PART

02

데이터 분석 기획

1장) 데이터 분석 기획의 이해

2장) 분석 마스터 플랜

1장. 데이터 분석 기획의 이해

2장. 분석 마스터 플랜

데이터 분석을 위해서는 이해관계자들이 분석의 목표 및 과정을 잘 이해하는 것이 중요하다. 본 장에서는 데이터 분석 기획을 위해 반드시 알아야 하는 분석 방 법론에 대해서 살펴본다. 그리고 다양한 방법으로 분 석 기회를 발굴하여 분석 과제를 정의하는 방법을 살 펴본다. 분석과제를 수행하기 위한 프로젝트 관리방안을 살펴 보고 중·장기적 측면에서 분석 마스터 플랜을 수립하 는 방법과 분석 거버넌스 체계를 어떻게 수립하는지 이해한다.

Learning Map

어떤 것을 학습하게 될지 살펴보자!

1장

데이터 분석 기획의 이해

- 분석 기획 방향성 도출 - 분석 과제 발굴

- 분석 방법론

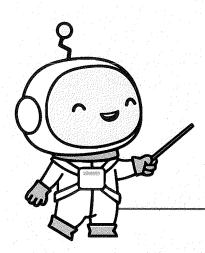
- 분석 프로젝트 관리 방안

2장

분석 마스터 풀랜

- 마스터 플랜 수립 프레임 워크

- 분석 거버넌스 체계 수립



데이터 분석 기획의 이해

4 DAY



2과목 데이터 분석 기획에서는 컨설팅회사나 기업에서 분석 프로젝트를 수행하기 위해 의사결정자를 설득할 때 꼭 필요한 과정을 학습하게 됩니다. 데이터를 분석하기 위해 필요한 방법론과 프로세스를 반드시 알아야 합니다. 또한 실제 비즈니스에서 분석하고자 하는 분석과제를 정의하는 방법과 마스터 플랜을 확정하는 방법을 이해해야 합니다. 대학의 어떤 학과에서도 이런 내용을 가르쳐주지 않기 때문에 용어가 낯설 것입니다. 용어를 익숙하게 하는 것이 가장 중요합니다.



の 計会 呈표

- 분석 기획 방향성 도출을 위한 분석 기획의 특징과 고려사항을 이해한다.
- 분석 방법론 중에서 KDD분석 방법론에 대해 이해한다.
- 분석 방법론 중에서 CRISP-DM 분석 방법론에 대해 이해한다.
- 빅데이터 분석 방법론을 이해하고 각 단계별 내용을 설명할 수 있다.

이 눈높이 체크

✓ 데이터 분석 방법론과 프로세스의 필요성을 이해하시나요?

- 최근 대용량 데이터베이스와 빅데이터를 통해 새로운 인사이트를 도출하고자 하는 시도가 증가하면서 데이터를 분석할 때, 어떤 방법론과 어떤 프로세스로 데이터를 분석하는 것이 효율적인지에 대한 관심이 증가하고 있다.
- 빅데이터나 대용량 데이터의 경우, 분석하고자 하는 목적에 따라 가장 적절한 방법 론을 찾는 것이 가장 중요하다. 또한 대용량 데이터를 분석하는 프로세스에서 중요
 한 과정을 생략하거나 중복할 경우 발생되는 비용은 엄청난 손실로 나타날 수 있기 때문에 효율적인 프로세스를 통해 분석 업무를 수행하는 것이 중요하다.

✔ KDD, CRISP-DM, 빅데이터 분석 방법론에 대해 들어보셨나요?

- o 대용량 데이터베이스와 빅데이터를 분석하기 위해서 어떤 프로세스로 작업을 하는 것이 가장 효율적인지에 대한 연구는 지금도 계속 진행되고 있다.
- o 대용량 데이터베이스를 통해 정형화된 데이터베이스를 분석하는 정형 데이터 마이닝 프로세스로 가장 많이 활용되고 있는 프로세스는 KDD(Knowledge Discovery in Databases) 분석 방법론과 CRISP-DM 프로세스이다.
- o 최근 빅데이터를 통해 대용량이면서 비정형 데이터를 어떤 프로세스를 통해 분석 하는 것이 효과적임지를 고민하면서 빅데이터 분석 프로세스가 발전하고 있다.



1장 데이터 분석 기획의 이해

부식기회 방향성 도출



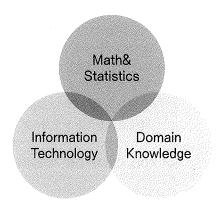
분석기획의 특징

가. 분석기획이란?

- 실제 분석을 수행하기에 앞서 분석을 수행할 **과제를 정의**하고, 의도했던 **결과를 도출**할 수 있도록 이를 적절하게 **관리**할 수 있는 방안을 **사전에 계획**하는 일련의 작업이다.
- 분석과제 및 프로젝트를 직접 수행하는 것은 아니지만, 어떠한 목표(What) 를 달성하기 위하여(Why) 어떠한 데이터를 가지고 어떤 방식으로(How) 수 행할 지에 대한 일련의 계획을 수립하는 작업이기 때문에 성공적인 분석결 과를 도출하기 위한 중요한 사전 작업이다.

나. 데이터 사이언티스트의 역량

• 데이터 사이언티스트는 수학/통계학적 지식 및 정보기술(IT기술, 해킹기술, 통신기술 등) 뿐만 아니라 해당 비즈니스에 대한 이해와 전문성을 포함한 3가지 영역에 대한 고른 역량과 시각이 요구된다.



• 다시 말해, 분석을 기획한다는 것은 해당 문제 영역에 대한 전문성 역량 및 수학/통계학적 지식을 활용한 분석 역량과 분석의 도구인 데이터 및 프로 그래밍 기술 역량에 대한 균형 잡힌 시각을 가지고 방향성 및 계획을 수립 해야 한다는 것을 의미한다.

• 분석은 분석의 대상(What)과 분석의 방법(How)에 따라서 4가지로 나누어진다.

• 특정한 분석 주제를 대상으로 진행할 경우에도, 분석 주제 및 기법의 특성상 이러한 4가지 유형을 넘나들면서 분석을 수행하고 결과를 도출하는 과정을 반복한다.

분석 대상과 방법

출 제 포인트

분석 기획을 위해 필요한 3가지 역량과 분석 대상과 방법에 따른 분석 유형으로 출제가능성이 높습니다.



2

분석의 대			
Known	Un-Known		
Optimization	Insight	Known	분석의 방법
Solution	Discovery	Un-Known	(How)

• 목표시점 별로는 당면한 과제를 빠르게 해결하는 "과제 중심적인 접근 방식과 지속적인 분석 내재화를 위한 "장기적인 마스터 플랜 방식"으로 나눌 수 있다.

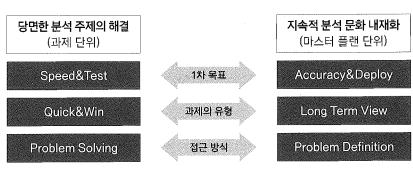
목표 시점 별 분석 기획 방안

• 분석기획에서는 문제해결(Problem Solving)을 위한 단기적인 접근방식과 분석과제 정의(Problem Definition)를 위한 중장기적인 마스터 플랜 접근방식을 융합하여 적용하는 것이 중요하다.



목표 시점별 분석 방향, 목표, 유형, 접근 방식에 대한 특성을 묻는 문제가 출제 될 수 있습니다. 내용을 이해하시기 바랍니다.





〈목표 시점별 분석 기획 방안 〉

• 의미있는 분석을 위해서는 분석 기술, IT 및 프로그래밍, 분석 주제에 대한 도메인 전문성, 의사소통이 중요하고 분석대상 및 방식에 따른 다양한 분석 주제를 과제 단위 혹은 마스터 플랜 단위로 도출할 수 있어야 한다.

0

분석 기획시 고려사항

가. Available Data

- Transaction data
- · Human-generated data
- · Mobile data
- Machine and sensor data 등

나. Proper Business Use Case

- · Customer analytics
- Social media analytics
- Plant and facility management
- · Pipeline management
- · Price optimization
- Fraud detection 등

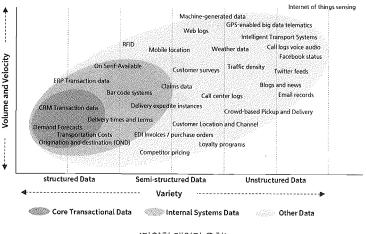
다. Low Barrier of Execution

- Cost
- Simplicity
- Performance
- Culture 등

성공적 분석

예시

- 정형 데이터(Structured Data, DB로 정제된 데이터)
- 반정형 데이터(Semi-structured Data, 센서 중심으로 스트리밍되는 머신데이터)
- 비정형 데이터(Unstructured Data, email, 보고서, 소셜미디어 데이터)





〈다양한 데이터 유형〉

- 가. 분석의 기본인 **가용** 데이터(Available Data)에 대한 고려가 필요하다.
 - 분석을 위한 데이터의 확보가 우선적이며, 데이터의 유형에 따라 적용 가능 한 솔루션 및 분석 방법이 다르기 때문에 유형에 대한 분석이 선행적으로 이루어져야 한다.
- 나. 분석을 통해 가치가 창출될 수 있는 적절한 활용방안과 유즈케이스

(Proper Business Use Case) 탐색이 필요하다.

• "바퀴를 재발명하지 마라"라는 격언처럼 기존에 잘 구현되어 활용되고 있는 유사 분석 시나리오 및 솔루션을 최대한 활용하는 것이 중요하다.

다. 분석 수행시 발생하는 장애요소들에 대한 사전계획 수립이 필요하다.

(Low Barrier Of Execution)

• 일회성 분석으로 그치지 않고 조직의 역량으로 내재화하기 위해서는 충분하고 계속적인 교육 및 활용방안 등의 변화 관리(Change Management)가고려되어야 한다.

참고

	종류	정형 데이터	반정형 데이터	비정형 데이터
	특징	· 데이터 자체로 분석 가능 · RDB구조의 데이터 · 데이터베이스로 관리	· 데이터로 분석이 가능하 지만 해석이 불가능하며 메타정보를 활용해야 해 석이 가능	 데이터 자체로 분석이 불가능 특정한 처리 프로세스를 거쳐 분석데이터로 변경 후 분석
b	유형	· ERP, CRM, SCM 등 정보시스템	· 로그데이터, 모바일데이터, 센싱데이터	· 영상, 음성, 문자 등
_				





1장 데이터 부석 기회의 이해

분석 방법론



데이터 분석 방법론을 정의할 때 반드시 필요한 내용인 절차, 방법, 도구와 기법, 템 플릿과 산출물을 기억하고 적용업무의 특성에 따른 모델들도 기억하시기 바랍니다.





분석 방법론 개요

가. 개요

- 데이터 분석이 효과적으로 기업 내에 정착하기 위해서는 이를 **체계화한 절 차와 방법**이 정리된 데이터 **분석 방법론의 수립**이 필수적이다.
- 프로젝트는 개인의 역량이나 조직의 우연한 성공에 기인해서는 안 되고, 일 정한 수준의 품질을 갖춘 산출물과 프로젝트의 성공 가능성을 확보하고 제시할 수 있어야 한다.
- 방법론은 상세한 **절차(Procedures)**, **방법(Methods)**, 도구와 기법 (Tools&Techniques), 템플릿과 산출물(Templates&Outputs)로 구성 되어 어느 정도의 지식만 있으면 활용이 가능해야 한다.



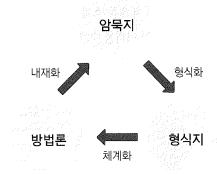
데이터 기반 의사결정의 필요성에서 합리적 의사결정을 가로막는 장애요소에 대한 문제가 자주 출제되오니 기억하시기 바랍니다.



나. 데이터 기반 의사결정의 필요성

- 1) 경험과 감에 따른 의사결정 → 데이터 기반의 의사결정
- 2) 기업의 합리적 의사결정을 가로막는 장애요소 : 고정 관념(Stereotype), 편향된 생각(Bias), 프레이밍 효과(Framing Effect : 문제의 표현 방식에 따라동일한 사건이나 상황임에도 불구하고 개인의 판단이나 선택이 달라질 수 있는현상) 등

다. 방법론의 생성과정



甲基	의미	예	특징	상호작용
암묵지	학습과 경험을 통해 개인에게 체화되어 있지만 겉으로 드러 나지 않는 지식	김치 담그기, 자전거 타기	사회적으로 중요하지 만 다른 사람에게 공유 되기 어려움	공통화 내면화
형식지	문서나 매뉴얼처럼 형상화된 지식	교과서, 비디오, DB	전달과 공유가 용이함 	표출화 연결화



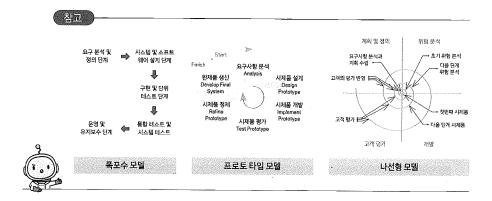
- 1) 폭포수 모델(Waterfall Model)
 - 단계를 순차적으로 진행하는 방법으로, 이전 단계가 완료되어야 다음 단계로 진행될 수 있으며 문제가 발견될 시 피드백 과정이 수행된다.(기존 IT의 SW개발 방식)

2) 프로토타입 모델(Prototype Model)

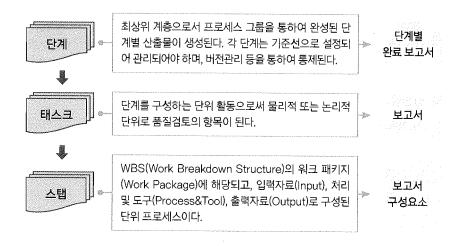
• 폭포수 모델의 단점을 보완하기 위해 점진적으로 시스템을 개발해 나가 는 접근 방식으로, 고객의 요구를 완전하게 이해하고 있지 못하거나 완 벽한 요구 분석의 어려움을 해결하기 위해 일부분을 우선 개발하여 사용자에게 제공한다. 시험 사용 후 사용자의 요구를 분석하거나 요구 정 당성을 점검, 성능을 평가하여 그 결과를 통한 개선 작업을 시행하는 모델이다.

3) 나선형 모델(Spiral Model)

• 반복을 통해 점증적으로 개발하는 방법으로, 처음 시도하는 프로젝트 에 적용이 용이하지만 관리 체계를 효과적으로 갖추지 못한 경우 복잡도 가 상승하여 프로젝트 진행이 어려울 수 있다.



마. 방법론의 구성





KDD 분석 방법론

가. 개요

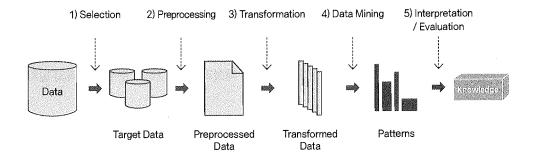
• KDD(Knowledge Discovery in Databases)는 1996년 Fayyad가 프로파 일링 기술을 기반으로 데이터로부터 통계적 패턴이나 지식을 찾기 위해 활 용할 수 있도록 체계적으로 정리한 데이터 마이닝 프로세스이다. 데이터마 이닝, 기계학습, 인공지능, 패턴인식, 데이터 시각화 등에서 응용 될 수 있는 구조를 갖고 있다.



프로세스에 해당하거나 해당하지 않는 분석절차에 대한 부분과 분석절차 설명 후 어느 단계에 대한 설명인지 맞추는 부분이 자주 출제되므로 기억하시기 바랍니다.



나. KDD 분석 절차



1) 데이터셋 선택(Selection)

- 데이터셋 선택에 앞서 분석 대상의 비즈니스 도메인에 대한 이해와 프로 젝트 목표 설정이 필수이며 데이터베이스 또는 원시 데이터에서 분석에 필요한 데이터를 선택하는 단계이다.
- 데이터마이닝에 필요한 목표데이터(Target Data)를 구성하여 분석에 활용한다.

2) 데이터 전처리(Preprocessing)

- 추출된 분석 대상용 데이터 셋에 포함되어 있는 **잡음(Noise)과 이상치** (Outlier), 결측치(Missing Value)를 식별하고 필요시 제거하거나 의미 있는 데이터로 재처리하여 데이터 셋을 정제하는 단계이다.
- 데이터 전처리 단계에서 **추가로 요구되는** 데이터 셋이 필요한 경우 데이터 선택 프로세스를 재실행한다.

3) 데이터 변환(Transformation)

- 데이터 전처리 과정을 통해 정제된 데이터에 분석 목적에 맞게 변수를 생성, 선택하고 데이터의 차원을 축소하여 효율적으로 데이터마이닝을 할수 있도록 데이터에 변경하는 단계이다.
- 데이터마이닝 프로세스를 진행하기 위해 학습용 데이터(Training Data) 와 시험용 데이터(Test Data)로 데이터를 분리하는 단계이다.

4) 데이터 마이닝(Data Mining)

- 학습용 데이터를 이용하여 분석목적에 맞는 데이터마이닝 기법을 선택하고, 적절한 알고리즘을 적용하여 데이터마이닝 작업을 실행하는 단계이다.
- 필요에 따라 데이터 **전처리**와 데이터 **변환 프로세스**를 추가로 실행하여 최 적의 결과를 산출한다.

5) 데이터 마이닝 결과 평가(Interpretation/Evaluation)

- 데이터마이닝 결과에 대한 해석과 평가, 그리고 분석 목적과의 일치성을 확인한다.
- 데이터마이닝을 통해 발견한 지식을 업무에 활용하기 위한 방안 마련의 단계이다.
- 필요에 따라 데이터 선택 프로세스에서 데이터마이닝 프로세스를 반복 수행한다.

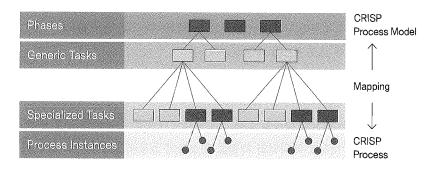


CRISP-DM 분석 방법론

가. 개요

CRISP-DM(Cross Industry Standard Process for Data Mining)은 1996년
 유럽연합의 ESPRIT에서 있었던 프로젝트에서 시작되었으며, 주요한 5개의 업체들(Daimler-Chrysler, SPSS, NCR, Teradata, OHRA)이 주도하였다.
 CRISP-DM은 계층적 프로세스 모델로써 4개 레벨로 구성된다.

나. CRISP-DM의 4레벨 구조



- 최상위 레벨은 여러 개의 단계(Phases)로 구성되고 각 단계는 일반화 태스 크(Generic Tasks)를 포함한다. 일반화 태스크는 데이터마이닝의 단일 프로세스를 완전하게 수행하는 단위이며, 이는 다시 구체적인 수행 레벨인 세분화 태스크(Specialized Tasks)로 구성된다.
- 예를 들어 데이터 정제(Data Cleansing)라는 일반화 태스크는 범주형 데이터 정제와 연속형 데이터 정제와 같은 세분화 태스크로 구성된다.
- 마지막 레벨인 프로세스 실행(Process Instances)은 데이터마이닝을 위한 구체적인 실행을 포함한다.

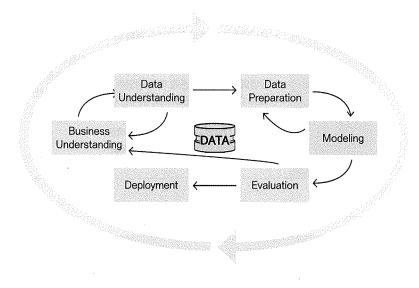


CRISP-DM 프로세스의 4가지 레벨과 6단계 그리고 각 단계별 업무 내용이 시험에 자주 출제되는 부분입니다. 반드시 기억해 두세요.



다. CRISP-DM의 프로세스

• CRISP-DM 프로세스는 6단계로 구성되어 있으며, 각 단계는 단방향으로 구성 되어 있지 않고 **단계 간 피드백**을 통하여 단계별 완성도를 높이게 되어 있다.



