

분석배경 및 제안

- 기획 재정부에 따르면 향후 5년간 공공기관의 주산과 부채 규모가 증가하고, 부채비율은 '24년까지 171.4% 수준이 될 것으로 전망
- 현재 공공기관의 효율적인 예산 활용이 요구되고 있는 상황

'20~'24년 부채비율 전망(조원, %, %p)

구 분	'19년(실적)	'20년	'21년	'22년	'23년	'24년
자산 (조원)	794.6	824.5	860.4	901.0	934.2	975.1
부채 (조원)	497.2	521.6	540.8	571.0	591.1	615.8
자본 (조원)	297.5	302.9	319.6	330.0	343.2	359.3
부채비율(%)	167.1	172.2	169.2	173.0	172.2	171.4

'20~'24년 금융부채 전망(조원)

구 분	'19년(실적)	'20년	'21년	'22년	'23년	'24년
금융부채(조원)	378.4	400.8	415.2	438.4	456.6	476.6
(총자산 대비, %)	47.6	48.6	48.3	48.7	48.9	48.9

※ 자료: 기획재정부, 「'20~'24년 공공기관 중장기 재무관리계획」, 2020.09



조달청의 조달 업무에 있어 공공기관의 시간과 비용을 획기적으로 줄이는 것이 목표



관리도 / 추정가격예측모델

• 조달청 공고 현황 정보 미흡

입찰 공고를 확인했을 때 공고 기관과 수요 기관이 분리되어 있기 때문에 정확한 수요기관의 예산 지출 현황(물품/공사/용역)을 파악하기가 어려움이 있다고 판단하였다.

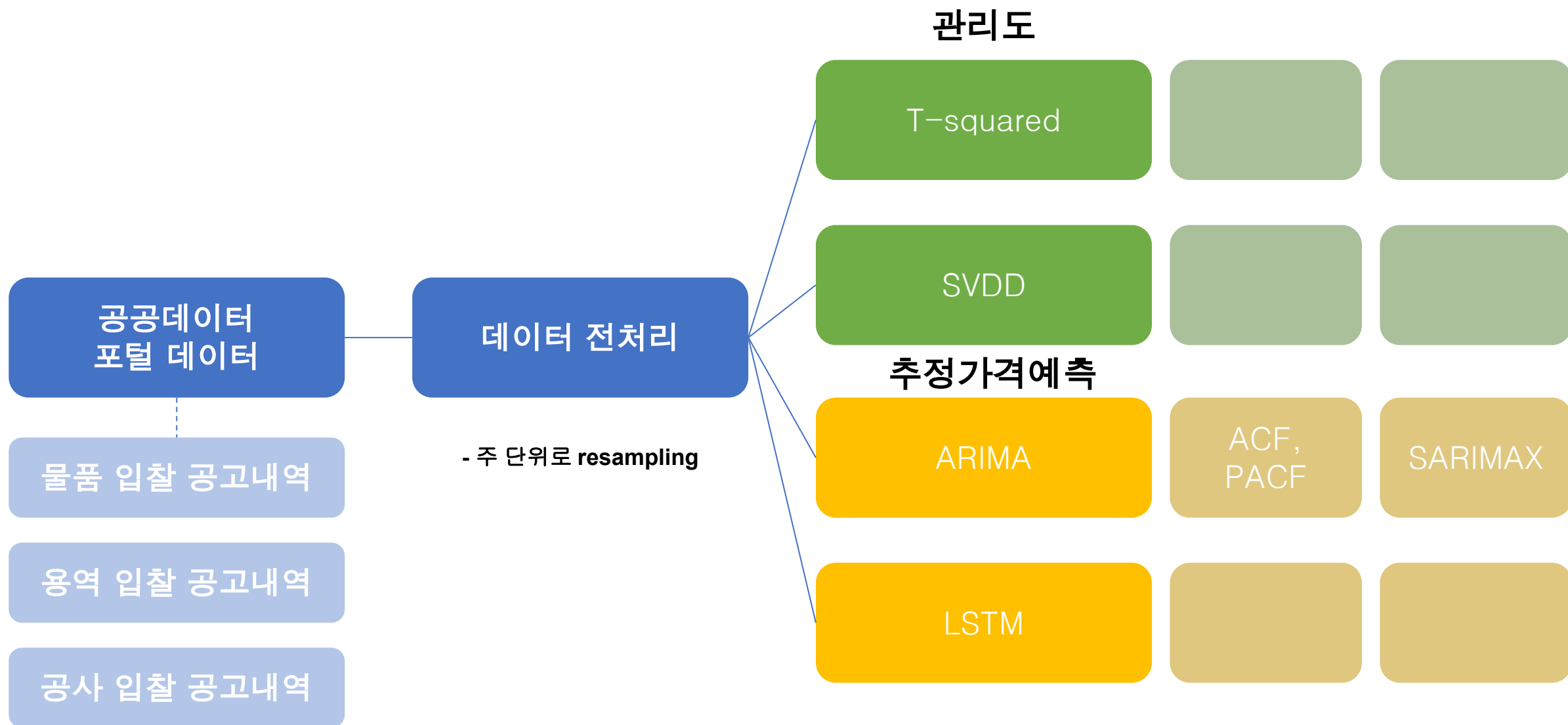
수요기관별 공고 현황에 대해 과거부터 누적된 현황을 한 눈에 파악할 수 있는 지표 요구

• 기관의 효율적인 예산 관리 지원

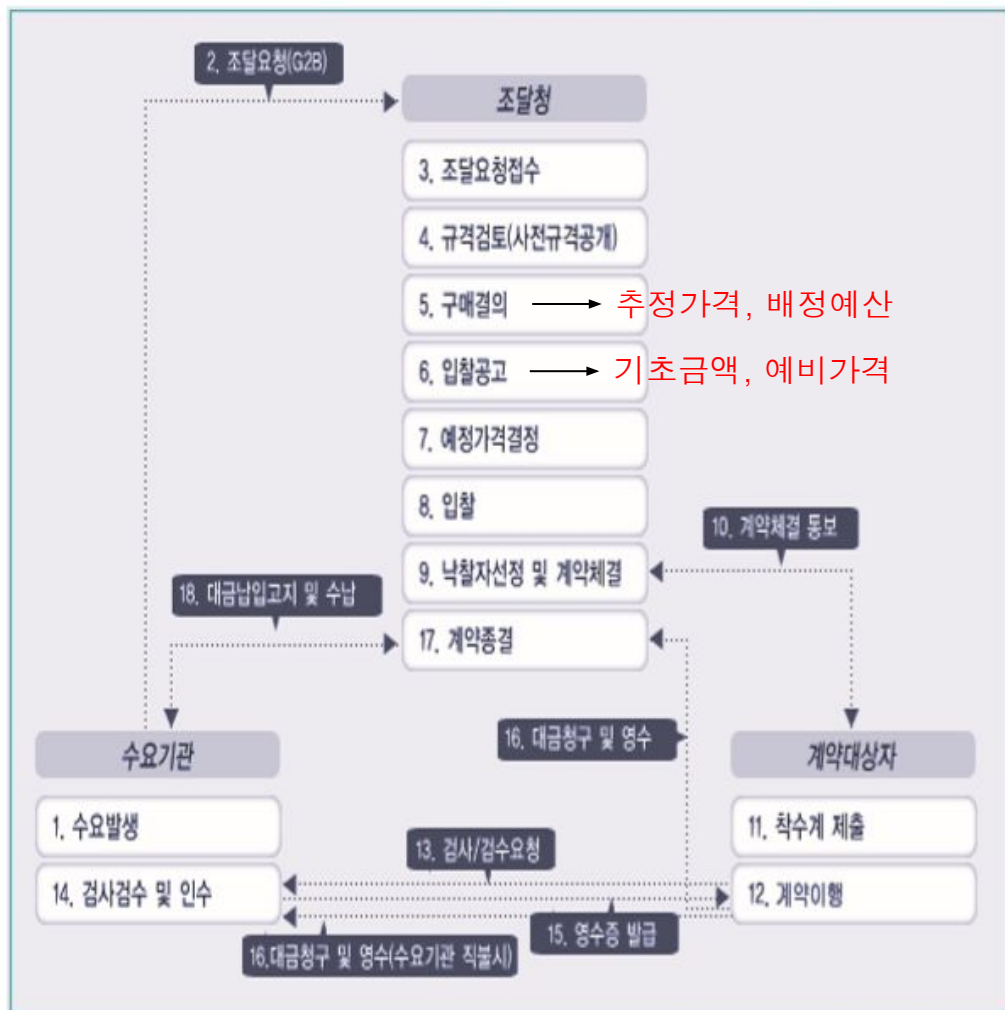
기간별로 평균적인 금액보다 높은 금액이 게시된 구간에 알람을 주어 특수 원인 변동을 식별

