

시계열 분류에 따른 부품의 주문량 예측 모델 구축

팀 아베퍼

(이재헌, 송요셉, 조성혜, 최찬영)

데이터 및 과제

- 서비스 부품의 수요 예측 및 공급량 관리 모델 구축
 - 과제 1: LG 사의 서비스 부품 **미래수요를 정확하게 예측**하여 서비스 지연에 따른 고객 만족도 향상
 - 과제 2: 재고 부족/과잉으로 인한 **비용을 고려**하여 부품의 **공급량을 관리**하여 최적의 비용 시스템 구축
- 훈련 데이터
 - 3362개 부품의 일간/주간 수요 데이터 (기간: 2018.07.27 - 2023.08.04)
 - 3362개 부품의 운송 및 재고 유지 비용 데이터
- 제출 데이터
 - 선정된 10개 부품에 대한 과제 1, 과제 2 예측 값 (기간: 2023.07.31 ~ 2023.10.29)
- 모델 성능 평가 방법

task1: relative MSE compared to zero

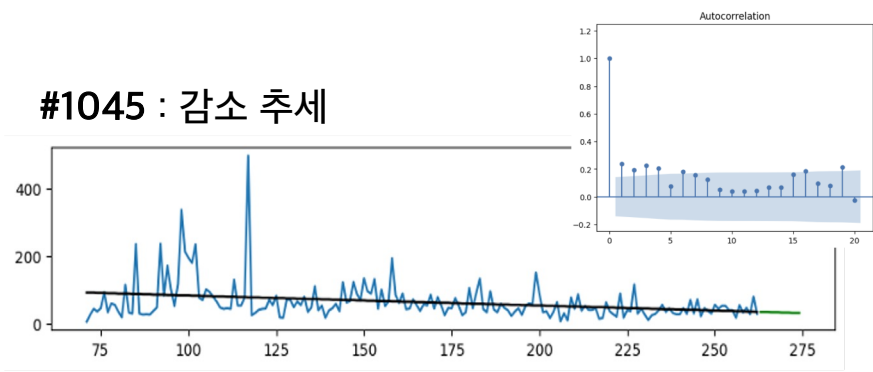
$$\begin{aligned} \bullet \ rMSE_i &= \frac{\sum_{t=1}^T (d_{i,t} - p_{i,t})^2}{\sum_{t=1}^T d_{i,t}^2} \\ \bullet \ \text{Score1} &= \sum_{i \in \text{Parts}} rMSE_i / |\text{Parts}| \end{aligned}$$

task2: relative cost compared to zero shipping

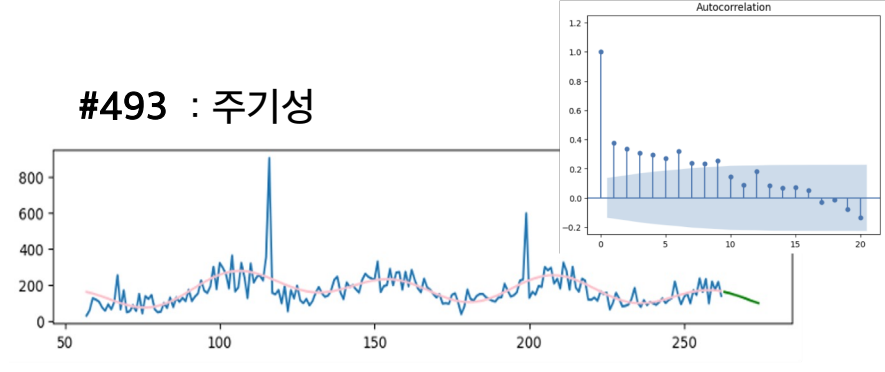
$$\begin{aligned} \bullet \ rCost_i &= \frac{\sum_{t=1}^T cost_{i,t}}{c_{air} \sum_{t=1}^T d_{i,t}} \\ \bullet \ \text{Score2} &= \sum_{i \in \text{Parts}} rCost_i / |\text{Parts}| \end{aligned}$$

