

안녕하세요, 임형빈입니다.



데이터에 기반한 문제해결 역량을 쌓아 왔습니다.

데이터 구축부터 탐색적 데이터 분석, 드릴다운을 통해 문제를 발굴하고, Process를 Inovation 하기 위한 인사이트 도출과 방법론을 제시하겠습니다.

인공지능을 활용한 기존 프로세스 개선에 관심이 있습니다.

ML/DL 모델링, RAG 및 Agent 등 급변하는 AI 트렌드에 적응하며, 사용자 경험을 개선할 수 있는 프로세스 개선에 기여하겠습니다.

SKILL

Language

Python 하 중 상 SQL 하 중 상

Cloud & Server / DB

Linux AWS Dataiku Docker ElasticSearch MariaDB LanceDB

Framework

Langchain(Langgraph) FastAPI Transformers

Colaboration & Other Tools

Git Bitbucket Jira ERDCloud Postman Swagger Grafana Streamlit Gradio

OA

Excel PowerPoint

WORK EXPERIENCE

SK주식회사 AX 애커튼파트너스, AI PI Consultant

2025. 09. 01 - 2026. 02. 28 (6개월)

Boston Consulting Group, Research Analyst

2025. 03. 21 - 06. 17 (3개월)

MindsNCompany, DataScience Intern

2024. 03. 04 - 06. 28 (4개월)

PROJECT

SK Hynix SCM KPI 선행관리 고도화, KPI 기반 모니터링 시스템 구축

SK주식회사 AX 애커튼파트너스, AI PI Consultant (25.09~26.02)

BCG: LLM Application

Bonston Consulting Group, Research Analst (25.03~06)

멀티모달을 활용한 국내 관광지 추천 시스템 설계 및 구현

2024 국민대학교 캡스톤디자인 경진대회 우수상 (2025.01)

예술의 전당의 고객만족도, 공공성, 수익성 증진을 위한

좌석 재분할에 따른 효과적 가격 모델 수립

2023 빅콘테스트 정형데이터분석 어드벤스드리그 최우수상 (2023.12)

HONORS & AWARDS

2024 국민대학교 캡스톤디자인 경진대회 우수상

2025.01, 국민대학교

국민대학교 기후변화대응 비즈니스 아이디어 공모전 대상

2024.10, 국민대학교, 인문사회융합인재양성사업 환경컨소시엄사업단 (HUSS)

2023 빅콘테스트 정형데이터분석 어드벤스드리그 최우수상

2023.12, 과학기술정보통신부, 예술의전당

2021 병영문학상 수필 부문 가작

2021.12, 국방부 장관

EDUCATION

국민대학교

AI빅데이터융합경영학과, 데이터사이언스융합전공 학사 (3.69 / 4.5)

2019 - 2025

ACTIVITY

Proact0 (가짜연구소 소속 오픈소스 프로젝트) CrewMember

2025. 03 - 2025.08

인공지능 연합동아리 투빅스, 21기 부원

2024

교내 인공지능 학회 X:AI, 5기 운영진

2024

교내 빅데이터 학회 DnA, 학회원

2022 - 2023

CERTIFICATE

빅데이터분석기사

2025.07

ADsP (데이터분석준전문가)

2023.09

OPIc IM3

2024.09

SK Hynix SCM KPI 선행관리 고도화, KPI 기반 모니터링 시스템 구축

SK주식회사 AX 애커튼파트너스, AI PI Consultant (25.09~ 26.02)

OVERVIEW

SK hynix SCM KPI 선행관리 체계 고도화

SCM Value Chain의 판매·생산 계획 KPI를 선행 관리 지표로 재설계
KPI와 재무 성과간 정량적 관계 분석을 통해 경영지표 활용도 강화

KPI 기반 모니터링 시스템 구축

설계 KPI의 지속적 관리 체계 구축
Nexacro기반 System UI 설계
Raw Data → 모니터링 결과값 간의 데이터 모델 설계

GOAL

계획 수립 및 관리에서의 선행성 확보
SCM Value Chain 내 계획 연계성 및 투명성 강화
KPI, 재무 성과 간 연결 구조 정량화를 통한 S&OP 의사결정 지원
모니터링 시스템 구축을 통한 지속적 그룹 SCM 성숙도 관리

활용 기술 (OA)

 Python  SQL  Nexacro  Excel  PowerPoint

ROLE

- KPI 현황 분석 및 Pain Point 도출
- 데이터 기반 KPI 설계
- 현업 및 개발 대상 커뮤니케이션 담당
- KPI - 재무지표 관계 분석 (회귀 / ML)
- KPI 모니터링 시스템 UI 및 데이터 모델 설계

CONTRIBUTION

KPI 선행 관리 체계 설계

- 판매계획정확도 / 생산계획준수율 AS-IS 분석 및 TO-BE 설계
- 단일 시점 계획 → 다기간 (과거 N월) 계획 기반 KPI 구조 전환

CEO KPI 선정

- KPI vs O/I 재무지표 (영업이익 / FCF) 간 회귀, ML 분석을 통한 기여도 계수 산출
- 월 별 KPI 1% 변화에 따른 재무 영향도 계수 (N% 효과) 산출
- 조직 SCM 성숙도 판단의 기준 지표

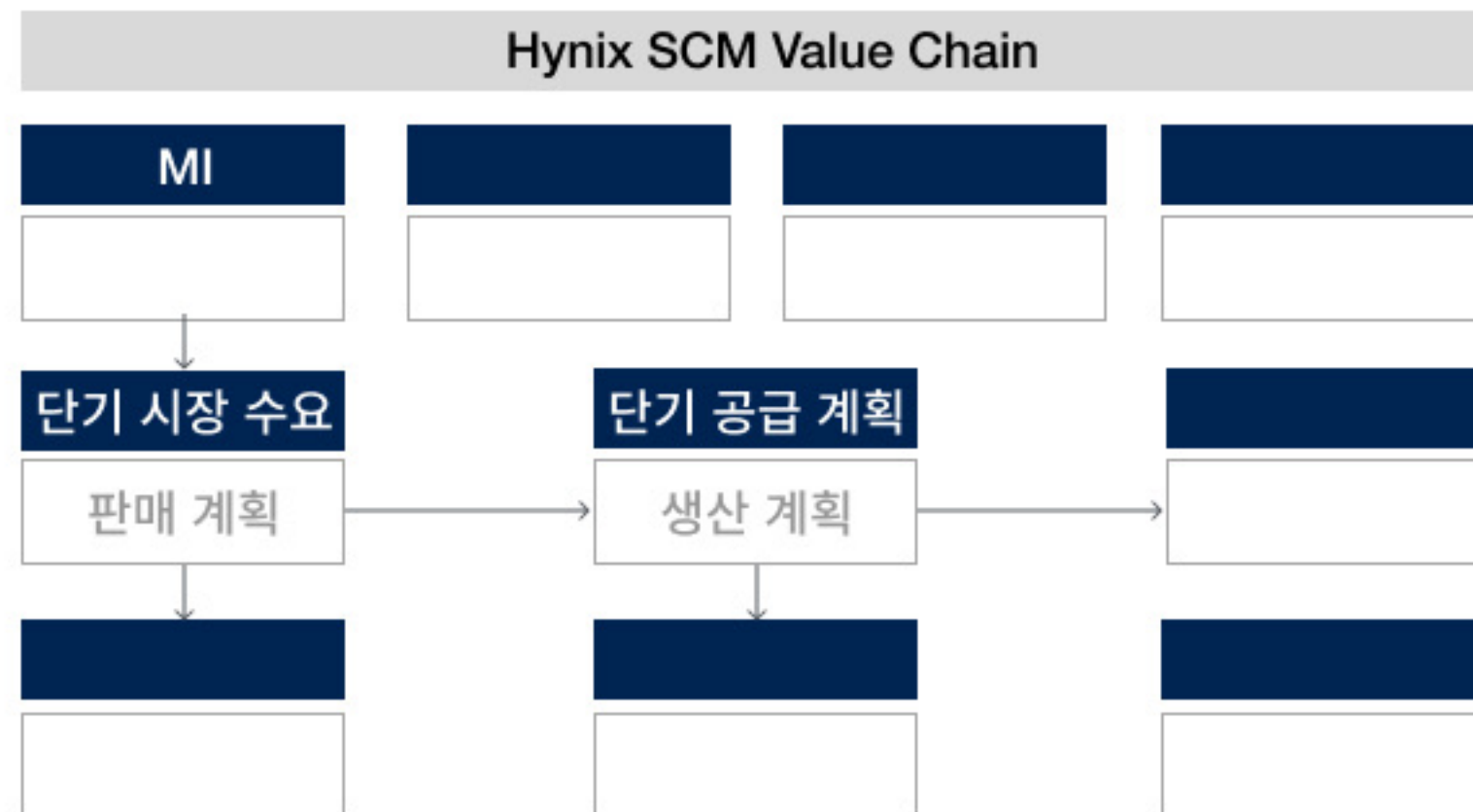
지속적 KPI 운영 체계 구축

- KPI 산출 Data Flow 설계
- KPI CEO 보고를 위한 지표 산출 자동화 (Raw Data → KPI)를 통한 산출 시간 90% 단축
- KPI 모니터링 시스템 UI / Data Model 설계