추후 발생될 입찰 공고의 추정가격 예측을 통해 효율적인 예산 관리 지원

분석 개요_flow



입찰 프로세스



- 규격검토 및 구매결의 □ 입찰 공고 □ 예정가격 결정 □ 입찰 순

- 추정가격 : 기관에서 사용한 것으로 나타내는 제일 처음 금액 □ 기관에서 사용하려는 '예산'으로 정의
기관의 입장에서 예산관리를 도우는 것이 목적이므로, '추정가격'을 사용

'일주일 치'의 합산 추정가격으로 선정

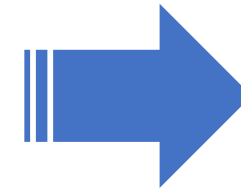
- 기관이 입찰공고를 올리기 일주일 전에 발주계획을 세우는 경우가 많기 때문에
- 언제 예산이 많이 사용되었는지 보기 쉬움
- 추정금액의 변동이 줄어들어 일주일 치의 추정 금액의 변동을 관리할 수 있음

데이터 선정_추정가격예측 공통



수요기관명	
방위사업청	62468
한국수자원공사	13366
서울특별시	11746
한국도로공사	9715
서울교통공사	8109

전체 데이터 셋에서
공고 개수가 많은
기관을 대상으로 선정
=> **한국수자원공사**

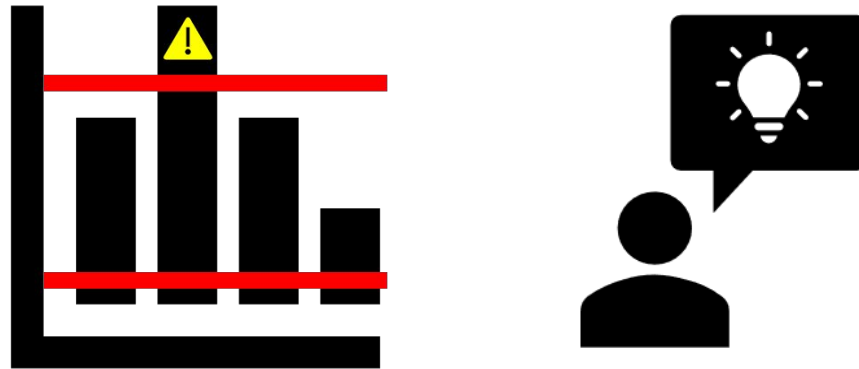


추정가격	
date	
2018-04-08	1213394500.00000
2018-04-15	2100809455.00000
2018-04-22	5161653780.00000
2018-04-29	1889171056.00000
2018-05-06	8386696600.00000
...	...
2021-12-05	15323871280.00000
2021-12-12	15572623196.00000
2021-12-19	23078195505.00000
2021-12-26	6937094040.00000
2022-01-02	4566138800.00000

분석 목적_관리도

입찰공고의 추정가격 특수 원인 변동 식별

: 기관 및 대상별로 입찰 공고에 게시된 추정 가격을 바탕으로 과거 누적 데이터 보다 예산 소모가 크게 발생할 경우(공정이탈상태) Alarm을 일으켜 원인 식별



입찰 공고의 추정가격이 안정상태에 있는지 판별
또 안정상태로 유지함으로써 입찰 공고의 효과적인 관리

활용 데이터 및 전처리

활용데이터	활용내용
물품 입찰공고 내역(2017~2021)	물품의 추정가격 및 입찰공고 게시일자를 알기 위해서 사용하였음
용역 입찰공고 내역(2017~2021)	용역의 추정가격 및 입찰공고 게시일자를 알기 위해서 사용하였음
공사 입찰공고 및 진행 내역(2017~2021)	공사의 추정가격 및 입찰공고 게시일자를 알기 위해서 사용하였음

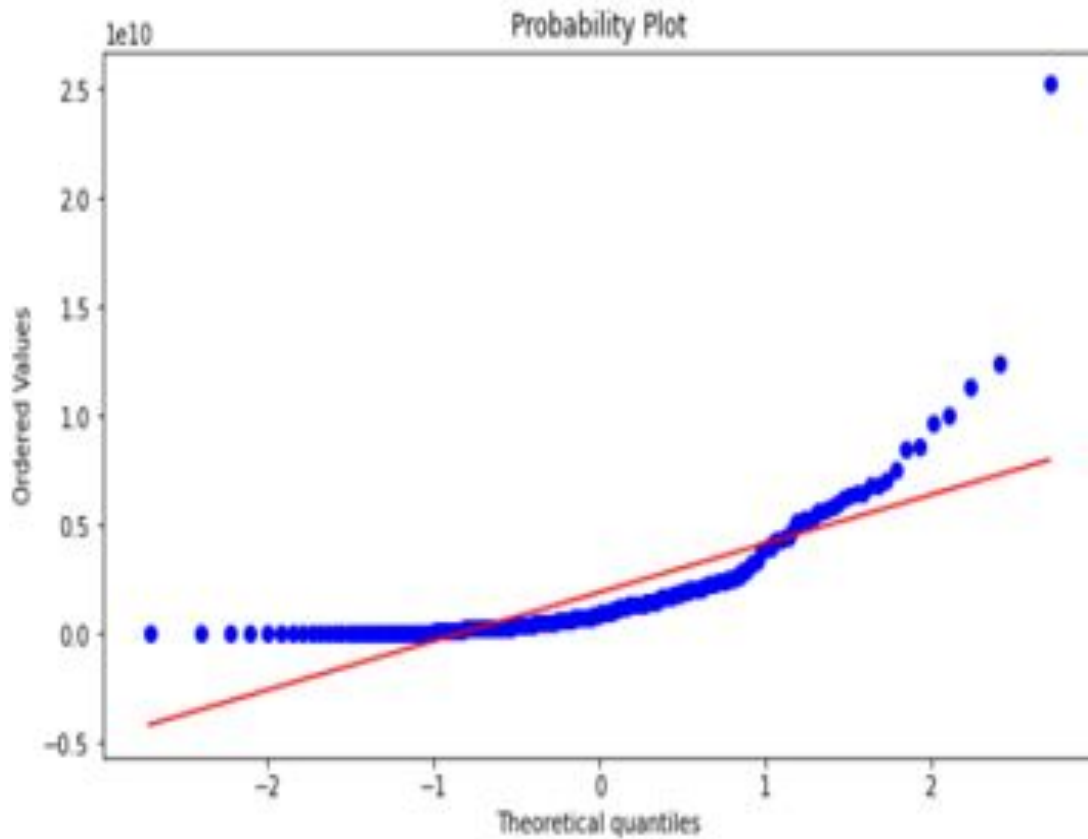
등록유	공고게시일	자	입찰공고	수요기관명	수요기관코	긴급공고	정정공고	취소공고	추정가격	계약방	낙찰자결정방법	입찰계약방법	예가방	업종제한내용	지역제한내용	
0	나라장터 (G2B)	2020-03-31	자제조	0	전라북도 정읍시	4690000	Y	N	N	87100000.0	전자입찰	추정가격 1억미만(PQ비대상)	제한경쟁	복수예가	NaN	전라북도[45000]
1	나라장터 (G2B)	2020-03-31	자제조	0	울산광역시 북구	3720000	N	N	N	42171818.0	전자입찰	제한최저	수의(종역소역수의	복수예가	[학술 연구용역(1169)]	부산광역시[26000], 울산광역시[31000], 경상남도[48000]
2	나라장터 (G2B)	2020-03-31	자제조	0	서울특별시	6110000	N	N	N	50000000.0	전자입찰	제한최저	수의(종역소역수의	복수예가	NaN	서울특별시[11000]
3	나라장터 (G2B)	2020-03-31	자제조	1	충청북도교육청 충청북도재건교육지원청	8031000	Y	N	Y	62630000.0	전자입찰	추정가격 1억미만(PQ비대상)	제한경쟁	복수예가	[안전진단전문기관(건축)(1387)]	충청북도[43000]
4	나라장터 (G2B)	2020-03-31	자제조	0	충청북도교육청 충청북도재건교육지원청	8031000	Y	N	N	70980000.0	전자입찰	추정가격 1억미만(PQ비대상)	제한경쟁	복수예가	[안전진단전문기관(건축)(1387)]	충청북도[43000]

- 각 수요기관에 대해서 입찰 공고 게시 건수가 **1000회** 이상인 기업이 전체 공고의 **60%**를 차지

□ 이에 대한 기관을 주 **target**으로 하여 진행

- 입찰 공고는 있으나, 추정 금액이 **0원**으로 아직 산정되지 않은 값들 □ 관리상한선 선정에 영향을 끼치므로 제거
- 기관에서 입찰공고에 명시한 “추정금액”을 기관에서 사용하려는 “예산”으로 고려하여 관리도를 구축

T-square + Bootstrap



T-square 다변량 관리도는 각 변수에(물품/용역/공사) 대해서 정규분포를 따라야 사용 가능하다.