단계	내용	수행업무
업무이해 Business Understanding	 비즈니스 관점에서 프로젝트의 목적과 요구사항을 이해하기 위한 단계 도메인 지식을 데이터 분석을 위한 문제 정의로 변경하고 초기 프로젝트 계획을 수립하는 단계 	업무 목적 파악, 상황 파악, 데이터 마이닝 목 표 설정, 프로젝트 계획 수립

데이터 이해 Data Understanding	 분석을 위한 데이터를 수집하고 데이터 속성을 이해하기 위한 단계 데이터 품질에 대한 문제점을 식별하고 숨겨져 있는 인사이트를 발견하는 단계 	초기 데이터 수집, 데 이터 기술 분석, 데이터 탐색, 데이터 품질 확인
데이터 준비 Data Preparation	 분석을 위하여 수집된 데이터에서 분석 기법에 적합한 데이터를 편성하는 단계 (많은 시간이 소요 될 수 있음) 	분석용 데이터 셋 선택, 데이터 정제, 분석용 데이터 셋 편성, 데이터 통합, 데이터 포맷팅
모델링 Modeling	 다양한 모델링 기법과 알고리즘을 선택하고 모델링 과정에서 사용되는 파라미터를 최적화해 나가는 단계 모델링 과정에서 데이터 셋이 추가로필요한 경우 데이터 준비 단계를 반복수행할 수 있으며, 모델링 결과를 테스트용 데이터 셋으로 평가하여 모델의과적합(Overfitting) 문제를 확인 	모델링 기법 선택, 모 델 테스트 계획 설계, 모델 작성, 모델 평가
평가 Evaluation	 모델링 결과가 프로젝트 목적에 부합하는지 평가하는 단계로 데이터마이닝 결과를 최종적으로 수용 할 것인지 판단 	분석결과 평가, 모델 링 과정 평가, 모델 적용성 평가
전개 Deployment	 모델링과 평가 단계를 통하여 완성된 모델을 실 업무에 적용하기 위한 계획 을 수립하는 단계 모니터링과 모델의 유지보수 계획 마련 → 모델에 적용되는 비즈니스 도메인 특성, 입력되는 데이터의 품질 편차, 운영모델의 평가기준에 따라 생명주기(Life Cycle)가 다양하므로 상세한 전개계획이 필요 CRISP-DM의 마지막 단계, 프로젝트 종료 관련 프로세스를 수행하여 프로젝트 마무리 	전개 계획 수림, 모니 터링과 유지보수 계획 수립, 프로젝트 종료 보고서 작성, 프로젝 트 리뷰



KDD와 CRISP-DM≌ ЫIJ

KDD	CRISP-DM
분석대상 비즈니스 이해	업무 이해(Business Understanding)
데이터셋 선택(Data Selection)	데이터의 이해(Data Understanding)
데이터 전처리(Preprocessing)	⊔ o c ≃ o o (Data Onderstanding)
데이터 변환(Transformation)	데이터 준비(Data Preparation)
데이터 마이닝(Data Mining)	모델링(Modeling)
데이터 마이닝 결과 평가 (Interpretation/Evaluation)	평가(Evaluation)
데이터 마이닝 활용	전개(Deployment)

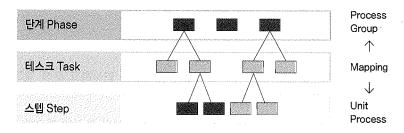
가. 빅데이터 분석의 계층적 프로세스



빅데이터 분석을 위한 3개의 계층과 내용이 시험 에 자주 출제됩니다.







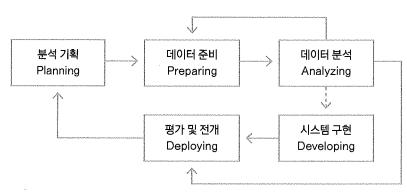
- 1) 단계(Phase): 프로세스 그룹(Process Group)을 통하여 완성된 단계별 산출물이 생성된다. 각 단계는 기준선(Baseline)으로 설정되어 관리되어야 하며, 버전관리(Configuration Management) 등을 통하여 통제가 이루어져야 한다.
- 2) 태스크(Task): 각 단계는 여러 개의 태스크(Task)로 구성된다. 각 태스크는 단계를 구성하는 단위 활동이며, 물리적 또는 논리적 단위로 품질 검토의 항목이 될 수 있다.
- 3) 스텝(Step): WBS(Work Breakdown Structure)의 워크 패키지(Work Package)에 해당되고 입력자료(Input), 처리 및 도구(Process&Tool), 출력자료(Output)로 구성된 단위 프로세스(Unit Process)이다.

나. 빅데이터 분석 방법론 - 5단계



빅데이터 분석 방법론의 5단계와 각 단계별 주요 업무는 이 장에서 가장 중요하고 시험에도 자주 나오는 부분입니다. 아래 그림을 반드시 기억하셔야 합니다.





문석 기획	데이터 순비	데이터 문석	시스템 구현	평가 및 선개
비지니스 이해 및 범위 설정 프로젝트 정의 및 계획 수립 프로젝트 위험 계획 수립	필요 데이터 정의 데이터 스토어 설계 데이터 수집 및 정합성 점검	분석용 데이터 준비 텍스트 분석 탐색적 분석 모델링 모델 평가 및 검증 모델 적용 및 운영방안 수립	・설계 및 구현 ・시스템 테스트 및 운영	모델 발전계획 수립 프로젝트 평가 및 보고

- 1) 분석기획(Planning): 비즈니스 도메인과 문제점을 인식하고 분석 계획 및 프로젝트 수행계획을 수립하는 단계이다.
- 2) 데이터 준비(Preparing) : 비즈니스 요구사항과 데이터 분석에 필요한 원천 데이터를 정의하고 준비하는 단계이다.
- 3) 데이터 분석(Analyzing): 원천 데이터를 분석용 데이터 셋으로 편성하고 다양한 분석 기법과 알고리즘을 이용하여 데이터를 분석하는 단계이다. 분석 단계를 수행하는 과정에서 추가적인 데이터 확보가 필요한 경우 데이터 준비 단계로 피드백(Feedback)하여 두 단계를 반복하여 진행한다.
- 4) 시스템 구현(Developing): 분석 기획에 맞는 모델을 도출하고 이를 운영 중인 가동 시스템에 적용하거나 시스템 개발을 위한 사전 검증으로 프로토 타입 시스템을 구현한다.
- 5) 평가 및 전개(Deploying): 데이터 분석 및 시스템 구현 단계를 수행한 후, 프로젝트의 성과를 평가하고 정리하거나 모델의 발전 계획을 수립하여 차기 분석 기획으로 전달하고 프로젝트를 종료하는 단계이다.



분석 기획의 세부단계와 실제 업무를 정확히 이해 해야 하며 업무별 출력자료를 기억하셔야 합니다.



다. 단계별 세부단계 및 실제 업무

1) 분석 기획(Planning)

분석 기획	데이터 준비	데이터 분석	시스턴	ll 구현 평가 및 전개
비지니스 이해 및 범위 설정	\rightarrow	비즈니스 이해	->	프로젝트 범위 설정
프로젝트 정의 및 계획 수립	\rightarrow	데이터 분석 프로젝트 정의		프로젝트 수행 계획 수립
프로젝트 위험계획 수립	$\rightarrow \boxed{}$	데이터 분석 위험 식별	$\bigg]\to\bigg $	위험 대응 계획 수립

	계	내용	입력자료	프로세스 및 도구	출력자료
비즈니스 이해 및 범위설정	비즈니스 이해	내부 업무 매뉴얼과 관련자료, 외부의 관련 비즈니스 자 료를 조사하고 향후 프로젝트 진행을 위한 방향을 설정 한다.		자료 수집 및 비즈니스 이해	비즈니스 이해 및 도메인 문제점
	프로젝트 범위 설정	빅데이터 분석 프로젝트의 대상인 비즈니스에 대한 이해 와 프로젝트 목적에 부합하는 범위를 설정하고 프로젝트 범위 정의서인 SOW(Statement Of Work)를 작성한다.	- 중장기 계획서 - 빅데이터 분석 프로젝트 지시서 - 비즈니스 이해 및 도메인 문제점	- 자료 수집 및 비즈니스 이해 - 프로젝트 범위 정의서 작성 절차	프로젝트 범위 정의서(SOW)
(12) (12) (2) (2) (2) (2) (2) (2) (2) (2) (2) (2) (2) (2) (2)	데이터 분석 프로젝트 정의	프로젝트의 목표 및 KPI, 목표 수준 등을 구체화하여 상 세 프로젝트 정의서를 작성하고 프로젝트 목표를 명확화 하기 위한 모델 운영 이미지 및 평가 기준을 설정한다.	- 프로젝트 범위 정의서(SOW) - 빅데이터 분석 프로젝트 지시서	- 프로젝트 목표 구체화 - 모델 운영 이미지 설계	- 프로젝트 정의서 - 모델 운영 이미지 설계서 - 모델 평가 기준
정의 및 계획수립	프로젝트 수행 계획수립	프로젝트의 목적 및 배경, 기대효과, 수행방법, 일정 및 추진조직, 프로젝트 관리방안을 작성하고 WBS는 프로 젝트 산출물 위주로 작성되어 프로젝트의 범위를 명확하 게 한다.	- 프로젝트 범위 정의서(SOW) - 모델 운영 이미지 설계서 - 모델 평가 기준	- 프로젝트 범위정의서 (SOW) - WBS 작성	- 프로젝트 수행계획서 - WBS
프로젝트 위험계획	데이터 분석 위험 식별	앞서 진행된 프로젝트 산출물과 정리 자료를 참조하고 전 문가의 판단을 활용해 프로젝트를 진행하며 발생 가능한 위험을 식별한다. 식별된 위험은 위험의 영향도와 빈도, 발생가능성에 따라 위험의 우선순위를 설정한다.	- 프로젝트 범위 정의서(SOW) - 프로젝트 수행 계획서 - 선행 프로젝트 산출물 및 정리자료	- 위험 식별 절차 - 위험 영향도 및 발생 가능성 분석 - 위험 우선순위 판단	식별된 위험 목록
수립	위험 대응 계획 수립	식별된 위험은 상세한 정량적, 정성적 분석을 통해 위험 대응방안을 수립한다. 예상되는 위험에 대해 회피 (Avoid), 전이(Transfer), 완화(Mitigate), 수용(Accept) 으로 구분하여 위험관리 계획서를 작성한다.	- 식별된 위험 목록 - 프로젝트 범위 정의서(SOW) - 프로젝트 수행계획서	- 위험 정량적 분석 - 위험 정성적 분석	위험관리 계획서

2) 데이터 준비(Preparing)

분석기획 테이	기터 준비	데이터 분석	시스팀	비구현 평가 및 전개 -
필요 데이터 정의	\rightarrow	데이터 정의	$\bigg] \to$	데이터 획득방안 수립
데이터 스토어 설계		데이터 스토어 설계	$\bigg] \to$	비정형 데이터 스토어 설계
데이터 수집 및 정합성 점검	→ G (이터 수집 및 저장] ->	데이터 정합성 점검

단계	10-10	내용	입력자료	프로세스 및 도구	출력자료
필요	데이터 정의	시스템, 데이터베이스, 파일, 문서 등의 다양한 내·외부 원천 데이터 소스(Raw Data Source)로부터 분석에 필 요한 데이터를 정의한다.	- 프로젝트 수행 계획서 - 시스템 설계서 - ERD - 메타데이터 정의서 - 문서 자료	- 내·외부 데이터 정의 - 정형·비정형·반정형 데이터 정의	데이터 정의서
	데이터 희득방안 수립	-내·외부의 다양한 데이터 소스로부터 정형·비정형·반정 형 데이터를 수집하기 위한 구체적인 방안을 수립한다. - 내부 데이터 획득에는 부서 간 업무협조와 개인정보보 호 및 정보 보안과 관련한 문제점을 사전에 점검한다. 외부 데이터의 획독은 다양한 인터페이스 및 법적인 문 제점을 고려하여 상세한 데이터 획득 계획을 수립한다.	- 데이터 정의서 - 시스템 설계서 - ERD - 메타데이터 정의서 - 문서 자료 - 데이터 구입	데이터 획득 방안 수립	데이터 획득 계획서
	정형 데이터 스토어 설계	정형 데이터는 일반적으로 관계형 데이터베이스인 RDBMS(Relational Data Base Management System) 룰 사용하고 데이터의 효율적인 저장과 활용을 위하여 데 이터스토어의논리적, 물리적 설계를 구분하여 설계한다.	- 데이터 정의서 - 데이터 획득 계획서	- 데이터베이스 논리, 물리 설계 - 데이터 매핑	- 정형 데이터 스토어설계서 - 데이터 매핑 정의서
설계	비정형 데이터 스토어 설계	하둡, NoSQL 등을 이용하여 비정형 또는 반정형 데이터를 저장하기 위한 논리적, 물리적 데이터 스토어를 설계한다.	- 데이터 정의서 - 데이터 획득 계획서	비정형·반정형 데이터 논리, 물리 설계	- 비정형 데이터 스토어 설계서 - 데이터 매핑 정의서
	데이터 수집 및 저장	크롤링 등 데이터 수집을 위한 ETL 등과 같은 다양한 도 구와 API, 스크립트(Script) 프로그램 등을 이용하여 데 이터를 수집하고, 수집된 데이터를 설계된 데이터 스토어 에 저장한다.	- 데이터 정의서 - 데이터 획득 계획서 - 데이터 스토어 설계서	- 데이터 크롤링 도구 - ETL 도구 - 데이터 수집 스크립트	수집된 분석용 데이터
점검	데이터 정합성 점검	데이터 스토어의 품질 점검을 통하여 데이터의 정합성을 확보하고 데이터 품질개선이 필요한 부분에 대하여 보완 작업을 한다.	수집된 분석용 데이터	- 데이터 품질 확인 - 정합성 점검 리스트	정합성 점검 보고서

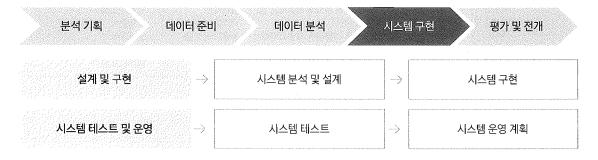
3) 데이터 분석(Analyzing)

분석 기회	데이터 준비	데이터 분석	시스템 구현	면 평가 및 전개
분석용 데이터 준비	\rightarrow	비즈니스 룰 확인	->	분석용 데이터셋 준비
텍스트 분석	$\rightarrow $	텍스트 데이터 확인 및 추출		텍스트 데이터 분석
탐색적 분석	\rightarrow	탐색적 데이터 분석] -> [데이터 시각화
모델링	$\rightarrow \boxed{}$	데이터 분할	→ [데이터 모델링
모델 평가 및 검증	$\rightarrow $	모델 평가	_> [모델 검증

Ę	:계	내용	입력자료	프로세스 및 도구	출력자료
분석용 데이터 준비	비즈니스 룰 확인	- 비즈니스 이해, 도메인 문제점 인식, 프로 젝트 정의 등을 이용하여 프로젝트의 목표를 정확하 게 인식한다. - 세부적인 비즈니스 룰을 파악하고 분석에 필요한 데 이터의 범위를 확인한다.	- 프로젝트 정의서 - 프로젝트 수행 계획서 - 데이터 정의서 - 데이터 스토어	- 프로젝트 목표 확인 - 비즈니스 룰 확인	- 비즈니스 룰 - 분석에 필요한 데이터 범위
분석용 데이터 준비	분석용 데이터 셋 준비	- 데이터 스토어로부터 분석에 필요한 정형·비정형 데 이터를 추출한다. 필요시 적절한 가공을 통하여 분석 도구 입력 자료로 사용될 수 있도록 편성한다. - 추출된 데이터는 데이터베이스나 구조화된 형태로 구 성하고 필요시 분석을 위한 작업 공간(Play Ground, Sandbox 등)과 전사 차원의 데이터 스토어로 분리할 수도 있다.	- 데이터 정의서 - 데이터 스토어	- 데이터 선정 - 데이터 변환 - ETL 도구	분석용 데이터 셋
	텍스트 데이터 확인 및 추출	데이터 스토어(Data Store)에서 필요한 텍스트 데이터 를 추출한다.	비정형 데이터 스토어	- 분석용 텍스트 데이터 확인 - 텍스트 데이터 추출	분석용 텍스트 데이터
텍스트 분석	텍스트 데이터 분석	 추출된 텍스트 데이터를 분석 도구로 적재하여 다양한 기법을 분석하고 모델을 구축한다. 텍스트 분석을 위해 용어사전을 사전에 확보하고 업무 도메인에 맞도록 작성해야한다. 구축된 모델은 시각화 도구를 이용하여 모델의 의미전달을 명확하게한다. 	- 분석용 텍스트 데이터 - 용어사전 (유의어 사전, 불용어 사전 동)	- 분류체계 설계 - 형태소분석 - 키워드 도출 - 토픽분석 - 감성분석, 의견분석 - 네트워크 분석	텍스트 분석 보고서

	탐색적 데이터 분석	다양한 관점 별로 기초 통계량 (평균, 분산, 표준편차, 최대값, 최소값)을 산출하고, 데이터의 분포와 변수간 의 관계 등 데이터 자체의 특성 및 데이터의 통계적 특 성을 이해하고 모델링을 위한 기초자료로 활용한다.	분석용 데이터 셋	- EDA도구 - 통계분석 - 연관성 분석 - 데이터 분포 확 인	데이터 탐색 보고서
탐색적 분석	데이터 시각화	- 탐색적 데이터 분석을 위한 도구로 활용한다. 그러나 모델의 시스템화를 위한 시각화를 목적으로 활용할 경우 시각화 기획, 시각화 설계, 시각화 구현 등의 별도의 프로세스를 따라 진행되어야 한다. - 탐색적 데이터 분석을 진행하면 수행된 데이터 시각화는 모델링 또는 향후 시스템 구현을 위한 사용자 인터페이스 또는 프로토타입(Prototype)으로 활용될 수도 있다.	분석용 데이터 셋	- 시각화 도구 - 시각화 패키지 - 인포그래픽 - 시각화 방법론	데이터 시각화 보고서
	데이터 분할	모델의 과적합과 일반화를 위하여 분석용 데이터 셋을 모델 개발을 위한 훈련용 데이터와 모델의 검증력을 테 스트하기 위한 테스트용 데이터로 분할한다. 모델에 적 용하는 기법에 따라 데이터 분할 또는 검증 횟수, 생성 모델 개수 등을 설정한다.	분석용 데이터 셋	데이터 분할 패키지	- 훈련용 데이터 - 테스트용 데이터
모델링	데이터 모델링	기계학습 등을 이용한 데이터 모델링은 훈련용 데이터 를 활용하여 분류, 예측, 군집 등의 모델을 만들어 가 동중인 운영 시스템에 적용한다. 필요시 비정형 데이 터 분석결과를 통합적으로 활용하여 프로젝트 목적에 맞는 통합 모델을 수행한다.	분석용 데이터 셋	- 통계 모델링 기 법 - 기계학습 - 모델 테스트	모델링 결과 보고서
	모델 적용 및 운영 방안	모델을 가동중인 운영시스템에 적용하기 위해서는 모델에 대한 상세한 알고리즘 설명서 작성이 필요하다. 알고리즘 설명서는 시스템 구현 단계에서 중요한 입력 자료로 활용되므로 필요시 의사코드 수준의 상세한 작성이 필요하다. 또한 모델의 안정적 운영을 모니터링하는 방안도 수립한다.	모델링 결과 보고서	- 모니터링 방안 수립 - 알고리즘 설명 서 작성	- 알고리즘 설명서 - 모니터링 방안
모델 평가	모델 평가	프로젝트 정의서의 모델 평가 기준에 따라 모델을 객관 적으로 평가하고 품질관리 차원에서 모델 평가 프로세 스를 진행한다. 모델 평가를 위해 모델 결과 보고서 내 의 알고리즘을 파악하고 테스트용 데이터나 필요시 모 델 검증을 위한 별도의 데이터를 활용한다.	- 모델링 결과 보고서 - 평가용 데이터	- 모델평가 - 모델 품질관리 - 모델 개선작업	모델 평가 보고서
및 검증	모델 검증	모델의 실적용성을 검증하기 위해 검증용 데이터를 이용해 모델 검증 작업을 실시하고 모델링 검증 보고서를 작성한다. 검증용 데이터는 모델 개발 및 평가에 활용된 훈련용이나 테스트용 데이터가 아닌 실 운영용 데이터 를 확보하여 모델의 품질을 최종 검증한다.	- 모델링 결과 보고서 - 모델 평가 보고서 - 검증용 데이터	모델 검증	모델 검증 보고서

4) 시스템 구현(Developing)



	계	내용	입력자료	프로세스 및 도구	출력자료
설계	시스템 분석 및 설계	가동중인 시스템을 분석하고 알고리즘 설명서에 근거하 여 응용시스템 구축 설계 프로 세스를 진행한다. 시스템 분석과 설계는 사용 중인 정보시스템 개발 방법론을 커스 터마이징하여 적용할 수 있다.	- 알고리즘 설명서 - 운영중인 시스템 설계서	정보시스템 개발 방법론	시스템 분석 및 설계 서
및 구현	시스템 구현	시스템 분석 및 설계서에 따라 BI 패키지를 활용하거나 새 롭게 시스템을 구축하거나 가동중인 운영 시스템의 커스 터마이장 등을 통해 설계된 모델을 구현한다.	- 시스템 분석 및 설계서 - 알고리즘 설명서	- 시스템 통합 개발 도구(IDE) - 프로그램 언어 - 패키지	구현 시스템
시스템 테스트	시스템 테스트	구축된 시스템의 검증을 위하여 단위 테스트, 통합 테스트, 시스템 테스트 등을 실시한다. 시스템 테스트는 품질 관리 차원에서 진행함으로써 적용된 시스템의 객관성과 완전성을 확보한다.		품질관리 활동	시스템 테스트 결과보고서
및 운영	시스템 운영 계획	구현된 시스템을 지속적으로 활용하기 위해 시스템 운영 자, 사용자를 대상으로 필요한 교육을 실시하고 시스템 운영계획을 수립한다.	- 시스템 분석 및 설계서 - 구현 시스템	- 운영계획 수립 - 운영자 및 사용자교육	- 운영자 매뉴얼 - 사용자 매뉴얼 - 시스템 운영 계획서

5) 평가 및 전개(Deploying)

분석 기획	데이터 준비	데이터 분석		네 구현 평가 및 전개
모델 발전 계획 수	-d →	모델 발전 계획		
프로젝트 평가 및 5	≠12 →	프로젝트 성과 평가	->	프로젝트 종료

단계	내용	입력자료	프로세스 및 도구	출력자료
모델 발전 모델 발전 계획 수립 계획	개발된 모델의 지속적인 운영과 기능 향상을 위한 발전계획을 상세하게 수립하여 모델의 계속성을 확보한다.	i	모델 발전 계획 수립	모델 발전 계획서
프로젝트 성과 평가 프로젝트	프로젝트의 정량적 성과와 정성적 성과로 나누어 성과 평가서를 작성한다.	- 프로젝트 산출물 - 품질관리 산출물 - 프로젝트 정의서 - 프로젝트 수행 계획서	- 프로젝트 평가기준 - 프로젝트 정량적 평가 - 프로젝트 정성적 평가	프로젝트 성과 평가서
평가 및 보고 프로젝트 종료	프로젝트 진행과정의 모든 산출물 및 프로세스를 지식 자산화하고 최종 보고서를 작성하여 의사소 통 절차에 따라 보고하고 종료한다.	- 프로젝트 산출물 - 품질관리 산출물 - 프로젝트 정의서 - 프로젝트 수행 계획서 - 프로젝트 성과 평가서	- 프로젝트 지식자산화 작업 - 프로젝트 종료	프로젝트 최종 보고서



1장 데이터 분석 기획의 이해



분석 과제 발굴은 비즈니스 상에서 필요한 데이터 분석 과제를 발굴하는 방법론을 설명하고 있다. 분석 대상과 방법을 활용한 분석 방법론은 시험에 자주 나오는 부분입니다. 꼭 기억하세요.

