

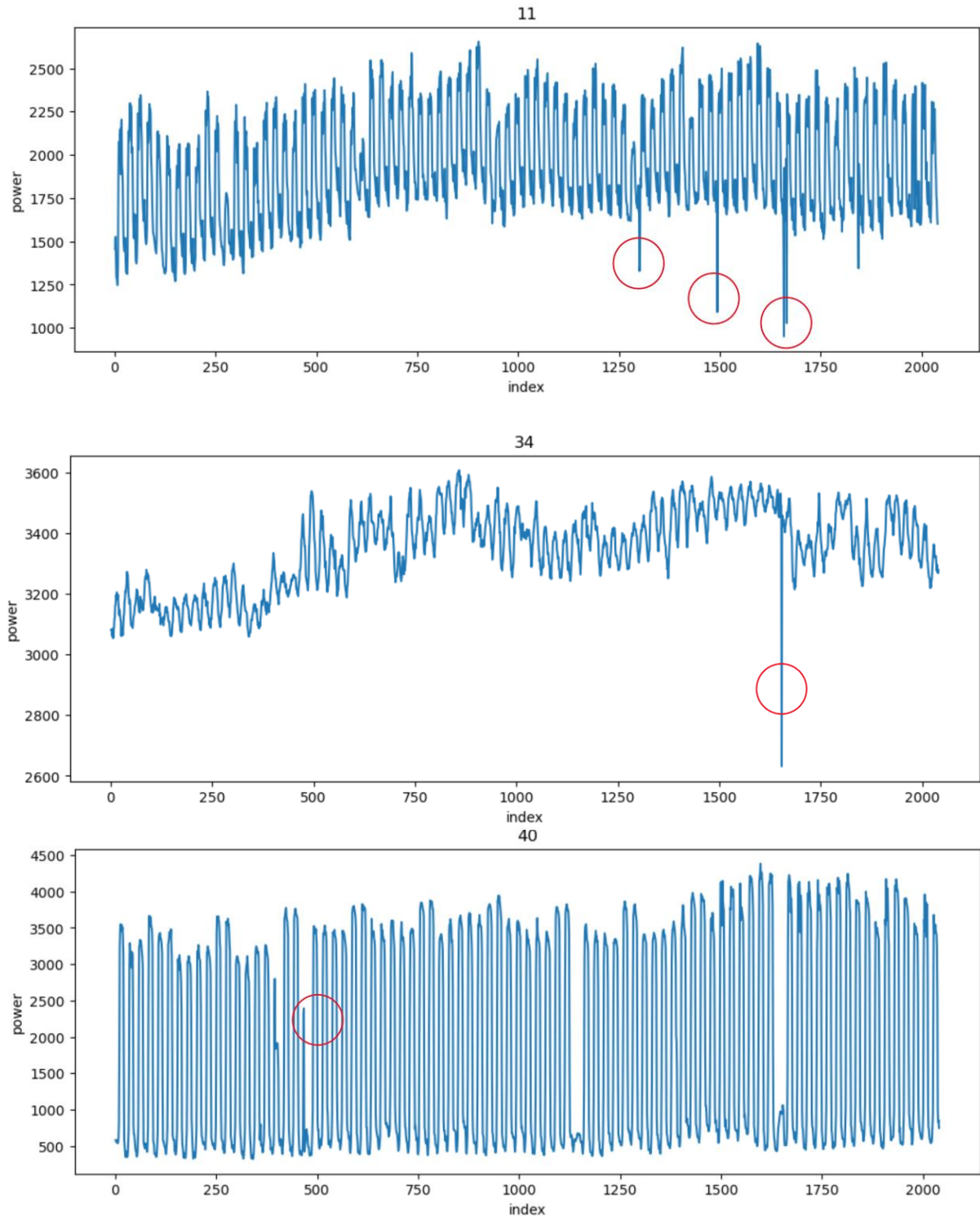
Remove Outliers

이상치를 제거하는 알고리즘도 있습니다만,

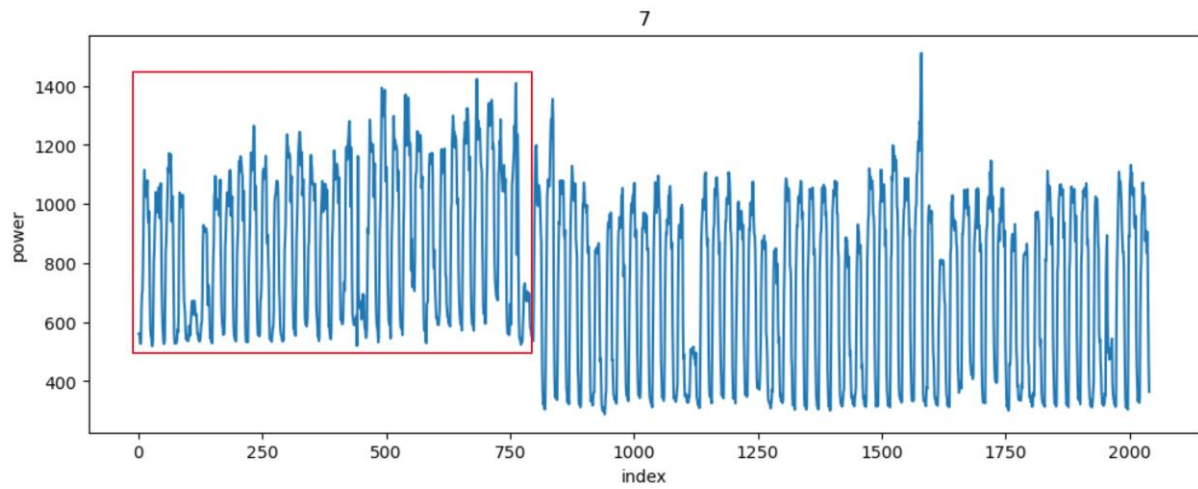
건물마다 전력량의 범위, 전력량의 패턴, 이상치들의 타입 모두가 달랐기 때문에

그래프를 통해 직접 제거하는 것이 가장 좋은 방법이라고 생각했습니다.