□ QQplot을 그려보았으나, 실제 데이터는 정규분포를 따르지 않음

❖ بوت스트랩 방식 : 분포를 따른다는 가정이 필요 없고, 정상데이터에 **alpha**값이 고정되므로, 우리가 원하는 성능의 **alpha**값을 가지는 관리도가 만들어진다.

데이터 분석 - t-square 관리도 제작

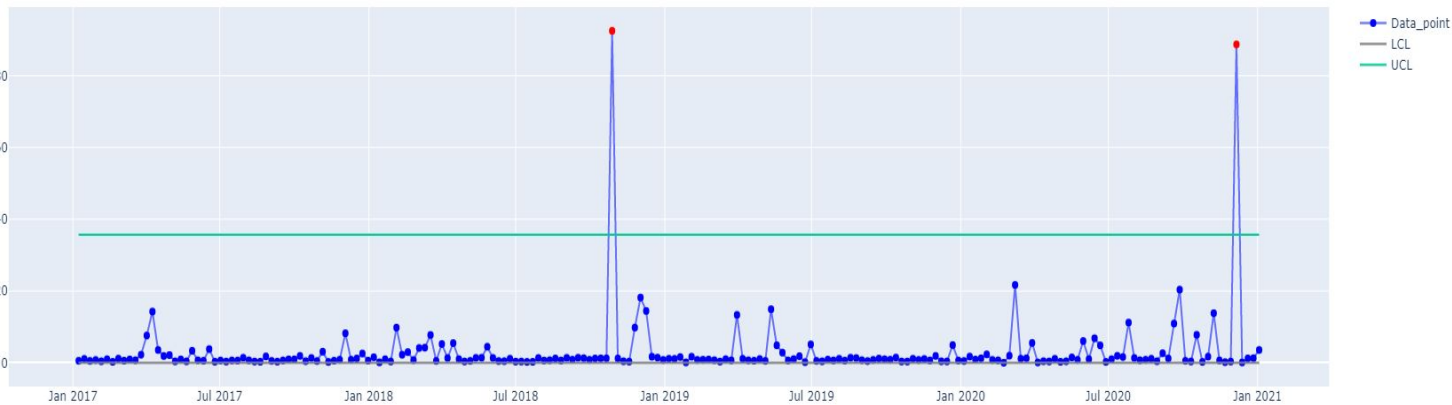
```
a, b, problem_row, total, f, df_공사_lw, lst_problem_공사, df_T_square_Text, df_t_s = t_square_interpret('5710000', alpha=0.01, bs_size=1000)
```

```
df_용역_입찰공고_내역_3['수요기관코드'].value_counts().index
```

```
Index(['1690000', '6110000', 'B500001', '1613000', 'B500004', 'B551',  
      'B552584', '6500000', '6460000', '5710000',
```

```
...  
'Z014609', 'Z003100', '7191208', '8530063', '7451043', '5150047',  
'Z024461', '5340050', 'Z039055', '7679031'],  
      dtype='object', length=17407)
```

5710000기관의 2017 ~ 2020의 물품 / 용역 / 공사의 통합 T_square 관리도 (Apply bootstraps)



목적 : 기관의 입장에서 예산관리를 진행

□ 데이터 내의 “예산” 사용은 물품/공사/용역 분야로 나누어져 있다. 이를 통합할 경우,

① 각 분야에 쓰이는 예산 규모가 다름

② 어느 분야에서 예산 사용량이 큰지 확인이 불가하기 때문에 각 분야를 나누어 관리도를 구축하는 것을 목표

기관의 예산 사용 데이터

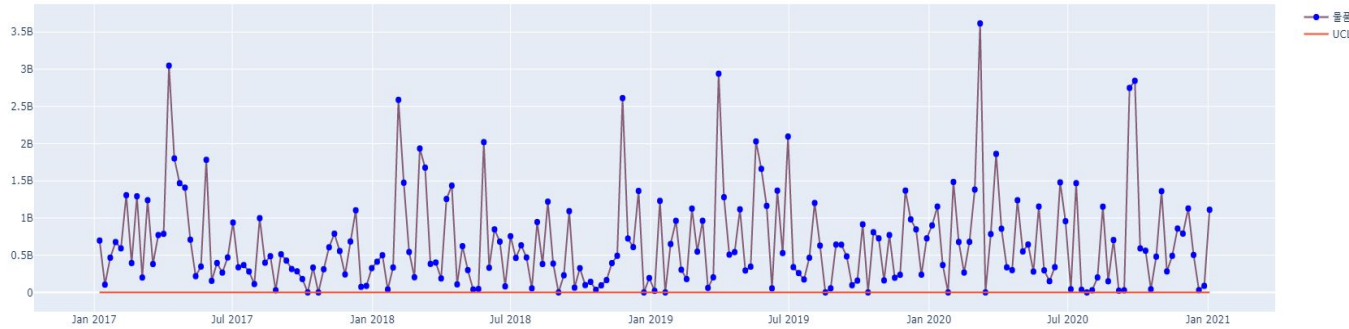
- 물품/공사/용역 각 분야에 쓰이는 예산 규모가 다름

- 예산 사용량이 일반적인 경우보다 큰지 확인이 불가

□ 물품/공사/용역 각 분야를 나누어서 관리도를 구축

데이터 분석 - t-square 관리도 결과

5710000기관의 2017 ~ 2020의 물품 관리도 (Apply bootstraps)



표준화된 평균과의 거리를 나타내고, 변수 간 관계까지 고려하는 관리도

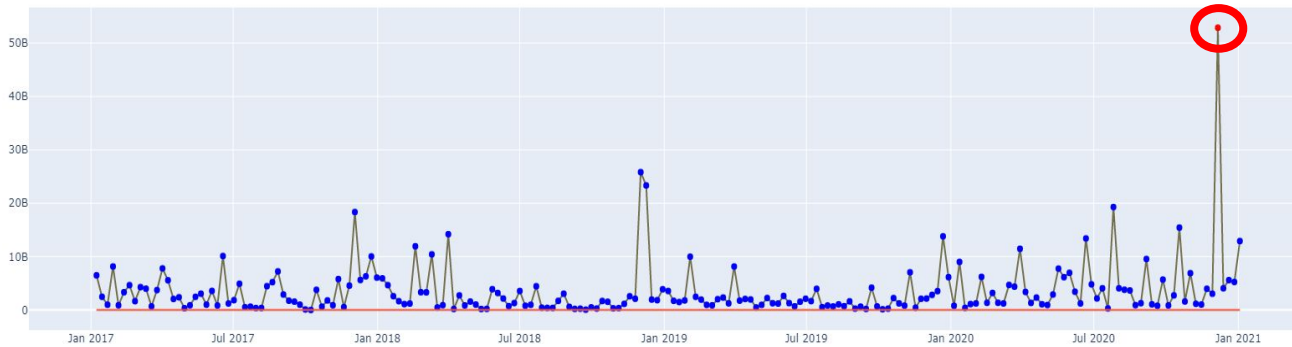
- '우리가 관리하려는 값' □ 관리도의 '하나의 점'
- 하나의 점이 관리한계선을 넘게 된다면 이상치로 인식