SK Hynix SCM KPI 선행관리 고도화

판매계획 정확도, 생산계획 준수율 설계

Hynix SCM Value Chain에서 MI 이후 수립 및 관리되며,
이후 단기 공급 및 월별 생산, 자재 등의 S&OP 운영에 활용되는
판매 계획과 생산 계획에 대한 Process Inovation을 목표로 진행



PROBLEM

- 당월 계획 수립 시점에 당월 실적의 약 50%가 포함되어 선행관리 지표로 활용이 제한
- 기존 생산 계획 KPI의 경우 달성률로 관리되었기 때문에 최근 N년간 Up turn 환경에서 대부분의 제품 초과 생산 발생 및 100%가 넘게 측정되어 활용도가 떨어짐.
- KPI가 사용률이 저조한 내부 SCM 시스템에서만 공유되어 Value Chain 내 타 부서 활용도 낮음

APPROACH

당월 계획 → 2~4개월 계획 구간으로 확장

- 판매 계획의 경우 수립 시점이 Value Chain 내 앞단에 위치하므로, 이후 Chain과의 연계성 강화를 위해 A,B 공정 투입 시점인 3개월 전 시점으로 확장
- 생산 계획의 경우 이전월 계획을 포함하여 산정

산출 로직

공정 투입 중요도 기반 월별 가중치 설정, 최종 지표는 월별 가중합으로 산정

- 판매계획정확도 = $\text{계획판매량} / \text{N월 계획}$
 - i. 판매 계획과 실적을 비교해 계획판매량 산출
 - ii. 특정 거래선의 경우 분기별 관리되어 분기초 이상치로 발생 → Moving Average로 관리
- 생산계획준수율 = $\text{준수율} / \text{N월 계획}$
 - i. 생산 계획대비 실적의 준수량 산출
 - ii. Up Turn 기간임을 고려해 무투자 증산의 초과 생산의 경우 N%의 Tolerance로 보장하도록 설계

UTILAZATION

KPI 공유 체계

- Value Chain 내 연관 부서인 영업 기획, 제조 운영, 전략 기획, S&OP 등 대상 조직 선정 후 월 2회 산출 자료 송부

KPI Drill-down 분석

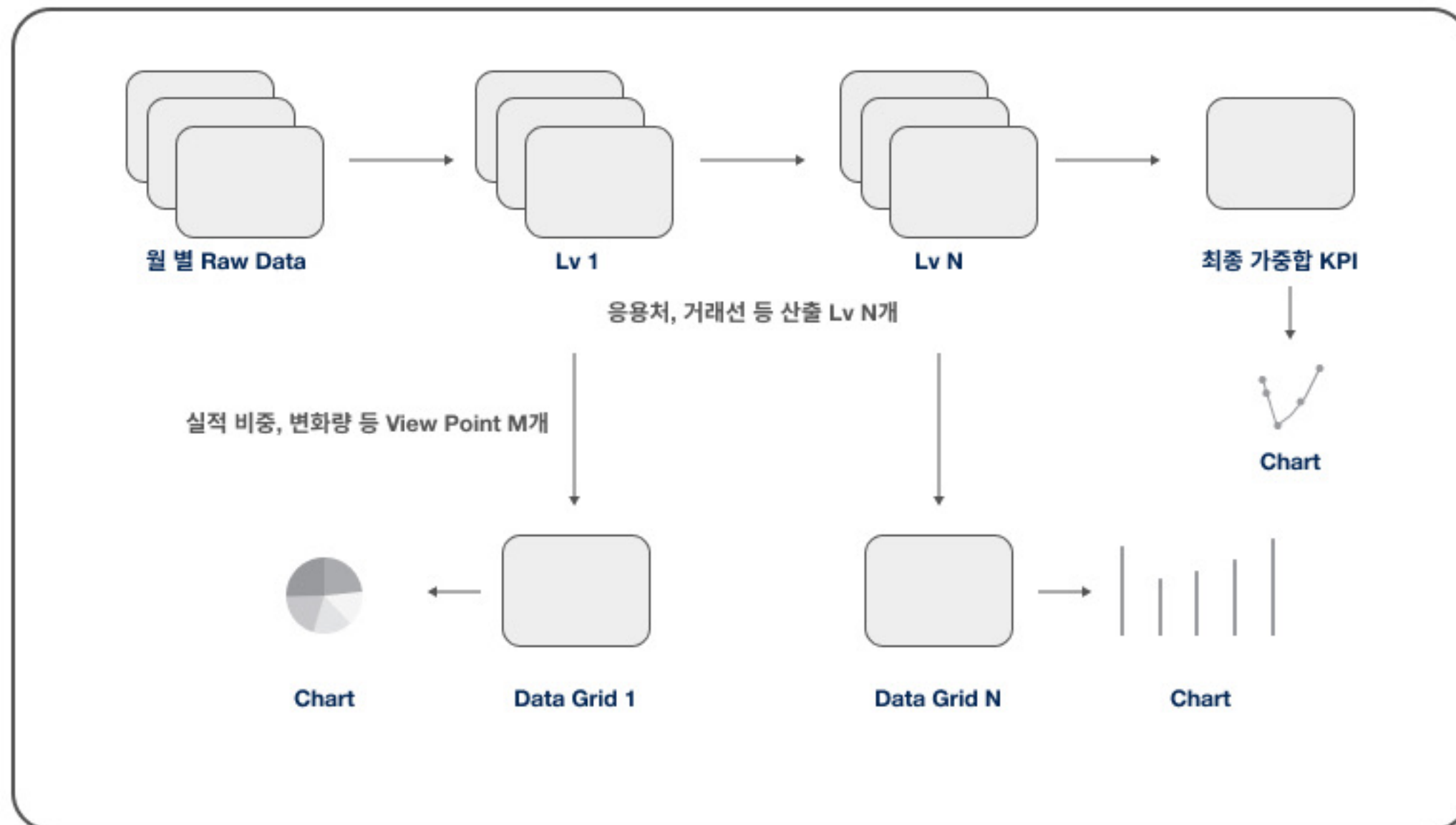
- 월별 Trend와 응용처, 거래선, 제품군 등 N개 Lv에서의 관리 Point 설정
- 실적 비중, 계획 대비 변화량 등 M개 Lv에서의 Top/Bottom N 제품 산출, 분석 후 사유 송부

KPI와 재무 지표(FCF, 영업이익) 간 정량적 관계를 산출하여 S&OP 운영 기준으로 활용

- KPI와 재무 데이터간 회귀, 머신러닝 분석을 통한 기여도 계수 산출 (KPI 1% 변화 시 재무 효과 N%)

KPI 기반 모니터링 시스템 구축

KPI Calculation Flow



UI DESIGN

설계 KPI에 대해, 지속적 관리를 위한 기존 Nexacro 기반 SCM System 내 UI 설계서 작성

조회부, 차트부, 그리드부 3개 구조로 설계

- System 내 Query 분석을 통해, 데이터 산출을 위한 GROUP BY 조건 리스트업 후 조회부 설계
- KPI 별 Trend 추세선과, Lv 별 관리 Point에 따라 차트 설계
- 조건에 따라 KPI 별 표기가 필요한 필드 정의 등 Data Grid 설계

