



SPARTA

빅데이터 기반 품질관리(QA, QC) 양성과정 교육 과정 소개서

과정 개요



과정명

빅데이터 기반 품질관리(QA, QC) 실무 인재 양성과정

운영 목적

제조 산업의 디지털 전환에 발맞춰, 데이터 분석과 품질관리 역량을 교육현장에 즉시 투입 가능한 실무형 인재를 양성하여 산업 경쟁력 강화에 기여

대상자

공학계열 전공의 취업준비생 중

- 제조·생산·품질관리 분야에 관심이 있으며
- Python, SQL 등 데이터 활용 역량을 갖추거나 성장 의지가 있는 인재

운영 기간

약 5개월간 진행되는 온라인 기반 풀타임 집중 교육

09:00 - 21:00 / 일일 12시간 학습 진행

커리큘럼 개요



제조 공정 데이터 분석을 활용한 품질 관리(불량률 예측 및 이상 탐지)

통계 분석 기법

- 평균, 중앙값, 분산 등 기초 통계량 분석
- 제조 데이터를 활용한 가설 검정, 상관 분석 등 통계 분석 방법론
- 변수 간 상관관계 분석 및 단순/다중 회귀 분석 활용한 인과 관계 탐색
- 독립표본 T-test, ANOVA를 통한 집단 간 차이 분석 및 유의성 해석
- 품질 관리에 필요한 통계적 사고 및 해석 역량

대시보드 구축 & AI 활용 역량 강화

- Tableau, Streamlit를 활용한 대시보드 설계
- 주요 지표(KPI) 및 제조/생산 품질 데이터 시각화
- 시계열 데이터를 활용한 생산 관리 실시간 모니터링
- 생성형 AI를 활용한 품질 리포트 자동 생성



Python 문법 & 데이터 분석 기초

- 데이터 조회, 결측치 및 이상치 처리
- Pandas를 활용한 데이터 전처리
- Matplotlib을 활용한 기초 시각화
- 데이터 리터러시

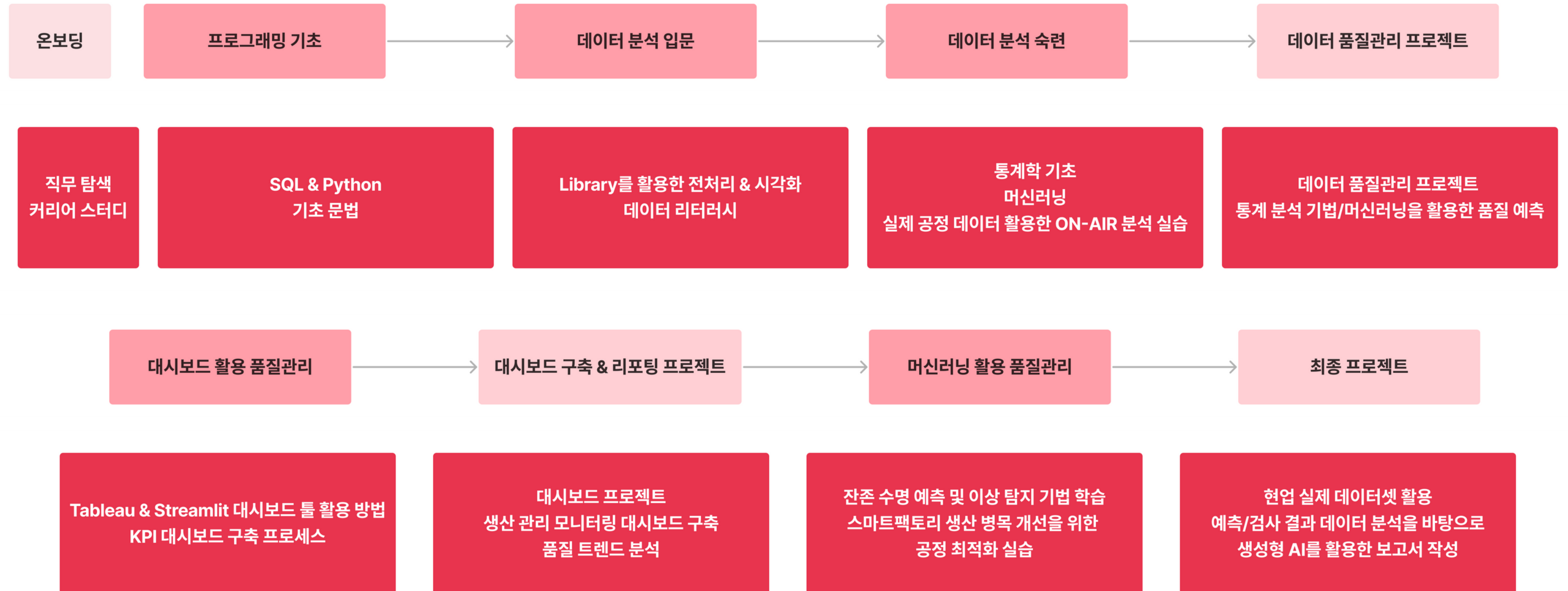
머신러닝 기반 데이터 분석 심화

- 시계열 센서 데이터를 활용한 불량 예측 및 이상 탐지
- 회귀, 분류, 예측 모델을 활용한 품질 예측 모델링
- 성능 지표 기반 모델 평가 및 하이퍼파라미터 최적화
- 모델 해석을 통해 제조 현장 의사결정에 활용 가능한 인사이트 도출

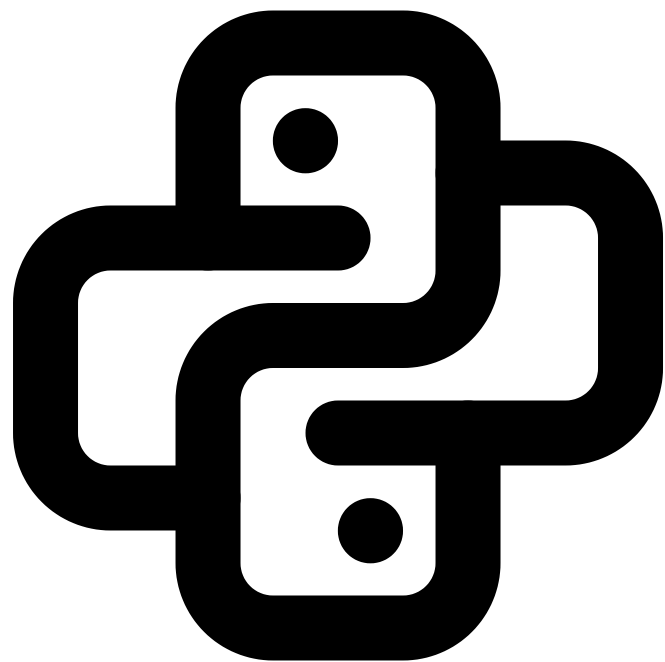
커리큘럼 개요



4번의 학습주차 개인과제 + 3번의 단기 팀 프로젝트 → 최종 프로젝트



프로그래밍 기초 → 데이터 분석 입문



Python 문법 기초 학습

자료형, 변수, 연산자, 조건문, 반복문, 함수 등 Python 기본 문법 학습

데이터 분석 라이브러리 활용

Pandas를 활용한 데이터프레임 생성, 조회, 결측치 및 이상치 처리 실습
Matplotlib과 Seaborn을 활용한 기초 통계 시각화 학습 및 실습

데이터 리터러시

데이터의 구조와 유형 이해, 변수와 관측치 구분 등 분석 전 필수 개념 학습
기초 통계량 분석(EDA)을 활용한 데이터 해석 역량 강화
데이터 기반 의사결정을 위한 사고 체계 및 관점 정립

