오기 전의 제품의 원형으로 개발 검증과 양산 검증을 거쳐야 시제품이 될 수 있다
 - '정보시스템의 미완성 버전 또는 중요한 기능들이 포함되어 있는 시스템의 초기모델'이다

나선형 모델링 - 주관식 문항으로 출제 (19회)

프로토타이핑 - 틀린 것 찾는 객관식 (17회)

프로토타이핑 - 적절한 설명 찾기, 적절하지 않은 설명 찾기 (18회, 23회)

▼ 07 KDD 분석

- 분석 대상의 비즈니스 도메인에 대한 이해와 프로젝트 목표를 정확하게 설정
- 데이터셋 선택
- **데이터 전처리**: 데이터 세트에 포함되어있는 **잡음(Noise)**과 **이상값(Outlier)**, **결측치(Missing Value)**를 식별하고 필요시 제거하거나 의미 있는 데이터로 처리하는 데이터 세트 정제작업 단계
- **데이터 변환**: 분석 목적에 맞는 변수 선택, 데이터의 차원 축소
- **데이터 마이닝**: 데이터의 패턴을 찾거나 데이터를 분류 또는 예측 등의 마이닝 작업 시행
- 데이터 마이닝 결과 평가

설명을 써주고 '데이터 전처리' 찾기 (23회)

▼ 08 CRISP-DM 분석 (22년 2월 18일 수정)

- **업무 이해**(Business Understanding)

- 주요구성 : 업무 목적 파악 - 상황 파악 - 데이터 마이닝 목표 설정 - 프로젝트 계획 수립
- 데이터 이해
 - KDD의 데이터 셋 준비, 데이터 전처리 == CRISP-DM의 데이터 이해
- 데이터 준비
 - KDD의 데이터 변환 == CRISP-DM의 데이터 준비
- 모델링(Modeling)
 - 데이터 분석 방법론, 머신러닝을 이용한 수행 모델을 만들거나 데이터를 분할하는 부분 (단답형 지문)
 - 모델링 기법 선택, 모델링 작성, 모델 평가
- 평가(Evaluation)
 - 분석 결과 평가, 모델링 과정 평가, 모델 적용성 평가
- 전개

KDD의 데이터 전처리와 유사한 CRISP-DM 분석을 주관식으로 (16회)

업무 이해에 대한 - 주요 구성 찾기 (21회)

구성 내용 아닌 것 찾기 (21회)

- 모델 적용성 평가가 ... 아닌 것이었음 (-.-)

모델링 주관식 문제로 출제 (18회)

▼ 09 분석기획(Planning) 단계의 Task

- 비즈니스 이해 및 범위 설정 -> 프로젝트 정의 및 계획 수립 -> 프로젝트 위험계획 수립
- 비즈니스 이해 및 범위 설정 : SOW(statement of work) 작성
- 프로젝트 정의 및 계획 수립 : WBS(work breakdown structure) 작성
- 프로젝트 위험 계획 수립 :
 - 위험에 대한 대응 방법: 회피(avoid), 전이(Transfer), 완화(Mitigate), 수용(Accept)

분석기획 단계의 순서 찾기 객관식 (17회)

빅데이터 분석 기획 단계에서의 Task 아닌 것 (20회, 22회)

프로젝트 위험 계획 수립시 위험에 대한 대응 방법 아닌 것 고르기 (20회, 22회)

- 관리(Management)라는 것 틀림

10 SOW : 작업기술서 또는 업무기술서

- SOW를 통해 프로젝트의 목적달성을 위해서 해야 할 일에 대해 이해할 수 있음
- 일정, 원가, 품질에 대한 요구 사항 및 고객, 사용자들을 명시한 포괄적인 문서
- WBS(Work Breakdown Structure) 범위 기술서에 정의된 high level의 정보를 바탕으로 구체적인 업무 범위를 설정하는 것

11 데이터 준비 단계

- 필요 데이터 정의
- 데이터 스토어 설계
- 데이터 수집 및 정합성 점검

▼ 12 데이터 분석 단계

- 분석용 데이터 준비
- 텍스트 분석
- 탐색적 분석(EDA)
- 모델링
- 모델 평가 및 검증

추가적인 데이터 확보가 필요한 경우 반복적인 피드백을 수행하는 구간

데이터 준비 ~ 데이터 분석

데이터 분석 주관식 문제로 출제 (18회) => ??