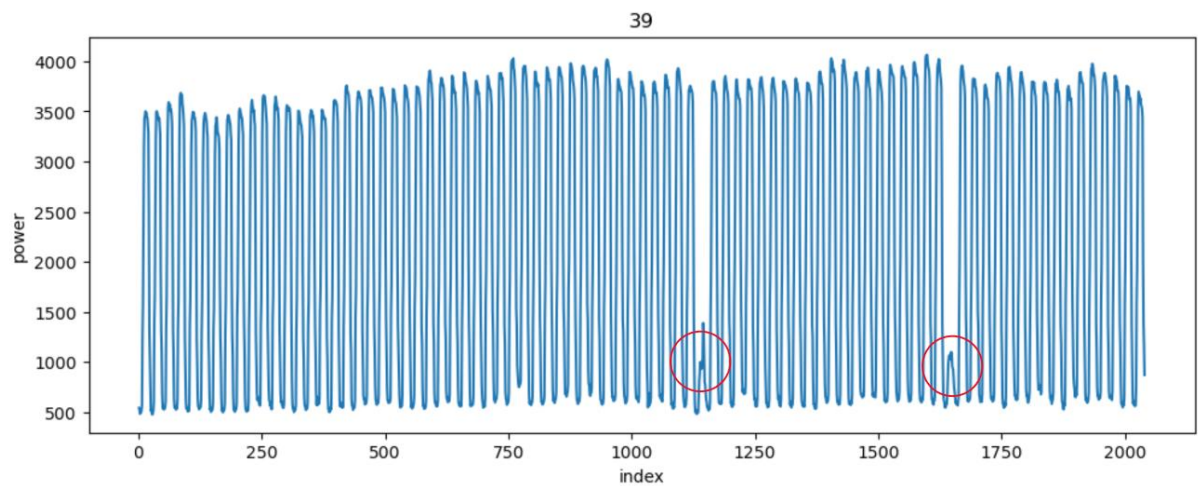
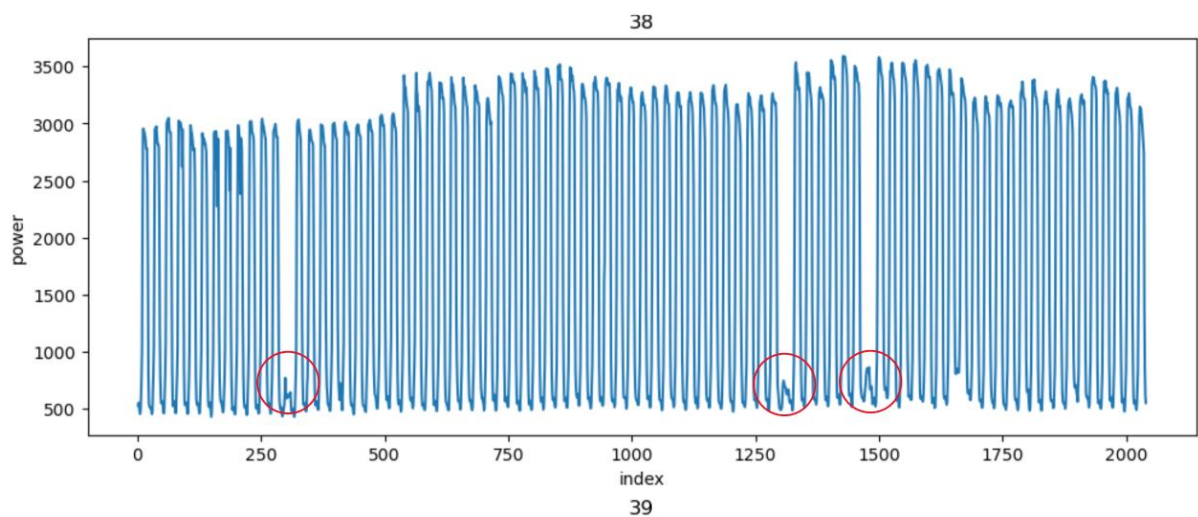
Type1. 긴 꼬리