데이터 분석 숙련 → 데이터 품질관리 프로젝트



통계 기반 데이터 분석 기초

평균, 중앙값, 표준편차 등 기초 통계량을 활용한 데이터 분포 및 특성 이해
IQR, Z-score 등을 활용한 이상치 탐지 및 품질 이상 데이터 식별
기초 통계 해석 역량을 바탕으로 데이터 기반 의사결정 능력 함양

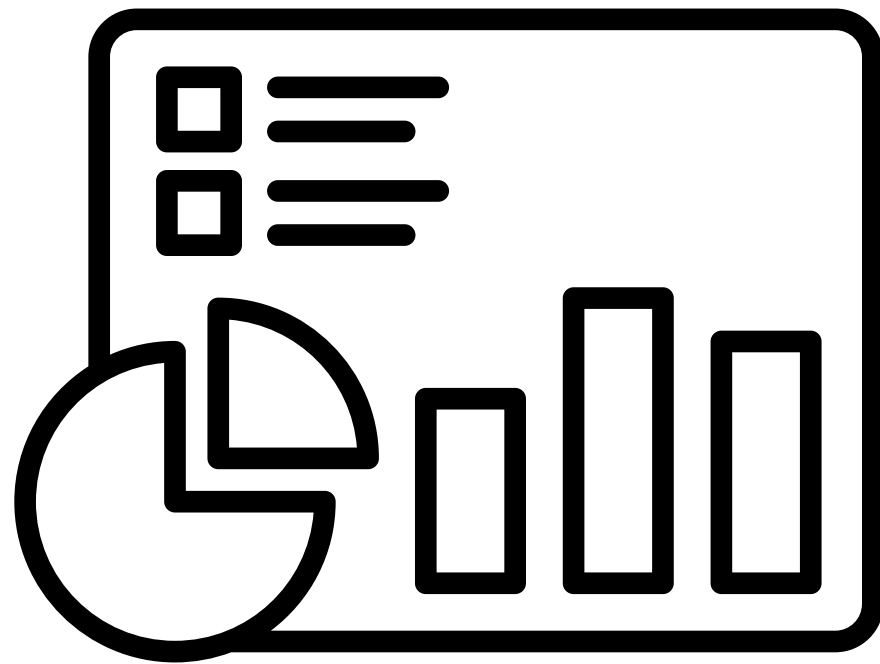
회귀 분석 및 가설 검정

공정 변수 간 상관관계 분석을 통해 선형 패턴 파악
단순/다중 회귀 모델을 활용하여 주요 변수의 영향 분석
T-test, ANOVA를 활용하여 그룹 별 통계적 차이 검증
분석 목적에 따른 적절한 가설 수립 및 유의성 검정 실습

머신러닝 라이브러리 활용 입문

지도학습, 비지도학습, 강화학습 등 머신러닝 개념과 차이에 대한 이해
군집(비지도학습)을 통해 그룹별 데이터 특성 파악
Scikit-learn을 활용한 예측 모델 구성 및 성능 평가 실습

대시보드 활용 품질관리 → 대시보드 프로젝트



대시보드 툴 활용 기초

Tableau 및 Streamlit의 사용법 학습을 통해 시각화 구성 요소 이해
시계열, 파이, 바 차트 등 다양한 시각화 방식 실습으로 표현력 강화
주요 지표를 효과적으로 보여줄 수 있는 대시보드 레이아웃 설계 학습

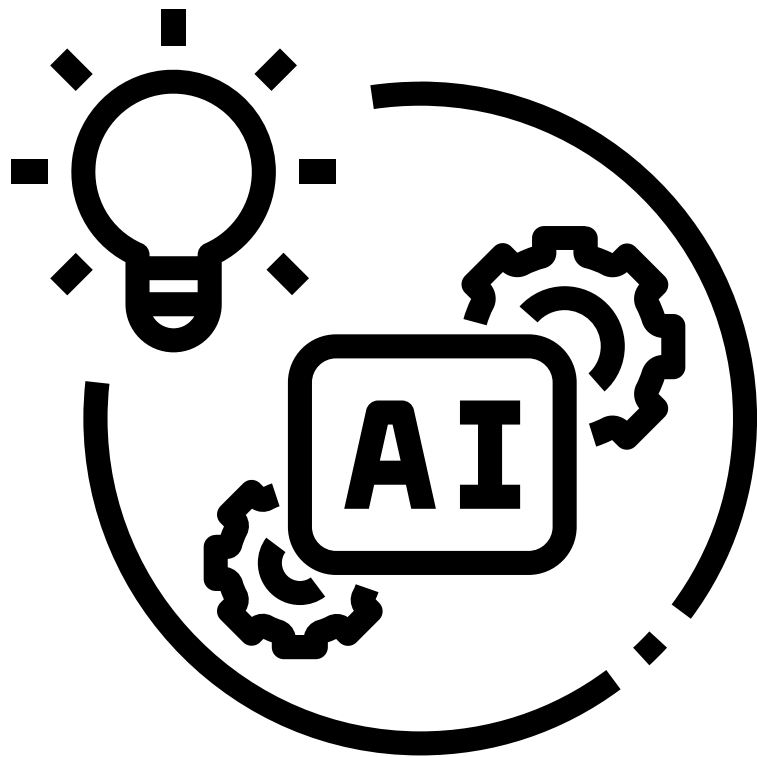
생산/품질 모니터링 대시보드 설계

생산 공정 데이터 및 품질 지표(KPI)를 기반으로 한 대시보드 기획
시계열 데이터를 활용한 트렌드 분석 및 이상 징후 시각화 구성
사용자 관점에서 모니터링이 용이한 정보 구조 설계 및 구현

생산/품질관리 대시보드 구축 프로젝트

실제 산업 데이터를 활용한 데이터 분석, 모델링, 대시보드 구축 등 전 과정 수행
Streamlit/Tableau를 활용해 사용자 맞춤형 필터링 및 지표 조회 기능 구현
제조 산업군 특화 인사이트를 공정 최적화를 지원하는 대시보드 완성

머신러닝 활용 품질관리 → 최종 프로젝트



머신러닝/딥러닝 기반 품질 관리 심화

시계열 센서 데이터를 활용한 이상 탐지 및 불량 예측 모델 구축
이미지 데이터를 활용한 딥러닝 기반 양품/불량 분류
목적에 맞는 ML/DL 모델 개발 및 성능 분석 및 최적화

품질 관리를 위한 고급 통계 분석

SPC(통계적 공정관리), FMEA(고장 모드 및 영향 분석), 6-Sigma 이해
ISO 9001, 공정능력 분석, MSA 등 필수 품질관리 도구 학습
제조 공정 데이터 기반 품질 개선 전략 수립

생성형 AI를 활용한 리포팅 자동화

Chat-GPT API를 활용해 텍스트 요약, 시각화, 이미지 생성 자동화 실습
Streamlit 기반 대시보드 내 자연어 데이터 분석 및 보고서 자동 생성 구현
Langchain과 RAG 구조를 활용한 문서 기반 질의응답 및 통합 리포트 작성 실습

주차별 커리큘럼

기타 - 단계별 팀 프로젝트

-  [제조] 제품 결함 데이터 분석
-  [차량] 친환경 차량 성능 데이터 분석
-  [의류] 섬유 가공 품질보증 데이터 분석
-  [생산] 세계 식량 생산 데이터 탐색
-  [에너지] 태양광 에너지 발전량 데이터

기초 프로젝트

-  [반도체] 공정 센서 데이터 불량 검출 분석
-  [스마트팜] 식물 생산 관련 데이터 분석
-  [모빌리티] 엔진 센서 데이터 기반 결함 예측 및 품질관리 분석
-  [제조] 도금욕 공정 데이터 기반 품질 예측
-  [제조] 다이캐스팅 공정 데이터 기반 품질 예측
-  [바이오] 약물 사용 데이터 기반 위험도 예측 및 분류
-  [바이오] 의료 데이터 기반 뇌졸중 발생 예측 분석

심화(데이터 품질관리) 프로젝트

-  [2차전지] 리튬 이온 배터리 데이터셋 : 잔존 수명 예측
-  [모빌리티] 전기차 센서 데이터셋 : 주행 거리 및 배터리 성능 예측
-  [바이오] 유전체 정보 데이터셋 : 품종 분류
-  [스마트팜] 축산 착유기 사용 현황 데이터셋 : 착유량 예측
-  [제조/광물] 실시간 광물 품질 데이터셋 : 농도 예측
-  [운송] 자재 이송 물류 데이터셋 : 지연 예측
-  [반도체] 웨이퍼 센서 데이터셋 : 이상 탐지
-  [제조/스테인리스] 강판 표면 결함 데이터셋 : 품질 예측

실전(품질관리 대시보드) 프로젝트

프로젝트 우수작 - 기초 프로젝트

기초 프로젝트에서는 분석 목적에 맞는 데이터 전처리와 변수 해석을 통해 통계 기반 인사이트 도출 역량을 기릅니다.