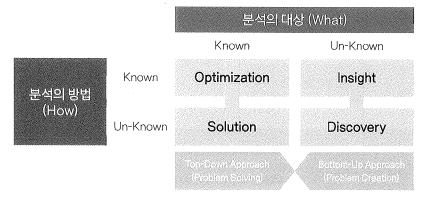




분석과제 발굴 방법론

가. 개요

- 분석 과제는 풀어야 할 다양한 문제를 데이터 분석 문제로 변환한 후 관계 자들이 이해하고 프로젝트로 수행할 수 있는 **과제 정의서 형태로 도출**된다.
- 분석과제를 도출하기 위한 방식으로는 크게 **하향식 접근 방법**(Top Down Approach)과 **상향식 접근 방법**(Bottom up Approach)이 있다.

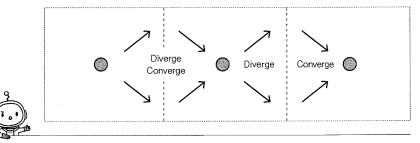


〈 분석 과제 도출의 2가지 유형 〉

- 문제가 주어져 있는 상태에서 답을 구하는 하향식 접근 방식이 전통적으로 수행되었던 분석 과제 발굴 방식이다. 그러나 대규모의 다양한 데이터를 생 성하고 빠르게 변하는 기업 환경에서는 문제 자체의 변화가 심해 정확하게 문제를 사전에 정의하는 것이 어려워지고 있다.
- 분석 과제 발굴을 두가지 방식으로 나누었지만, 실제 새로운 상품 개발이나 전략 수립 등 중요한 의사결정을 할 때 하향식 접근 방법과 상향식 접근 방 법이 혼용되어 사용되며, 분석의 가치를 높일 수 있는 최적의 의사결정은 두 접근 방식이 상호 보완 관계에 있을 때 가능하다.

• 디자인 사고(Design Thinking) :

상향식 접근 방식의 발산 단계와 하향식 접근 방식의 수렴단계를 반복적으로 수행하는 식의 상호 보완적인 동적 환경을 통해 분석의 가치를 높일 수 있는 최적의 의사 결정 방식



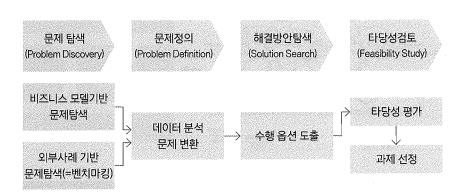
나. 하향식 접근법(Top Down Approach)

• 하향식 분석 접근법은 현황 분석을 통해 기회나 문제를 탐색(Problem Discovery)하고, 해당 문제를 정의(Problem Definition), 해결방안을 탐 색(Solution Search)한다. 그리고 데이터 분석의 타당성 평가(Feasibility Study)를 거쳐 분석 과제를 도출하는 과정으로 구성된다.



하향식 접근방식의 정의와 프로세스는 시험에 자주 나오 는 부분입니다. 특히 문제 탐색 단계는 반드시 기억하세요.





〈데이터 분석 기획 단계(하향식 접근법) 〉

1) 문제 탐색(Problem Discovery)단계 - 하향식 접근법 1단계

- 전체적인 관점의 기준 모델을 활용하여 빠짐없이 문제를 도출하고 식별 하는 것이 중요하다.
- 전체적인 관점의 기준 모델로는 기업 내·외부 환경을 포괄하는 비즈니스 모델과 외부 참조 모델이 존재하다.
- 과제 발굴 단계에서는 세부적인 구현 및 솔루션에 초점을 맞추는게 아니라, 문제를 해결함으로써 발생하는 가치에 중점을 두는 것이 중요하다.



비즈니스 모델 캔버스를 활용한 과제 발굴 방법 5가지 영역이 시험에 자주 출제 됩니다.

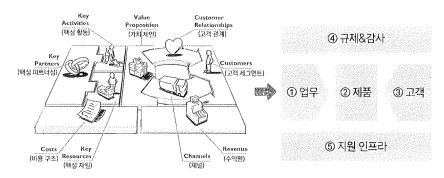


비즈니스 모델 캔버스 채널 (Channel)의 특징

- · 기업이 제공하는 상품이나 서비스에 대한 고객의 이해 를 높임
- · 기업이 전달하는 밸류 프로 포지션을 고객들이 평가할 수 있도록 함
- · 고객이 특정한 상품이나 서 비스를 구매하게 도와줌
- · 고객에게 밸류 프로포지션 을 전달
- · 구매 고객에 대한 애프터 서 비스를 제공

가) 비즈니스 모델 기반 문제 탐색

• 기업 내·외부 환경을 포괄하고 있는 비즈니스 모델이라는 틀 (Frame)을 활용하여 비즈니스모델 캔버스의 9가지 블록을 단순 화하여 업무(Operation), 제품(Product), 고객(Customer) 단위로 문제를 발굴하고, 이를 관리하는 두 가지의 영역인 규제와 감사 (Regulation&Audit)영역과 지원 인프라(IT&Human Resources)영역에 대한 기회를 추가로 도출하는 작업을 수행한다.



과제발굴방법	내용	예
· 업무 (Operation)	제품 및 서비스를 생산하기 위 해서 운영하는 내부 프로세스 및 주요 자원(Resource) 관 련 주제 도출	- 생산 공정 최적화 - 재고량 최소화
제품 (Product)	생산 및 제공하는 제품·서비 스를 개선하기 위한 관련 주 제 도출	- 제품의 주요기능 개선 - 서비스 모니터링 지표도출

고객 (Customer)	제품·서비스를 제공받는 사용 자 및 고객, 이를 제공하는 채 널의 관점에서 관련 주제 도출	- 고객 Call 대기 시간 최소화 - 영업점 위치 최적화
규제와 감사 (Regualtion& Audit)	제품 생산 및 전달과정 프로 세스 중에서 발생하는 규제 및 보안의 관점에서 주제 도출	- 제공 서비스 품질의 이상 징후 관리 - 새로운 환경 규제 시 예상 되는 제품 추출 등
지원 인프라 (IT&Human Resources)	분석을 수행하는 시스템 영역 및 이를 운영·관리하는 인력 의 관점에서 주제 도출	- EDW 최적화 - 적정 운영 인력 도출 등

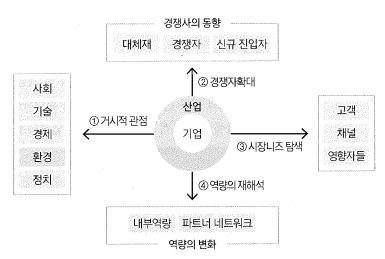
• 현재 사업을 영위하고 있는 환경, 경쟁자, 보유하고 있는 역량, 제공하고 있는 시장을 넘어서 거시적 관점의 요인, 경쟁자의 동향, 시장의 니즈 변화, 역량의 재해석 등 새로운 관점의 접근을 통해 새로운 유형의 분석 기회 및 주제 발굴을 수행해야 한다.



분석 기회 발굴의 범위 확장시 내용에서 4가지 관점과 내용이 시험에 자주 출제 됩니다.



나) 분석 기회 발굴의 범위 확장



① 거시적 관점의 메가트랜드

• 조직 및 해당 산업에 폭넓게 영향을 미치는 사회·경제적 요인을 STEEP으로 요약되는 Social(사회), Technological(기술), Economic(경제), Environmental(환경), Political(정치) 영역으로 폭넓게 나눈다.

8	내용	O
Social (사회)	비즈니스 모델의 고객(Custo -mer) 영역에 존재하는 현재 고 객을 확장하여 전체 시장을 대상 으로 사회적, 문화적, 구조적 트 렌드 변화에 기반한 분석 기회를 도출	- 노령화 - 밀레니엄 세대 등장 - 저출산에 따른 사업모델 변화
Technological (기술)	과학, 기술, 의학 등 최신 기술의 등장 및 변화에 따른 역량 내 재 화와 제품·서비스 개발에 대한 분석 기회를 도출	- 나노 기술 - IT 융합 기술 - 로봇 기술의 고도화로 인한 제품의 Smart화
Economic (경제)	산업과 금융 전반의 변동성 및 경제 구조 변화 동향에 따른 시 장의 흐름을 파악하고, 이에 대한 분석 기회를 도출	원자재 가격, 환율, 금리 변동에 따른 구매전략의 변화 등
Environmental (환경)	환경과 관련된 정부, 사회단체, 시민사회의 관심과 규제 동향을 파악하고 이에 대한 분석 기회를 도출	탄소 배출 규제 및 거래 시장 등장에 따른 원가 절감 및 정보 가시화 등
Political (정치)	주요 정책방향, 정세, 지정학적 동향 등의 거시적인 흐름을 토대 로 한 분석 기회를 도출	대북관계 동향에 따른 원자재 구매 거래선의 다변화 등

② 경쟁자 확대 관점

• 현재 수행하고 있는 사업 영역의 직접 경쟁사 및 제품·서비스뿐만 아니라 대체재와 신규 진입자 등으로 관점을 확대하여 위협이 될 수 있는 상황에 대한 분석 기회 발굴의 폭을 넓혀서 탐색한다.

영역	내용	0
대체재 (Substitute)	융합적인 경쟁 환경에서 현재 생 산을 수행하고 있는 제품·서비스 를 온라인으로 제공하는 것에 대 한 탐색 및 잠재적 위험을 파악	오프라인 제공 서비스 → 온라인제공에 대한 탐색 및 잠재적 위협 파악
경쟁자 (Competitor)	현재 생산하고 있는 제품·서비 스의 주요 경쟁자에 대한 동향 을 파악하여 이를 고려한 분석 기회를 도출	식별된 주요 경쟁사의 제품·서비스 카탈로그 및 전략 분석을 통한 잠 재적 위협 파악
신규 진입자 (New Entrant)	향후 시장에 대해서 파괴적인 역 할을 수행할 수 있는 신규 진입 자에 대한 동향을 파악하여 이를 고려한 분석 기회를 도출	새 제품에 대한 클라우 드 소싱 서비스인 킥 스 타터의 유사 제품을 분 석하고 잠재적위협 파악

③ 시장의 니즈 탐색 관점

• 현재 수행하고 있는 사업에서의 직접 고객뿐만 아니라 고객과 접 촉하는 역할을 수행하는 채널(Channel) 및 고객의 구매와 의사 결정에 영향을 미치는 영향자들

명 명	내용	예
고객 (Customer)	고객의 구매 동향 및 고객의 컨텍스트 를 더욱 깊게 이해하여 제품·서비스 의 개선에 필요한 분석 기회를 도출	철강 기업의 경우 조선 산 업과 자동차 산업의 동향 및 주요 거래선의 경영 현 황 등을 파악하고 분석 기회 도출
채널 (Channel)	영업사원, 직판 대리점, 홈페이지 등의 자체적으로 운영하는 채널뿐만 아니라 최종 고객에게 상품·서비스를 전달하는 것에 가능한 경로를 파악 하여 해당 경로에 존재하는 채널별로 분석 기회를 확대하여 탐색	은행의 경우 인터넷 전문 은행 등 온라인 채널의 등장에 따른 변화에 대한 전략 분석 기회 도출
영향자 (Influencer)	기업 의사결정에 영향을 미치는 주 주·투자자·협회 및 기타 이해 관계 자의 주요 관심사항에 대해서 파악 하고 분석기회를 탐색	M&A 시장 확대에 따른 유사, 업종의 신규기업 인수기회 탐색

④ 역량의 재해석 관점

• 현재 해당 조직 및 기업이 보유한 역량뿐만 아니라 해당 조직의 비즈 니스에 영향을 끼치는 파트너 네트워크를 포함한 활용 가능한 역량 을 토대로 폭넓은 분석 기회를 탐색한다.

영역	내용	예
내부 역량 (Competency)	지적 재산권, 기술력 등 기본적인 것 뿐만 아니라 중요하면서도 자 첫 간과하기 쉬운 지식, 기술 등의 노하우와 인프라적인 유형 자산 에 대해서 재해석하고 해당영역 에서 분석 기회를 탐색	자사 소유 부동산을 활 용한부가 가치 창출 기회 발굴
파트너와 네트워 크 (Partners& Network)	자사가 직접 보유하고 있지는 않 지만 밀접한 관계를 유지하고 있 는 관계사와 공급사 등의 역량을 활용해 수행할 수 있는 기능을 파 악해보고 이에 대한 분석 기회를 추가적으로 도출	수출입·통관·노하우를 활용한 추가 사업기회 탐색

다) 외부참조 모델기반 문제탐색

• 유사·동종 사례 벤치마킹을 통한 분석기회 발굴은 제공되는 산업별, 업무 서비스별 분석 테마 후보 그룹 (Pool)을 통해 "Quick&Easy" 방식으로 필요한 분석기회가 무엇인지에 대한 아이디어를 얻고, 기업에 적용할 분석테마 후보 목록을 워크숍 형태의 브레인스토밍(Brain storming)을 통해 빠르게 도출하는 방법이다.

교통	안전 행정 보건	복지 교통 도시	병역 일자리 정책	정치 환경 기타
	교통사고 감소를 위한 빅데이터 예보 서비스	빅데이터를 활용한 보다 안전한 도로관리	국민 참여형 어린이 안전 및 교통사고 원인 분석	사고 행동을 분석하여 적극적인 예방을 계획
A Hijo	빅데이터로 전기차 충전 인프라 설치 입지 선정	서울시 심야버스 노선 확 정을 위한 빅데이터 분석	지능형 교통안내 시스템 위한 센서데이터 분석	포트홀 사고 방지를 위한 스마트폰의 GPS 정보 분 석
	대중교통 서비스개선을 위한 택시 및 전차시간 실시간 분석	서울시민의 대중교통 이용 실태분석을 위한 KCB 융합데이터 분석	빅데이터로 위험한 도로를 피하자	빅데이터를 이용한 교통 및 범죄정보 관리
안	골든 타임 확보로 응급환자 생존률 향상	빅데이터 기반 지능형 전기 화재 예방플랫폼 구축	빅데이터 기반 지능형 도시가스 배관 위험 예측	빅데이터로 산불 피해 최소화
졘	동물 및 인간 감영병 확 산 대응 지원 체계 구축	질병도 이젠 빅데이터로 예측하고 예방	상수도 누수지역 탐지	
<u>쥲</u> 0	빅데이터로 행정 부서간 갈등전 예방	공동주택 부조리 분석 시스템 개발	체납액 회수도 과학적인 방법으로	근로감독 사업장 선정 과학화
정	실업급여 부정수급 악용 빅데이터로 뿌리 뽑기	빅데이터를 통한 고용취약자 파악	사회 취약계층 선제적 발견	

• 현재 환경에서는 데이터를 활용하지 않은 업종 및 업무 서비스가 사실상 존재하지 않기 때문에 데이터 분석을 통한 인사이트(Insight)를 도출하고 업무에 활용하는 사례들을 발굴하고, 자사의 업종 및 업무 서비스에 적용하며 평상시 지속적인 조사와 데이터 분석을 통한 가치 발굴 사례를 정리하여 풀(Pool)로 만들어 둔다면 과제 발굴 및 탐색시 빠르고 의미있는 분석 기회도출이 가능하다.

금융	농축수산	문화관광 에너지	유통	의료 제조	IT 기타
의료	인공지능 기반 약효 예측 모델 개발	국민건강주의예보 서비스 제공을 위한 빅데이터 분석	OH OH	빅데이터를 활용한 종합식품업체의 성장	프로세스 체질개선을 위한 빅데이터 활용
농축 수산	농산물 수급조절의 효율성 제고	참치·넙치 스마트양식하는 청색형명 온다.	문화 관광	빅데이터와 함께 안전한 여행	연예인 마케팅에도 필요한 빅데이터

보 현	고객 위치에 따른 보험 정보 제공을 위한 고객 위치 분석	보험사기 방지를 위한 보험사고 데이터 분석	자동차 보험료 선정 및 손실 률 축소를 위한 자동차 센서 정보 (운행기록장치) 분석	콜센터직원 배분을 위한 고객 데이터 분석
여 마	고객의 금융 습관 개선을 위한 고객 행동패턴 분석	금융사기 관리를 위한 빅데이터 분석	기업 이미지 관리를 위한 평판 분석	빅데이터에 의한 경기지표 산출을 통한 경기 현황 분석
iii	신규 금융 서비스 기회 발전을 위한 고객정보 분석	신용 위험도 파악을 위한 고객생활패턴 분석	신용평가 모델 수립을 위한 대출 신청자 행동패턴 분석	최적의 투자 상품 추천 을 위한 고객 경험 분석

라) 분석 유즈 케이스(Analytics Use Case)

• 현재의 비즈니스 모델 및 유사·동종사례 탐색을 통해서 빠짐없이 도출한 분석 기회들을 구체적인 과제로 만들기 전에 분석 유즈 케이스로 표기하는 것이 필요하다. 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.

엄마	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
717 (F)	서비스 수준 유지	서비스별로 달성 수준을 측정하고 평가한 뒤 목표 수준을 벗어나면 경보발생	• 품질수준 재고 • 고객만족 재고
고객 -	고객 만족 달성	고객 세그먼트별로 만족 수준을 측정하고 이상이 있으면 원인을 분석하여 대책 강구	• 고객만족 재고 • 고객유지 향상
20 30 40 7 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	파이프라인 최적화	파이프라인 단계별로 고객 상태를 파악하고 수주 규모를 예상하고 필요한 고객 기회를 추정하여 영업 촉진	목표 매출 달성고객반응률 향상
<u>판매</u> -	영업성과 분석	영업 직원별 사용 원가(급여 포함)와 실적을 분석하고 부 진한 영업 직원 세그먼트를 식별하여 영업 정책에 반영	• 영업 수율 향상 • 영업 직원 생산성 제고



문제 정의 단계의 목적과 표현 방식이 출제됩니다.

