

데이터 분석 전문가 가이드

과목 III 데이터 분석 기획 제 1장 데이터 분석 기획의 이해

2017.01.17

김가경

제1절 분석 기획 방향성 도출

1. 분석 기획의 특징

- 빅데이터 분석의 특징은 수학/통계학적 지식 및 해킹 기술(IT 기술 등) 뿐만 아니라 해당 비즈니스에 대한 이해와 전문성을 포함하여 3가지 영역에 대한 고른 역량과 시각이 요구됨.

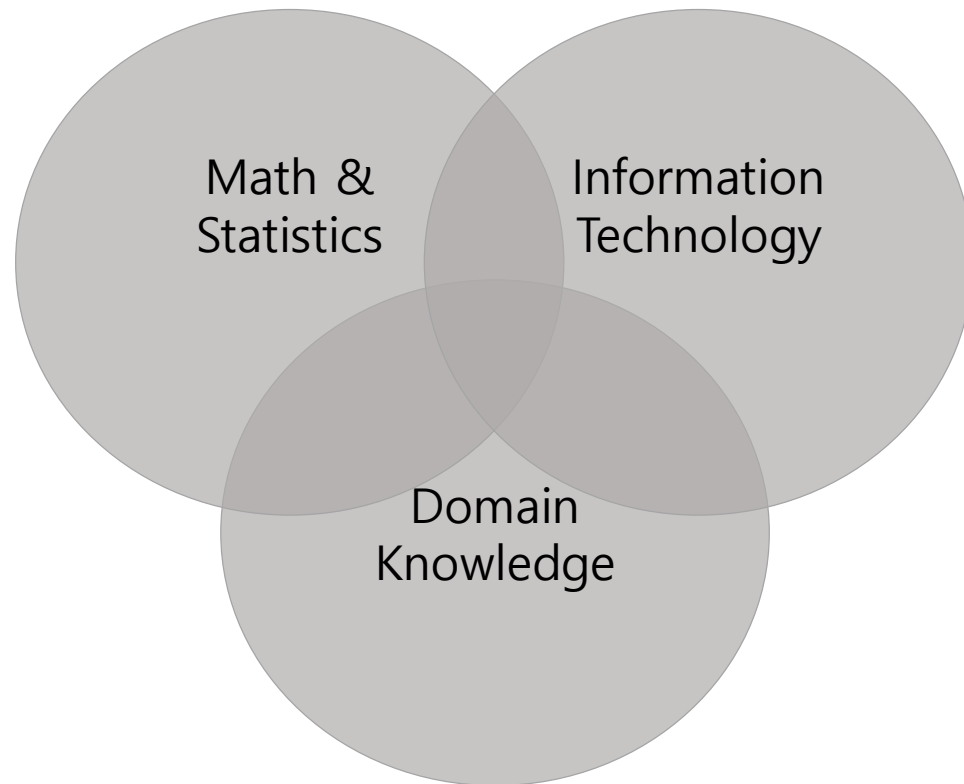


그림 1. 데이터 사이언스 역량

- 분석은 분석의 대상(What) 및 분석의 방법(How)에 따라 4가지로 나뉨

		분석의 대상 (What)	
		<i>Known</i>	<i>Un-Known</i>
분석의 방법 (How)	<i>Known</i>	Optimization (최적화)	Insight (통찰)
	<i>Un-Known</i>	Solution	Discovery (발견)

그림 2. 분석 주제 유형

- 특정한 분석 주제를 대상으로 진행할 경우에도 4가지 유형을 넘나들면서 분석을 수행하고 결과를 도출하는 과정을 반복

- 목표 시점 별로는 당면한 과제를 빠르게 해결하는 “과제 중심적인 접근 방식”과 지속적인 분석 내재화를 위한 “장기적인 마스플랜 방식”으로 나뉨

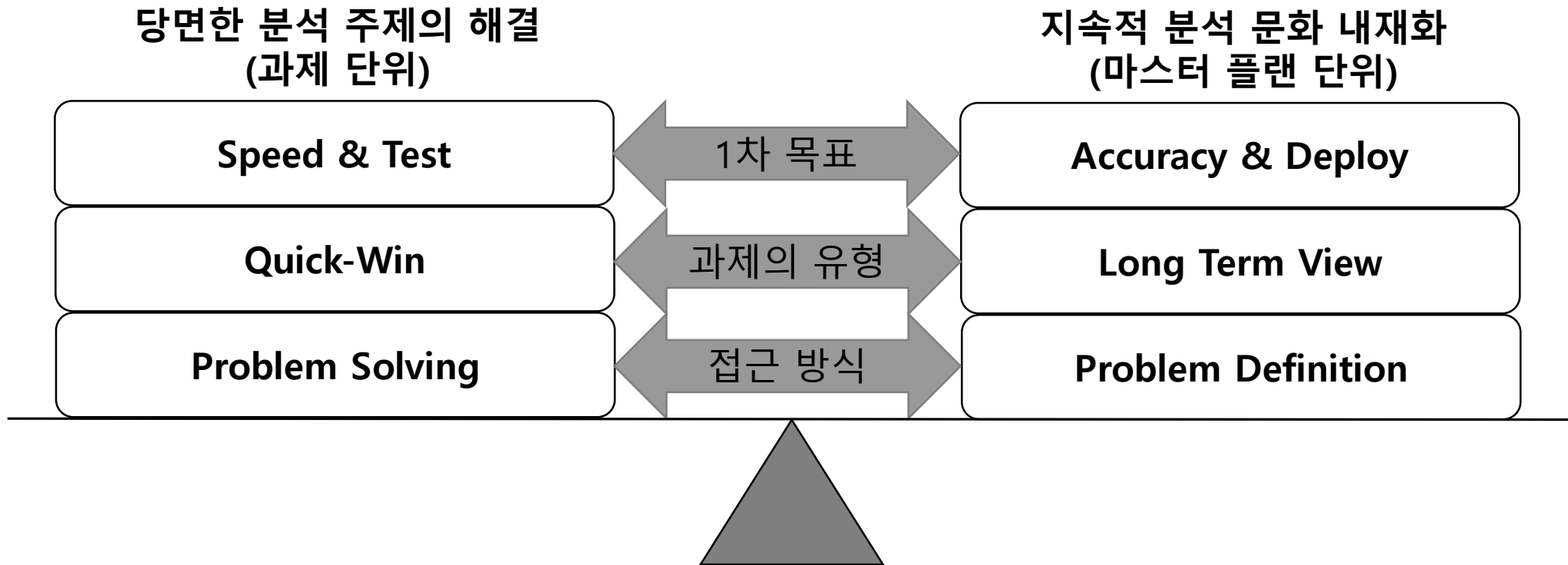


그림 3. 목표 시점 별 분석 기획 방안

- 분석가는 분석 기술, IT 및 프로그래밍, 분석 주제에 대한 도메인 전문성, 의사소통이 중요. 또한 프로젝트 관리 역량, 리더십 역량이 필요.

