

2022 데이터 청년 캠퍼스

거래추적 솔루션의 시계열 성능 데이터 이상 탐지 모델

팀명 : H2O_MCM

팀원 : 박재석, 양창혁, 장지원, 진조아

CONTENTS



01

프로젝트 정의

- 01. 개발 배경
 - 02. OpenMCM
 - 03. 프로젝트 진행 과정
 - 04. 활용 데이터 정의
-

개발 배경



데이터 제공 기업

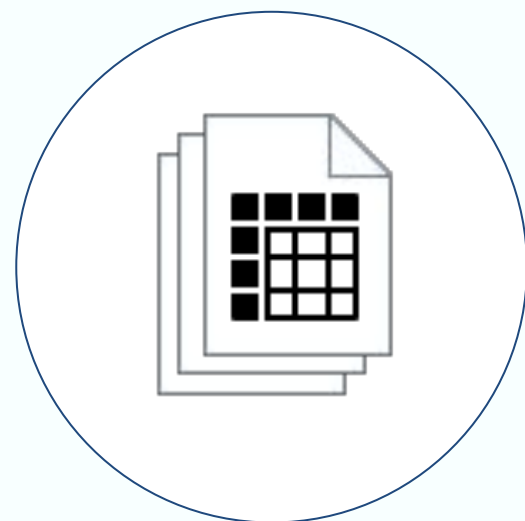
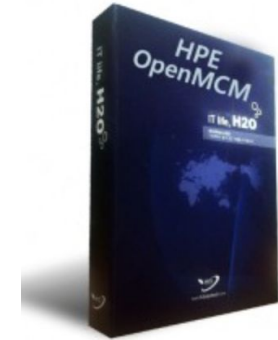
❑ (주) 에이치투오 시스템테크놀로지

: 다양한 이벤트를 대용량으로 처리해야하는 IT 인프라 환경에서 분산 미들웨어 기술을 바탕으로 실시간 거래 추적 모니터링 등의 서비스와 문제 해결 솔루션을 제공하는 기업

Open MCM란?

- ❑ 실시간 End-to-End 트랜잭션 관리 및 동작제어
- ❑ 운영시스템 테스트를 위한 최적 가이드 제공
- ❑ 직관적인 장애 모니터링 제공

E2E/APM 결합 솔루션으로 시장점유 1위



활용 데이터

- ❑ Open MCM 삼성화재 정상 트랜잭션 로그데이터
- ❑ 제공받은 데이터를 기반으로 만든 이상발생 데이터

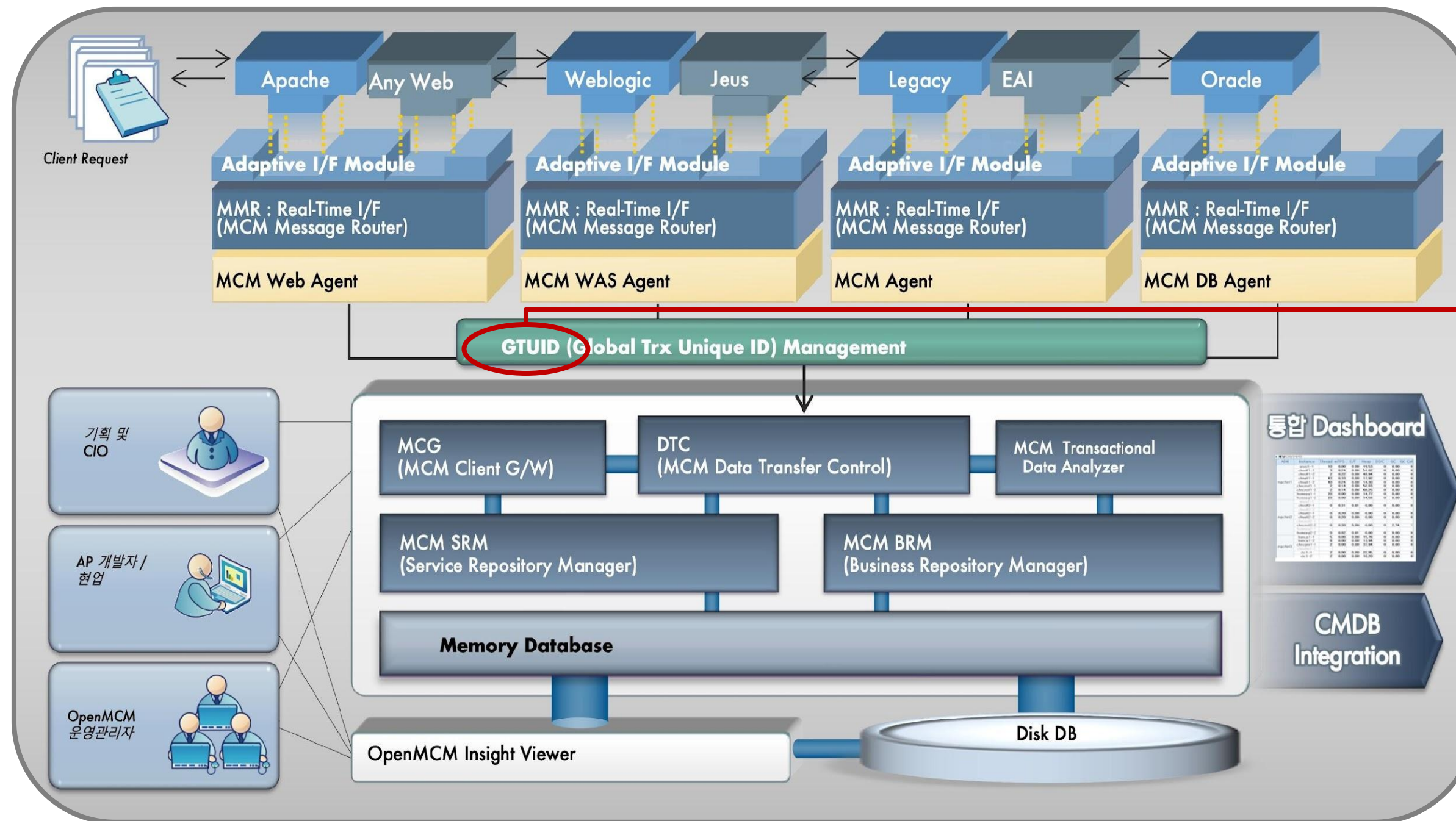
Open MCM 서비스 고객층

- ❑ 보험사, 카드사, 증권사, 은행, 공공기관 등
실시간 트랜잭션 관리가 필요한 기업
- ex. 삼성화재, 신한은행, 한국투자 증권, 국민연금공단...



Open MCM이란?

OpenMCM은 대용량 고속 처리를 위한 OpenMCM Master Server (Core Engine)와 Web, WAS, MCA, EAI 및 DBMS와의 인터페이스 처리를 위한 MCM Agent로 구성된 대용량 초고속 처리 아키텍처이다.



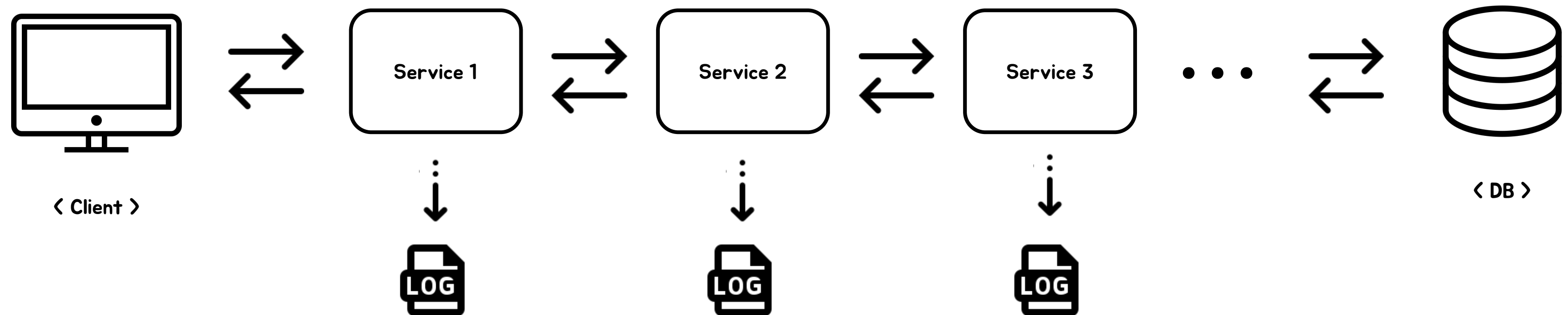
트랜잭션이란?

시스템에서 '더 이상 쪼갤 수 없는 업무 처리의 최소 단위'로, 하나의 트랜잭션이 곧 하나의 거래내역이다.

GTUID 란?

전체 트랜잭션 통합 ID로 트랜잭션 건수 별로 발생한다.