9조 : 태양광 발전소 최적의 위치 도출을 위한 기후 환경 요인과 에너지의 상관 관계 분석

정교한 EDA 및 시각화를 통한 데이터 해석

각 컬럼의 의미를 분석하고 히스토그램, 산점도 등을 활용해 변수 간 관계와 데이터 분포를 효과적으로 파악하여 논리적으로 설명

도메인 지식을 활용한 파생 변수 설계 및 인사이트 도출

'GHI'와 같은 파생 변수를 설계하여 타깃 변수와의 상관관계를 규명하고, 분석 목적에 부합하는 의미 있는 인사이트를 도출

외부 데이터 연계 및 분석 주제의 확장성

기상청 데이터를 직접 수집·활용하여 기존 분석을 '국내 태양광 발전 최적 지역 분석'으로 확장, 주제의 적용성과 프로젝트 완성도를 높임

구구절절 잘하조

01 주제 소개



Q-Energy 데이터 분석가

태양광 발전을 통해 친환경 에너지를 생산하는 기업

환경 데이터를 활용하여 발전량과의 관계를 분석하고
성능 개선 방안을 도출

환경 데이터 기반 발전량 분석 프로젝트

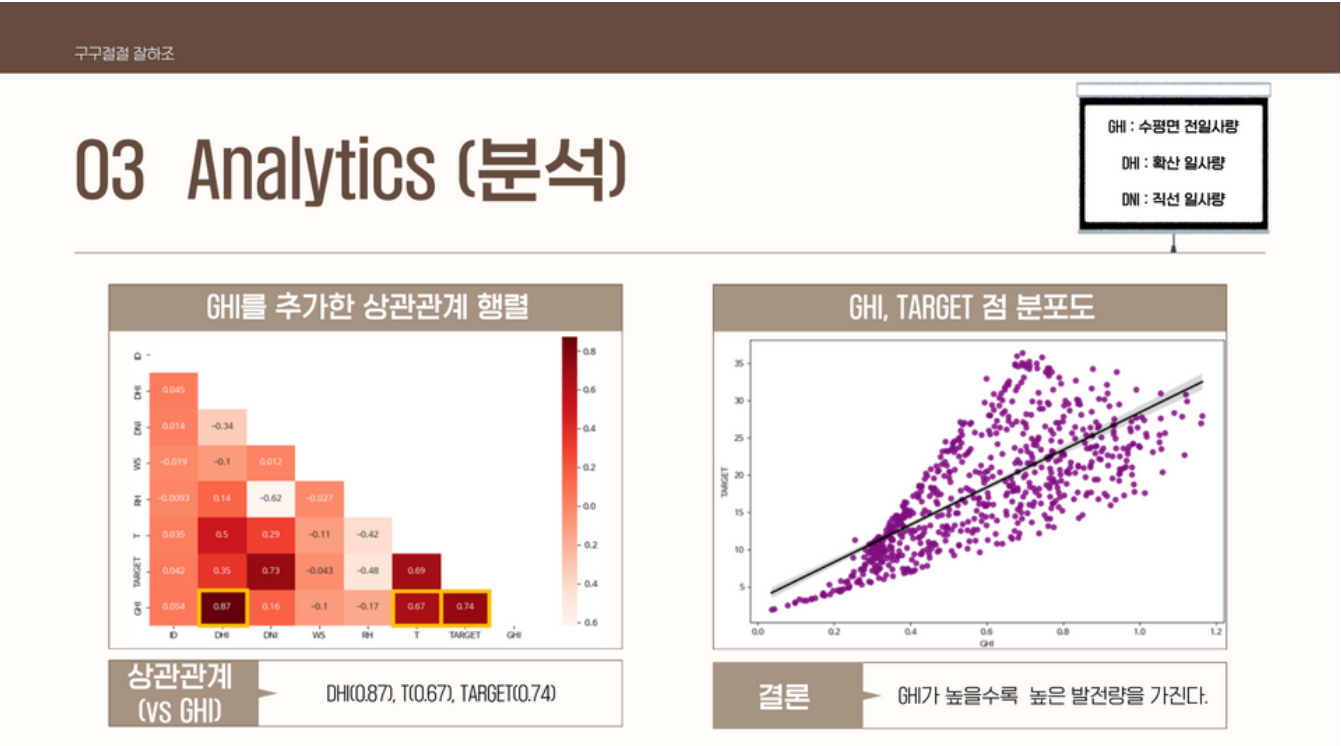
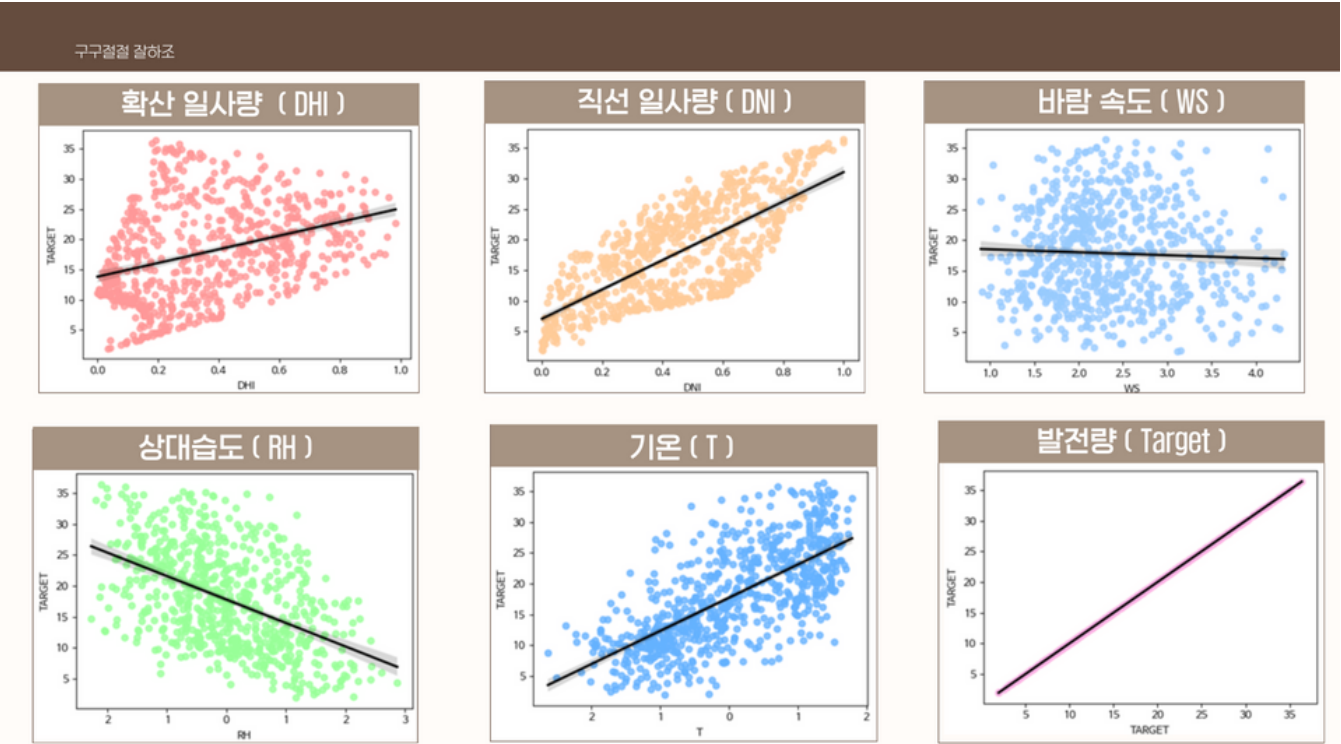
태양광 발전이란 ?

