

시간대별 수요 예측을 통한 인력자원관리

국민대학교 권유진
국민대학교 이경욱

WE DESIGN LOGISTICS



목차

I. 과제 개요

- 배경
- 목적

II. 과제 수행

- EDA
- BASELINE 작성
- ML
- SARIMA
- PROPHET
- DNN

III. 활용방안 및 기대효과

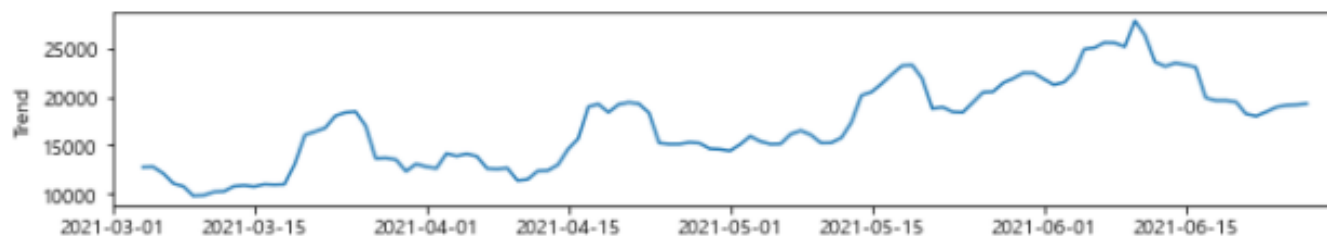
IV. 소스코드 시연 동영상

V. Q&A

— 과제 개요

과제 개요 - 배경

주문 수 증가하는 추세



익일배송 서비스 확대

기존에는 주문한 다음날 배송을 받기 위해 오후 3시 정도까지는 주문을 해야 했는데, CJ대한통운 e-풀 필먼트에서는 밤 12시에 주문해도 다음날 받아볼 수 있다.

업체들은 안정적인 배송과 더 길어진 주문 여유시간의 효과를 톡톡히 보고 있다. CJ대한통운이 고객사 대상으로 실시한 설문조사에 따르면, 업체 중 77%가 서비스 개시 이후 25% 이상 판매량이 늘었으며 일부 업체는 100% 이상 증가한 경우도 있었다.

