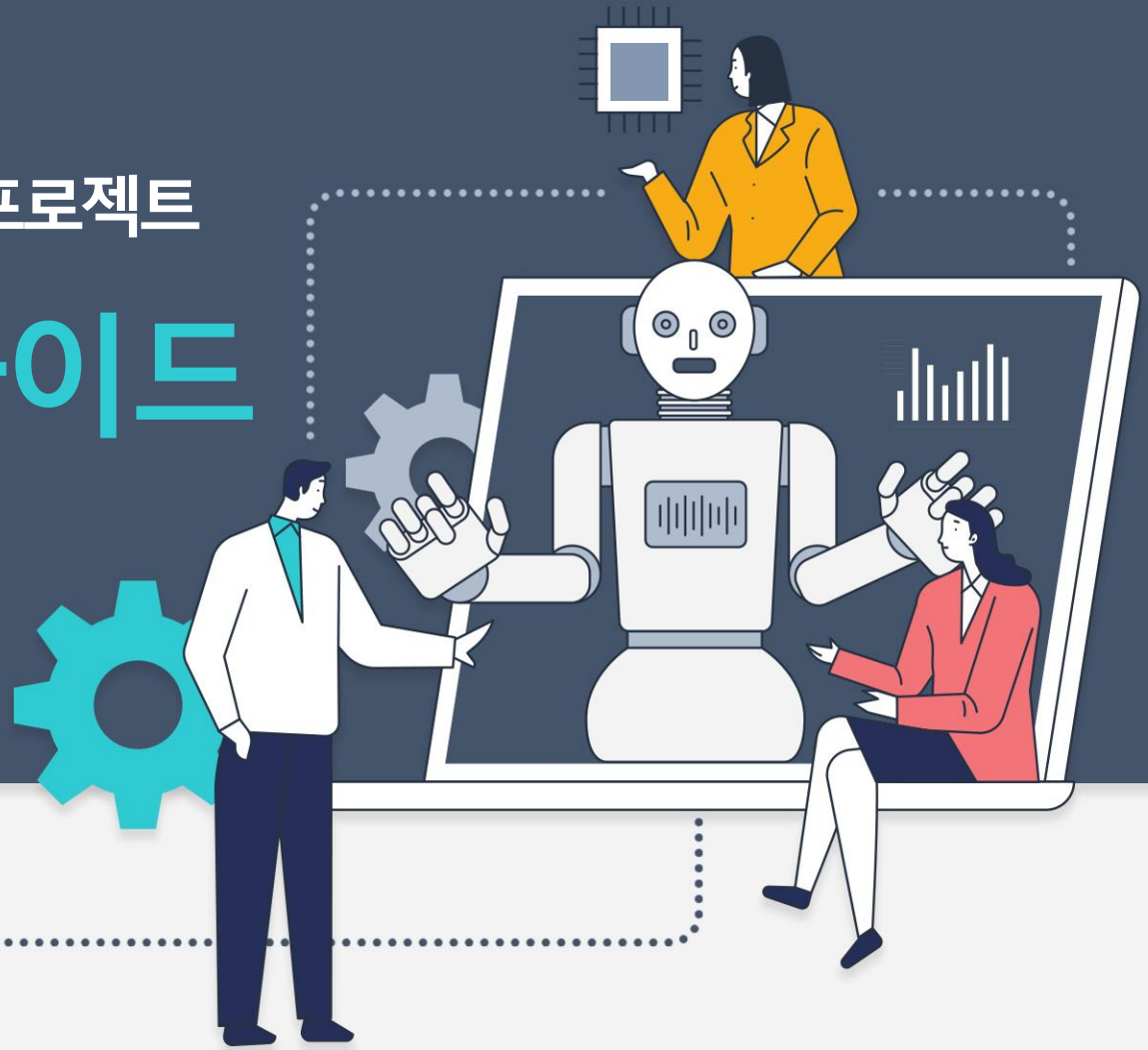


문제해결을 위한 데이터분석 프로젝트

포트폴리오 가이드

-데이터분석 프로젝트



1. 포트폴리오는 **팀별로 작성하여 개별 제출**
(LC 과제제출에는 개별 업로드)
2. 제공된 목차 및 작성요령을 참고하여 작성하되,
템플릿/디자인/구성은 변경 및 추가 가능
3. 포트폴리오를 바탕으로 프로젝트 발표 진행
 - 10~15 슬라이드
 - 발표는 결과 위주로
 - Q&A 포함 20분 이내

