



태양광 패널을 사용하는 고객의 에너지 소비량 및 생산량 예측

ESTsoft Wassup 1기 2회차 프로젝트 - 5조

목차

- 팀원 소개
- 주제 소개
- 데이터 살펴보기
- 프로젝트 관리
- 앞으로의 방향
- Q&A

팀원 소개



김선들



김종성



안지우



최세준 🏰

주제 소개

출처

- kaggle competition, by Enefit

배경

- Enefit과 계약한 고객들 중
태양광 패널을 이용하여 전력을 생산하고
동시에 소비하는 고객 (Prosumer)

목표

- 생산량과 소비량의 불균형 해소와 비용
절감을 위해 태양광 패널을 사용하는 고
객의 에너지 소비량 및 생산량 예측

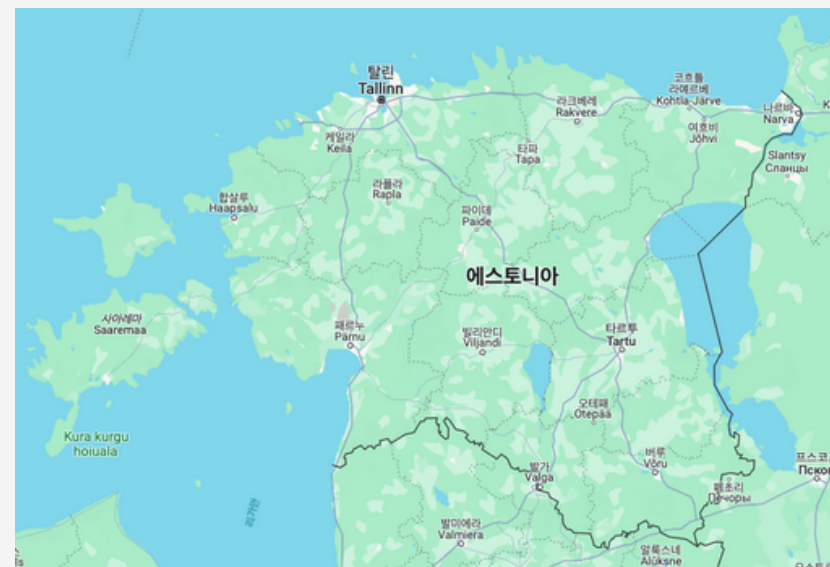


에스토니아 살펴보기

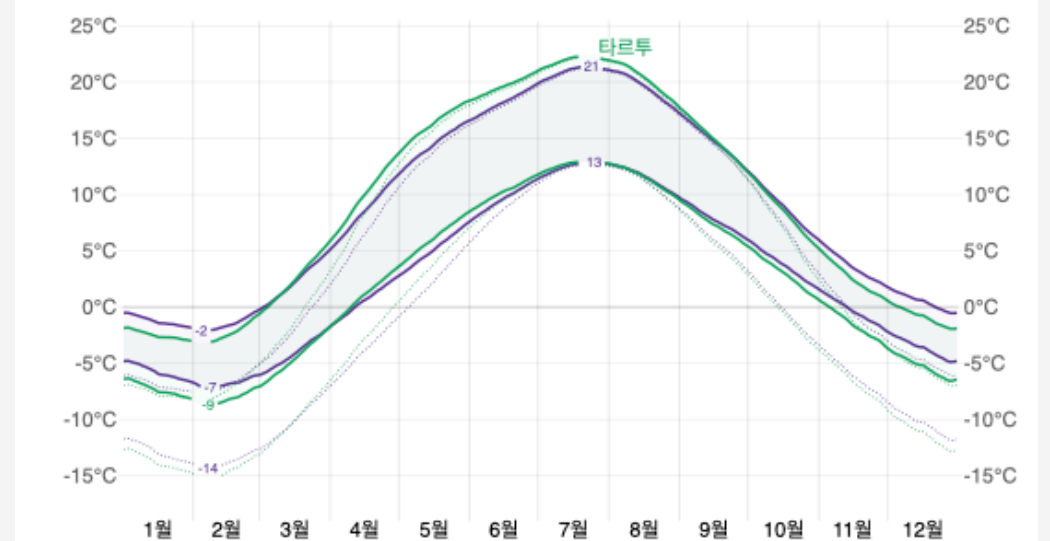


에스토니아

- 북유럽, 평야지대
- 냉량습윤한 해양성기후
- 연강수량은 600~750mm



에스토니아의 평균 고온 저온



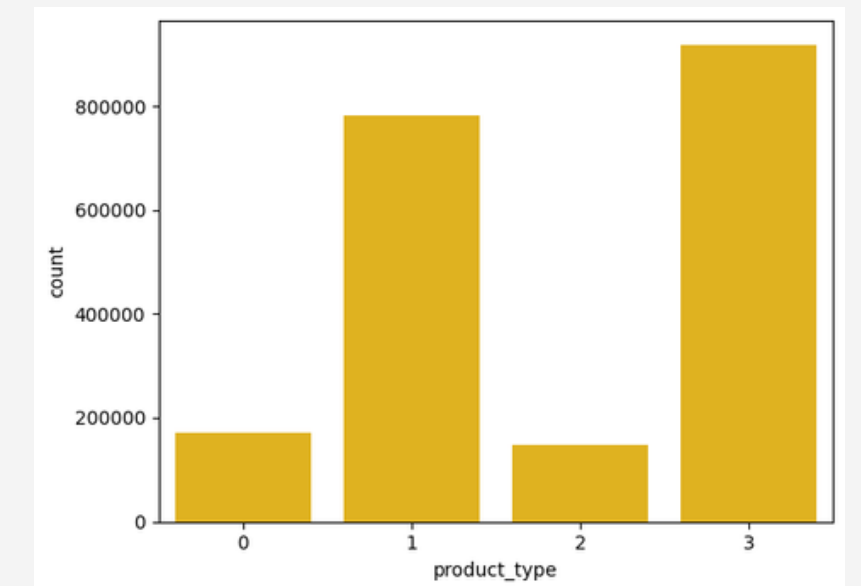
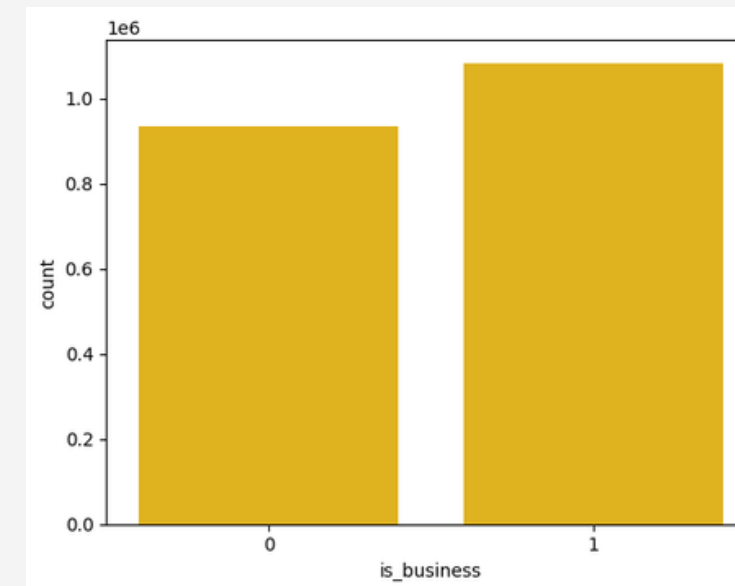
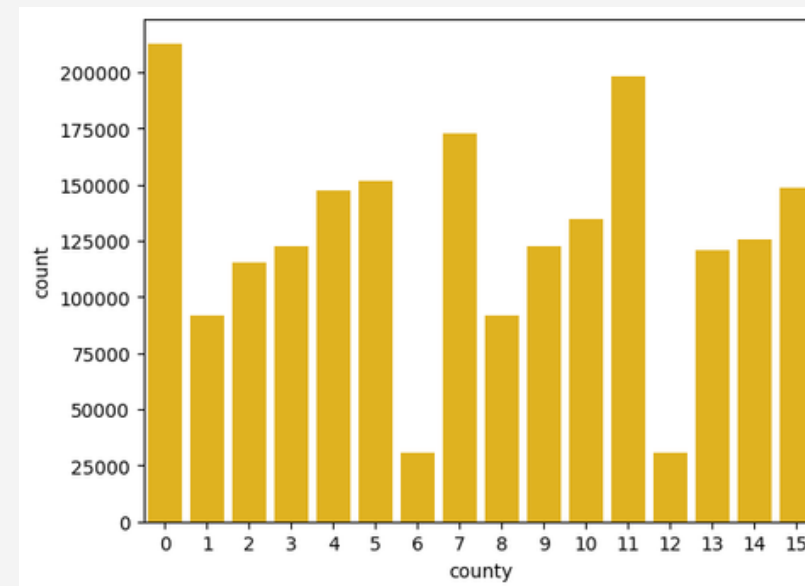
데이터 살펴보기 - 세그먼트

