

## 제 2과목 : 데이터 분석 기획

### 01장. 데이터 분석 기획의 이해

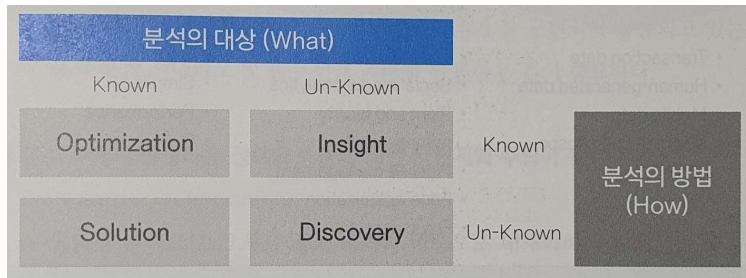
#### 분석기획

: 실제분석을 수행하기 앞서 분석할 가제를 정의하고, 의도했던 결과를 도출할 수 있도록 이를 적절하게 관리할 수 있는 방안을 사전에 계획하는 일련의 작업

- 분석과제 및 프로젝트를 직접 수행하는 것은 아니지만, 어떠한 목표를 달성하기 위해 어떠한 데이터를 가지고 어떤 방식으로 수행할지 일련의 계획을 수립하는 작업

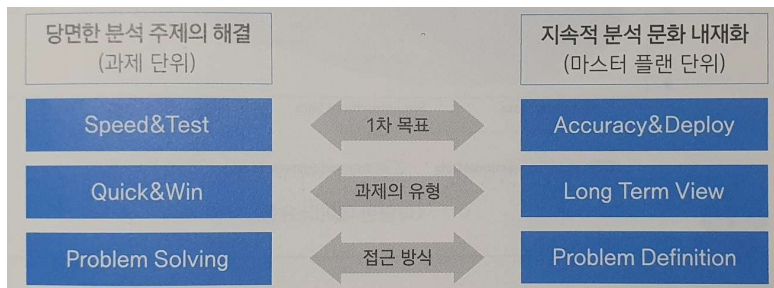
데이터 사이언티스트는 **수학/통계학적 지식** 및 **정보기술(IT)** 과 해당 **비즈니스에 대한 이해와 전문성**을 포함한 3가지 영역에 대한 고른 역량과 시각이 요구

분석의 대상과 방법에 따라 분석이 나누어진다.



목표 시점 별 분석 기획 방안

1. 과제 중심적인 접근 방식
2. 장기적인 마스터 플랜 방식



#### 분석 기획시 고려사항

1. 분석의 기본인 **가용 데이터(Available Data)**에 대한 고려가 필요
  - 분석을 위한 데이터 확보가 우선이며, 데이터의 유형에 따라 분석 방법이 달라 유형에 대한 분석이 선행적으로 이루어져야함
2. 분석을 통해 가치가 창출될 수 있는 **적절한 활용방안과 유즈케이스(Proper Business Use Case)** 탐색이 필요
3. 분석 수행시 발생하는 장애요소들에 대한 사전계획 수립이 필요(Low Barrier Of Execution)

#### ※참고

- 정형 데이터는 데이터 자체로 분석 가능
- 반정형은 메타정보를 활용해야 해석 가능
- 데이터 자체로 분석이 안되어서 특정한 처리 프로세스를 거쳐서 분석

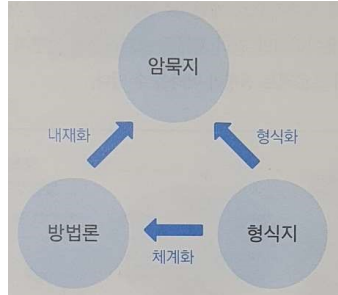
#### 분석 방법론

- 데이터 분석이 효과적으로 기업 내에 정착하기 위해서는 이를 **체계화한 절차와 방법**이 정리된 **데이터 분석 방법론**의 수립이 필요하다.
- 방법론은 상세한 **절차(Procedures)**, **방법(METHODS)**, **도구와 기법(Tool&Techniques)**, **템플릿과 산출물(Templates&Outputs)**로 구성되어 어느 정도의 지식만 있으면 활용이 가능해야 한다.

기업의 합리적 의사결정을 가로막는 장애요소

1. 고정관념(Stereotype)
  2. 편향된 생각(Bias)
  3. 프레임링 효과(Framing Effect) 등
- > 개인에 판단이나 선택이 달라질 수 있는 현상을 뜻함

## 방법론의 생성 과정



참고				
구분	의미	예	특징	상호작용
암묵지	학습과 경험을 통해 개인에게 체화되어 있지만 겉으로 드러나지 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

## 방법론의 적용 업무 특성에 따른 모델

1. 폭포수 모델
2. 프로토타입 모델
3. 나선형 모델 : 처음 시도하는 프로젝트에 유리하나 관리 못하면 복잡도 상승

## 방법론의 구성 (각 순서로)

1. 단계 -> 단계별 완료 보고서 생성
2. 태스크 -> 보고서 (물리적/논리적 단위의 품질검토 항목이 됨)
3. 스텝 -> 보고서 구성요소 산출

## 가. KDD(Knowledge Discovery in Databases) 분석 방법론

- 프로파일링 기술을 기반으로 데이터로부터 통계적 패턴이나 지식을 찾기 위해 활용할 수 있도록 체계적으로 정리한 데이터 마이닝 프로세스

## KDD 분석 절차

1. 데이터셋 선택(Selection)
  - 데이터셋 선택에 앞서 분석 대상의 **비즈니스 도메인**에 대한 이해와 **프로젝트 목표 설정**이 필수이며 데이터베이스 또는 원시 데이터에서 분석에 필요한 데이터 선택
  - 데이터마이닝에 필요한 **목표데이터(Target Data)**를 구성하여 분석에 활용
2. 데이터 전처리(Processing)
  - 추출된 분석 대상용 데이터 셋에 포함되어 있는 **노이즈와 이상치(Outlier)**, 결측

**치(Missing Value)**를 식별하고 필요시 제거하거나 의미있는 데이터로 재처리하여 데이터셋을 정제하는 과정

## 3. 데이터 변환 (Transformation)

- 정제된 데이터에 분석 목적에 맞게 변수를 생성, 선택하고 데이터의 차원을 축소하여 효율적으로 데이터마이닝을 할 수 있도록 데이터에 변경하는 단계
- 데이터마이닝 프로세스를 진행하기 위해 학습용 데이터와 시험용 데이터로 데이터를 분리하는 단계

## 4. 데이터 마이닝(Data Mining)

- 학습용 데이터를 이용해 분석목적에 맞는 **데이터마이닝 기법**을 선택하고, 적절한 알고리즘을 적용하여 데이터마이닝 작업을 실행하는 단계
- 필요에 따라 데이터 전처리와 데이터 변환 프로세스를 추가로 진행.