Open MCM 설명



프로젝트 진행과정

데이터 분석

주제 선정

데이터 시각화

모델 선택

모델링

시스템 구축

가상 데이터 생성

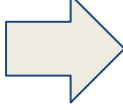
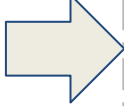
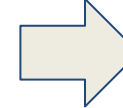
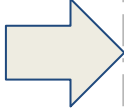
테스트

02

주제 선정

- 01. 주제 선정
- 02. 이상탐지 기준과 방법
- 03. 서버 장애의 원인

주제 선정

상황	기업 요구사항	주제 선정
<p>-기업에서 제공하는 Open MCM 서비스에서 매일 다량의 데이터가 발생함.</p> 	<p>-낭비되는 로그 데이터로 실무에 활용할 수 있는 분야를 찾고 싶음.</p> 	<p><u>-실시간 TPS 추이를 통한 시스템의 이상 분석 및 예측</u></p>
<p>-기존의 서비스는 오류의 탐지를 위한 임계치를 <u>임의로</u> 설정하여서 사용하고 있음</p> 	<p>-자원을 보다 효율적으로 배치하기 위해 임의로 설정한 값이 아닌 근거가 있는 값을 임계치로 설정하고 싶음</p> 	<p><u>-미래의 TPS 추이 예측하고, 예측을 바탕으로 BI 창출</u></p>

서버 장애의 원인

1. 리소스의 부족

2. 시스템 에러

3. 하드웨어 문제



서버 사용자 및 현재 서버의 상태로
예측이 가능

서버 장애의 원인 - 리소스의 부족

1. 리소스의 부족



[뉴스투데이=황수분 기자] 증권사들이 기업공개(IPO) 때마다 고질적인 문제로 지적돼 왔던 '전산장애' 이슈가 최근 다시 수면위로 떠오르고 있다.

여의도에 본사들 둔 한국투자증권이 이번주 서울과 수도권에 내린 기록적인 폭우로 홈트레이딩시스템(HTS)과 모바일트레이딩시스템(MTS)이 15시간 이상 '먹통' 사태를 빚은 것이 HTS, MTS에 대한 투자자들의 불만을 촉발한 것이다.

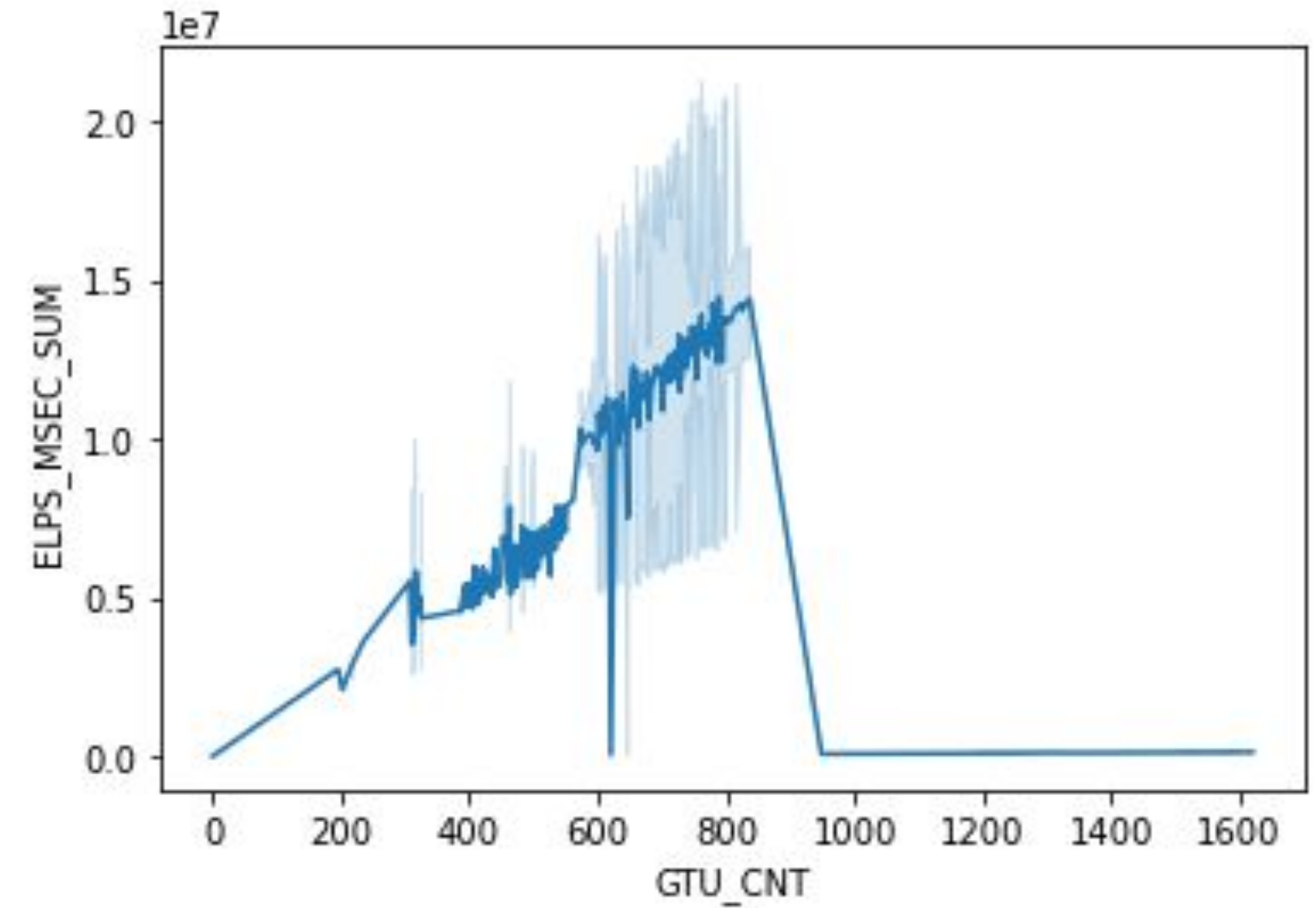
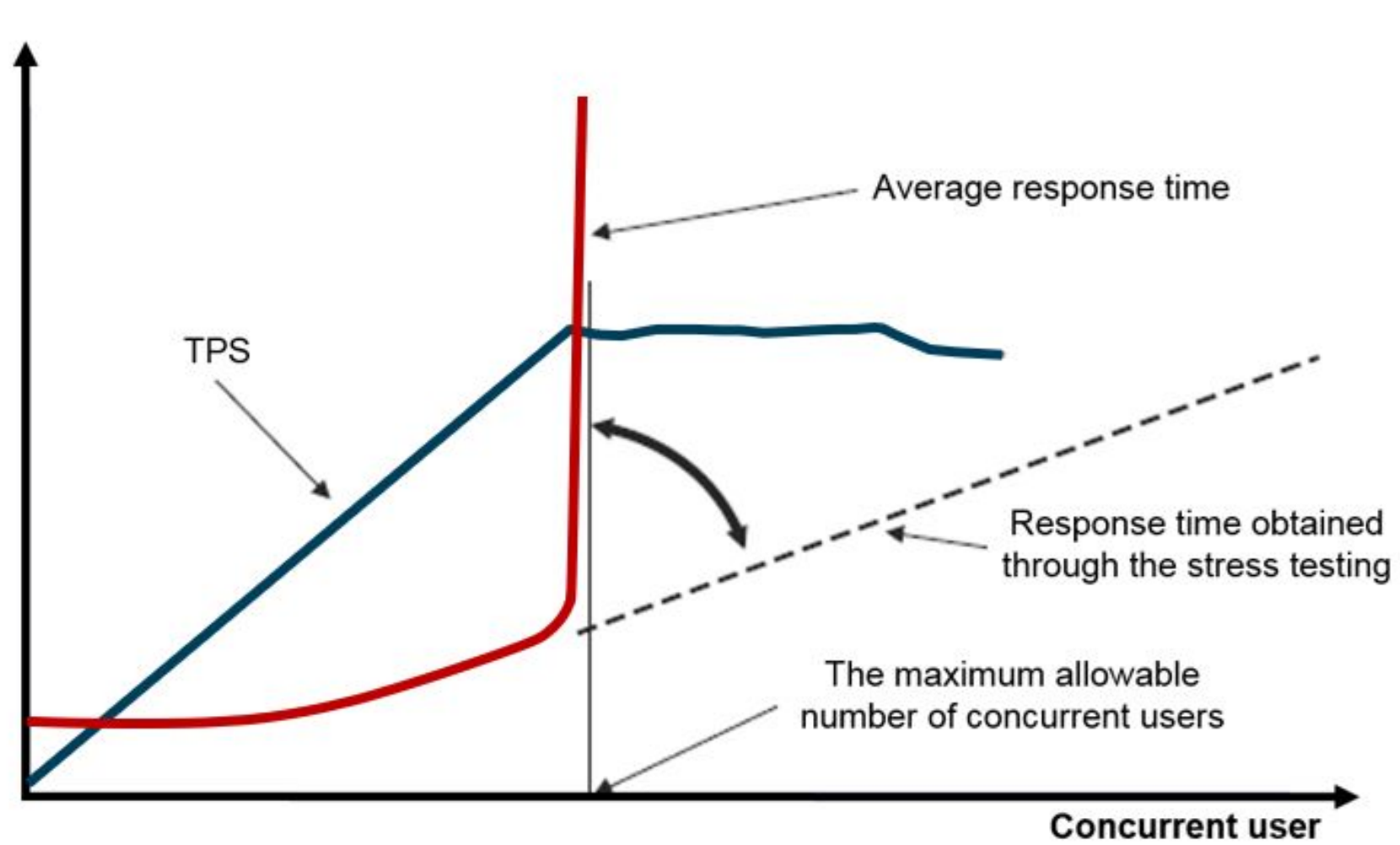
14일 금감원에 따르면 지난해 상반기 증권사 민원건수는 5212건으로 전년 동기 대비 20.5% 늘었고, 이 중 증권사 HTS·MTS 장애와 관련된 '내부통제·전산장애' 유형은 전년 동기 대비 140.1% 증가했다.