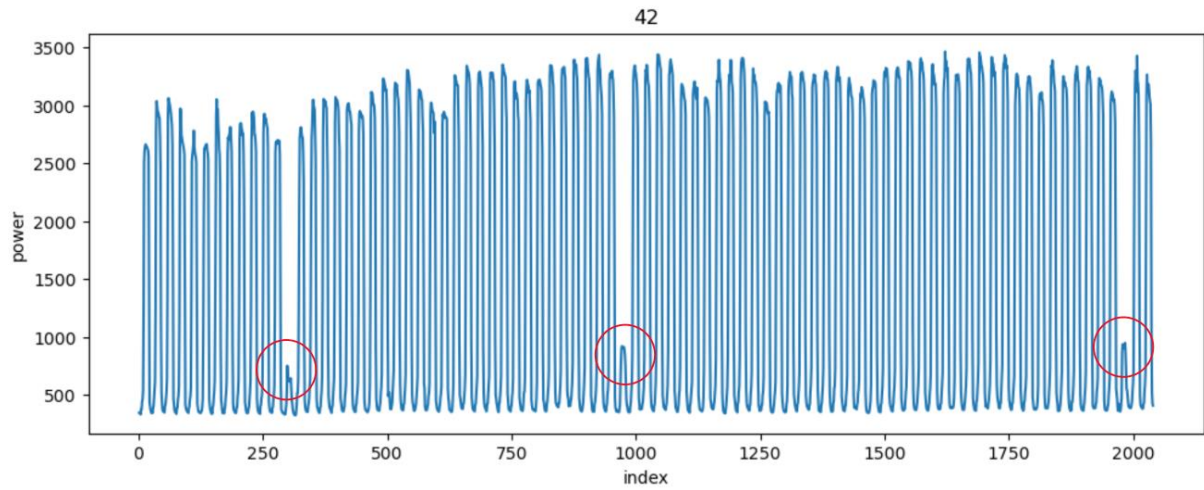


Type2. 범위이동

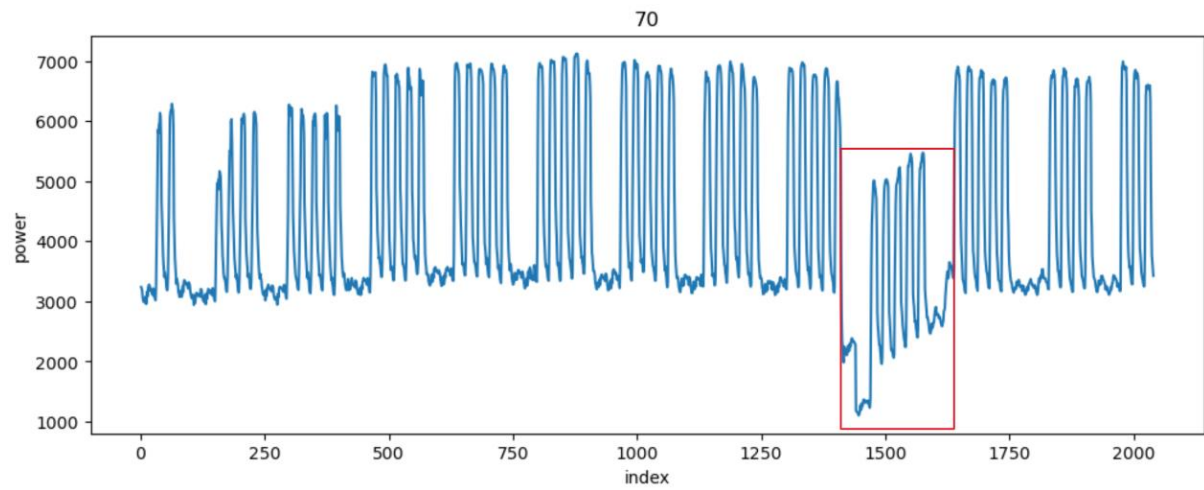


Type3. 불규칙적인 휴무일