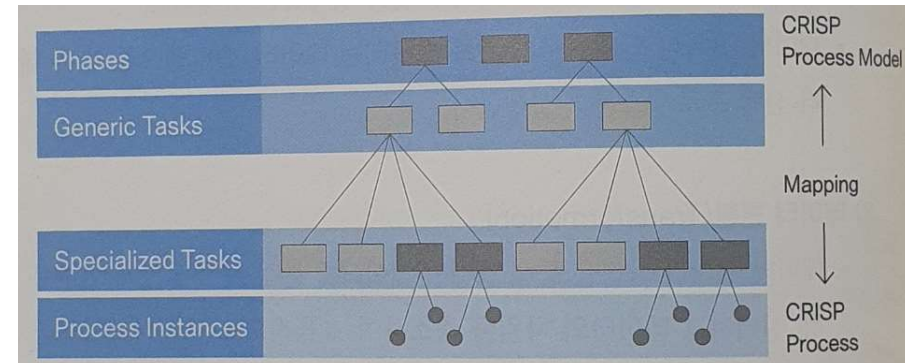
## 5. 데이터 마이닝 결과 평가(Interpretation/Evaluation)

- 데이터 마이닝 **결과에 대한 해석과 평가**, 그리고 **분석 목적과의 일치성**을 확인한다,
- 데이터마이닝을 통해 발견한 지식을 업무에 활용하기 위한 방안 마련의 단계

## 나. CRISP-DM(Cross Industry Standard Process for Data Mining)분석방법론

- 유럽연합 ESPRIT에서 있었던 프로젝트에서 시작
- **계층적 프로세스 모델**로써 4개 레벨로 구성

## CRISP-DM의 4레벨 구조



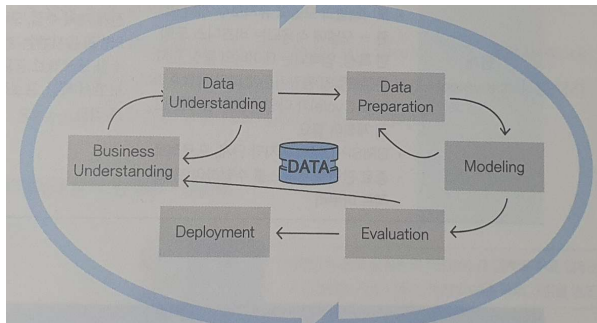
- 최상위 레벨은 여러 개의 단계(Phase)로 구성되고 각 단계는 일반화 태스크(Generic Tasks)를 포함한다. 일반화 태스크는 데이터마이닝의 단일 프로세스를 완전하게 수행하는 단위이며, 이는 다시 구체적인 수행 레벨인 세분화 태스크(Specialized Tasks)로 구성된다.

- 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-DM의 프로세스

- 6단계로 구성, 각 단계는 단방향으로 구성되어 있지 않고 단계가 ↳ 피드백을 통한 단계별 완성도를 높게 되어 있음



단계	내용	수행업무
업무이해 Business Understanding	○비즈니스 관점에서 <b>프로젝트의 목적과 요구사항</b> 을 이해하기 위한 단계 ○도메인 지식을 데이터 분석을 위한 문제 정의로 변경하고 초기 프로젝트 계획을 수립하는 단계	-업무 목적 파악 -상황 파악 -데이터마이닝 목표 설정 -프로젝트 계획 수립
데이터 이해 Data Understanding	○분석을 위한 데이터를 수집하고 <b>데이터 속성</b> 을 이해하기 위한 단계 ○데이터 품질에 대한 문제점을 식별하고 숨겨져 있는 <b>인사이트</b> 를 발견하는 단계	-초기 데이터 수집 -데이터 기술 분석 -데이터 탐색 - 데이터 품질 확인

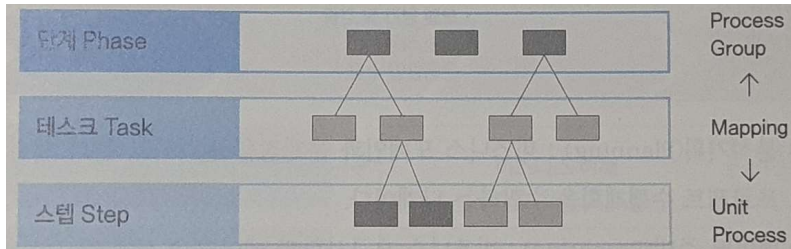
데이터 준비 Data Preparation	○분석을 위해서 수집된 데이터에서 분석 기법에 적합한 데이터를 편성하는 단계 (많은 시간이 소요될 수 있음)	-분석용 <b>데이터 셋 선택</b> -데이터 정제 -분석용 <b>데이터 셋 편성</b> -데이터 통합 -데이터 포매팅
모델링 Modeling	○다양한 모델링 기법과 알고리즘을 선택하고 모델링 과정에서 사용되는 파라미터를 최적화해 나가는 단계 ○모델링 과정에서 데이터 셋이 추가로 필요한 경우 데이터 준비 단계를 반복 수행, 모델링 결과를 테스트용 데이터 셋으로 평가하여 모델의 과적합(Overfitting) 문제를 확인	- 모델링 기법 선택 - 모델 테스트 계획 설계 - 모델 작성 - <b>모델 평가</b>
평가 Evaluation	○모델링 결과가 프로젝트 목적에 부합하는지 평가하는 단계로, 데이터마이닝 결과를 최종적으로 수용할 것인지 판단	- <b>분석결과 평가</b> - <b>모델링 과정 평가</b> - <b>모델 적용성 평가</b>
전개 Deployment	○모델링과 평가 단계를 통하여 완성된 모델을 <b>실 업무에 적용</b> 하기 위한 계획을 수립하는 단계 ○CRISP-DM의 마지막 단계, 프로젝트 종료 관련 프로세스를 수행하여 프로젝트 마무리	-전개 계획 수립 -모니터링과 유지보수 계획 수립 -프로젝트 종료 보고서 작성 -프로젝트 리뷰