시차에 따른 자기상관성을 기준으로 정상(Stationary)/비정상 시계열 분류

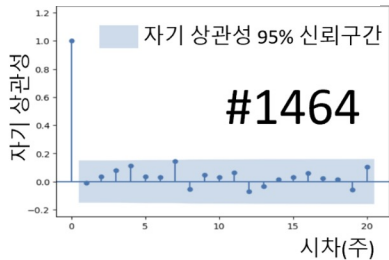
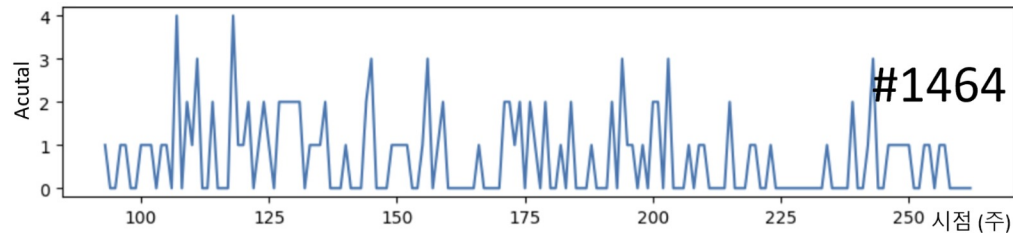
#1045 : 감소 추세



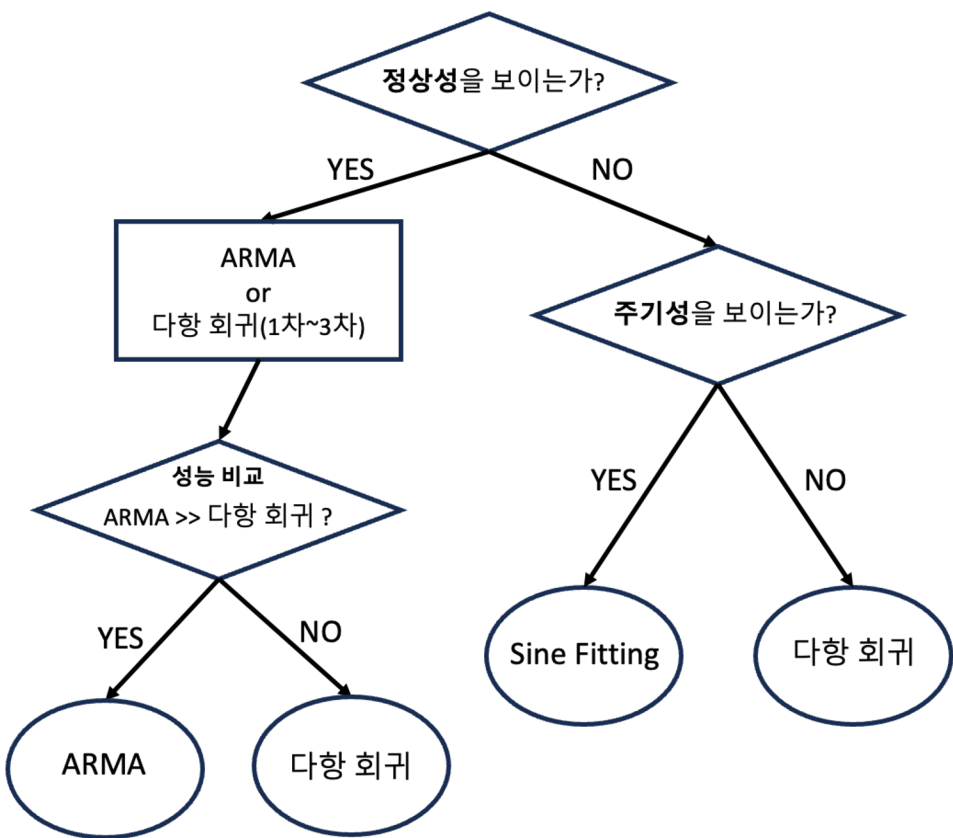
#493 : 주기성



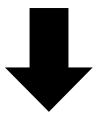
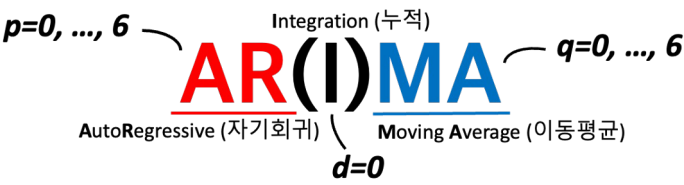
정상 시계열: 모든 시차에 대해 자기 상관성을 보이지 않음