증권업계가 사상 최대 실적을 경신하며 새 역사를 썼으나, 급증한 전산장애 민원 해결이라는 과제를 남긴 셈이다.

폭우 이슈를 떠나 공모주 상장일 시스템 접속 급증으로 인한 민원이 크게 늘어 문제 해결이 시급하다는 목소리가 다시 부각됐다.

서버 사용자 및 현재 서버의 상태로 예측이 가능

서버 장애의 원인 - 리소스의 부족

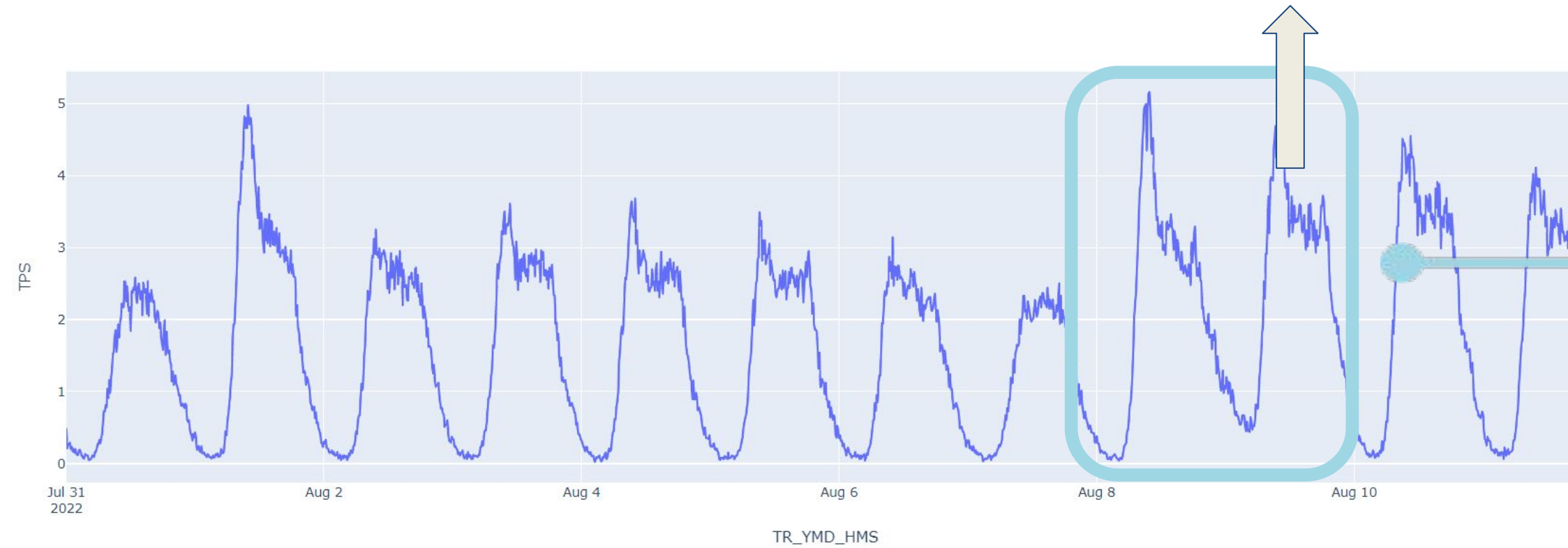


03

데이터 처리방안 및 활용기법

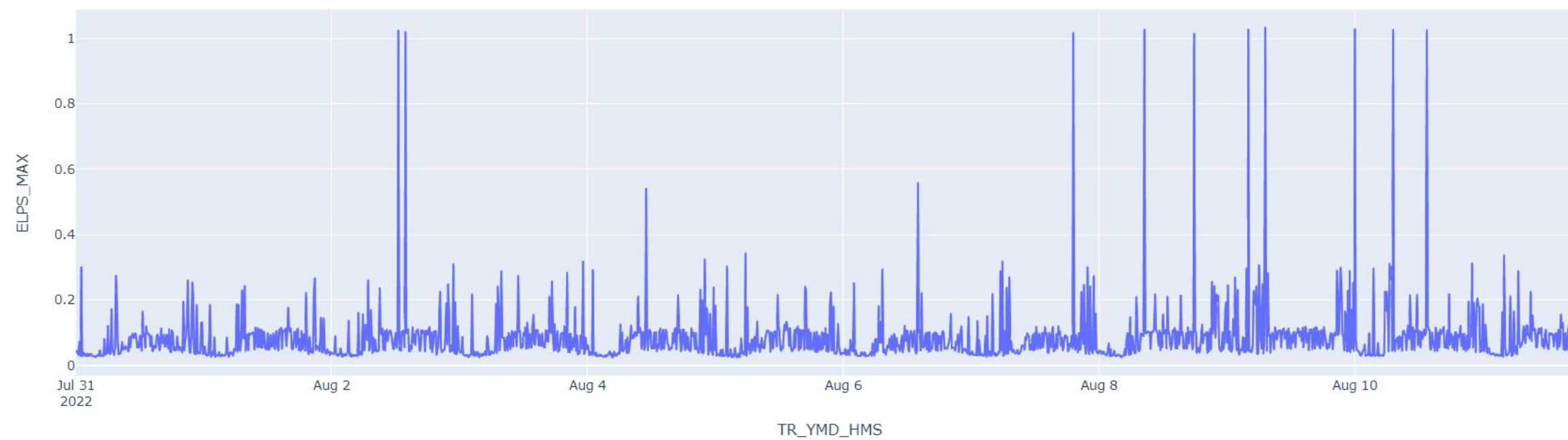
- 01. 시계열 데이터 시각화 (EDA)
- 02. 모델 선정
- 03. 모델링 결과

시계열 데이터 시각화 - EDA



TPS

TPS는 영업시작 시간에 급증하여, 영업종료 시간이 다가올때 급감하는 형태의 그래프를 그린다.