세그먼트 결정요소

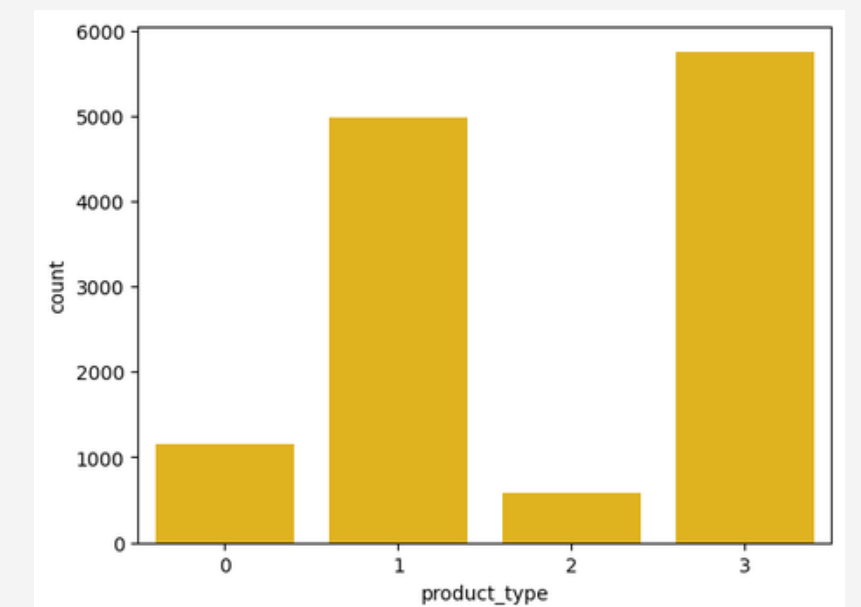
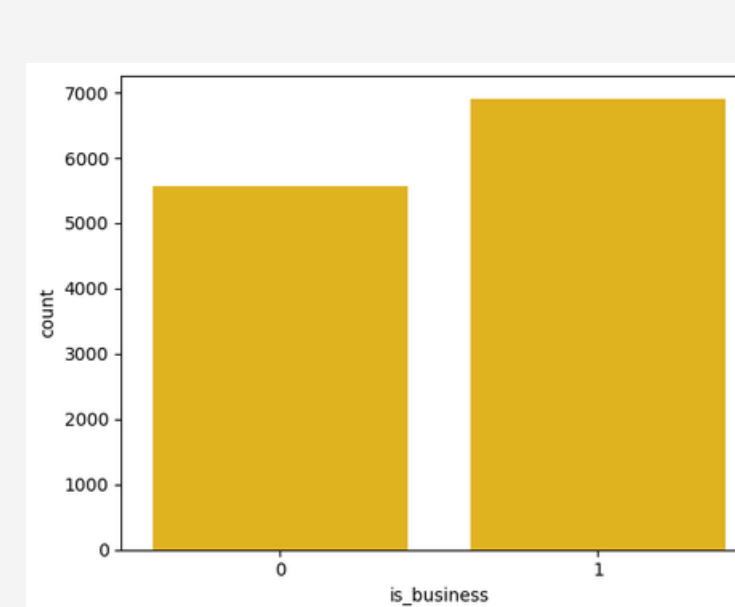
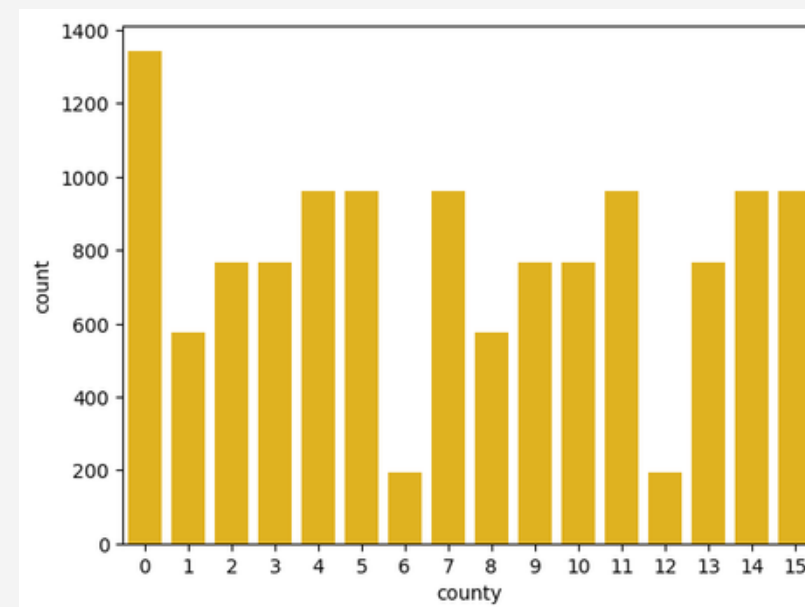
- county
- is_business
- product_type