2. 분석 기획 시 고려사항

1) 가용한 데이터 (그림 5 참조)

- 정형 데이터 (Structured Data)
- 비정형 데이터 : e-mail, 보고서, 소셜 미디어 데이터 등
- 반정형 데이터 : 센서를 중심으로 스트리밍 되는 머신 데이터

2) 적절한 유스케이스

- 기존에 잘 구현되어 활용되고 있는 유사 분석 시나리오 및 솔루션 활용
- Customer&Social media analytics
- Plant and facility&pipeline management
- Price optimization
- Fraud detection

3) 분석과제 수행을 위한 장애요소

- 정확도와 비용 상승 간의 고려
- 사용자가 쉽게 이해하고 활용할 수 있는 방안
- 설치 환경에 따른 성능 문제 고려
- 지속적인 교육 및 활용 방안 등의 변화 관리 (Change Management)

제2절 분석 방법론

1. 분석 방법론 개요

- 상세한 절차(Procedures), 방법(Methods), 도구화 기법(Tools & Techniques), 템플릿과 산출물 (Templates & Outputs) 로 구성되어 어느 정도의 지식이 있으면 데이터 분석 활용이 가능해야 함.
- 개인의 암묵지가 조직의 형식지로 발전하는 형식화 과정을 거치고, 이를 체계화하여 문서화한 최적화된 형식지로 전개됨으로써 방법론이 만들어짐. 방법론은 다시 개인에게 전파되고 내재화 하는 과정을 거쳐 암묵지로 발전하는 선순환 과정이 진행.

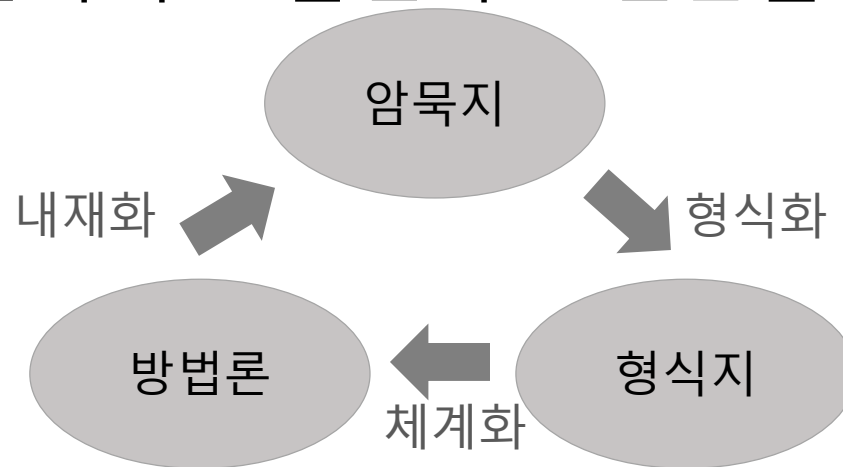


그림 7. 방법론 생성과정

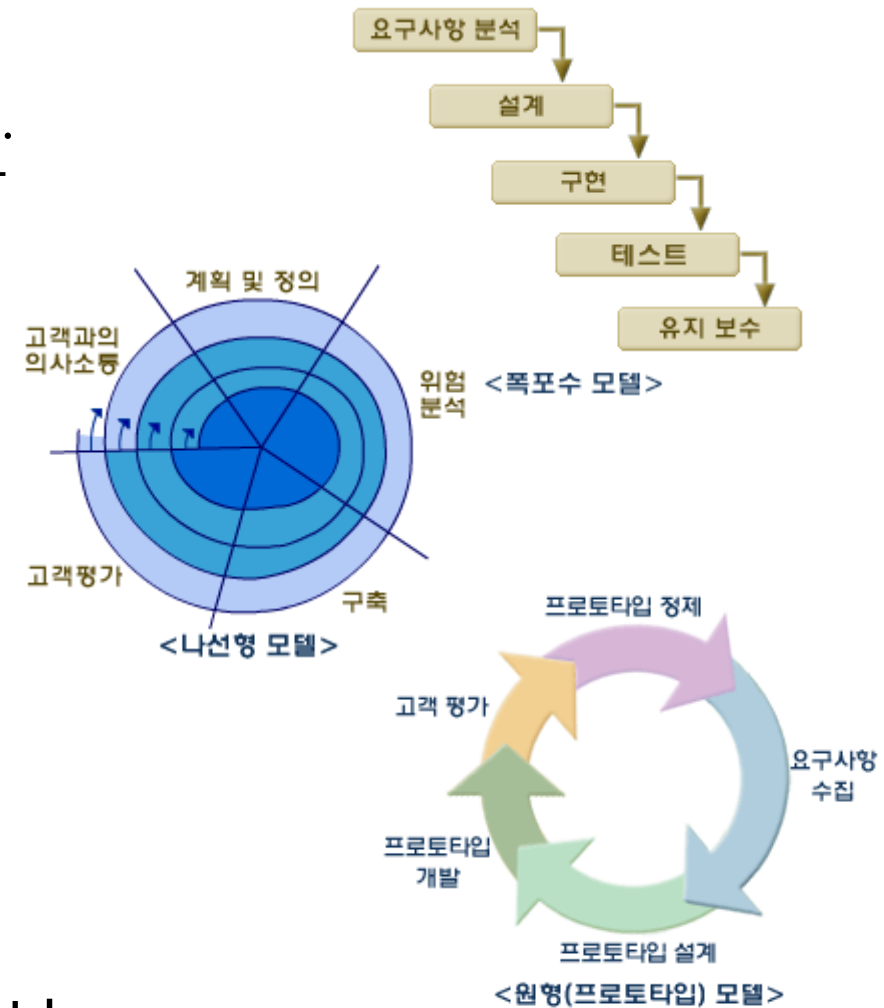
• 방법론의 모델

- 폭포수 모델 (Waterfall Model): 순차적으로 단계를 진행. 하향식 (Top Down)으로 진행되나 문제나 개선사항이 발견되면 전 단계로 돌아가는 피드백 과정이 수행
- 나선적 모델 (Spiral Model): 반복을 통해 점증적으로 개발. 처음 시도하는 프로젝트에 적용이 용이하나 복잡도가 상승
- 프로토타입 모델: 개발초기에 시스템의 모형(원형, prototype)을 간단히 만들어 사용자에게 보여 주고, 사용자가 정보시스템을 직접 사용해 보게 함으로써 기능의 추가, 변경 및 삭제 등을 요구하면 이를 즉각 반영하여 정보시스템 설계를 다시 하고 프로토타입을 재구축하는 과정을 사용자가 만족할 때까지 반복해 나가면서 시스템을 개선시켜 나가는 방식