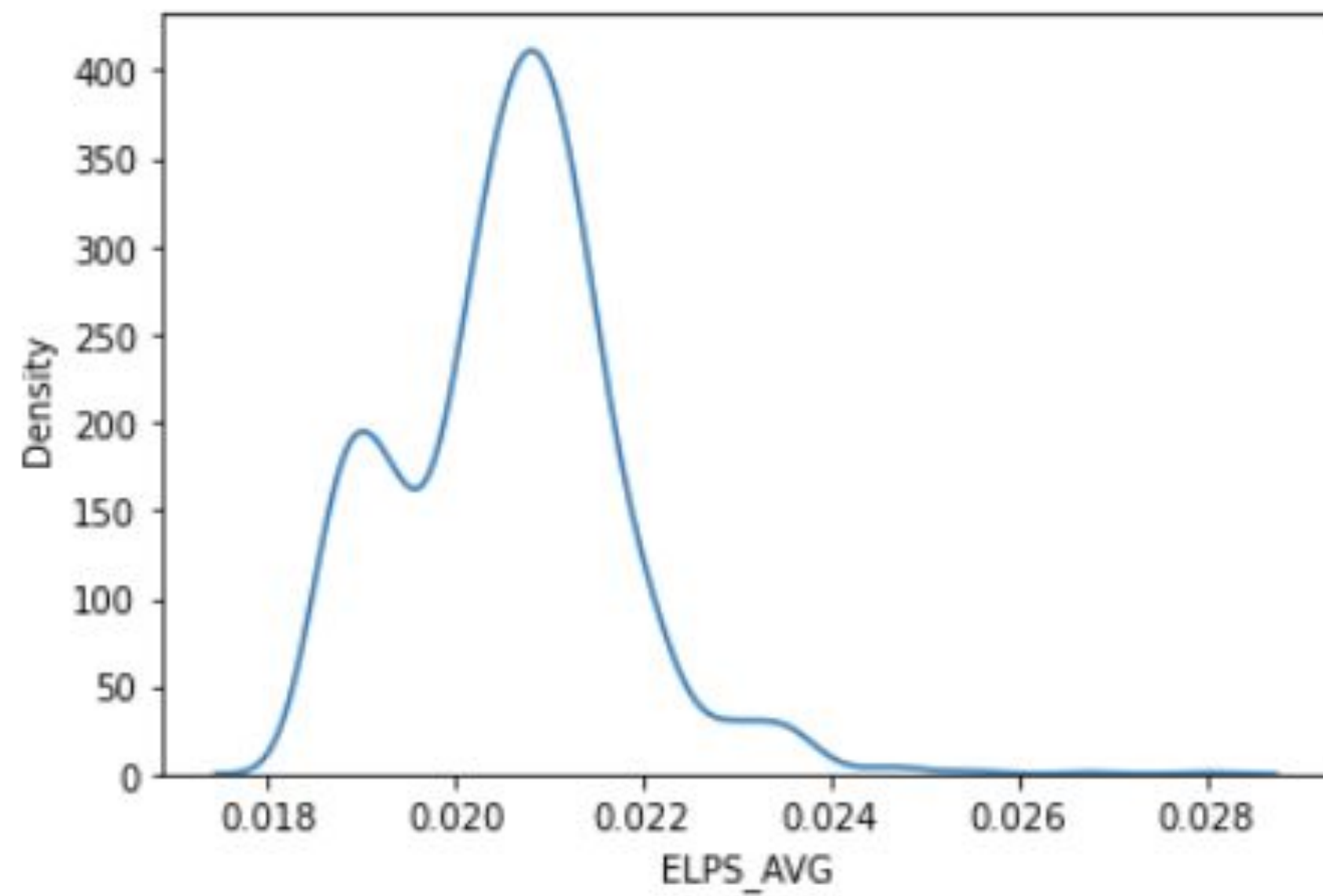


ELPS_MAX

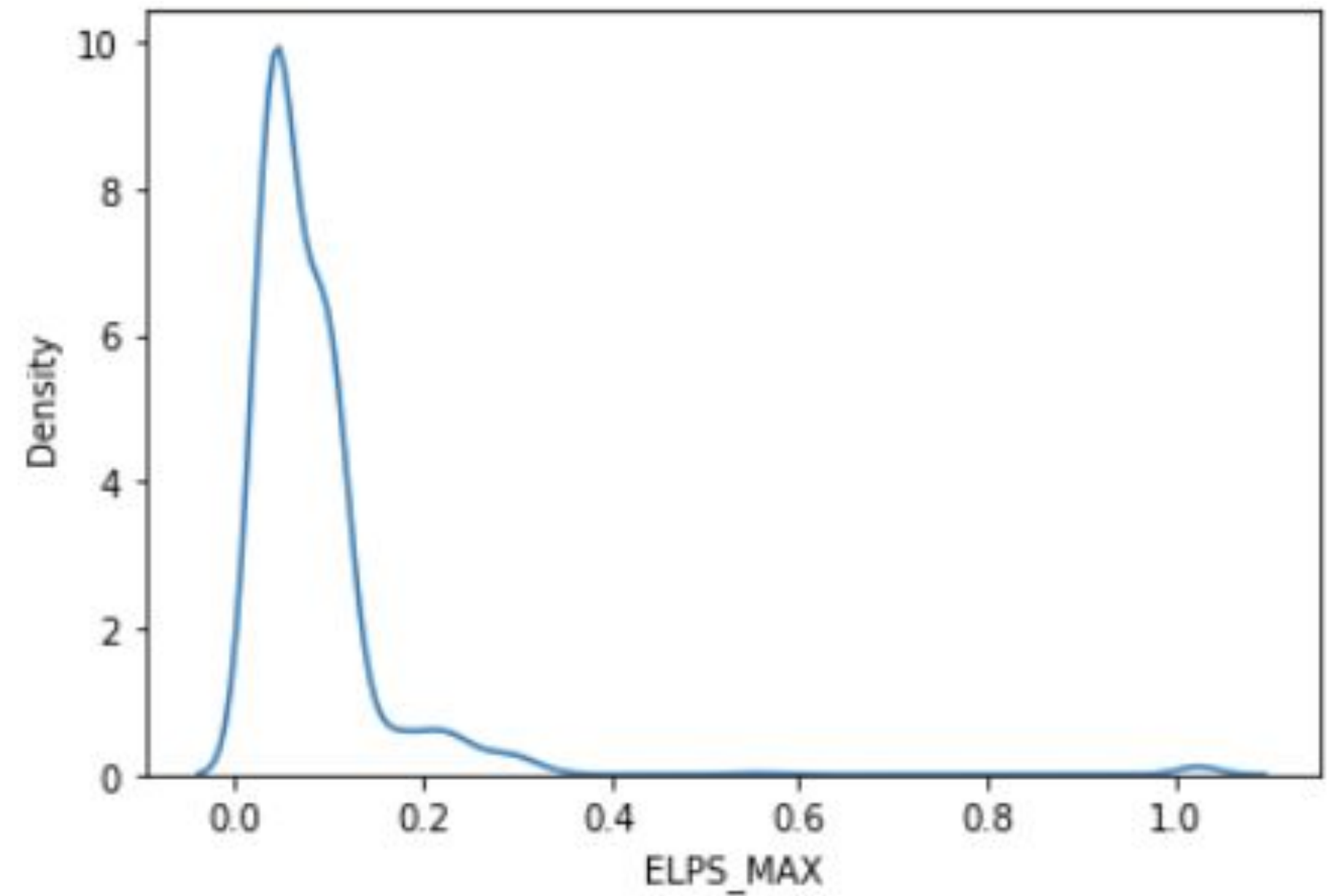
ELPS_MAX는 대체로 TPS커질때 값이 될 확률이 높다

시계열 데이터 시각화 - EDA

평균지연시간 확률 밀도

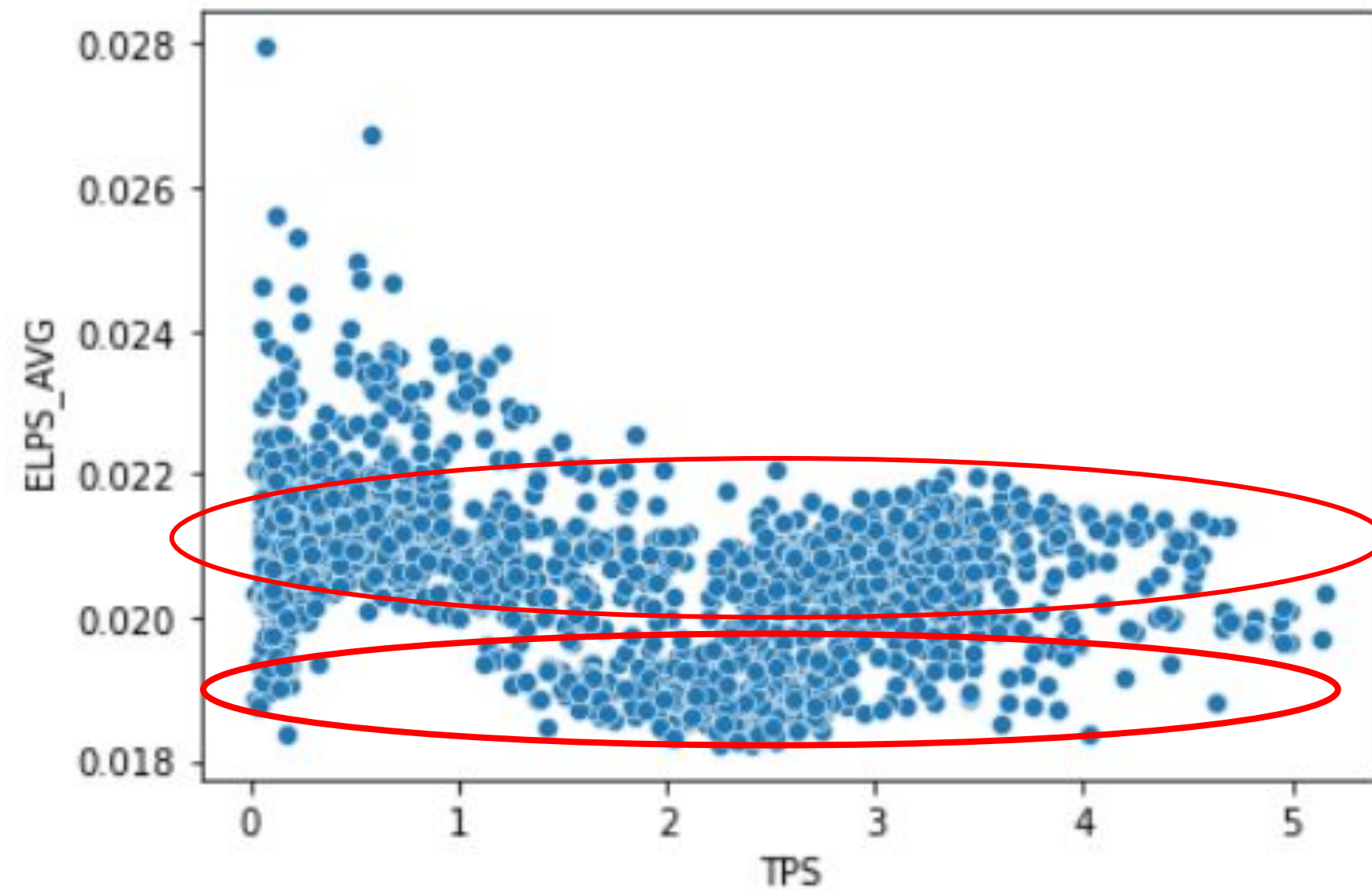


최대지연시간 확률 밀도

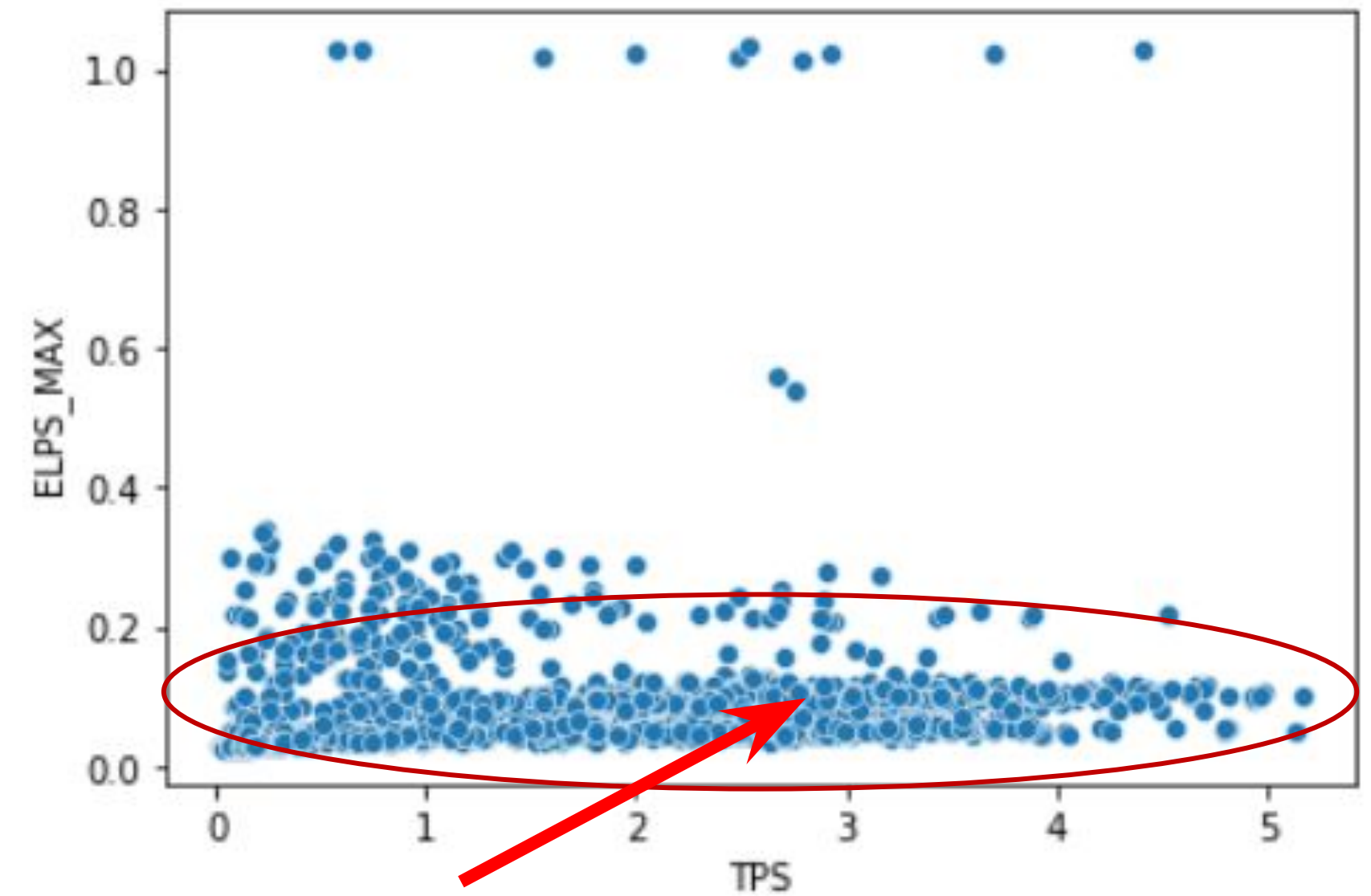


시계열 데이터 시각화 - EDA

TPS/평균 응답시간



TPS/최대 응답시간



99% 구간

모델 선정 시 고려사항

높은 정확도

- ❑ MAE, RMSE, RMSE, R2, RMSLE, MAP, TT 등 모델 성능 지표를 통해 모델의 예측 정확도 판별 및 정확도가 높은 모델 선정

- ❑ 독립변수 X: 시간, 날짜, 이벤트, 요일 등....

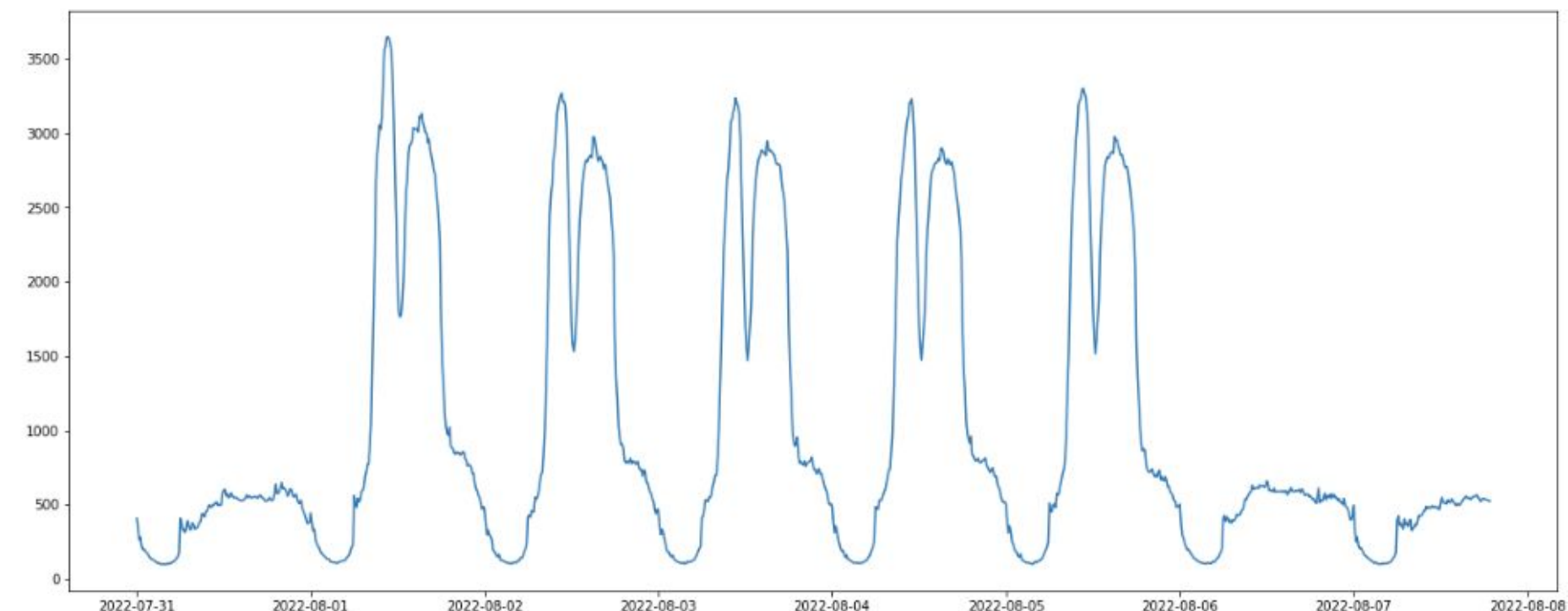
- ❑ 종속변수 Y: TPS

- ❑ 독립변수 X: TPS

- ❑ 종속변수 Y: ELPS Time

시계열 분석

- ❑ 시간대별로 변하는 TPS를 예측해야함
- ❑ 평일보다 거래량이 없는 주말이나 특정 이벤트가 발생하는 날에 TPS 예측하기에 용이해야함



모델 선정하기

TIME	TPS
20220201	264.78
20220202	327.71
20220203	1413.01
20220204	1292.70
20220205	342.94
20220206	321.18
20220207	1352.69
20220208	1250.99
20220209	1238.53
20220210	1355.28
20220211	1213.59