UI 개발을 위한 Data Model 설계

- KPI Data Model을 개발 전 상세 설계 산출물로 전달
 - Raw Data Table → Data Grid, Chart 까지의 Data Flow 정의
 - Table 및 Field 명, GROUP BY 및 ORDER BY 조건을 포함한 명세서 작성

CALCULATION AUTOMATION

Raw Data → KPI 까지의 산출 자동화(w/ Python, Excel)

- CEO KPI 자료 및 부서 송부용 자료 제작을 위해 데이터 산출 자동화 양식 제작
- 부서에 KPI 시뮬레이션을 위한 양식 제작
- Data Model 설계 시에 참고, 부서 별 KPI 산출 설명회에 참고
- 기존 수기 작업 대비 산출 시간 90% 감소 (약 4일 → 30분)

BCG: LLM Application

Bonston Consulting Group, Research Analst (25.03~06)



활용 기술

- Python SQL AWS Dataiku Docker Grafana Streamlit
- Langchain(Langgraph) FastAPI MariaDB LanceDB
- Git Bitbucket Jira ERDCloud Postman Swagger

ROLE

- AI 개발 (Agent, RAG) Model Deploy (Dataiku, EC2) FastAPI 개발
- 앱 로그 데이터 분석, Grafana 대시보드 문서작업 (API 명세, 온보딩, RDB Schema)

OVERVIEW

MR 업무 지원을 위한 콜 활동기록, 제품정보 조회, 일정 관리, AI 챗봇 기능 개발에 참여
제품 공정 엔지니어의 기계 일탈 대응에 도움을 줄 수 있는 MFG Chatbot 개발에 참여

CONTRIBUTION

담당한 Agent Chain의 Test Case 성능 평균 90% 달성

- RDB에서 정보를 추출해 답변을 생성하는 Chain, MR 활동 기록 정보 추출 및 요약 Chain 외 5개의 Chain을 개발
- Chain 별 Test Case 증강 및 Test Code를 작성해 고도화
- 데모 런칭 시 고객 정성평가 기준 평균 90%, 최대 95%의 성능 달성

고객사 내부망과 모델 서버인 Dataiku와 AWS EC2 환경에 Deploy

- 기능별, Chain 별 응답을 FastAPI로 구현해 Ops에서의 Deploy를 담당
- BE, FE 담당자와의 원활한 소통을 위한 API 명세를 작성

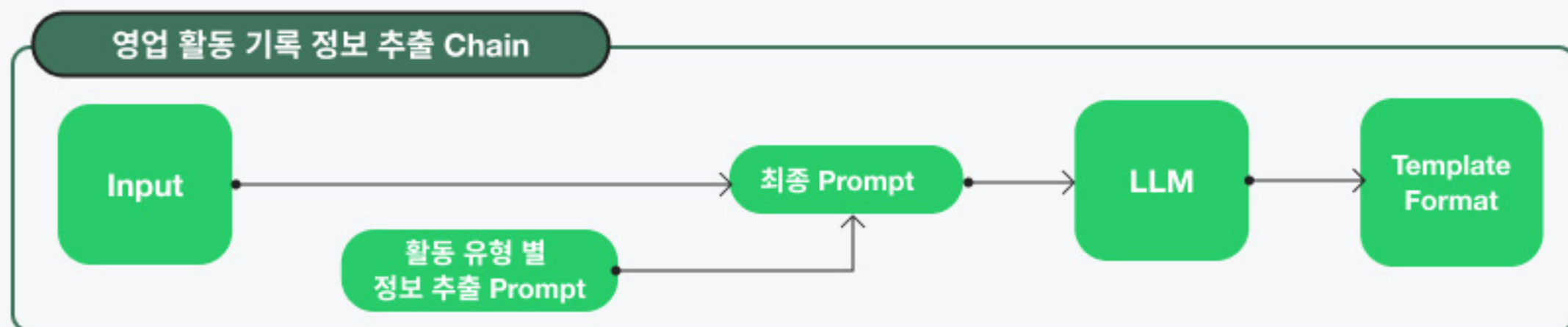
기능별 사용되는 문서에 대해 RAG Pipeline 구현

- 공정 메뉴얼, 제품 메뉴얼, Compliance 등의 문서를 AWS Textract API를 통해 파싱 및 청킹해 S3에 적재
- Chain 별 검색되는 문서의 수, Reranker에 차등을 줌으로써 기능별 시간 및 성능 요구사항을 만족

BCG: LLM Application

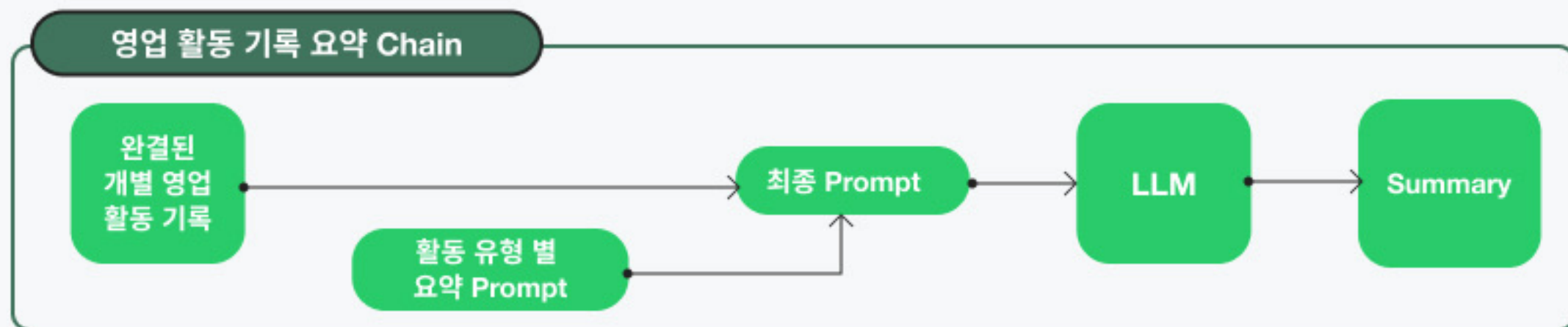
AGENT 개발

- RDB에서 정보를 추출해 답변을 생성하는 Chain, 영업 활동 기록 정보 추출 및 요약 Chain 외 5개의 Chain 개발
- FastAPI 코드 및 API 명세를 작성해 BE, FE와 모델 서버를 연동
- Pydantic을 통해 LLM Output을 Json으로 통제해, Agent 및 BE와의 연결에서 에러 최소화
- 요구사항을 만족하기 위해 실험을 통해 Token 수, 성능과 시간의 Trade-Off를 조절



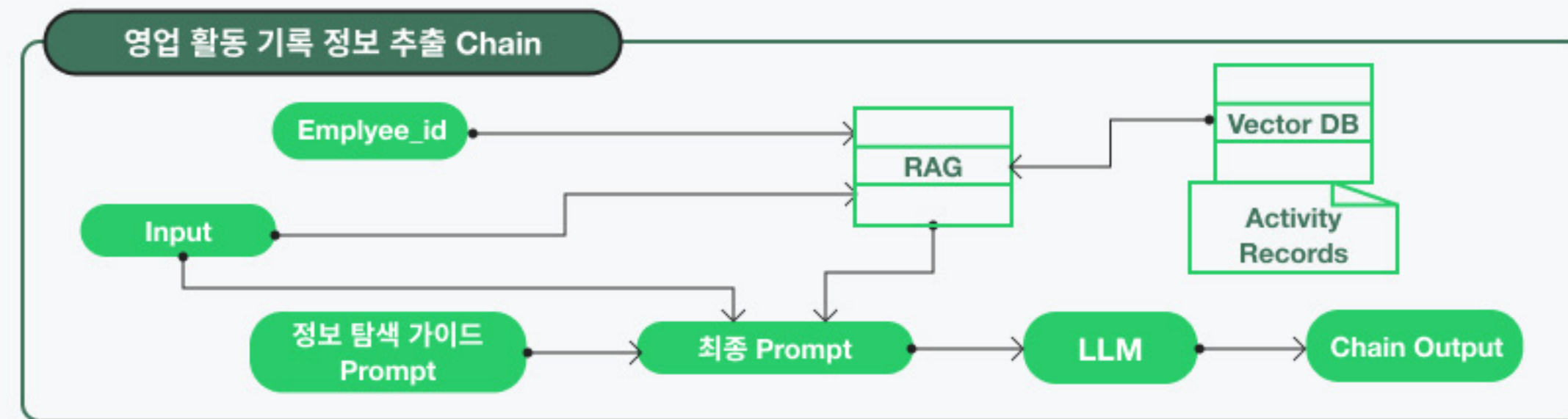
고성과자의 암묵지를 반영한 영업 활동 기록 정보 추출 포맷 정의

다양한 사용자의 영업 기록을 통일된 포맷으로 정보를 추출 및 요약해야하는 요구사항을 반영하기 위해 고성능자의 영업기록들을 타겟으로 포맷을 정의했고, 암묵지 또한 반영했다는 평가를 받았습니다.