• 일반적 방법론은 계층적 프로세스의 형태로 구성

- [최상위] 단계 (버전 관리 등을 통한 단계별 완료보고서) -> 태스크 (보고서, 물리적, 논리적 단위로 품질검토의 항목) -> 스텝 (보고서 구성요소, 입력자료, 처리 및 도구, 출력자료로 구성된 단위 프로세스) [최하위]

방법론 참고 자료 <http://gisulsa.tistory.com/>



2. KDD 분석 방법론

- KDD(Knowledge Discovery in Database): 1996년 Fayyad가 체계적으로 정리한 데이터 마이닝 프로세스



- 분석 대상 비즈니스 도메인의 이해
- 분석 대상 데이터셋 선택과 생성
- 데이터에 포함되어 있는 노이즈(Noise)와 이상값(Outlier)등을 제거하는 정제작업이나 선처리
- 분석 목적에 맞는 변수를 찾고 (필요시 데이터의 차원을 축소) 데이터 마이닝 기법과 알고리즘 선택
- 데이터 마이닝 시행 후 결과 해석
- 데이터 마이닝에서 발견된 지식 활용

3. CRISP-DM 분석 방법론

- CRISP-DM (Cross Industry Standard Process for Data Mining): 1996년 유럽연합의 ESPRIT에서 있었던 프로젝트에서 시작하여 DaimlerChrysler, SPSS, NCR 등이 참여하여 1999년 첫 버전 발표
- 최상위 레벨은 여러 단계(Phase)로 구성되고 각 단계는 일반화 태스크(Generic Tasks)를 포함
- 세 번째 레벨은 세분화 태스크(Specialized Tasks)로 일반화 태스크를 구체적으로 수행
- 마지막 레벨인 프로세스 실행(Process Instances)은 데이터 마이닝을 위한 구체적인 실행을 포함.
- 다음 6단계 프로세스로 구성되었으며 각 단계 간의 피드백을 통해 완성도를 높임
 - 업무 이해, 데이터 이해, 데이터 준비, 모델링, 평가(Evaluation), 전개(Deployment)

4. 빅데이터 분석 방법론

- 빅데이터를 분석하기 위한 방법론은 3개의 계층적 프로세스 모델 (Stepwised Process Model)
 - 단계 (Phase): 프로세스 그룹을 통해 완성된 단계별 산출물이 생성. 각 단계는 기준선으로 설정되어 버전 관리 등을 통해 통제
 - 태스크: 태스크는 단계를 구성하는 단위 활동. 물리적 또는 논리적 단위로 품 질검토의 항목
 - 스텝: WBS(Work Breakdown Structure)의 워크패키지(Work Package)에 해당 되고 입력자료(Input), 처리 및 도구(Process&Tool), 출력자료(Output)로 구성된 단위 프로세스(Unit Process)
- 빅데이터 분석 방법론 참조 모델

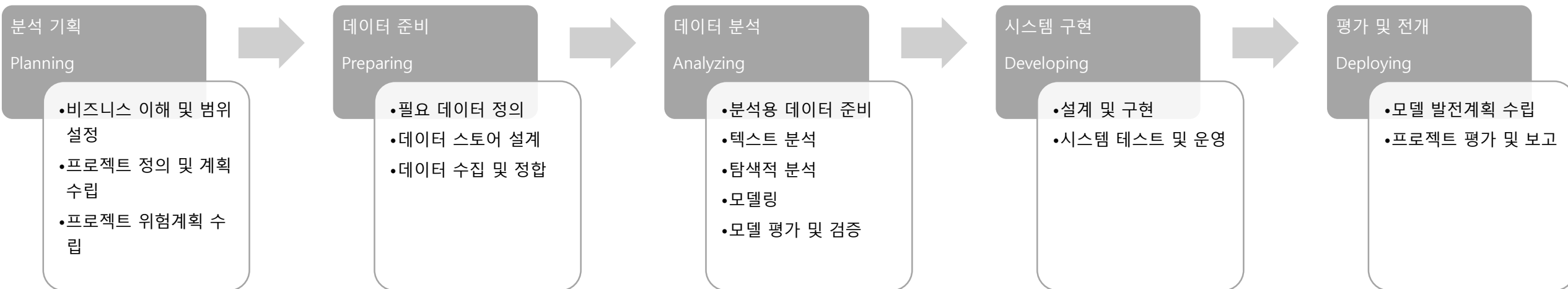


그림 14. 빅데이터 분석 방법론

5. 분석 계획 (Planning)

가. 비즈니스 이해 및 범위 설정

1) 비즈니스 이해

- 입력자료: 업무 매뉴얼, 업무전문가의 지식, 빅데이터 분석 대상 도메인에 대한 관련 자료
- 프로세스 및 도구: 자료 수집 및 비즈니스 이해
- 출력자료: 비즈니스 이해 및 도메인 문제점

2) 프로젝트 범위 설정

- 입력자료: 중장기 계획서, 빅데이터 분석 프로젝트 지시서, 비즈니스 이해 및 도메인 문제점
- 프로세스 및 도구: 자료 수집 및 비즈니스 이해, 프로젝트 범위 정의서 작성 절차
- 출력자료: 프로젝트 범위 정의서(SOW)

나. 프로젝트 정의 및 계획 수립

1) 데이터 분석 프로젝트 정의

- 입력자료: 프로젝트 범위 정의서, 빅데이터 분석 프로젝트 지시서
- 프로세스 및 도구: 프로젝트 목표 구체화, 모델 운영 이미지 설계
- 출력자료: 프로젝트 정의서, 모델 운영 이미지 설계서, 모델 평가 기준

2) 프로젝트 수행 계획 수립

- 입력자료: 프로젝트 정의서, 모델 운영 이미지 설계서, 모델 평가 기준
- 프로세스 및 도구: 프로젝트 수행 계획 작성, WBS 작성 도구, 일정계획 수립 도구
- 출력자료: 프로젝트 수행 계획서, WBS

다. 프로젝트 위험계획 수립

1) 데이터 분석 위험 식별

- 입력자료: 프로젝트 정의서, 프로젝트 수행 계획서, 선행 프로젝트 산출물 및 정리자료
- 프로세스 및 도구: 위험 식별 절차, 위험 영향도 및 발생 가능성 분석, 위험 우선순위 판단
- 출력자료: 식별된 위험 목록