<http://www.100ssd.co.kr/news/articleView.html?idxno=76331>

- F/C 시설 확충
- 새벽배송/당일배송 서비스 예정
- 고객사 증가



시장 점유율 증가 예상



정확한 수요예측에 기반한
인력 자원관리 모델 필요

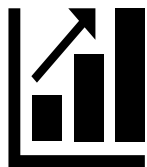


- 효율적인 인력 자원 관리 필요
- 비용 절감 VS 안전 문제 + 고객 신뢰

과제 개요 - 목적



확장될 시간대별 세분화 서비스 모델에 대한 선행 프로젝트



시간대별 수요예측모델수립

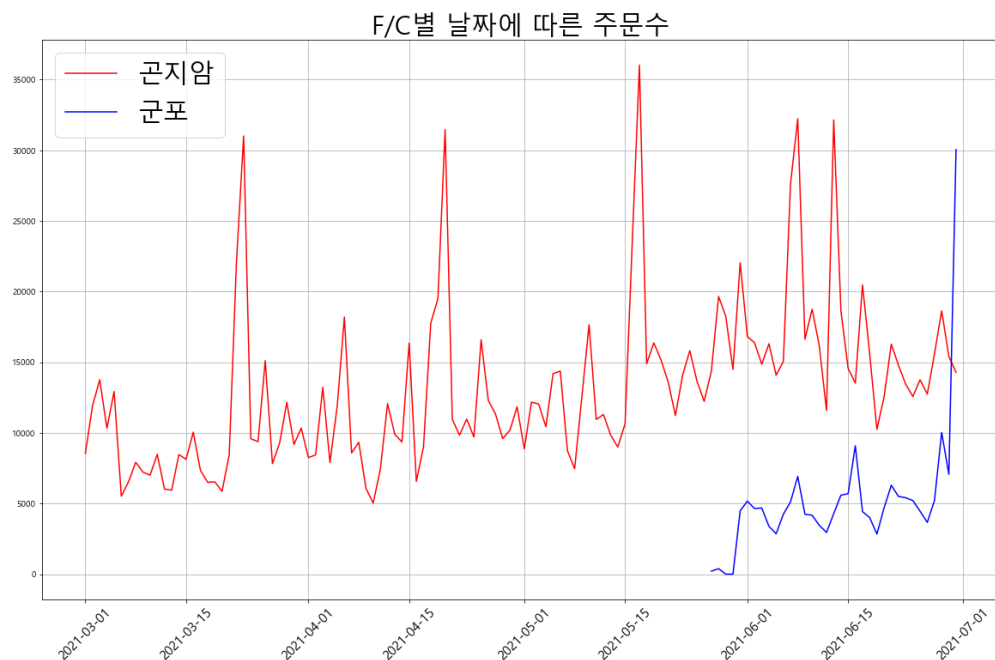


예측수요기반 작업 사이클 별 필요인력 계산모델 수립

— 과제 수행

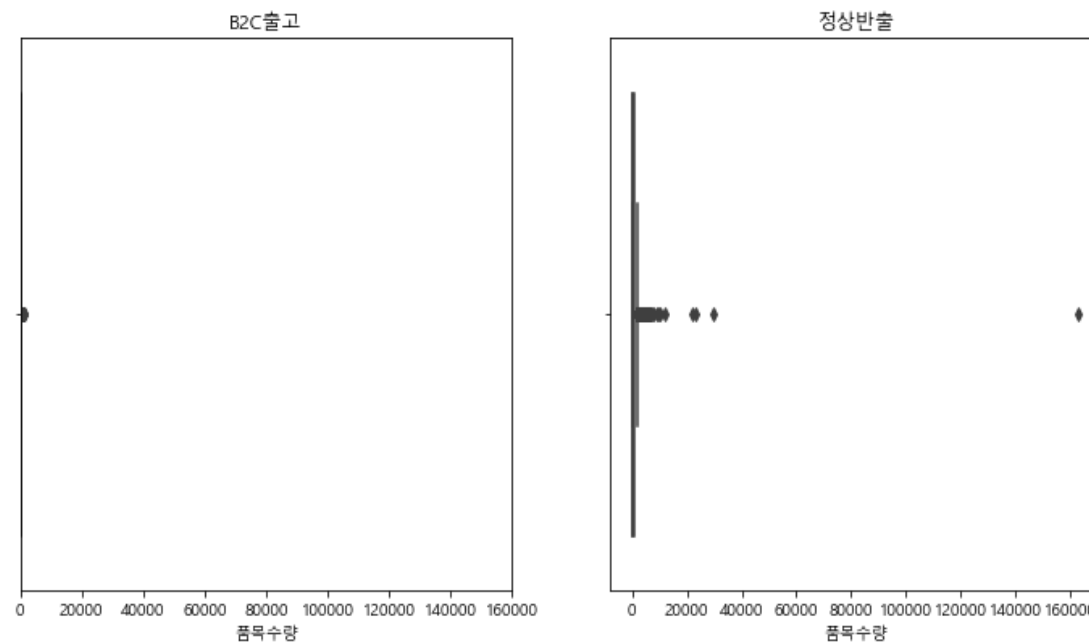
과제 수행 - EDA

1. F/C별 날짜에 따른 주문수 시각화



→ 군포 F/C의 데이터는 사이즈가 작음

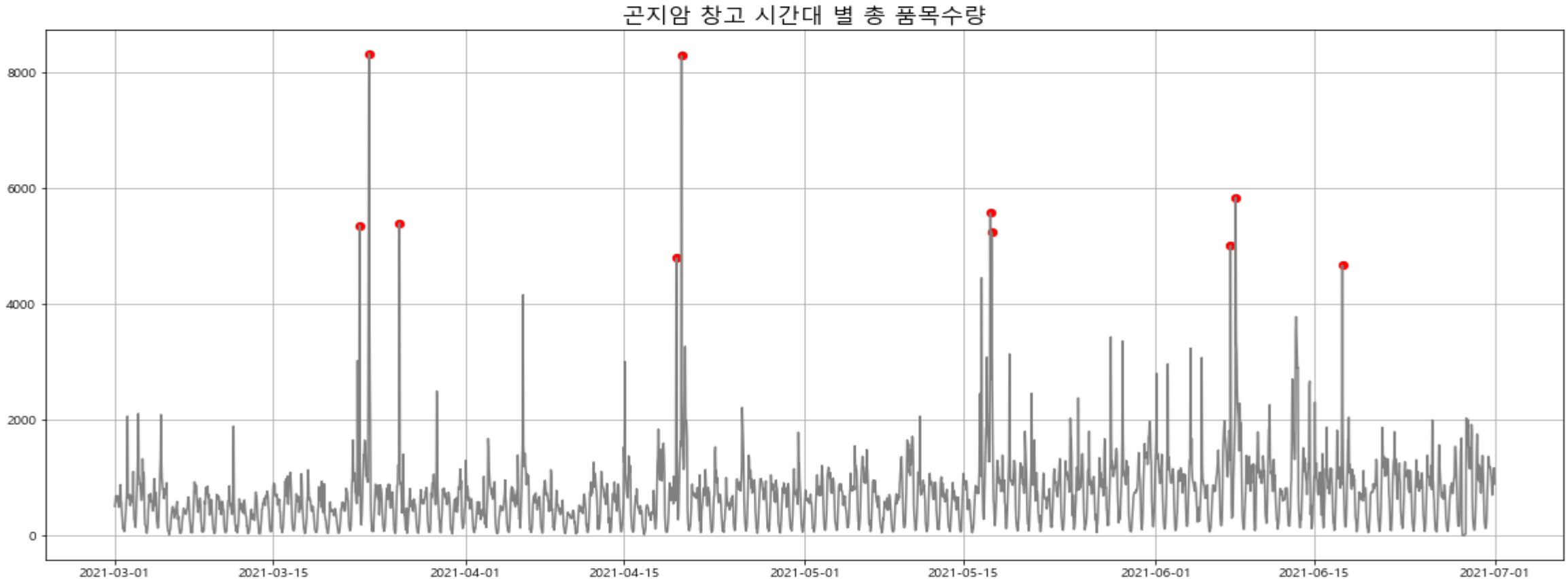
2. B2C출고와 정상반출된 품목수량 분포



→ 정상반출인 경우 이상치가 많고 B2C출고에 대한 수요예측에 적합하지 않은 데이터임

과제 수행 - EDA

■ 곤지암 F/C의 주문날짜-주문시간대별 B2C출고 품목수량

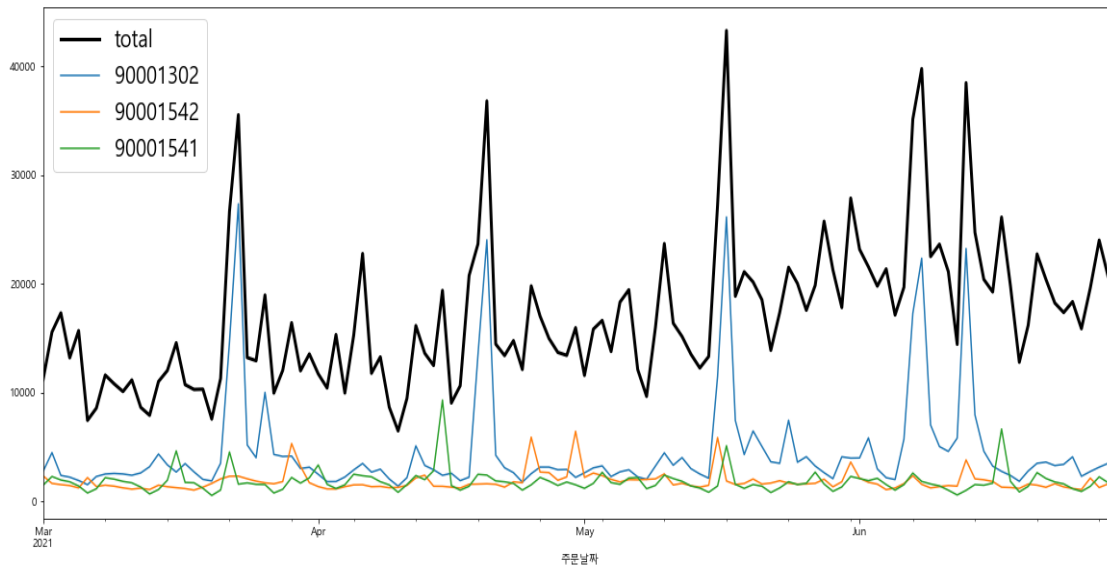


→ 수요가 급격히 상승하는 부분(이상치)에 대한 탐색이 필요해 보임

과제 수행 - EDA

■ 이상치 탐색

1. 수요 상위 3개의 고객사코드의 품목수량과 전체 품목수량 시간대별 추세



- 이상치에 같은 움직임을 보이는 추세
- 해당 고객사의 특정 이벤트가 영향을 준 것이라 추정

2. 상위 3개 고객사코드의 입력자ID를 통해 고객사 추정

```
90001302  
['iNEXBatch' 'sibizi22' 'sibizi28' 'sibizi56' 'BATCH' 'sibizi47'  
 'sibizi49' 'sibizi53' 'sibizi27' 'AK04']  
90001541  
['iNEXBatch' 'NESTLE11' 'NESTLE09' 'BATCH']  
90001542  
['iNEXBatch' 'nutrione06' 'nutrione01' 'BATCH' 'nutrione10']
```

- 90001302는 입력자ID를 통해 추정불가 품목수량합계 1위로 LG생활건강이라 추정
- 90001541과 90001542은 각각 NESTLE와 Nutrione으로 추정



<https://brand.naver.com/lgcaremall?NaPm=ct%3Dkvnj5ci5%7Cc%3Dcheckout%7Ctr%3Dds%7Ctr%3D%7Chk%3D2c63d1b3e27ac52f939efe736eb961bc8b25be4e>



<https://brand.naver.com/nestlestore?NaPm=ct%3Dkvnj5ci5%7Cc%3Dshop%7Ctr%3Dds%7Ctr%3D%7Chk%3D3774c0d1628c15f9f1c3d33f3e74a62daa5b0e12%7Ctr%3Dundefined>



<https://brand.naver.com/nutrione?NaPm=ct%3Dkvnj5ci5%7Cc%3Dcheckout%7Ctr%3Dds%7Ctr%3D%7Chk%3Df76b83b805e90cbcae50d4c447798db577776fe3>

- 네이버 브랜드스토어에서 추정된 고객사 확인

과제 수행 - EDA

■ 이상치 탐색

	Nestle	Nutrione	LG생활건강
3월	22일	-	22 ~ 28일
4월	15일	25, 30일	20일
5월	17일	16, 31일	17일
6월	17일	13, 27일	7 ~ 14일