- 01 프로젝트 배경
- 02 팀 구성 및 역할
- 03 수행절차 및 방법
- 04 결론 및 향후 과제
- 05 느낀점

작성요령

- ❖ [프로젝트 배경]은 아래와 같은 내용 등으로 구성하여 작성한다.
- ❖ 주제에 대한 한줄 설명 및 시스템 환경 설명
 - 프로젝트 주제
 - 프로젝트 개요
 - 컨셉, 훈련 내용과의 관련성, 개발 환경 등

작성요령

- ❖ [팀 구성 및 역할]은 훈련생 별로 해당 프로젝트를 진행하면서 주도적으로 참여한 부분을 중심으로 작성한다.
- 표, 도식 활용하여 팀원 각각의 역할을 작성

작 성 요 령

- ❖ [수행절차 및 방법]은 프로젝트 수행절차 및 방법을 제시한다.
(구성요소를 포함하여 예시 표와 다르게 수정하여 작성 가능함)

03-1 프로젝트 work-flow

구분	기간	활동	도구
사전 기획	• O/O(월) ~ O/O(목)	• 프로젝트 기획 및 주제 선정 • 기획안 작성	
	• O/O(금)	• 프로젝트 주제 & 아이디어 발표	
전처리 및 모델링	• O/O(월) ~ O/O(목)	• 데이터 전처리 • 모델링 (시각화 포함)	
수정/보완	• O/O(월) ~ O/O(금)	• 피드백 의견 반영하여 프로젝트 고도화	
프로젝트 발표	• O/O(월)		