2) 위험 대응 계획 수립: 예상되는 위험에 대한 대응은 회피, 전이, 완화, 수용으로 구분하여 위험 관리 계획서를 작성

- 입력자료: 식별된 위험 목록, 프로젝트 정의서, 프로젝트 수행 계획서
- 프로세스 및 도구: 위험 정량적 분석, 위험 정성적 분석
- 출력자료: 위험관리 계획서

6. 데이터 준비

가. 필요 데이터 정의

1) 데이터 정의

- 입력자료: 프로젝트 수행 계획서, 시스템 설계서, ERD, 메타데이터 정의서, 문서 자료
- 프로세스 및 도구: 내,외부 데이터 정의, 정형/비정형/반정형 데이터 정의
- 출력자료: 데이터 정의서

2) 데이터 획득방안 수립

- 입력자료: 데이터 정의서, 시스템 설계서, ERD, 메타데이터 정의서, 문서 자료, 데이터 구입
- 프로세스 및 도구: 데이터 획득 방안 수립
- 출력자료: 데이터 획득 계획서

나. 데이터 스토어 설계

1) 정형 데이터 스토어 설계 – 관계형 데이터베이스인 RDBMS(Relational Data Base Management System)를 사용

- 입력자료: 데이터 정의서, 데이터 획득 계획서
- 프로세스 및 도구: 데이터베이스 논리설계, 데이터베이스 물리설계, 데이터 매핑
- 출력자료: 정형 데이터 스토어 설계서, 데이터 매핑 정의서

2) 비정형 데이터 스토어 설계 – 하둡, NoSQL 등을 이용

- 입력자료: 데이터 정의서, 데이터 획득 계획서
- 프로세스 및 도구: 비정형,반정형 데이터의 논리 및 물리 설계
- 출력자료: 비정형 데이터 스토어 설계서, 데이터 매핑 정의서

다. 데이터 수집 및 정합성 점검

1) 데이터 수집 및 저장

- 입력자료: 데이터 정의서, 데이터 획득 계획서, 데이터 스토어 설계서
- 프로세스 및 도구: 데이터 크롤링 도구, ETL(Extract Transform Load) 도구, 데이터 수집 스크립트
- 출력자료: 수집된 분석용 데이터

2) 데이터 정합성 점검: 데이터 스토어의 품질 점검을 통해 정합성을 확보하고 품질개선 필요한 부분의 보완 작업

- 입력자료: 수집된 분석용 데이터
- 프로세스 및 도구: 데이터 품질 확인, 데이터 정합성 점검 리스트
- 출력자료: 데이터 정합성 점검 보고서

7. 데이터 분석 (1)

가. 분석용 데이터 준비

- 1) 비즈니스 룰 확인
 - 입력자료: 프로젝트 정의서, 프로젝트 수행 계획서, 데이터 정의서, 데이터 스토어
 - 프로세스 및 도구: 프로젝트 목표 확인, 비즈니스 룰 확인
 - 출력자료: 비즈니스 룰, 분석에 필요한 데이터 범위
- 2) 분석용 데이터셋 준비
 - 입력자료: 데이터 정의서, 데이터 스토어
 - 프로세스 및 도구: 데이터 선정, 데이터 변환, ETL 도구
 - 출력자료: 분석용 데이터셋

나. 텍스트 분석

- 1) 텍스트 데이터 확인 및 추출
 - 입력자료: 비정형 데이터 스토어
 - 프로세스 및 도구: 분석용 텍스트 데이터 확인, 텍스트 데이터 추출
 - 출력자료: 분석용 텍스트 데이터
- 2) 텍스트 데이터 분석
 - 입력자료: 분석용 텍스트 데이터, 용어사전/용어 동의어 사전, 불용어 사전 등
 - 프로세스 및 도구: 분류체계 설계, 형태소 분석, 키워드 도출, 토픽 분석, 감성 분석, 오피니언 분석, 네트워크 분석
 - 출력자료: 텍스트 분석 보고서

다. 탐색적 분석

- 1) 탐색적 데이터 분석
 - 입력자료: 분석용 데이터셋
 - 프로세스 및 도구: EDA 도구, 통계 분석, 변수간 연관성 분석, 데이터 분포 확인
 - 출력자료: 데이터 탐색 보고서
- 2) 데이터 시각화
 - 입력자료: 분석용 데이터셋
 - 프로세스 및 도구: 시각화 도구 및 패키지, 인포그래픽, 시각화 방법론
 - 출력자료: 데이터 시각화 보고서

7. 데이터 분석 (2)

라. 모델링

1) 데이터 분할

- 입력자료: 분석용 데이터셋
- 프로세스 및 도구: 데이터 분할 패키지
- 출력자료: 훈련용 데이터, 테스트용 데이터

2) 데이터 모델링

- 입력자료: 분석용 데이터셋
- 프로세스 및 도구: 통계 모델링 기법, 기계학습, 모델 테스트
- 출력자료: 모델링 결과 보고서

3) 모델 적용 및 운영 방안

- 입력자료: 분석용 데이터셋
- 프로세스 및 도구: 모니터링 방안 수립, 알고리즘 설명서 작성
- 출력자료: 알고리즘 설명서, 모니터링 방안

마. 모델 평가 및 검증

1) 모델 평가

- 입력자료: 모델링 결과 보고서, 평가용 데이터
- 프로세스 및 도구: 모델 평가, 모델 품질관리, 모델 개선작업
- 출력자료: 모델 평가 보고서

2) 모델 검증

- 입력자료: 모델링 결과 보고서, 모델 평가 보고서, 검증용 데이터
- 프로세스 및 도구: 모델 검증
- 출력자료: 모델 검증 보고서

8. 시스템 구현

가. 설계 및 구현

1) 시스템 분석 및 설계

- 입력자료: 알고리즘 설명서, 운영중인 시스템 설계서
- 프로세스 및 도구: 정보시스템 개발방법론
- 출력자료: 시스템 분석 및 설계서

2) 시스템 구현

- 입력자료: 시스템 분석 및 설계서, 알고리즘 설명서
- 프로세스 및 도구: 시스템 통합개발도구(IDE), 프로그램 언어, 패키지
- 출력자료: 구현 시스템

나. 시스템 테스트 및 운영

1) 시스템 테스트

- 입력자료: 구현 시스템, 시스템 텍스트 계획서
- 프로세스 및 도구: 품질관리 활동
- 출력자료: 시스템 테스트 결과보고서

2) 시스템 운영 계획

- 입력자료: 시스템 분석 및 설계서, 구현 시스템
- 처리 및 도구: 운영계획 수립, 운영자 및 사용자 교육
- 출력자료: 운영자 매뉴얼, 사용자 매뉴얼, 시스템 운영 계획서