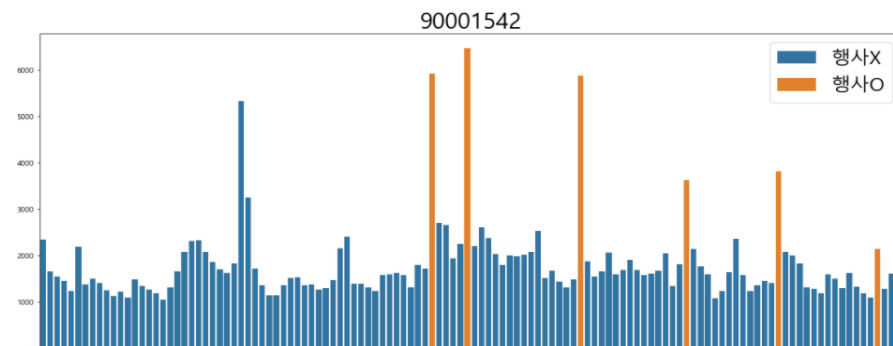
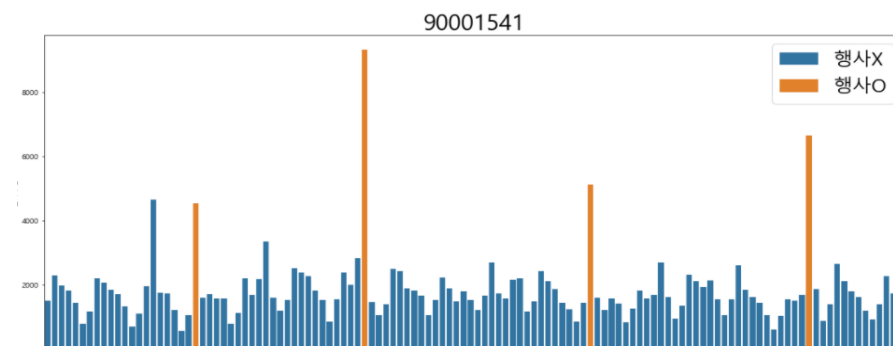
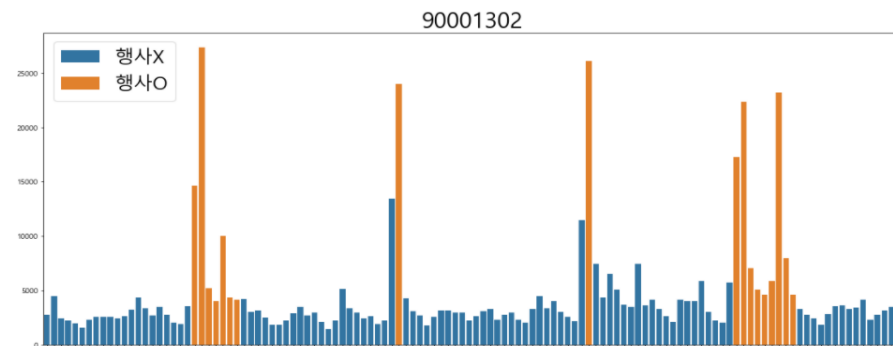
- 각 고객사 공식몰을 통해 확인한 이벤트 날짜
- LG생활건강 : LG생건X네이버 레드워크, 브랜드데이
- 네슬레 : 브랜드데이, 구매왕 이벤트
- 뉴트리원 : 쇼핑라이브 경품 이벤트

■ ANOVA TEST

귀무가설: 고객사 별 이벤트는 시간대 별 총 품목 수량에 영향을 미치지 않는다.

	df	sum_sq	mean_sq	F	PR(>F)
nestle	1.0	2.372464e+07	2.372464e+07	72.647177	2.446179e-17
nut_one	1.0	1.728581e+07	1.728581e+07	52.930848	4.416771e-13
lg	1.0	4.226676e+07	4.226676e+07	129.424959	2.248843e-29
Residual	2924.0	9.549008e+08	3.265735e+05	NaN	NaN

- 귀무가설이 기각되었고,
행사일들이 총 품목수량에 통계적으로 유의하다

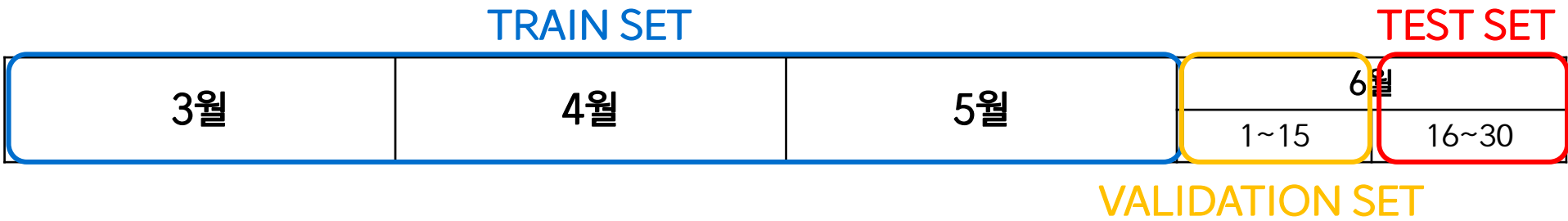


과제 수행 – BASELINE 작성

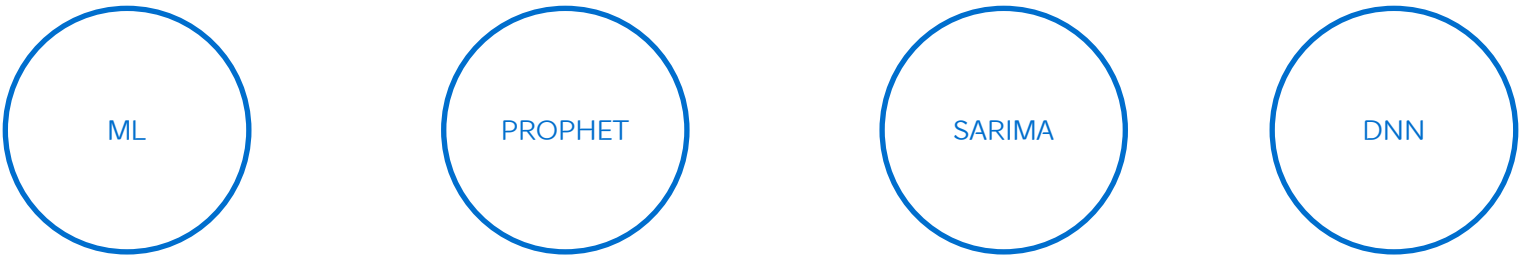
- DATA SET: 곤지암 창고의 B2C출고 데이터의 주문 날짜, 시간 별 상품 총 구매 수량

6/1																							
0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	2	2	2	2	2
0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9	0	1	2	3
6/2																							
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9	0	1	2	3
6/3																							
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9	0	1	2	3

- TRAIN TEST SPLIT



- PREDICTION METHOD



— 과제 수행 - ML

과제 수행 – MACHINE LEARNING

■ 핵심요인과 변수(FEATURE)선정

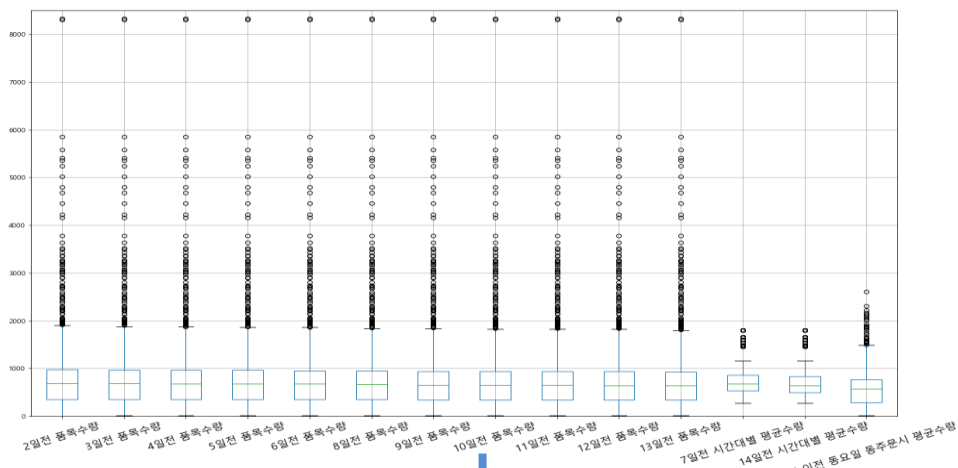
핵심 요인	선정한 변수(feature)
최근 물량 트렌드	동요일을 제외한 최근 2주간 시간대별 품목수량
	최근 2주 동요일 하루 간 시간대별 평균 품목수량
캘린더 적용치	고객사 3사 이벤트일
요일별 특성치	14일 이전의 모든 동요일 동주문시간에 대한 평균 품목수량

주문시	주문요일	주문일	2일전 품목 수량	3일전 품목 수량	4일전 품목 수량	5일전 품목 수량	6일전 품목 수량	8일전 품목 수량	9일전 품목 수량	...	11일 전 품 목 수 량	12일 전 품 목 수 량	13일 전 품 목 수 량	7일전 시간대 별 평균수량	14일전 시 간대별 균수량	14일 이 전 동요일 동주 문시 평균 수량	레드 위크	네 슬 레	뉴 트 리 원	
0	0	0	22	452	509	594	521	611	445	351	...	580	710	544	501.333333	484.0	500.0	1	1	0
1	1	0	22	579	844	683	720	771	456	434	...	610	681	828	501.333333	484.0	575.0	1	1	0
2	2	0	22	455	689	1134	690	944	560	453	...	591	664	780	501.333333	484.0	679.0	1	1	0
3	3	0	22	424	609	915	587	712	563	285	...	860	619	787	501.333333	484.0	671.0	1	1	0
4	4	0	22	451	656	616	595	981	602	325	...	646	681	721	501.333333	484.0	599.0	1	1	0