### KDD와 CRISP-DM 비교

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

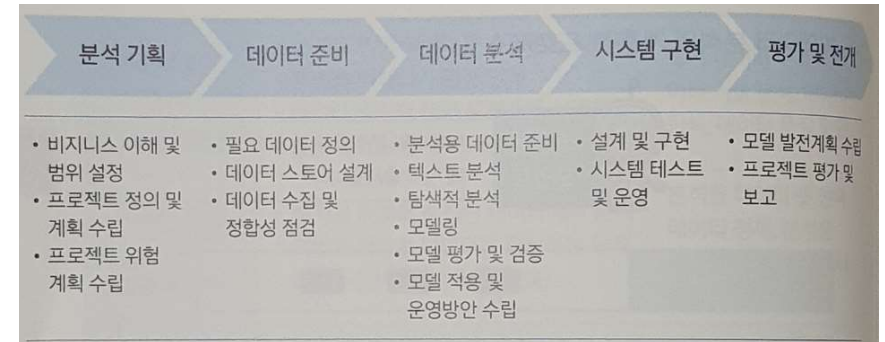
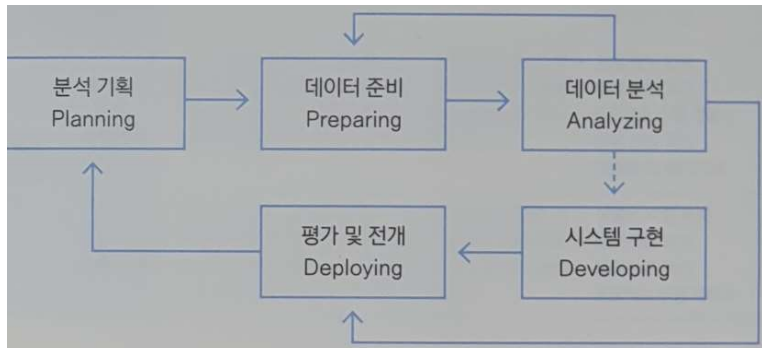
## 다. 빅데이터 분석 방법론

빅데이터 분석의 프로세스는 3개의 계층으로 구성된다



1. 단계(Phase) : 프로세스 그룹을 통해 완성된 단계별 산출물이 생성. 각 단계는 기준선(Baseline)으로 설정되어 관리되어야하며, 버전관리 등을 통하여 통제가 이루어져야 한다.
2. 태스크(Task) : 각 단계는 **여러 개의 태스크**로 구성된다. 각 태스크는 단계를 구성하는 단위활동이며, 물리적 또는 논리적 단위로 품질 검토의 항목이 될 수 있다.
3. 스텝(Step) : WBS(Work Breakdown Structure)의 워크 패키지에 해당되고 입력자료, 처리 및 도구, 출력자료로 구성된 단위 프로세스이다.

### 빅데이터 분석 방법론- 5단계



※페이지 p141~145까지 빅데이터 분석 방법론의 실제 업무가 나옴. 필히 확인!

※특히 각 단계별 세부단계의 순서를 아는 것이 좋음.

### 분석 과제 발굴 (방법론)

: 분석 과제는 풀어야 할 다양한 문제를 데이터 분석 문제로 변환한 후 관계자들이 이해하고 프로젝트로 수행할 수 있는 **과제 정의서 형태로** 도출

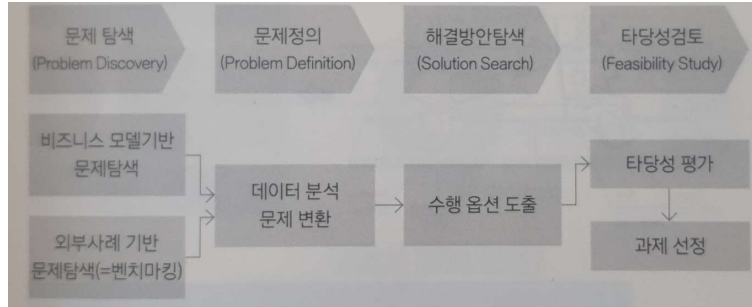
- 분석과제를 도출하기 위한 방식으로는 크게 하향식/상향식 접근으로 나뉨
- 분석의 대상이 알려져 있으면(최적화/솔루션) 하향식, 알려지지 않았으면(통찰/발견) 상향식 접근방법에 해당된다.
- 전통적인 것은 문제가 주어진 상태에서 답을 찾는 하향식. 현재는 문제 자체에 대한 변화가 심해 정확하게 정의하기가 힘들
- 하지만 실제로는 상향식/하향식이 모두 혼용되어 쓰임. 분석의 가치를 높일 수 있는 최적의 의사결정은 두 접근 방식이 상호 보완 관계에 있을 때

※디자인 사고 : 상향식과 하향식으로 각각 반복적으로 수행하는 최적의 의사 결정 방식



## 하향식 접근법 (Top-Down Approach)

순서 : 문제 탐색 -> 문제 정의 -> 해결 방안 탐색 -> 타당성 검토



1단계 - 문제 탐색 : 전체적인 관점의 기준 모델을 활용하여 빠짐없이 문제를 도출하고 식별하는 것이 중요하다.

- 세부적인 구현 및 솔루션에 초점을 맞추는 것이 아니라, **문제를 해결함으로써 발생하는 가치에** 중점을 두는 것이 중요

가) 비즈니스 모델 문제 탐색 : 기업 내·외부를 포괄한 비즈니스 모델이라는 틀을 사용해 비즈니스모델 캔버스의 9가지 블록을 단순화하여 **업무,제품,고객** 단위로 문제를 발굴하고, 이를 관리하는 두 가지 영역인 **규제&감사**와 **자원 인프라** 영역에 대한 기회를 추가로 도출하는 작업을 수행한다.