9. 평가 및 전개(Deploying)

가. 모델 발전 계획 수립

1) 모델 발전 계획

- 입력자료: 구현 시스템, 프로젝트 산출물
- 프로세스 및 도구: 모델 발전 계획 수립
- 출력자료: 모델 발전 계획서

나. 프로젝트 평가 및 보고

1) 프로젝트 성과 평가

- 입력자료: 프로젝트 산출물, 품질관리 산출물, 프로젝트 정의서, 프로젝트 수행 계획서
- 프로세스 및 도구: 프로젝트 평가 기준, 프로젝트 정량적 평가, 프로젝트 정성적 평가
- 출력자료: 프로젝트 성과 평가서

2) 프로젝트 종료

- 입력자료: 프로젝트 산출물, 품질관리 산출물, 프로젝트 정의서, 프로젝트 수행 계획서, 프로젝트 성과 평가서
- 프로세스 및 도구: 프로젝트 지식자산화 작업, 프로젝트 종료
- 출력자료: 프로젝트 최종 보고서

제3절 분석 과제 발굴

- 하향식 접근 방식(Top Down Approach)
문제가 주어지고 이에 대한 해법을 찾기 위하여 각 과정이 체계적으로 단계화 되어 수행하는 방식
- 상향식 접근 방식(Bottom Up Approach)
문제의 정의 자체가 어려운 경우 데이터를 기반으로 문제의 재정의 및 해결방안을 탐색하고 이를 지속적으로 개선하는 방식
- 중요한 의사 결정을 할때 상향식 접근 방식의 발산(Diverge) 단계와 도출된 통찰을 분석하고 검증하는 하향식 접근 방식의 수렴(Converge) 단계를 반복할 수 있는 최적의 의사결정 방식

분석의 대상 (What)

분석의 방법 (How)