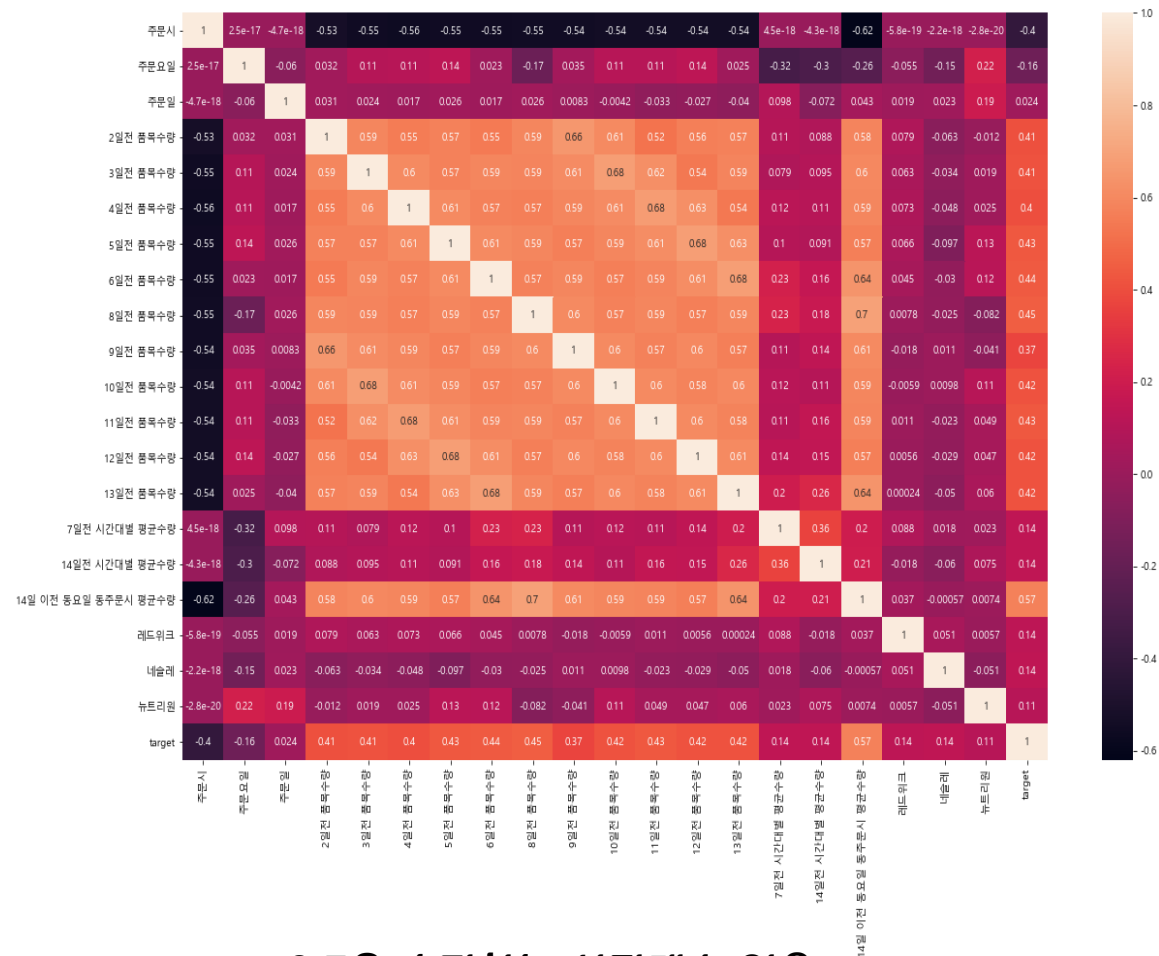
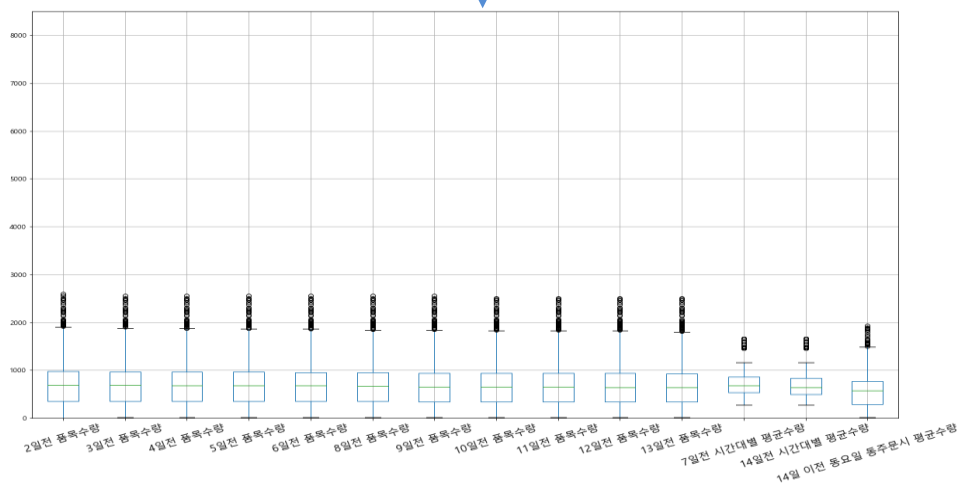
5 rows × 21 columns

과제 수행 — MACHINE LEARNING

■ 변수(FEATURE) 이상치 치환 및 상관관계 확인



15백분위수와 85백분위수 밖의 값들은
그 경계값으로 치환



■ 0.7을 초과하는 상관계수 없음

과제 수행 – MACHINE LEARNING

- 개별 모델링과 OOF(OUT-OF-FOLD)

〈검증데이터 RMSE〉

	개별 모델링	OOF
XGBoost	806.05	460.06
Lightgbm	612.14	433.31
ExtraTrees	665.27	475.76
Catboost	609.26	410.74
RandomForest	629.58	485.06