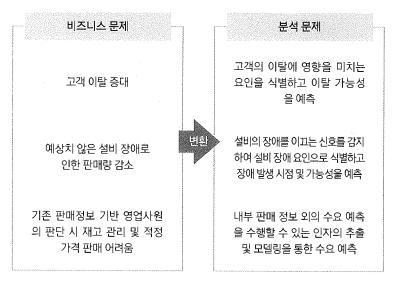


侧周

'고객 이탈의 증대'라는 비즈 니스 문제는 '고객 이탈에 영 향을 미치는 요인을 식별하고 이탈 가능성을 예측'하는 데 이터 분석 문제로 변환 될 수 있다.

2) 문제 정의(Problem Definition) 단계 - 하향식 접근법 2단계

- 식별된 비즈니스 문제를 데이터의 문제로 변환하여 정의 하는 단계이며, 앞서 수행한 문제 탐색의 단계가 무엇을(What) 어떤 목적으로(Why) 수 행해야 하는지에 대한 관점이었다면, 본 단계에서는 이를 달성하기 위해 서 필요한 데이터 및 기법(How)을 정의하기 위한 데이터 분석의 문제로 의 변환을 수행하게 된다.
- 데이터 분석 문제의 정의 및 요구사항 : 분석을 수행하는 당사자뿐만 아니라 해당 문제가 해결되었을 때 효용을 얻을 수 있는 최종사용자(End User) 관점에서 이루어져야 한다.
- 데이터 분석 문제가 잘 정의되었을 때 필요한 데이터의 정의 및 기법 발굴이 용이하기 때문에 가능한 정확하게 분석의 관점으로 문제를 재정의할 필요가 있다.



〈 비즈니스 문제를 분석 문제로 변환 〉

3) 해결방안 탐색(Solution Search) 단계 - 하향식 접근법 3단계

• 이 단계에서는 정의된 **데이터 분석 문제를 해결**하기 위한 다양한 방안 이 모색된다.

- 가) 기존 정보시스템의 단순한 보완으로 분석이 가능한지 고려
- 나) 엑셀 등의 간단한 도구로 분석이 가능한지 고려
- 다) 하둡 등 분산병렬처리를 활용한 빅데이터 분석 도구를 통해 보다 체계적이고 심도있는 방안 고려

		분석역량 (Who)		
		확보	미확보	
분석 기법 및	기존 시스템	기존 시스템 개선 활용	교육 및 채용을 통한 역량확보	
시스템 (How)	신규 도입	시스템 고도화	전문업체 Sourcing	

〈해결 방안 탐색 영역〉

• 분석역량을 기존에 가지고 있는 지의 여부를 파악하여 보유하고 있지 않은 경우에는 교육이나 전문인력 채용을 통한 역량을 확보하거나 분석 전문업체를 활용하여 과제를 해결하는 방안에 대해 사전 검토를 수행한다.

4) 타당성 검토(Feasibility Study) - 하향식 접근법 4단계

• 도출된 분석 문제나 가설에 대한 대안을 과제화하기 위해서는 다음과 같은 다각적인 타당성분석이 수행되어야 한다.



타당성 검토 단계에서 검토 대상 3개(경제성, 더 이터, 기술)와 관련 내용이 출제 됩니다.



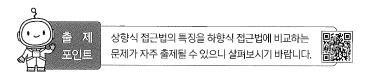
가) 경제적 타당성

• 비용대비 편익 분석 관점의 접근이 필요하다. 비용 항목은 데이터, 시스템, 인력, 유지보수 등과 같은 분석 비용으로 구성되고, 편익으 로는 분석 결과를 적용함으로써 추정되는 실질적 비용 절감, 추가 적 매출과 수익 등과 같은 경제적 가치로 산출된다.

나) 데이터 및 기술적 타당성

 데이터 분석에는 데이터 존재 여부, 분석 시스템 환경 그리고 분석 역량이 필요하다. 특히, 분석 역량의 경우 실제 프로젝트 수행시 걸 림돌이 되는 경우가 많기 때문에 기술적 타당성 분석시 역량 확보 방안을 사전에 수립하고 이를 효과적으로 평가하기 위해서는 비즈 니스 지식과 기술적 지식이 요구된다.

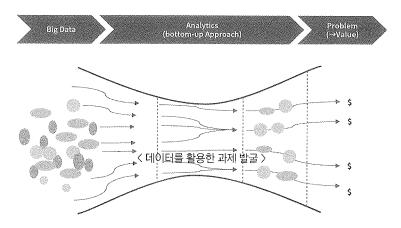
- 위의 타당성 검토를 통해 도출된 대안을 통해
 - ① 평가 과정을 거쳐 가장 우월한 대안을 선택한다.
 - ② 도출한 데이터 분석 문제 및 선정된 솔루션 방안을 포함한다.
 - ③ 분석과제 정의서의 형태로 명시하는 후속작업을 시행한다.
 - ④ 프로젝트 계획의 입력물로 활용한다.



다. 상향식 접근법(Bottom-Up Approach)

1) 정의

- 한 의약제조사는 특허기간이 만료된 의약품 약 2천 종류의 데이터를 분석, 상호 결합하여 새로운 의약품을 개발하려고 시도했다.
- 의약품 집합으로부터 두 개의 조합을 선택할 수 있는 방법은 백만 개 이 상이기 때문에 이 회사는 새로운 결합의 효과성을 검증하기 위해 다양 한 기법을 적용하여 데이터를 분석했다.
- 여기에서는 전통적인 하향식 문제 해결방식과 대비하여 기업에서 보유하고 있는 다양한 원천 데이터로부터의 분석을 통하여 통찰력과 지식을 얻는 상향식 접근방법을 기술한다. 상향식 접근방법은 아래 그림처럼 다양한 원천 데이터를 대상으로 분석을 수행하여 가치있는 모든 문제를 도출하는 일련의 과정이다.

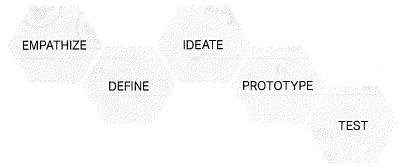


2) 기존 하향식 접근법의 한계를 극복하기 위한 분석 방법론

• 기존 접근방법인 논리적인 단계별 접근법은 문제의 구조가 분명하고 문제를 해결하고 해결책을 도출하기 위한 데이터 분석가 및 의사결정자에게

주어져 있음을 가정하고 있기 때문에 솔루션 도출에는 유효하지만 새로 우 문제의 탐색에는 한계가 있다.

- 따라서 기존의 논리적인 단계별 접근법 기반의 문제해결 방식은 최근 복 잡하고 다양한 환경에서 발생하는 문제에는 적합하지 않을 수 있다.
- 이를 해결하기 위해서 스탠포드 대학의 d. school(Institute of Design at Stanford)은 디자인 사고(Design Thinking) 접근법을 통해서 전통적 인 분석적 사고를 극복하려고 한다.
- 통상적인 관점에서는 분석적으로 사물을 인식하려는 'Why'를 강조하지 만, 이는 우리가 알고있다고 가정하는 것이기 때문에 문제와 맞지 않는 솔루션인 경우 오류가 발생할 소지가 있다. 그렇기 때문에, 답을 미리 내는 것이 아니라 사물을 있는 그대로 인식하는 'What' 관점에서 보아야한다는 것이다.
- 객관적으로 존재하는 데이터 그 자체를 관찰하고 실제적으로 행동에 옮김으로써 대상을 좀 더 잘 이해하는 방식으로의 접근을 수행하는 것이다.



〈디자인 사고(Design Thinking)의 프로세스 〉

• 이와 같은 점을 고려하여 d. school에서는 첫 단계로 감정이입(Empathize)을 특히 강조하고 있다.

3) 비지도 학습과 지도 학습

- 가) 비지도 학습(Unsupervised Learning)
 - 일반적으로 **상향식 접근방식의 데이터 분석**은 비지도 학습(Unsupervised Learning) 방법에 의해 수행된다.
 - 비지도 학습은 데이터 분석의 목적이 명확히 정의된 형태의 특정 필 드의 값을 구하는 것이 아니라 데이터 자체의 결합, 연관성, 유사성 등을 중심으로 데이터의 상태를 표현하는 것이다.

• 비지도 학습의 데이터 마이닝 기법의 예 - 장바구니 분석, 군집 분석, 기술 통계 및 프로 파일링 등

나) 지도 학습(Supervised Learning)

이 목적이다.

• 명확한 목적 하에 데이터분석을 실시하는 것은

예세

- 지도학습과 비지도학습의 비교 그림에서 o와 x를 구분 짓게 하는 분류 (Classification)는 지도학습에 해당되고, 인자들 간의 유사성을 바탕으로 수행하는 군집화(Clustering)는 비지도 학습에 해당한다.
 - → 지도학습의 경우 결과로 도**출되는 값에 대하여 사전에 인지**하고 어떠한 데이터를 넣었을 때 어떠한 결과가 나올지를 예측하는 것이다.
 - →비지도학습의 경우 목표 값을 사전에 정의하지 않고 데이터 자체만을 가지고 그 룹들을 도출함으로써 해석이 용이하지는 않지만 새로운 유형의 인사이트를 도출하기에 유용한 방식으로 활용할 수 있다.

지도학습(Supervised Learning)이라고 하며, 분류, 추측, 예측, 최적화를 통해 사용자의 주도하에 분석을 실시하고 지식을 도출하는 것

• 통계적 분석에서는 인과관계 분석을 위해 가설을 설정하고 이를 검정하기 위해 모집단으로 부터 표본을 추출하고 그 표본을 이용한 가설 검정을 실시하는 방식으로 문제를 해결하였다. 그러나 빅데이터 환경에서는 이와 같은 논리적인 인과관계 분석뿐만 아니라 상관관계 분석 또는 연관 분석을 통하여 다양한 문제 해결에 도움을 받을 수 있다. 인과관계로부터 상관관계 분석으로의 이동이 빅데이터 분석에서의 주요 변화라고 할 수 있다. 다량의 데이터 분석을 통해서 "왜" 그러한 일이 발생하는지 역으로 추적하면서 문제를 도출하거나 재정의 할 수 있는 것이 상향식 접근법이다.

출 제 포인트

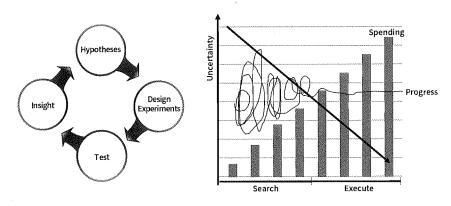
프로토타이핑 접근법의 방법론과 특징은 자주 출 제되므로 꼭 기억해 주세요



4) 시행착오를 통한 문제 해결

가) 정의

- 프로토타이핑 접근법은 사용자가 요구사항이나 데이터를 정확히 규정하기 어렵고 데이터 소스도 명확히 파악하기 어려운 상황에서 일단 분석을 시도해 보고 그 결과를 확인해 가면서 반복적으로 개선해 나가는 방법을 말한다.
- 하향식 접근방식은 문제가 정형화되어 있고 문제해결을 위한 데이터가 완벽하게 조직에 존재할 경우에 효과적이다.
- 이에 반하여 프로토타이핑 방법론은 비록 완전하지는 못하다 해도 신속하게 해결책이나 모형을 제시함으로써, 이를 바탕으로 문제를 좀 더 명확하게 인식하고 필요한 데이터를 식별하여 구체화할 수 있 게 하는 유용한 상향식 접근 방식이다.
- 프로토타이핑 접근법의 기본적인 프로세스는 가설의 생성, 디자인에 대한 실험, 실제 환경에서의 테스트, 테스트 결과에서의 통찰도출 및 가설 확인으로 구성된다.



나) 빅데이터 분석 환경에서 프로토타이핑의 필요성

- ① 문제에 대한 인식 수준
 - 문제 정의가 불명확하거나 이전에 접해보지 못한 새로운 문제일 경우 사용자 및 이해관계자는 프로토타입을 이용하여 문제를 이 해하고, 이를 바탕으로 구체화하는데 도움을 받을 수 있다.

이 가능하다.

이동통신사에서 수집하는 사 용자의 위치 데이터는 사용자

의 호출을 효율적으로 처리하 기 위한 원래의 목적 뿐만 아

니라, 사용자들이 특정시간에

많이 모이는 장소가 어디인가 를 분석하는 정보로도 활용

② 필요 데이터 존재 여부의 불확실성

• 문제해결을 위해 필요한 데이터의 집합이 모두 존재하지 않을 경우, 그 데이터의 수집을 어떻게 할 것인지 또는 그 데이터를 다른 데이터로 대체할 것인지 등에 대한 사용자와 분석가 간의 반복적이고 순환적인 협의 과정이 필요하다. 대체 불가능한 데이터가 존재하는지 사전에 확인한다면 불가능한 프로젝트를 수행하는 리스크를 사전에 방지할 수 있다.

③ 데이터 사용 목적의 가변성

• 데이터의 가치는 사전에 정해진 수집목적에 따라 확정되는 것이 아니고, 그 가치가 지속적으로 변화할 수 있다. 따라서 조직에서 보유 중인 데이터라 하더라도 기존의 데이터 정의를 재검토하여 데이터의 사용 목적과 범위를 확대할 수 있을 것이다.

라. 분석과제 정의

• 분석과제 정의서를 통해 분석별로 필요한 소스 데이터, 분석방법, 데이터 입수 및 분석의 난이도, 분석 수행주기, 분석결과에 대한 검증 오너십, 상세 분석 과정 등을 정의한다. 분석 데이터 소스는 내·외부의 비구조적인 데이터와 소셜 미디어 및 오픈 데이터까지 범위를 확장하여 고려하고 분석방법 또한 상세하게 정의한다.

(明月)

분석과제 정의서

분/	석명	분석정의			
해지상담접	촉패턴 분석	기 해지 계약건 발생 고객의 해지 시점 상담정보 분석을 통해 해지 고객의 상 담특성을 발굴하는 분석			
소스데이터	데이터 입수 난이도	분석방법			
접촉채널, 건수,	서비스 수준 유지	해지로 이어지는 해지 상담의 유의미 속성을 요인분석을 통해 발굴하고 클			
접촉평균 시간 최종 접촉 이후 해지까지 시간	데이터 입수사유	스터링 분석을 통해 영향요인을 그룹 하고, 그룹핑된 요인그룹이 해지에 치는 영향도를 회귀분석함			
상담인력 업무 능숙도	N/A				
분석적용난이도	분석적용 난이 사유	분석주기	분석결과 검증 Owner		
중	접촉 로그 등의 비구조적 데이터 분석 필요	월별 업데이터	해지방어팀		
		j			





1장 데이터 분석 기획의 이해

분석 프로젝트 관리 방안



분석 프로젝트 관리를 위한 5가지 주요 속성법의 방법론과 특징은 자주 출제되므로 꼭 기억하세요.



• 과제 형태로 도출된 분석기회는 프로젝트를 통해서 그 가치를 증명하고 목표를 달성해야한다. 분석프로젝트는 다른 프로젝트 유형처럼 범위, 일정, 품질, 리스크, 의사소통 등 영역별 관리가 수행되어야 할 뿐 아니라 다양한 데이터에 기반한 분석기법을 적용하는 특성 때문에 5가지의 주요 속성을 고려한 추가적인 관리가 필요하다.

분석과제 관리를 위한 5가지 주요 영역

Data Complexity

Data Size

분석과제관리 5 Analytic Dimensions

Accuracy & Precision

Analytic Complexity

Speed

〈 분석과제 관리를 위한 5가지 영역 〉

1장 데이터 분석 기획의 이해

8 명	내용
Data Size	분석하고자 하는 데이터의 양을 고려한 관리 방안 수립이 필요하다. 하둡 환경에서의 엄청난 데이터양을 기반으로 분석하는 것과 기존 정형 데이터베이스에 있는 시간 당 생성되는 데이터를 분석할 때의 관리방식은 차이가 날 수밖에 없다.
Data Complexity	BI(Business Intelligence) 프로젝트처럼 정형 데이터가 분석 마트로 구성 되어 있는 상태에서 분석을 하는 것과 달리 텍스트, 오디오, 비디오 등의 비정 형 데이터 및 다양한 시스템에 산재되어 있는 원천 데이터들을 통합해서 분석 프로젝트를 진행 할 때는, 초기 데이터의 확보와 통합뿐 아니라 해당 데이터에 잘 적용될 수 있는 분석 모델의 선정 등에 대한 사전 고려가 필요하다.
Speed	분석결과가 도출되었을 때 이를 활용하는 시나리오 측면에서의 속도를 고려해야 한다. 일 단위, 주 단위 실적의 경우에는 배치(Batch)형태로 작업되어도 무방하지만 실시간으로 사기(Fraud)를 탐지하거나 고객에게 개인화된 상품·서비스를 추천하는 경우에는 분석 모델의 적용 및 계산이 실시간으로 수행되어야하기 때문에 프로젝트 수행시 분석 모델의 성능 및 속도를 고려한 개발 및 테스트가 수행되어야한다.
Analytic Complexity	분석 모델의 정확도와 복잡도는 트레이드 오프(Trade off)관계가 존재한다. 분석 모델이 복잡할수록 정확도는 올라가지만 해석이 어려워지는 단점이 존 재하므로 이에 대한 기준점을 사전에 정의해 두어야 한다. 고객의 신용을 평 가하는 마케팅 시나리오에서 분석모델을 활용하여 신용점수가 낮게 나올 때 어떠한 변수에 기인했는지를 모델에서 설명해 줄 수 없으면 영업·마케팅 직원 입장에서는 해당 고객과의 소통이 어려워지는 단점이 존재하므로, 해석이 가 능하면서도 정확도를 올릴 수 있는 최적모델을 찾는 방안을 사전에 모색해야 한다.
Accuracy& Precision	Accuracy는 모델과 실제 값 사이의 차이가 적다는 정확도를 의미하고 Precision은 모델을 지속적으로 반복했을 때의 편차의 수준으로써 일관적 으로 동일한 결과를 제시한다는 것을 의미한다. 분석의 활용적인 측면에서 는 Accuracy가 중요하며, 안정성 측면에서는 Precision이 중요하다. 그러나 Accuracy와 Precision은 트레이드오프가 되는 경우가 많기 때문에 모델의 해석 및 적용 시 사전에 고려해야 한다.



Low accuracy Low precision



Low accuracy High <u>precision</u>



High <u>accuracy</u> Low precision



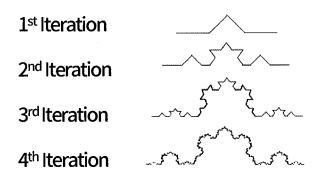
High <u>accuracy</u> High <u>precision</u>

〈Accuracy와 Precision의 관계〉

가. 개요

- 분석가의 목표: 분석의 정확도를 높이는 것이지만 프로젝트의 관점에서는 도출된 분석 과제를 잘 구현하여 원하는 결과를 얻고 사용자가 원활하게 활용할 수 있도록 전체적인 과정을 고려해야하기 때문에 개별적인 분석 업무 수행뿐만 아니라 전반적인 프로젝트 관리 또한 중요하다.
- 분석가의 입장: 데이터의 원천을 다루는 데이터 영역과 결과를 활용할 비즈니스 영역의 중간에서 분석 모델을 통한 조율을 수행하는 조정자의 역할 이 핵심이 된다. 특히 분석 프로젝트에서는 데이터 영역과 비즈니스 영역의 현황을 이해하고 프로젝트의 목표인 분석의 정확도 달성과 결과에 대한 가치 이해를 전달하는 조정자로서의 분석가의 역할이 중요하다. 조정자로서의 분석가가 해당 프로젝트의 관리자까지 겸임하게 되는 경우가 대부분이므로, 프로젝트 관리방안에 대한 이해와 주요 관리 포인트를 사전에 숙지하는 것이 필수적이다.
- 분석 프로젝트는 도출된 결과의 재해석을 통한 지속적인 반복 및 정교화가 수행되는 경우가 대부분이므로, 프로토타이핑 방식의 애자일(Agile) 프로젝트 관리방식에 대한 고려도 필요하다. 데이터 분석의 지속적인 반복 및 개선을 통하여 의도했던 결과에 더욱 가까워지는 형태로 프로젝트가 진행될 수 있도록 적절한 관리 방안 수립이 사전에 필요하다.

The Data Scientist as builder



〈 분석 프로젝트의 반복적 개선 특성 〉

• 분석 프로젝트는 데이터 영역과 비즈니스 영역에 대한 이해뿐만 아니라 지속적인 반복이 요구되는 분석 프로세스의 특성을 이해한 프로젝트 관리방안을 수립하는 것이 중요하다.

