

데이터 분석 프로젝트 가이드

분석 개요, 프로세스 및 역할

2022년 6월 15일
Kim Jin Soo

I. 데이터 분석 개요

II. 분석과제 프로세스

III. R & R

IV. 분석 관련 팁

데이터 분석이란 데이터를 보고 통계적인 지식 또는 시각화 도구를 활용하여 새로운 사실을 알아내거나 특정한 인사이트를 얻는 행위로, 데이터의 특성을 파악하고 원인을 파악하기 위한 모형(model)을 만드는 기술을 필요

데이터 분석 프로세스

문제정의

- 업무의 목적, 이유, 비즈니스에 미치는 영향, 구체적인 설계와 지표, 일정과 예상 Output 등에 대해 협의하는 단계

데이터 수집

- 파일 형태 또는 API 형태로 데이터를 수집할 수 있으며, 수집된 데이터를 불러오고 프로그램에 저장하는 단계

데이터 처리

- 데이터 필터링, 그룹핑, 조인 등 데이터 조합 또는 이상치 제거, 표준화, 카테고리화, 차원 축소 등 데이터 정리의 단계

데이터 분석

- 탐색적분석, 변수간 관계 파악, 통계 모형 개발 등을 통해 사업과 관련된 지표 개발, 산출, 문제해결 및 추천하는 단계

리포팅

- 적절한 시각화 방법을 활용하여 분석결과를 사업관계자에게 간결하고 명확한 메시지의 형태로 전달하는 단계

데이터 분석 기법 종류

탐색적/
상관
분석
(cnt, μ)

- 다양한 변수를 조합해가며 특이점 파악

통계
분석
(ρ , H_0)

- 데이터를 요약하여 숫자를 산출하거나 모수를 추정

데이터
마이닝
(pred)

- 데이터의 패턴, 규칙 등 학습을 통해 예측&분류

시각화 :
차트

- 차트나 도표 등으로 분석결과를 표현하여 의미 전달

시각화 :
지도

- 지도 위에 변수들을 시각화하여 표현

I. 데이터 분석 개요

II. 분석과제 프로세스

III. R & R

IV. 분석 관련 팁

I. 데이터 분석 개요

II. 분석과제 프로세스

III. R & R

IV. 분석 관련 팁

1. 아이디어 구상

2. 분석시나리오 선정

3. 분석결과 활용

4. 분석시나리오 예시

5. 실습 : 분석시나리오 작성

주제선정 – 아이디어 구상

데이터 분석을 진행하기 위해서는 먼저 분석을 하게 된 배경과 해결방법, 그리고 활용방안을 제시해야 합니다. 현재 거의 모든 분야에서 데이터 분석이 진행되고 있으며, 각 상황에 맞는 분석으로 인사이트를 도출합니다.

사업 분야	분석 목표	분석 내용
의료	예측	▪ 건강보험심사평가원 명세서 데이터를 분석해 환자들의 결핵 발병을 예측
의료	예측	▪ 데이터 분석을 통해 폐의 CT 스캔에 기초한 환자의 폐 기능 저하 정도 예측
의료	추천	▪ 데이터 분석을 통해 유방암 여부를 예측함으로써 조기 발견·치료 제언
스포츠	예측	▪ NFL 선수가 핸드오프(공을 건네받음)를 받은 후 몇 야드를 달릴 것인지 예측
스포츠	예측	▪ 농구선수들의 여러 기록 데이터를 분석해 공헌도 평가라는 기준으로 농구선수 TOP 5 를 예측
교육	추천	▪ 교육 과정 개편에 따라 데이터 분석 결과를 근거로 자녀가 진학할 고등학교 선택을 할 수 있도록 추천
교육	예측	▪ 사회연결망 분석을 통해 국내 유아 영어교육에 사회적 인식 예측
...
유통	예측	▪ 유통업계의 지속적인 구매감소 고객의 구매패턴을 분석해 구매감소 고객을 미리 예측
농업	예측	▪ 기상 데이터, 양파 재배면적, 생산량과 SNS 연관 데이터 복합 분석을 통해 양파 생산량 예측
제조	추천	▪ 자동차 엔진오일의 교환 유형을 분석한 결과를 토대로 고객 이탈 예측과 방지 대책 추천
부동산	영향분석	▪ 경기도 신도시 아파트 데이터를 분석하여 매매가격 변동에 영향을 미치는 주요 환경요인들이 무엇인지 분석
안전	추천	▪ 경찰서 신고데이터를 기반으로 순찰차의 최적 순찰경로를 추천
금융	예측	▪ 스타트업마다 투자 받는 금액의 차이에 대한 이유를 분석하여 투자할 만한 스타트업을 미리 예측
SNS	영향분석	▪ 주어진 트윗에 대한 표지의 감정을 보고 어떤 단어나 구절이 감성분석을 가장 잘 지지하는지 분석
...

I. 데이터 분석 개요

II. 분석과제 프로세스

III. R & R

IV. 분석 관련 팁

1. 아이디어 구상

2. 분석시나리오 선정

3. 분석결과 활용

4. 분석시나리오 예시

5. 실습 : 분석시나리오 작성

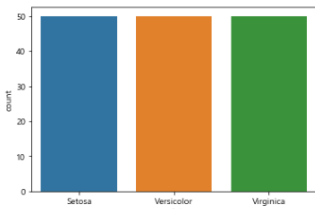
주제선정 - 분석시나리오 선정

분석과제 프로세스
분석시나리오 선정

데이터 분석을 수행할 때 어떤 주제를 기반으로 할 것인지 정해졌다면, 해당 주제에 대해 구체적인 분석목표와 이를 달성하기 위한 세부 프로세스를 수립

EDA(Exploratory Data Analysis)