□ 빨간색 점으로 바뀌고 Alarm 일으킴

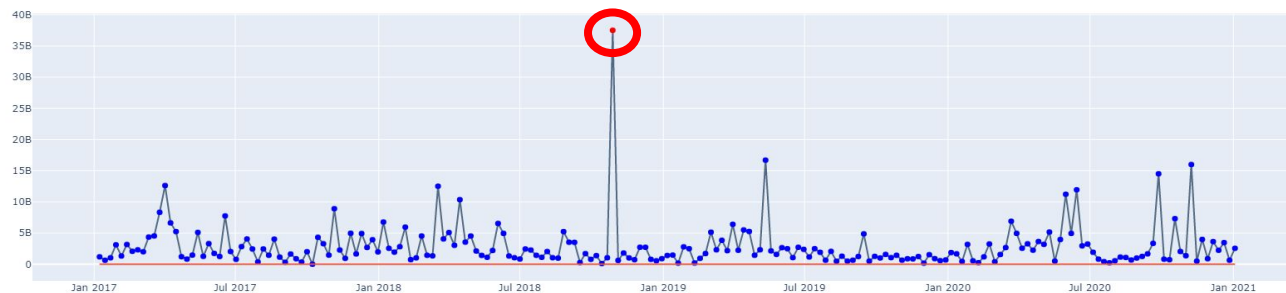
□ 알람이 일어난 점에 대해서 일반적인 값들과의 차이점 확인할 수 있도록 예산 시계열 그래프 구현

또한, 표의 형태로 '어느 분야의 예산'에서 문제가 있는지 알 수 있음

5710000기관의 2017 ~ 2020의 용역 관리도 (Apply bootstraps)



5710000기관의 2017 ~ 2020의 공사 관리도 (Apply bootstraps)



	용역_t	공사_t	물품_t	용역_real	공사_real	물품_real
공고게시일자						
2018-10-28	0.912021	91.424912	.751189	309000000.000000	37521366363.000000	95311818.000000
2020-12-06	88.188989	0.048389	0.006626	52876390611.000000	3625478185.000000	1128080000.000000

데이터 분석 - SVDD 관리도 제작

- 사용자는 사용 목적에 따라 관리도를 선택하여 사용할 수 있다.
- 기존 T-square 관리도 : '표준화시킨 평균과의 거리'를 중심으로 해서 관리도를 구축
- SVDD 관리도 : 정상데이터를 포함하는 최소 부피의 입체 도형을 찾는 관리도

* 구현 과정

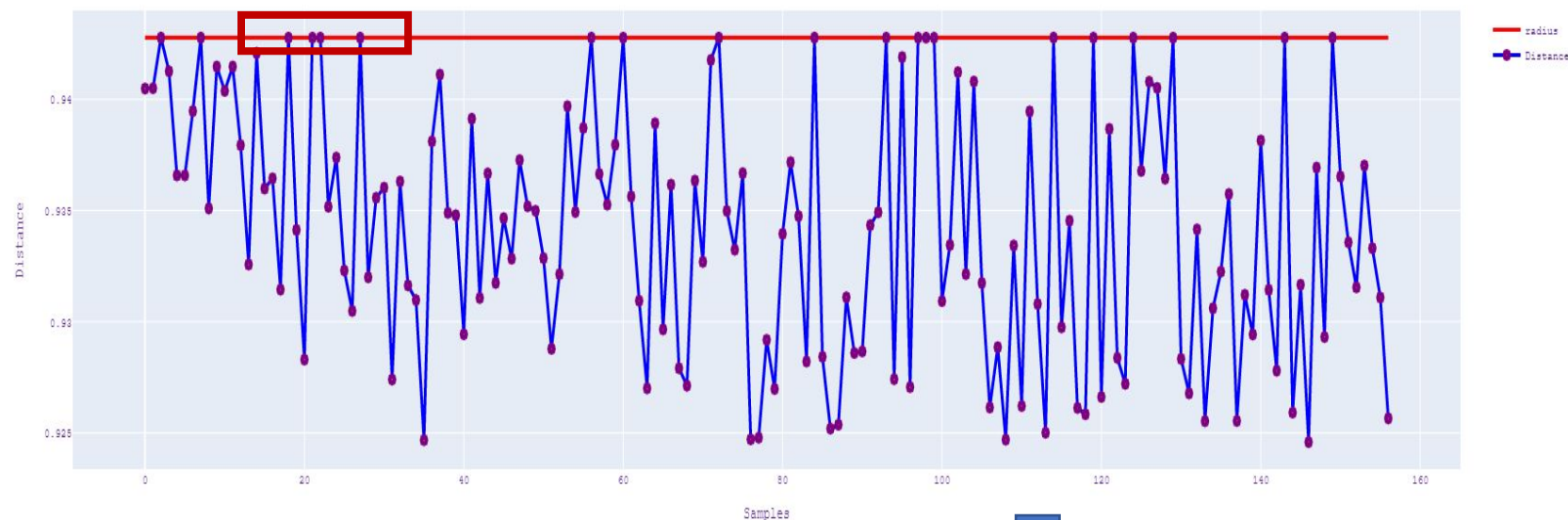


* 이 과정을 매해 갱신해 나간다.

이후, 현재 데이터로 정상/비정상 판별

데이터 분석 - SVDD 관리도 결과

Control Chart



확대



*** Fitting of the SVDD model is completed. ***

running time	= 0.1287 seconds
kernel function	= rbf
iterations	= 11
number of samples	= 157
number of features	= 3
number of SVs	= 20
ratio of SVs	= 12.7389 %
accuracy	= 95.5414 %

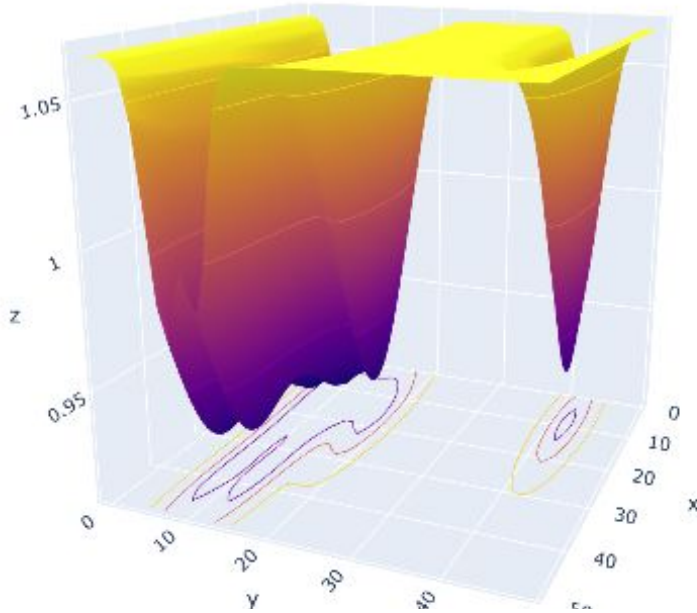
*** Prediction of the provided data is completed. ***