분석 프로젝트의 특성

2

• 분석 과제정의서를 기반으로 프로젝트를 시작하되 지속적인 개선 및 변경을 염두에 두고 기간 내에 가능한 최선의 결과를 도출할 수 있도록 프로젝트 구성원들과 협업하는 것이 분석 프로젝트의 특징이다.



분석 프로젝트의 관리방안

• 분석 프로젝트의 영역별 주요 관리 항목과 내용은 아래와 같다.

주제 그룹 범위 (Scope)	개념 및 관련 프로세스 • 분석 기획단계의 프로젝트 범위가 분석을 진행하면서 데이터의 형태와 양 또는 적용되는 모델의 알고리즘에 따라 범위가 빈번하게 변경됨 • 분석의 최종 결과물이 분석 보고서 형태인지 시스템인지에 따라서 투입되는 자원 및 범위 또한 크게 변경되므로 사전에 충분한 고려가 필요함
시간 (Time)	• 데이터 분석 프로젝트는 초기에 의도했던 결과(모델)가 나오기 쉽지 않기 때문에 지속적으로 반복되어 많은 시간이 소요될 수 있음 • 분석 결과에 대한 품질이 보장된다는 전제로 Time Boxing 기법으로 일정관리를 진행하는 것이 필요함
원가 (Cost)	 외부 데이터를 활용한 데이터 분석인 경우 고가의 비용이 소요될 수 있으므로 사전에 충분한 조사가 필요함 오픈 소스 도구(Tool) 외에 프로젝트 수행 시 의도했던 결과를 달성하기 위하 여 상용 버전의 도구(Tool)가 필요할 수 있음 ex) 가시화를 위한 BI 솔루션, 지리정보 표기를 위한 GIS 솔루션 등
품질 (Quality)	•분석 프로젝트를 수행한 결과에 대한 품질 목표를 사전에 수립하여 확정해야 함 •프로젝트 품질은 품질 통제(Quality Control)와 품질보증(Quality Assurance)으로 나누어수행되어야 함
통합 (Integration)	• 프로젝트 관리 프로세스들이 통합적으로 운영될 수 있도록 관리해야 함
조달 (Procurement)	• 프로젝트 목적성에 맞는 외부 소싱을 적절하게 운영할 필요가 있음 • PoC(Proof of Concept) 형태의 프로젝트는 인프라 구매가 아닌 클라우드
	등의 다양한 방안을 검토할 필요가 있음
자원 (Resource)	등의 다양한 방안을 검토할 필요가 있음 •고급 분석 및 빅데이터 아키텍쳐링을 수행할 수 있는 인력의 공급이 부족하므로 프로젝트 수행 전 전문가 확보에 대한 검토가 필요함
. —	•고급 분석 및 빅데이터 아키텍쳐링을 수행할 수 있는 인력의 공급이 부족하므
(Resource) 리스크	 고급 분석 및 빅데이터 아키텍쳐링을 수행할 수 있는 인력의 공급이 부족하므로 프로젝트 수행 전 전문가 확보에 대한 검토가 필요함 분석에 필요한 데이터 미확보로 분석 프로젝트 진행이 어려울 수 있으므로 관련 위험을 식별하고 대응방안을 사전에 수립해야함 데이터 및 분석 알고리즘의 한계로 품질 목표를 달성하기 어려울 수 있어 그에

〈 분석 프로젝트 영역별 주요 관리 항목 〉

• 분석가가 분석 프로젝트에서 프로젝트 관리자의 역할을 수행하는 경우가 대부 분이기 때문에 프로젝트 관리 영역에 대한 주요한 사항들을 체크포인트 형태 로 관리해서 발생할 수 있는 이슈와 리스크를 숙지하고 미연에 방지할 필요가 있다.

참고

프로젝트관리 지침의 프로젝트 관리 체계

통합(Integration), 이해관계자(Stakeholder), 범위(Scope), 자원(Resource), 시간 (Time), 원가(Cost), 리스크(Risk), 품질(Quality), 조달(Procurement), 의사소통 (Communication)의 10개의 주제 그룹으로 구성되어 있다. (KSA ISO 21500의 프로 젝트 관리 주제 그룹)

MEMO



분석 마스터 플랜에서는 분석과제들의 우선순위 기준, 우선순위 평가기준, 우선순위 설정 방법, 이행계획수립방법이 중요합니다.



- 데이터 분석을 위한 마스터 플랜 수립을 이해한다.
- 분석과제의 시급성과 난이도에 따른 분석과제 우선순위를 선정 할 수 있다.
- 데이터 분석 거버넌스 체계를 이해한다.
- 데이터 분석 조직구조와 교육내용을 이해한다.

0 法長り利ヨ

✔ 데이터분석을 위한 마스터 플랜 수립이 필요한 이유를 알고 있으신가요?

데이터 분석을 구현하기 위한 다양한 기준들을 통해 데이터 분석을 위한 로드맵을 수립합니다.

✓ 분석과제의 우선순위를 선정할 수 있으신가요?

비즈니스 관점에서 도출된 다양한 분석과제들을 기업에 적용시키기 위해서는 적용 우 선순위를 평가해야 합니다.

✔ 데이터 분석 거버넌스 체계가 필요한 이유를 알고 있으신가요?

분석 거버넌스 체계를 통하여 기업의 현 분석수준을 정확히 진단하고, 분석 조직 및 분석 전문인력 배치, 분석 관련 프로세스 및 분석 교육 등의 관점에서 정의할 수 있습니다.

데이터 분석 조직구조와 교육내용을 이해하고 있으신가요?

데이터 분석 조직은 기업의 경쟁력 확보를 위해 필요하며, 데이터 분석을 통해 의미있는 인사이트를 찾아 실행하는 역할을 수행할 수 있어야 합니다. 데이터 분석 교육을 통해 조직원의 분석 역량 향상을 도모할 수 있습니다.



1장 데이터 분석 기획의 이해

마스터 플랜 수립 프레임 워크



우리는 앞 장에서 기업의 여러 가지 환경들을 분석해 분석기회를 도출하고 분석과제를 정의했습니다. 제1절을 통해 정의된 업무별 분석과제들의 우선순위를 평가하는 방법과 그 기준을 알아봅시다. 우선순위 평가 후 단계적인 구현 로드맵과 전체적인 분석 마스터 플랜을 수립하는 방법에 대해 알아봅시다.



(1)

분석 마스터

프레임 워크

플랜 수립

가. 마스터 플랜 수립 개요

- 데이터 기반 구축을 위해서 분석 과제를 대상으로 전략적 중요도, 비즈니스 성과 및 ROI, 분석 과제의 실행 용이성 등 다양한 기준을 고려해 적용 우선순위를 설정한다.
- 업무내재화 적용 수준, 분석데이터 적용 수준, 기술적용 수준 등 분석 적용 범위 및 방식에 대해서 종합적으로 고려하여 데이터 분석 구현을 위한 로 드맵을 수립해야 한다.

우선순위 고려요	注
① 전략적 중요도	
② 비즈니스 성과/ROI	적용 우선 순위 설정
③ 실행 용이성	

적용범위 / 방식 고려	병요소
① 업무 내재화 적용 수준	01-41
② 분석 데이터 적용 수준	Analytics 구현 로드맵 수립
③ 기술 적용 수준	TH

절고

기업 및 공공기관에서는 시스템의 중장기 로드맵을 정의하기 위한 정보전략계획인 ISP를 수행한다.

※ ISP(Information Strategy Planning): 정보기술 또는 정보시스템을 전략적으로 활용하기 위하여 조직 내·외부 환경을 분석하여 기회나 문제점을 도출하고 사용자의 요구사항을 분석하여 시스템 구축 우선순위를 결정하는 등 중장기 마스터 플랜을 수립하는 절차이다.



※분석 마스터 플랜: 일반적인 ISP 방법론을 활용하되 데이터 분석 기획의 특성을 고려하여 수행하고 기업에서 필요한 데이터 분석 과제를 빠짐없이 도출한 후 과제의 우선순위를 결정하고 단기 및 중·장기로 나누어 계획을 수립한다.

2장 분석 마스터 플랜



분석과제들의 우선순위를 평가하는 기준들이 어떻게 되어 있는지, 분석과제들을 평가하는 평가기준이 어떻게 만들어 졌는지 흐름을 이해하고 그 평가기준에 대해 숙지합시다.



나. 수행 과제 도출 및 우선순위 평가

1) 우선순위 평가 방법 및 절차

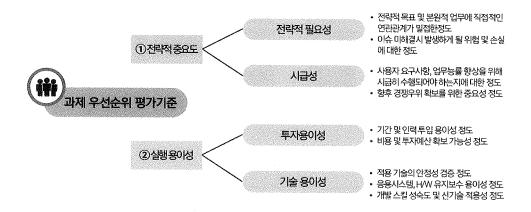
- 우선순위 평가의 경우 정의된 데이터 과제에 대한 실행 순서를 정하는 것이다.
- 업무별 도출된 분석과제를 우선순위 평가 기준에 따라 평가한 뒤, 과제 수행의 선·후행 관계를 고려하여 적용순위를 조정해 최종 확정한다.

분석 과제 도출 📦 우선순위 평가 🕸 우선순위 정련

• 과제우선순위 기준 수립 • 분석 과제 수행의 선후 관계 분석을 통해 순위조정

2) 일반적인 IT프로젝트의 우선순위 평가 예시

• 정보전략계획(ISP)과 같은 일반적인 IT프로젝트 과제의 우선순위 평가를 위해 전략적 중요도, 실행 용이성 등 기업에서 고려하는 중요 가치기준에 따라 다양한 관점에서의 우선순위 기준을 수립하여 평가한다.

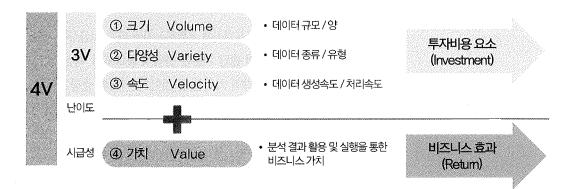




빅데이터의 특징이 분석 ROI 요소로 어떻게 매칭이 되는지 알아야 합니다. 3V는얼마나 많은 데이터를 가지고 어떠한 데이터를 구성해서 얼마나 빨리 처리하여결과를 도출하는지에 초점을 둔 것이기 때문에 투자비용 요소에 해당됩니다. 4V는 가치를 창출 한다는 의미를 가지고 있으므로 비즈니스 효과에 해당합니다.



3) ROI 관점에서 빅데이터의 핵심 특징



가) 투자비용(Investment)요소

- ① 크기(Volume)
 - 데이터의 규모 및 양을 의미, 대용량 데이터를 저장 처리하고 관리하기 위해서는 새로운 투자가 필요하다.
- ② 다양성(Variety)
 - 다양한 종류와 형태를 가진 데이터를 입수하는데에 있어 투자가 필요하다.
- ③ 속도(Velocity)
 - 데이터 생성 속도 및 처리속도를 빠르게 가공·분석하는 기술이 요구되다.

나) 비즈니스 효과(Return)요소

- (4) 가치(Value)
 - 분석 결과를 활용하거나 실질적인 실행을 통해 얻게 되는 비즈니스 효과 측면의 요소로, 기업데이터 분석을 통해 추구하거나 달성하고자 하는 목표 가치를 의미한다.



난이도는 해당 기업의 현 상황에 따라 조율할 수 있습니다. 뒤의 제3절 '분석 거버넌스 체계'에서 제시하는 분석 준비도 및 성숙도 진단 결과에 따라 해당 기업의 분석수준을 파악하고, 이를 토대로 분석 적용 범위 및 방법에 따라 난이도를 조정할 수 있습니다.



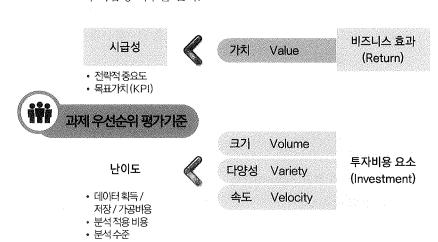
4) 데이터 분석과제 추진시 고려해야 하는 우선순위 평가 기준

가) 시급성

• 전략적 중요도와 목표가치에 부합하는지에 따른 시급성이 가장 중요한 기준이다. 시급성의 판단 기준은 전략적 중요도가 핵심이며, 이는 현 재의 관점에서 전략적 가치를 둘 것인지, 미래의 중장기적 관점에 전 략적인 가치를 둘 것인지를 고려하고, 분석 과제의 목표가치(KPI)를 함께 고려하여 시급성 여부를 판단할 수 있다.

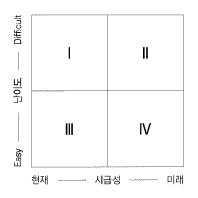
나) 난이도

데이터를 생성, 저장, 가공, 분석하는 비용과 현재 기업의 분석 수준을 고려한 난이도 역시 중요한 기준이다. 난이도는 현 시점에서 과제를 추진하는 것이 적용 비용 측면과 범위 측면에서 바로 적용하기 쉬운 것인지 또는 어려운 것인지에 대한 판단기준으로서, 데이터 분석의 적합성 여부를 본다.



5) 포트폴리오 사분면 분석을 통한 과제 우선순위 선정

• 우선순위 선정 기준을 토대로 난이도 또는 시급성을 고려하여 분석 과 제를 4가지 유형으로 구분하여 분석 과제의 적용 우선순위를 결정한다.





- 전략적 중요도가 높아 경영에 미치는 영향이 크므로 현재 시급하게 추진이 필요함
- 난이도가 높아 현재 수준에서 과제를 바로 적용하기에 어려움



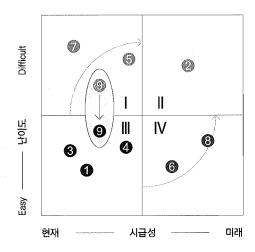
- 현재 시점에서는 전략적 중 요도가 높지 않지만 중장기 적 관점에서는 반드시 추진 되어야 함
- 분석과제를 바로 적용하기 에는 난이도가 높음



- 전략적 중요도가 높아 현재 시점에 전략적 가치를 두고 있음
- 과제 추진의 난이도가 어렵 지 않아 우선적으로 바로 적 용 가능할 필요성이 있음



- 전략적 중요도가 높지 않아 중장기적 관점에서 과제 추 진이 바람직함
- 과제를 바로 적용하는 것은 어렵지 않음



- 사분면 영역에서 가장 우선적인 분석 과제 적용이 필요한 영역은 3사분 면이다.
- 전략적 중요도가 현재 시점에는 상대적으로 낮은 편이지만 중장기적으로는 경영에 미치는 영향도가 높고, 분석 과제를 바로 적용하기 어려워 우선순위가 낮은 영역은 2사분면이다.
- 분석과제의 적용 우선순위 기준을 '시급성'에 둔다면 Ⅲ→Ⅳ→Ⅱ 영역 순이며, 우선순위 기준을 '난이도'에 둔다면 Ⅱ→Ⅰ→Ⅱ 영역 순으로 의사결정을 할 수 있다.

• ⑨번 과제와 같이 1사분면에 위치한 분석과제는 데이터 양, 데이터 특성, 분석범위 등에 따라 난이도를 조율함으로써 적용 우선순위를 조정할 수 있다. 예를 들어 분석에 필요한 데이터양이 수 TB 규모라면, 분석 대상 이 되는 소스 데이터를 내부 데이터 관점에서 우선 분석할 수 있도록 데 이터의 양을 줄여 난이도를 낮출 수 있다. 이를 통해 궁극적으로는 1사분 면(Ⅰ영역)에서 3사분면(Ⅱ영역)으로 분석 적용의 우선순위를 조정하여 추진할 수 있다.



단계적 구현 로드맵은 3가지의 추진단계로 구성되어 있으며 각 단계마다 어떤 목표와 추진과제를 가지고 있는지 숙지해야 합니다. 보기 중 잘못 연결되어 있는 것을 체크할 수 있어야 합니다.



다. 이행계획 수립

1) 로드맵 수립

- 가) 분석 과제에 대한 포트폴리오 사분면(Quadrant) 분석을 통해 과제의 1차 적 우선순위를 결정한다.
- 나) 분석 과제별 적용범위 및 방식을 고려하여 최종적인 실행 우선순위를 결정한 후 단계적 구현 로드맵을 수립한다.
- 다) 단계별로 추진하고자 하는 목표를 정의한다.
- 라) 추진 과제별 선·후행 관계를 고려하여 단계별 추진 내용을 정렬한다.

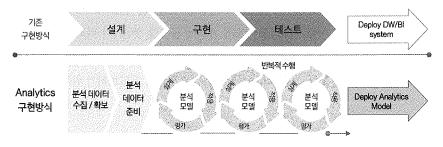
	Stage 1	Stage 2	Stage 3		
추진단계	데이터 분석체계 도입	데이터분석유효성검증	데이터분석확산및고도화		
단계별 추진목표	박데이터의 성공적인 도입을 위해, 비즈니스 Pain Point가 무엇인지 식별하고 이를 해결 해나가는 관점에서 분석기회 를 발굴하여 분석 과제로 정 의하고 마스터 플랜을 수립함	분석과제에 대한 Piot을 수행하여 비즈니스적인 유효성과 타당성을 검증하고 기술적인 실현가능성을 검증함 Piot 수행에 필요한 분석 일고 리즘 및 아키텍처를 설계함	Plot을 통해 검증된 분석과제를 업무 프로세스에 내재화하기 위한 PD와 변화관리실시 Plot 검증결과를 소사에 확산하는 관점에서 박데이터 분석-활용시스템을 구축하고 유관시스템을 고도화함		
추진 과제	추진과제 0 분석기회발굴및 마스터플랜 분석과제정의 수립	추진과제 1 분석알고리즘 및 분석과제 아키텍처설계 Plot수행	추진과제 2 업무프로세스 내재화를 변화 위한 Process Innovation 관리 추진과제 3 백데이터 분석 - 유관시스템 활용시스템구축 고도화		



폭포수 모델(Water-Fall): 순차적인 소프트웨어 개발 프로세스로 개발의 흐름이 마치 폭포수처럼 지속적으로 아래로 향하는 것처럼 보인다는 것에서 이름이 붙여졌다. 예를 들어 소프트웨어 요구사항 기술, 소프트웨어 설계, 소프트웨어 구현, 통합 시험과 디버 깅 설치, 소프트웨어 유지보수의 단계로 프로세스가 이루어지는 경우가 이에 해당한다.

2) 세부 이행계획 수립

- 데이터 분석체계는 고전적인 폭포수(Water-Fall) 방식도 있으나 반복적인 정련과정을 통하여 프로젝트의 완성도를 높이는 방식을 주로 사용한다.
- 반복적인 분석 체계는 모든 단계를 반복하기보다 데이터 수집 및 확보 와 분석데이터를 준비하는 단계를 순차적으로 진행하고, 모델링 단계는 반복적으로 수행하는 혼합형을 많이 적용하며, 이러한 특성을 고려하 여 세부적인 일정계획도 수립해야 한다.



Refine Analytics Model

추진기간		201	3년							201	4년					
향후 추진과제	М1	M2	МЗ	M4	M5	М6	M7	М8	М9	M10	M11	M12	M13	M14	M15	M16
추진과제 1-1. 분석 알고리즘 및		: 3개월	!												5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5	
분석 아키텍처 설계									ļ				ļ			
추진과제 1-2. 분석 과제 Pilot 수행 •		47	월													
추진과제 2-1. 업무 프로세스 내재화를 위한 Process Innovation						3개월										
추진과제 2-2. 변화관리							(67	H월					
추진과제 3-1. 빅데이터 분석 활용 시스템 구축							67	H월								
추진과제 3-2. 유관시스템 고도화									•		47	∦월				

〈세부 추진 일정 계획의 예시 〉



1장 데이터 분석 기획의 이해

분석 거버넌스 체계 수립



분석 거버넌스 체계를 수립하기 위해선 기업 내의 전체적인 분석 기준들과 환경들을 분석하여 우리가 가지고 있는 현재 자원이 타 경쟁사와 유사 업종과 비교해서 어느정도의 수준에 있는지 평가를 해야 합니다. 또한 분석을 할 수 있는 분석조직, 분석인력에 대한 교육을 통해 분석 거버넌스를 수립하는 것이 중요합니다.