▼ 13 하향식 접근 모델

- 하향식 접근법의 데이터 분석 기획 단계
 - Problem Discovery -> Problem Definition -> Solution Search -> Feasibility Study
 - 문제탐색 -> 문제정의 -> 해결방안탐색 -> 타당도검토
- 하향식 접근방식의 문제 탐색 단계(Problem Discovery)
 - **비즈니스 모델 기반 문제 탐색**: 비즈니스 모델 틀을 활용하여 가치가 창출될 문제를 누락없이 도출할 수 있다

비즈니스 모델 관점에서는 해당 기업의 사업 모델을 도식화한 비즈니스 모델 캔버스 블록을 단순화하여 **업무, 제품, 고객** 단위로 문제를 발굴하고 이를 관리하는 **지원 인프라, 규제와 감사** 영역에 대한 기회를 추가로 도출하는 작업 수행

- **분석 기회 발굴의 범위 확장**: 환경과 경쟁구도의 변화, 역량의 재해석을 통해 분석 기회를 추가 도출한다

- **거시적 관심의 요인**: 사회, 기술, 경제, 환경, 정치 영역
- **경쟁자 확대 관점**: 대체재 영역, 경쟁자 영역, 신규진입자 영역
- **시장의 니즈 탐색**: 고객(소비자)영역, 채널 영역, 영향자들 영역
- **역량의 재해석 관점**: 내부역량 영역, 파트너 네트워크 영역

- 외부 참조 모델 기반 문제 탐색
- 분석 유즈 케이스(Analytics Use Case) 기반 문제 탐색

현재의 비즈니스 모델 및 유사, 동종사례 탐색을 통해서 빠짐없이 도출한 분석기회들을 구체적인 과제로 만들기 전에 **분석 유즈 케이스**로 표기하는 것이 필요하다. 풀어야 할 문제에 대한 상세서명 및 해당 문제를 해결했을 때 발생하는 효과를 명시함으로써 향후 데이터 분석 문제로의 전환 및 적합성 평가에 분석 유즈 케이스를 활용하도록 한다

- 문제 탐색의 도구
 - 비즈니스 모델 기반 문제 탐색, 외부 참조 모델 기반 문제 탐색, 분석 유즈 케이스

하향식 분석 기획 단계 올바른 것 찾기 (20회, 22회)
 하향식 접근 방식의 문제 탐색 단계 설명 아닌 것 고르기 (18회)
 분석 유즈 케이스 단답형 (23회)
 비즈니스 모델 관점 구성단위 찾는 객관식 문항 (17회, 22회)
 문제 탐색의 도구가 아닌 것 찾기 (19회)
 - 탐색적 문제 발견이라해서 틀림
 틀린 것 고르기 (16회, 20회)
 - 소비자(Consumer)영역은 '시장 니즈 탐색 관점의 영역'이다
 - 경쟁자 관점은 '대체제', '신규진입자', '경쟁자' 영역등이 있다

▼ 14 상향식 접근 모델

- 디자인 싱킹의 발산 단계에 해당함
- 인사이트 도출 후 반복적인 시행착오를 통해 수정하며 문제를 도출하는 일련의 과정
- 문제의 정의 자체가 어려운 경우 상향식 접근 방식 사용
- 데이터를 기반으로 문제의 재정의 및 해결방안을 탐색하고 이를 지속적으로 개선하는 방식
- 일반적으로 상향식 접근 방식의 데이터 분석은 비지도학습(Unsupervised Learning) 방법에 의해 수행됨

상향식 접근방식의 특징으로 올바른 것 찾기 (23회)

▼ 15 분석 과제 접근 방법

- 하향식 접근 방식
 - 문제가 확실할 때 사용함
 - 문제가 주어지고 해법을 찾기 위해 사용함
- 상향식 접근 방식
 - 문제의 정의 자체가 어려운 경우 사용함
- 디자인 사고(Design Thinking)

- 상향식과 하향식을 **반복적**으로 사용하기 쉬움
- 기존의 논리적인 단계별 접근법에 기반한 문제해결 방식은 최근 복잡하고 다양한 환경에서 발생하는 문제에 적합하지 않을 수 있다, 이를 해결하기 위해 "디자인 사고" 접근법을 통해 전통적인 분석적 사고를 극복하려고 한다
- 상향식 방식의 발산(Diverge)단계와 도출된 옵션을 분석하고 검증하는 하향식 접근 방식의 수렴단계를 반복하여 과제를 발굴한다

틀린 것 고르는 객관식 (16회)

- 문제가 확실할 때 하향식인데 상향식으로 써서 틀림!