**업무(Operation)** : 제품 및 서비스를 생산하기 위한 내부 프로세스 및 주요 자원 관련 주제 도출

**제품(Product)** : 생산 및 제공하는 제품·서비스를 개선하기 위한 관련 주제 도출

**고객(Customer)** : 제품·서비스를 제공받는 사용자 및 고객, **이를 제공하는 채널**의 관점에서 주제 도출

**규제와 감사(Regulation&Audit)** : 제품 생산 및 전달과정 프로세스 중에서 발생하는 규제 및 보안의 관점에서 주제 도출

**자원 인프라(IT&Human Resources)** : 분석을 수행하는 시스템 영역 및 이를 운영·관리하는 인력의 관점에서 주제 도출

-> 다양한 관점의 접근을 통해 새로운 유형의 분석 기회 및 주제 발굴을 수행

## ※분석 기회 발굴의 범위 확장

1. 거시적 관점 (사회 기술 경제 등)
2. 경쟁자 확대 (대체자, 경쟁자, 신규 진입자)
3. 시장니즈 탐색 (고객 채널 영향자들)
4. 역량의 재해석 ( 내부역량, 파트너 네트워크)

## 나) 외부 참조 모델 탐색

: 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 브레인스토밍을 통해 빠르게 도출하는 방법이다,

- 다양한 산업( 교통, 안전 , 행정, 보건, 도시, 정치, 일자리 등) 다양

## 다) 분석 유즈 케이스 (Analytics Use Case)

: 현재의 비즈니스 모델 및 유사·동종사례 탐색을 통해 빠짐없이 도출한 분석 기회들을 구체적인 과제로 만들기 전에 분석 유즈 케이스로 표기하는 것이 필요

-> 풀어야할 문제에 대한 상세한 설명 및 해당 문제를 해결했을 때 발생하는 효과를 명시함으로써 향후 데이터 분석 문제로의 전환 및 합성 평가에 활용함

-> 업무 - 유즈케이스 - 설명 - 효과를 명시함

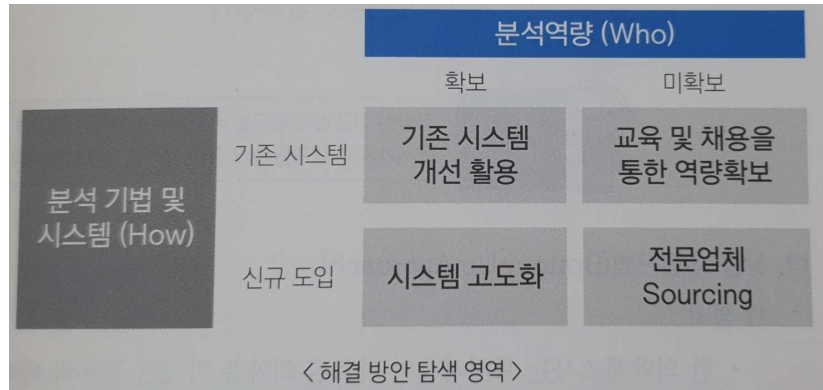
2단계 - 문제 정의 : 식별된 **비즈니스 문제를 데이터의 문제로** 변환하여 정의하는 단계이며, 앞서 수행한 문제 탐색의 단계가 무엇(What) 어떤 목적(Why)로 수행해야 하는지에 대한 관점이었다면, 본 단계에서는 이를 달성하기 위해서 필요한 데이터 및 기법(How)를 정의하기 위한 데이터 분석기 문제로 변환 수행

- 데이터 분석 문제 정의 및 요구사항 : 분석을 수행하는 당사자 뿐만 아닌 이 문제가 해결되었을 때 호용을 얻을 수 있는 최종사용자 관점에서 이루어져야 한다

- 가능한 정확하게 분석의 관점으로 문제를 재정의할 필요가 있음

3단계 - 해결방안 탐색 : 정의된 **데이터 분석 문제를 해결**하기 위한 다양한 방안이 모색

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



해당 영역에 분석역량을 보유하지 않은 경우, 전문인력이나 교육으로 확보하거나 분석 전문 업체를 이용하여 과제를 해결하는 방안에 대해 사전 검토를 수행

4단계 - 타당성 검토

- 가) 경제적 타당성 : 비용대비 편익 분석 관점의 접근이 필요
  - 나) 데이터 및 기술적 타당성
- (총 3가지 중요)

상향식 접근법 (Bottom-Up Approach)

- : 의약품제조사 가 수많은 의약품 데이터를 통해 새로운 의약품 개발
- : 기업이 보유한 다양한 원천 데이터로부터 분석을 통해 통찰력과 지식을 얻는 상향식 접근방법. -> 가치있는 모든 문제를 도출하는 일련의 과정 (빅데이터 -> 분석 -> 문제 정의)

기존 하향식 접근법의 한계를 극복하기 위한 분석 방법론

- 기존 접근 방법인 논리적인 단계별 접근은 문제의 구조가 분명하고, 해결하기 위한 데이터 분석가 및 의사결정자에게 주어져 있다고 가정하므로 솔루션 도출에는 유효하나 문제 탐색에는 한계가 있음
- 따라서 최근 복잡하고 다양한 환경에서 발생하는 문제에 부적합
- 디자인사고 접근법을 통해 극복
- 사물을 있는 그대로 인식하는 What 관점에서 보아야함
- 관찰하여 대상을 본질적으로 이해하는 방식의 접근 수형

지도 학습과 비지도 학습

1. 비지도 학습(Unsupervised Learning)

- 일반적으로 상향식 접근 방식의 데이터 분석은 비지도 학습에 해당
- 비지도 학습은 데이터 분석의 목적이 명확히정의된 형태의 특정 필드의 값을 구하는 것이 아닌 데이터 자체의 **결합, 연관성, 유사성** 등을 중심으로 데이터의 상태를 표현하는 것이다.
- 예시) 장비가니 분석, 군집분석, 기술 통계 및 프로 파일링

2. 지도 학습 (Supervised Learning)

- 지도 학습은 명확한 목적 하에 데이터 분석을 실시하는 것이다.
- 분류, 추측, 예측, 최적화를 통해 사용자의 주도하에 분석을 실시하고 지식 도출이 목적이다.