- 태양의 빛 에너지를 전환시켜 전기를 생산하는 발전기술
- 태양광 패널은 일반적으로 효율이 20% 이상
- 태양광 발전의 효율을 높이는 방법은 매우 다양

이번 분석에서는 환경 데이터를 기반으로 분석을 진행해보자!



프로젝트 우수작 - 기초 프로젝트



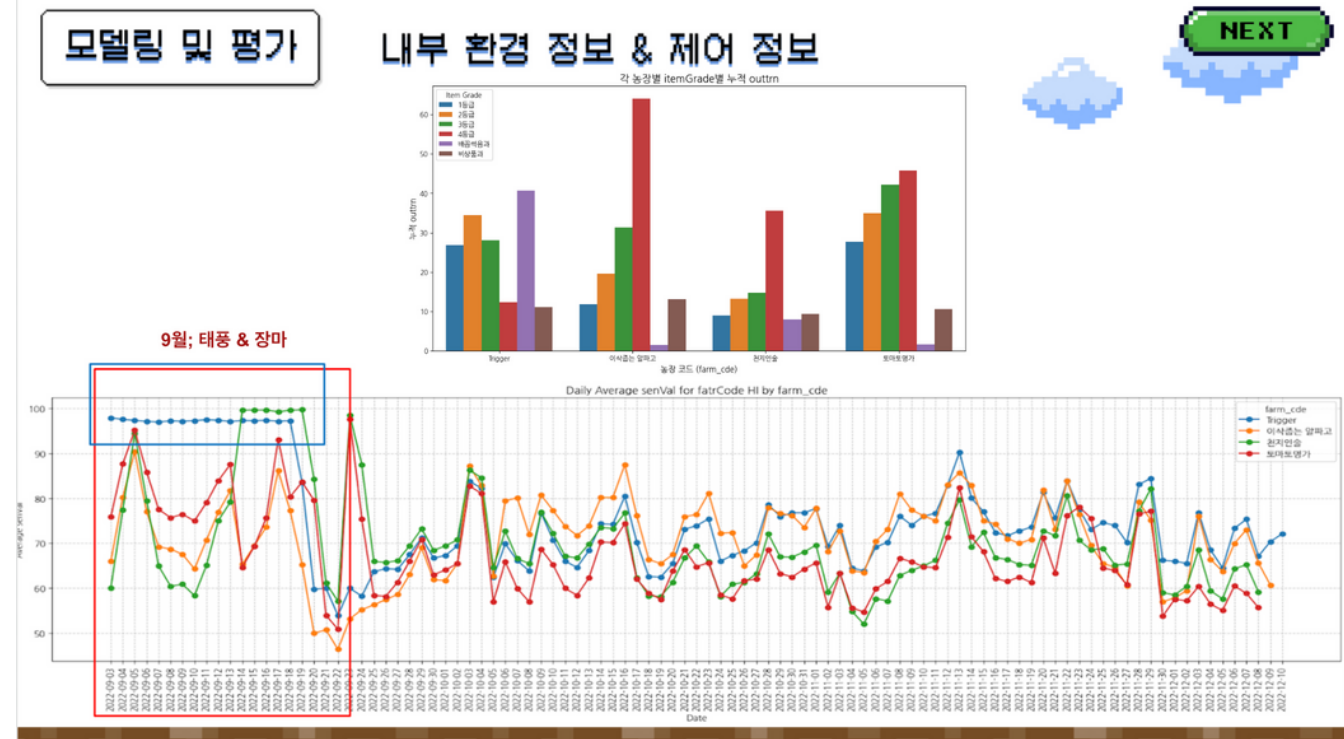
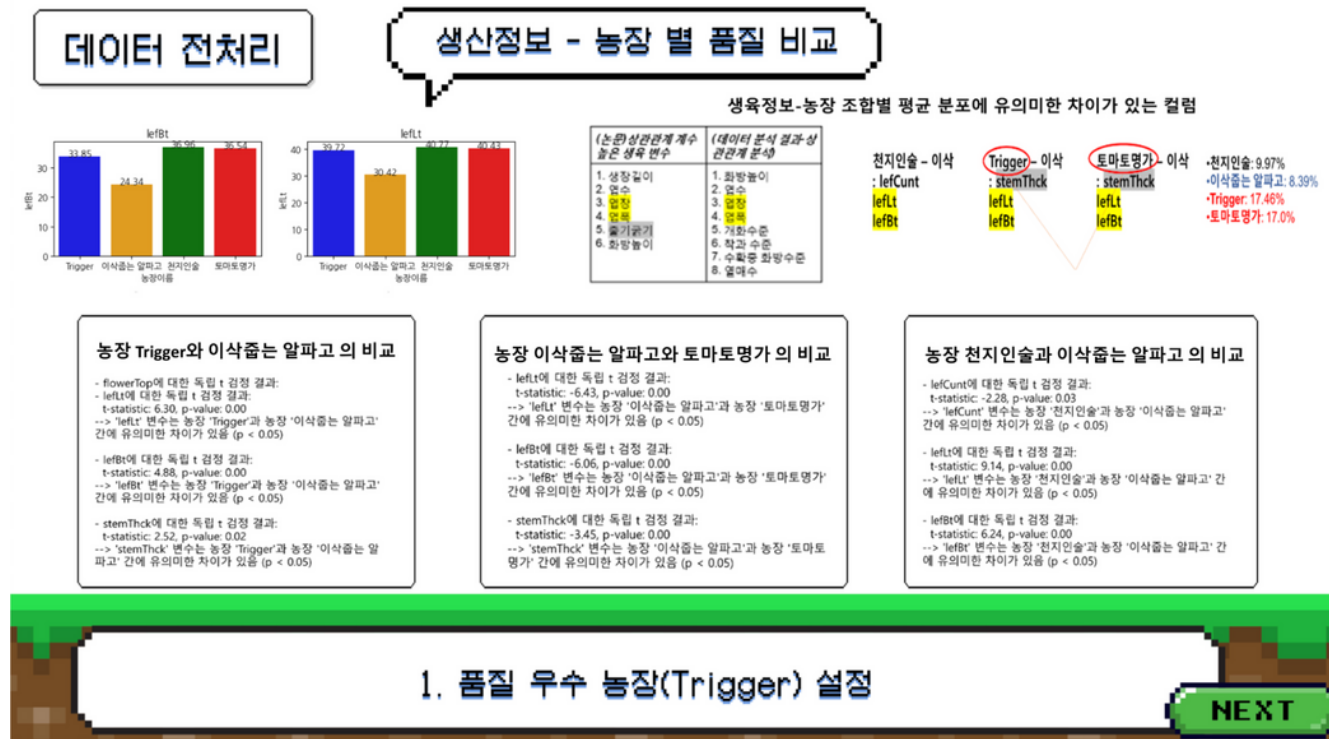
데이터 품질관리 프로젝트에서는 공정, 모빌리티, 바이오, 스마트팜 등 제조 데이터를 활용해 통계 분석과 머신러닝 기반 예측 모델 개발을 수행합니다.