디자인 사고(Design Thinking) (21회)

- 주관식 문제

▼ 16 분석 프로젝트의 일반적인 특징

- 분석 프로젝트는 다른 프로젝트 유형처럼 범위, 일정, 품질, 리스크, 의사소통 등 영역별 관리가 수행되어야 한다
- 또한 다양한 데이터에 기반한 분석 기법을 적용하는 특성 때문에 5가지 주요 특성을 고려하여 추가적 관리가 필요하다
- **분석 과제 주요 특성**으로 **Data Size, Data Complexity, speed, Analytic Complexity, Accuracy, Precision** 등이 있다 (18회)
 - Accuracy : 분석의 활용적인 측면 (모델과 실제값의 차이)
 - Precision: 안정성 측면 (모델을 반복했을 때의 편차)
 - **Accuracy, Precision은 트레이드 오프 경우가 많다**
 - **정확도와 복잡도는 트레이드 오프관계가 존재한다**
- 분석 프로젝트는 도출된 결과의 재해석을 통한 지속적인 반복 및 정규화가 수행되기도 한다
- 분석 프로젝트의 경우 관리 영역에서 일반 프로젝트와 다르게 의유해야 할 요소가 존재한다
 - **분석 프로젝트 관리 영역 10개**: **시간, 범위, 품질, 통합, 이해관계자, 자원, 원가, 리스크, 조달, 의사소통**
- 분석 결과를 도출하였을 때 이를 활용하는 시나리오 측면에서 속도를 고려해야 함

틀린 것 찾는 객관식 (17회)

- 분석 프로젝트가 다른 유형과 다르고 추가적 관리사항이 필요없다로 틀림

분석 과제 수행시 고려하는 5요소 아닌 것 찾기 (18회)

- 5요소 기억 : 데이터 크기, 데이터 복잡성, 스피드, 분석 복잡도, accuracy & precision

분석 프로젝트 관리의 영역이 아닌 것 찾기 (19회)

- 영역별 주요 관리 항목 10가지 암기 필요

적절하지 않음 설명 찾기(21회)

- 일정계획을 지속적인 철저한 통제와 관리로 일정 관리 진행이라해서 틀림

적절하지 않은 설명 찾기(21회)

- Accuracy, Precision 관계가 트레이드 오프 되지 않는다 해서 틀림

▼ 17 분석 과제 관리 프로세스

- 분석 과제 관리 프로세스는 크게 과제발굴과 (과제수행) 및 모니터링으로 나누어진다.
- 과제발굴 단계에서는 개별 조직이나 개인이 도출한 분석 아이디어를 발굴한다
- 분석 과제 확정 후 과제 제안자는 (팀구성)을 하고 (분석과제실행)시 지속적인 모니터링과 과제 결과를 공유하고 개선하는 절차를 수행한다

괄호 채우기 주관식 (21회)

▼ 18 분석 마스터 플랜

- 데이터 분석 구현을 위한 적용 우선순위 평가 고려사항
 - 전략적 중요도
 - 실행 용이성
 - 비즈니스 성과/ROI(투자자본수익률)
- 적용 범위 및 방식의 고려요소
 - 업무 내재화 적용 수준
 - 분석 데이터 적용 수준
 - 기술 적용 수준
- 중장기적 마스터 플랜 수립을 위해서는 분석 과제를 대상으로 다양한 기준을 고려해 적용할 우선순위를 설정할 필요가 있음
- 분석 과제의 적용 범위 및 방식에 대해서도 종합적으로 고려하여 결정함
- 분석 과제 수행의 **선후행 관계를 고려하여 우선순위를 조정**해 나간다
- 일반적인 IT 프로젝트의 우선순위로는 전략적 중요도와 실행 용이성이 있다