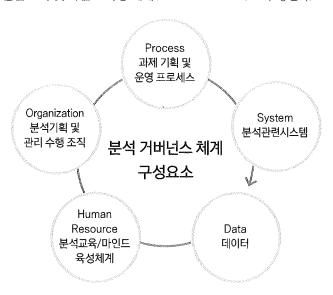
거버넌스 체계

가. 개요

• 기업에서 데이터를 이용한 의사결정이 강조될수록 데이터 분석과 활용을 위한 체계적인 관리가 중요해진다. 단순히 대용량 데이터를 수집·축적하는 것보다는 어떤 목적으로 어떤 데이터를 어떻게 분석에 활용할 것인가가 더욱 중요하기 때 문이다. 그리고 조직 내 분석 관리체계를 수립해야 하는 이유는 데이터 분석을 기 업의 문화로 정착하고 데이터 분석업무를 지속적으로 고도화하기 위해서이다.

나. 구성요소

• 마스터 플랜 수립 시점에서 데이터 분석의 지속적인 적용과 확산을 위한 거버넌스 체계는 **분석 기획 및 관리를 수행하는 조직(Organization), 과제** 기획 및 운영 프로세스(Process), 분석 관련 시스템(System), 데이터(Data), 분석 관련 교육 및 마인드 육성 체계(Human Resource)로 구성된다.



가. 개요

- 기업들은 데이터 분석의 도입 여부와 활용에 명확한 분석 수준을 점검할 필요가 있다. 데이터분석의 수준진단을 통해 데이터 분석 기반을 구현하기 위해서는 무엇을 준비하고 보완해야 하는지 등 분석의 유형 및 분석의 방향성을 결정할 수 있다.
- 데이터 분석 수준 진단을 위한 **분석 준비도**(Readiness)와 **분석 성숙도** (Maturity)의 구성은 아래와 같다.



(2)



분석 거버넌스 체계를 수립하기 위해 가장 먼저 선행되어야 하는 부분은 데이 터 분석 수준을 진단하는 것입니다. 분석 준비도와 분석 성숙도를 통해 기업의 분석 수준을 진단할 수 있습니다. 주관식으로 출제 될 수 있으니 꼭 기억합시다.



분석 준비도 Readiness

분석업무 분석 분석기법 인력,조직 분석데이터 분석문화 분석인프라

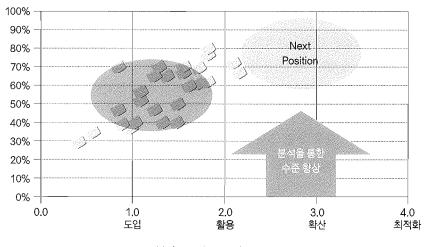


도입	활용	확산	최적화	
스니즈니브	조직및	역량	ìΤ	

나. 수준 진단 목표 2가지

1) 정의

- 기업의 현재 분석 수준을 명확히 이해하고, 수준진단 결과를 토대로 미래의 목표수준을 정의한다.
- 데이터 분석을 위한 기반 또는 환경이 유사업종 또는 타 경쟁사에 비해 어느 정도 수준이고, 데이터를 활용한 분석의 경쟁력 확보를 위해 어떠 한 영역에 선택과 집중을 해야 하는지, 어떤 관점을 보완해야 하는지 등 개선방안을 도출한다.



〈분석 목표 수준 정의〉

2) 분석 준비도

- 가) 목표: 기업의 데이터 분석 도입의 수준을 파악하기 위한 진단방법
- 나) 구성 : 총 6가지(분석업무파악, 인력 및 조직, 분석기법, 분석 데이터, 분석문화, IT 인프라)

다) 진단 과정

- ① 영역별로 세부 항목에 대한 수준파악
- ② 진단결과 전체 요건 중 일정 수준이상 충족하면 분석업무 도입
- ③ 충족하지 못할 시 분석 환경 조성



Q. 분석업무를 파악하고 인력·조직에 대해 평가하고 분석기법, 분석데이터, 분석문화, IT 인프라를 평가함으로써 기업의 분석 수준이 어느 정도이고 목표를 어떻게 세울 것인지에 대한 분석 거버넌스 체계를 수립하는데 활용되는 방법론은 무엇일까?



A. 분석준비도. 이런 주관식문제 뿐만 아니라 객관식문제도 출제될 수 있습니다. 6가지 요소 중 한가지요소의 내용을 제시하고 틀린 설명을 고르라는 문제 출제 될 수 있으니 각각의 세부내용을 자세히 살펴봅시다.

분석업무 파악

인력 및 조직

분석기법

- 발생한 사실 분석업무
- 예측 분석업무
- 시뮬레이션 분석업무
- 최적화 분석업무
- 분석업무 정기적 개선
- 분석 전문가 직무 존재
- 분석 전문가 교육 훈련 프로그램
- 관리자들의 기본적 분석 능력
- 전사 분석업무 총괄 조직 존재
- 경영진의 분석 업무 이해 능력
- 업무별 적합한 분석기법 사용
- 분석업무 도입 방법론
- 분석기법 라이브러리
- 분석기법 효과성 평가
- 분석기법 정기적 개선

분석 데이터

- 분석업무를 위한 데이터 충분성
- 분석업무를 위한 데이터 신뢰성
- 분석업무를 위한 데이터 적시성
- 비구조적 데이터 관리
- 외부 데이터 활용 체계
- 기준 데이터 관리(MDM)

분석 문화

- 사실에 근거한 의사결정
- 관리자의 데이터 중시 정도
- 회의 등에서 데이터 활용 상황
- 경영진의 직관 vs 데이터 기반의 의사 결정
- 데이터 공유 및 협업 문화

IT 인프라

- 운영시스템 데이터 통합
- EAI, ETL 등 데이터유통체계
- 분석 전용 서버 및 스토리지
- 빅데이터 분석 환경
- 통계 분석 환경
- 비쥬얼 분석 환경

3) 분석 성숙도 모델

- 가) 조직의 성숙도 평가 도구
 - CMMI(Capability Maturity Model Integration) 모델
- 나) 성숙도 수준분류
 - 도입단계, 활용단계, 확산단계, 최적화단계
- 다) 분석 성숙도 진단 분류
 - 비즈니스 부문, 조직·역량부문, IT부문



목표 시점별 분석 방향, 목표, 유형, 접근 방식에 대한 특성을 묻는 문제가 출제 돨 수 있습니다. 내용을 이해하시기 바랍니다.

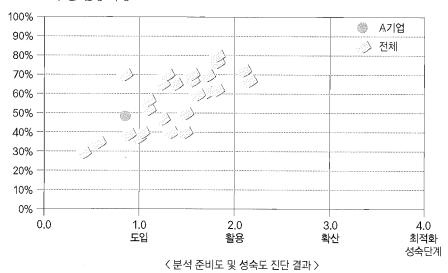


단계	도입단계	활용단계	확산단계	최적화단계
설명	분석을 시작하여 환경과 시스템을 구축	분석 결과를 실제 업무에 적용	전사 차원에서 분석을 관리하고 공유	분석을 진화시켜서 혁신 및 성과 향상에 기여
비즈니스 부문	• 실적분석 및 통계 • 정기보고 수행 • 운영 데이터 기반	미래 결과 예측시뮬레이션운영 데이터 기반	 전사성과실시간 분석 프로세스 혁신 3.0 분석규칙 관리 이벤트 관리 	• 외부 환경분석 활용 • 최적화업무적용 • 실시간 분석 • 비즈니스 모델 진화
조직역량 부문	• 일부 부서에서 수행 • 담당자 역량에 의존	전문 담당부서에서 수행 분석기법 도입 관리자가 분석 수행	• 전사 모든 부서 수행 • 분석 COE 조직 운영 • 데이터 사이언티스트 확보	• 데이터 사이언스 그룹 • 경영진 분석활용 • 전략 연계
∏부문	• 데이터 웨어하우스 • 데이터 마트 • ETL/EAI • OLAP	• 실시간 대시보드 • 통계분석환경	 빅데이터 관리 환경 시뮬레이션·최적화 비주얼 분석 분석 전용 서버 	분석 협업환경분석 Sandbox프로세스 내재화빅데이터 분석

〈 분석 성숙도 모델 〉

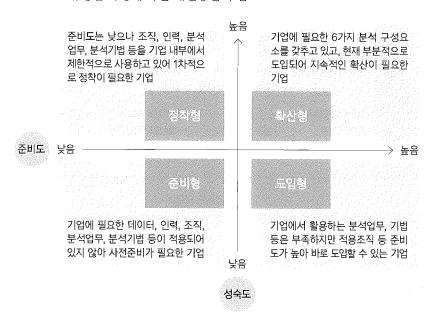
4) 분석 수준 진단 결과

- 기업의 현재 분석 수준을 객관적으로 파악
- 경쟁사의 분석 수준과 비교하여 분석 경쟁력 확보 및 강화를 위한 목표 수준 설정 가능



가) 분석 관점에서의 사분면 분석

- 분석 수준 진단결과를 구분
- 데이터 분석 수준에 대한 목표 방향을 정의
- 유형별 특성에 따른 개선방안 수립

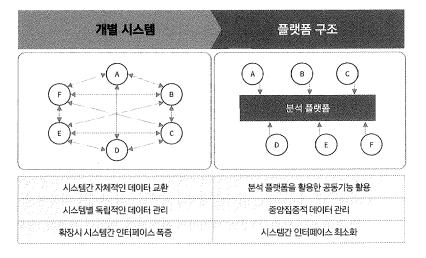


〈사분면 분석(Analysis Quadrant) 〉

가. 개요

• 분석과제 단위별로 별도의 분석시스템을 구축하는 경우, 관리의 복잡도 및 비용의 증대라는 부작용이 나타나게 된다. 따라서 분석마스터 플랜을 기획 하는 단계에서부터 장기적이고 안정적으로 활용할 수 있는 확장성을 고려 한 플랫폼 구조를 도입하는 것이 적절하다. 분석지원 인프라 방안 수립

(3)



나. 플랫폼

- 단순한 분석 응용프로그램뿐만 아니라 분석 서비스를 위한 응용프로그램 이 실행될 수 있는 기초를 이루는 컴퓨터 시스템을 의미한다.
- 일반적으로 하드웨어에 탑재되어 데이터 분석에 필요한 프로그래밍 환경과 실행 및 서비스 환경을 제공하는 역할을 수행한다.
- 분석플랫폼이 구성되어 있는 경우에는 새로운 데이터 분석 니즈가 존재할 경우 개별적인 분석 시스템을 추가하는 방식이 아닌 서비스를 추가적으로 제공하는 방식으로 확장성을 높일 수 있다.

	분석 서비스 제공 엔진				
광의의 분석 플랫폼	분석 어플리케이션				
	분석 서비스 제공 API				
협의의	데이터 처리 Framework				
분석 플랫폼	분석 엔진 분석 라이브러리				
народно удиници до повите и менент равно дите и изголиционного и и и и усторите до 18 менения и 18 менения и и	운영체제 (OS)				
하드웨어					

〈분석 플랫폼 구성 요소 〉



데이터 거버넌스 체계 수립



데이터 거버넌스 체계와 관련된 단답형 문제가 자주 출제되니 숙지하시기 바랍니다.



가, 데이터 거버넌스의 개요

- 1) 전사 차원의 모든 데이터에 대하여 정책 및 지침, 표준화, 운영조직 및 책임 등의 표준화된 관리체계를 수립하고 운영을 위한 프레임워크(Framework) 및 저장소(Repository)를 구축하는 것을 말한다.
- 2)마스터 데이터(Master Data), 메타 데이터(Meta Data), 데이터 사전(Data Dictionary)은 데이터 거버넌스의 중요한 관리 대상이다.
 - 기업은 데이터 거버넌스 체계를 구축함으로써 데이터의 가용성, 유용성, 통합성, 보안성, 안전성을 확보할 수 있다.
 - 데이터 거버넌스는 독자적으로 수행될 수도 있지만 전사 차원의 IT 거버넌 스나 EA(Enterprise Architecture)의 구성요소로써 구축되는 경우도 있다.
 - 빅데이터 거버넌스는 이러한 데이터 거버넌스의 체계에 대하여 빅데이터 의 효율적인 관리, 다양한 데이터의 관리체계, 데이터 최적화, 정보보호, 데이터 생명주기 관리, 데이터 카테고리별 관리 책임자(Data Steward) 지정 등을 포함한다.

나. 데이터 거버넌스 구성요소

1) 개요

• 구성요소인 원칙(Principle), 조직(Organization), 프로세스(Process) 는 유기적으로 조합하고 효과적으로 관리하여, 데이터를 비즈니스 목적 에 부합하도록 하고 최적의 정보 서비스를 제공할 수 있도록 한다.

2) 구성3요소

가) 원칙(Principle)

- 데이터를 유지·관리하기 위한 지침과 가이드
- 보안, 품질 기준, 변경관리

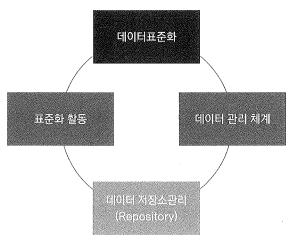
나) 조직(Organization)

- 데이터를 관리할 조직의 역할과 책임
- 데이터 관리자, 데이터베이스 관리자, 데이터 아키텍트(Data Architect)

다) 프로세스(Process)

- 데이터 관리를 위한 활동과 체계
- 작업 절차, 모니터링 활동, 측정 활동

다. 데이터 거버넌스 체계



〈데이터 거버넌스 체계〉

1) 데이터 표준화

- 데이터 표준화는 데이터 표준 용어 설정, 명명 규칙(Name Rule)수립, 메타 데이터(Meta Data)구축, 데이터 사전(Data Dictionary)구축 등의 업무로 구성된다.
- 데이터 표준용어는 표준 단어사전, 표준 도메인사전, 표준 코드 등으로 구성되며 사전간 상호 검증이 가능하도록 점검 프로세스를 포함해야 한 다.
- 명명 규칙은 필요시 언어별(한글, 영어 등)로 작성되어 매핑 상태를 유지해야 한다.

2) 데이터 관리 체계

- 데이터 정합성 및 활용의 효율성을 위하여 표준 데이터를 포함한 메타데이터(Meta Data)와 데이터 사전(Data Dictionary)의 관리 원칙을 수립한다.
- 수립된 원칙에 근거하여 항목별 상세한 프로세스를 만들고 관리와 운영을 위한 담당자 및 조직별 역할과 책임을 상세하게 준비한다.

• 빅데이터의 경우 데이터양의 급증으로 데이터의 생명 주기 관리방안 (Data Life Cycle Management)을 수립하지 않으면 데이터 가용성 및 관리비용 증대 문제에 직면하게 될 수 있다.

3) 데이터 저장소 관리(Repository)

- 메타데이터 및 표준 데이터를 관리하기 위한 전사 차원의 저장소를 구성한다.
- 저장소는 데이터 관리 체계 지원을 위한 워크플로우(Workflow) 및 관리용 응용 소프트웨어(Application)를 지원하고 관리 대상 시스템과의 인터페이스를 통한 통제가 이루어져야한다.
- 데이터 구조 변경에 따른 사전 영향 평가도 수행되어야 효율적인 활용이 가능하다.

4) 표준화 활동

- 데이터 거버넌스 체계를 구축한 후 표준 준수 여부를 주기적으로 점검하고 모니터링을 실시한다.
- 거버넌스의 조직 내 안정적 정착을 위한 계속적인 변화 관리 및 주기적 인 교육을 진행한다.
- 지속적인 데이터 표준화 개선 활동을 통하여 실용성을 높여야 한다.



데이터 조직 및 인력방안 수립

가. 현황

• 빅데이터 등장에 따라 기업의 비즈니스도 많은 변화를 겪고 있는데, 이러한 비즈니스 변화를 인식하고 기업의 차별화된 경쟁력을 확보하는 수단으로서 데이터 과제 발굴, 기술 검토 및 전사 업무적용계획 수립 등 데이터를 효과 적으로 분석·활용하기 위해 기획, 운영 및 관리를 전담할 수 있는 전문 분석 조직의 필요성이 제기되고 있다.

나. 분석 조직의 개요

데이터 분석 조직은 기업의 경쟁력 확보를 위해 데이터 분석의 가치를 발견하고, 이를 활용하여 비즈니스를 최적화하는 목표를 갖고 구성되어야 한다.
 이를 위해 기업의 업무 전반에 걸쳐 다양한 분석 과제를 발굴해 정의하고, 데이터 분석을 통해 의미있는 인사이트를 찾아 실행하는 역할을 수행할 수 있어야 한다. 다양한 분야의 지식과 경험을 가진 인력과 업무 담당자 등으로 구성된 전사 또는 부서 내 조직으로 구성할 수 있다.

지업의 경쟁력 확보를 위하여 비즈니스 질문(Question)과 이에 부합하는 가치(Value)를 찾고 비즈니스를 최적화(Optimization)하는 것

역할

전자 및 부서의 분석 업무를 발굴하고 전문적 기법과 분석 도구를 활용하여 기업 내 존재하는 빅데이터 속에서 Insight를 전파하고 이를 Action화 하는 것

구성

기초통계학 및 분석 방법에 대한 지식과 분석 경험을 가지고 있는 인력으로 전사 또는 부서 내 조직으로 구성하여 운영

〈분석 조직의 개요〉

다. 조직 및 인력 구성 시 고려사항

1) 주요 고려사항

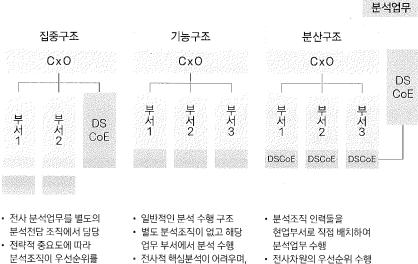
	주요 고려사항
조직 구조	 비즈니스 질문(Question)을 선제적으로 찾아 낼 수 있는 구조인가? 분석 전담조직과 타 부서간 유기적인 협조와 지원이 원활한 구조인가? 효율적인 분석 업무를 수행하기 위한 분석 조직의 내부 조직구조는? 전사 및 단위부서가 필요 시 접촉하며 지원할 수 있는 구조인가? 어떤 형태의 조직(중앙집중형, 분산형)으로 구성하는 것이 효율적인가?
인력 구성	 비즈니스 및 IT 전문가의 조합으로 구성되어야 하는가? 어떤 경험과 어떤 스킬을 갖춘 사람으로 구성해야 하는가? 통계적 기법 및 분석 모델링 전문 인력을 별도로 구성해야 하는가? 전사 비즈니스를 커버하는 인력이 없다. 그렇다면? 전사 분석업무에 대한 적합한 인력 규모는 어느 정도인가?



특정 구조에 대한 설명을 주고 어떤 구조에 대한 설명인지 고르는 문제가 출제 될 수 있습니다. 또는 그림을 제시하고 일어나는 현상 중 틀린 것을 고르게 하는 문제가 나올 수 있으니 본 내용를 완벽히 숙지하도록 합시다!



2) 분석을 위한 3가지 조직 구조



M DSCoE: Data Science Center of Excellence

- 전사차원의 우선순위 수행
- 분석결과에 따른 신속한 Action 가능
- 베스트 프랙티스 공유 가능
- 부서 분석업무와 역할 분담 명확히 해야함(→업무과다 이원화 가능성)

〈분석 조직 구조〉

부서 현황 및 실적 통계 등

가능성 높음

과거 실적에 국한된 분석 수행

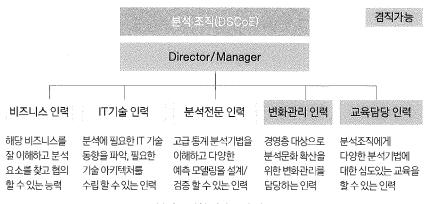
3) 분석조직의 인력구성

정해서 진행 가능

• 현업 업무부서의 분석업무와

이중화/이원화 가능성 높음

• 전문역량을 갖춘 각 분야의 인재들을 모아 조직을 구성하여 분석조직의 경쟁력을 극대화 할 수 있다.



〈분석 조직의 인력 구성 예시 〉

가. 현황

• 분석 마스터 플랜이 수립되고 초기 데이터 분석 과제가 성공적으로 수행되는 경우, 지속적인 분석 니즈 및 기회가 분석 과제 형태로 도출될 수 있다. 이런 과정에서 분석 조직이 수행할 주요한 역할 중의 하나가 분석 과제의 기획 및 운영이므로 이를 체계적으로 관리하기 위한 프로세스를 수립해야 한다.

분석 과제 관리 프로세스 수립

6

나. 과제 관리 프로세스

1) 과제 발굴

• 개별 조직이나 개인이 도출한 분석 아이디어를 발굴하고 이를 과제화하여 분석 과제 풀(Pool)로 관리하면서 분석 프로젝트를 선정하는 작업을 수행 한다.