20220212	323.22
20220213	309.19

일	월	화	수	목	금	토
30	31	1 설날	2 음 1.1	3	4 입춘	5
6	7	8	9	10	11	12
13	14 별런타인...	15 정월 대보...	16	17	18	19 우수
20	21	22	23	24	25	26
27	28 2.28 민주...	1	2	3	4	5

설날의 경우. 평일이지만
주말 수준으로
transaction이 작게
일어남을 볼 수 있다

모델 선정하기

	Model	MAE	MSE	RMSE	R2	RMSLE	MAPE	TT (Sec)
xgboost	Extreme Gradient Boosting	0.2426	0.1370	0.3658	0.9032	0.1596	0.4520	0.165
rf	Random Forest Regressor	0.2264	0.1380	0.3636	0.9026	0.1587	0.4855	0.573
gbr	Gradient Boosting Regressor	0.2627	0.1399	0.3707	0.9010	0.1660	0.5617	0.104
lightgbm	Light Gradient Boosting Machine	0.2932	0.1608	0.3988	0.8850	0.1685	0.5808	0.043
dt	Decision Tree Regressor	0.2840	0.2590	0.5017	0.8119	0.2134	0.4075	0.027
ada	AdaBoost Regressor	0.5293	0.3808	0.6156	0.7293	0.2734	1.4449	0.104
knn	K Neighbors Regressor	0.5492	0.5945	0.7614	0.5716	0.3175	1.2872	0.069