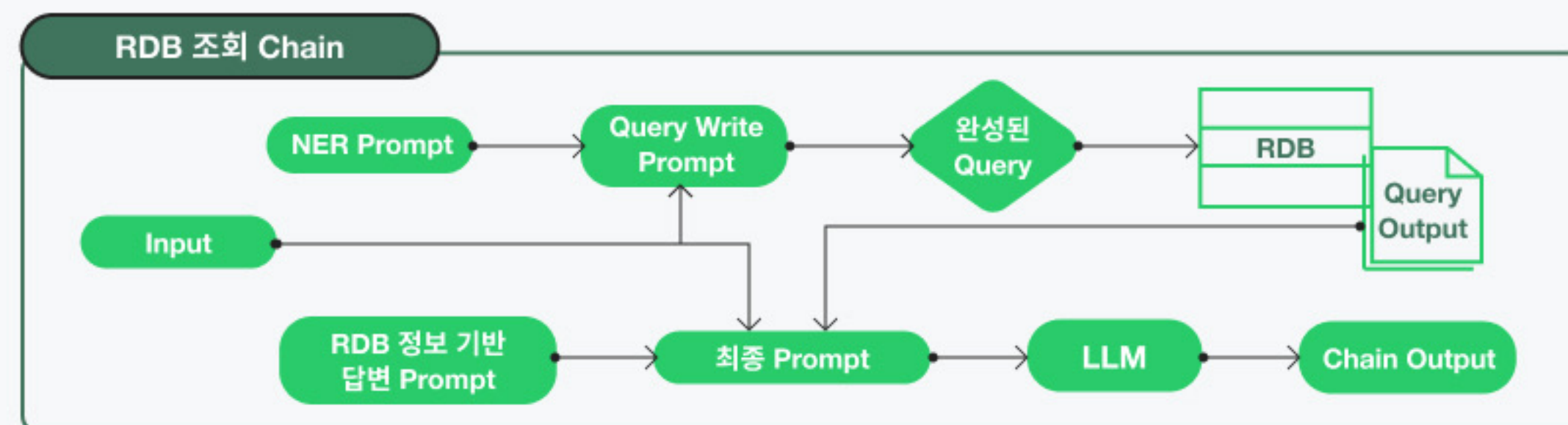
TEST CASE 및 고성능자의 영업기록 데이터를 바탕으로 영업 종류 별 프롬프트를 작성

약 7개의 영업 종류 별 입력되는 정보가 달랐기에, 데이터를 바탕으로 일반화된 프롬프트를 작성했습니다.



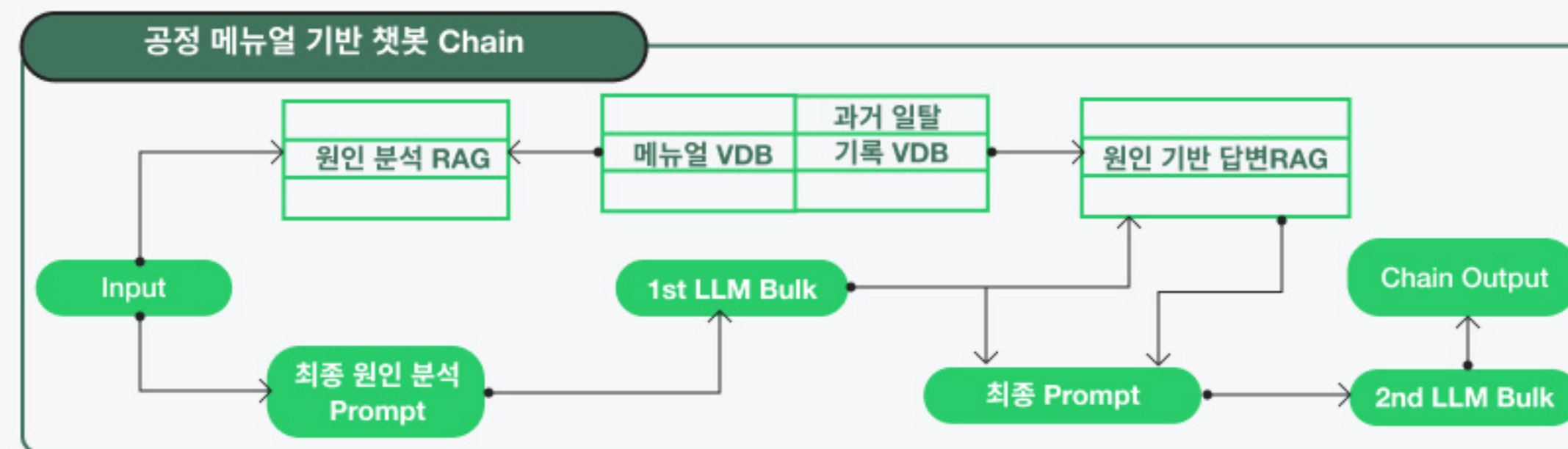
EMPLOYEE ID 기반 VDB 조회 (RAG 검색 필터링)

FE에서 EmployeeID를 입력받아, 유저의 질문을 토대로 적재된 영업 기록 정보를 조회하도록 구현



NER 기반 QUERY 포맷 생성 및 RDB TABLE 명 추출

- Test Case를 바탕으로 쿼리 시나리오를 작성, 인풋에서 NER을 진행해 쿼리의 파라미터를 입력받아 RDB를 조회하도록 구현
- Token 수를 줄이기 위한 시나리오 별 쿼리 포맷을 최적화



다중호출을 통한 N개의 원인 분석, N번의 RAG를 통해 레퍼런스를 탐색하는 COT 구조

- 유저 질문에서 N개의 원인을 분석하고, 원인 별 메뉴얼 VDB를 검색해 레퍼런스를 탐색하도록 구현
- 체계적인 원인 분석과, 각각에 해당하는 레퍼런스에 기반한 최종 답변을 통해 성능을 고도화