※ 통계적 분석은 인과관계 분석을 위해 가설설정 후 가설검정을 통해 문제를 해결 그러나 빅데이터 환경에서는 인과관계 분석뿐만 아니라 상관관계 분석 또는 연관 분석을 통해 다양한 문제 해결에 도움을 받을 수 있다

다량의 데이터 분석을 통해 ‘왜’ 그러한 일이 발생하는지 역으로 추적하면서 문제를 도출하거나 재정의 할 수 있는 것이 상향식 접근법이다

### 시행착오를 통한 문제 해결

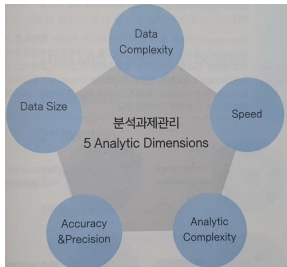
- 프로토타이핑 접근법은 사용자가 요구사항이나 데이터를 정확히 규정하기 어렵고 데이터 소스도 명확히 파악하기 어려운 상황에서 일단 분석을 시도해보고 그 결과를 확인해가면서 반복적으로 개선해나가는 방법.
- 하향식은 문제가 정형화 되어 있고 데이터가 보통 완벽할 때 적합
- 프로토타이핑 접근법은 완전하지 못해도 신속히 해결책 및 모형 제시함으로써 문제를 정확히 인식하고 필요한 데이터를 식별하는 유용한 상향식 접근 방식
- 프로토타이핑 접근법의 프로세스 : 가설의 생성 -> 디자인에 대한 실험 -> 실제 환경에서의 테스트 -> 테스트 결과에서의 통찰도출 및 가설 확인

### 빅데이터 분석 환경에서 프로토타이핑의 필요성

1. 문제에 대한 인식 수준 : 문제의 정의가 불명확 -> 문제 이해 및 구체화
2. 필요 데이터 존재 여부의 불확실성 : 데이터가 온전치 않아 수집할 것인지 대체할 것인지에 대한 사용자와 분석가 간의 반복적이고 순환적인 협의 과정 필요
3. 데이터 사용 목적의 가변성 : 데이터의 가치는 정해진 수집 목적에 따라 확정되는 것이 아님.->기존의 데이터 정의를 재검토하여 **데이터의 사용 목적과 범위 확대**

### 분석 프로젝트 관리 방안

- 과제 형태로 도출된 분석 기회는 프로젝트를 통해 그 가치를 증명하고 목표를 달성해야한다.
- 분석 프로젝트는 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



Data Size, Data Complexity, Speed,  
Analytic Complexity, Accuracy&Precision

영역	내용
Data Size	<b>분석하고자하는 데이터의 양</b> 을 고려한 관리 방안 수립 필요. 하둡에서 엄청난 양의 데이터로 분석하는 것과 기존 정형 데이터베이스로 분석하는 것에 있어 관리방식은 차이가 날 수 밖에 없다
Data Complexity	BI 프로젝트처럼 정형 데이터가 분석 마트로 구성되어 있는 상태에서 분석하는 것과 달리 텍스트, 오디오, 비디오 등의 비정형 데이터 및 다양한 시스템에 산재되어 있는 원천 데이터들을 통합해서 분석할때는, 해당 <b>데이터가 잘 적용될 수 있는</b> 분석 모델의 선정등의 고려가 필요하다
Speed	분석결과가 도출되었을 때 이를 활용하는 시나리오 측면에서의 속도를 고려해야한다. 일/주 단위 실적 같은 경우에는 배치(Batch)형태로 작업되어도 무방하나 실시간으로 사기(Fraud)를 탐지하거나 고객에게 개인화된 상품·서비스를 추천하는 경우에는 분석 모델의 적용 및 계산이 실시간으로 수행되어야 하므로, 프로젝트 수행 시 <b>분석 모델의 성능 및 속도를 고려한 개발 및 테스트</b> 가 수행되어야 한다.
Analytic Complexity	분석 모델의 정확도와 복잡도는 트레이드 오프 관계가 존재. 따라서 기준점을 사전에 정의되어야함. 해석이 가능하면서도 정확도를 올릴 수 있는 최적모형을 찾는 방안 모색
Accuracy & Precision	정확도는 예측값과 실제값의 차이의 정도를 뜻하고, Precision은 모델을 지속적으로 반복했을 때 편차의 수준으로써 일관적으로 동일한 결과를 제시한다는 것을 의미한다. 분석의 활용적인 측면에서는 Accuracy가 안정성 측면에서는 Precision이 중요하다.

데이터의 원천을 다루는 데이터 영역과 결과를 활용할 비즈니스 영역의 중간에서 분석 모델을 통한 조율을 수행하는 조정자 역할의 핵심

### 분석 프로젝트의 관리 방안

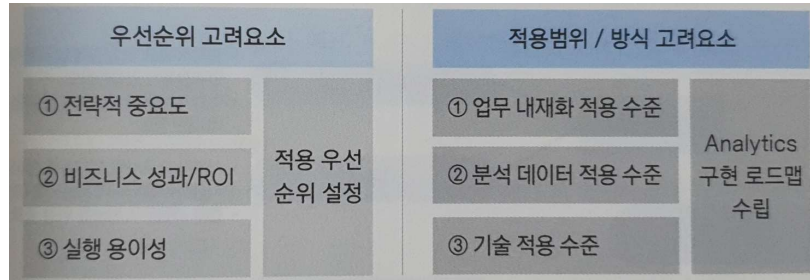
: 범위, 시간, 원가, 품질, 통합, 조달, 자원, 리스크, 의사소통, 이해관계자

이러한 사항들을 체크리스트로 관리하여 발생할 이슈 및 리스크를 숙지하고 미연에 방지할 필요가 있다

분석 마스터 플랜(제일 가장 기본이 되는 플랜, 또는 그런 설계) - 기본계획

: 데이터 분석을 구현하기 위한 다양한 기준들을 통해 데이터 분석을 위한 로드맵을 수립하기 위해 작성해야한다.

- 데이터 기반 구축을 위해서 분석 과제를 대상으로 전략적 중요도, 비즈니스 성과 및 ROI, 분석 과제의 실행 용이성 등 다양한 기준을 고려해 **적용 우선순위를** 설정



※ISP(Information Strategy Planning) : 정보기술 또는 정보시스템을 전략적으로 활용하기 위하여 조직 **내·외부 환경을 분석**하여 기회나 문제점을 도출하고 사용자의 요구사항을 분석하여 **시스템 구축 우선순위를 결정**하는 등 중·장기 마스터 플랜을 수립하는 절차.