MAE가 가장 높은 모델 :

MSE가 가장 높은 모델 :

RMSE가 가장 높은 모델 :

R2가 가장 높은 모델 :

RMSLE가 가장 높은 모델 :

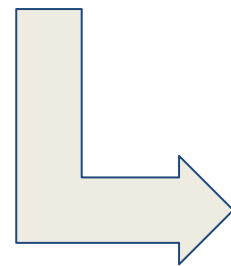
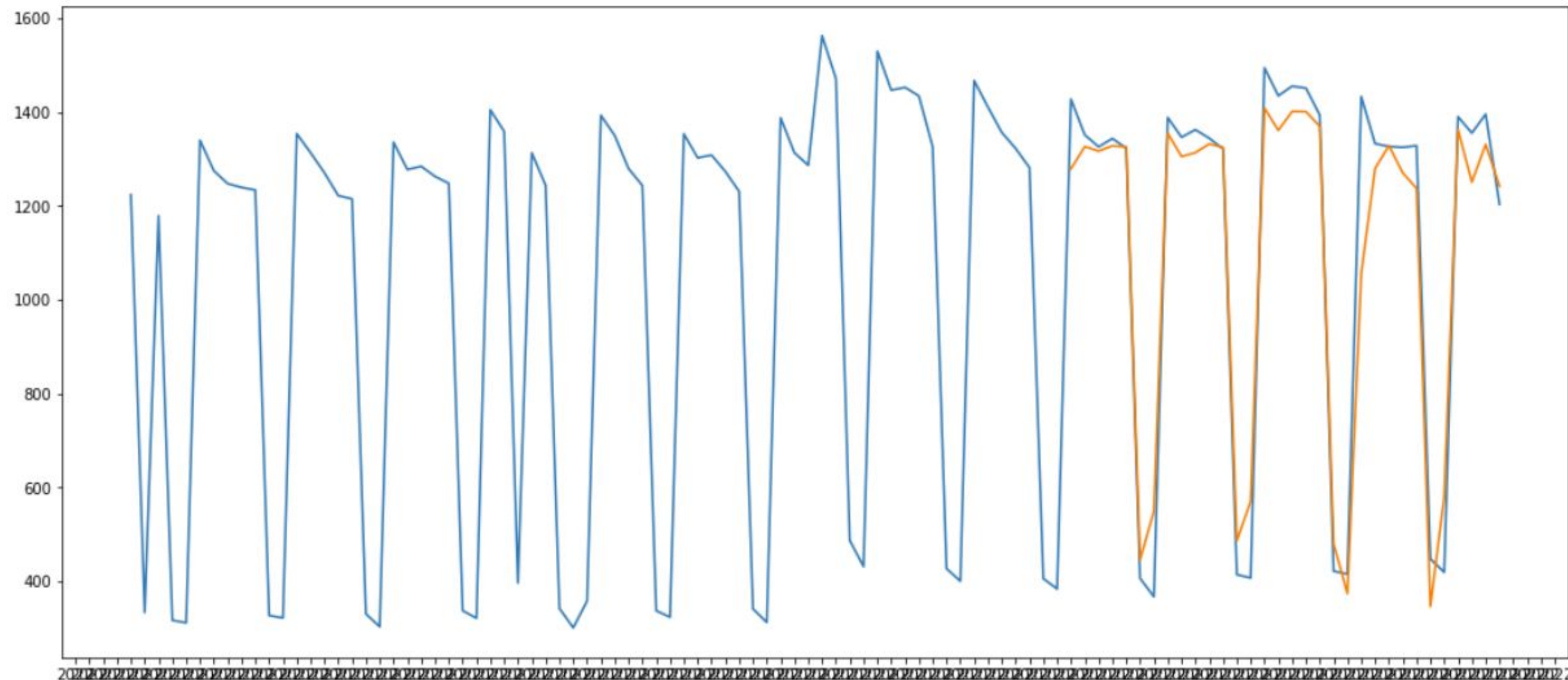
MAPE가 가장 높은 모델 :

TT가 가장 높은 모델:

최종 선정 모델 : xgboost

선정 이유 : xgb는 카테고리컬로 구분할 수 있기때문에, 거래량이 없는 주말이나 특정 이벤트가 발생하는 날에 TPS 예측하기에 더 용이함.

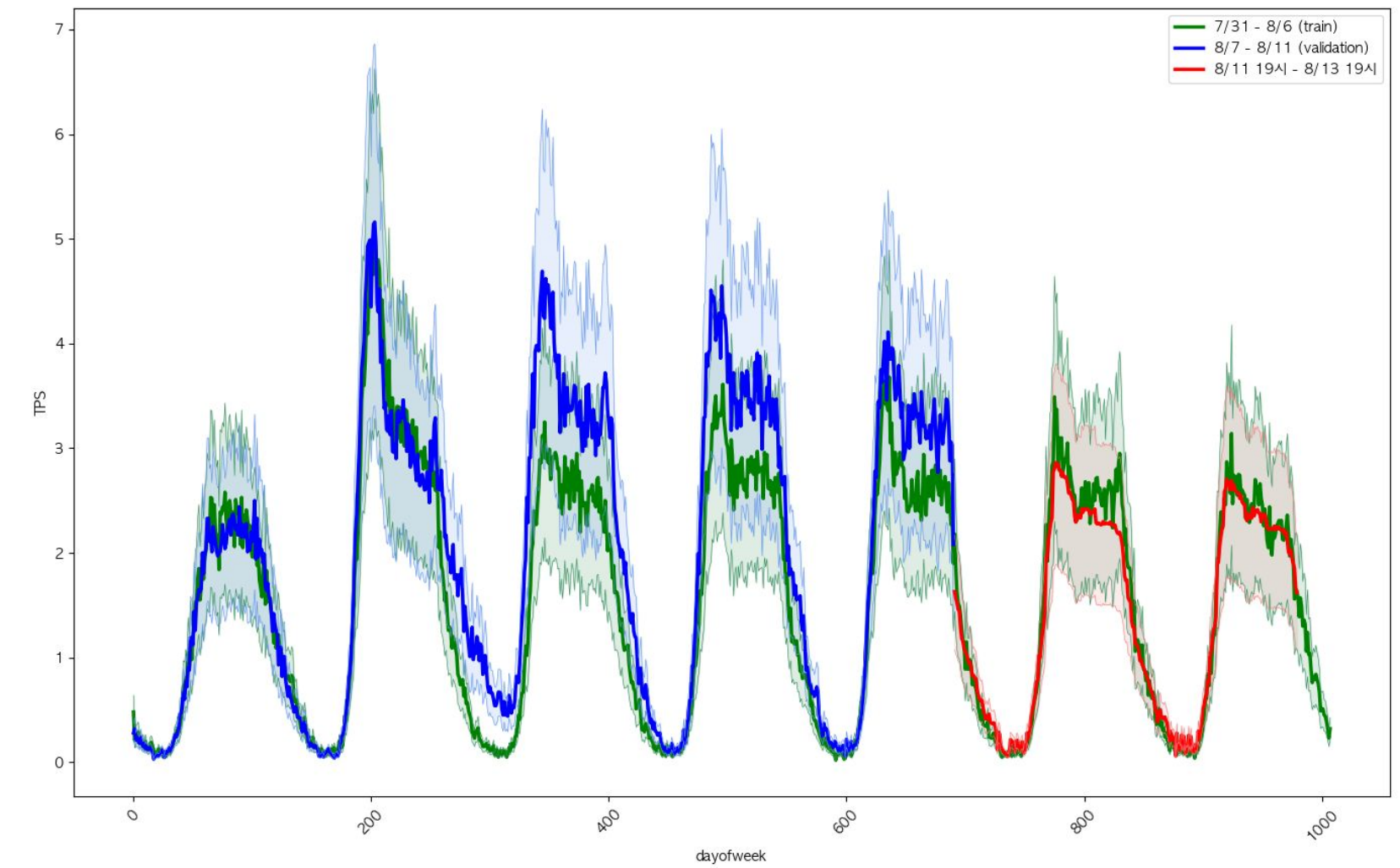
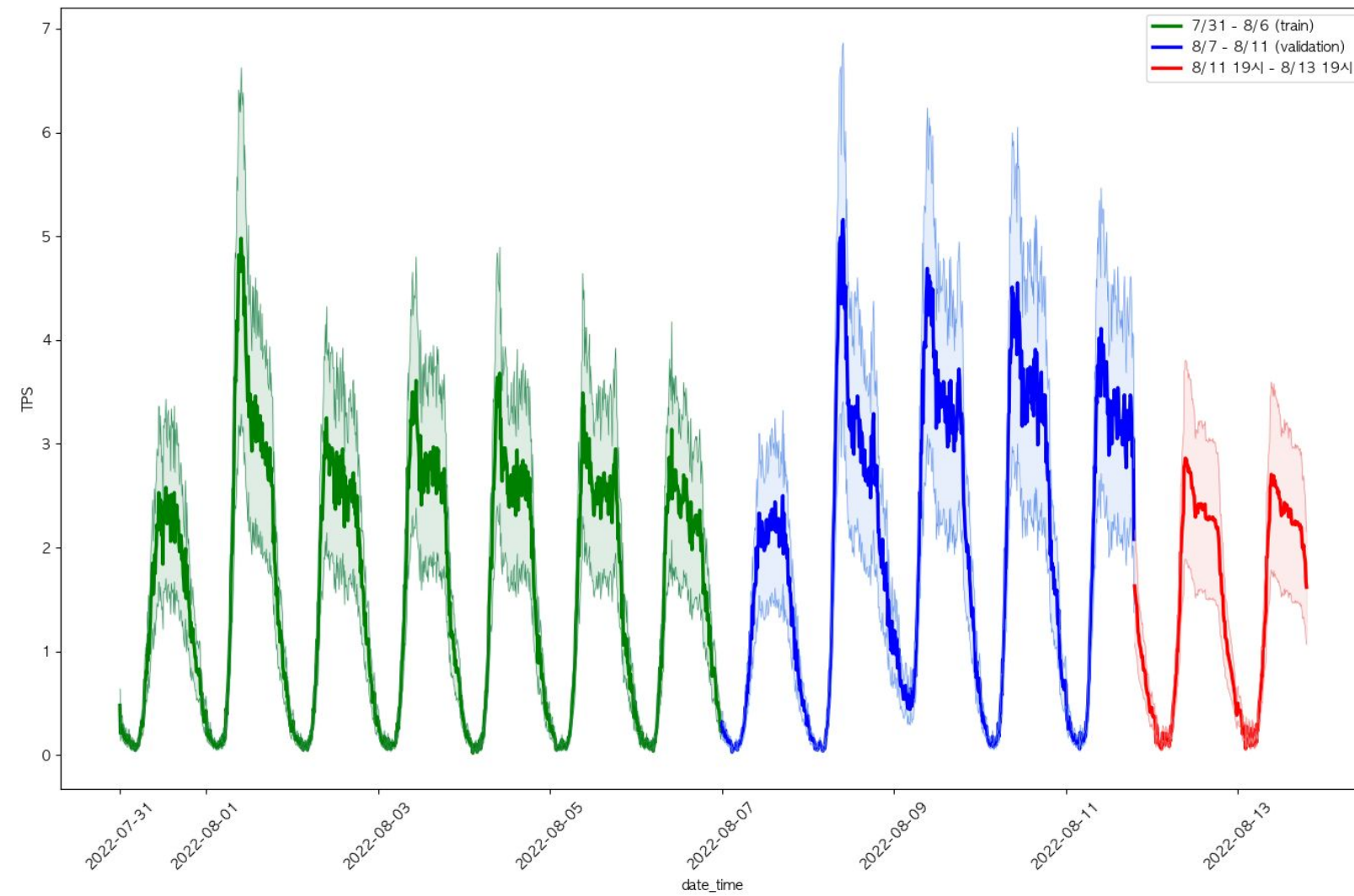
XGBRegressor를 통한 예측 모델링