BCG: LLM Application

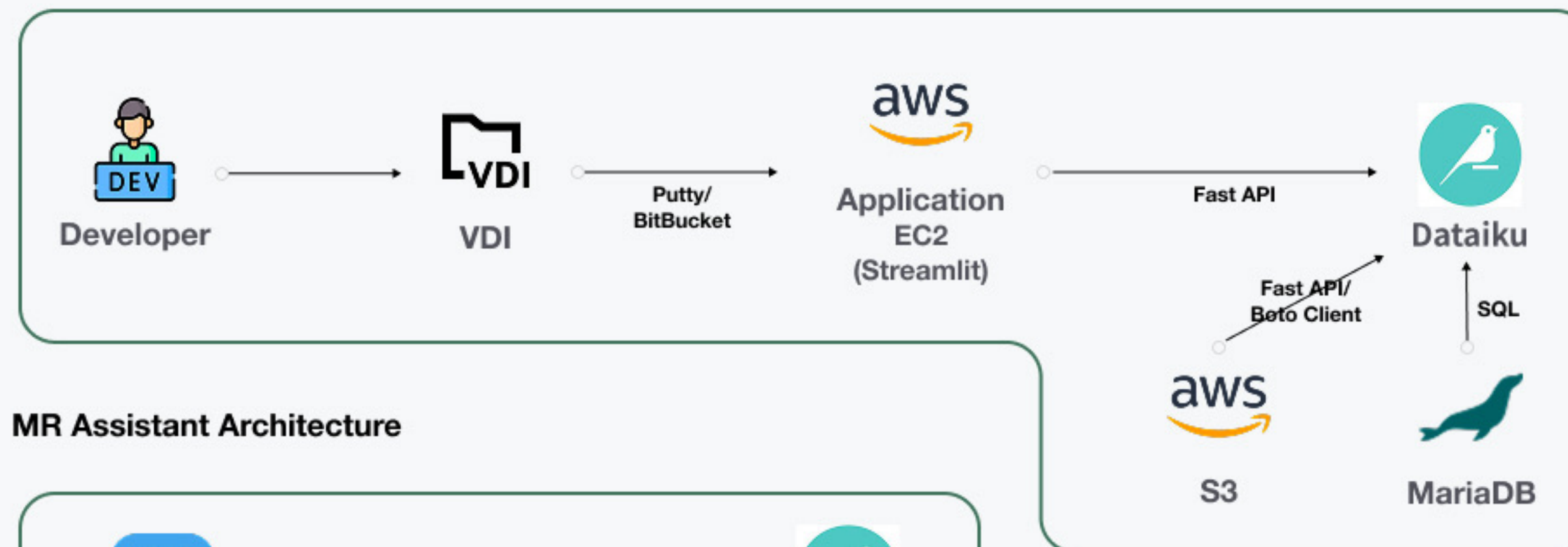
DEPLOY

- 고객사 Model Server인 Dataiku, EC2에 Deploy를 담당
- Model Server와 F/B 연결을 위한 FastAPI 개발 및 API 명세 작성을 담당
- Application 및 기능 별 AWS Key 관리를 위한 Config 및 모듈 파일을 작성 및 관리

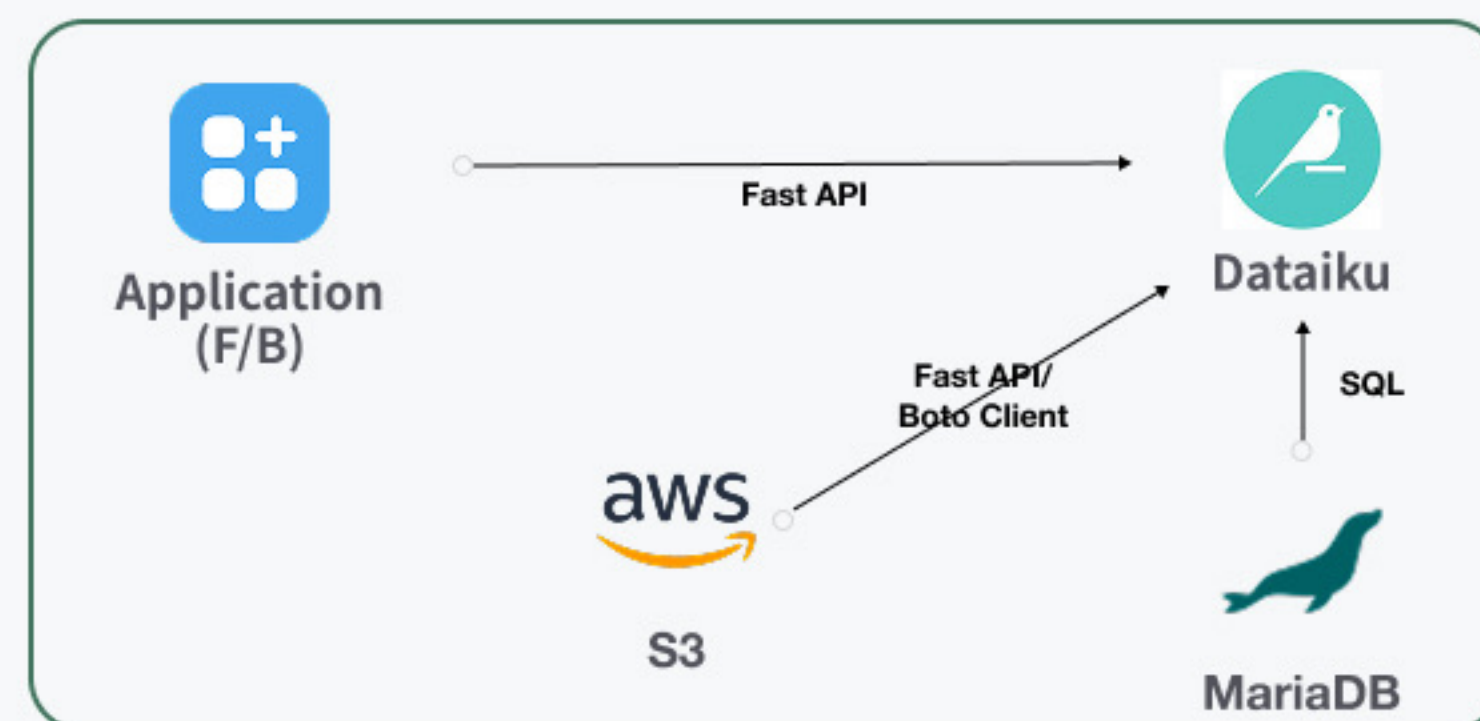
QA(JIRA, GRAFANA DASHBOARD)

- 데모 런칭 이후, 그랜드 런칭을 위한 QA 작업에 참여
- Jira 피드백 티켓 별 QA를 통해 프롬프트 엔지니어링, 구조 변경 등을 통해 고도화
- 일부 티켓에 대해서 DevinAI를 사용해 QA 자동화를 진행
- BE 적재되는 로그 DB 모니터링 및 그랜드 오픈에서 사용할 대시보드 구현에 참여
- Chain 별 QA를 위한 적재되는 로그를 확인할 수 있는 Query 작성해 Table 추가
- Table 결과값들을 사용해 QA를 위한 Test Data 증강 진행