running time	= 0.0040 seconds
number of samples	= 157
number of alarm	= 7

이상치가 나올 확률 = 4.4586

관리상한선을 벗어난 이상치

데이터 분석 - SVDD 관리도 결과



*** Fitting of the SVDD model is completed. ***

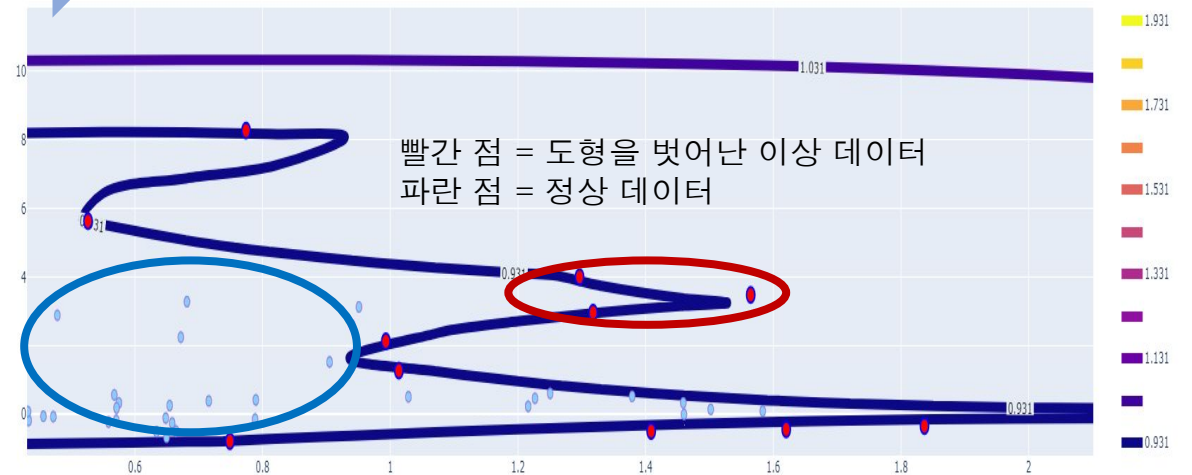
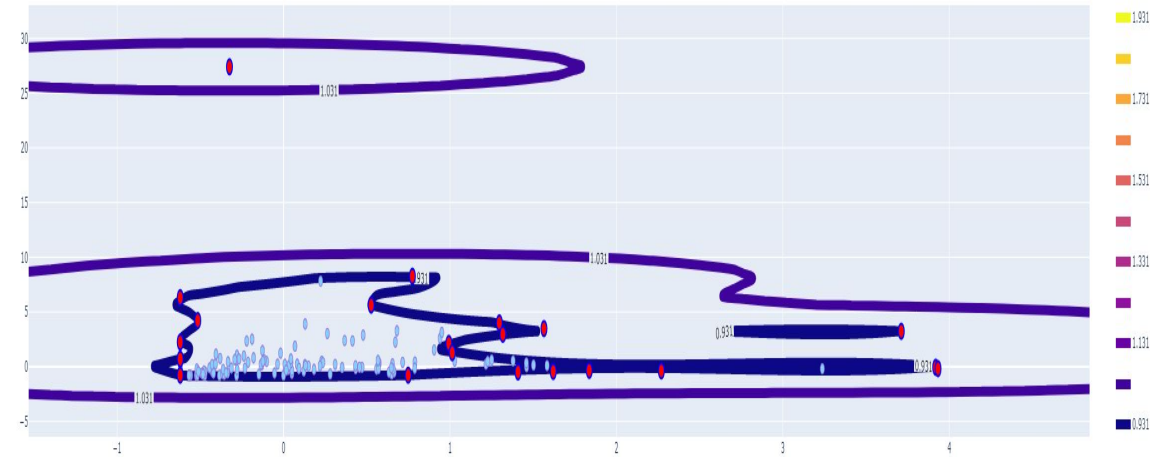
running time = 0.0439 seconds
kernel function = rbf
iterations = 11
number of samples = 157
number of features = 2
number of SVs = 21
ratio of SVs = 13.3759 %
accuracy = 89.1720 %

*** Prediction of the provided data is completed. ***

running time = 0.0040 seconds
number of samples = 157
number of alarm = 17

Calculating the grid scores (0050*0050)...

Calculation of the grid scores is completed. Time cost 0.1546 seconds

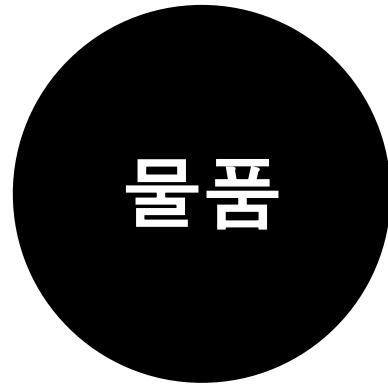


빨간 점 = 도형을 벗어난 이상 데이터
파란 점 = 정상 데이터

분석 목적_추정가격예측 공통

대상(물품/용역/공사)별 추정 가격 예측

: 대상별로 입찰 공고에 게시된 추정 가격을 바탕으로 미래의 가격 형성 범위를 예측



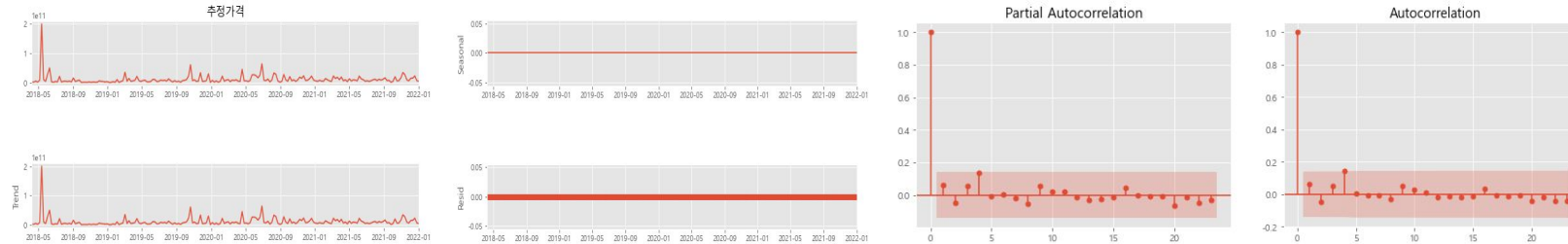
조달 예산 관리에 효율성 제고

과거 데이터 기반으로 추정가격이 높게 형성되는 구간에 대한 예측 정보 제공

분석 과정_데이터 전처리 및 EDA

1. 정상시계열 검증

: Data decomposition / ACF / PACF



✓ 계절성이나 특이 추세성 확인

✓ 시계열 데이터 정상성 확인

✓ ARIMA Model 파라미터 계수 추측

단위근 검정 ¶

```
from statsmodels.tsa.stattools import adfuller
result = adfuller(weekly_df_fin[1:])
print('ADF Statistic: %f'%result[0])
print('p-value : %f'%result[1])
print('Critical values:')
for key, value in result[4].items():
    print('%t%s: %3f'% (key, value))
```

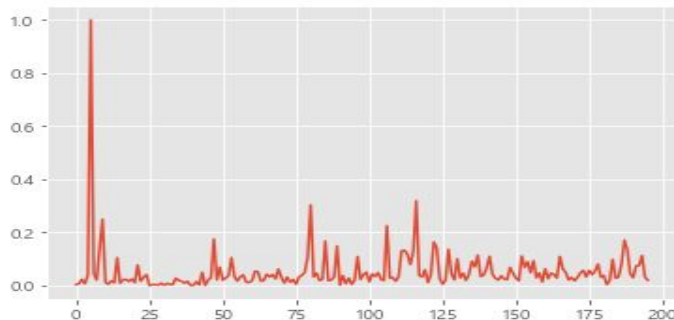
ADF Statistic: -13.060865
p-value: 0.000000
Critical values:
1%: -3.465
5%: -2.877
10%: -2.575

ADF Static value가 Critical values
보다 작은 값을 가지면 해당 데이터는
정상 시계열 데이터로 간주한다.

✓ 단위근 검정은 (비정상) 시계열에 대해 확률적 추세 여부를 검정한다.
즉, 비정상 시계열인지 검정하는 문제로 귀결된다.

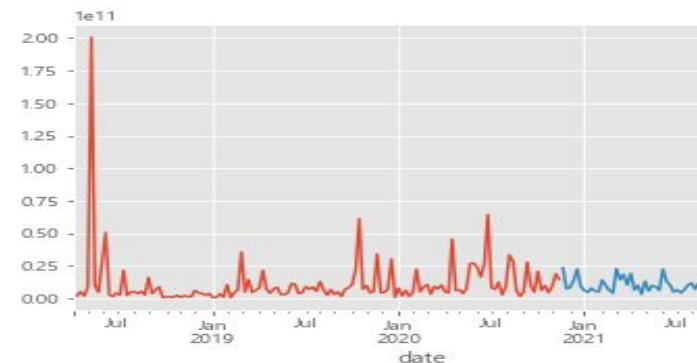
2. Feature scaling(minmax Scaling)