수행 과제 도출 및 우선순위 평가

가. 우선순위 평가 방법 및 절차

- 우선순위 평가의 경우 정의된 데이터 과제에 대한 실행 순서를 정함

: **분석 과제 도출 -> 우선순위 평가 -> 우선순위 정렬**(선·후 관계에 따라 순위 조정)

나. 일반적인 IT 프로젝트의 우선순위 평가 예시

- 전략적 중요도 / 실행 용이성 등

다. ROI 관점에서 빅데이터의 핵심 특징

3V(Volume, Variety, Velocity) -----> 투자비용 요소로 난이도를 결정

1V(Value) -----> 비즈니스 요소로 시급성을 결정

(총 4V)

1. 투자비용(Investment) 요소

- 크기 : 대용량 데이터를 저장·관리하는 데 새로운 투자 필요

- 다양성 : 다양한 종류의 데이터를 입수하는데 투자가 필요

- 속도 : 데이터 생성 및 처리 속도를 빠르게 가공·분석하는데 기술 필요

2. 비즈니스 효과(Return) 요소

- 가치 : 분석 결과를 활용하거나 실질적인 실행을 통해 얻게 되는 비즈니스 효과 측면의 요소로, 기업데이터 분석을 통해 추구하거나 달성하고자 하는 목표 가치를 의미

라. 데이터 분석과제 추진시 고려해야 하는 우선순위 평가 기준

1. 시급성 : 전략적 중요도와 목표가치에 부합하는지에 따른 시급성이 가장 중요한 기준이다. 시급성의 판단 기준은 전략적 중요도가 핵심이며, 이는 현재의 관점에서 전략적 가치를 둘 것인지, 미래의 중장기적 관점에 전략적인 가치를 둘 것인 지를 고려하고, 분석 과제의 목표가치 (KPI)를 함께 고려하여 시급성 여부 판단

2. 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 판단 기준으로서, 데이터 분석의 적합성 여부를 본다.

마. 포트폴리오 4분면 분석을 통한 과제 우선순위 선정

난이도 (쉬움 어려움), 시급성(현재, 미래)를 따져 현재 급하고 어려운 과제가 제일 우선순위가 된다.

- 물론 우선순위를 시급성에 기준으로 두면 3->4->2 사분면 순서이고,

- 우선순위 기준을 난이도에 두면 3-> 1->2 사분면 순서이다.

※만약 데이터의 양이 많은 1사분면 과제를 데이터의 양을 줄여 3사분면으로 이동시킬 수 있음



## 이행계획 수립

### 1. 로드맵 수립

- 분석과제에 대한 포트폴리오 사분면 분석을 통해 과제의 1차적 우선순위 결정
- 분석 과제별 적용범위 및 방식을 고려하여 최종적인 실행 우선순위를 결정한 후 단계적 구현 로드맵을 수립
- 단계별로 추진하고자 하는 목표 정의
- 추진 과제별 선·후행 관계를 고려하여 단계별 추진 내용 정렬

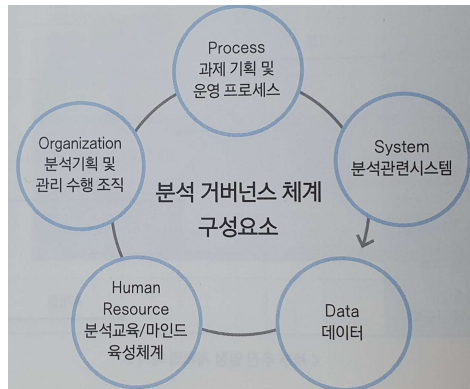
### 2. 세부 이행 계획 수립

- 데이터 분석 체계는 고전적인 폭포수 모델도 있으나 반복적인 정련 과정을 통해 프로젝트의 완성도를 높이는 방식을 주로 사용
- 보통 데이터 수집/준비는 순차적으로 진행하고 모델링 단계를 반복적으로 수행

## 분석 거버넌스 체계 수립

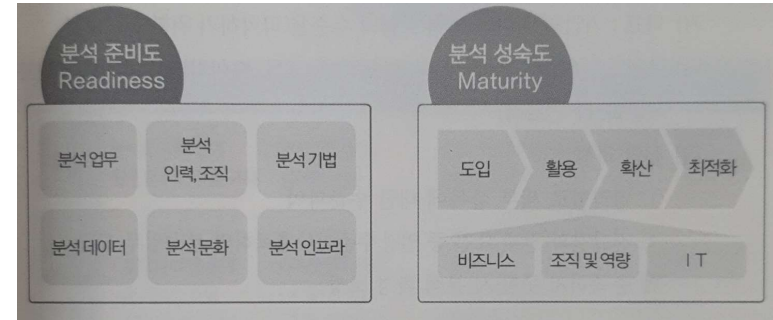
- : 기업에서 데이터를 이용한 의사결정이 강조될수록 데이터 분석과 활용을 위한 체계적인 관리가 중요해진다. 단순한 대용량 데이터를 수집·축적하는 것보다는 어떤 목적으로 어떤 데이터를 어떻게 분석에 활용할지가 더 중요하기 때문
- 또한 분석 관리체계를 수립하는 이유는 데이터 분석을 기업의 문화로 정착하고 데이터 분석업무를 지속적으로 고도화하기 위해서이다.

## 거버넌스 체계 구성요소



## 데이터 분석 수준진단

- 기업들은 데이터 분석의 도입 여부와 활용에 명확한 수준을 점결할 필요가 있음.
- 수준진단을 통해 데이터 분석 기반을 구현하기 위해 필요하고 무엇을 보완해야할지 결정
- 데이터 분석 수준 진단을 위해 **분석 준비도(Readiness)**와 **분석 성숙도 (Maturity)**



## 수준 진단 목표 2가지 (분석 준비도, 분석 성숙도)

- 1. 정의 :** 기업의 현재 분석 수준을 명확히 이해하고, 수준진단 결과를 토대로 미래의 목표 수준을 정의한다.

- 데이터 분석을 위한 수준 또는 환경이 유사업종 또는 타 경쟁사에 비해 어느 수준이고, 어떤 영역에 선택, 집중 및 어떤 관점을 보완해야 하는지 등 개선방안 도출

## 2. 분석 준비도

- : 기업의 데이터 분석 도입의 수준을 파악하기 위한 진단 방법
- 6가지로 구성(분석업무파악, 인력 및 조직, 분석기법, 분석 데이터, 분석문화, IT 인프라)
  - 진단과정

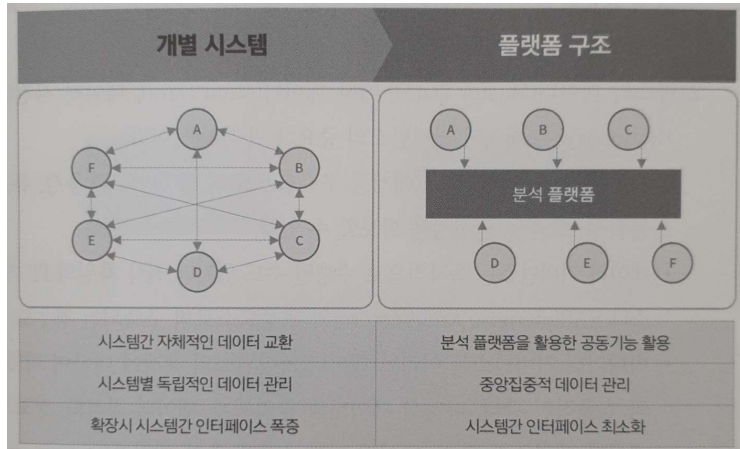
- 영역별 세부 항목에 대한 수준 파악
- 진단결과 전체 요건 중 일정 수준이상 충족하면 분석업무 도입
- 충족하지 못할 시 분석 환경 조성