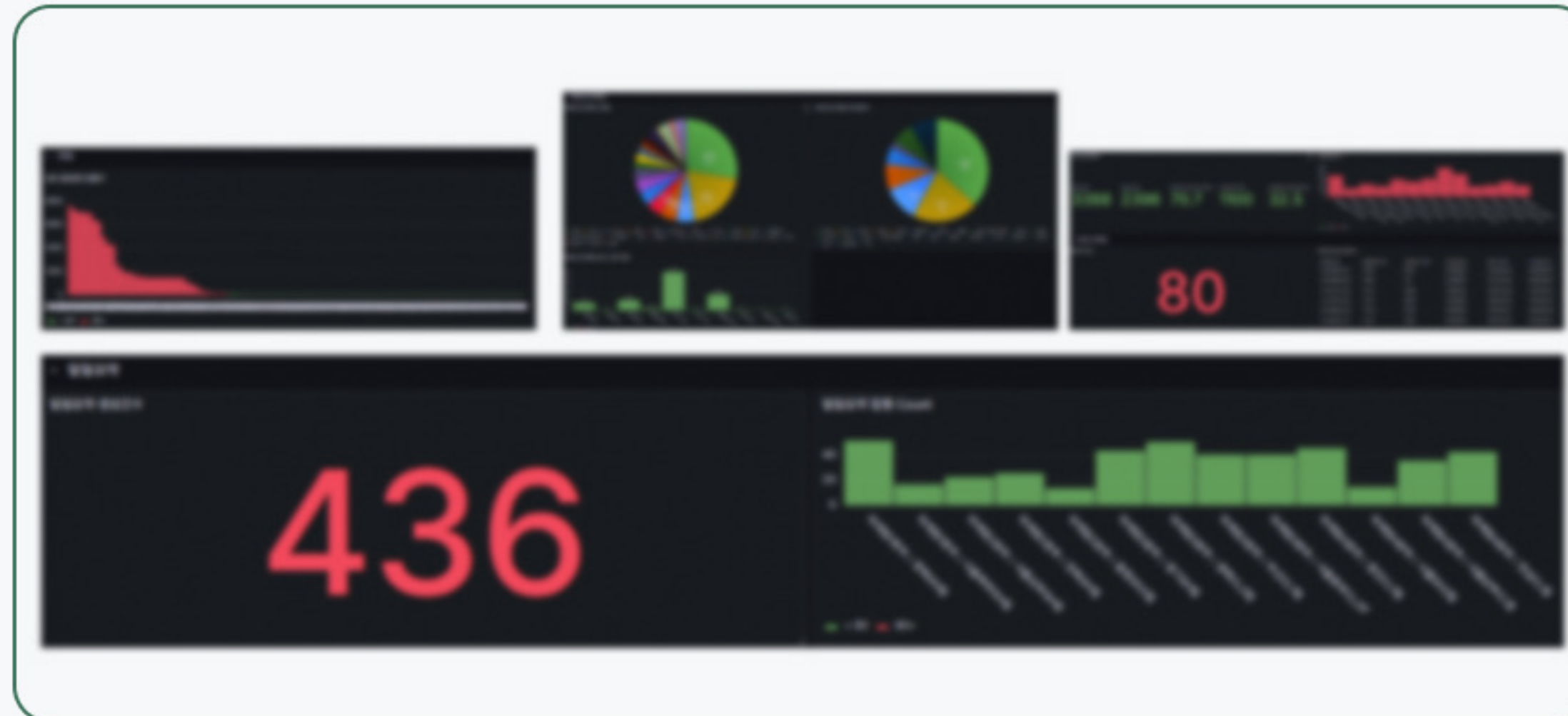
MFG Chatbot Architecture



MR Assistant Architecture

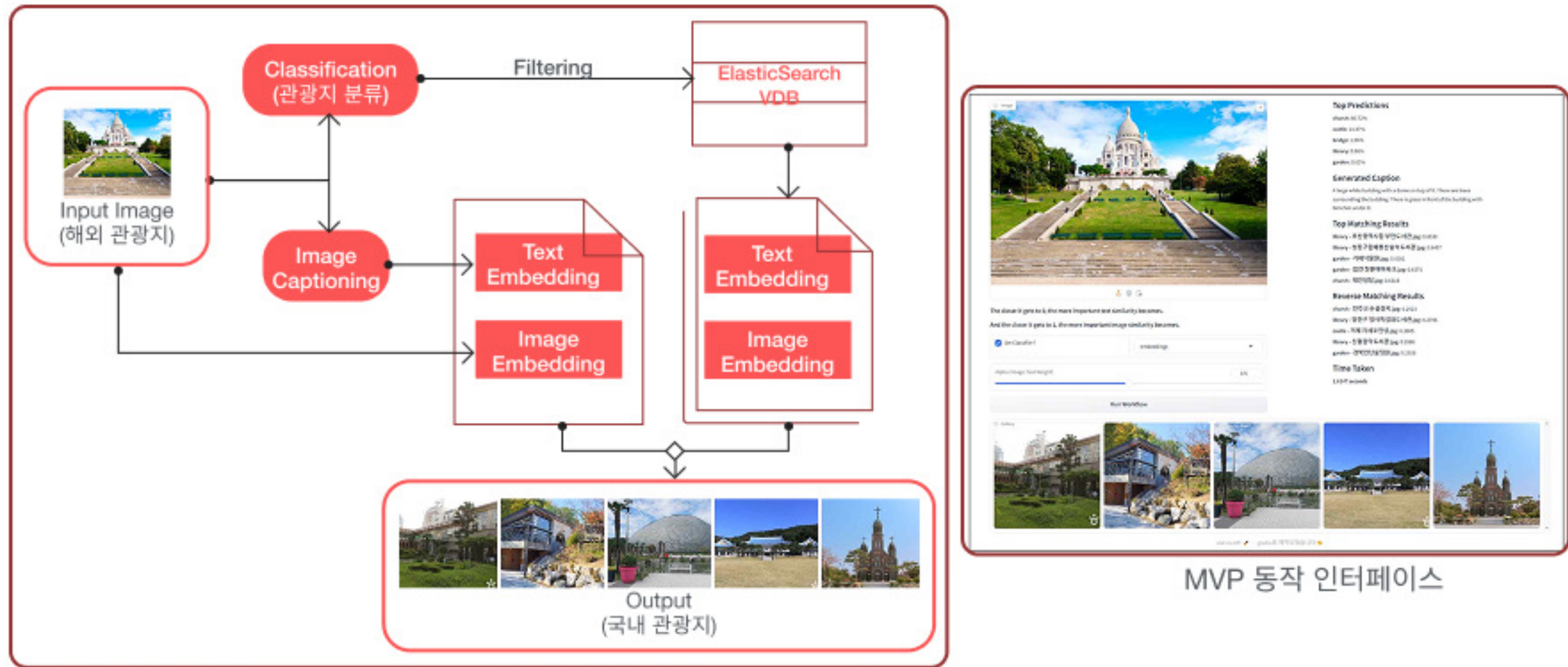


Grafana Dashboard



멀티모달을 활용한 국내 관광지 추천 시스템 설계 및 구현

2024 국민대학교 캡스톤디자인 경진대회 우수상 (2025.01)



관광지 추천 시스템 아키텍처

활용 기술

- Linux
- Huggingface Transformers
- Pytorch
- ElasticSearchDB
- Gradio

OVERVIEW

멀티모달 기반 추천 시스템 도입을 통한
관광지 추천 알고리즘의 현 기술적 한계 개선 및 관광지 추천 만족도 개선

ROLE

- 이미지 캡셔닝 모델링
- 관광지 추천 도메인 리서치
- 관광지 이미지 데이터 수집 (TourAPI, GoogleLandmark)
- Gradio 데모 제작

CONCLUSION

관광지 추천 산업에 멀티모달 추천 시스템을 도입해, 실질적 문제를 해결함으로써
딥러닝 모델링 프로젝트 전 과정을 체험

- Torch DDP를 이용한 분산 학습, Huggingface Transformers를 사용한 커스텀 파인튜닝 코드를 제작해 사용함으로써 코드레벨에서의 딥러닝 프로젝트 경험
- 관련 연구 및 서비스를 분석해, 데이터에 기반한 도메인 속 문제를 정의
- 단일 모달리티 시스템이 주인 관광지 추천 도메인에서 이미지-텍스트 기반의 멀티모달 추천 시스템을 도입해 사용자 경험을 개선