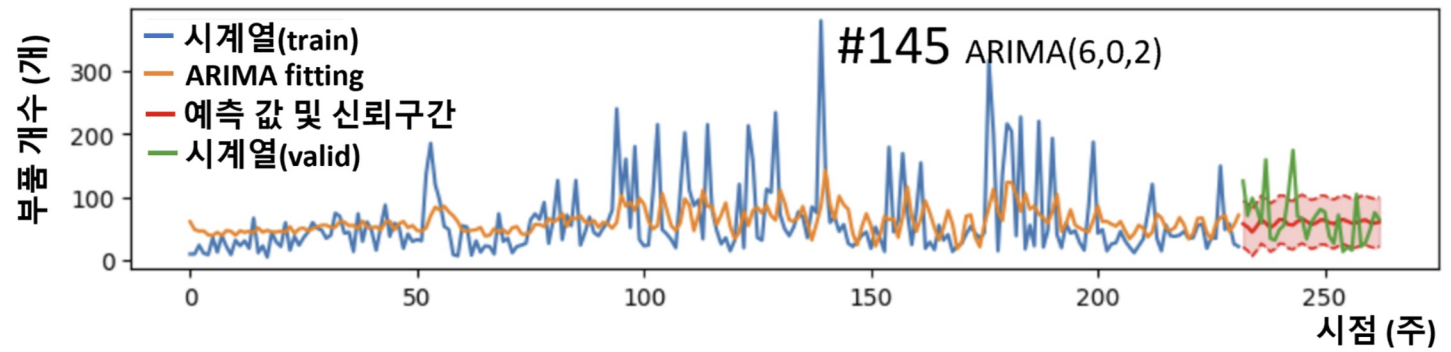
정상성/비정상성에 따라 모델을 선정함



정상(Stationary) 시계열에는 ARMA 모델을 적용함



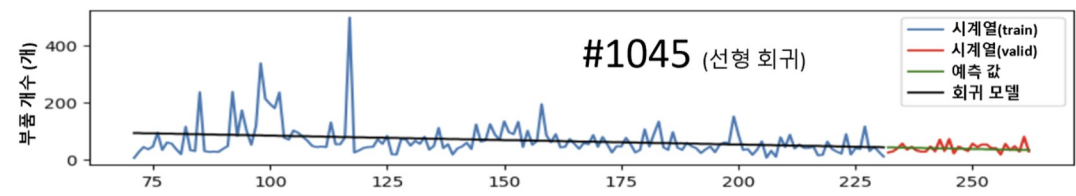
- Training Set : 2018 ~ 2022년의 주문량
- Validation Set : 2023년의 주문량