: 추정가격(feature)의 크기를 통일하기 위해 크기 단위 조정



3. Train / Test set 분리

: 모델 적합 전 train set, test set 을 분리



분석 과정_모델링

4. 모델적합

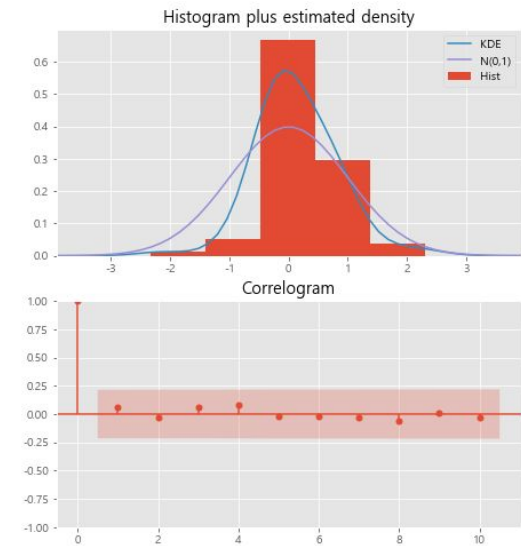
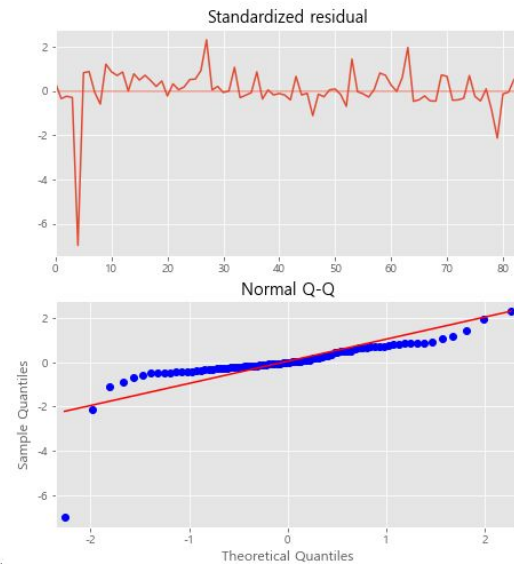
: auto_arima library 사용

SARIMAX Results						
=====						
Dep. Variable:	y	No. Observations:	137			
Model:	SARIMAX	Log Likelihood	-3441.947			
Date:	Fri, 18 Feb 2022	AIC	6887.894			
Time:	15:22:52	BIC	6893.734			
Sample:	0	HQIC	6890.267			
	- 137					
Covariance Type:	opg					
=====						
	coef	std err	z	P> z	[0.025	0.975]

intercept	1.106e+10	1.68e+09	6.567	0.000	7.76e+09	1.44e+10
sigma2	3.888e+20	0.294	1.32e+21	0.000	3.89e+20	3.89e+20
=====						
Ljung-Box (L1) (Q):	0.38	Jarque-Bera (JB):	21800.33			
Prob(Q):	0.54	Prob(JB):	0.00			
Heteroskedasticity (H):	0.18	Skew:	6.88			
Prob(H) (two-sided):	0.00	Kurtosis:	63.25			

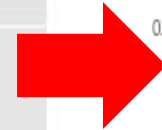
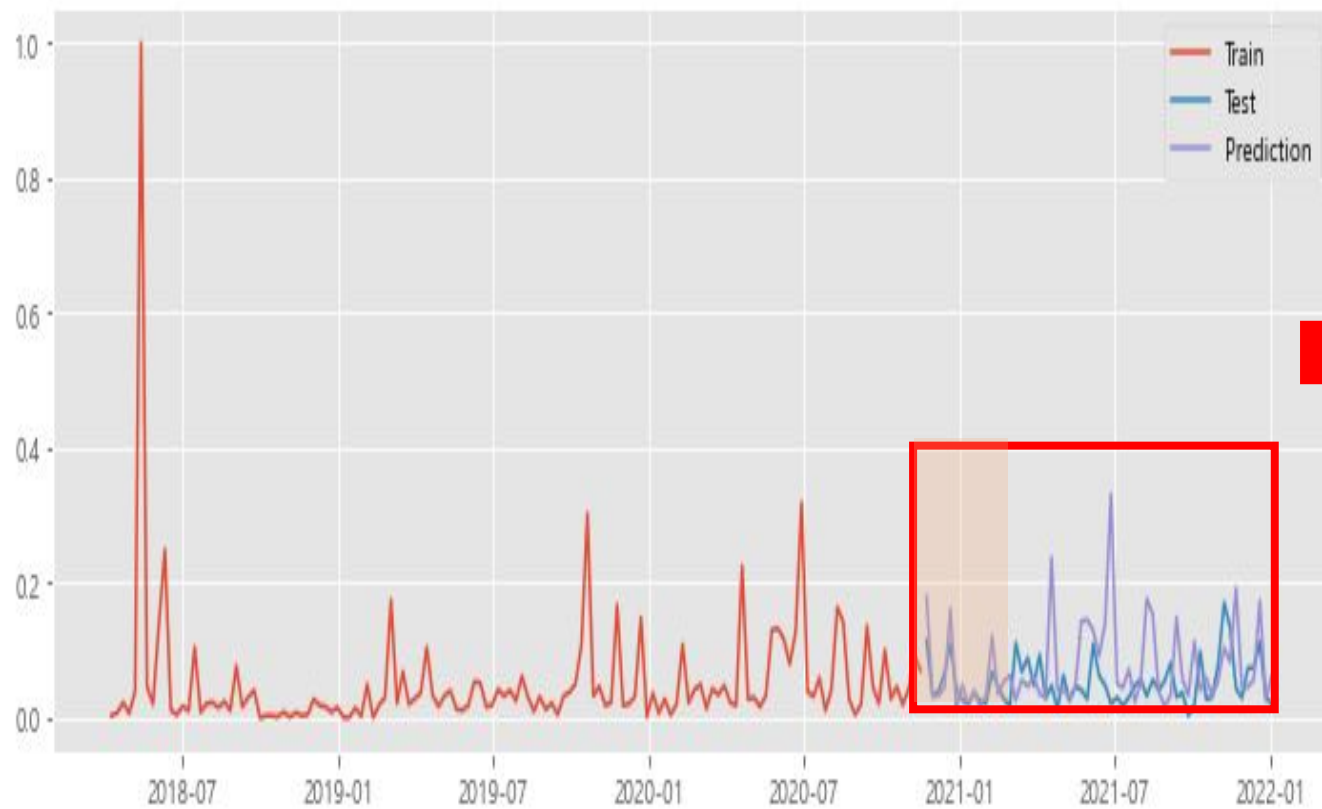
Warnings:

- [1] Covariance matrix calculated using the outer product of gradients (complex-step).
- [2] Covariance matrix is singular or near-singular, with condition number 4.8e+35. Standard errors may be unstable.