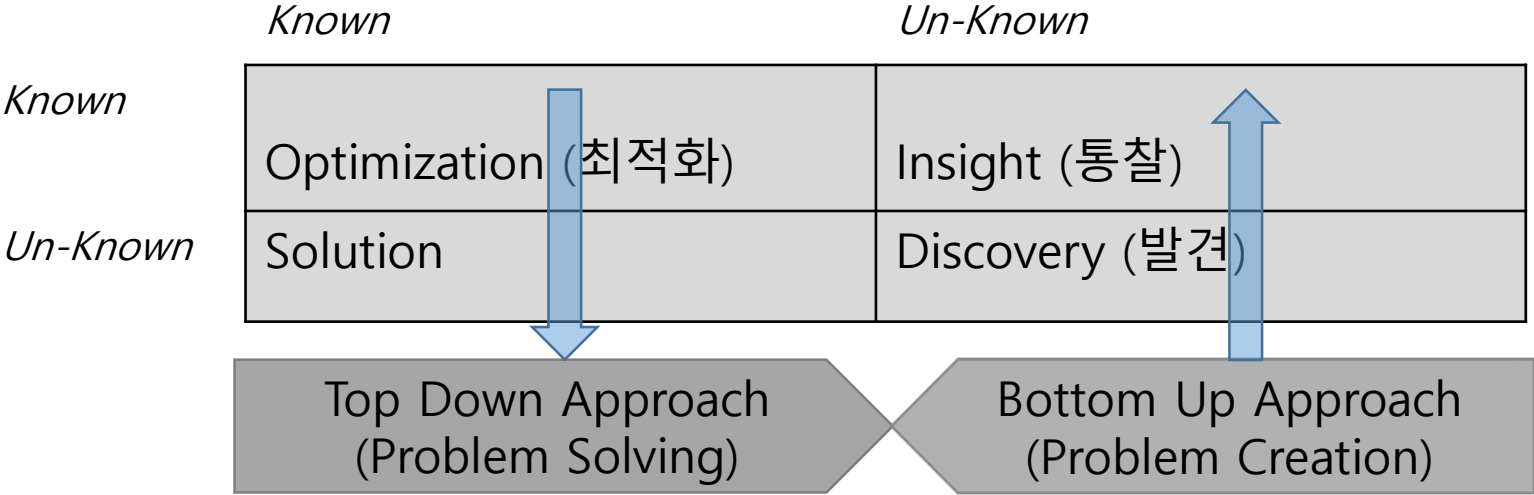
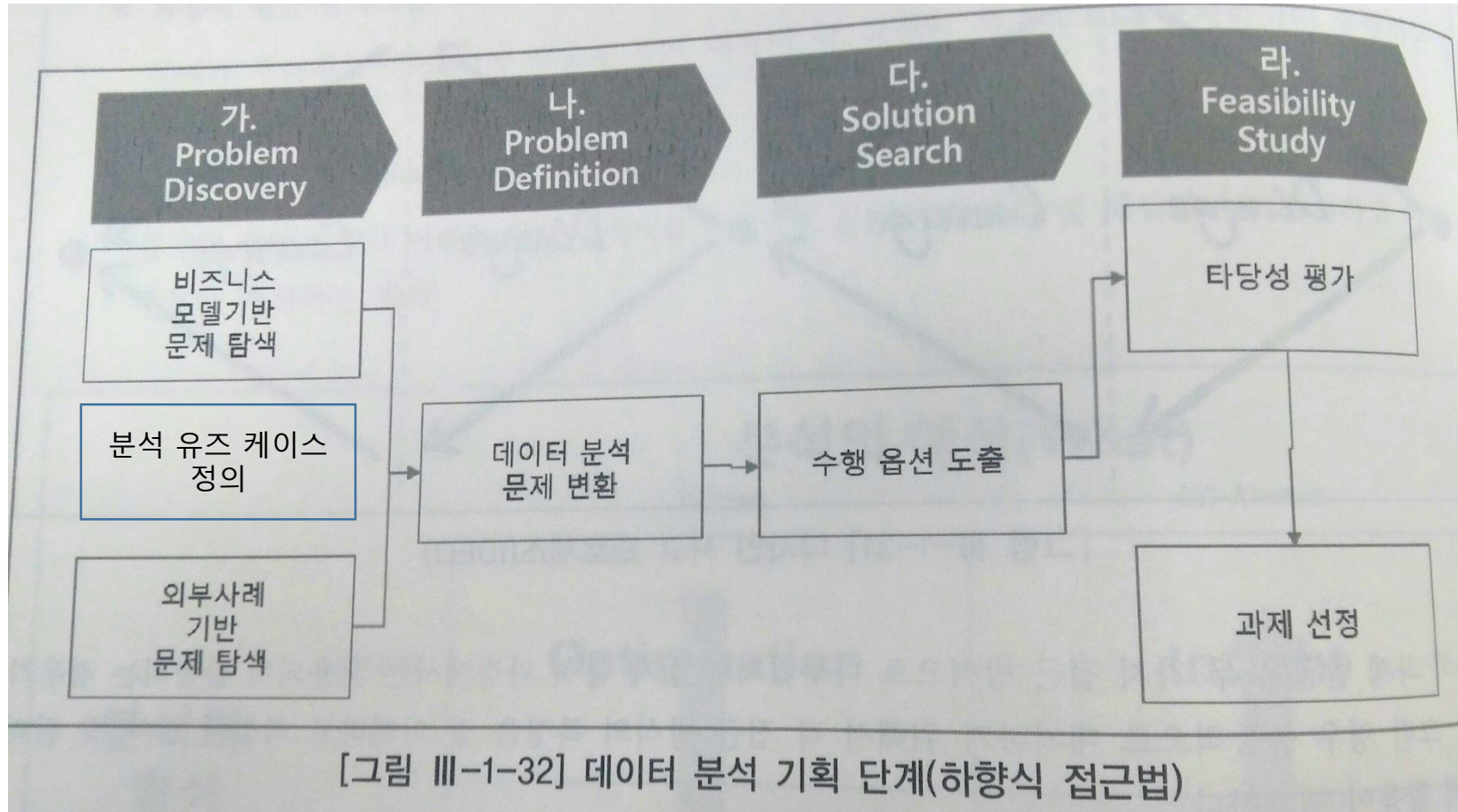
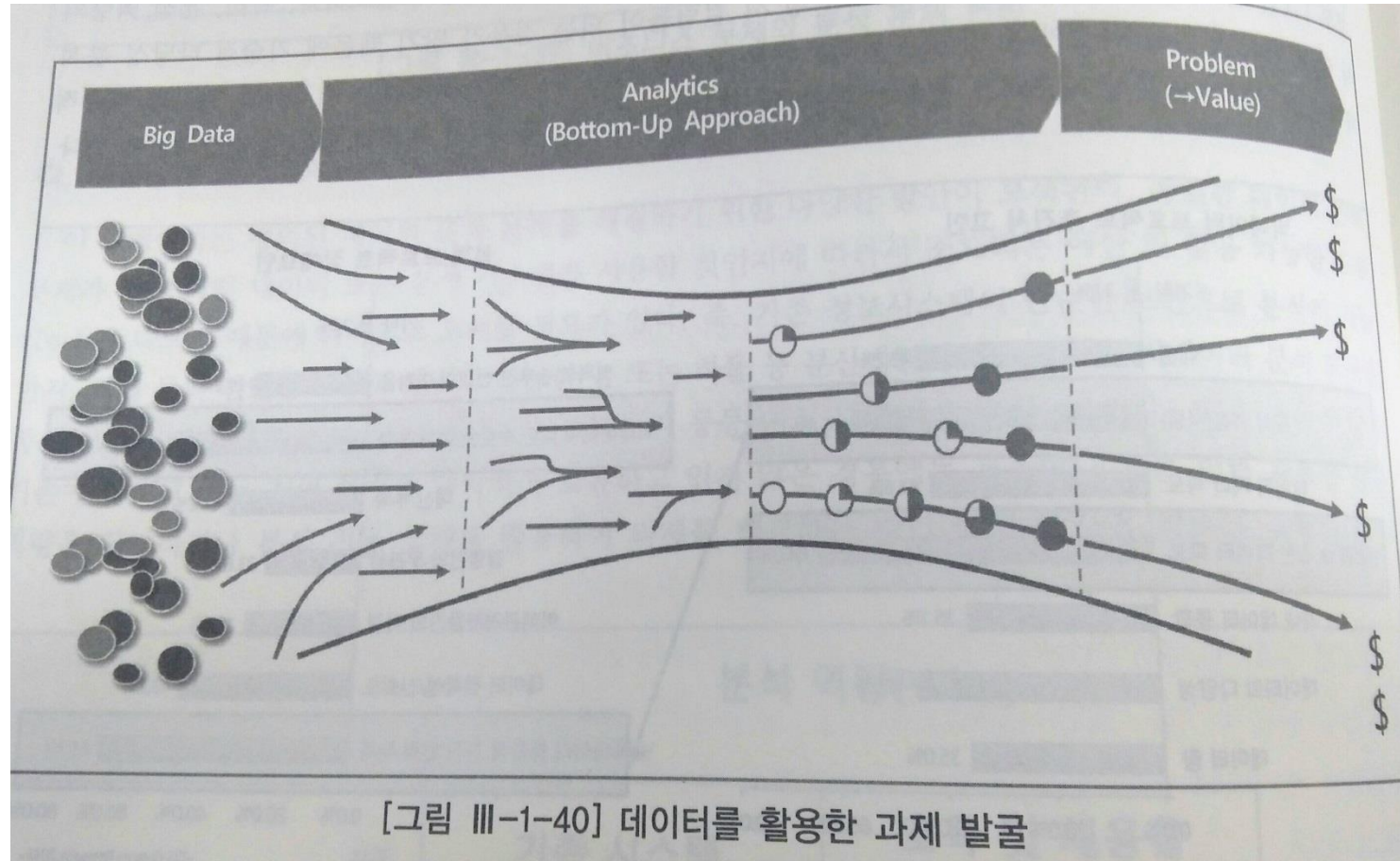