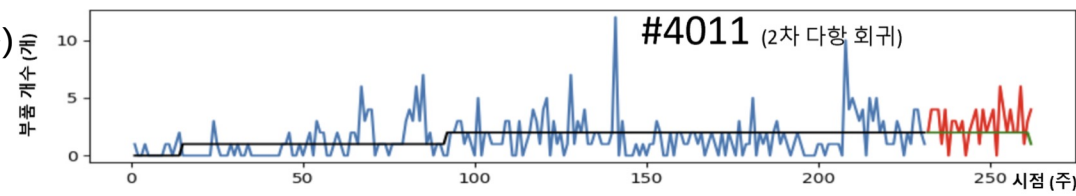
비정상 시계열에는 각 시계열의 추세를 반영한 회귀 모델을 적용함

- Training Set : 2018 ~ 2022년의 주문량
- Validation Set : 2023년의 주문량

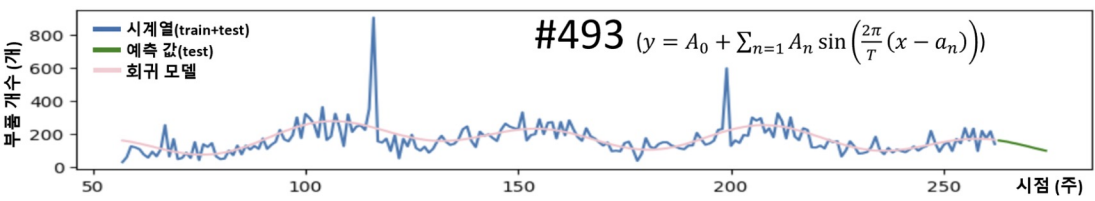
- 선형 회귀



- 다항 회귀 (2차~)



- 사인파 회귀



4개의 예측 모델 적용, Validation 결과가 가장 좋은 모델을 사용함

부품	데이터 특징	Linear	Polynomial	ARMA	Sine Fit	최종 모델
#4011	정상	0.2934	0.2460	0.4290	0.3481	다항 회귀 (2차)
#1045	비정상 (감소 추세)	0.1388	0.6343	0.3533	None	선형 회귀
#141	비정상 (증가 추세)	0.3801	0.3742	0.3972	None	다항 회귀 (2차)
#145	정상	0.2931	0.4832	0.2899	None	ARIMA (p=6, d=0, q=2)
#1464	정상	10	0.9583	0.9583	None	다항 회귀 (3차)
#3569	정상	0.375	0.5529	0.5385	None	선형 회귀
#4290	정상	0.2760	0.2436	0.3680	None	다항 회귀 (2차)
#493	비정상 (주기성)	0.2789	0.2050	0.0879	0.0653	Sine Fitting
#523	정상	0.5026	0.4870	0.4627	0.4152	ARIMA (p=2, d=0, q=1)
#779	정상	0.4803	0.2163	0.2360	None	다항 회귀 (2차)

Relative Total Cost 최소화 전략 - 비용의 상대적 비율과 재고 비축

모든 제품에 대해서 carrying, shipping, airplane cost의 비율은 다음과 같음.

Type	Carrying	Shipping	Airplane
Relative Cost	1	14.58	65.64

비행기 배송 비용 $\uparrow \Rightarrow$ 재고가 부족하지 않도록 미리 넉넉하게 주문해야 함.

Shipping $\times 4 \approx$ Airplane \Rightarrow 너무 많이 주문하면 주문 비용과 보관 비용이 커짐.

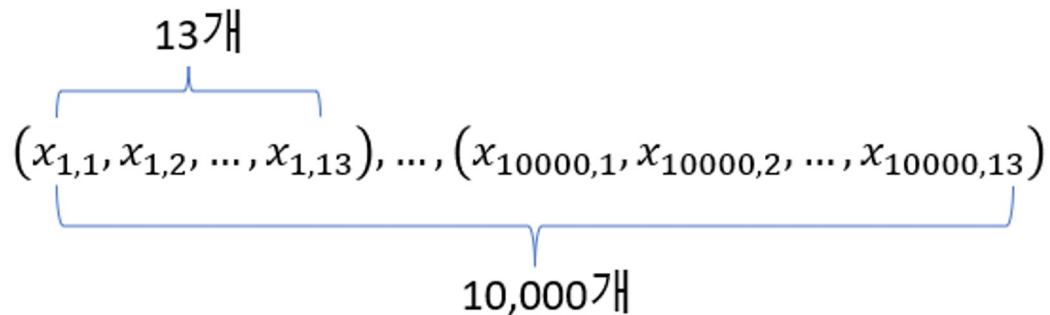
그러므로, 재고를 비축하되 적당하게 해야된다. \Rightarrow 얼마만큼이 적당한가?

재고를 얼마나 비축할 것인가?