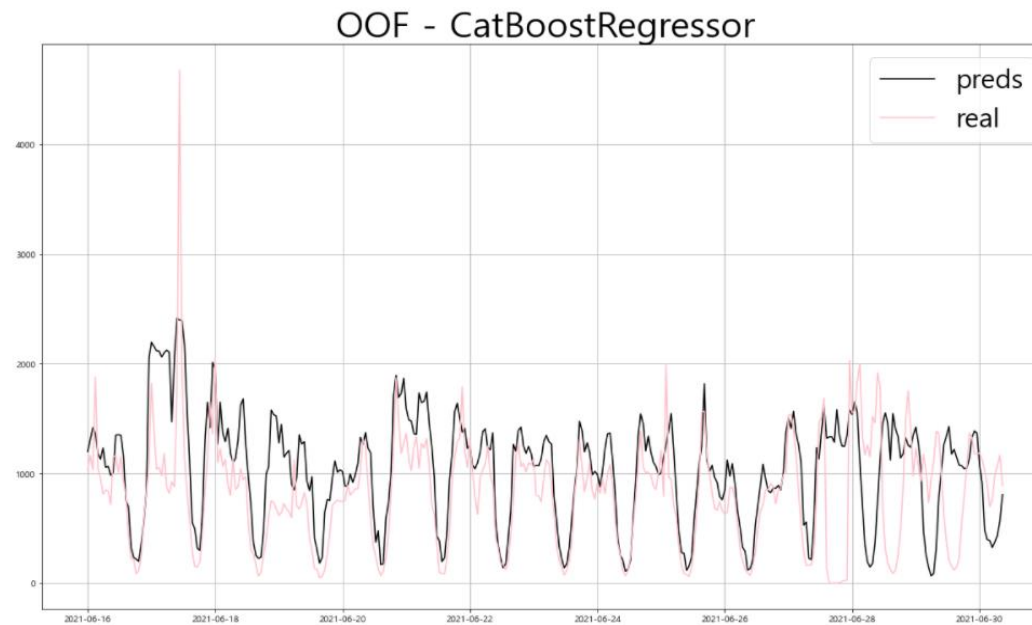
각 모델 베이지안 최적화를 통해
하이퍼 파라미터 최적화 후 모델링



OOF – Catboost Model Test data RMSE → 510.05

과제 수행 — MACHINE LEARNING

- 최고 성능 모델 예측값 시각화

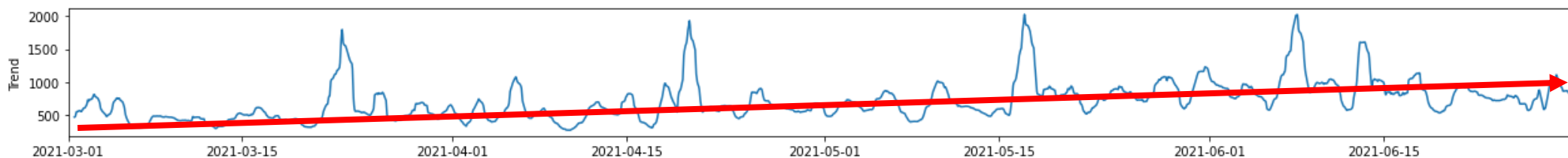
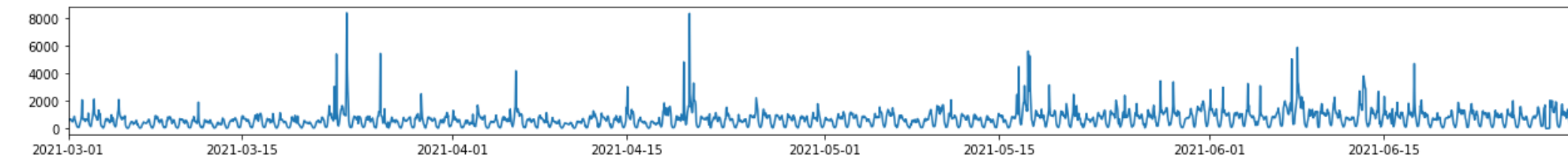


- 시간대별 계절적 추세를 잘 맞춤
- 비정상적으로 수요가 증가하는 시간대(이상치)는 잘 맞추지 못함

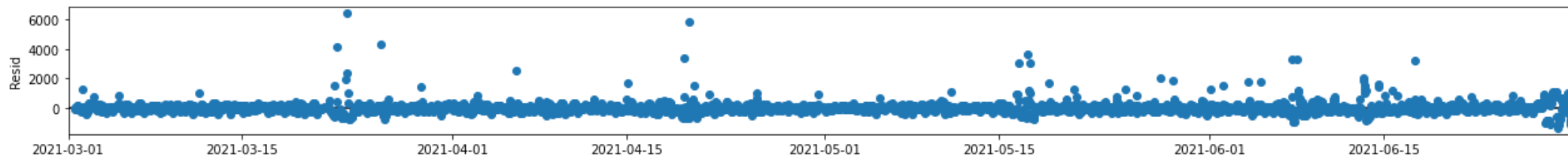
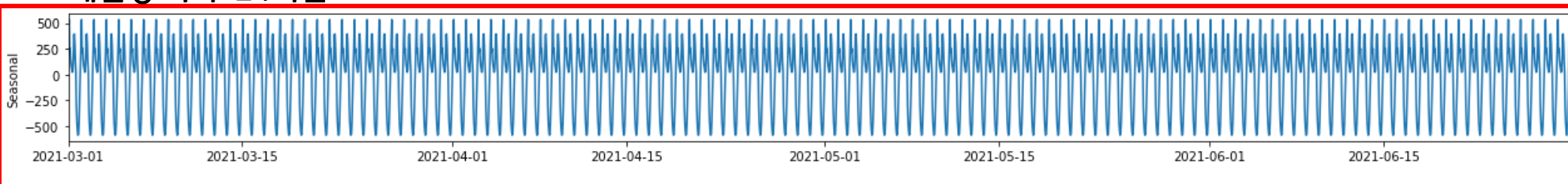
— 과제 수행 - SARIMA