3가지 조합

Train.csv

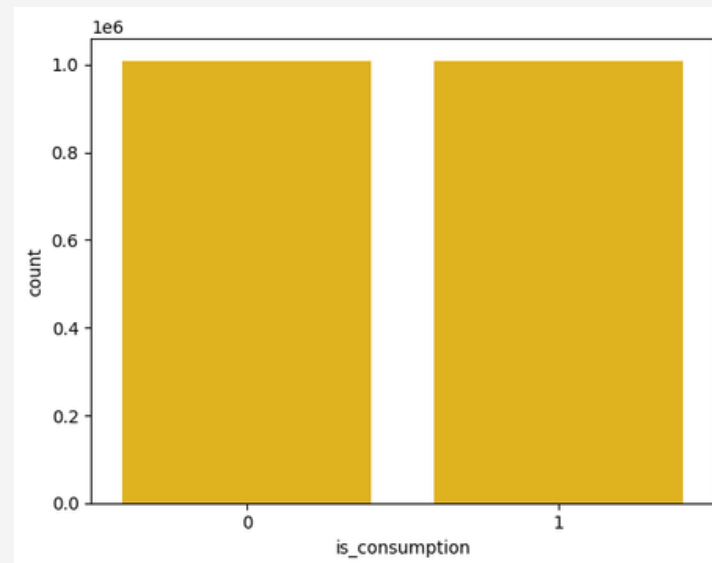


Test.csv

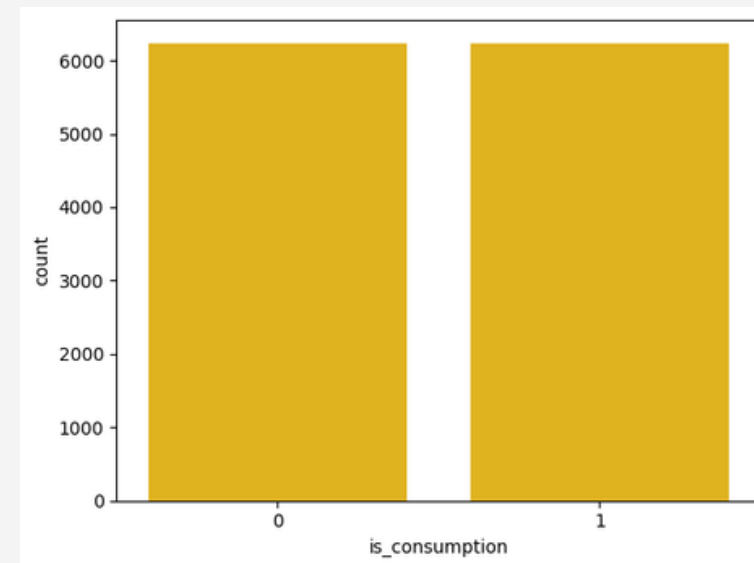


데이터 살펴보기 - 주요 Column

train

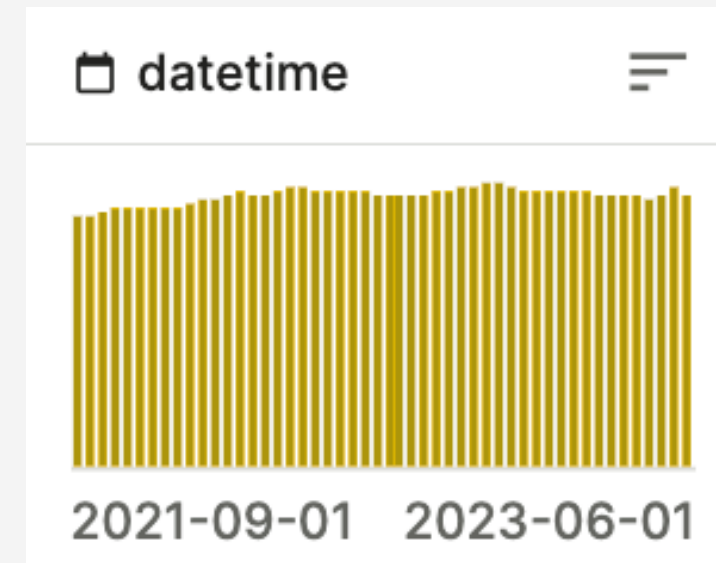


test

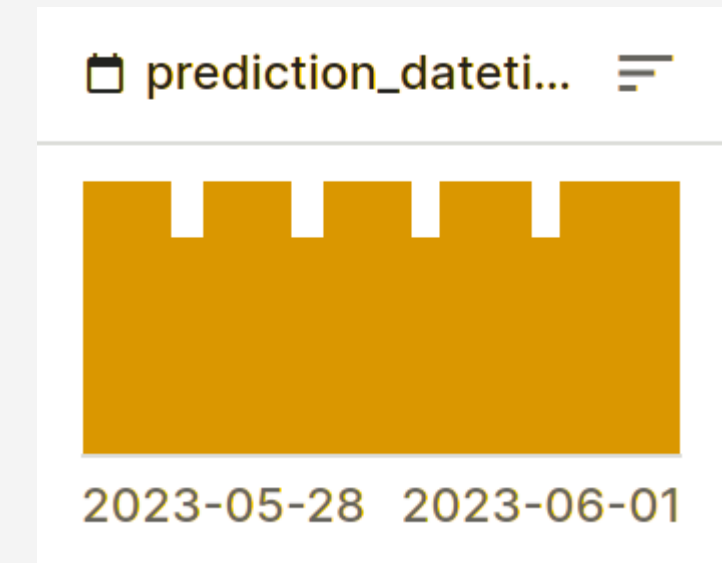