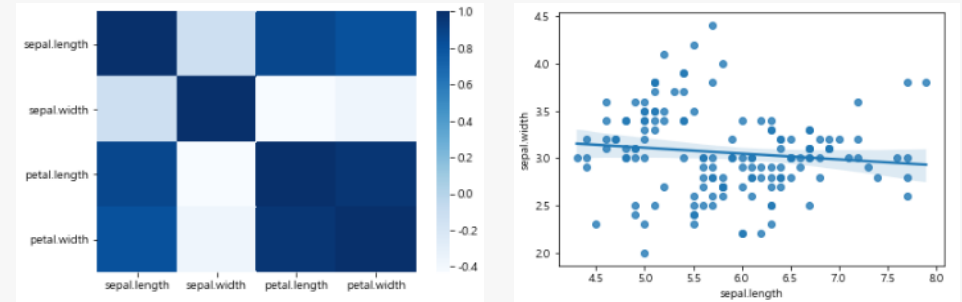
- 데이터 내 몇몇 요소들의 기초통계를 파악하거나 그래프를 통해 패턴 등을 파악하고 분석에 용이한 요소인지 판단



	sepal.length	sepal.width	petal.length	petal.width
variety				
Setosa	5.006	3.428	1.462	0.246
Versicolor	5.936	2.770	4.260	1.326
Virginica	6.588	2.974	5.552	2.026

상관분석(자체가설검증)

- 데이터 내 2개 이상의 변수 간의 관계를 파악하기 위해 상관분석을 진행하고 상관도에 따라 분석 시 종속변수~설명변수 관계를 형성



변수 선택

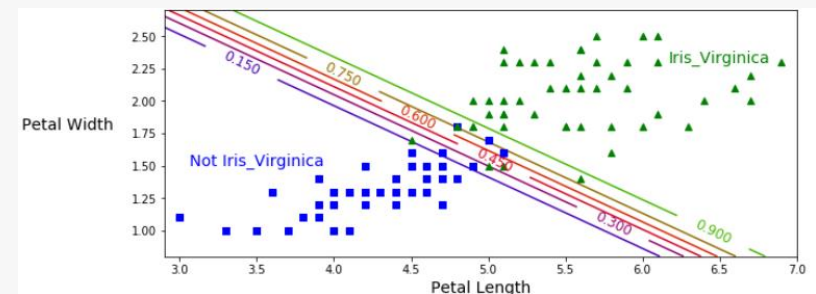
- 분석목적과 상관도 분석을 통해 최종적으로 학습할 변수(x)와 분석 결과를 산출하고자 하는 변수(y)를 조합하여 나열

$$y \sim ax_1 + bx_2 + \dots + \varepsilon$$

$$\text{sepal.length} \sim \text{sepal.width} + \text{petal.width}$$

모델링 및 학습

- 분석목표에 따라 선택한 종속변수-독립변수 관계의 데이터를 학습시킨 후, 실험데이터에 적용하여 분류, 예측, 추천 값 등을 산출




I. 데이터 분석 개요

II. 분석과제 프로세스

III. R & R

IV. 분석 관련 팁

- 
1. 아이디어 구상
 2. 분석시나리오 선정
 3. 분석결과 활용
 4. 분석시나리오 예시
 5. 실습 : 분석시나리오 작성

데이터 분석을 통해 결과값을 산출한 이후에 사업 관련 고객 또는 이해관계자에게 전달할 자료를 작성하고, 완성된 데이터와 결과물을 다시 저장하는 업무를 수행

결과물 산출 및 보고

- 분석을 마치고 나면 분석 결과(예측치)가 포함된 데이터셋 테이블과 차트 등의 산출물 정리
- 분석결과 보고 시에는 사업고객 또는 이해관계자에게 발표를 하거나, 해당 부처에 보고서 형태로 제출



분석코드.ipynb



분석시나리오.csv



분석결과차트.pdf



분석보고서.pdf



분석발표회
일시: 20XX. XX. XX
장소: OO본사 회의실

결과물 배치등록(구축)

- 분석결과물로는 분석언어 스크립트(코드), 분석데이터셋, 모델 알고리즘 등이 존재
- 사업주체 내 정보시스템에 분석파일을 등록하거나, 공개포털 등의 개발 알고리즘을 적용하여 시각화면에 표출



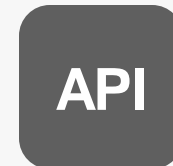
분석코드.ipynb



분석데이터셋.csv



분석결과차트.pdf



웹 화면 표출
및 공유



개방형포털에 분석
시스템 적용



file.zip으로
저장




데이터베이스 내
/data/analysis/file.zip
으로 적재

I. 데이터 분석 개요

II. 분석과제 프로세스

III. R & R

IV. 분석 관련 팁

- 
1. 아이디어 구상
 2. 분석시나리오 선정
 3. 분석결과 활용
 4. 분석시나리오 예시
 5. 실습 : 분석시나리오 작성

추진 필요성

여가활성화를 위한 정부/지자체 역할 증대

- 국민여가활성화기본법제정(15년)
- 국민여가활성화기본계획(18~22년)수립
- 국민의 삶의 질 향상의 범정부적 의무규정
- 지자체 역할(기반구축 관련)



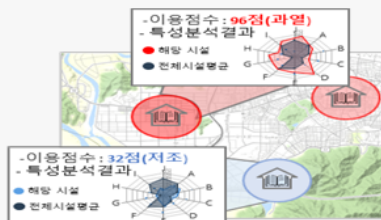
1-3. 일상의 여가공간 확대		
1-3-1. 생활밀착형 여가공간 확대		문체부, 지자체
1-3-2. 국공유지를 활용한 여가접근성 확대		기재부, 국토부, 지자체
1-3-3. 여가친화도시 구축 지원		문체부(지자체)
1-3-4. 여가공간 관리체계 구축		문체부, 지자체

- 대구시는 기존 시설 활성화와 신규 공공시설 확충 필요
- 대구시민의 요구가 반영된 수요자 중심의 정책 개발 필요

분석 목표

"공공 여가기반시설 종합 점수 산출 및 입지 추천"