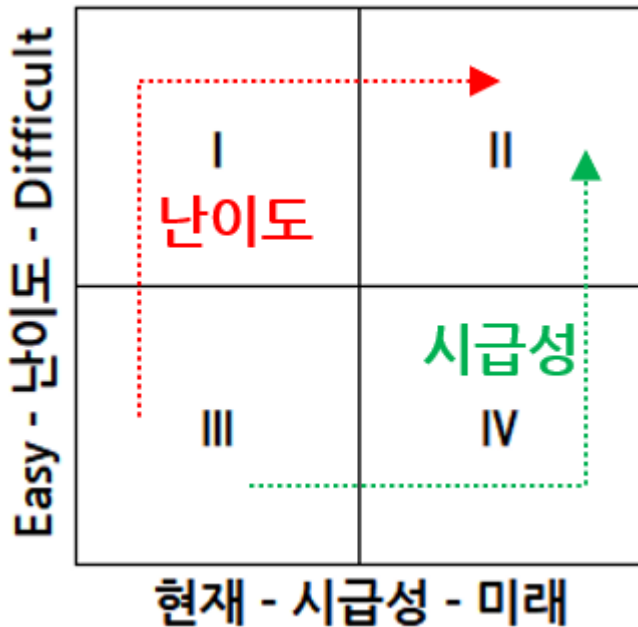
우선순위 평가 시 고려요소 아닌 것 고르기 객관식 (16회)

마스터 플랜 설명 틀린 것 찾기 (17회)

- 분석 과제 수행의 선후관계 분석을 통해 전체 과제를 반복적이고 순환적으로 작성한다 해서 틀림
- 마스터 플랜 수립시 적용 범위 및 방식의 고려요소 아닌 것 찾기 (20회, 22회)
- 투자 비용 수준이 있어서 틀림

▼ 19 분석 과제의 우선순위 선정 기법

- 포트폴리오 사분면을 통한 과제 우선순위 선정



- 3사분면 : 일반적으로 가장 먼저 하는 것
- 우선순위를 '시급성'에 둔다면 Ⅲ - IV - Ⅱ 순서 진행
- 우선순위를 '난이도'에 둔다면 Ⅲ - I - Ⅱ 순서 진행
- **시급성의 판단 기준은 전략적 중요도가 핵심**
- **난이도**는 현시점에서 과제를 추진하는 것이 **비용과 범위 측면**에서 쉬운(Easy) 것인지 어려운(Difficult) 것인지에 대한 판단 기준
- **시급성이 높고 난이도가 높은 영역은 경영진 또는 실무 담당자의 의사결정에 따라 적용 우선순위를 조정할 수 있음**

일반적으로 가장 먼저 하는 것 : 3사분면 (19회)

우선 순위를 '시급성'에 둔다면 결정해야 할 우선 순위는? (20회, 22회)

- 3 -> 4 -> 2 순서

우선 순위를 '난이도'에 둔다면 결정해야 할 우선 순위는?

- 3 -> 1 -> 2 순서

우선 순위를 '시급성'과 '난이도' 모두 둘때는? (23회)

- 시급성 - 현재, 난이도 - Easy (3)

적절하지 않은 설명 (16회, 21회)

- 시급성에 비용측면 없음

▼ 20 데이터 분석을 위한 조직 구조

- **집중형 조직 구조**
 - 조직내에 별도의 독립적인 분석 전담조직이 구성하고 **회사의 모든 분석 업무를 전담 조직에서 담당**함
 - 일부 협업 부서와 분석 업무가 **중복 또는 이원화될 가능성**이 있음
- **기능중심 조직 구조**
 - 별도로 분석 조직을 구성하지 않고 **각 해당 업무부서에서 직접 분석하는 형태**임
 - 일반적인 분석 수행구조, 전사적 핵심 분석이 어려움

- **분산 조직구조**

- 조직의 인력들을 협업부서에 배치가 되어 **신속한 업무에 적합**
- 전사 차원의 우선순위 수행, 부서 분석업무와 역할 분담 명확히 해야함

틀린 것 찾기 (17회)

- 분산된 조직 구조를 신속한 업무에 적합하지 않다고 함 (WRONG)

집중구조 설명 써놓고 찾기 (19회, 21회)

분산구조 설명 써놓고 찾기 (23회)

▼ 21 빅데이터 거버넌스의 특징

- 기업이 가진 과거 및 현재의 모든 데이터를 분석하여 비즈니스 인사이트를 찾는 노력은 비용면에서 효율적이지 못함 -> **분석 대상 및 목적을 명확히 정의하고, 필요한 데이터를 수집, 분석하여 점진적으로 확대해 나가는 것이 좋음**
- 빅데이터 분석에서 품질관리도 중요하지만, **데이터 수명주기 관리방안을 수립하지 않으면 데이터 가용성 및 관리 비용 증대 문제에 직면할 수 있음**
- ERD는 운영 중인 데이터베이스와 일치하기 위해 **계속해서 변경사항을 관리**하여야 함
- 산업 분야별, 데이터 유형별, 정보 거버넌스 **요소별로 구분하여 작성**함
- 적합한 분석 업무를 도출하고 가치를 높여줄 수 있도록 분석 조직 및 인력에 대해 **지속적인 교육과 훈련을 실시**함
- 빅데이터 거버넌스와 비교할 때 데이터 거버넌스와의 차이점
 - 데이터 생명주기 관리, 개인정보보호 및 보안, 데이터 품질기준, 변경관리
 - **데이터 백업 주기 변경은 차이점 아님**