1. 예측할 13주의 주문량의 확률 분포를 구한다.

e.g. Test 141 order : X , $x \sim p(x)$

2. 이 확률 분포를 기반으로 13주의 주문량 샘플을 10,000개 생성한다. (i.i.d)



재고를 얼마나 비축할 것인가? - 최적의 k, m 찾기

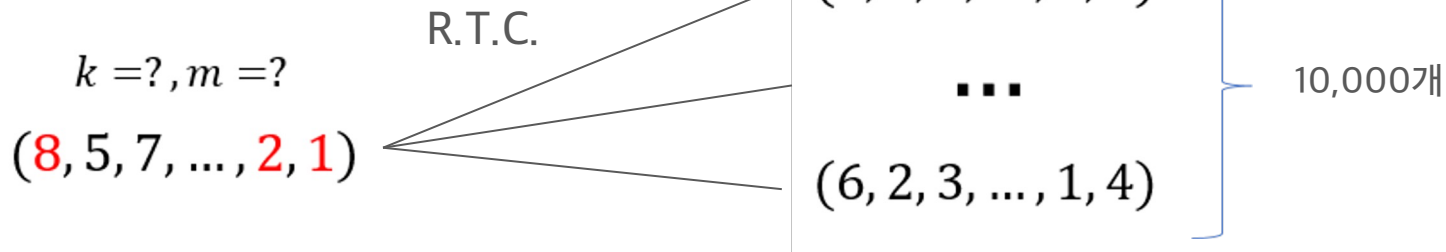
3. Task 1 예측 결과의 첫 주에 k 개를 더하고, 마지막 두 주에서 m 개씩 빼준다.

$$k = 4, m = 2$$

$$(4, 5, 7, \dots, 4, 3) \rightarrow (8, 5, 7, \dots, 2, 1)$$

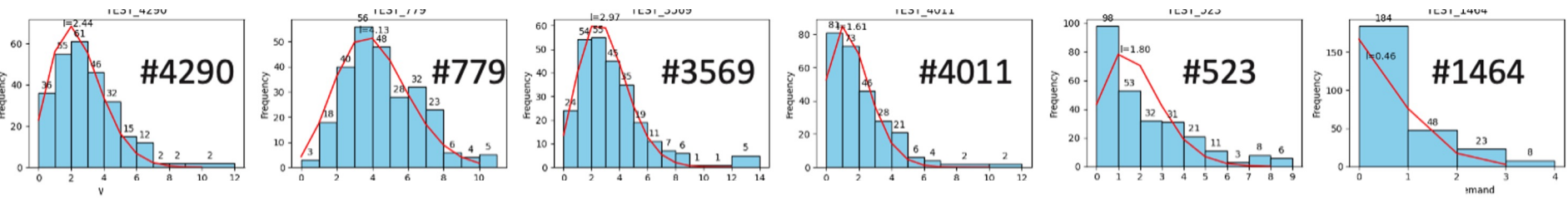
4. k, m 을 바꿔가며 랜덤 샘플과의 Relative Total Cost를 구한 후 평균 R.T.C가

가장 낮았던 k, m 을 찾는다.



정상성을 보이는 부품들은 푸아송 분포를 따른다고 가정

- 아래 6개의 부품은 특별한 추세나 주기성을 띄지 않고, **정상성을 보임**
: #4290, #779, #3569, #4011, #523, #1464 (총 6개)
- 수요가 완전히 무작위(random)로 발생한다고 가정
⇒ 수요 발생 간격이 푸아송 분포를 따름



— 푸아송 분포 ■ 주별 주문량의 도수 분포 (frequency distribution)

랜덤 샘플(step 1)을 위한 확률 분포 결정

예측할 13주의 주문량의 확률 분포 정하기

1. 주문량이 많은 부품(4개): 2023년도 주문량(최근 30주)의 주문량 분포 사용
ex) (200, 200, 100, 50) $\rightarrow p(200) = 0.5, p(100) = 0.25, p(50) = 0.25$
2. 주문량이 적은 부품(6개): 2023년도의 주문량 평균을 따르는 푸아송 분포 사용
ex) (2, 1, 0) \rightarrow 푸아송 분포 with $\lambda=1$