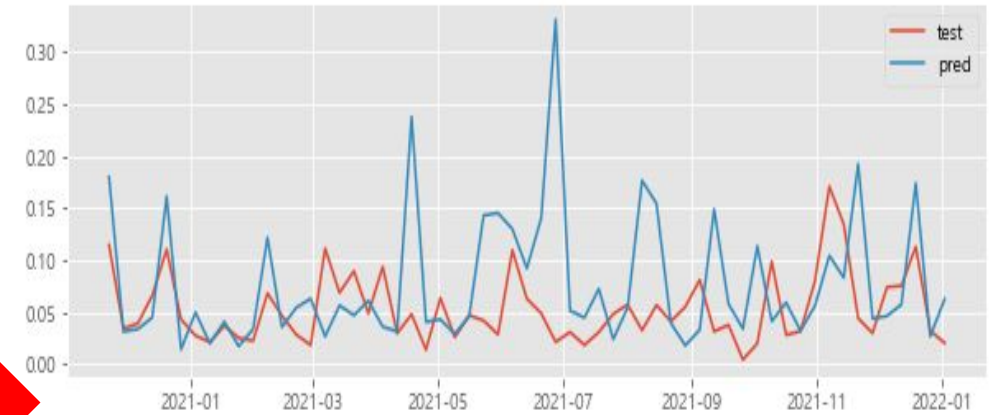


분석 과정_모델링

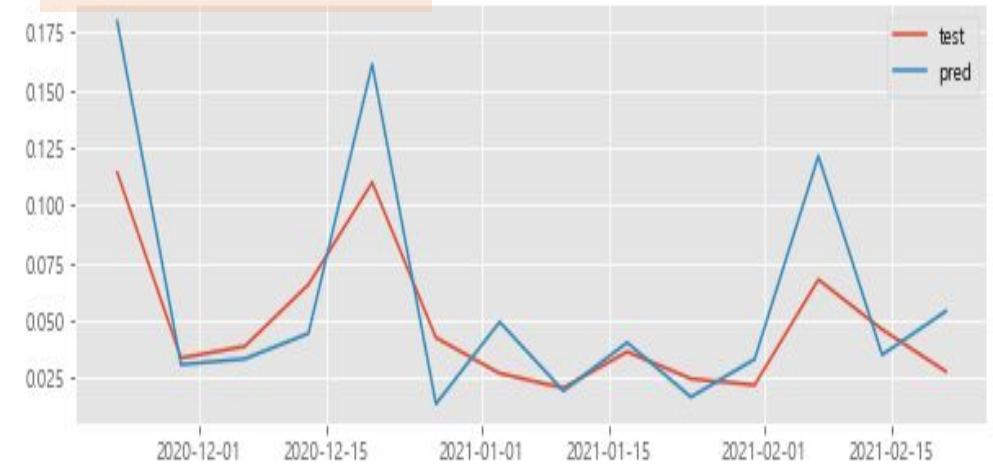
5. 모델test



Test/Predict_전체 RMSE : 0.0709



Test/Predict_축소 RMSE : 0.0300



5. 모델test



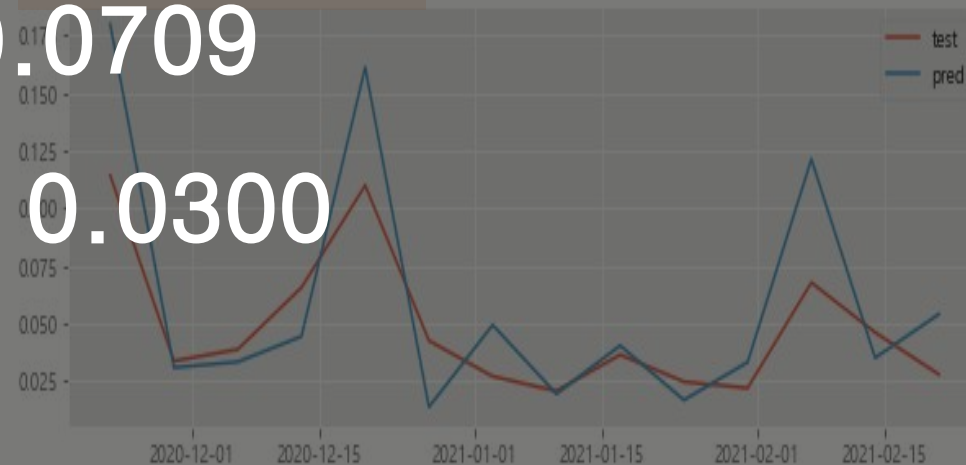
학습 이후 “단기간”의 예측 결과
훌륭한 성능을 나타내었다.

full_RMSE : 0.0709
3week_RMSE : 0.0300

Test/Predict_전체



Test/Predict_축소



Machine Learning Model- 데이터 전처리 및 EDA

Scaling

안정적인 학습 모델을 구축하기 위함

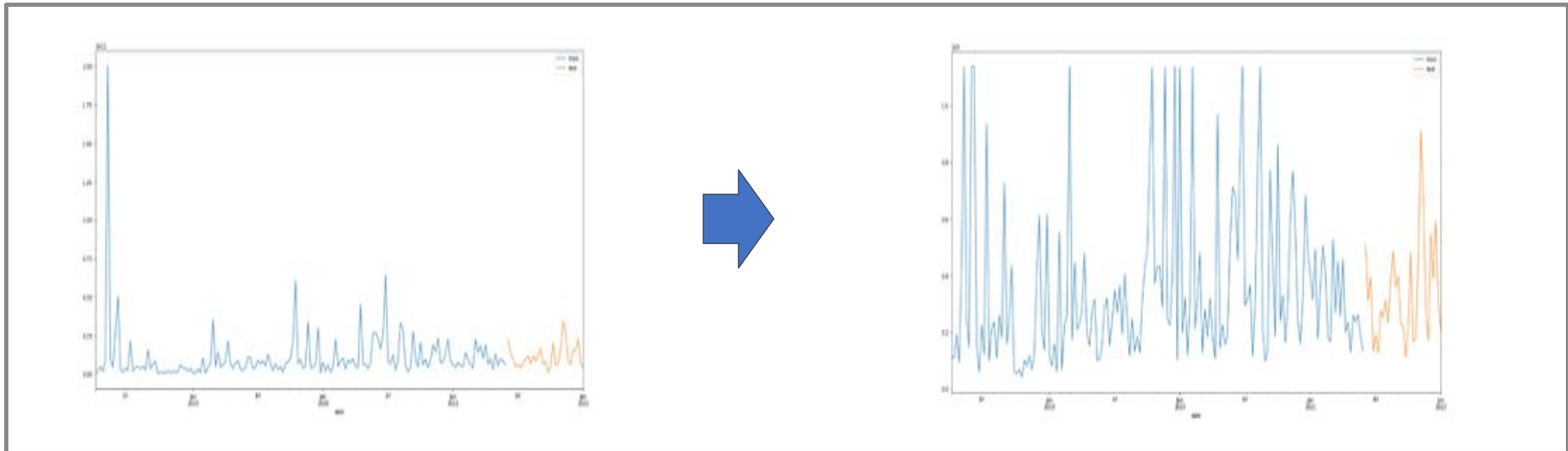
Min Max scaling ➡ Log Transformation robust scaler

: 전체 RMSE 값이 0.03793 -> 0.0274 로 감소

=> 예측값은 역 scaling으로 도출

이상치 제거

0.95 percentile 을 기준으로 이상치 판단

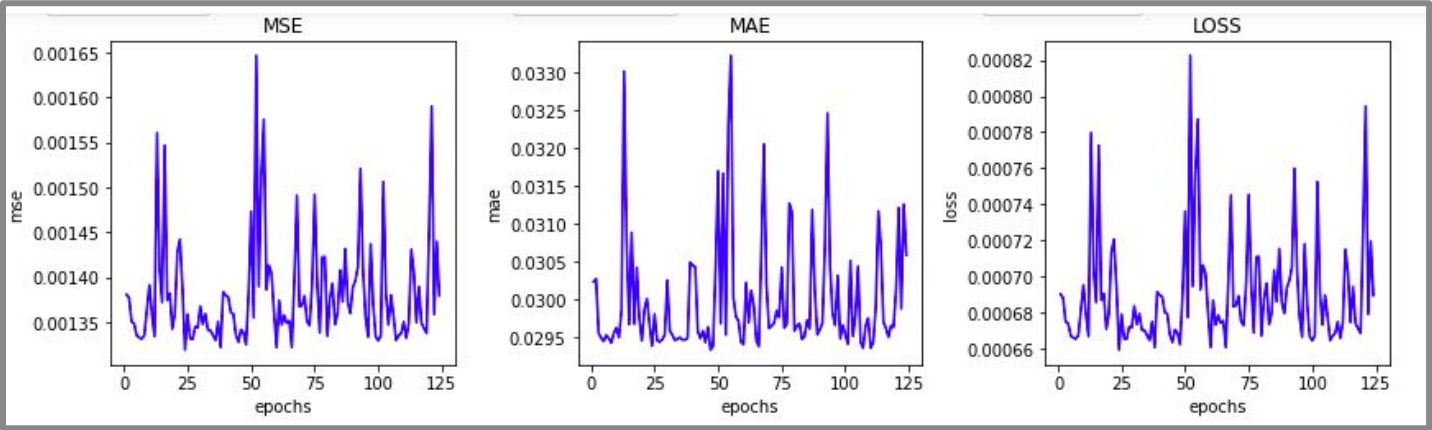


Machine Learning Model - 모델링

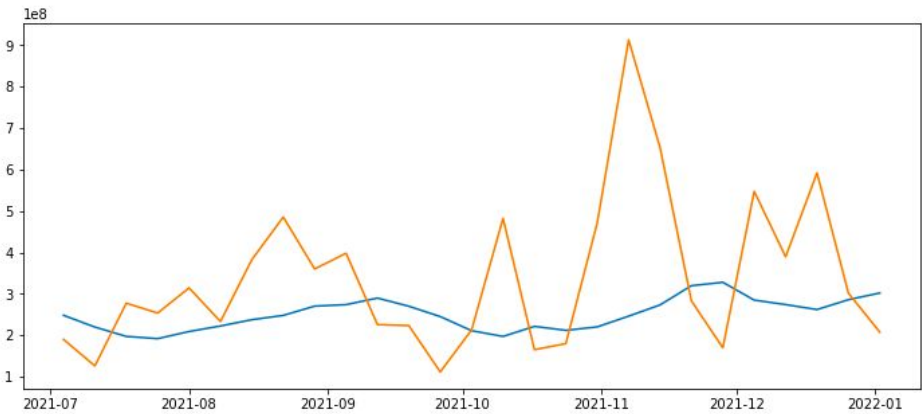
Model: "sequential"