is_consumption

train

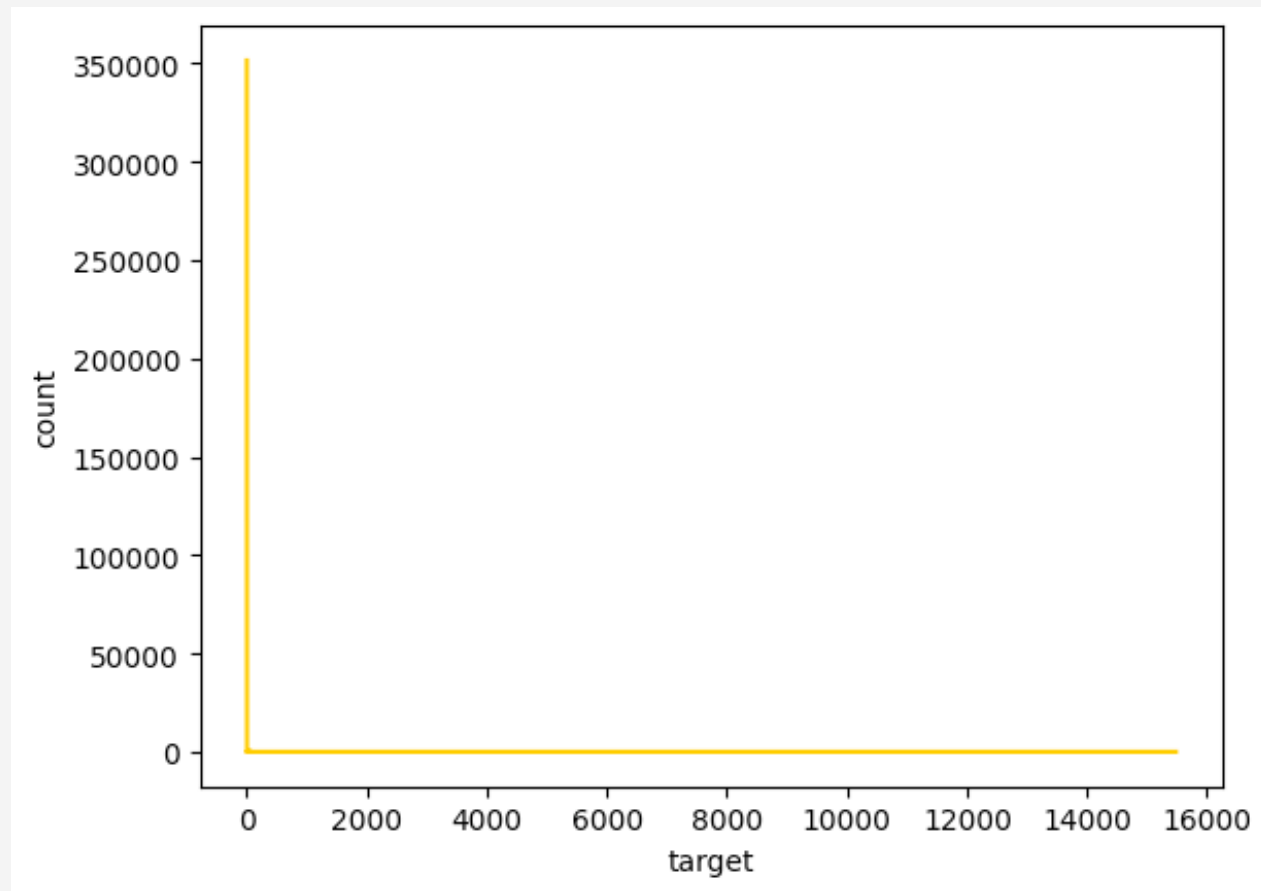


test

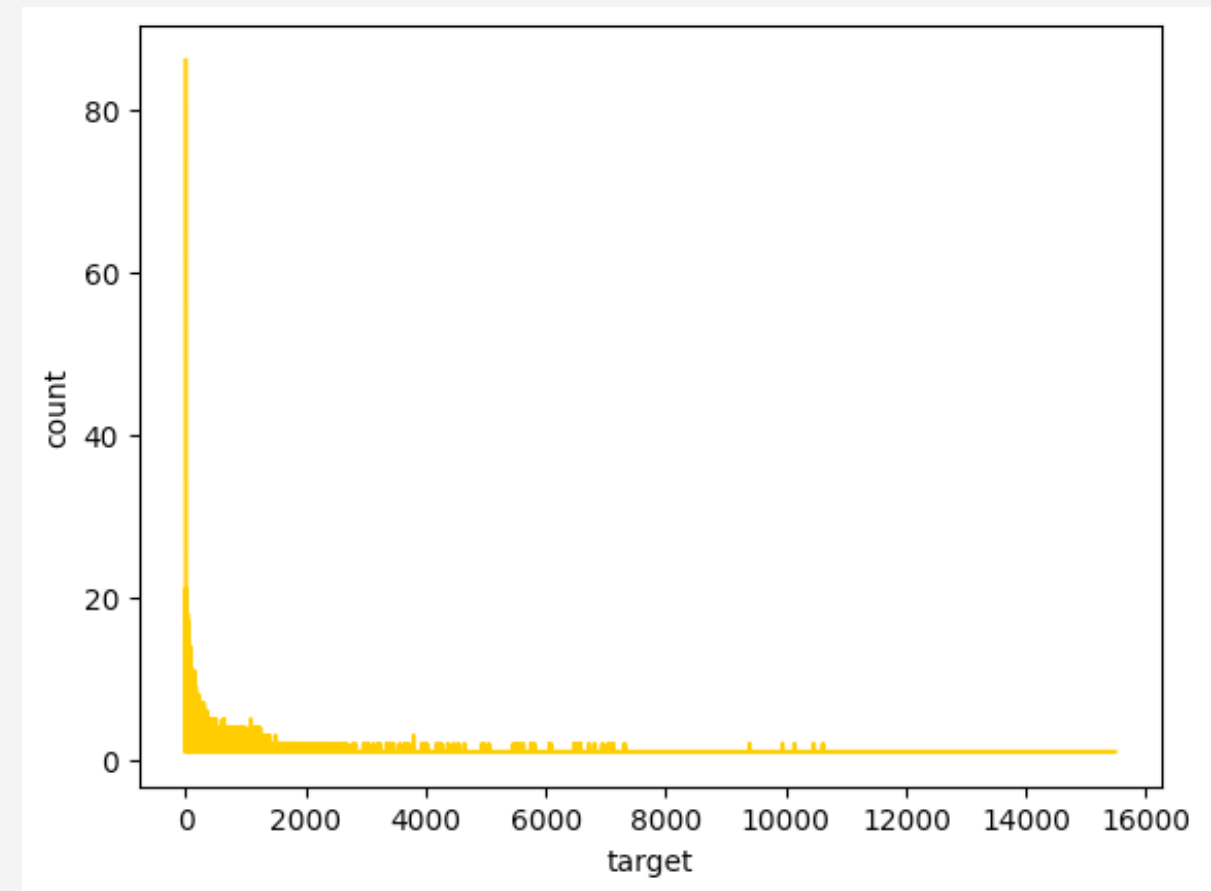


datetime

데이터 살펴보기 - 타겟



is_consumption: 0



is_consumption: 1

타겟

시간별 에너지 소비량 또는 생산량(MWh)