## 3. 분석 성숙도 모델

- 조직의 성숙도 평가 도구 : CMMI(Capability Maturity Model Integration) 모델
- 성숙도 수준분류 : 도입단계, 활용단계, 확산단계, 최적화단계
- 비즈니스 부문, 조직·역량부문, IT부문

분석 관점에서의 사분면 분석

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



분석지원 인프라 방안 수립

: 분석과제 단위별로 별도의 분석시스템을 구축하는 경우, 관리의 복잡도 및 비용의 증대라는 부작용이 나타남 따라서 분석마스터 플랜을 기획하는 단계부터 장기적이고 안정적인 확장성을 고려한 **플랫폼 구조를 도입**하는 것이 적절

## 플랫폼

- 단순한 분석 응용프로그램뿐만 아니라 분석 서비스를 위한 응용 프로그램이 실행될 수 있는 기초를 이루는 컴퓨터 시스템을 의미한다
- 일반적으로 하드웨어에 탑재되어 데이터 분석에 필요한 프로그래밍 환경과 실행 및 서비스 환경을 제공하는 역할 수행
- 분석플랫폼이 구성되어 있는 경우에는 새로운 데이터 분석 니즈가 존재할 경우 개별적인 분석 시스템을 추가하는 것이 아닌 서비스를 추가적으로 제공하는 방식으로 확장성 상승

## 데이터 거버넌스 체계 수립 ★★★

- : 전사 차원의 모든 데이터에 대하여 정책 및 지침, 표준화, 운영조직 및 책임 등의 표준화된 관리체계를 수립하고 운영을 위한 프레임워크 및 저장소(Repository)를 구축하는 것
- 마스터 데이터, 메타 데이터, 데이터 사전은 데이터 거버넌스의 중요한 관리 대상

- 기업은 거버넌스 체계를 구축함으로써 데이터의 가용성, 유용성, 통합성, 보안성, 안정성 확보할 수 있다.
- 데이터 거버넌스는 독자적으로도 가능하나 전사 차원의 IT 거버넌스나 EA의 구성요소로 구축될 수 있음
- 빅데이터 거버넌스는 이러한 데이터 거버넌스의 체계의 대하여 빅데이터의 효율적인 관리, 다양한 데이터의 관리체계, 데이터 최적화, 정보보호, 데이터 생명주기 관리, 데이터 카테고리별 관리 책임자 지정 등을 포함한다

## 데이터 거버넌스 구성요소

### 1. 원칙(Principle)

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

### 2. 조직(Organization)

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

### 3. 프로세스(Process)

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

※이 3가지를 유기적으로 조합하고 효과적으로 관리하여, 데이터를 비즈니스 목적에 부합하도록 하고 최적의 정보 서비스 제공할 수 있도록 함

## 데이터 거버넌스 체계

: 데이터 표준화, 데이터 관리 체계, 데이터 저장소 관리, 표준화 활동

### 1. 데이터 표준화

- 데이터 표준 용어 설정, 명명 규칙, 데이터 사전 구축 등의 업무로 구성
- 반드시 상호 검증이 가능하도록 점검 프로세스 포함해야함

### 2. 데이터 관리 체계

- 데이터 정합성 및 활용의 효율성을 위하여 표준 데이터를 포함한 메타 데이터와 데이터 사전의 관리 원칙을 수립한다
- 수립된 원칙에 근거하여 항목별 상세한 프로세스를 만들고 관리와 운영을 위한 담당자 및 조직별 역할과 책임을 상세하게 준비한다.
- 빅데이터는 데이터 생명 주기 관리방안을 수립해야 데이터 가용성 및 관리비용 문제X

### 3. 데이터 저장소 관리

- 메타데이터 및 표준 데이터를 관리하기 위한 전사 차원의 저장소 구성
- 저장소는 데이터 관리 체계 지원을 위한 워크플로우 및 관리용 응용 소프트웨어를 지원하고 관리 대상 시스템과의 인터페이스를 통한 통제가 이루어져야 한다.
- 데이터 구조 변경에 따른 사전 영향 평가도 수행되어야 효율적 활용 가능

### 4. 표준화 활동

- 데이터 거버넌스 체계를 구축한 후 표준 준수 여부를 주기적으로 점검하고 모니터링 실시
- 거버넌스의 조직 내 안정적인 정착을 위한 지속적인 변화 관리 및 주기적 교육 진행
- 지속적인 데이터 표준화 개선 활동을 통해 실용성 상승

### 데이터 조직 및 인력방안 수집

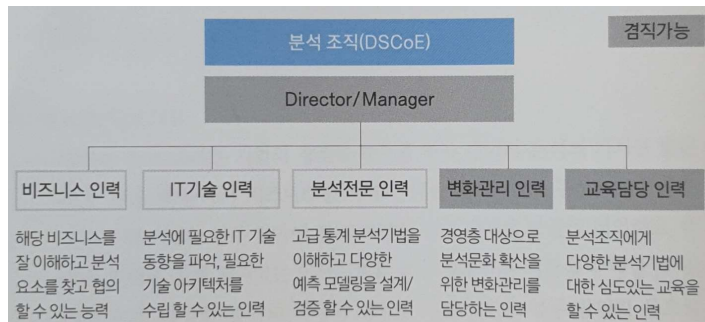
: 빅데이터 등장에 따라 기업 비즈니스에 큰 변화가 나타나고 있음 이러한 비즈니스 변화를 인식하고 기업의 차별화된 경쟁력을 확보하는 수단으로써 데이터 분석·활용등을 위한 전문 분석 조직의 필요성 제기