Grid Search로 최적의 k , m 값을 구함

부품	데이터 분포	Best K	Best M
#4011	푸아송 분포 ($\lambda=1.61$)	3	3
#1045	Empirical	57	0
#141	Empirical	4937	1571
#145	Empirical	127	32
#1464	푸아송 분포 ($\lambda=0.46$)	×	×
#3569	푸아송 분포 ($\lambda=2.97$)	4	1
#4290	푸아송 분포 ($\lambda=2.44$)	6	2
#493	Empirical	166	18
#523	푸아송 분포 ($\lambda=1.80$)	5	1
#779	푸아송 분포 ($\lambda=4.13$)	5	1

추가적인 고려사항

✓ #145: 분산에 비해 K 가 너무 크다.

⇒ 재고유지비를 줄이기 위해,
첫 주에 K 를 모두 더하는 대신 1~3 주에 K 를 나눠서 더해주었다.

✓ #1464: 추출된 표본 값이 모두 0인 경우 자주 존재,
Relative Total Cost를 계산할 수 없었다.

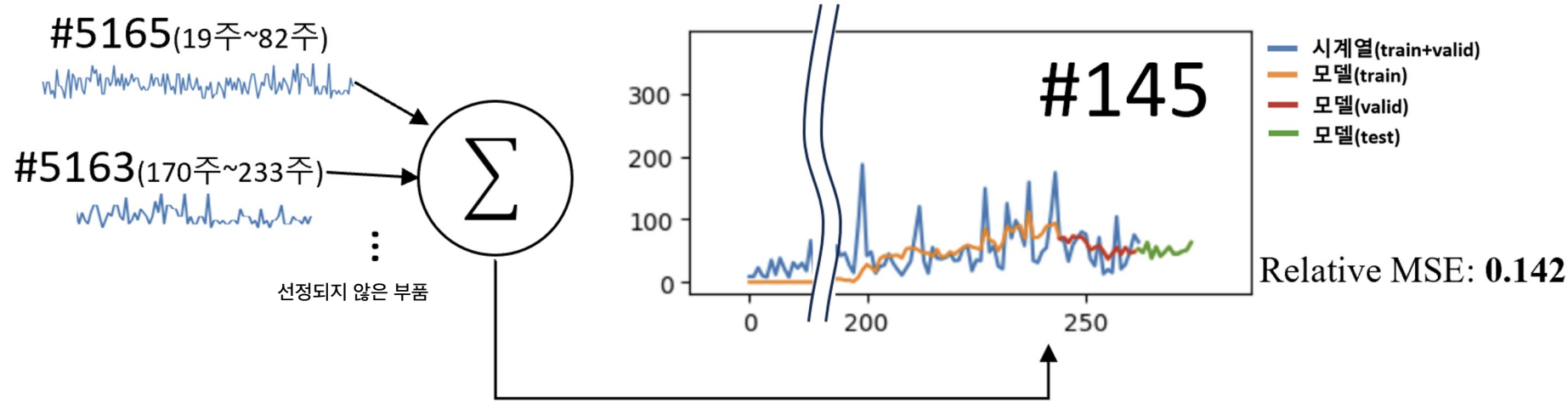
⇒ 분산을 고려해 첫 세 주에 (3, 2, 1)을 예측하였다.

결론 및 한계

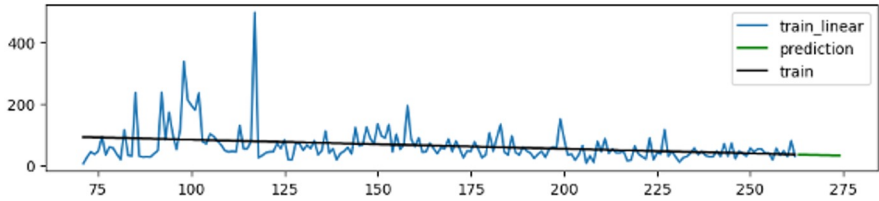
- 과제 1에서는, 정상/비정상 시계열을 분류하여 각 분류에 따라 모델을 선정하였음.
 - 주문량이 많고, 주거나 추세가 명확한 부품은 안정적인 예측 결과를 보임. (ex. #493, #1045)
 - (한계점) 회귀 모델의 경우 추세를 설명하기에는 적합할 수 있으나, underfitting이 발생했을 우려가 있기 때문에 ARMA 모델과의 조합(ensemble)을 통해 모델 복잡도를 높여 이를 해결할 수 있을 것임.
- 과제 2에서는, 정상/비정상 시계열 분류에 따라 각각 Empirical / Poisson 분포를 적용하여 최적의 주문량 조정 값 (k , m)을 찾아내었음.
 - (한계점) Empirical / Poisson 분포 모두 적은 기간의 시계열(2023년 시계열)에서 얻었기 때문에 예측 기간에 발생할 부품 수요를 모두 예측하는 데 한계가 있음.