습도 예측 모델 구축을 넘어 베이지안 최적화를 통해 각 변수의 이상적 조합을 도출하고, ICT 기반 실시간 제어 프로세스 및 대시보드 구성으로 현업 적용 가능성까지 고려한 분석 구조 마련

CO2, 일사 센서



프로젝트 우수작 - 데이터 품질관리 프로젝트



프로젝트 우수작 - 데이터 품질관리 프로젝트

데이터 품질관리 프로젝트에서는 공정, 모빌리티, 바이오, 스마트팜 등 제조 데이터를 활용해 통계 분석과 머신러닝 기반 예측 모델 개발을 수행합니다.

1조: [모빌리티] 엔진 센서 데이터 기반 결함 예측 및 품질관리 분석

변수 해석과 결함 유형에 대한 정교한 분석

엔진 센서 데이터를 기반으로 주요 변수의 의미를 명확히 파악하고, 결함 유형별 분포와 특성을 분석하여 다중 분류 문제를 설계. 결함 간 유사성 파악 및 논리적인 데이터 기반 인사이트 도출

다중공선성 및 불균형 문제 대응을 통한 모델 최적화

상관계수, VIF 기반 피처 제거 및 다중공선성 제거, 클래스 불균형 문제를 인식하고 적절한 분류 모델 선정, 파라미터 튜닝, confusion matrix 분석을 통해 모델 성능을 극대화

문제 재정의 및 모델 구조 개선을 통한 예측력 향상 시도

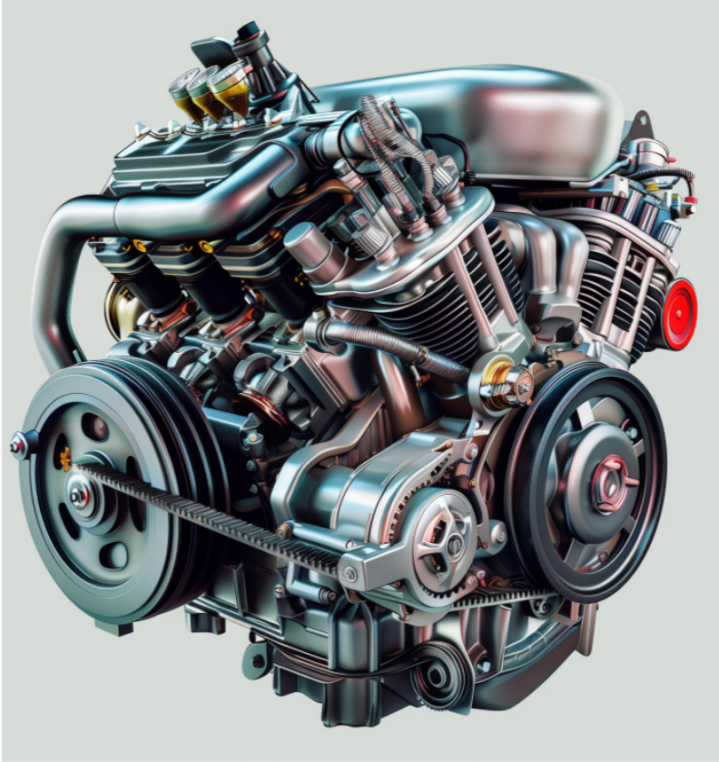
결함 유형 구분이 어려운 경우를 재정의하여 2단계 모델링 구조(1차 범주 분류 → 2차 세부 분류)로 확장하는 제안이 탁월했으며, 향후 연구 가능성과 응용력 측면에서도 높은 평가를 받을 수 있는 분석 흐름을 구성

○○○

모빌리티_1조

003

CHAPTER 1. 프로젝트의 목적



프로젝트 목적

'자동차 엔진 센서 데이터'를 활용해 결함 예측 및 품질 개선

- 에코 모빌리티의 데이터 분석 팀(Data Analyst)
- 결함 상태(정상/결함)를 사전에 예측하는 전략 제안
- EDA 수행, 통계분석, 머신러닝모델을 활용

프로젝트 우수작 - 데이터 품질관리 프로젝트

○○○

모빌리티_1조

007

CHAPTER 2-1. 데이터의 주요 변수

- 주요 변수(Performance, Fuel Consumption, Exhaust Gas)

CO

배기가스 중 일산화탄소 비율
연료 혼합비 이상 여부 유추 가능

HC

배기가스 중 탄화수소 농도
불완전 연소 시 높게 나타남

CO2

배기가스 중 이산화탄소 비율
연소 효율 및 환경 배출 지표

Lambda

공기-연료비의 이론값(1)에 대한
실제값의 비율
적정 혼합비 평가

O2

배기가스 중 산소 비율
연소 상태 평가

AFR

실제 공기-연료비
엔진 연소 상태를 직접적으로 보여줌

○○○

모빌리티_1조

010

CHAPTER 2-3. EDA

- 상관관계 분석

- 1) Performance Features
- 2) Fuel Consumption Features
- 3) Exhaust Gas Features

유사한 특성의 변수들 간에 높은 상관관계를 보임

○○○

모빌리티_1조

023

CHAPTER 3-2. 모델 성능 지표

- 모델 성능 평가(3가지 모델)

Model Accuracy Comparison: All feature vs selected feature

Model	Default Accuracy	Tuned Accuracy
Random Forest	74.55%	84.62%
XGBoost Model	75.11%	84.44%
SVM	60.54%	78.17%

정확성 84%로 향상!

전체 변수 모델 정확성 : 74.55%, 75.11%, 60.54%

변수 조정 후 모델 정확성 : 84.62%, 84.44%, 78.17%

→ ['Force', 'Power', 'CO', 'O2']를 이용하여, 새로운 변수를 만들었을 때, 정확성 ↑

→ Random Forest > XGBoost > SVM 순서로 모델 정확성 ↑

○○○

모빌리티_1조

027

CHAPTER 4-2. 모델의 개선점 및 추가 연구 가능성

- 모델의 개선점 및 추가 연구 가능성

개선점

- 1) Fault 2, 3 분류 성능 개선
 - 하이퍼 파라미터 튜닝의 최적화
 - 다양한 변수 엔지니어링으로 설명력 강화
- 2) 모델 해석성 개선
 - SHAP(Shapley Additive Explanations) 활용

추가 연구 가능성

- 1) 다중 모델 앙상블 기법
 - 1차 모델을 통한 거친 분류(0,1,2,3) → 2차 모델을 통한 정밀한 분류(2 vs 3)로 최적화
- 2) 데이터 수집 및 다양성 향상
 - 다양한 자동차 모델과 결함 유형을 포함한 더 많은 데이터를 수집
 - 데이터 증강 기법을 사용하여 훈련 데이터의 다양성을 향상

프로젝트 우수작 - 품질관리 대시보드 프로젝트

품질관리 대시보드 프로젝트에서는 Tableau와 Streamlit을 활용하여 주요 지표(KPI)를 실시간으로 모니터링할 수 있는 대시보드를 설계 및 구현합니다.