- (여가기반) 공공시설 이용지수를 산정함으로써 해당 시설 활성화를 위한 객관적인 분석 기반 마련
- 이용객 특성을 도출함에 따라, 수요자 중심의 효과적인 공공시설 유형별 입지 추천



분석 영역

여가 활동 분류

구분	활동 내용
문화예술관람활동	전시회, 박물관, 연주회, 연극, 영화 등 관람
문화예술참여활동	글짓기, 독서토론, 악기연주, 사진, 춤 등
스포츠관람활동	농구, 야구, 축구, 격투기 등 경기 관람
스포츠참여활동	구기운동, 골프, 수영, 육상, 헬스, 자전거 등
휴식(공원 only)	산책, 목욕, 낮잠, TV 시청, 라디오 청취 등
취미오락활동	게임, 쇼핑, 인터넷 활동, 요리, 반려동물 돌보기, 등산, 낚시, 음주, 독서 등
관광활동	문화유적방문, 캠핑, 여행, 소풍, 드라이브 등
사회및기타 활동	사회봉사, 종교활동, 친구만남, 계모임, 가족/친지방문 등

연간 가장 많이 참여한 유형별 여가활동 (1~5순위)



※ Source : 문화체육관광부, 2019 국민여가활동조사

선정 사유

- 적극적 여가활동 활성화 필요
- 적극적 여가활동은 시설이나 공간이 필요

주제선정 - 분석시나리오 예시

분석과제 프로세스
분석시나리오 예시

분석 대상 시설

<유형별 시설 분포 예상도>

- 문화시설 (Cultural Facility)
- 휴식시설 (Leisure Facility)
- 체육시설 (Sports Facility)

	구분	부가 설명
문화	과학관/미술관/박물관	-
	도서관	-
	문예회관	공연장을 중심으로 하는 복합적 성격의 지역문화예술시설 eg. 아트센터, 예술회관
	지방문화원	지역축제, 민속행사 등 전통문화의 발굴·보존과 각종 문화행사를 주최 eg. 중구문화원, 동구팔공문화원
체육	국민체육센터	일반인과 장애인을 위한 생활체육시설
	실내체육관	실내구기종목 및 투기종목
	단일종목시설	
	종합운동장	육상경기 트랙 및 필드 경기장을 갖춘 종합체육시설
휴식	공원	-

- ✓ 문화/체육/휴식의 3개 영역, 세부 9개 유형의 시설을 대상으로 분석 진행
- ✓ 시설 운영/관리 주체가 지자체/중앙정부기관/시설공단/위탁 등 다양하며, 관리항목 및 구분 기준이 상이하여 데이터 가공 및 방문인구를 추정하기 위한 **융복합 분석**이 필요
- ✓ 시설 유형별로 복합적인 특성 존재로 분석 시, 시설의 다양한 이용 요인 및 이용 패턴 도출 등의 **insight**를 제공할 수 있을 것으로 예상