Thank you :)

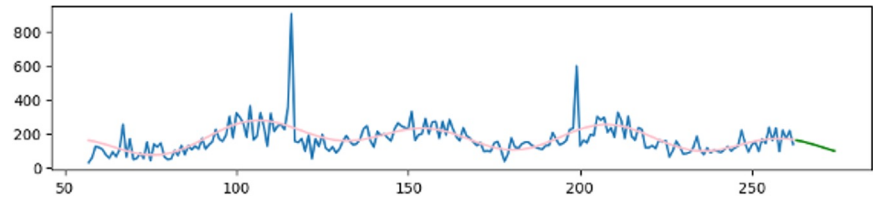
선택되지 않은 부품의 시계열을 활용하고자 Cross-Correlation 값이 높은 시계열을 조합하여 서비스 부품 수요를 예측하였지만, overfitting이 발생하여 예측 성능이 크게 떨어짐



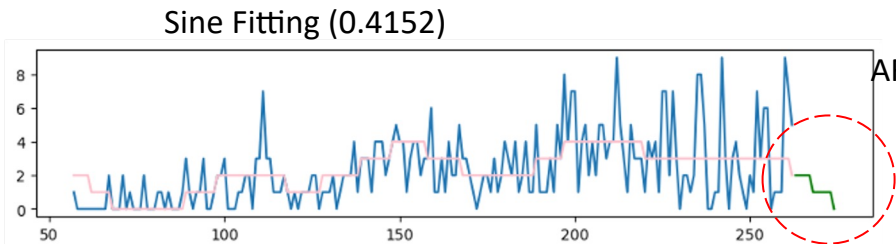
특수한 사례들



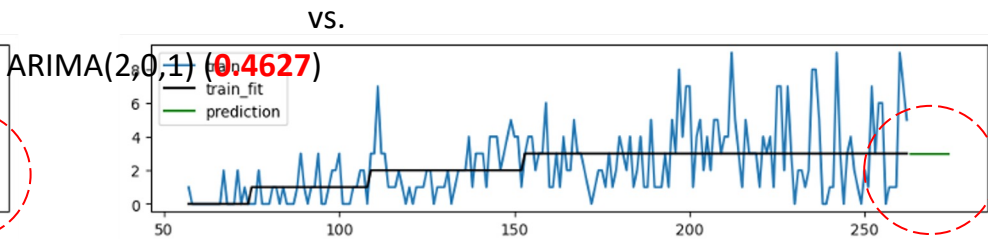
✓ (#1045) 선형 모델에 감소 추세가 잘 반영됨



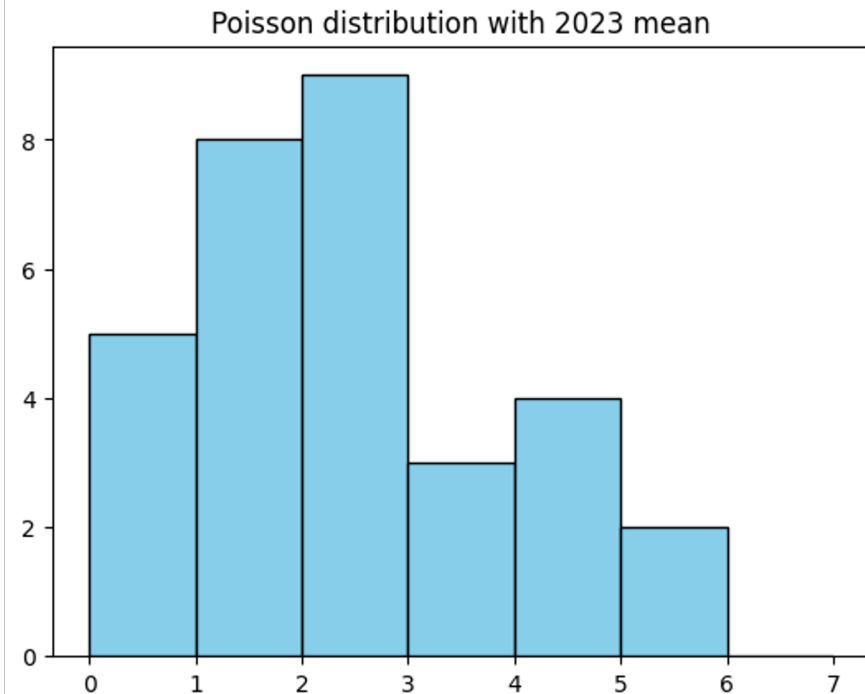
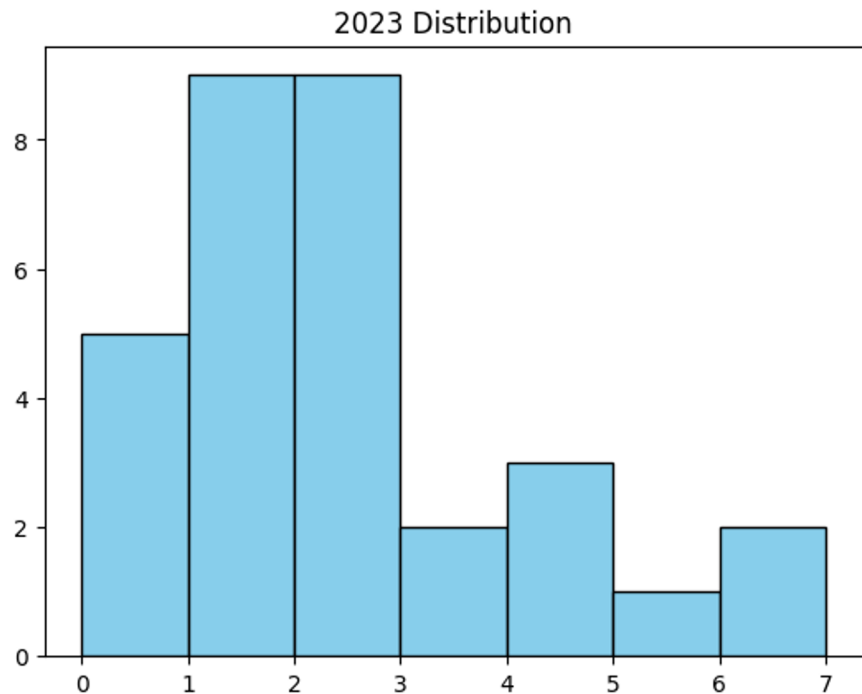
✓ (#493) Sine Fitting 모델에 주기성이 잘 반영됨



✓ (#523) Sine Fitting 모델이 훈련셋에 overfit 됨
⇒ Sine Fitting Model의 Validation Score가 더 좋음에도 ARMA를 사용



2023년도 주문량이 푸아송 분포를 따르는가?



왜 첫 주와 마지막 두 주 값만 조정했을까?

- 재고 유지비를 감안하더라도 비행기 운송비가 매우 크기 때문에, 첫 주에 peak가 떠서 재고가 부족한 상황을 방지하고자 첫 주에 k 를 모두 배치하였다.
- 재고 유지비를 줄이기 위해 마지막 두 주에서 모두 m 씩 예측값을 빼주었다.
- 실험적으로 1~3주에 나눠 배치하는 방식 등 다양한 방식을 비교해 보았으나, 유의미한 차이가 보이지 않았다.