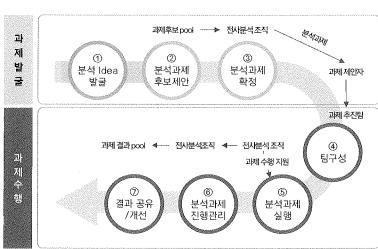
2) 과제 수행

• 분석을 수행할 팀을 구성하고 분석 과제 실행 시 지속적인 모니터링과 과제결과를 공유하고 개선하는 절차를 수행한다.



분석 과제 관리 프로세스에 대한 잘못된 것을 선택하는 문제가 자주 출제되니 숙지할 수 있도록 합시다.





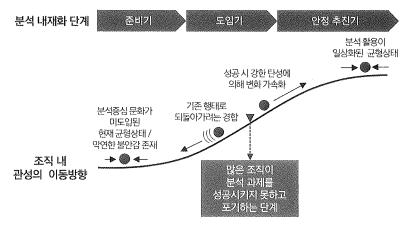
- 분석 조직이 지속적이고 체계적인 분석 관리 프로세스를 수행함으로써 조직 내 분석 문화 내재화 및 경쟁력을 확보할 수 있다.
- 해당 과제를 진행하면서 만들어진 시사점(Lesson Learned)을 포함한 결과물을 풀(Pool)에 잘 축적하고 관리함으로써 향후 유사한 분석과제 수행 시 시행착오를 최소화하고 프로젝트를 효율적으로 진행할 수 있다.



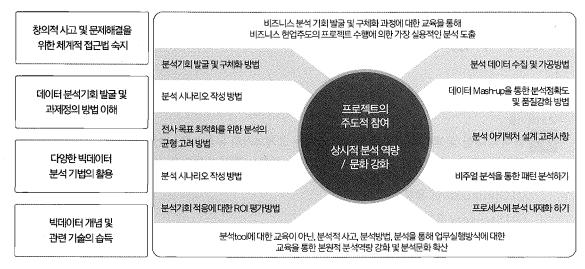
분석 교육 및 변화관리

가, 개요

- 빅데이터의 등장으로 많은 비즈니스 영역에서 변화를 가져왔다. 이러한 변화에 보다 적극적으로 대응하기 위해서는 기업에 맞는 적합한 분석 업무를 도출하고, 가치를 높여줄 수 있도록 분석 조직 및 인력에 대한 지속적인 교육과 훈련을 실시하여야 한다. 또한 경영층이 사실 기반(Fact-Based) 의사결정을 할수 있는 문화를 정착시키는 등 지속적인 변화관리를 계획하고 수행하여야 한다.
- 새로운 체계의 도입시에는 저항 및 기존 행태로 되돌아가는 관성이 존재하기 때문에 분석의 가치를 극대화하고 내재화하는 안정적인 추진기로 접어들기 위해서는 분석에 관련된 교육 및 마인드 육성을 위한 적극적인 변화관리가 필요하다.



〈 분석 도입에 대한 문화적 대응 〉



〈데이터 분석 방법 및 분석적 사고 교육 〉

나. 분석 교육의 목표

• 단순한 툴 교육이 아닌 분석역량을 확보하고 강화하는 것에 초침을 맞추어 진행되어야 한다.

참교

① 분석 기획자: 데이터 분석 큐레이션 교육



- ② 분석 실무자: 데이터 분석 기법 및 툴에 대한 교육
- ③ 업무 수행자 : 분석기회 발굴, 구체화, 시나리오 작성법 등
- 분석적인 사고를 업무에 적용할 수 있도록 다양한 교육을 통해 조직 구성원 모두에게 분석기반의 업무를 정착시키고 이를 통해 데이터를 바라보는 관 점, 데이터 분석과 활용 등이 기업 문화로 자연스럽게 확대되어야 한다.

참교

빅데이터 시대의 변화에 적극적인 대응방법

- ① 기업에 맞는 적합한 분석업무의 수행
- ② 분석 조직 및 인력에 대한 지속적인 교육과 훈련 실시
- ③ 경영층이 사실기반(Fact Based)의사결정을 할 수 있는 문화 정착



④ 지속적인 변화관리를 계획하고 수행·업무 수행자에 대한 분석기회 발굴, 구체화, 시나리오 작성법 등

데이터 분석 기획

✔ 21회 기출

- ①1 분석의 대상 및 방식에 따라서 분석의 주제는 크게 4가지 유형으로 나뉜다. 이 중 분석 대상은 명확하지만 분석 방식이 명확하지 않은 경우 수행하는 주제 유형으로 가장 적절한 것은?
 - ① Optimization 유형
 - ② Discovery 유형
 - ③ Insight 유형
 - ④ Solution 유형

✔ 18회 기출

02 아래 (가)와 (나)에 순서대로 들어갈 내용으로 적절한 것은?

분석은 분석 대상(What) 및 분석 방법(How)에 따라서 4가지로 나눌 수 있다. 분석 대상이 명확하게 무엇인지 모르는 경우에는 기존 분석 방식을 활용하여 (가)을(를) 도출해냄으로 써 문제의 도출 및 해결에 기여하거나 (나) 접근법으로 분석 대상 자체를 새롭게 도출할 수 있다.

① 최적화 - 통찰

② 솔루션 - 통찰

③ 통찰 - 발견

④ 발견 - 솔루션

✔ 18회 기출

- 03 분석 기획 고려사항 중 장애요소에 대한 설명으로 부적절한 것은?
 - ① 비용대비 효과의 적정한 비용
 - ② 분석 모형의 안정적 성능 확보
 - ③ 이해도 높은 모형보다는 복잡하고 정교한 모형
 - ④ 조직 역량으로 내재화를 위한 변화 관리

✔ 14회 기출

- 04 다음 중 성공적인 분석을 위해서 고려해야 할 요소로 가장 부적절한 것은?
 - ① 관련 데이터의 파악
 - ② 원접에서 솔루션 탐색
 - ③ 비즈니스 케이스 확보
 - ④ 이행 저해 요소 관리

✔ 25회 기출

0.5 다양한 데이터 유형 중 정형 데이터 - 반정형 데이터 - 비정형데이터 순서로 가장 적절한 것은?

- ① Demand Forecasts Competitor pricing Email records
- ② Facebook status Weather data Web logs
- ③ RFID Internet of things sensing Loyalty program
- ④ CRM Transaction data Twitter feeds Mobile location

✔ 15회 기출

- 06 다음 분석 방법론 중 순차적으로 진행되면서 이전 단계가 완료된 후 다음 단계로 진행하는 하향식 (Top Down)으로 진행되는 특징을 가지고 있는 모델은?
 - ① 프로토타입(Prototype) 모델
 - ② 폭포수(Waterfull) 모델
 - ③ 나선형(Spiral) 모델
 - ④ 애자일(Agile) 모델

✔ 25회 기출

- 07 다음 중 CRISP-DM 방법론의 모델링 단계에서 수행하는 테스크(task)로 적절하지 않은 것은 무엇인가?
 - ① 모델 테스트 계획 설계
 - ② 모델 평가
 - ③ 모델링 기법 선택
 - ④ 모델 적용성 평가

✔ 17회 기출

- 08 빅데이터 분석 방법론의 분석기획 단계 순서가 바르게 연결된 것은?
 - ① 프로젝트 범위 설정-데이터 분석 프로젝트 정의-프로젝트 수행계획 수립-데이터 분석 위험 식별
 - ② 프로젝트 범위 설정-데이터 분석 프로젝트 정의-데이터 분석 위험 식별-프로젝트 수행계획 수립
 - ③ 데이터 분석 위험 식별-프로젝트 범위 설정-프로젝트 수행계획 수립-데이터 분석 프로젝트 정의
 - ④ 데이터 분석 위험 식별-프로젝트 범위 설정-데이터 분석 프로젝트 정의-프로젝트 수행계획 수립

✔ 25회 기출

🕡 🔾 다음 중 빅데이터 분석 방법론에서 단계 간 피드백이 반복적으로 많이 발생할 수 있는 단계는?

- ① 분석 기획 단계 → 데이터 준비 단계
- ② 데이터 준비 단계 → 데이터 분석 단계
- ③ 데이터 분석 단계 → 시스템 구현 단계
- ④ 시스템 구현 단계 → 평가와 전개 단계

✔ 13회 기출

🌃 0 다음 중 기업에서 데이터에 기반한 의사결정을 방해하는 요소들로 구성된 것은?

- ① 바이어스, 비편향적 사고
- ② 프레이밍 효과. 고정관념
- ③ 프레이밍 효과, 직관력
- ④ 직관력, 비편향적 사고

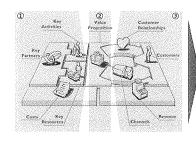
✔ 14회 기출

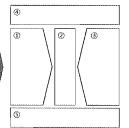
11 비즈니스 모델 캔버스의 채널(Channels)에 대한 기능으로 가장 부적절한 것은?

- ① 해당고객에게 접근하는 유통 채널을 공급한다.
- ② 고객에게 밸류 프로포지션을 전달한다.
- ③ 구매 고객에 대한 에프터서비스(A/S)를 제공한다.
- ④ 기업이 제공하는 상품이나 서비스에 대한 고객의 이해를 높여준다.

✔ 22회 기출

12 아래 그림은 비즈니스 모델 캔버스를 활용한 과제 발굴 기법을 보여주고 있다. 각 빈칸에 들어갈 말로 적절한 것은?





- ① 1: 업무, 2: 제품, 3: 고객, 4: 규제&감사, 5: 지원 인프라
- ② 1: 업무, 2: 가치, 3: 고객, 4: 규제&감사, 5: 지원 인프라
- ③ 1: 업무, 2: 제품, 3: 고객, 4: 관리&감독, 5: 지원&자산
- ④ 1: 업무, 2:가치, 3:고객, 4:자원&자산, 5:지원 인프라

✔ 15회 기출

13 데이터 분석에서는 하향식 접근 방식과 상향식 접근 방식으로 분석과제를 발굴하게 되는데, 다음 중 하향식 접근 방식의 단계에서 타당성 평가에 대한 설명으로 가장 부적절한 것은?

- ① 도출된 분석 문제에 대한 대안을 과제화하기 위해서는 다각적 타당성 검토가 필요하다.
- ② 경제적 타당성은 비용대비 효익의 관점에서 평가한다.
- ③ 데이터 타당성 확보를 위하여 문제발생 포인트에 대한 데이터 확보가 중요하다.
- ④ 기술적 타당성 분석 시 적용 가능한 요소기술 확보 방안에 대한 사전 고려가 필요하다.

✔ 18회 기출

1 가장식 데이터 분석 기획에서 문제 탐색 단계에 대한 설명으로 가장 부적절한 것은?

- ① 빠짐없이 문제를 도출하고 식별하는 것이 중요
- ② 문제를 해결함으로써 발생하는 가치에 중점을 두는 것이 중요
- ③ 비즈니스 모델 캔버스는 문제 탐색 도구로 활용
- ④ 문제 탐색은 유스케이스 활용보다는 새로운 이슈탐색이 우선

✔ 25회 기출

15 다음 중 '분석과제 정의서'에 대한 설명으로 가장 적절한 것은?

- ① '분석과제 정의서'에는 소스 데이터, 데이터 입수 및 분석의 난이도, 분석방법 등에 대한 항목이 포함되어야 한다.
- ② '분석과제 정의서'는 프로젝트를 수행하는 이해관계자가 프로젝트의 방향을 설정하고 성공 여부를 판별할 수 없는 자료이다.
- ③ '분석과제 정의서'에는 분석모델에 적용될 알고리즘과 분석모델의 기반이 되는 Feature가 포함되어야 한다.
- ④ '분석과제 정의서'는 프로젝트 계획서를 작성하기 위한 중간 결과로써 구성 항목(Configuration Item)으로 도출할 필요는 없다.

✔ 15회 기출

16 다음 중 분석 마스터플랜의 세부 이행계획 수립시 고려해야 할 데이터 분석체계(분석 방법론)에 대한 설명으로 가장 적절한 것은?

- ① 데이터 분석 체계는 이해도가 높은 폭포수 모델이 가장 적정하다.
- ② 반복적 정련 방식은 데이터 수집 및 확보 단계를 반복적으로 수행한다.
- ③ 프로토타입(Prototype) 모델은 데이터 분석체계로는 적절하지 못하다.
- ④ 프로젝트의 세부 일정계획도 데이터 분석체계를 고려하여 작성한다.

✔ 21회 기출

◢ ◢ 다음 중 분석 프로젝트 관리에 대한 설명으로 가장 부적절한 것은?

- ① 데이터 분석 모델의 품질을 평가하기 위해서 SPICE를 활용할 수 있다.
- ② 분석 프로젝트 관리는 KSA ISO 21500:2013를 가이드로 활용할 수 있다.
- ③ 분석 프로젝트의 일정계획 수립시 데이터 수집에 대한 철저한 통제와 관리가 필요하다.
- ④ 분석 프로젝트의 최종 산출물이 보고서인지 또는 시스템인지에 따라 프로젝트 관리에 차이가 있다.

✔ 13회 기출

🌃 🎖 다음 중 데이터 분석 과제에서 프로젝트 관리에 대한 설명으로 가장 부적절한 것은?

- ① 분석 과제는 분석 전문가의 상상력을 요구하므로 일정을 제한하는 입정계획은 적절하지 못하다.
- ② 분석 과제에는 많은 위험이 있어 사전에 위험을 식별하고 대응방안을 수립해야 한다.
- ③ 분석과제는 적용되는 알고리즘에 따라 범위가 변할 수 있어 범위관리가 중요하다.
- ④ 분석과제에서 다양한 데이터를 확보하는 경우가 있어 조달관리 또한 중요하다.

✔ 14회 기출

다음 중 데이터 분석을 위한 수준진단에서 '분석 준비도'의 분석 데이터의 진단 항목으로 가장 부적절한 것은?

- ① 분석 업무를 위한 데이터 충실성, 신뢰성, 적시성
- ② 내부데이터 집중 활용 체계
- ③ 기준데이터 관리(MDM)
- ④ 비구조적 데이터 관리

✔ 23회 기출

10 다음 빅데이터의 특징 중 투자비용 요소 또는 난이도를 평가하는 요소가 아닌 것은?

- ① Volume
- ② Variety
- ③ Value
- 4 Velocity

✔ 23회 기출

- 21 다음 중 난이도와 시급성을 고려하였을 때 우선적으로 추진해야 하는 분석 과제로 적절한 것은?
 - ① 난이도 : 쉬움(Easy), 시급성 : 현재
 - ② 난이도 : 어려움(Difficult), 시급성 : 미래
 - ③ 난이도 : 쉬움(Easy), 시급성 : 미래
 - ④ 난이도 : 어려움(Difficult), 시급성 : 현재

✔ 19회 기출

- 22 지속적인 분석 내재화를 위한 "장기적인 마스터 플랜 방식"에 비하여 "과제 중심적인 접근 방식"의 특징으로 가장 적절하지 못한 것은?
 - ① Quick-Win
 - ② Accuracy & Deploy
 - ③ Problem Solving
 - 4 Speed & Test

✔ 13회 기출

- 23 기업의 데이터 분석과제 수행을 위한 수준을 평가하기 위하여 분석 준비도(Readiness)를 파악해야 한다. 다음 중 데이터 분석 준비도 프레임워크에서 분석 업무 파악 영역으로 가장 부적절한 것은?
 - ① 최적의 분석 업무
 - ② 업무별 적합한 분석 기법
 - ③ 예측 분석 업무
 - ④ 발생한 사실 분석 업무

✔ 25회 기출

24 다음 중 아래의 데이터 거버넌스 체계가 설명하는 항목은?

메타데이터 관리, 데이터 사전관리, 데이터 생명주기 관리

- ① 데이터 표준화
- ② 데이터 관리 체계
- ③ 데이터 저장소 관리
- ④ 표준화 활동

✔ 21회 기출

데이터 분석을 위한 조직 구조 중 아래 보기에 해당하는 것은?

- 전사 분석업무를 별도의 분석 전담 조직에서 담당
- 전략적 중요도에 따라 분석조직이 우선순위를 정해서 진행 가능
- 현업 업무부서의 분석업무와 이중화/이원화 가능성 높음
- ① 집중구조
- ② 기능구조
- ③ 분산구조
- ④ 복합구조

✔ 18회 기출

②6) 분석 마스터 플랜 수립에서 과제 우선순위 결정과 관련한 내용으로 부적절한 것은?

- ① 가치는 투자비용 요소이다.
- ② 전략적 중요도, ROI, 실행 용이성은 분석과제 우선순위 결정에 고려할 사항이다.
- ③ 시급성과 전략적 필요성은 전략적 중요도의 평가 요소이다.
- ④ 적용 기술의 안전성 검증은 기술 용이성의 평가 요소이다.

✔ 17회 기출

27 분석 마스터 플랜에 대한 설명으로 가장 부적절한 것은?

- ① 과제 우선순위 평가는 비즈니스 효과인 시급성과 투자비용 요소인 난이도에 근거하여 결정 된다.
- ② 분석 마스터 플랜은 분석과제 도출, 우선순위 결정, 중장기 마스터 플랜 수립과제 도출 순으 로 된다.
- ③ 과제별 데이터 분석 체계는 폭포수 방식도 있으나 반복적인 정련과정을 통하여 과제의 완성 도를 높이는 방식으로 많이 사용한다.
- ④ 분석 과제 로드맵은 과제의 우선순위를 고려하여 작성하되 과제별 선후관계를 감안하여 반 복이 없는 계획을 작성한다.

✔ 2.5회 기출

28 다음 중 분석 기회 발굴의 범위 확장 방법에 관한 설명으로 부적절한 것은?

- ① 거시적 관점의 메가트랜드에서는 현재의 조직 및 해당 산업에 폭넓게 영향을 미치는 사회·경제적 요인을 사회·기술·경제·환경·정치 영역으로 나누어서 좀 더 폭넓게 기회 탐색을 수행한다.
- ② 경쟁자 확대 관점에서는 현재 수행하고 있는 사업 영역의 직접 경쟁사 및 제품, 서비스를 중심으로 현 상황에 대한 분석 기회 발굴의 폭을 넓혀서 탐색한다.
- ③ 시장의 니즈 탐색 관점에서는 현재 수행하고 있는 사업에서의 직접 고객뿐만 아니라 고객과 접촉하는 역할을 수행하는 채널 및 고객의 구매와 이사결정에 영향을 미치는 영향자들에 대한 폭넓은 관점을 바탕으로 분석 기회를 탐색한다.
- ④ 역량의 재해석 관점에서는 현재 해당 조직 및 기업이 보유한 역량뿐만 아니라 해당 조직의 비즈니스에 영향을 끼치는 파트너 네트워크를 포함한 활용 가능한 역량을 토대로 폭넓은 분석 기회를 탐색한다.

✔ 23회 기출

29 아래의 내용 중 빅데이터 거버넌스에 대한 설명으로 가장 적절한 것은?

- ⊙ 빅데이터 분석은 다양한 데이터를 활용하기 위하여 회사 내 모든 데이터를 활용해야 한다.
- © ERD는 운영 중인 데이터베이스와 일치하기 위하여 철저한 변경관리가 필요하다.
- ② 빅데이터 거버넌스는 산업분야별, 데이터 유형별, 정보 거버넌스 요소별로 구분하여 작성한다.
- ② 빅데이터 분석은 고품질의 데이터 확보가 필요하므로 데이터 수명 주기 관리 보다는 품질관리가 중요하다.
- 1) (7), (1)
- (2) (E) (E)
- (3) (1), (12)
- 47.2

✔ 15회 기출

- 30 빅데이터를 활용한 비즈니스는 기업에 많은 변화를 가져오고 있다. 다음 중 기업에서 이러한 변화를 수용하기 위한 중장기적 대응 방안으로 가장 거리가 먼 것은?
 - ① 분석 조직 및 인력에 대한 교육과 훈련
 - ② 데이터 기반의 의사결정문화 정착
 - ③ 데이터 분석 도구 기반의 교육
 - ④ 분석역량 강화를 위한 체계적인 계획 및 시행

✔ 13회 기출

31. 다음 중 분석과제 관리 프로세스에 대한 설명으로 가장 부적절한 것은?

- ① 과제발굴단계에는 분석 아이디어 발굴, 분석과제 후보제안, 분석과제 확정 프로세스가 있다.
- ② 분석과제로 확정된 분석과제를 풀(Pool)로 관리한다.
- ③ 분석과제 중에 발생된 시사점과 분석 결과물은 풀(Pool)로 관리하고 공유된다.
- ④ 과제수행단계에서는 팀 구성, 분석과제 실행, 분석과제 진행관리, 결과 공유 프로세스가 있다.