멀티모달을 활용한 국내 관광지 추천 시스템 설계 및 구현

기존 관광지 추천 알고리즘의 문제점

1. 현존 관광지 추천 서비스 및 연구는 단일 모달리티 기반의 시스템만 존재

- 관광지 추천은 시각적 특징과 문화 및 역사적 배경, 분위기 등의 맥락적 정보가 복합적으로 고려 필요

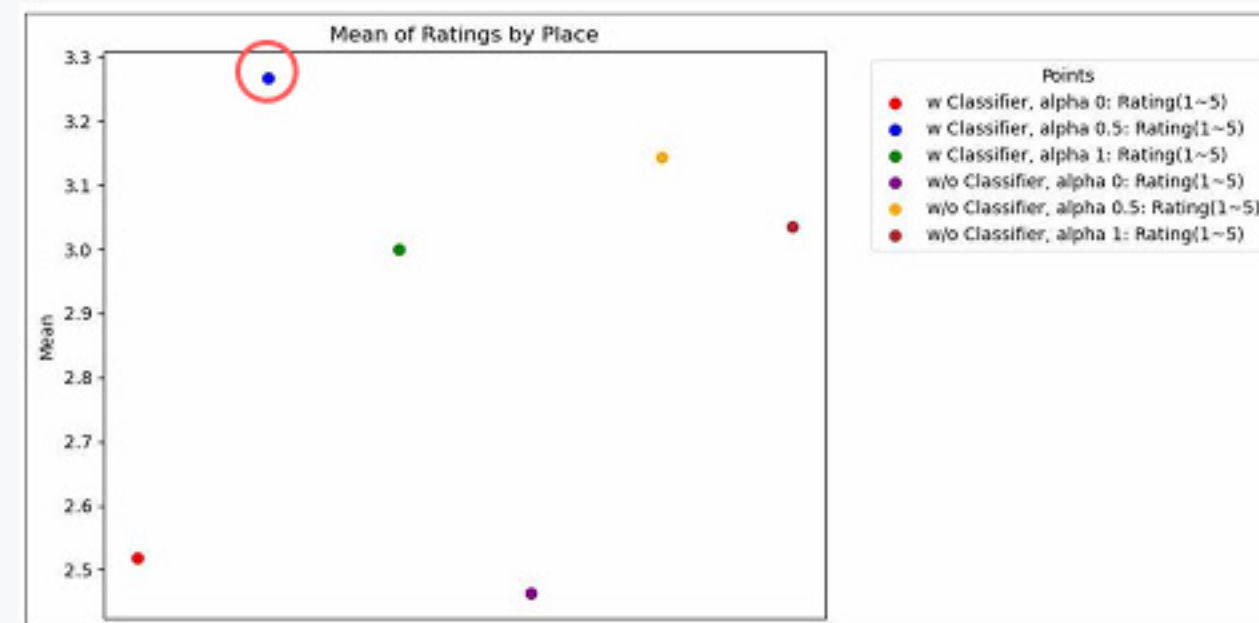
2. 관광지 이미지는 여러 객체에 대한 이해가 필요

- 장면은 객체에서 장면의 각 부분 → 재료 → 질감 순으로 세부화되어 묘사 필요

“ 관광지 이미지 정보와 이미지에 대한 세부적인 묘사(텍스트 캡셔닝) 두가지 모달리티의 접근을 통해 관광지 추천에서의 사용자 경험을 개선할 수 있을 것 ”

	COCO-Caption	Localized Narratives
	"A motorcycle with its brake extended standing outside"	"In this image, theres grass on the ground, in the middle there is a bike which is in black color, in the background there are some green color plants and there are some homes."

Model	Florence2-COCO (Ours)	Florence2-Flickr	Florence2-ft	BLIP-image-captioning-large-COCO	Bottom-up-attention (Localized Narratives Baseline-standard captioning)
Bleu1	0.439	0.328	0.272	0.341	0.322
Bleu1	0.285	0.211	0.149	0.227	-
Bleu1	0.188	0.139	0.082	0.154	-
Bleu1	0.125	0.092	0.043	0.104	0.081
METEOR	0.181	0.158	0.122	0.170	-
ROUGE_L	0.341	0.316	0.232	0.343	0.317
CIDEr	0.336	0.214	0.124	0.266	0.293
SPICE	0.232	0.211	0.136	0.234	0.257



데이터셋 선정

- 다양하게 등장하는 객체 정보를 학습할 수 있도록, 풍부한 텍스트 캡션을 가진 Localized Narratives를 선정
- COCO dataset 대비
 - 같은 이미지에 대해 생성되는 캡션 양이 약 200% 증가
 - 주 객체 뿐 아니라 이미지의 배경, 세부 객체에 대한 표현 수집

모델링 실험

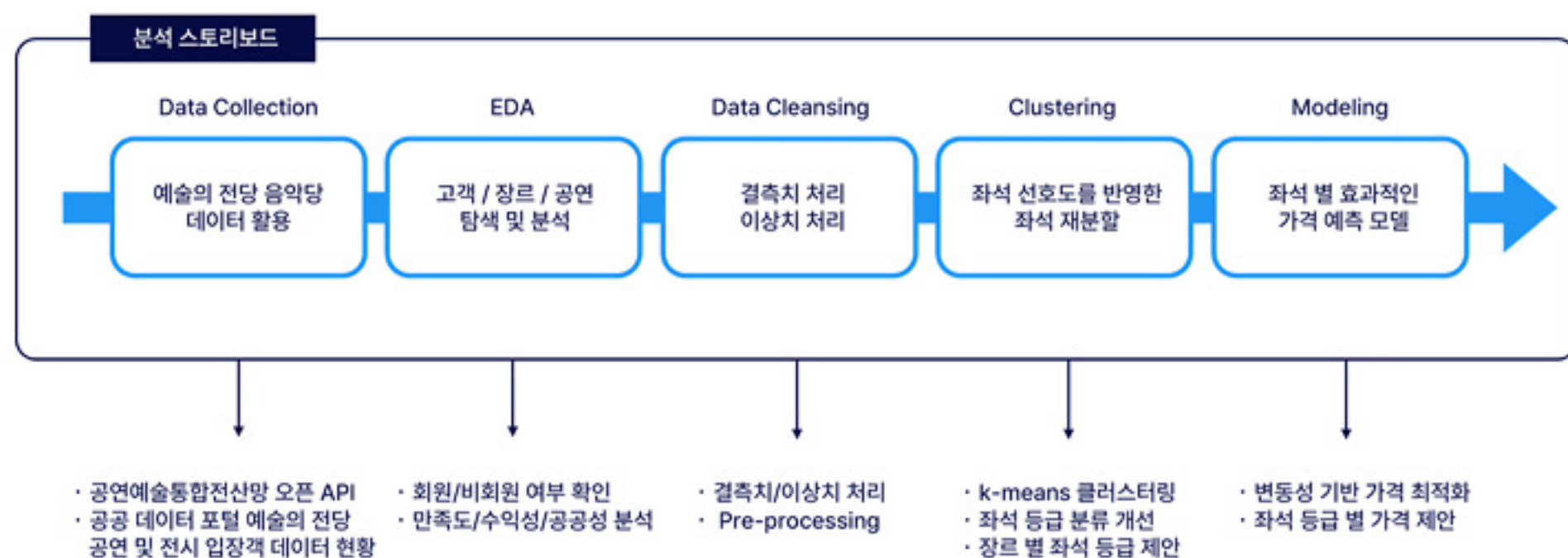
- 관련 연구를 참조하여, Visual Grounding을 잘 표현할 수 있는 Florence2와 비교군으로 BLIP2를 선정
- Bleu 외 총 8개의 평가지표를 사용해 생성된 텍스트의 문법적, 의미적 성능을 다각화하여 측정
- Torch DDP를 사용해 분산학습을 반영한 파인튜닝 코드를 작성

결과 분석

- Localized Narratives의 Baseline 대비 정량지표 평균 21% 개선**
Bleu1 36%, Bleu4 54%, RougeL 8% 등 8개 지표 중 6개의 지표에서 가장 높은 성능을 기록
- 정성평가 설문 결과, 최고 점수로 멀티모달 관광지 추천 방법론의 사용자 경험 개선 효과를 입증**
분류기(w/o), 유사도(이미지, 이미지-텍스트, 텍스트) 총 6가지 방법론에 대한 30명 대상 정성평가 설문 결과, 평균 3.27(5점 만점)로 가장 높게 나타남

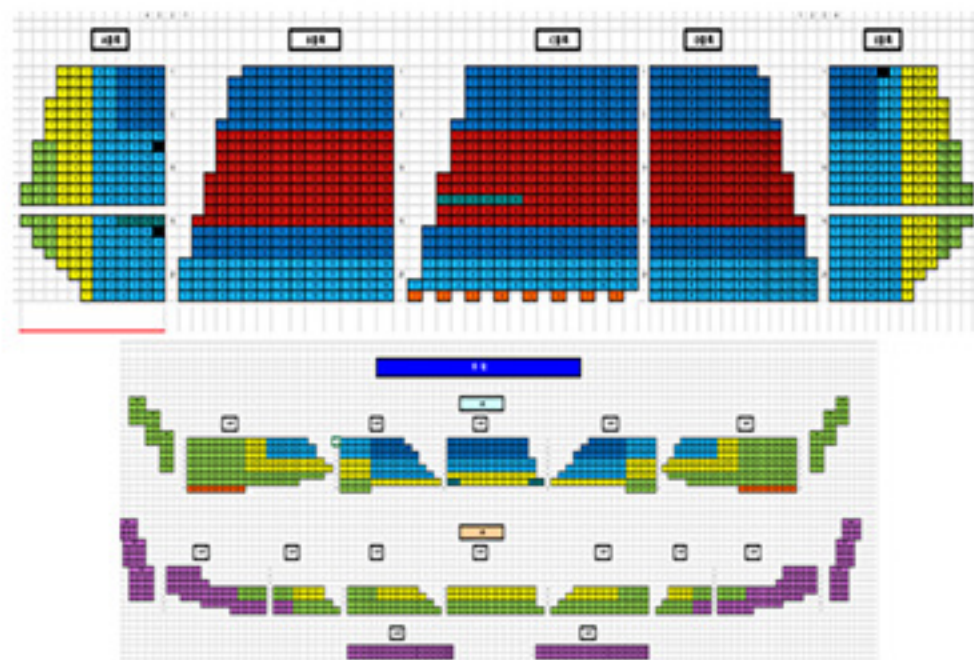
예술의 전당의 고객만족도, 공공성, 수익성 증진을 위한 좌석 재분할에 따른 효과적 가격 모델 수립