과제 수행 — SARIMA

■ 시계열 분해

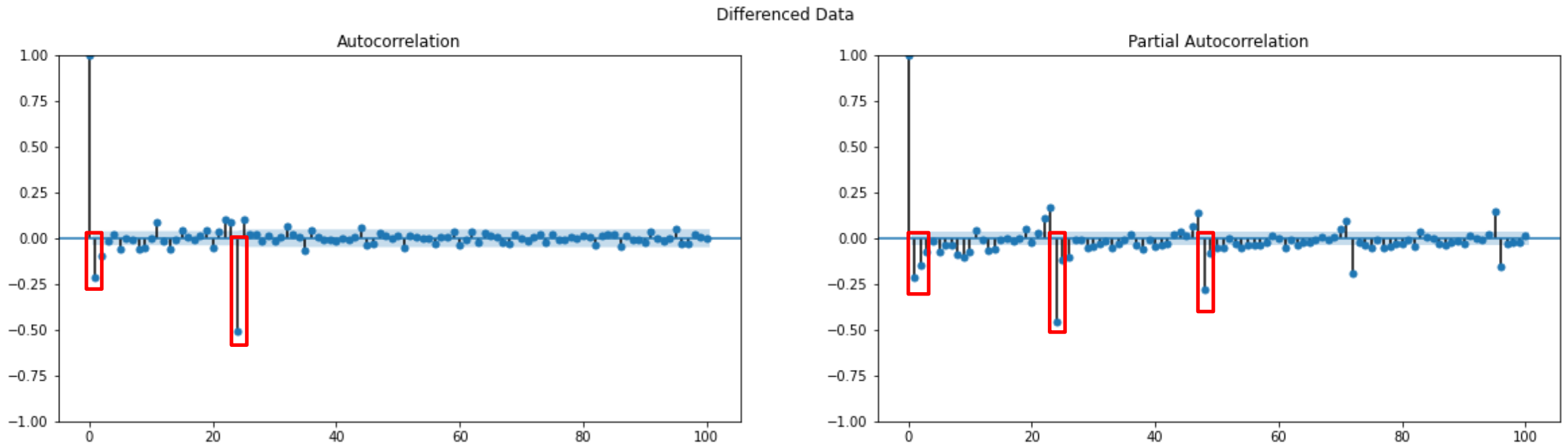


계절성 시차: 24시간



과제 수행 – SARIMA

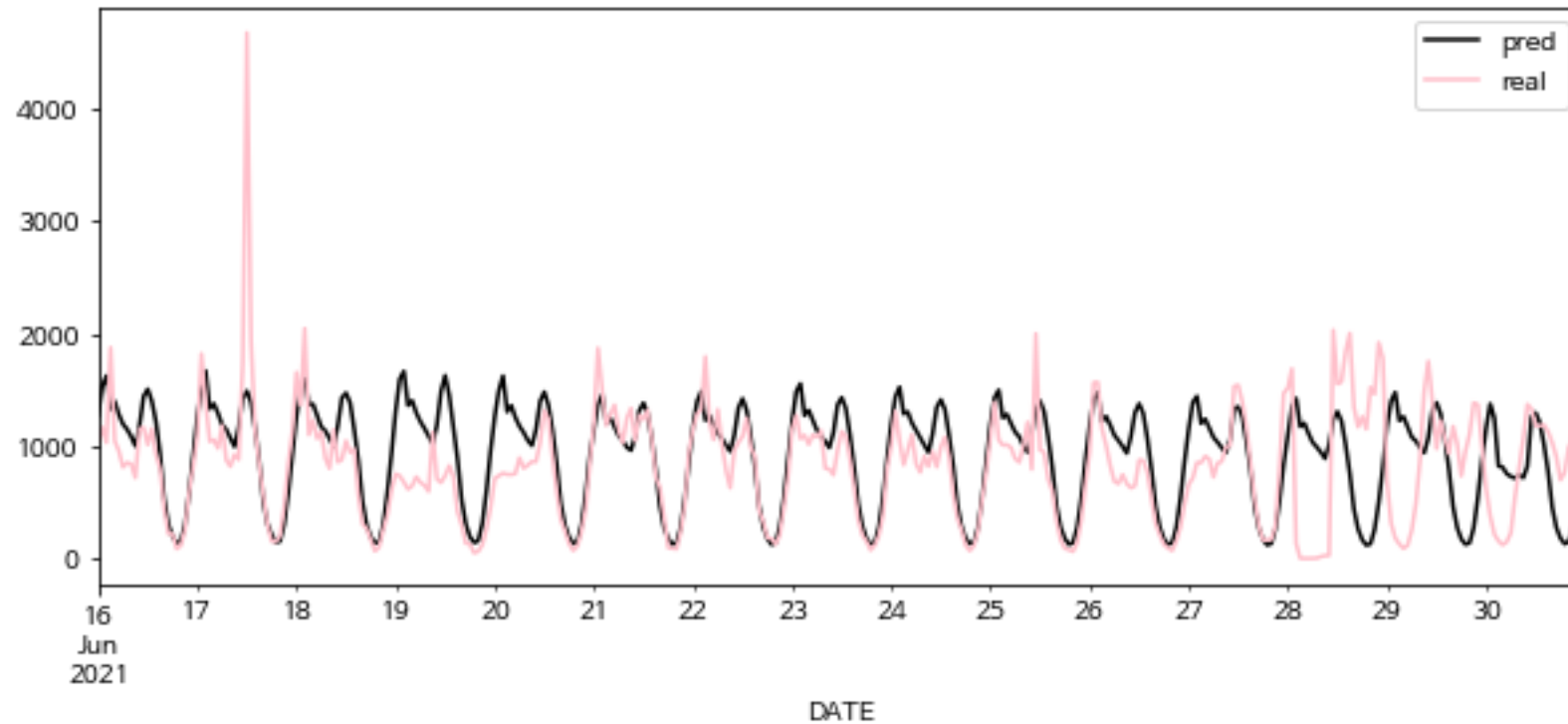
- 정상성 만족을 위해 로그 변환 시행



- SARIMA (2,1,1)(2,1,1,24)

과제 수행 — SARIMA

- SARIMA 시계열 예측 결과



- RMSE : 474.54
- TREND와 SEASONAL TREND는 잘 반영
- 정확한 예측 X

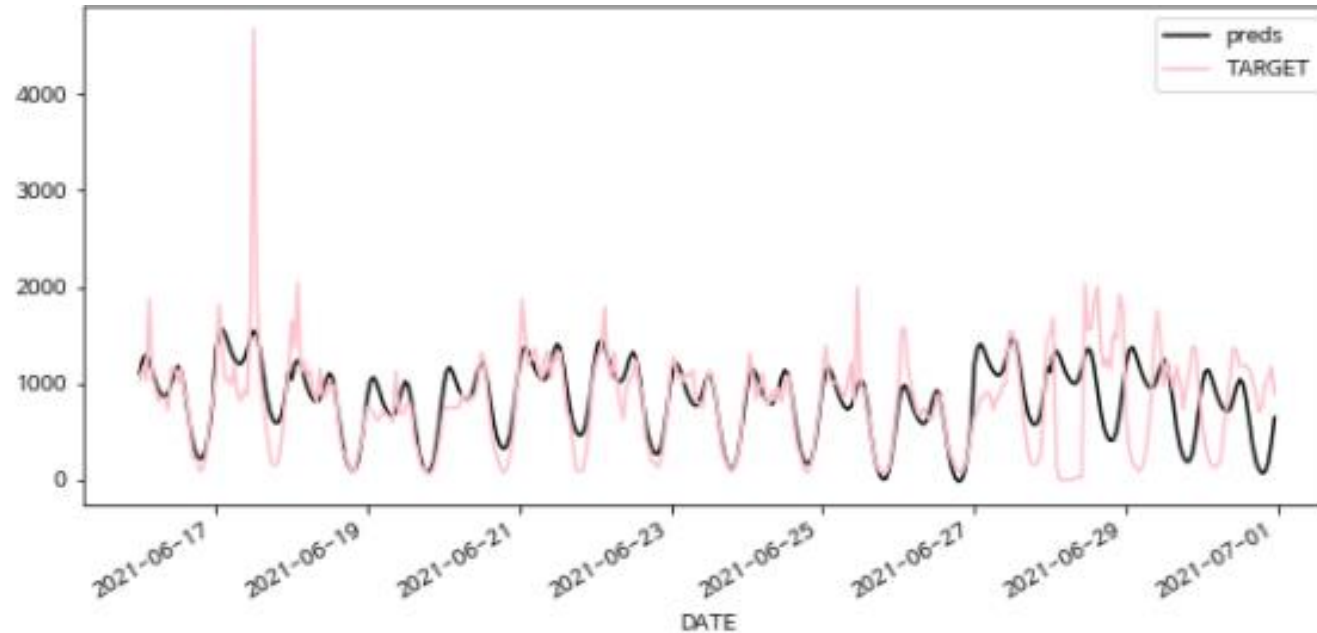
— 과제 수행 - PROPHET

■ PROPHET 주요 구성요소



6/1	6/2	6/3	6/4	6/5	6/6	6/7	6/8	6/9	6/10
TRAIN					PREDICT				
TRAIN						PREDICT			
TRAIN							PREDICT		
TRAIN								PREDICT	

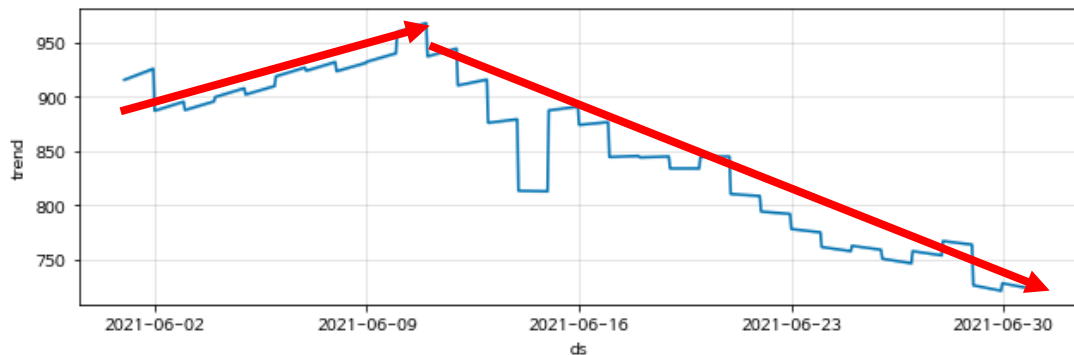
과제 수행 — PROPHET



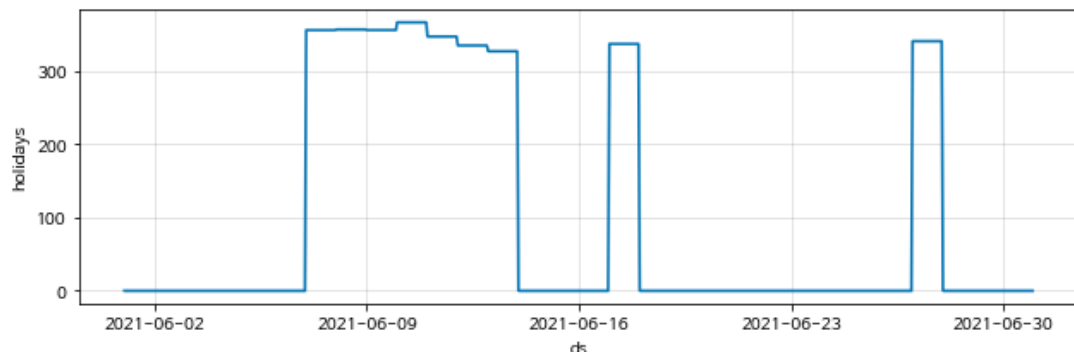
- RMSE: 429.193
- 대부분의 추세 정확히 맞춤.
- 비정상적으로 변화하는 시간(이상치)는 예측 실패