분석 과정

1. 데이터 수집

시설 정보

유동 인구

기상 정보

인구 통계

상권 정보

<데이터 수집>

- 기관 내 데이터
- 기관 외 데이터
- 민간 데이터

<데이터 전처리>

<데이터마트 구성>

2. 데이터 분석

<분석 내용>

- 이용점수 도출
- 시설 특성분석
- 입지분석

<분석 방법>

- 빈도분석
- 회귀분석
- 상관분석
- 패턴분석

3. 시각화

<전달 메시지별 다양한 시각화>

- 시간: 막대/점
- 분포: 파이 차트, 트리맵
- 관계: 스캐터플롯, 버블차트
- 비교: 히트맵, 평행좌표계
- 공간: 지도 맵핑

4. 활용 및 고도화

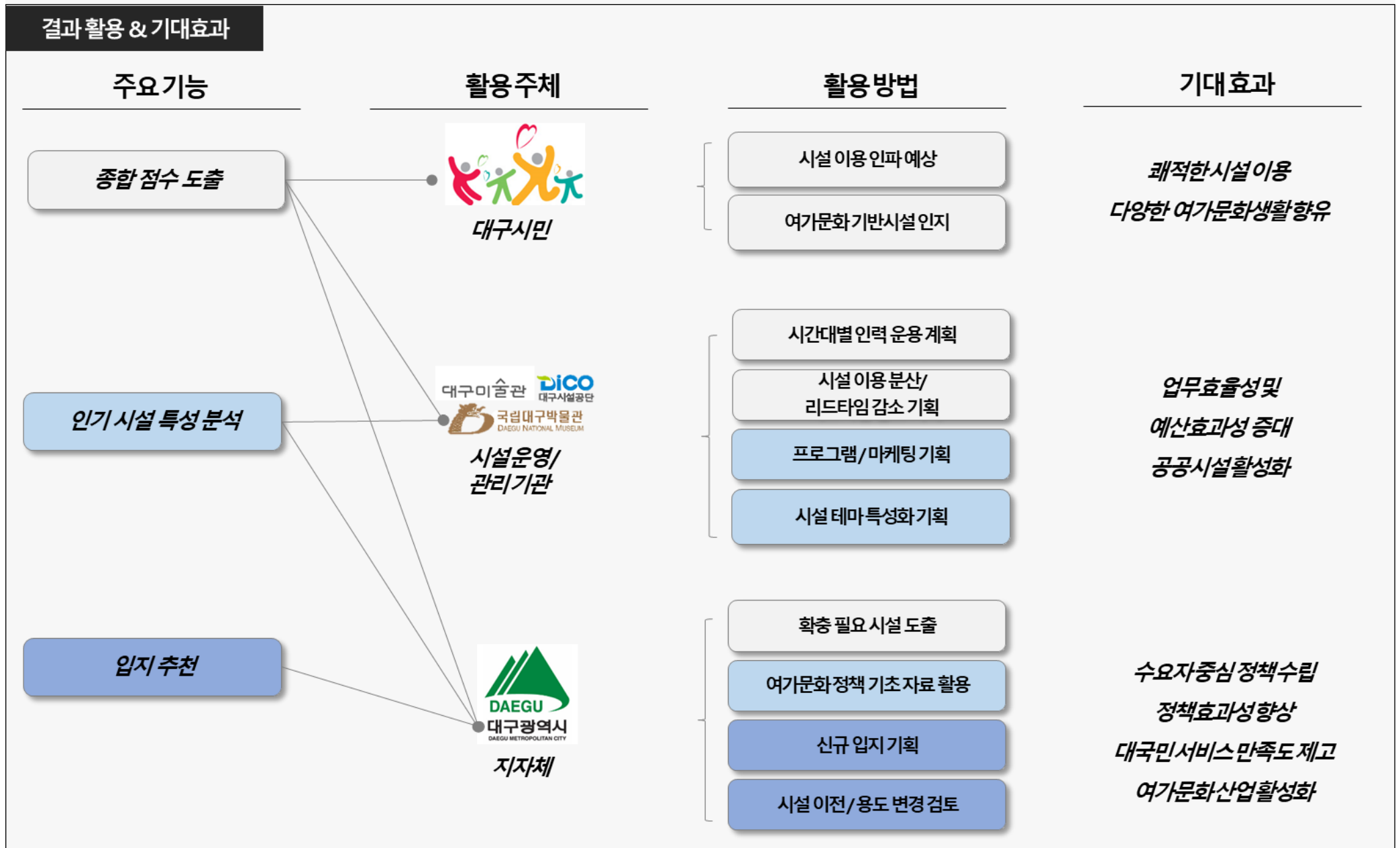
<데이터허브 활용>

- 융복합 데이터셋 제공
- 빅데이터 분석 활용 사례

<공공시설 이용/관리 시스템 구축>

주제선정 - 분석시나리오 예시

분석과제 프로세스
분석시나리오 예시




I. 데이터 분석 개요

II. 분석과제 프로세스

III. R & R

IV. 분석 관련 팁

- 
1. 아이디어 구상
 2. 분석시나리오 선정
 3. 분석결과 활용
 4. 분석시나리오 예시
 5. 실습: 분석시나리오 작성

각 팀별로 주어진 주제 또는 분석하고 싶은 주제를 선정하여 시나리오를 작성합니다. 분석을 하게 된 배경부터 분석 이후의 상황까지 구체적으로 작성합니다.

분석배경

- 문제 정의
- (Output) 분석 필요성 설명

분석목표

- 예측/추천/분류 등
- (Output) 목표를 위한 단계별 업무 정리

필요데이터 목록

- 분석을 위해 필요한 데이터 선정
- (Output) 수집된 데이터

데이터 수집& 탐색 방안

- 데이터 전처리, 상관분석 등
- (Output) 변수 선택 방법과 선택된 변수 목록

분석&시각화

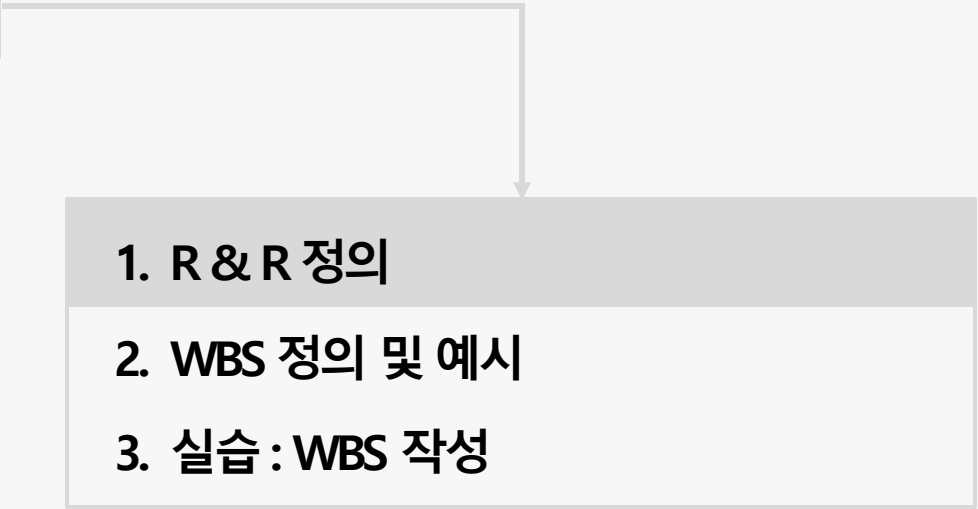
- 분석모델 선택, 알고리즘 실행
- (Output) 모델 결과, 예측결과 데이터

기대효과

- 기대효과 설명
- (Output) 분석결과 활용방안 제시

- I. 데이터 분석 개요
- II. 분석과제 프로세스
- III. R & R
- IV. 분석 관련 팁

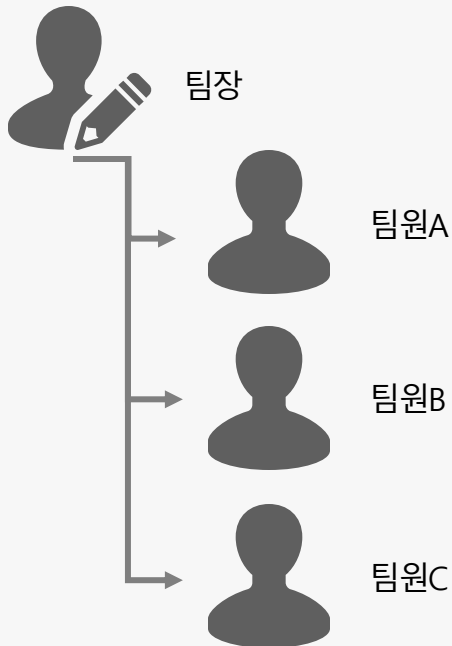
- I. 데이터 분석 개요
- II. 분석과제 프로세스
- III. R & R
- IV. 분석 관련 팁