2023 빅콘테스트 정형데이터분석 어드벤스드리그 최우수상 (2023.12)

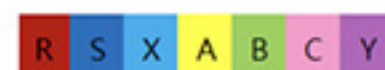


프로젝트 파이프라인

독주 예시



제안한 독주 장르 좌석 재분할 예시



OVERVIEW

클래식 공연 활성화를 위한 예술의전당 효과적 가격 모델 수립

ROLE

- 좌석 클러스터링
- 수익성, 공공성 기반 피쳐 엔지니어링
- EDA 및 도메인 리서치
- 가격 ML 모델링
- 외부 데이터 수집(OpenAPI, 공공데이터포털)

CONCLUSION

전체 13개 장르 중 11개에서 만석 기준 최대 15%의 수익성 향상

- 원가에 대해 고객 만족도, 수익성, 공공성 기반의 파생변수 기반 가중치를 계산 후 최종 가격을 설정
- RMSE, R2 등의 평가지표를 기준으로 회귀 모델링을 통해 가격 모델을 설명
- 고객만족도, 수익성, 공공성의 3가지를 만족할 수 있는 가격 모델을 제안

획일화된 장르별 좌석 재분할을 통해, 기존 좌석안 대비 최대 4 종류 증가

- 전체 13개 장르에 대하여, 5~7 종류의 획일화된 좌석 재분할을 제안
- 클러스터링을 활용한 좌석 재분할을 통해, 장르별 특징이 고려되지 않은 기존의 문제점을 개선
- 좌석 선택의 다양성을 증진함으로써 클래식 공연에서의 고객 경험을 개선 시도

활용 기술



Python

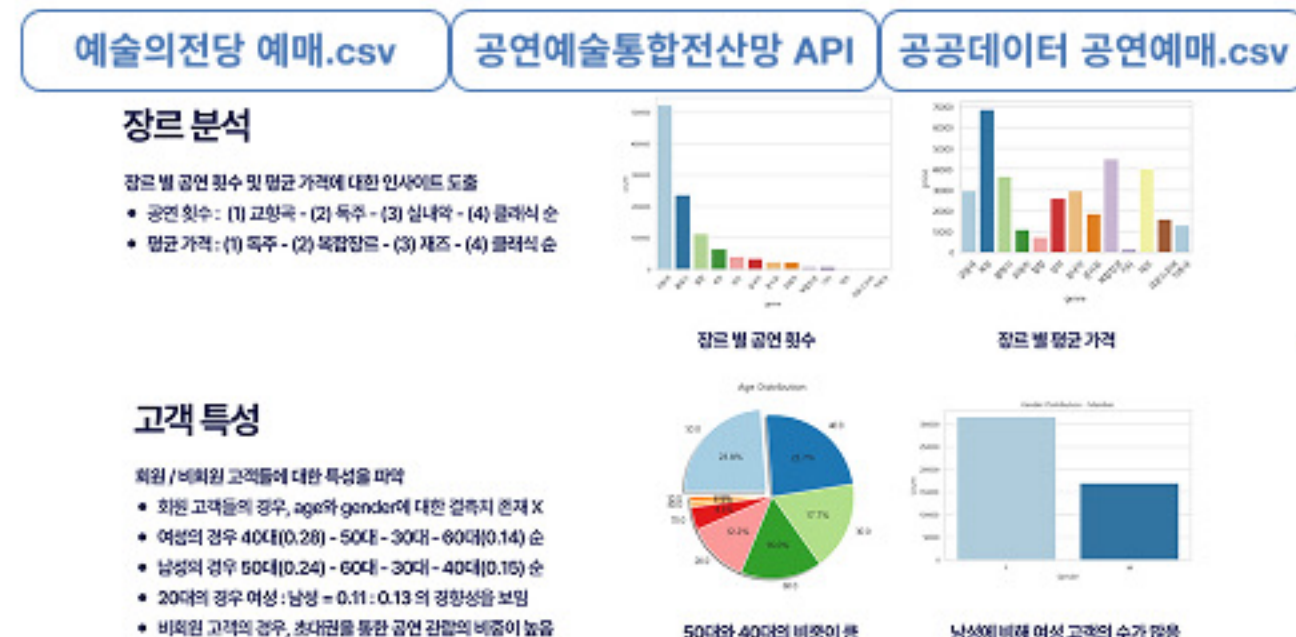


Git

예술의 전당의 고객만족도, 공공성, 수익성 증진을 위한 좌석 재분할에 따른 효과적 가격 모델 수립

기존 예술의 전당의 PainPoint 도출 및 문제 해결 방안

EDA



인사이트

1. 예술의 전당 공연들의 기존 좌석 등급에 대한 보편적 기준이 모호.

- 기존 좌석안에 장르별 특징이 고려되지 않음
- 표준좌석등급(3~5)를 따르지만, 비슷한 규모의 해외 공연장 대비 좌석의 선택 다양성이 부족

2. 공공성을 띄는 예술의전당 특성 상, 수익성 증진을 위해 여러 요소가 고려됨.

- 가격 모델이 수익성을 만족하며, 공공성 또한 만족해야 한다고 판단
- 예술 공연 가격 수립 도메인 조사 및 수요 증진을 위한 고객 만족도라는 요인 또한 고려

접근 방식 도출

“ 기존 가격안에 대한 수익성 분석 결과,
‘장르’에 따른 좌석 분류 및
가격 모델 수립이
수익성 증대에 효과적일 것 ”

클러스터링



- 목적
가격 모델 수립에 개인별 좌석에 따른 선호도 반영
장르에 구분없는 보편적 좌석 등급 산정
- 기법
관련 연구 참고, FINCH와 K-means 클러스터링 2가지 기법을 통한 실험
Yello bricks 기반 주관적 군집 수 설정 한계 개선 및 설명 용이성으로 K-means 선정
- 기준 변수
좌석의 위치 정보(열,블록 등)와 선호도 기반(판매 속도 및 구매 빈도 등) 파생변수 생성

가격 ML 모델링

- 고객만족도, 수익성, 공공성을 나타낼 수 있는 피쳐 생성
공연 취약 계층 비율, 장르별 평균가, 재관람여부 등 기존 데이터로부터 엔지니어링 수행
- Boosting 계열 모델링 (LGBM, XGB, Catboost Regressor)**
 - 원가에 대해, 파생변수 기반 가중치를 부여 후 최종 가격 설정
 - RMSE, R2 등의 회귀 평가지표 사용
 - R2 0.88 등 지표가 가장 높고, 고차원 데이터에서 강건한 LGBM을 최종 모델로 선정

결과 분석

- 장르적 특징과 고객 선호도를 반영한 좌석 등급 재분류**
 - 기존 표준좌석등급분류(3~5) → 장르별 등급 재분류(6~7)를 통한 좌석 다양화
- 좌석 등급 별 객관적 가격 모델 수립**
 - 원가에 대해 고객 만족도, 수익성, 공공성 기반의 파생변수 기반 가중치를 계산 후 최종 가격 설정
 - RMSE, R2 등의 평가지표 기준으로 회귀 모델링을 통한 가격 모델 설명
 - 전체 13개 장르 중 11개에서 만석 기준 최대 15%의 수익성 향상