과제 수행 – PROPHET

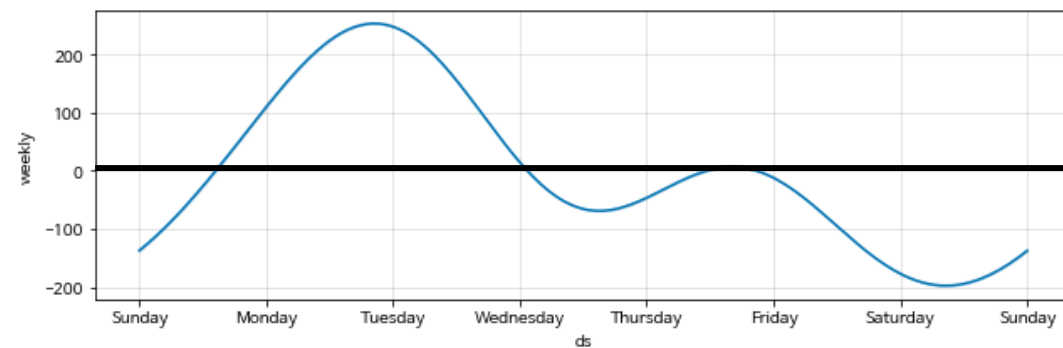
COMPONENTS 분석



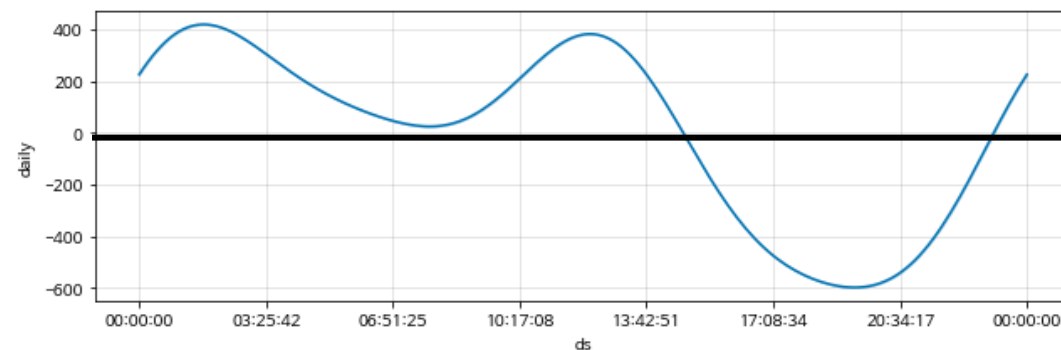
- 6월 10일까지 증가하다가 감소하는 트렌드



- 이벤트일(HOLIDAY)은 약 300~400만큼 증가에 영향



- 월, 화에는 주문 수가 증가하는 영향
- 목, 토, 일에는 감소하는 영향



- 약 14시 부터 23시까지는 예측에 음의 영향
- 그 외의 시간은 양의 영향

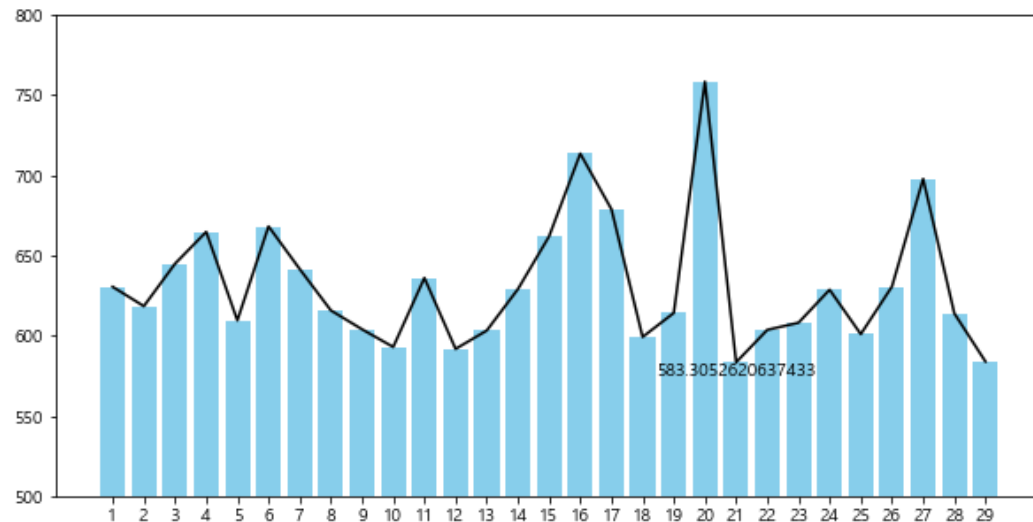
— 과제 수행 - DNN

과제 수행 – DNN

▪ SLIDING WINDOW

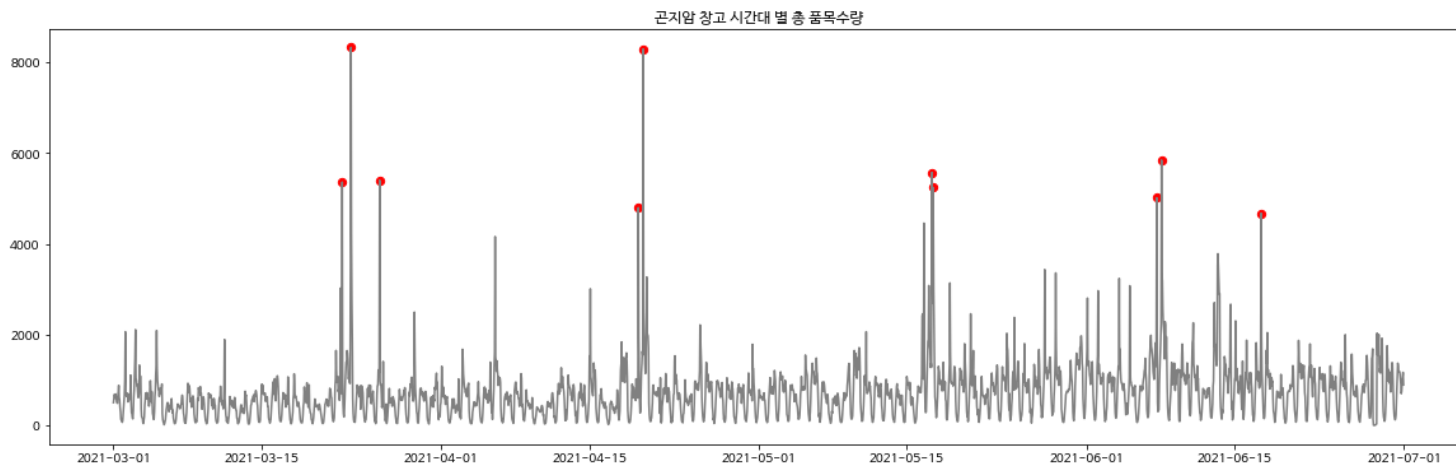
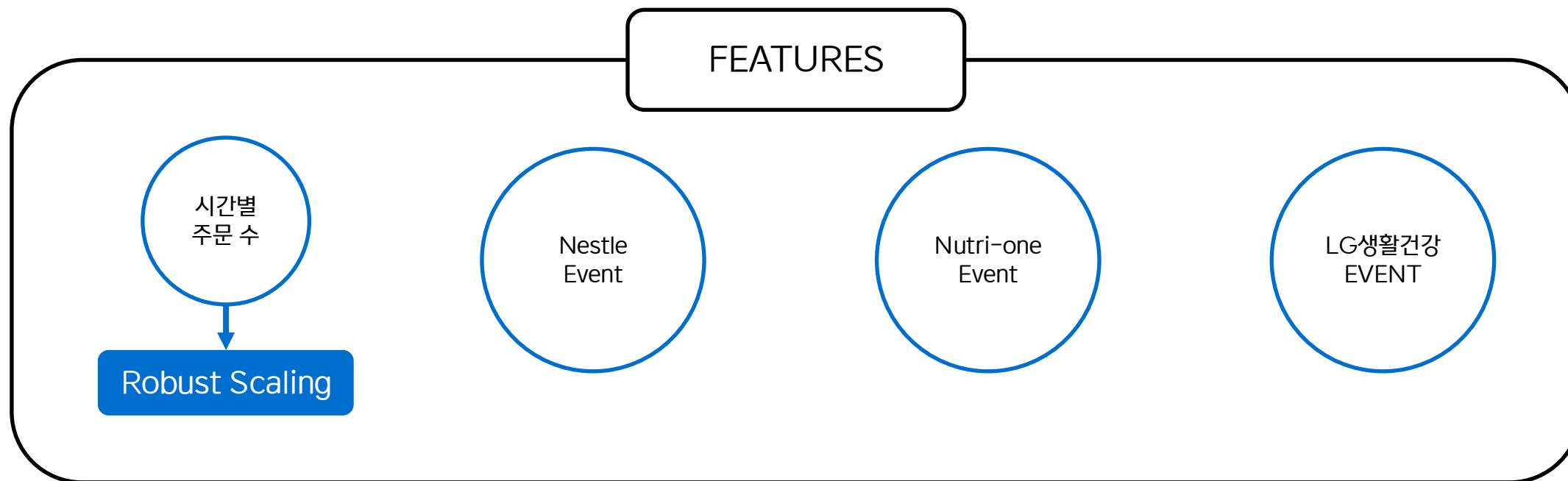
6/1	6/2	6/3	6/4	6/5	6/6	6/7	6/8	6/9	6/10
INPUT						PRED			
	INPUT						PRED		
		INPUT						PRED	
			INPUT						PRED