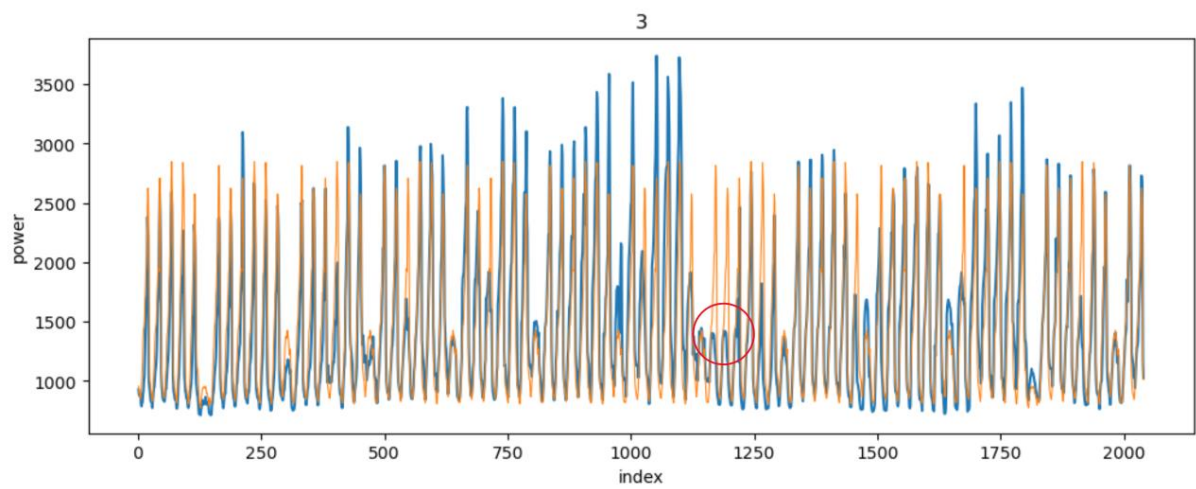


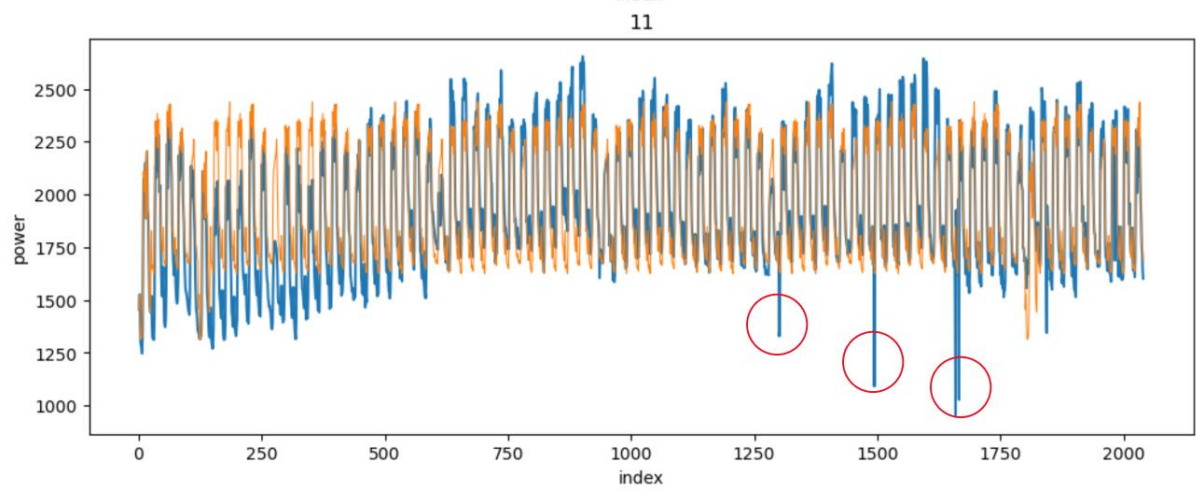
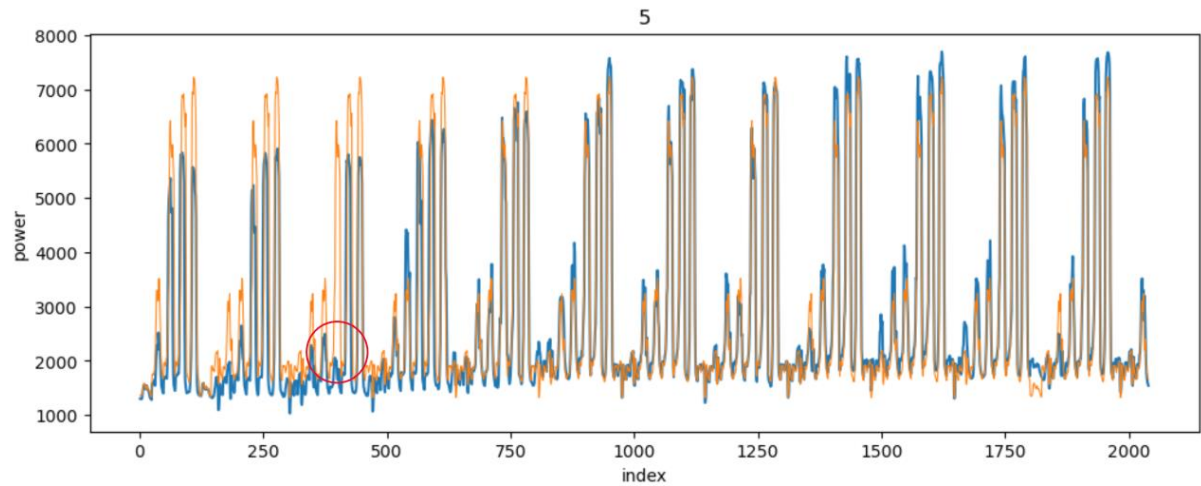
Type4. 휴가