4조: [공정] 철광석 부유선별 공정 최종 제품의 실리카 불순물 함량 공정 최적화 분석

공정 특성을 반영한 전처리 및 파생 변수 설계

비선형적 공정 특성을 고려하여 주요 변수 간 비율로 변수를 변환해 차원을 축소하고 다중공선성을 완화. 시간 변수 제거, 센서 이상치 대응 등으로 데이터의 품질을 높였으며, 도메인 지식을 바탕으로 파생 변수 설계

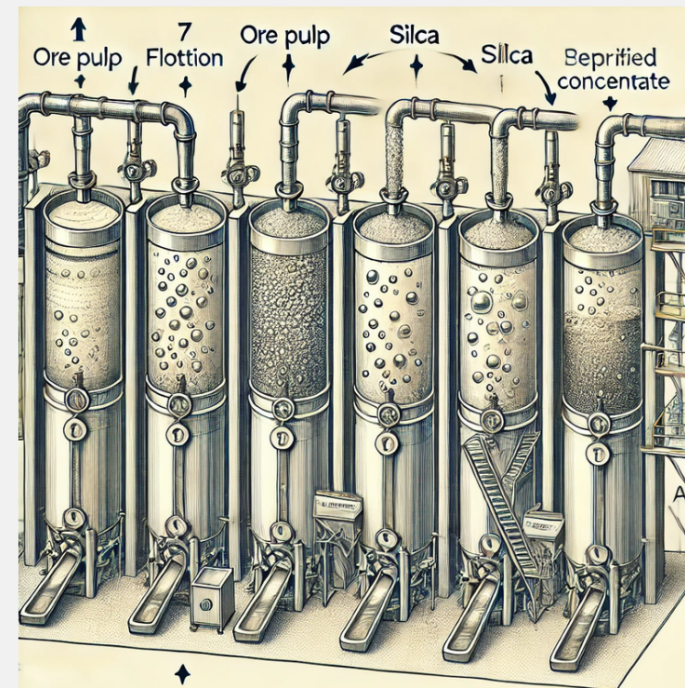
시계열성과 공정 특성을 반영한 고도화된 모델링

트리 기반 앙상블 모델(RandomForest, XGBoost)부터 시계열 정보를 반영한 LSTM까지 다양한 알고리즘을 적용. 각 모델별 성능을 비교하고 하이퍼파라미터 조정 및 EarlyStopping 등을 활용해 성능 최적화에 주력

공정 모니터링을 위한 대시보드 구축

분석 결과를 시각화한 Streamlit 대시보드를 통해 실리카 불순물 함량의 변화 추이와 주요 변수 간 관계를 실시간으로 모니터링. 사용자(엔지니어) 관점을 고려하여 공정 최적화 조건을 제시하여 현장 활용성을 높임

02. About 부유선별 공정



부유선별 컬럼(Flotation Column)

부유선별(Flotation) 공정에서 사용되는 세로형 탱크로, 광석에서 불순물을 제거하는 역할을 하는 장치.

기포를 이용하여 원하는 광물(철)과 불순물(실리카)을 분리함.

Data

부유선별 컬럼 내 공기 주입량 (Air Flow)

Flotation Column 01~07 Air Flow

부유선별 컬럼 내 액체 레벨 (Column Level)

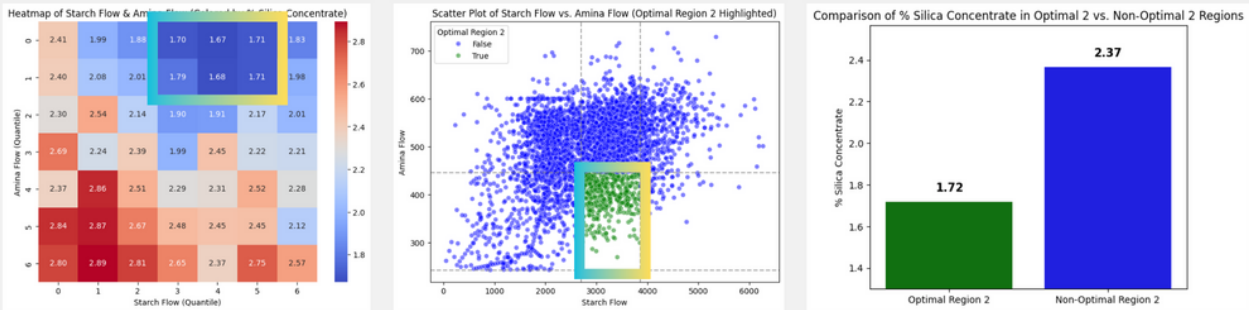
Flotation Column 01~07 Level

프로젝트 우수작 - 품질관리 대시보드 프로젝트

03. EDA 분석 - 공정 최적화 가능성 분석

화학적 첨가제(Flotation Reagents)

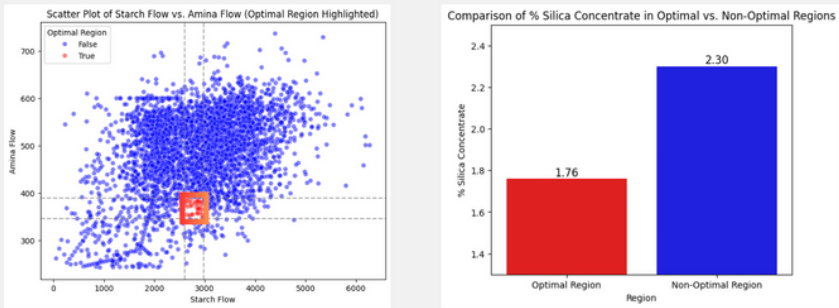
2D Heatmap을 통한 공정최적화2구간 starch는 2698~3865, amina는 243~446



03. EDA 분석 - 공정 최적화 가능성 분석

화학적 첨가제(Flotation Reagents)

Starch Flow의 최적 구간 2602~2975, Amina Flow의 최적 구간은 346~390

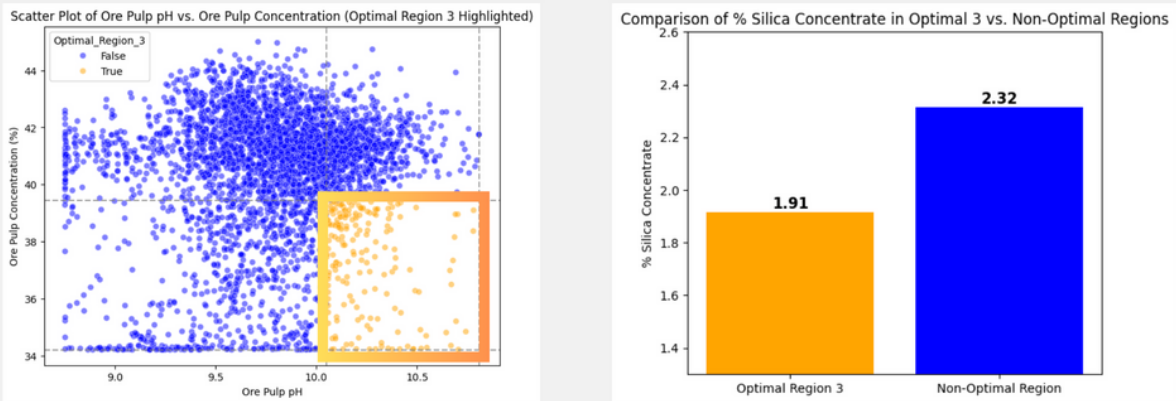


공정 최적화 구간과 비최적화 구간의 % Silica Concentrate 평균 차이에 대한 T-검정 결과
 귀무가설(H₀): 최적화된 공정 구간과 비최적화 구간의 % Silica Concentrate 평균에 통계적으로 유의미한 차이가 없다.
 대립가설(H₁): 최적화된 공정 구간과 비최적화 구간의 % Silica Concentrate 평균에 통계적으로 유의미한 차이가 있다.
 T-통계량: -4.1673
 P-값: 0.0002
 결론: 귀무가설(H₀) 기각 → 최적화된 공정 구간과 비최적화 구간의 % Silica Concentrate 평균 차이는 통계적으로 유의하다.

03. EDA 분석 - 공정 최적화 가능성 분석

광액(Ore pulp)

광액의 특성 변수의 산포에서 공정최적화구간 확인



국가별 광산물 수입

2023년 아 연 광(MT) 총 수입금액
2조 2985억 원

아래에서 광종을 선택하면 해당 광종의 수입 정보를 확인할 수 있습니다.