프로젝트 진행 일정 설명

03-2 데이터명세

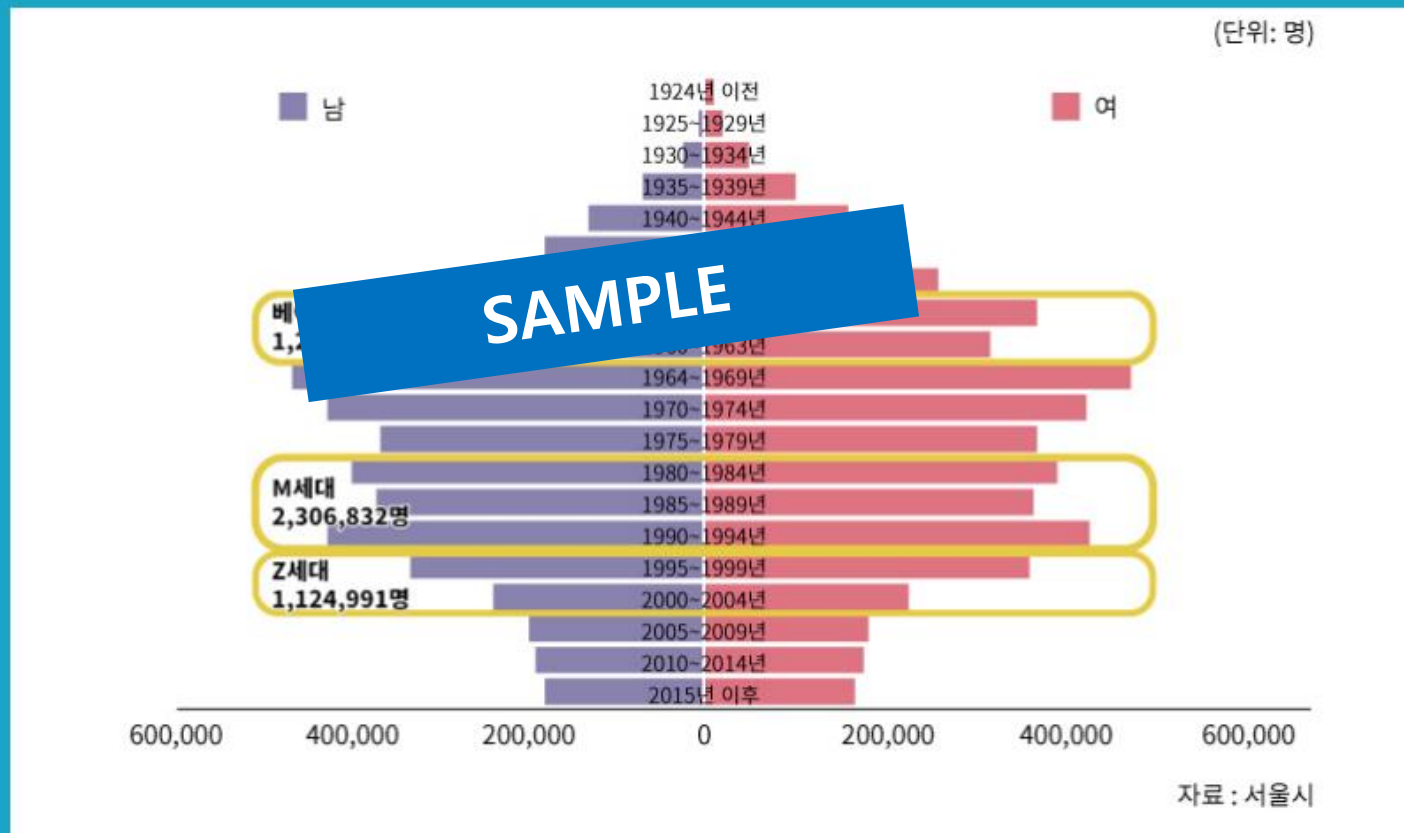
출처	데이터이름	제공 형태	요약
KT-빅데이터센터	시간-지역별 배달 주문건수	CSV	지역-시간-업종별 주문배달건수 기록 자료 (1957315 × 6) 기간 : 2019.07.17 – 2020.09.30
기상자료개방포털	종관기상관측 (ASOS)	CSV	시간대별 기상 요소 관측 자료 (275231 × 15)
에어코리아	미세먼지 데이터		시간대별 미세먼지 관측 자료 (21216 × 9)
공공데이터 서울시 <u>확진자</u> 현황 경기도 <u>감염병관리지원단</u>	코로나 <u>확진자수</u>	csv/API	일별, 지역별 코로나 <u>확진자수</u> 기록 자료 (827 × 3)

SAMPLE

03-3 현황 분석

수집 및 선택 데이터에 대한 현황분석
데이터를 시각화를 통해 설명(데이터간의 관계 등)

출생 연도에 따른 인구 피라미드(2020년 기준)