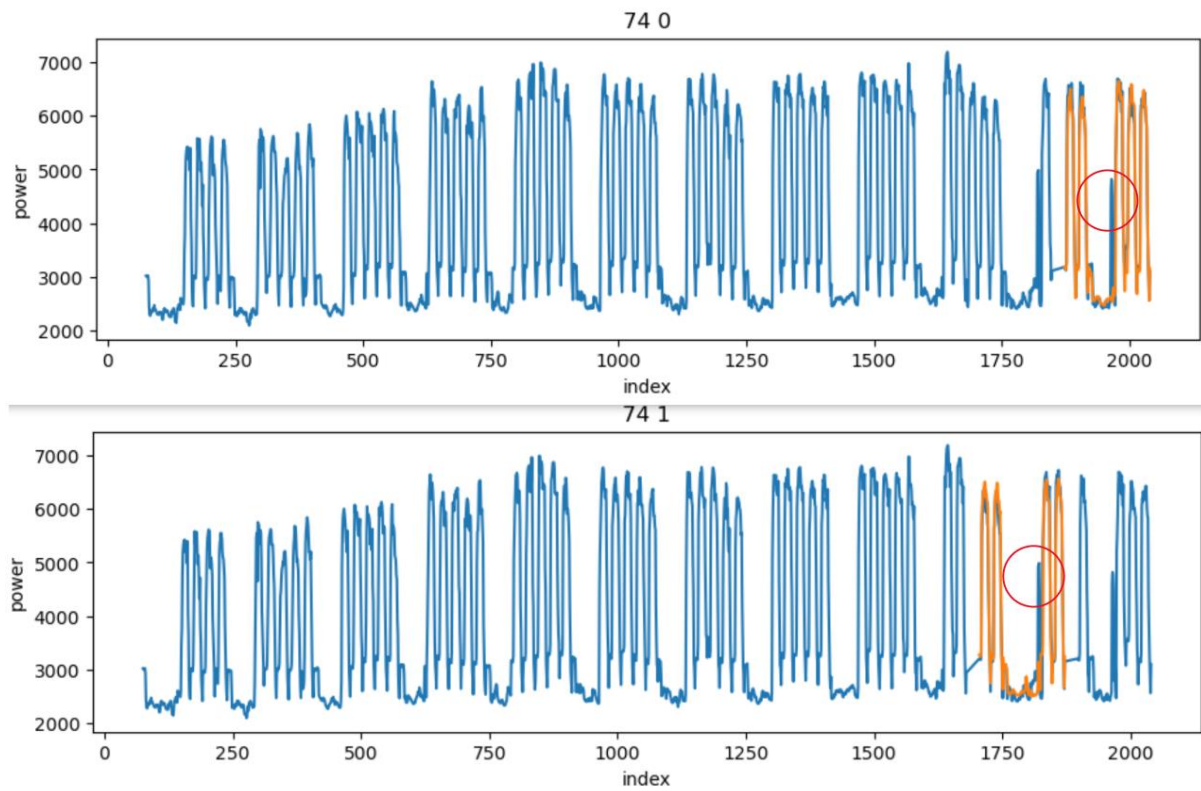
아래의 방법을 통해 직접 탐지하고 제거하였습니다.

방법1. dhc, hour별 전력량 평균 이용





방법2. Validation set 예측값 모니터링



위 방법들을 통해 총 1941 개의 데이터를 드랍하였습니다.