데이터 살펴보기 - 추가데이터

추가로 주어진 데이터

- client data
- gas prices
- electricity prices
- forecast weather
- historical weather

추가로 알아볼 데이터

- 에스토니아 공휴일정보
 - 국경일에 따라서
에너지 소비 현황이 달라질 수 있음

데이터 살펴보기 – train & test

train set

- 2021/09/01 ~ 2023/05/31
- 약 200만 rows
- 결측값: 528 rows

test set

- 2023/5/28~23/5/31 (샘플 테스트 셋)
- 리더보드는 일정 기간마다 주최 측에서 미래의 데이터를 가지고 예측을 진행한다고 함.

데이터 살펴보기 - 데이터 합치기

데이터가 별개의 csv파일로 나뉘어진 상태
-> Merge하는 작업 필요

- 2,018,352 rows, 57 columns
- dtypes: float64(35), int64(15), object(7)

시계열 기준을 잘 맞춰주어야함

- data_block_id: 같은 data_block_id는 예측할 날짜(forecast time)을 기준으로 봤을 때 이용가능한 모든 정보. 모든 csv파일의 모든 rows에서 해당 id가 존재함.

데이터 살펴보기 - 평가 지표

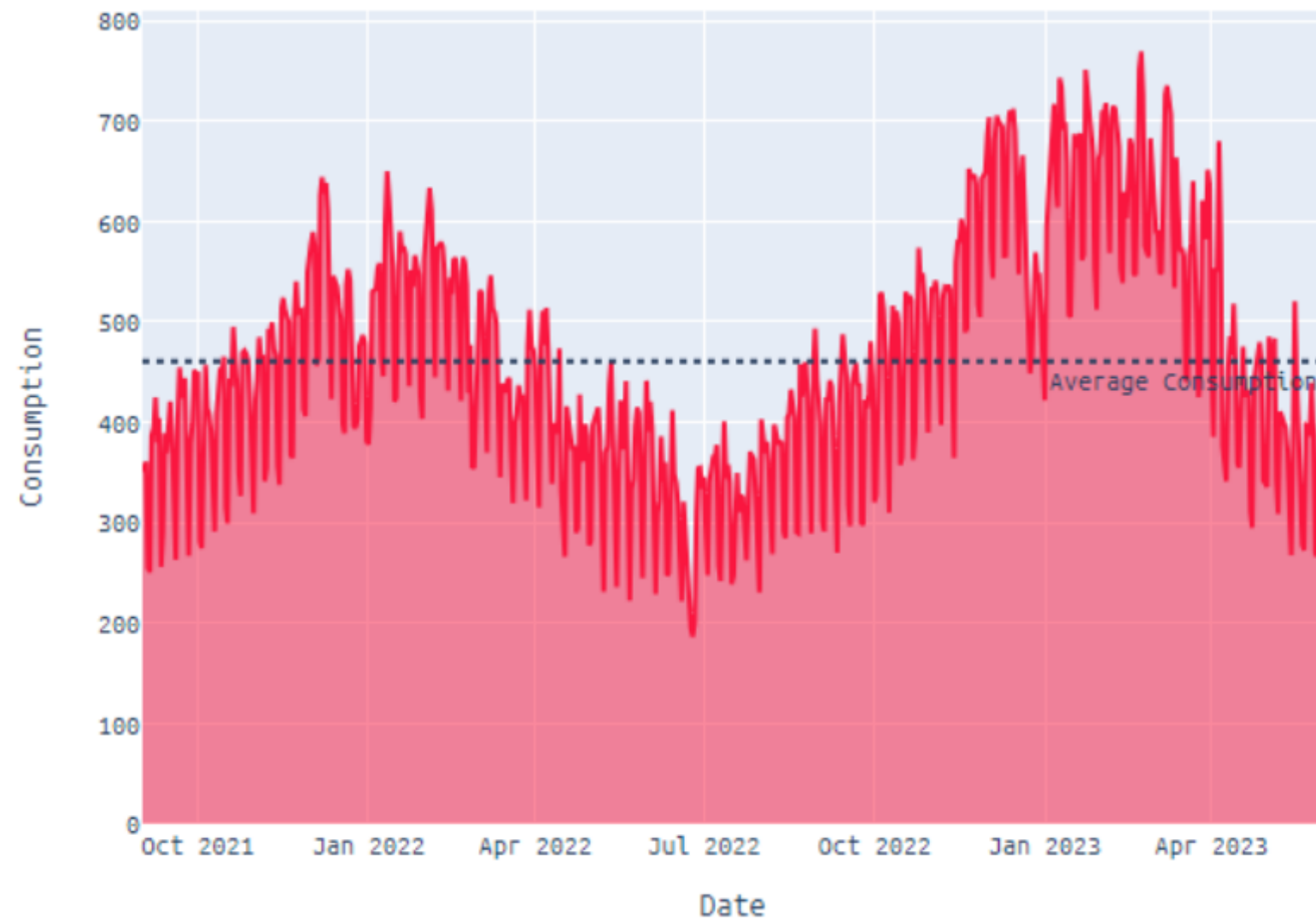
Mean Absolute Error (MAE)

- between the predicted return and the observed target
- n is the total number of data points.
- y_i is the predicted value for data point i .
- x_i is the observed value for data point i .