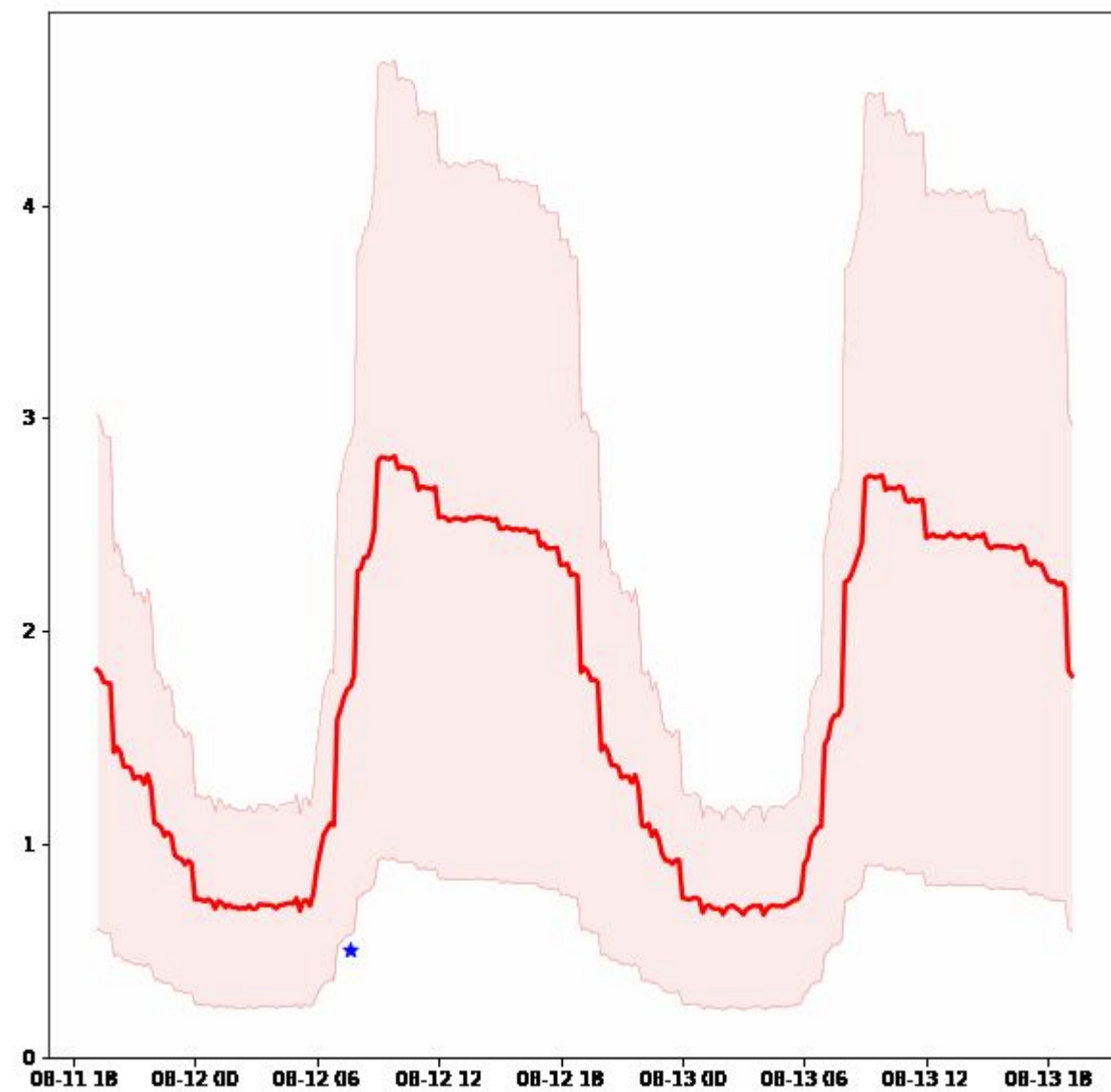
일별예측을 통해 대략적인 추세 파악

해당 일에 기존보다 과하게 높거나 낮게 예측되는경우 임계치를 그에 맞게 조정

XGBRegressor를 통한 예측 모델링



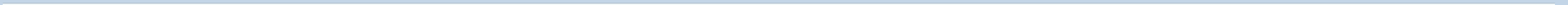
XGBRegressor를 통한 예측 모델링



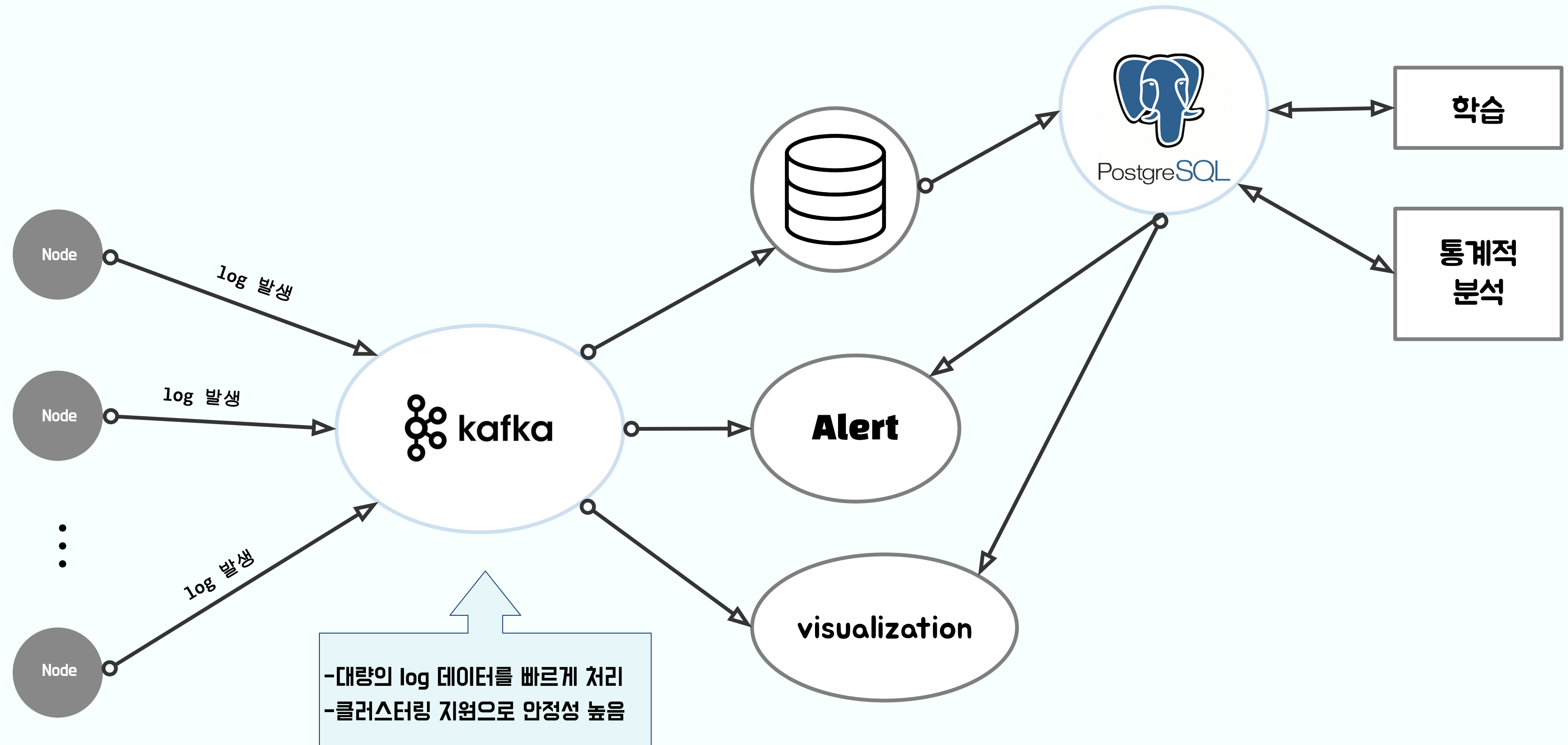
04

시스템 개발 및 테스트

- 01. 시스템 설계도
- 02. 개발 결과 (시연)



프로토타입 설계도



개발 결과

```
producer connected to broker
{'TIME': 20220808000000, 'TPS': 476.42, 'ELPS_AVG': 0.175856, 'ELPS_MIN': 0.0, 'ELPS_MAX': 108.452386}
TIME_LOG
0
0
{'TIME': 20220808001000, 'TPS': 325.64, 'ELPS_AVG': 0.2172, 'ELPS_MIN': 0.0, 'ELPS_MAX': 91.222438}
TIME_LOG
0
1
```

1

select * from mint_log

Data output

Messages

Notifications

	tr_ymd_hms bigint	tps double precision	elps_avg double precision	elps_min double precision	elps_max double precision
1153	202208080000	324.8875501	0.207758783	0	225.227558
1154	202208080010	291.4280641	0.203830933	0	215.8436242
1155	202208080020	288.8101575	0.200224841	0	207.2335396
1156	202208080030	253.3742594	0.196890079	0	198.930418
1157	202208080040	220.9801	0.193714025	0	190.4497865
1158	202208080050	192.7434262	0.190534048	0	181.3760877
1159	202208080100	175.2305025	0.187156466	0	171.439497
1160	202208080110	155.729459	0.183380416	0	160.5726985
1161	202208080120	144.8672771	0.179024119	0	148.9397184
1162	202208080130	130.2097726	0.173950619	0	136.932391

1

select * from tps_predict

Data output

Messages

Notifications

	time bigint	tps double precision	elps_avg double precision	elps_min double precision	elps_max double precision
1	20220807191000	528.7451	0.23950762	6.6424595e-09	204.45634
2	20220807192000	519.94147	0.23950762	6.6424595e-09	204.45634
3	20220807193000	505.08145	0.23520531	6.6424595e-09	210.21727
4	20220807194000	506.30685	0.23520531	6.6424595e-09	210.21727
5	20220807195000	507.09564	0.23520531	6.6424595e-09	210.21727
6	20220807200000	539.37213	0.2408432	6.6424595e-09	200.10254
7	20220807201000	541.39966	0.2438509	6.6424595e-09	200.10254
8	20220807202000	559.5282	0.2438509	6.6424595e-09	193.39317
9	20220807203000	560.849	0.2438509	6.6424595e-09	193.39317
10	20220807204000	564.6225	0.24495049	6.6424595e-09	193.39317
11	20220807205000	564.6636	0.24495049	6.6424595e-09	193.39317
12	20220807210000	561.021	0.25440177	6.6424595e-09	196.22354
13	20220807211000	549.9582	0.25440177	6.6424595e-09	202.9329
14	20220807212000	541.4987	0.25323695	6.6424595e-09	202.9329
15	20220807213000	530.10205	0.25323695	6.6424595e-09	207.28671
16	20220807214000	512.6125	0.25104308	6.6424595e-09	213.04764