- 
- 1. R & R 정의
 - 2. WBS 정의 및 예시
 - 3. 실습 : WBS 작성

R&R(Role & Responsibility)은 기업 조직에서 개별 프로세스 및 조직의 구성원들이 수행해야 할 '역할'과 '책임' 관계의 정립을 의미하며, 담당자들에게 Role과 책임범위 그리고 각 업무의 기능을 제공하고 이를 토대로 업무 분장 및 조직도 작성

R & R 작성

- 관리자(Manager)는 모든 담당자들의 명단과 모든 업무(Task)와 역할(Role)에 대해 작성
- 각 구성원들은 역할배치를 통해 프로젝트 기획과 진행을 지원



업무 분장

- 업무목표 달성을 위해 업무를 체계적으로 배분
- 각 개인에게 각자의 업무를 배정하게 함으로써 업무의 효율성을 향상시킬 수 있는 방안 제시

예시

구분	주요 업무	담당자
기 획	▷ 뽕 모비스 업무 창구 ▷ 사업계획 수립 및 관리 ▷ 전문점 영업정책 수립 ▷ 가격책정 수립 및 운영	
마케팅	▷ 신규 ITEM 발굴 및 판매전략 수립 ▷ 시장 규모 및 M/S 분석 ▷ 전략품목 운영전략 수립 및 실시 ▷ 대형부품 매출 활성화 방안 수립	
시스템	▷ 채널링시스템 관리 ▷ PDA 시스템 관리 ▷ 결품자동출고시스템 운영 ▷ 유통정보시스템 신규 개발	
영업운영	▷ 전문점 채권관리 (수금, 연체료, 매출할인)	

조직도

- 조직 구조와 관계, 그리고 상대적 순위, 직위 등을 보여주는 다이어그램
- 모든 부서와 팀과 사람들사이에서 내부 관계를 정의하는 도구로 권한과 책임의 범위, 그리고 조직의 기능에 대해 이미지화

예시



- I. 데이터 분석 개요
- II. 분석과제 프로세스
- III. R & R
- IV. 분석 관련 팁

1. R & R 정의

2. WBS 정의 및 예시

3. 실습 : WBS 작성

R & R – WBS 정의 및 예시

WBS(Work Breakdown Structure)는 프로젝트의 범위와 최종산출물을 세부요소로 분할한 계층적 작업분해도로써 전체업무를 분류하여 구성요소로 만든 후 각 요소를 일정 별로 계획하여 각 담당자에게 할당하는 역할

WBS 구성항목

- 분류번호(1.1, 1.2 등)
- 시작일
- 진척도
- 태스크(과업)
- 종료일
- 담당자
- 세부과업
- 기간
- 상태

WBS의 역할

- 프로젝트에서 수행할 업무를 식별
- 전체일정 진행상황 파악
- 일정과 원가, 자원요구사항
- 프로젝트의 정보 추적 통제
- 일정계획 및 산정
- 고객, 팀 간 의사소통 링크

분석과제 업무분장 예시

성명	구분	주요업무
○○○	기획	· 빅데이터 공모전 사례 조사
	수집	· 공공데이터 포털 조사
		·
△△△	기획	· 스마트시티 사례 조사
	기획	· 지자체 데이터 조사
		·
□□□	기획	· 뉴스 기반 문제정의
	수집	· API 데이터 수집방법 조사
	처리	· 막대그래프, 박스플롯 등 변수 탐색
		·

분석과제 WBS 예시

과업	세부 과업	세세부 과업	담당자	상태	8월																											
					10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28									
① 분석기획	전체	전체																														
	문제정의/시나리오 작성	R&R 및 WBS 작성	팀장	완료																												
		문제 정의(분석배경)	팀원A,B	진행중																												
		분석 방안 강구	팀원C,D	진행중																												
		분석시나리오 선정	공통	진행중																												
	분석범위 설정	전체																														
		분석대상 선정	공통	예정																												
		데이터 검색/수집	공통	예정																												
		분석범위 구체화 및 분석시나리오 수정	공통	예정																												
	전체	전체																														
② 분석수행	마트생성/전처리	전체 분석마트 설계	공통	예정																												
		데이터 추가 검색/수집(필요시)	팀원A	예정																												
		데이터 탐색(기초통계, 막대그래프, 가설수립 등)	팀원B	예정																												
		데이터 전처리(이상치, 결측값 및 기타)	팀원C	예정																												
		분석 데이터마트 생성(key값 설정 및 Join)	팀원D	예정																												
	분석알고리즘 개발	전체																														
		상관분석(데이터 탐색 시 선정된 가설검정)	팀원A,B	예정																												
		분석방법 선정 및 테스트	팀원C,D	예정																												
		최종모델 선정	공통	예정																												
		모델링 결과 요약	팀원A,B	예정																												
	마무리 및 보고자료	모델링 결과 시각화	팀원C,D	예정																												
		전체																														
		코드 취합	공통	예정																												
		산출물 보고(발표)	팀장	예정																												

- I. 데이터 분석 개요
- II. 분석과제 프로세스
- III. R & R
- IV. 분석 관련 팁

-
- ```
graph TD; A[III. R & R] --> B[1. R & R 정의]; B --> C[2. WBS 정의 및 예시]; C --> D[3. 실습 : WBS 작성];
```
- 1. R & R 정의
  - 2. WBS 정의 및 예시
  - 3. 실습 : WBS 작성

각 팀별로 작성한 시나리오에 대해 WBS를 작성합니다. 팀장과 팀원을 정하고, 각 인원이 분석과제를 수행하기 위해 역할을 정하고 일정을 계획합니다.