✔ 14회 기출

32 다음 중 기존에 행해졌던 데이터 분석과 비교한 빅데이터 분석의 특징으로 적절한 것은?

- (ㄱ) 분석대상 데이터를 모든 형태 및 내외부 데이터로 확대한다.
- (ㄴ) 데이터의 생산 시점에서부터 실시간에 가까운 분석이 가능하다.
- (디) 데이터 마트에 정형 데이터를 적재하고 데이터 분석을 통하여 모델을 만들 수 있다.
- (ㄹ) 고급 분석기법을 활용할 수 있다.
- ① (¬), (∟), (⊏)
- ②(¬),(∟),(≥)
- ③ (८), (੮), (₴)
- ④ (¬), (∟), (⊏), (≥)

✔ 15회 기출

33 다음 중 빅데이터 분석에서 Self Service Analytics 에 대한 설명으로 가장 부적절한 것은?

- ① Self Service Analytics에 포함되어야 하는 주요 기능은 BI도구, Ad hoc Report, OLAP, Visual Discovery, MachineLearning 등이다.
- ② Self Service Analytics의 성공적인 적용을 위해서는 Reference Method 의 작성 및 공유, 표준 데이터의 활용, 데이터 거버넌스, 도구 사용에 대한 지속적인 교육이 필요하다.
- ③ Self Service Analytics를 수행하기 위해서는 R, Python 등의 데이터 분석 언어와 많은 통계적 지식을 필요로 한다.
- ④ Self Service Analytics는 상업용뿐만 아니라 OSS(Open Service Software)로도 구현 되고, 분석의 성능 향상을 위하여 분산처리도 지원하고 있다.

✔ 14회 기출

- 34 4차 산업혁명은 기존 산업을 초월하는 Mass Production, Mass Customization, Servitization을 구현하게 되는데, 다음 중 Servitization에 대한 설명으로 가장 부적절한 것은?
 - ① Service Science의 하나로 제품과 서비스의 결합을 통한 기업의 새로운 비즈니스 모델
 - ② 제품의 서비스화와 서비스의 상품화를 모두 포함하는 결합 비즈니스 모델
 - ③ 파이프 라인 비즈니스를 넘어 플랫폼 비즈니스를 위한 모델
 - ④ 생산된 제품의 특화를 통한 경쟁우위 확보 모델

✔ 20회 기출

- 35 다음 중 빅데이터 분석 방법론의 분석 기획 단계에서 프로젝트 위험 대응 계획을 수립할 때 예상 되는 위험에 대한 대응 방법의 구분으로 부적절한 것은?
 - ① 회피(Avoid)
 - ② 관리(Manage)
 - ③ 완화(Mitigate)
 - ④ 수용(Accept)

✔ 25회 기출

- 36 프로토타이핑(Prototyping) 접근법에 대한 설명으로 가장 적절한 것은?
 - ① 문제가 정형화되어 있고 문제해결을 위한 데이터가 완벽하게 조직에 존재하는 경우 효과적이다.
 - ② 신속하게 해결책이나 모형을 제시함으로써 이를 바탕으로 문제를 좀 더 명확하게 인식하고 필요한 데이터를 식별하여 구체화할 수 있게 하는 유용한 상향식 접근 방법이다.
 - ③ 문제가 주어지고 이에 대한 해법을 찾기 위하여 각 과정이 체계적으로 단계화되어 수행하는 방식이다.
 - ④ 문제 정의가 불명확하거나 이전에 접하지 못한 새로운 문제일 경우에는 적용하기 어렵다.

✔ 15회 기출

37 아래 (가) 안에 들어갈 용어를 기입하시오.

분석 방법론의 "시스템 구현" 단계에서 시스템으로 구현된 모델은 검증(Verification&Validation)을 위하여 단위 테스트, 통합테스트, 시스템 테스트 등을 실시한다. 이중 (가) 테스트는 품질관리 차원에서 진행함으로써 적용된 시스템의 객관성과 안정성을 확보한다.

✔ 15회 기출

38 아래 (가) 에 들어갈 용어를 기입하시오.

데이터 거버넌스 체계에서 데이터 저장소(Repository) 관리란 메타데이터 및 표준 데이터를 관리하기 위한 전사 차원의 저장소로 구성된다. 저장소는 데이터 관리 체계 지원을 위한 워크 플로우 및 관리용 응용소프트웨어를 지원하고 관리 대상 시스템과의 인터페이스를 통한 통제가 이루어져야 한다. 또한 데이터 구조 변경에 따른 (가)도 수행되어야 효율적인 활용이 가능하다.

)

(

✔ 14회 기출

39 아래 (가) 안에 들어갈 용어를 기입하시오.

문제 탐색을 통해서 식별된 비즈니스 문제를 변환하는 단계로써, 문제 탐색 단계가 무엇을 어떤 목적으로 수행해야 하는가에 대한 관점이었다면, (가)단계는 이를 달성하기 위해서 필요한 데이터 및 기법(How)을 도출하기 위한 데이터 분석의 문제로의 변환을 수행하게 된다.

(

✔ 14회 기출

40 아래 (가) 에 들어갈 용어를 기입하시오.

분석 모델을 가동중인 운영시스템에 적용하기 위해서는 모델에 대한 상세한 "알고리즘 설명서" 작성이 필요하다. "알고리즘 설명서"는 "시스템 구현"단계에서 중요한 입력 자료로 활용되므로 필요시 (가) 수준의 상세한 작성이 필요하다.

.

✔ 21회 기출

41 아래 ()안에 각각 들어갈 용어로 적절한 것은?

분석 과제 관리 프로세스는 크게 과제 발굴과 (③)(으)으로 나누어진다. 조직이나 개인이 도출한 분석 아이디어를 발굴하고 이를 과제화하여 분석 과제 풀(Pool)로 관리하면서 분석 과제가 확정되면 (⑥), (⑥), 분석과제 결과 공유/개선의 분석과제 관리 프로세스를 수 행하게 된다.

(

✔ 13회 기출

아래 (⊙), (ⓒ) 안에 들어갈 정확한 용어를 기입하시오.

데이터 거버넌스 체계는 데이터 저장소(Repository)는 메타데이터 및 표준 데이터를 관리하

기 위한 전사 차원의 저장소로 구성된다. 저장소는 데이터 관리 체계 지원을 위한 (🗇) 및 관리용 응용소프트웨어를 지원하고 관리 대상 시스템과의 인터페이스를 통한 통제가 이루 어져야 한다. 또한 데이터 구조 변경에 따른 (①)도 수행되어야 효율적인 활용이 가능하다. (✔ 12회 기출 $m{43}$ 아래 (ㄱ), (ㄴ), (ㄷ) 안에 들어갈 용어를 순서대로 기입하시오. 비즈니스 모델 캔버스는 9가지 블록을 단순화하여 (ㄱ), (ㄴ), 고객단위로 문제를 발 굴하고 이를 관리하는 규제와 감사, (ㄷ) 영역으로 나눠 분석 기회를 도출한다. (✔ 16회 기출 4.4 KDD 분석 방법론에서 잡음, 이상치, 결측치를 식별하여 분석용 데이터셋을 선택하고 분석에 필 요한 변수 등을 선정하는 단계와 유사한 CRISP-DM 방법론의 단계는? () ✔ 15회 기출 🖊 5 합리적 의사결정을 방해하는 요소로써 표현방식 및 발표자에 따라 동일한 사실에도 판단을 달리 하는 현상을 무엇이라 하는가? ()

✔ 18회 기출

🖊 🐧 문제가 주어지고 이에 대한 해법을 찾기 위하여 각 과정이 체계적으로 단계화되어 수행하는 분 석과제 발굴 방식을 무엇이라고 하는가?

)

✔ 18회 기출

🋂 🖊 아래는 빅데이터 분석 프로세스에서 데이터 분석 단계 중 어떤 것에 대한 설명인가?

분석용 데이터를 이용한 가설 설정을 통하여 통계모델을 만들거나 기계학습을 이용한 데이 터의 분류, 예측, 군집 등의 기능을 수행하는 모델을 만드는 과정

✔ 19회 기출

43 아래는 여러 분석 방법론 중 하나에 대한 설명이다. () 안에 적절한 용어는 ?

)모델은 반복을 통하여 점증적으로 개발하는 방법으로 처음 시도하는 프로젝트에 적용이 용이하지만, 반복에 대한 관리 체계를 효과적으로 갖추지 못한 경우 복잡도가 상승 하여 프로젝트 진행이 어려울 수 있다.

(

)

)

✔ 17회 기출

🙎 🔾 소프트웨어와 시스템공학의 역량 숙성도를 측정하기 위한 모델로 소프트웨어 품질보증과 시스템 엔 지니어링 분야의 품질보증 기술을 통합하여 개발된 평가모델로 1~5단계로 구성된 성숙도 모델은?

✔ 22회 기출



50 아래() 안에 공통적으로 들어갈 용어는 ?

기업 및 공공기관에서는 시스템의 중장기 로드맵을 정의하기 위한 ()을(를) 수행한다.)은(는) 정보기술 또는 정보시스템을 전략적으로 활용하기 위하여 조직 내·외부 환경 을 분석하여 기회나 문제점을 도출하고 사용자의 요구사항을 분석하여 시스템 구축 우선순 위를 결정하는 등 중장기 마스터플랜을 수립하는 절차이다.

MEMO	
,	
	8
	甘学

2과목 Ⅰ 데이터 분석 기획

정답 및 해설

91	4	11	1	21	1	31	2	41	ㄱ: 과제수행 ㄴ,ㄷ: 팀구성, 분석과제 실행, 분석과제 진행관리
02	3	12	1	22	2	32	1	42	ㄱ:워크플로우 ㄴ:사전영향 평가
03	3	13	3	23	2	33	4	43	ㄱ:업무 ㄴ:제품 ㄷ:지원 인프라
04	2	14	4	24	2	34	4	44	데이터준비
05	1	15	1	25	1	35	2	45	프레이밍효과
06	2	16	4	26	1	36	2	46	하향식 접근 방식
07	4	17	3	27	4	37	시스템	47	모델링
80	1	18	1	28	2	38	사전 영향 평가	48	나선형 모델
0.9	2	19	2	29	3	39	문제 정의	49	능력 성숙도 통합모델 (CMMI, Capability Maturity Model Integration)
10	2	20	3	30	3	40	의사 코드	50	ISP(Information Strategy Planning)

- 01. 분석대상은 명확하지만 분석방식이 명확하지 않은 경우는 Solution에 해당한다. (정답:④)
- 02. 분석 대상이 명확하게 무엇인지 모르는 경우에는 기존분석 방식을 활용하여 통찰을 도출 해냄으로써 문제의 도출 및 해결에 기여하거나 발견 접근법으로 분석 대상 자체를 새롭게 도출할 수 있다. (정답:③)
- 03. 복잡하고 정교한 모형은 분석 기획 고려사항 중 장애요소에 해당하지 않는다. (정답:③)
- 04. 분석 기획시 고려사항은 다음과 같다. (정답:②)
 - (1). 분석의 기본이 되는 데이터에 대한 고려가 필요하다.
 - (2). 분석을 통해서 가치가 창출될 수 있는 적절한 활용방안과 활용 가능한 유즈케이스의 탐색이 필요하다.
 - (3). 분석을 수행함에 있어서 발생하는 장애요소들에 대한 사전 계획 수립이 필요하다.
- 05. 정형 데이터의 유형은 ERP, CRM, SCM 등 정보시스템이며 반정형 데이터의 유형은 로그 데이터, 모바일 데이터, 센싱 데이터이다. 비정형 데이터의 유형은 영상, 음성 문자 등이다. (정답: ①)
- 06. 폭포수 모델(Waterfall Model)은 단계를 순차적으로 진행하는 방법으로, 이전 단계가 완료 되어야 다음 단계로 진행될수 있으며 문제가 발견되며 피드백 과정이 수행되기도 한다. (정답:②)

- 07. CRISP DM 방법론의 모델링 단계에서 수행하는 태스크(task)는 모델링 기법 선택, 모델 테스트 계획 설계, 모델 작성, 모델 평가가 있다. (정답: ④)
- 08. 빅데이터 분석 방법론의 분석기획 단계는 '프로젝트 범위 설정 → 데이터 분석 프로젝트 정의 → 프로젝트 수행 계획 수립 → 데이터 분석 위험 식별' 순서이다. (정답:①)
- 09. 데이터 분석 단계를 수행하는 과정에서 추가적인 데이터 확보가 필요한 경우 데이터 준비 단계로 피드백하여 단계를 반복하여 진행한다. (정답:②)
- 10, 기업에서 데이터에 기반한 의사결정을 방해하는 요소는 고정 관념, 편향된 생각, 프레이밍 효과이다. (정답 : ②)
- 11. 채널은 기업이 고객세그먼트에게 가치를 제안하기 위해 커뮤니케이션을 하고 상품이나 서비스를 전달하는 방법을 의미한다. 커뮤니케이션, 물류, 판매채널 등 기업과 고객의 인터페이스 전반이 바로 채널이다. 유통 채널을 공급하는 것은 채널영역이 아니다. (정답:①)
- 12. 기업 내·외부 환경을 포괄하고 있는 비즈니스 모델이라는 특을 활용하여 비즈니스 모델 캔버스의 9가자 블록을 단순화하여 업무, 제품, 고객 단위로 문제를 발굴하고 이를 관리하는 두가지의 영역인 규제와 감사 영역과 지원 인프라 영역에 대한 기회를 추가로 도출하는 작업을 수행한다. 비즈니스 모델 캔버스 해당 내용을 번호 순서대로 나열하면 업무, 제품, 고객, 규제&감사, 지원인프라 순서이다. (정답:①)
- 13. 데이터 타당성에 대해서는 데이터 존재 여부, 분석 시스템 환경, 분석역량에 대한 검토가 필요하다. 문제발생 포인트에 대한 확보는 중요하지 않다. (정답:③)
- 14. 문제 탐색 단계에서 현재의 비즈니스 모델 및 유사·동종사례 탐색을 통해서 도출한 분석 기회들을 구체적인 과제로 만들기 전에 분석 유즈케이스로 표기하는 것이 필요하다. (정답:④)
- 15. 분석과제 정의서는 분석별로 필요한 소스 데이터, 분석방법, 데이터 입수 및 분석의 난이도, 분석 수행주기, 분석결과에 대한 검증 오너십, 상세 분석 과정 등을 정의한다. (정답:①)
- 16. 세부이행 계획수립에서 '데이터 분석체계의 특성을 고려하여 세부적인 일정계획을 수립해야한다.'의 내용처럼 프로젝트의 세부 일정계획도 데이터 분석 체계를 고려하여 작성해야한다. (정답: ④)
- 17. 데이터 분석 프로젝트에서는 분석 범위가 빈번하게 변경되므로 분석 프로젝트 관리에서의 일정계획보다 더 많은 시간이 소요될 수 있다. 따라서 Time Boxing 기법과 같은 방법으로 일정관리를 진행하는 것이 필요하다. (정답:③)
- 18. 분석 프로젝트 관리방안에서 시간관리는 프로젝트의 활동 일정을 수립하고 일정 통제의 진척상황을 관찰하는데 요구되는 프로세스이다. (정답:①)
- 19. '분석 준비도'의 분석 데이터의 진단 항목은 분석을 위한 데이터 충분성, 신뢰성, 적시성, 비구조적 데이터 관리, 외부 데이터 활용 체계. 기준 데이터 관리(MDM)이 있다. (정답:②)
- 20. 투자비용 요소 또는 난이도를 평가하는 요소는 Volume, Variety, Velocity이다. (정답:③)

- 21. 사분면 영역에서 난이도와 시급성을 모두 고려할 때 가장 우선적인 분석 과제 적용이 필요한 영역은 난이도 : 쉬움, 시급성 : 현재를 나타내는 3사분면이다. (정답:①)
- 22. 과제 중심적인 접근 방식의 특징에는 Speed & Test, Quick-Win, Problem Solving이 해당하며 Accuracy & Deploy 는 장기적인 마스터 플랜 방식에 해당하는 내용이다. (정답:②)
- 23. 데이터 분석 준비 프레임워크에서 분석 업무 파악 영역에는 발생한 사실 분석 업무, 예측 분석 업무, 시뮬레이션 분석 업무, 최적화 분석 업무, 분석 업무 정기적 개선이 있다. (정답:②)
- 24. 데이터 거버넌스 체계 중 데이터 관리 체계는 데이터 정합성 및 활용의 효율성을 위하여 표준데이터를 포함한 메타 데이터와 데이터 사전의 관리 원칙을 수립한다. 빅데이터의 경우 데이터 양의 급증으로 데이터의 생명 주기 관리 방안을 수립하지 않으면 데이터 가용성 및 관리비용 증대 문제에 직면하게 될 수 있다. (정답:②)
- 25. 집중 구조는 전사 분석 업무를 별도의 분석 전담 조직에서 담당하며 전략적 중요도에 따라 분석조직이 우선순위를 정해서 진행 가능하기 때문에 현업 업무부서의 분석 업무와 이중화/이원화 가능성이 높다. (정답:①)
- 26. 가치는 비즈니스 효과에 해당하며 크기, 다양성, 속도가 투자비용요소에 해당한다. (정답:①)
- 27. 분석 과제 로드맵은 과제의 우선순위를 고려하여 작성하되 과제별 선후관계를 감안하여 단계별 추진 내용을 정렬한다. (정답:④)
- 28. 경쟁자 확대 관점에서는 현재 수행하고 있는 사업 영역의 직접 경쟁사 및 제품·서비스뿐만 아니라 대체재와 신규 진입자 등으로 관점을 확대하여 위협이 될 수 있는 상황에 대한 분석 기회 발굴의 폭을 넓혀서 탐색한다. (정답:②)
- 29. 단순히 대용량 데이터를 수집, 축적하는 것보다는 어떤 목적으로 어떤 데이터를 어떻게 분석에 활용할 것인가가 더욱 중요하다. 그리고 빅데이터 분석에서 품질관리도 중요하지만, 데이터 수명주기 관리방안을 수립하지 않으면 데이터 가용성 및 관리비용 증대 문제에 직면하게 될 수 있다. (정답:③)
- 30. 빅데이터 등장으로 비즈니스 영역에서는 분석 조직 및 인력에 대한 지속적인 교육과 훈련을 실시하고 사실 기반(Fact-Based) 의사결정을 할 수 있는 문화를 정착시키고자 했다. 또, 분석 교육의 목표는 단순한 툴 교육이 아닌 분석역량을 확보하고 강화하는 것에 초점을 맞추어 진행되어야 한다. (정답:③)
- 31. 분석과제 중에 발생된 시사점과 분석 결과물이 풀(pool)로 관리하고 공유된다. 확정된 분석과제는 풀(pool)로 관리하지 않는다. (정답 : ②)
- 32. 고급분석 기법은 기존에 행해졌던 데이터 분석에서도 활용되었다. (정답:①)
- 33. Self Service Anaytics에서는 분산처리는 지원하지 않는다. (정답:④)
- 34. Servitization은 제품과 서비스의 결합(Product Servitization), 서비스의 상품화(Service Productization), 그리고 기존 서비스와 신규 서비스의 결합 현상을 포괄하는 개념이다. ④번은 Servitization의 설명이 아니다. (정답:④)

- 35. 분석 방법론 기획단계에서 프로젝트 위험 대응 계획을 수립할 때 예상되는 위험에 대해 회피(Avoid), 전이(Transfer), 완화(Mitigate), 수용(Accept)으로 구분하여 위험관리 계획서를 작성한다. (정답 : ②)
- 36. 프로토타이핑 접근법은 신속하게 해결책이나 모형을 제시함으로써 문제를 좀 더 명확하게 인식하고 필요한 데이터를 식별하여 구체화할 수 있게 하는 유용한 상향식 접근 방법이다. 나머지 항목은 하향식 접근 방법에 해당하는 내용이다. (정답:②)
- 37. 정답: 시스템
- 38. 정답: 사전 영향 평가
- 39. 정답: 문제 정의
- 40. 정답: 의사 코드
- 41. 정답: 그 과제수행 ㄴ, ㄷ 팀구성, 분석과제 실행, 분석과제 진행관리
- 42. 정답: ¬ 워크플로우, ∟ 사전영향 평가
- 43. 정답: ㄱ 업무, ㄴ 제품, ㄷ 지원 인프라
- 44. 정답:데이터준비
- 45. 정답: 프레이밍효과
- 46. 정답 : 하향식 접근 방식
- 47. 정답:모델링
- 48. 정답: 나선형 모델
- 49. 정답 : 능력 성숙도 통합모델(CMMI, Capability Maturity Model Integration)
- 50. 정답: ISP (Information Strategy Planning)