거버넌스(governance)

- "공통의 목표를 달성하기 위해 정부, 시민사회, 기업 등 관련 이해관계자가 함께 의사결정을 해나가는 :

빅데이터 거버넌스의 특징 올바른 것 찾기 (19회, 23회)

빅데이터 거버넌스, 데이터 거버넌스의 차이점에 해당하지 않는 것 찾기 (23회)

- 데이터 백업 주기 변경

▼ 22 데이터 거버넌스

- 데이터 거버넌스에서 데이터 저장소 관리
 - 메타데이터 및 표준 데이터를 관리하기 위한 전사 차원의 저장소를 구성함
 - 저장소는 데이터 관리 체계 지원을 위한 워크프로우 및 관리용 응용 소프트웨어를 지원하고 관리 대상 시스템과의 인터페이스를 통한 통제가 이루어져야 함
 - 데이터 구조 변경에 따른 **"사전영향평가"**도 수행되어야 효율적인 활용이 가능함 (단답형)
- **데이터 거버넌스 체계요소**

- **데이터 표준화**: 데이터 표준용어 설정, 명명규칙 수립, 메타 데이터 구축, 데이터 사전 구축
- **데이터 관리체계**: 메타데이터와 데이터 사전의 관리 원칙 수립
- **데이터 저장소 관리**: 메타데이터 및 표준 데이터를 관리하기 위한 전사 차원의 저장소를 구성
- **표준화 활동**: 데이터 거버넌스 체계 구축 후, 표준 준수 여부를 주기적으로 점검, 모니터링

사전영향평가 주관식 문항 (20회)

데이터 표준화를 고르는 객관식 문제 (16회, 18회)

데이터 표준화의 특징으로 바른 것 찾기 문제 (19회)

▼ 23 데이터 분석 준비도 프레임워크

- 분석 업무 파악에 관한 항목
 - 최적화 분석 업무, 시뮬레이션 분석 업무, 예측 분석 업무
- 분석기법 영역
 - 분석 기법 라이브러리

분석업무파악에 관한 항목 아닌 것 고르기 (17회)

▼ 24 분석 거버넌스 체제 고려 사항

- 분석 기획/관리 및 추진 조직
- 분석 프로세스
- 과제 기획/운영 프로세스
- IT 기술과 마인드 육성체계
- 분석 시스템
- 데이터 거버넌스

분석 거버넌스 체제 고려사항 아닌 것 고르기 (16회, 22회)

- 분석 비용 및 예산 아님!!

▼ 25 분석 성숙도 모델

- 분석 수준 진단 방법 중 조직의 분석 및 활용을 위한 역량 수준을 파악하기 위해 도입 -> (활용) -> 확산 -> 최적화의 분석 성숙도 단계 포지셔닝을 파악하게 된다

활용 : 주관식 괄호 채우기 (23회)

▼ 26 관련 용어

- **Servitization**: 제품과 서비스의 결합을 나타내는 용어
- **CoE (Center of Excellence)**
 - 구성원들이 비즈니스 역량, IT 역량 및 분석 역량을 고루 갖추어야 하며, 협업부서 및 IT 부서와의 지속적인 커뮤니케이션을 수행하는 조직 내 분석 전문조직을 말함
- **정보전략계획(ISP)**
 - 기업의 경영목표 달성에 필요한 전략적 주요 정보를 포착하고, 주요 정보를 지원하기 위해 전사적 관점의 정보 구조를 도출하며, 이를 수행하기 위한 전략 및 실행 계획을 수립하는 전사적인 종합추진 계획
- **프레이밍 효과(Framing Effect)**
 - 동일한 사안이라고 해도 제시되는 방법에 따라 그에 관한 해석이나 의사결정이 달라지는 왜곡 현상

Servitization ; 용어의 뜻을 써주고 찾는 문제 (23회)

CoE : 주관식 문제 출제 (22회)

ISP : 주관식 문항으로 출제됨 (19회, 22회)

프레이밍 효과 : 주관식 문항으로 출제 (20회)