## Task(과업)

- 분석 시 필요한 업무 나열
- ex 분석 기획 / 분석 개발

## 세부과업

- 과업을 구체적으로 나열
- ex 분석기획 - 아이디어 구상, 시나리오 작성, 데이터 수집

## 시작일 & 종료일

- 세부과업을 수행하는 기간
- ex 20XX. XX. XX ~ 20XX. XX. XX

## 담당자

- 팀장, 팀원, 또는 전원(공통)
- ex 아이디어 구상: 전원, 데이터 검색: 팀원A, 데이터 수집: 팀원B 등


## 상태

- 과업의 진행상황 작성
- ex 진행중, 완료, 이슈발생 등

- I. 데이터 분석 개요
- II. 분석과제 프로세스
- III. R & R
- IV. 프로젝트 관련 팁



- I. 데이터 분석 개요
- II. 분석과제 프로세스
- III. R & R
- IV. 프로젝트 관련 팁

- 
- 1. 효율적 기획방안
  - 2. 효율적 수집방안
  - 3. 효율적 분석방안

해당 조직이 처한 **문제의 정의**와 문제해결을 위한 빅데이터 **분석 기획**이 중요

## 빅데이터 분석 시나리오

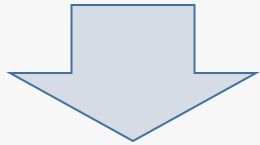
“어떤 데이터” 를 “어떻게 모아”  
“어떤 방법으로 분석, 모델링 하여”  
“어떠한 가치를 제공” 하겠다

## 문제 정의서 작성 도구

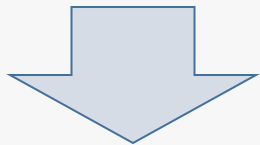
|                                                                                                                                                   |                                                                              |                                                                   |
|---------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------|-------------------------------------------------------------------|
| <b>Motivation</b> 해결하고 싶은 문제는 무엇인가? 그 문제를 인식하고 관심을 갖게 된 동기는 무엇인가?<br><br>나는 (언제) (어디서) 이러한 경험을 한적이 있습니다. 블라블라~~<br>(이런 사람)들의 (어떠한) 고통을 들어주고 싶습니다. |                                                                              |                                                                   |
| <b>Stakeholders</b><br><br>문제와 관련된 사람<br>다양한 이해관계자의 관점에서 문제를 바라보기                                                                                 | <b>Pain Point</b><br><br>이해관계자 별 불편 요소를 기술<br>페인 포인트(Pain Point)는 무엇인지 정리해보기 | <b>Priority</b><br><br>우선순위를 정해보기<br>누구의 어떤 문제가 가장 심각한지를 순차적으로 정리 |
| <b>Problem</b> 해결하고자 하는 문제를 한 문장으로 정의해 보기                                                                                                         |                                                                              |                                                                   |



분석기획 발굴



분석기획 구체화

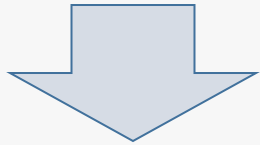


마스터 플랜 수립

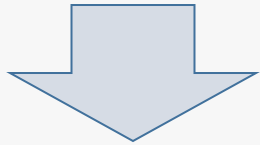
## 1단계: 분석기획 발굴

- 무슨 문제를 해결해야 하는가?
- 왜 그 문제를 해결해야 하는가?
- 문제해결을 통해 달성하고자 하는 목표는?

분석기획 발굴



분석기획 구체화

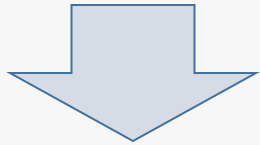


마스터 플랜 수립

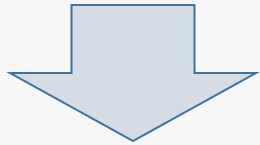
## 2단계: 분석기획 구체화

- 과제도출 단계에서 식별된 문제와 목표를 달성하기 위한 구체적인 방법을 제시
- 문제해결 수행절차, 적용할 방법이나 기술 등을 정의

분석기획 발굴



분석기획 구체화



마스터 플랜 수립

## 3단계: 마스터 플랜 수립

- 도출된 다양한 분석과제들을  
제한된 자원 내에서 효과/효율적  
으로 수행하기 위한 로드맵 및  
일정계획 수립

- 분석과제 추가 발굴 및  
지속적인 분석 고도화 등을 위해  
분석 거버넌스 체계 필요

## 착수보고서 포맷

| JBFG 데이터분석 팀프로젝트 |                     |     |                    |
|------------------|---------------------|-----|--------------------|
| < 프로젝트 착수보고서 >   |                     |     |                    |
| 주 제              | 구체적으로 쓰되 30자 이내로 작성 |     |                    |
| 팀 명              | OOOO팀               | 일 자 | 2022년 6월 17일       |
| 팀 장              | 이름                  | 팀 원 | 팀원명, 팀원명, 팀원명, 팀원명 |

### 1. 과제 개요

- 분석 배경 : 문제 정의 및 분석의 필요성
- 즉, 왜 분석을 해야하는지에 대한 명확한 이유와 목적이 있어야 한다.

### 2. 데이터 수집 전략

- 필요데이터 목록 : 분석에 필요한 데이터 선정
- 대상 데이터가 어디에 있을것인지를 리서치하고 어떻게 수집저장할것인지를 고민
- 분석을 위한 데이터 전처리 작업 및 단계별 데이터 관리 계획

### 3. 데이터 분석 전략

- 분석 시나리오 도출 및 예상
- 예측 결과 데이터 모델링 및 시각화 방안 검토

### 4. 결과 보고 예측 및 실행 계획

- 역할 분담 및 일정 계획
- 결과 도출 전략 및 가설
- 최종 분석완료 후의 기대효과 예측
- Lessons Learned : 프로젝트를 통해 얻고 싶은 것 (기대사항)