### 분석 조직의 개요

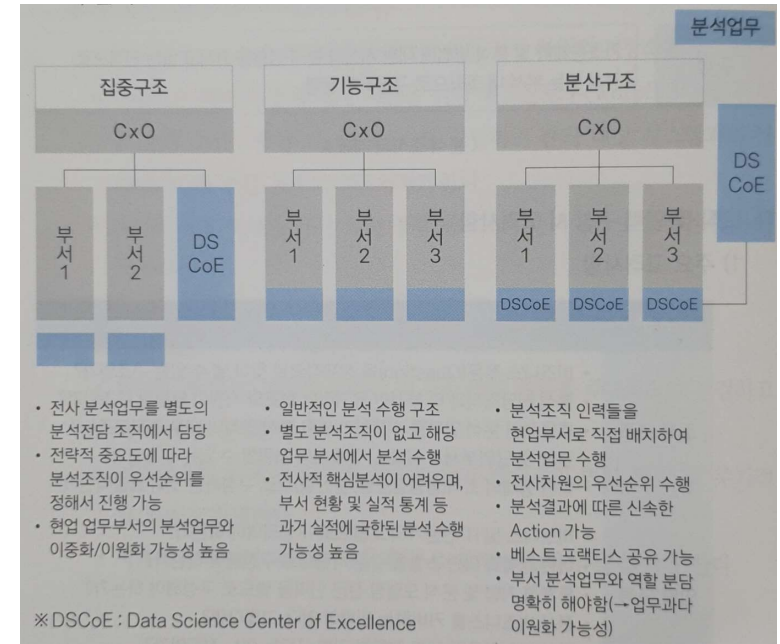
1. 목표 : 기업의 경쟁력 확보를 위하여 비즈니스 질문과 이에 부합하는 가치를 찾고 비즈니스를 최적화하는 것
2. 역할 : 전자 및 부서의 분석 업무를 발굴하고 전문적 기법과 분석 도구를 활용하여 기업 내 존재하는 빅데이터 속에서 Insight를 전파하고 이를 Action화 하는 것
3. 구성 : 기초통계학 및 분석 방법에 대한 지식과 분석 경험을 가지고 있는 인력으로 전사 또는 부서 내 조직으로 구성하여 운영

조직 및 인력 구성 시 고려사항 : 조직 구조 와 인력 구성

분석 조직의 인력 구성 : 전문역량을 갖춘 인재들을 모아 조직을 구성해 조직력 극대화



### 분석을 위한 3가지 조직 구조



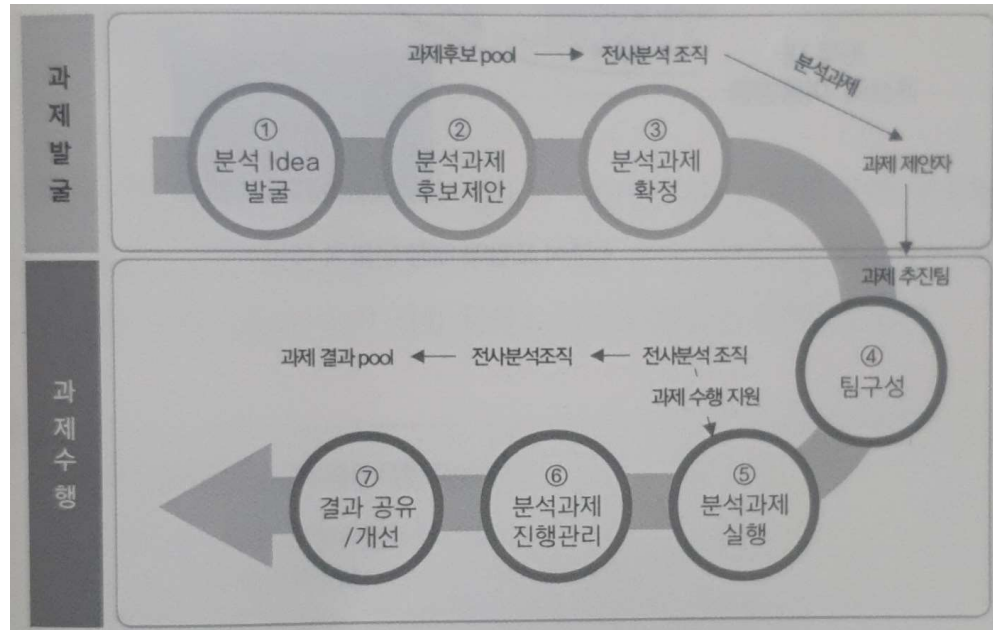
### 분석 과제 관리 프로세스 수립

현황 : 분석 마스터 플랜이 수립되고 초기 데이터 분석 과제가 성공저금로 수행되는 경우, 지속적인 분석 니즈 및 기회가 분석 과제 형태로 도출될 수 있음. 이런 과정에서 분석 과제의 기획 및 운영이므로 이를 체계적으로 관리하기 위한 프로세스 수립 필요

### 과제 관리 프로세스

- 1) 과제 발굴 : 개별 조직 혹은 개인이 도출한 분석 아이디어를 발굴하고 이를 과제화하여 분석 과제 풀로 관리하면서 분석 프로젝트를 선정하는 작업을 수행한다.
- 2) 과제 수행 : 분석을 수행할 팀을 구성하고 분석 과제 실행 시 지속적인 모니터링과 과제 결과를 공유하고 개선하는 절차를 수행한다
  - 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)를 포함한 결과물을 풀에 잘 축적하고 관리함으로써 향후 유사한 분석과제 수행 시 새행착오를 최소화하고 프로젝트를 효율적으로 진행

## 중요함



## 분석 교육 및 변화관리

: 많은 기업들은 새로운 체계의 도입시에는 저항 및 기존 행태로 돌아가는 관성이 존재하므로, 안정 추진기로 들어가기 위해서는 분석에 관련된 교육 및 마인드 육성을 위한 적극적인 변화 관리가 필요함

분석 교육의 목표 : 단순한 툴 교육이 아닌 분석역량 확보 및 강화

1. 분석 기획자 : 데이터 분석 큐레이션 교육
2. 분석 실무자 : 데이터 분석 기법 및 툴에 대한 교육
3. 업무 수행자 : 분석기획 발굴, 구체화, 시나리오 작성법 등

빅데이터 시대의 변화에 적극적인 대응 방법

1. 기업과 맞는 적합한 분석 업무 수행
2. 분석 조직 및 인력에 대한 지속적인 교육과 훈련
3. 경영층이 사실기반 의사결정을 할 수 있는 문화 정착
4. 지속적인 변화관리 계획하고 수행 업무 수행자에 대한 분석기획 발굴, 구체화, 시나리오 작성법 등