17	20220807215000	491.6646	0.251814	6.6424595e-09	216.39252
18	20220807220000	488.58575	0.25361544	6.6424595e-09	216.39252

1 select * from limits

Data output

Messages

Notifications

	tps double precision	elps_avg_upper double precision	elps_avg_lower double precision	elps_max_upper double precision	elps_max_lower double precision
5	40	0.105808606	0.105808606	116.0293316	116.0293316
6	50	0.116062132	0.102248299	128.0336732	128.0336732
7	60	0.116607466	0.072950726	147.5391601	147.5391601
8	70	0.120132699	0.067009996	155.6326435	155.6326435
9	80	0.119037522	0.064296709	150.8414719	150.8414719
10	90	0.126210896	0.063764708	167.9039456	167.9039456
11	100	0.134130928	0.064543278	184.6671599	184.6671599
12	110	0.144070604	0.06585297	222.6765468	222.6765468
13	120	0.168089996	0.078603994	246.8457584	246.8457584

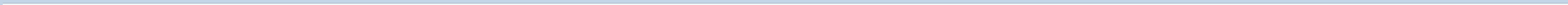
개발 결과

```
!!WARNING!! 20220808000000 : tps is different than predicted.
-----
!!WARNING!! 20220808000000 : ELPS_MAX is different than predicted.
-----
!!WARNING!! TPS and ELSP out of range: Check the system
current status {'TIME': 20220808001000, 'TPS': 325.64, 'ELPS_AVG': 0.2172, 'ELPS_MIN': 0.0, 'ELPS_MAX': 91.222438}
correct
current status {'TIME': 20220808002000, 'TPS': 264.84, 'ELPS_AVG': 0.220549, 'ELPS_MIN': 0.0, 'ELPS_MAX': 282.243435}
correct
current status {'TIME': 20220808003000, 'TPS': 262.51, 'ELPS_AVG': 0.185086, 'ELPS_MIN': 0.0, 'ELPS_MAX': 160.733486}
correct
current status {'TIME': 20220808004000, 'TPS': 227.51, 'ELPS_AVG': 0.215876, 'ELPS_MIN': 0.0, 'ELPS_MAX': 170.647935}
correct
current status {'TIME': 20220808005000, 'TPS': 200.41, 'ELPS_AVG': 0.204576, 'ELPS_MIN': 0.0, 'ELPS_MAX': 602.096439}
correct
current status {'TIME': 20220808010000, 'TPS': 193.35, 'ELPS_AVG': 0.20723, 'ELPS_MIN': 0.0, 'ELPS_MAX': 290.563067}
correct
current status {'TIME': 20220808011000, 'TPS': 180.08, 'ELPS_AVG': 0.190289, 'ELPS_MIN': 0.0, 'ELPS_MAX': 163.352333}
correct
```

05

서비스 활용 방안 및 기대 효과

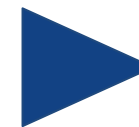
- 01. 기대 효과
- 02. 서비스 활용방안



기대 효과

기존

E-AMS 기능 \ AP 생명주기		개발/테스트
Controlling	실시간 장애 예측	X
	실시간 장애 예방	X
	실시간 장애 인지	O
	실시간 원인 추적	O



변경 후

E-AMS 기능 \ AP 생명주기		개발/테스트
Controlling	실시간 장애 예측	O
	실시간 장애 예방	X
	실시간 장애 인지	O
	실시간 원인 추적	O

서비스 활용 방안

OpenMCM 사용자

서버 증축 / 클라우드 이전에 필요한 자원량을 가시적이고 객관적인 지표로 확인 가능 -> 예산 편성에 유리.
기존에 임의로 정한 임계치를 모델링 결과를 바탕으로 더 정확하게 설정 가능

회사

고객 유치에 유리. 기존에 활용도가 적던 OpenMCM 로그데이터를 이용하여 실시간 오류 예측이라는 새로운 기능을 제공

감사합니다