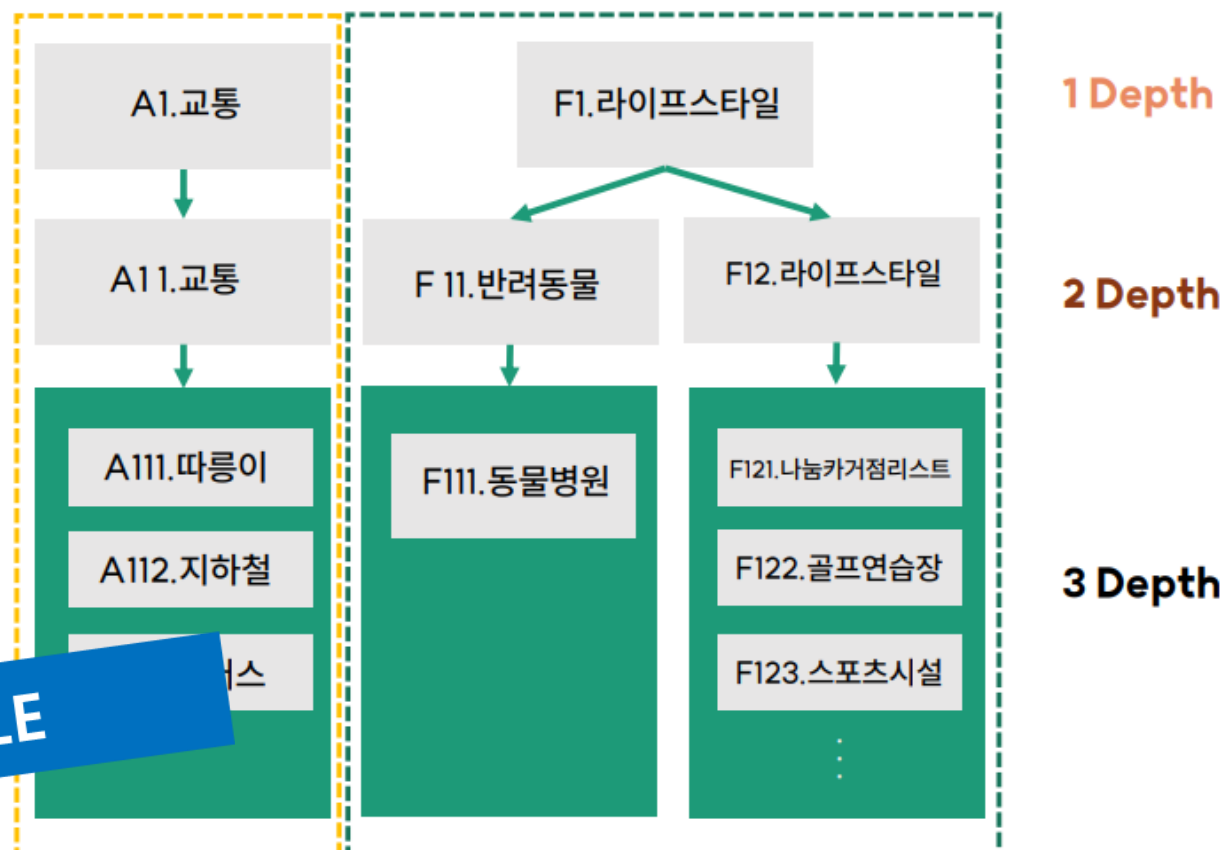
03-4. 데이터 전 처리 및 탐색

데이터 구조 변환 내용 설명

관리가 용이한 카테고리 형태
: 각 객체의 카테고리가 변형되고 확장될 수 있는 가능성을 고려하여 설계

CATE_C...	UPPER_CATE_CODE	CATE_NAME
A1	0	교통
A11	A1	교통
A111	A11	따릉이
A112	A11	지하철
A113	A11	버스
F1	0	라이프스타일
F11	F1	반려동물
F111	F11	동물병원
F12	F1	라이프스타일
F121	F12	나눔카거점리스트
F122	F12	골프연습장
F123	F12	스포츠시설
F124	F12	실내레저시설
F125	F12	헬스장

SAMPLE



03-5. Modeling

추천 시나리오 소개 후 추천 데이터 생성 과정 설명

데이터 활용에 대한 고찰

피쳐가중치

SAMPLE

카테고리 분류

03-5. Modeling

데이터 전처리 – 2. MZ 세대 트렌드와 피쳐 가중치



각 트렌드 키워드 끼리의 연관성을 찾아내어,
지역 추천시 사용자가
선택하지 않은 요소의 가중치도
악하여 추천 가능

SAMPLE



인스타그램 크롤링 활용

03-5. Modeling

추천 선택 process 및 가중치 적용 과정 설명

데이터 전처리 – 2. MZ 세대 트렌드와 피쳐 가중치

맥세권을 선택한 유저의 육아 가중치를 0.14 부여

	교통	방법 / 치안	건강	편의시설	교육	육아
교통	1		0.01			0.04
방법 / 치안		1				0.04
건강		0.02	1		0.02	0.07
편의시설	0.07				0.01	0.18
교육		0.01			1	0.035
육아					0.13	1
거주 MZ 세대 수	0			0	0	0
코리빙						
비건			0.01	0.02		0.04
키즈카페		0.003		0.003	0.02	0.61
공원	0.23			0.05	0.05	0.22
스타벅스	0.02					0.01
맥도날드	0.5					0.14
소음	0.11					0.04
안심택배				0.18		
실내레저시설						
헬스장				0.005		0.01
골프연습장				0.5		
나눔카	0.16					
반려동물						0.003