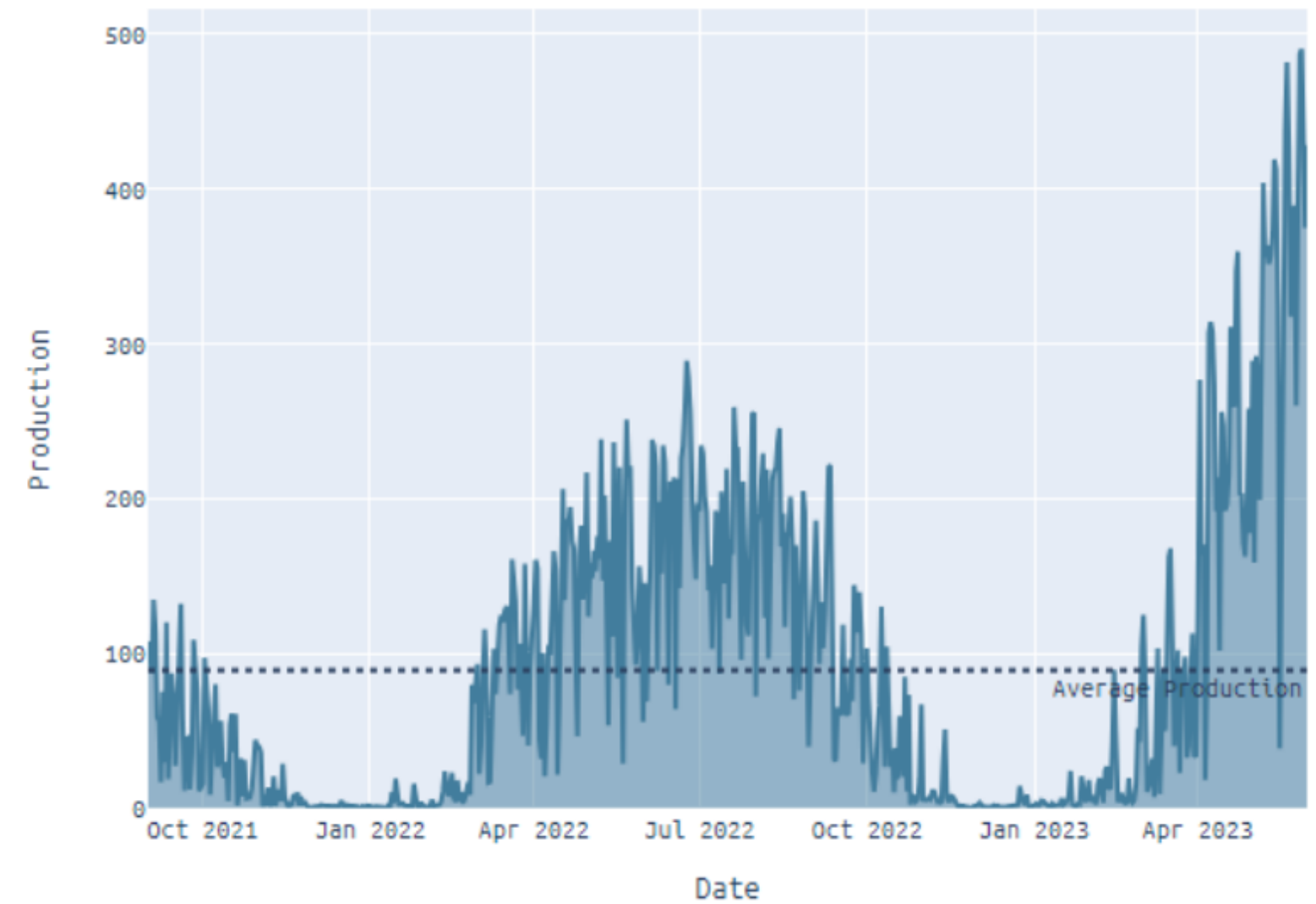
$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - x_i|$$

데이터 살펴보기 - 시각화

1. Daily Consumption Analysis



2. Daily Production Analysis



프로젝트 관리



기록 및 소통 방법

- 회의록, 기타 문서
 - 구글 드라이브
- 수업 시간 이외의 소통: 디스코드 서버 생성 후 이용
- 데일리 스크럼
 - 매일 수업 시작 직후와 수업 종료 직전에 각자 한 일, 할 일을 5분 내외로 이야기하기
 - 백로그 작업 선정

Github

- Main Branch
 - 최종 release 용
 - 실험 진행 branch
 - 모든 실험의 결과가 main branch에서 재현 가능해야 함.
- Dev Branch
 - Merge 용, conflict 해결 및 버그 수정
- Task 별 Branch
 - 기능 단위의 branch 생성

칸반 보드

- Backlog: 해야 할 일, 태스크, 하면 좋을 것 같은 일들 카드로 생성
- Ready: 하기로 정한 일
- In Progress: 작업 시작한 일
- In Review: 작업 끝나고 리뷰 중인 일
- Done: 작업 및 리뷰가 끝난 일

프로젝트 관리 - Github 칸반보드

pmz-q / Projects / est_wassup_02

est_wassup_02

KANBAN

Roadmap

Team items

In review

Priority board

My items

New view

-label:"timeline-event"

Backlog

1

Estimate: 0

...