Layer (type)	Output Shape	Param #
lstm (LSTM)	(None, 10)	480
dense (Dense)	(None, 1)	11

Total params: 491
Trainable params: 491
Non-trainable params: 0

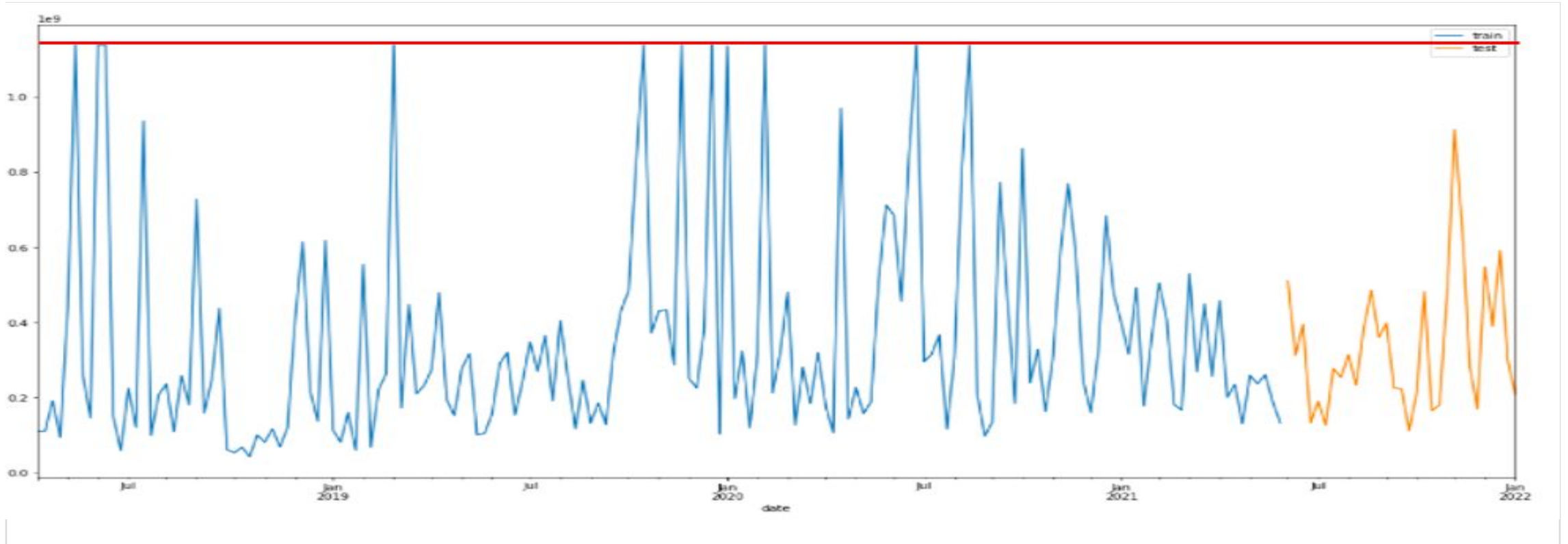


결과) Mean_Squared_log_error = 0.30123



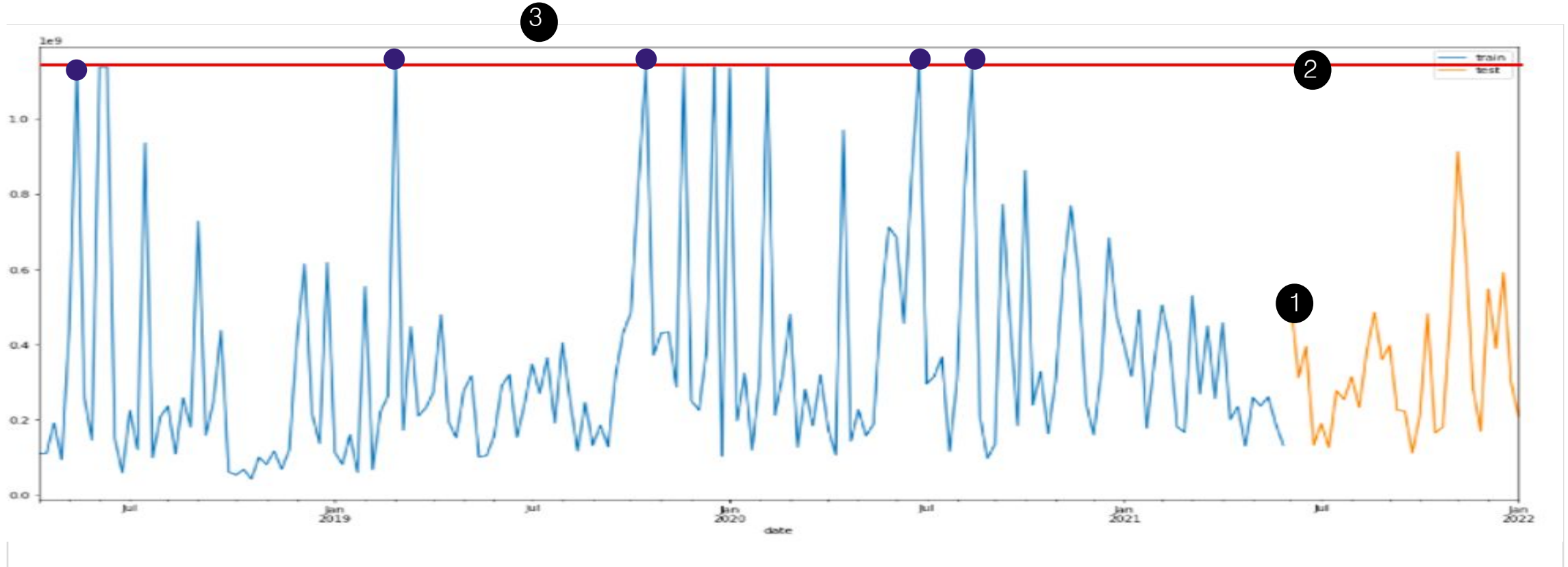
	mean_absolute_error	mean_squared_error	mean_squared_log_error	median_absolute_error
train	234959035.66158	271241698466206336.00000	0.62882	99967925.48148
test	145764996.40363	41568263341647224.00000	0.30123	94438572.36364

결론



이러한 관리도와 예측된 추정 가격 지표는 공공 기관의 기간별 예산을 배정함에 있어서 예산과 관련된 합리적인 의사결정을 내릴 수 있도록 도움을 줄 것으로 기대할 수 있다.

결론



1. 6개월치 추정 금액 예측
2. 과거 데이터를 통해 추정금액의 정상범위를 관리도를 통해 확인
3. UCL을 벗어난다면, 알람을 일으켜 인지함

➡ 이러한 관리도와 예측된 추정 가격 지표는 공공 기관의 기간별 예산을 배정함에 있어서 예산과 관련된 합리적인 의사결정을 내릴 수 있도록 도움을 줄 것으로 기대할 수 있다.

활용 방안 및 결론

- 사용자는 사용 목적에 따라 관리도를 선택하여 사용할 수 있다.
 - T-square 관리도는 '표준화시킨 평균과의 거리'를 중점으로 해서 관리도를 구축하였다면, SVDD는 정상데이터를 포함하는 최소 부피의 입체 도형을 찾는 것을 중점이라는 점에서 차이를 보인다.
 - SVDD 관리도는 hyper parameter(C , γ)를 변경해가며 좀 더 세밀하게 관리도를 조정하여 만들 수 있다. 이는 기관에 따라 작은 변동까지 확인하고 싶다면 hyper parameter 값을 변경하여 좀 더 민감한 관리도를 만들 수 있다. 그러나, 이 관리도에서는 어떤 변수가 원인이 되어서 alarm을 일으키는 지 알 수 없다는 단점이 있다.
-
- 과거데이터 기준이었는데 업데이트된 데이터 넣으면 되는가 ->논문, 이전 사례들 참고 결론부분
 - - 유지 보수; 조달청에서 이걸 사용하면 그만큼의 성능을 보일 수 있는가