- I. 데이터 분석 개요
- II. 분석과제 프로세스
- III. R & R
- IV. 프로젝트 관련 팁

- 
- 1. 효율적 기획방안
  - 2. 효율적 수집방안
  - 3. 효율적 분석방안

# 프로젝트 관련 팁 - 데이터 수집단계 Tip

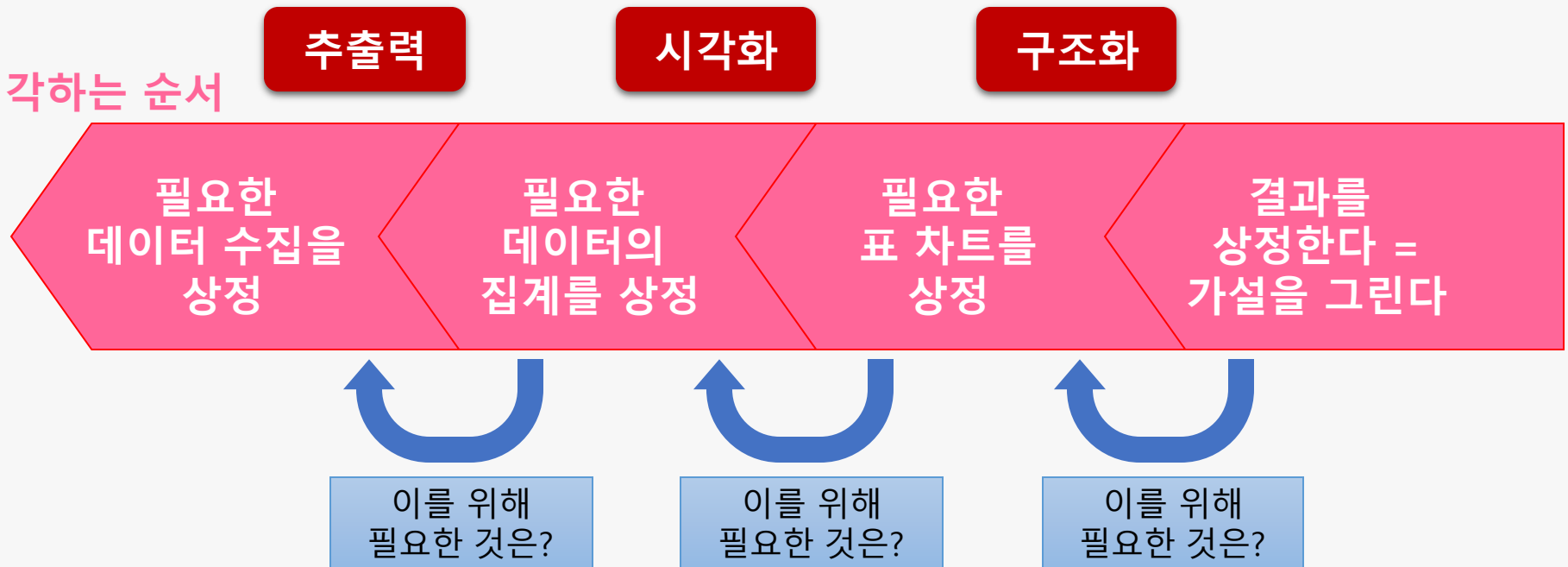
분석 관련 팁  
효율적 데이터수집방안

데이터분석 과정에서 데이터를 수집하기 위한 작업 순서는 우리가 보편적으로 생각하는 순서를 역으로 생각하면 필요한 데이터 수집을 위한 모델링을 보다 쉽게 접근할 수 있다.

## 작업 순서



## 생각하는 순서



- I. 데이터 분석 개요
- II. 분석과제 프로세스
- III. R & R
- IV. 프로젝트 관련 팁

- 
- 1. 효율적 기획방안
  - 2. 효율적 수집방안
  - 3. 효율적 분석방안

분석 진행 시 협업자와의 커뮤니케이션이 무엇보다 중요합니다. 분석 시 각자 맡은 업무를 수행하지만, 최종산출물은 하나로 합쳐지기 때문에, 보다 효율적인 소통과 최종산출물 취합을 위한 배려가 필요합니다.

## 표준데이터 및 변수 사용

- 각 분석가가 분석을 하기 위해 사용되는 데이터와 속성(변수)들은 표준값을 적용하여 진행
- 취합 시 시간과 비용을 절약할 수 있고, 불필요한 오류발생도 최소화할 수 있기 때문

### 데이터 구분

- 00.수집데이터\_마트
- 01.상관분석
- 02.데이터마트
- 03.예측분석
- 04.GIS시각화

### 표준 데이터 지정 및 사용

- DG\_REPORT\_FLOW\_SALE\_MART.csv
- DG\_REPORT\_FOREIGN\_MART.csv
- DG\_REPORT\_RESIDENT\_MART.csv
- DG\_REPORT\_WEATHER\_MART.csv

### 표준변수 기반 학습 모델 저장

- (코드)1.예측분석\_200522\_cl\_Linear\_RF.ipynb
- (코드)1.예측분석\_200522\_cl\_SVM\_LGBM.ipynb

## 코드 주석처리

- 분석코드에는 분석가 개개인의 스타일대로 개발을 하기 때문에, 코드 취합 시 또는 공유 시 해석에 어려움이 있을 수 있음
- 본인의 분석코드 옆에 주석처리(#)를 하여 해당 코드가 무엇을 의미하고 왜 해당 코드를 작성하게 되었는지 등을 공유

### 예시1

```
...
■ Krazy Coupon Lady
- 주요 매장들의 쿠폰북을 한데모아 제공하는 애플리케이션으로
 소비자들에게 효율적인 쇼핑을 도모

■ 주거문제와 고령화 문제를 함께 해결(프랑스 리옹)
- 세대 간 공동주거 제도, 학생들과 노인들이 함께 거주하며,
 학생들은 저렴한 가격으로 임대하면서 노인봉사활동 수료

■ 도시의 청결 문제 해소를 위한 프로젝트(프랑스 파리)
- 시민들을 청결대사로 임명, 쓰레기통 등 추가 설치

■ 대기오염 저감을 위한 슈퍼블록계획(스페인 바르셀로나)
- 버스도로와 자동차도로를 분리하며 교통체증을 완화하고,
 차량제한도로에는 문화, 레저, 지역사회 활동 등의 공간으로 탈바꿈
...
```