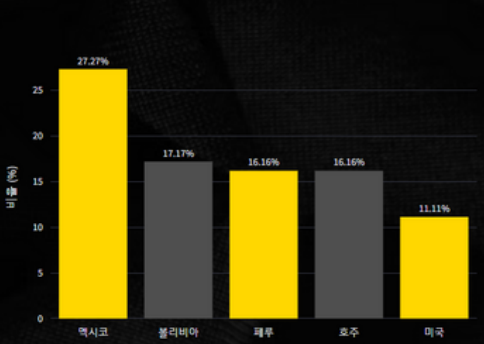
분석할 광종을 선택하세요:

아 연 광(MT)

아 연 광(MT) 수입 비율 지도



상위 5개국 수입 비율(%)



프로젝트 우수작 - 품질관리 대시보드 프로젝트

품질관리 대시보드 프로젝트에서는 Tableau와 Streamlit을 활용하여 주요 지표(KPI)를 실시간으로 모니터링할 수 있는 대시보드를 설계 및 구현합니다.

8조 : [운송] 2차전지 자재 이송 물류 데이터셋 : 지연 예측

정교한 전처리 및 지리정보 기반 데이터 구성

지연 예측에 필요한 변수 유형 정리, 결측 및 이상치 처리, 스케일링 등 전처리 과정이 탄탄하게 구성되었으며, 위도·경도 기반 위치 데이터를 지도 시각화에 맞게 변환하여 공간적 분석 기반을 마련

회귀 문제에 맞춘 모델링 및 평가 지표 활용

지연시간 예측이라는 회귀 문제에 적합한 다양한 모델을 비교 적용하고, MAE, RMSE, R² 등 회귀 특화 지표를 활용해 성능을 다각도로 분석함. 파라미터 조정 및 다중 모델 비교로 모델 성능 극대화를 시도

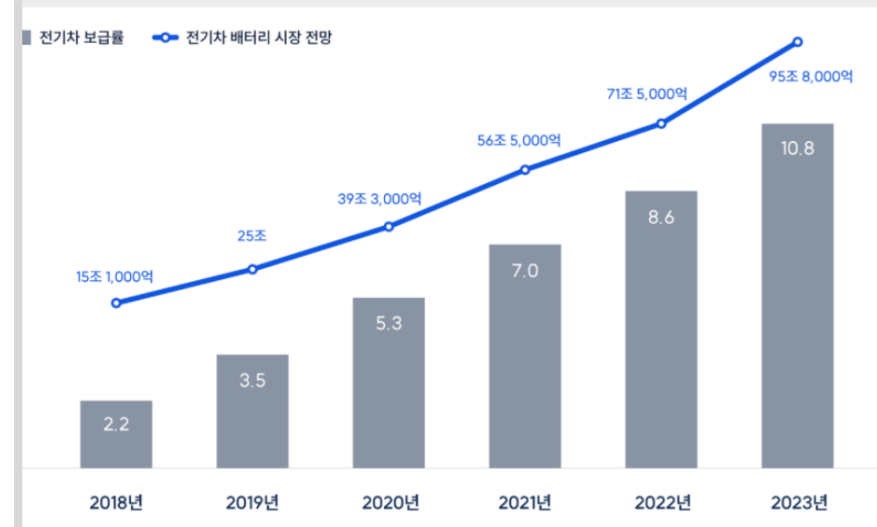
분석 결과의 해석 및 지연 개선 인사이트 도출

예측 결과를 지연 원인과 연계해 해석하며, 운송 효율 개선을 위한 실질적인 제안까지 도출. 지도 시각화를 통해 지연 발생 지역의 공간적 패턴을 드러내고, 분석 내용을 체계적으로 문서화하여 전달력 높은 결과물로 완성

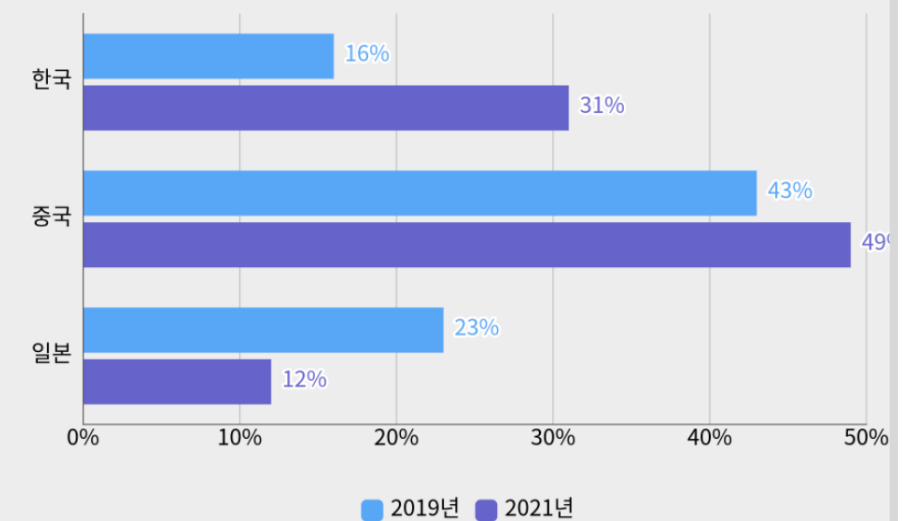
이차전지 물류 운송

2023년 1월, 라스베이거스에서 열린 세계 최대 가전·정보기술전시회 CES 2023에서도 확인할 수 있는 만큼 **전기차 배터리 시장이 계속해서 확장**되고 있으며 이와 함께 리튬 배터리의 수요 증가세도 주목받고 있음. 리튬 배터리는 운송 시 위험물로 분류되는 만큼, **까다로운 관리와 규정이 요구**됨!

전기차 보급률과 전기차 배터리 시장 규모(단위: 원, %)