그림 30. 분석 과제 도출의 두가지 유형

1. 하향식 접근 방식

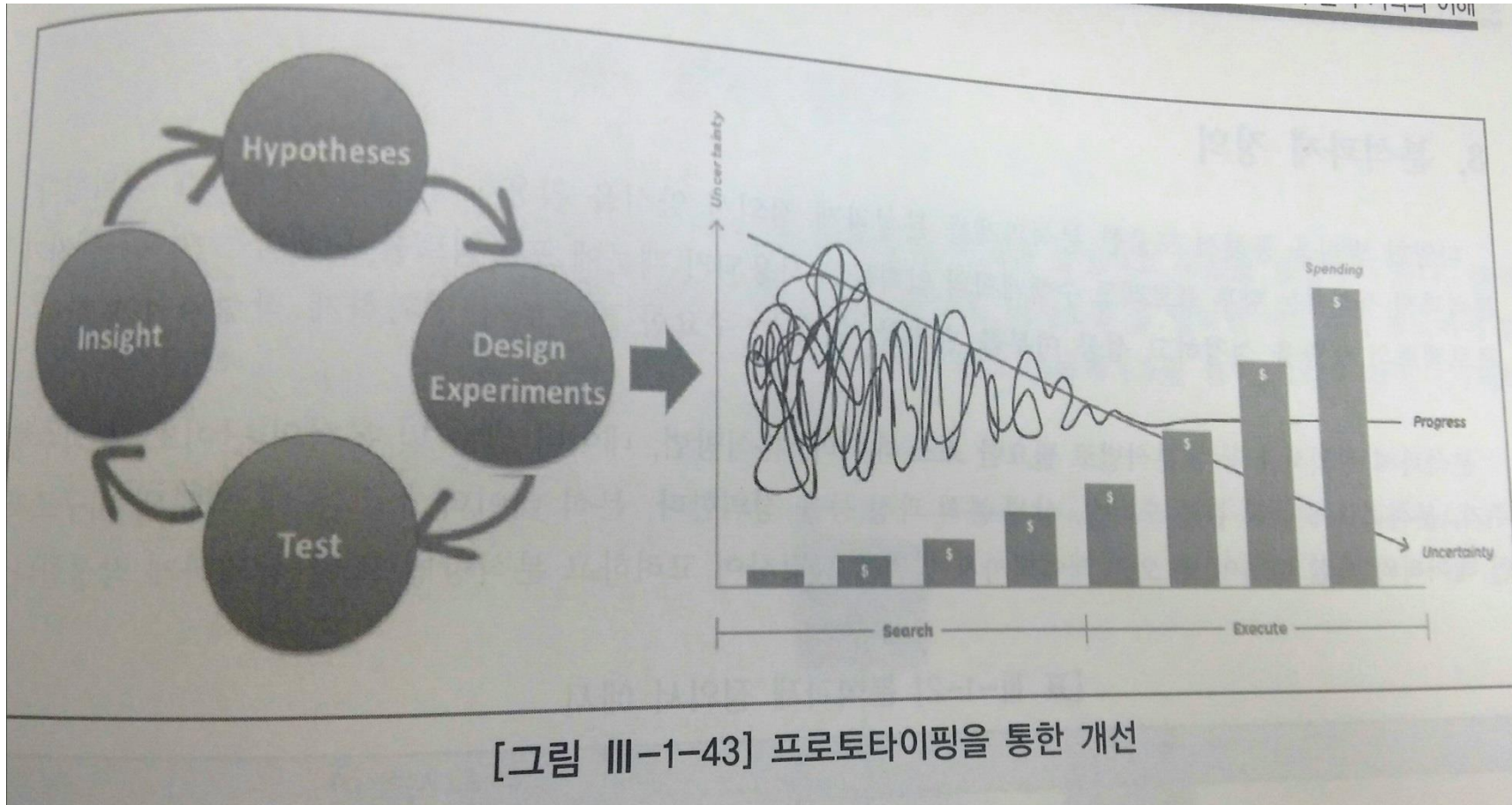


2. 상향식 접근 방식



빅데이터 분석 환경에서 프로토타이핑의 필요성

- 문제에 대한 인식 수준: 문제 정의가 불명확하거나 새로운 문제일 경우
- 필요 데이터 존재 여부의 불확실성
- 데이터 사용 목적의 가변성



3. 분석 과제 정의

- 분석 과제 정의서는 향후 프로젝트 수행계획의 입력물로 사용되기 때문에 프로젝트를 수행하는 이해관계자가 프로젝트의 방향을 설정하고 성공 여부를 판별할 수 있는 주요한 자료로서 명확하게 작성되어야 한다.

제4절 분석 프로젝트 관리 방안

1. 분석 프로젝트의 특성

- 분석 프로젝트는 데이터 영역과 비즈니스 영역에 대한 이해뿐만 아니라 지속적인 반복이 요구되는 분석 프로세스의 특성을 이해한 프로젝트 관리방안을 수립하는 것이 중요

2. 분석 프로젝트 관리방안

주제그룹		개념 및 관련 프로세스
총	합	프로젝트와 관련된 다양한 활동과 프로세스를 도출, 정의, 결합, 단일화, 조정, 통제, 종료에 필요한 프로세스
이	해관계자	프로젝트 스폰서, 고객사, 기타 이해관계자를 식별하고 관리하는데 필요한 프로세스
범	위	작업과 인도물을 식별하고 정의하는데 요구되는 프로세스
자	원	인력, 시설, 장비, 자재, 기반 시설, 도구와 같은 적절한 프로젝트 자원을 식별하고 확보하는데 필요한 프로세스
시	간	프로젝트 활동의 일정을 수립하고 일정 통제의 진척상황을 관찰하는데 요구되는 프로세스
원	가	예산 개발과 원가통제의 진척상황을 관찰하는데 요구되는 프로세스를 포함
리	스	위험과 기회를 식별하고 관리하는 프로세스
품	질	품질보증과 품질통제를 계획하고 확립하는데 요구되는 프로세스
조	달	계획에 요구된 프로세스를 포함하며 제품 및 서비스 또는 인도물을 인수하고 공급자와의 관계를 관리하는데 요구되는 프로세스
의	사소통	프로젝트와 관련된 정보를 계획, 관리, 배포하는데 요구되는 프로세스

[그림 III-1-47] KS A ISO 21500의 프로젝트 관리 주제 그룹

[표 III-1-3] 분석 프로젝트 영역별 주요 관리 항목

관리 영역	분석 프로젝트의 특성 및 주요 관리 항목
범위 (Scope)	<ul style="list-style-type: none"> - 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 범위가 크게 변경되므로 사전에 충분한 고려가 필요함
시간 (Time)	<ul style="list-style-type: none"> - 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)	<ul style="list-style-type: none"> - 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 위하여 상용 버전의 도구(Tool)가 필요할 수 있음 (예: 가시화를 위한 BI 솔루션, 지리정보 표기를 위한 GIS 솔루션 등)
품질 (Quality)	<ul style="list-style-type: none"> - 분석 프로젝트를 수행한 결과에 대한 품질목표를 사전에 수립하여 확정해야 함 - 프로젝트 품질은 품질통제(Quality Control)와 품질보증(Quality Assurance)으로 나누어 수행되어야 함
통합 (Integration)	<ul style="list-style-type: none"> - 프로젝트관리 프로세스들이 통합적으로 운영될 수 있도록 관리함
조달 (Procurement)	<ul style="list-style-type: none"> - 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)	<ul style="list-style-type: none"> - 고급 분석 및 빅데이터 아키텍처링을 수행할 수 있는 인력의 공급이 부족하므로 프로젝트 수행 전 전문가 확보 검토 필요
리스크 (Risk)	<ul style="list-style-type: none"> - 분석에 필요한 데이터 미확보로 분석 프로젝트 진행이 어려울 수 있어 관련 위험을 식별하고 대응방안을 사전에 수립해야 함 - 데이터 및 분석 알고리즘의 한계로 품질목표를 달성하기 어려울 수 있어 대응방안을 수립할 필요가 있음
의사소통 (Communication)	<ul style="list-style-type: none"> - 전문성이 요구되는 데이터 분석의 결과를 모든 프로젝트 이해관계자(Stakeholders)가 공유할 수 있도록 해야 함 - 프로젝트의 원활한 진행을 위한 다양한 의사소통체계 마련 필요
이해관계자 (Stakeholder)	<ul style="list-style-type: none"> - 데이터 분석 프로젝트는 데이터 전문가, 비즈니스 전문가, 분석 전문가, 시스템 전문가 등 다양한 전문가가 참여하므로 이해관계자의 식별과 관리가 필요함