하면 좋겠는 일들 작성해주세요

Draft

파이프라인 만들기

+ Add item

Ready

0

Estimate: 0

...

회의때 Backlog 에서 카드를 정리해서 Ready 로 옮겨주세요

+ Add item

In progress

3 / 13

Estimate: 90

...

Backlog 에서 In Progress 로 카드 옮길 때 디스코드에 "~ 합니다" 라고 말해주세요

est_wassup_02 #7

데이터셋 합치기

P1 10 S

est_wassup_02 #8

자료 조사

P1 30 M

est_wassup_02 #9

데이터 양상 정리

P1 50 M

+ Add item

In review

1

Estimate: 0

...

In Progress 에서 In Review 로 옮길 때 리뷰해달라고 디코서버에 말씀해주세요

est_wassup_02 #10

주제 발표 PPT 작성

P1 S

+ Add item

Done

0

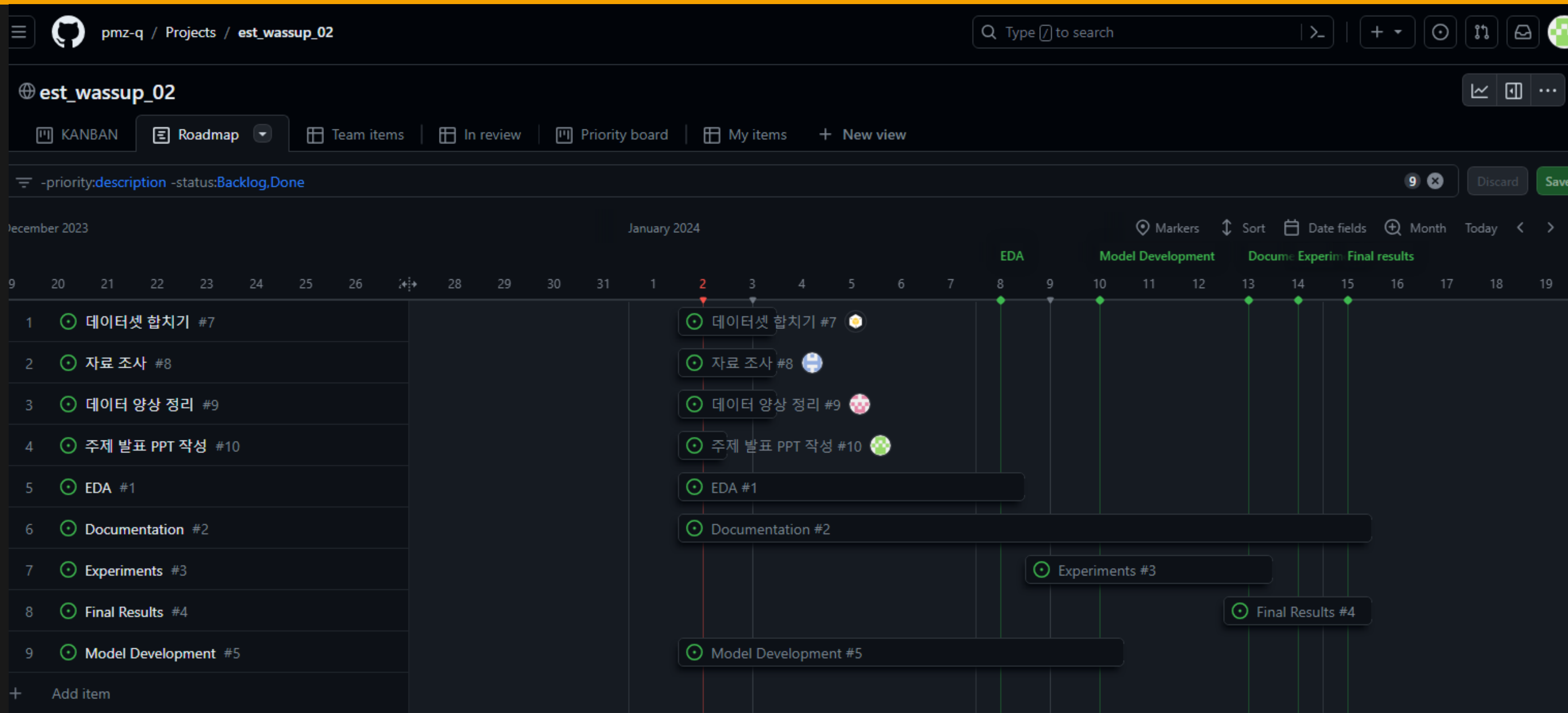
Estimate: 0

...

리뷰 끝나면 여기로 카드 옮겨주세요

+ Add item

앞으로의 방향 - 타임라인



앞으로의 방향



파이프라인 및 학습 검증

빠르게 파이프라인을 만들고, 전처리된 데이터로 학습이 잘 진행되는지를 확인합니다. (loss 감소 확인하기)



다양한 모델 활용

ARIMA, ANN, LSTM, PatchTST 와 같은 다양한 모델을 이용하고 비교합니다.



성능 끌어올리기

평가 지표(주로 MAE)를 기준으로 성능을 높이기 위해 할 수 있는 건 다 해봅니다. (파라미터 조정, 추가 데이터 전처리, 피쳐 선택, learning scheduler와 같은 기교 등)

Q&A



ESTSOFT