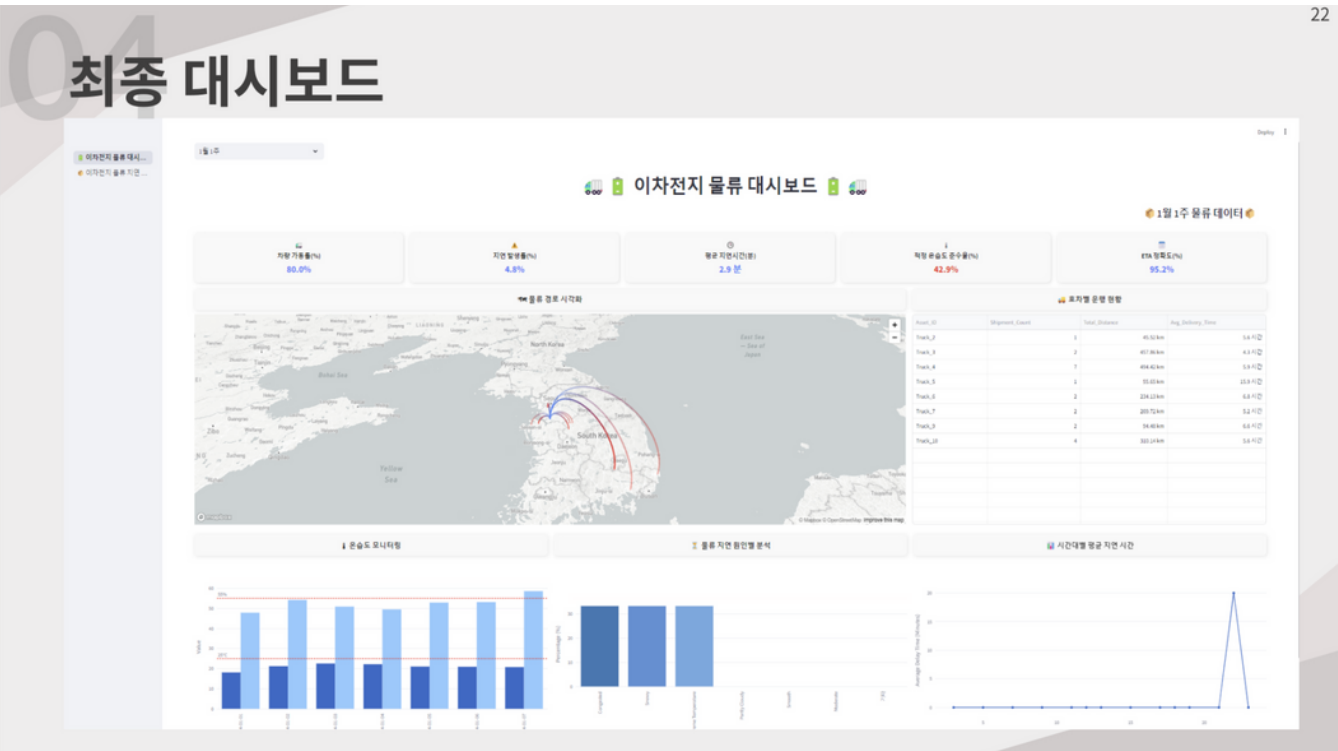
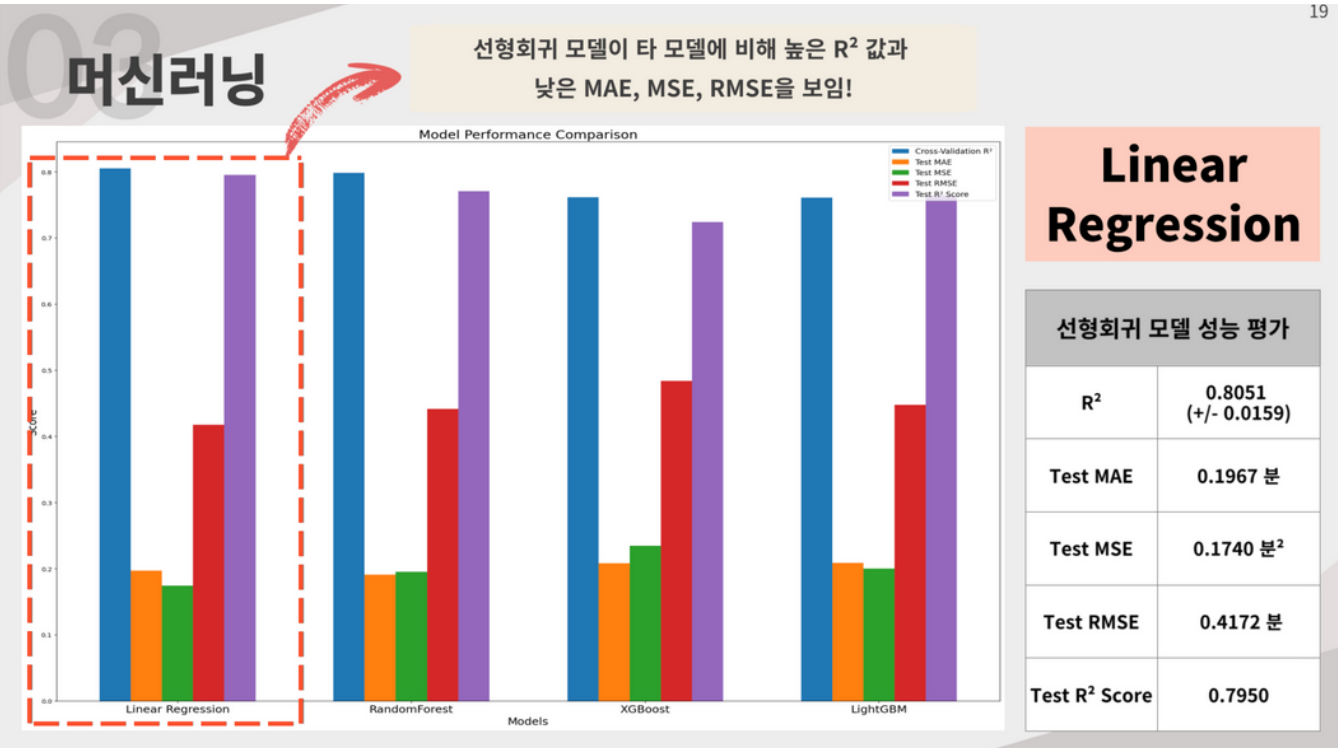
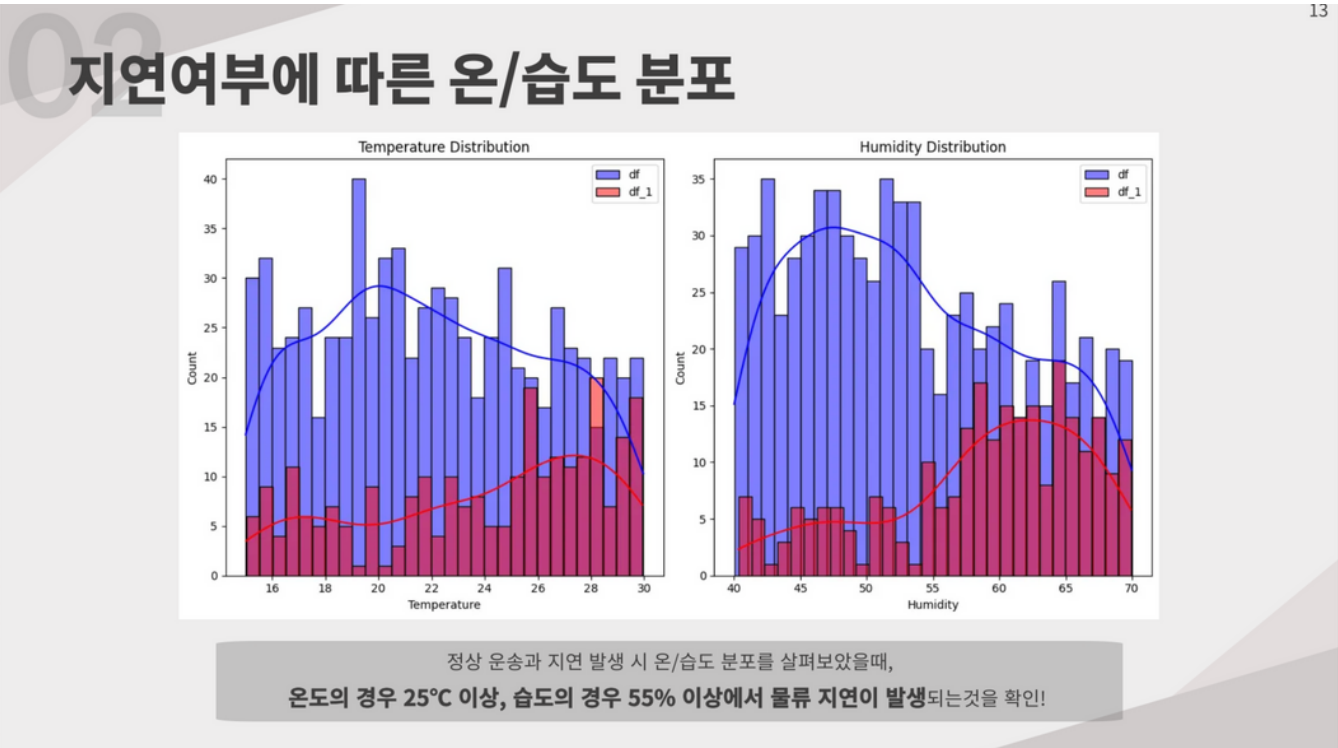


한중일 3개국 배터리 시장 점유율(%)



이러한 배터리 수요 증가는 곧, 운송 중 배터리 품질 관리에도 변화가 필요함을 뜻함.
수요가 증가할수록 기업 및 소비자의 품질 관리와 안전 배송에 대한 기준도 높아지기 때문!

프로젝트 우수작 - 품질관리 대시보드 프로젝트





감사합니다