1절 분석 기획 방향성 도출

분석 기획이란 실제 분석을 수행하기에 앞서 분석을 수행할 과제의 정의 및 의도했던 결과를 도출할 수 있도록 이를 적절하게 관리할 수 있는 방안을 사전에 계획하는 일련의 작업이다. 분석 과제 및 프로젝트를 직접 수행하는 것은 아니지만, 어떠한 목표(Why)를 달성하기 위하여 어떠한 데이터(What)를 가지고 어떤 방식으로(How) 수행할 지에 대한 일련의 계획을 수립하는 작업이기 때문에 성공적인 분석 결과를 도출하기 위한 중요한 사전 작업이다.

2절 분석 방법론

데이터 분석을 효과적으로 기업내 정착하기 위해서는 이를 체계화한 절차와 방법이 정리된 데이터 분석 방법론의 수립이 필수적이다. 프로젝트는 한 개인의 역량이나 또는 조직의 우연한 성공에 기인해서는 안 되고 일정한 수준의 품질을 갖춘 산출물과 프로젝트의 성공 가능성을 확보하고 제시할 수 있어야 한다. 따라서 방법론은 상세한 절차(Procedures), 방법(Methods), 도구와 기법(Tools & Techniques), 템플릿과 산출물(Templates & Outputs)로 구성되어 어느 정도의 지식만 있으면 활용이 가능해야 한다.

3절 분석 과제 발굴

분석 과제는 풀어야 할 다양한 문제를 데이터 분석 문제로 변환한 후 해당 이해관계자들이 이해하고 프로젝트로 수행할 수 있는 과제정의서 형태로 도출된다. 분석 과제를 도출하기 위한 방식은 크게 2가지로 나뉘어진다.

- 하향식 접근 방식(Top Down Approach)
문제가 주어지고 이에 대한 해법을 찾기 위하여 각 과정이 체계적으로 단계화되어 수행하는 방식
- 상향식 접근 방식(Bottom Up Approach)
문제의 정의 자체가 어려운 경우 데이터를 기반으로 문제의 재정의 및 해결방안을 탐색하고 이를 지속적으로 개선하는 방식

4절 분석 프로젝트 관리방안

과제 형태로 도출된 분석 기회는 프로젝트를 통해서 그 가치를 증명해야 한다. 분석 프로젝트는 일반적인 IT 프로젝트처럼 범위, 일정, 품질, 이슈/리스크, 의사소통 등 영역별 관리가 수행되어야 할 뿐 아니라 데이터를 다루면서 분석 모델을 생성하는 특성상 Data Size, Data Complexity, Speed, Model Complexity, Accuracy & Precision 등 5가지의 추가적인 중점 관리 영역에 대한 고려가 필요하다.

분석가의 경우 비즈니스 영역과 데이터 영역의 중재자로서 주로 분석 프로젝트의 관리자 역할을 수행하는 경우가 많기 때문에, 분석 프로젝트의 영역별 관리 포인트를 사전에 이해하고 적용하는 것이 프로젝트의 성공 가능성을 높이게 된다.

연습문제

문제 1. 데이터 분석을 통한 가치 발굴에서 필요한 주요 요소가 아닌 것은?

- ① Data
- ② Hadoop
- ③ Analytic Model
- ④ Analyst

문제 2. 다음 문장의 ()안을 기입하십시오.

분석 기획은 단기적으로는 (①)를 도출하여 프로젝트화 한 후 관리를 수행하여 분석결과를 도출하는 것이고, 중장기적으로는 (②) 수행을 지원할 수 있는 거버넌스 체계를 수립하는 것이다.

문제 3. 분석가가 가져야 할 가장 주요한 역량은?

- ① 문제에 대한 전문성 역량
- ② 수학/통계적 지식 등 분석 역량
- ③ 프로그래밍 등 기술 역량
- ④ 1), 2), 3)을 포괄하는 역량

문제 4. 데이터 분석 단계 모델링 태스크중 모델 적용 및 운영방안 스텝의 주요한 산출물은?

- ① 알고리즘 설명서
- ② 모델 검증 보고서
- ③ 모델 발전 계획서
- ④ 프로젝트 범위 정의서

문제 5. 분석 주제 유형 중, 문제를 잘 알고 있으면서 기존에 수행하고 있는 방법이 존재하는 경우에 해당하는 유형은?

- ① Optimization
- ② Solution
- ③ Discovery
- ④ Insight

문제 6. 다음 문장의 ()안을 기입하시오.

분석 과제를 도출하기 위한 방식은 문제가 주어진 경우 해법을 찾기 위하여 절차적으로 수행하는 (①

) 방식과 문제의 정의 자체가 어려운 경우 데이터를 기반으로 탐색하고 이를 지속적으로 개선해나가는 방식인 (②) 로 분류된다.

문제 7. 비즈니스 모델 캔버스를 활용한 분석 기회 발굴 시에 탐색하는 주요 영역이 아닌 것은?

- ① 경쟁자(Competitor)
- ② 업무(Operation)
- ③ 제품(Product)
- ④ 고객(Customer)

문제 8. “예상치 않은 설비장애로 인한 판매량 감소”라는 비즈니스 문제를 분석 문제로 변환하여 기입하시오.

문제 9. 다음 문장의 ()안을 기입하시오.

분석 프로젝트는 IT 프로젝트와 달리, 앞서 분석 방법론에서 살펴본 (①) 방식과 문제의 정의 자체가 어려운 경우 데이터를 기반으로 탐색하고 이를 지속적으로 개선해나가는 방식인 (②)로 분류된다.

문제 10. 분석 프로젝트 수행시 "Data, Business, 분석 등 다양한 영역의 사람들이 프로젝트에 참여하며, Project Sponsor 및 향후 분석 결과를 활용할 User 등 다양한 사람들의 니즈를 고려해야 한다"라는 관리 포인트에 대한 부분을 고려해야 하는 영역은?

- ① 원가(Cost)
- ② 이해관계자(Stakeholder)
- ③ 범위(Scope)
- ④ 품질(Quality)