03-5. Modeling

속성별 가중치 예시

<그림 1-4> 청년층의 문화시설 선호도

Base	주변 편의시설 (1000)	만족하는 편의시설 (996)	성별		연령				세대구성			
			남 (497)	여 (499)	20대 (248)	30대 (250)	40대	50대 (49)	미혼/ 독립가구 (171)	기혼/ 무자녀 (86)	기혼/ 유자녀 (501)	미혼/ 캥거루족 (177)
편의점	93.4	63.5	64.6	62.2	77.2	48.8	56.9	73.4				
버스 정류장	93.6				50.9	50.0	65.1	71.2				
숲/공원	65.9				25.1	46.5	49.1	45.2				
지하철 역	48.3				43.3	50.0	39.9	39.0				
프랜차이즈 커피 전문점	71.5				35.1	30.2	34.7	45.2				
대형마트	40.4	31.5	29.8	33.3	28.2	30.0	31.3	36.5	28.7	33.7	32.3	32.2
병원	68.1	31.3	26.8	35.9	18.5	29.2	35.7	41.8	17.5	18.6	40.9	25.4
패스트푸드점	58.9	28.1	26.0	30.3	32.3	26.0	23.3	30.9	28.1	19.8	27.5	35.6
초/중/고등학교	82.3	25.2	22.7	27.7	6.5	20.8	47.8	25.7	3.5	9.3	41.9	7.9
재래시장/전통시장	35.6	21.7	20.3	23.0	9.7	20.0	25.3	31.7	10.5	32.6	25.9	16.9
영화관	27.3	18.7	19.3	18.0	16.1	20.0	16.5	22.1	17.0	23.3	19.8	18.1
어린이집/유치원	70.9	12.3	11.5	13.2	1.6	20.4	18.5	8.8	2.3	0.0	22.4	1.1
백화점	12.4	7.4	7.8	7.0	6.0	7.2	6.0	10.4	5.8	5.8	9.0	5.6
복합 쇼핑센터	13.4	6.8	7.6	6.0	6.5	6.8	6.8	7.2	7.6	3.5	7.8	6.8

SAMPLE

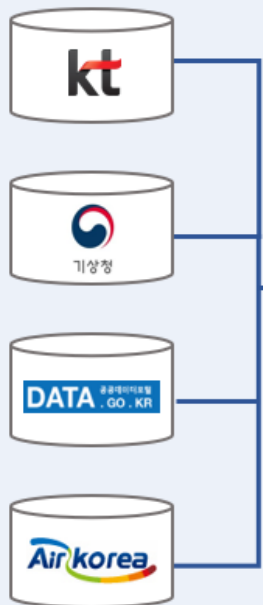
$$\text{편의점} * 0.44 + \text{카페} * 0.25 + (\text{대형마트} + \text{시장}) * 0.27 + \text{백화점} * 0.04$$

(논문 및 통계자료를 근거로 한 공식 추가)

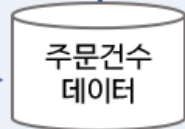
03-6 서비스 work-flow

2_Workflow

수집



병합
전처리



학습

Ridge

Random Forest

SAMPLE

모델링

업종별
예측 모델

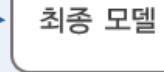
업종별
예측 모델

구현

DB



예측



예측값

저장

추출



WEB

웹 구현

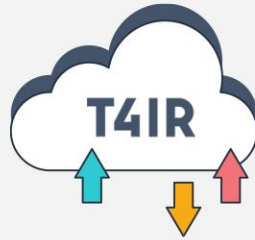
개발 서비스 소개

작 성 요 령

- ❖ [결론 및 향후과제] 현황분석 및 모델링에 대한 결과와 향후 과제를 제시한다
- ❖ 이전 서비스 활용방안으로 향후과제는 대체할 수 있음

작 성 요 령

- ❖ [느낀 점]은 프로젝트 수행에서 개인, 우리 팀이 잘한 부분과 아쉬운 점을 작성한다. 또한 프로젝트를 수행하면서 느낀 수행 상 어려움, 갈등요소 등을 작성하고 이를 해결한 방법을 작성한다.



THANK YOU

www.multicampus.com

(주)멀티캠퍼스



multicampus

Copyright by Multicampus Co., Ltd. All right reserved