### # 서울시 빅데이터 공모전

### 예시2

#### 서울시 내 소비패턴을 분석하여 가게입점 추천

문제 인식 → 분석시나리오 작성 → 데이터 수집/처리/분석

### 예시3

```
data_coor3[:count_finder].shape
좌표변환실패 갯수
fail_count = data_coor3[:count_finder]=='0.0'
sum(fail_count)
좌표변환성공률
print((count_finder-fail_count)/count_finder*100, '%')

결과저장
data_coor3.to_csv('../00.데이터셋/patrol_addr_reverse_190805.csv', index=False)
```

## 데이터분석 프로젝트를 성공시키기 위한 팀워크

### 다양한 것을 연결 지을 수 있다.

- 사람과 사람, 데이터와 데이터, 사업체와 사업체 등을 **가설이나 분석 결과를 근거**로 연결

### 비즈니스 과제에 대한 가설을 수립한다.

- 다각적인 시점에서 **비즈니스 과제를 해결할 수 있는 가설**을 수립
- 그 가설에 기초하여 얻은 데이터 분석 결과를 정확하게 해석

### PDCA로 프로젝트를 추진할 수 있다.

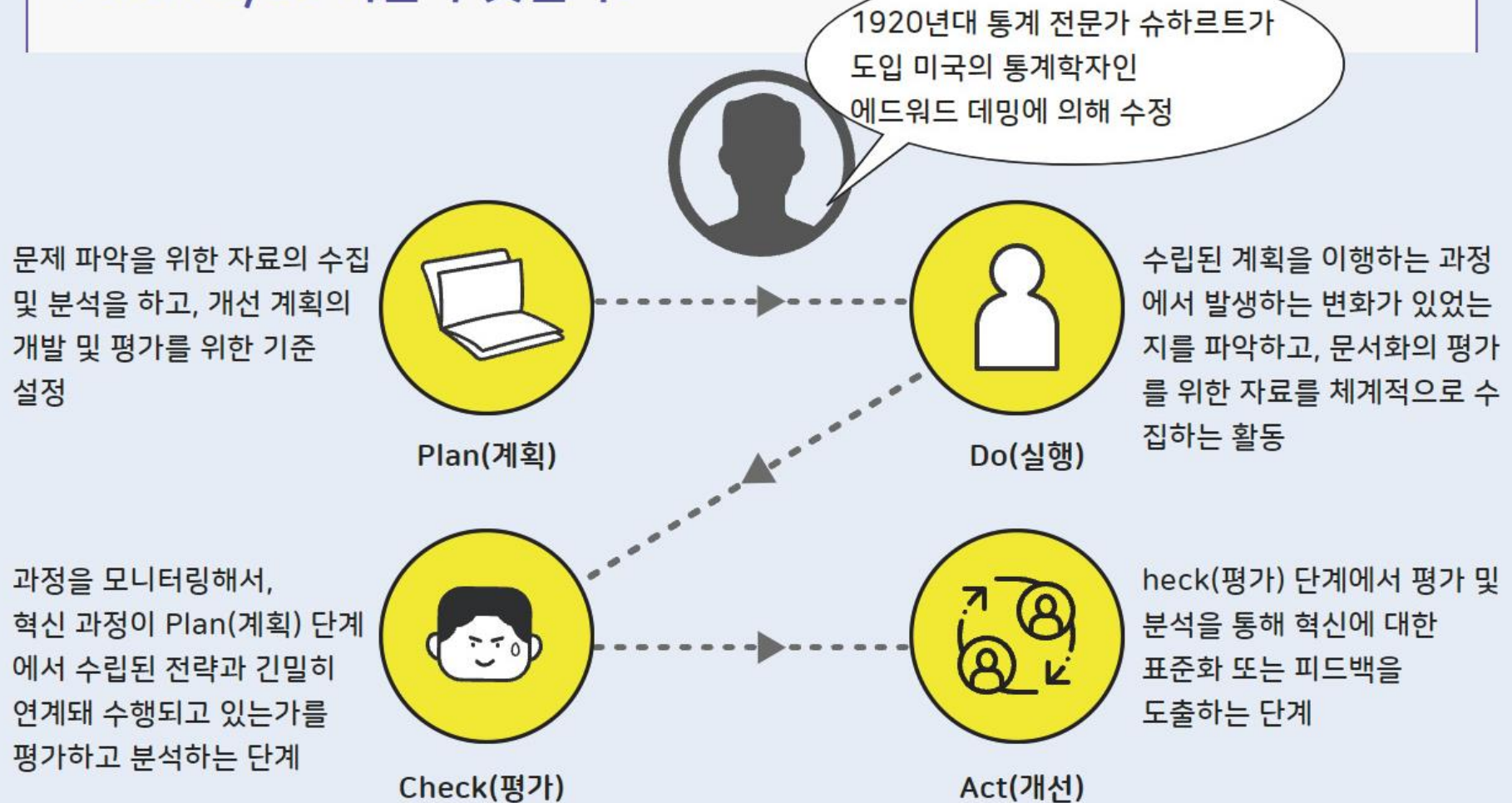
- 데이터의 **가시화, 분석, 고객에 대한 실행, 평가·개선**
- 일련의 PDCA 프로세스를 강력하게 추진

### 마케팅 실행 방안을 IT에 연계할 수 있다.

- 데이터 분석을 위한 테크놀로지를 이해하고 적절한 환경을 구축하는 능력이 필수
- 분석 시스템의 자동화를 기반으로 **마케팅과 IT를 연계**시키는 능력이 요구

## 데이터분석 프로젝트를 성공시키기 위한 팀워크

### PDCA Cycle이란 무엇인가?



마지막 제언!!!

데이터만으로  
모든 것을  
이해할 수 없다

현장을 무시하면서  
데이터만  
내세워서는 안된다.

**B.P.C**

**“Thank You”**