▪ INPUT WINDOW 정의



- LSTM을 사용한 GRID SEARCH
- HIDDEN LAYER: 1
- HIDDEN NODE: 64
- INPUT WINDOW가 21DAY일 때, 성능이 가장 높음.

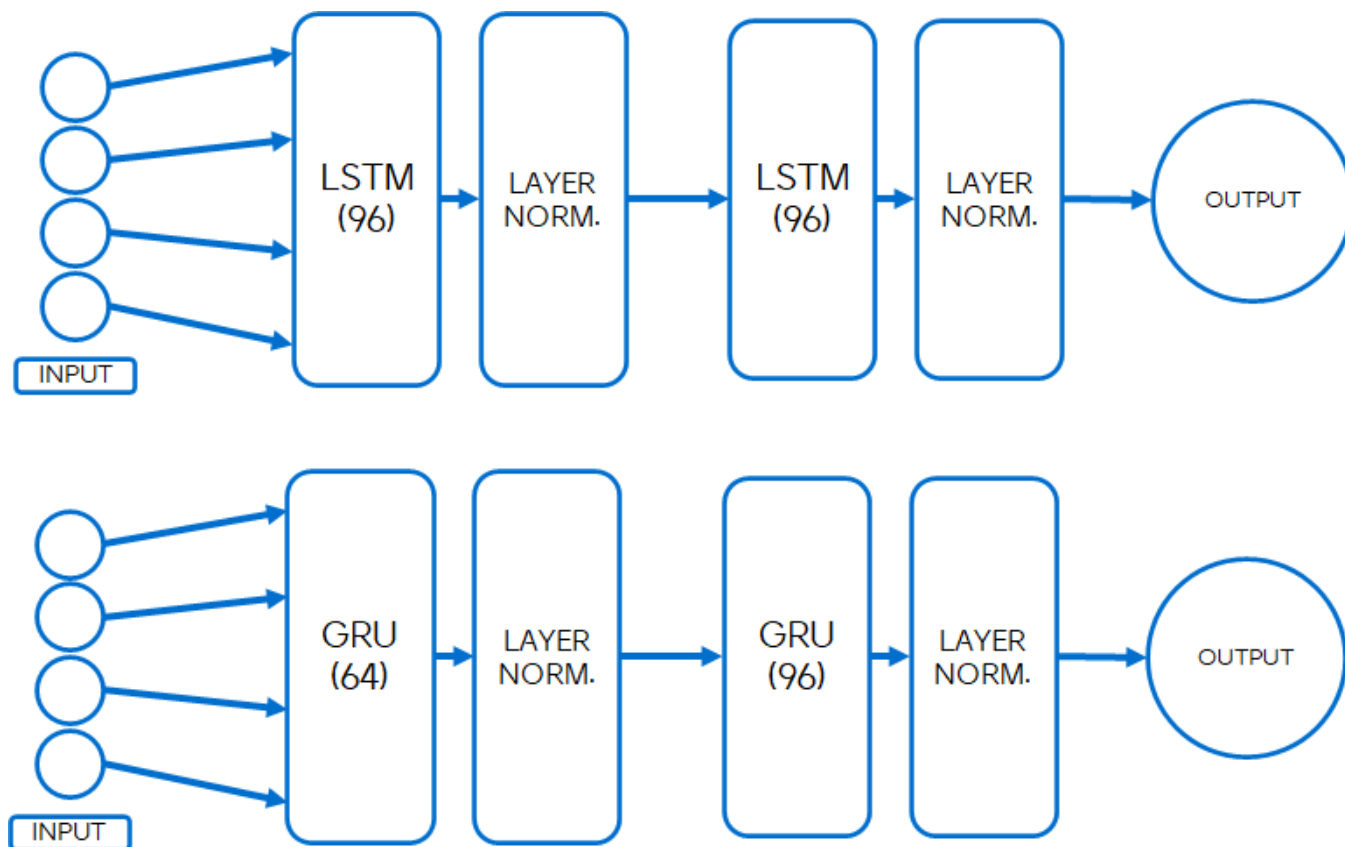
과제 수행 - DNN



- 시간별 이상치 존재
- 이상치는 상위 고객사 3개의 이벤트와 상관관계가 존재하므로 정확한 예측을 위해 FEATURE로 만들
- 각 FEATURE들의 SCALE을 맞추기 위해 SCALING 진행
- TRAIN에서 이상치의 영향을 줄이기 위해 ROBUST SCALING 시행

과제 수행 – DNN

- SEQ2SEQ 모델 사용
- RNN은 출력과 먼 위치에 있는 정보 잘 기억 못하기에 이를 보완한 LSTM, GRU 사용



과제 수행 — DNN

■ 딥러닝



- LSTM: 이상치에 소극적으로 예측 — RMSE: 357.871
- GRU: 이상치에 과감하게 예측 — RMSE: 407.653
- LSTM&GRU 앙상블 — RMSE: 326.683

과제 수행 - 최종 모델 선정

■ 모델 앙상블 결과

모델	ML(M)	SARIMA(S)	Prophet(P)	DNN(D)	M+S	M+P	M+D
RMSE	410.74	474.54	429.19	326.68	475.35	445.30	325.53
모델	S+P	S+D	P+D	M+D+P	M+S+D	S+P+D	M+S+P+D
RMSE	436.72	346.51	331.02	361.71	376.24	372.40	354.16

- 딥러닝(LSTM+GRU 앙상블) 단일모델을 사용하였을 때 성능이 가장 높다.
- 그러므로 딥러닝을 최종 모델로 선정

활용 방안 및 기대 효과

활용방안 및 기대효과

- 당일 배송, 새벽배송 실시할 것을 고려해 계산식 정의
- 가정

①

업무시간	오전조	오후조	야간조
	09~17	17~익일01	01~09
주문마감시간	익일배송	당일배송	새벽배송
↓ +2H	00시	12시	18시
출고시간 (차량출발시간)	익일배송	당일배송	새벽배송
	02시	14시	20시
업무자 담당 주문 시간	오전조	오후조	야간조
	07~15	15~23	23~익일07
UPH (unit per hour)	PICK	PACK	HUB(상차)
	100	100	200

②

주문부터 출고까지 두 시간 소요된다고 가정

↓

사원의 작업 물품은
두 시간 전에 주문 들어온 물품이라고 가정

③

$$y_{\text{오전}} = \frac{\sum_{i=7}^{15} x_i}{100 \times 8} + \frac{\sum_{i=7}^{15} x_i}{100 \times 8} + \frac{x \sum_{i=7}^{15} x_i}{200 \times 8}$$

$$y_{\text{오후}} = \frac{\sum_{i=15}^{23} x_i}{100 \times 8} + \frac{\sum_{i=15}^{23} x_i}{100 \times 8} + \frac{x \sum_{i=15}^{23} x_i}{200 \times 8}$$

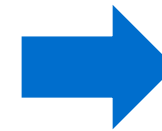
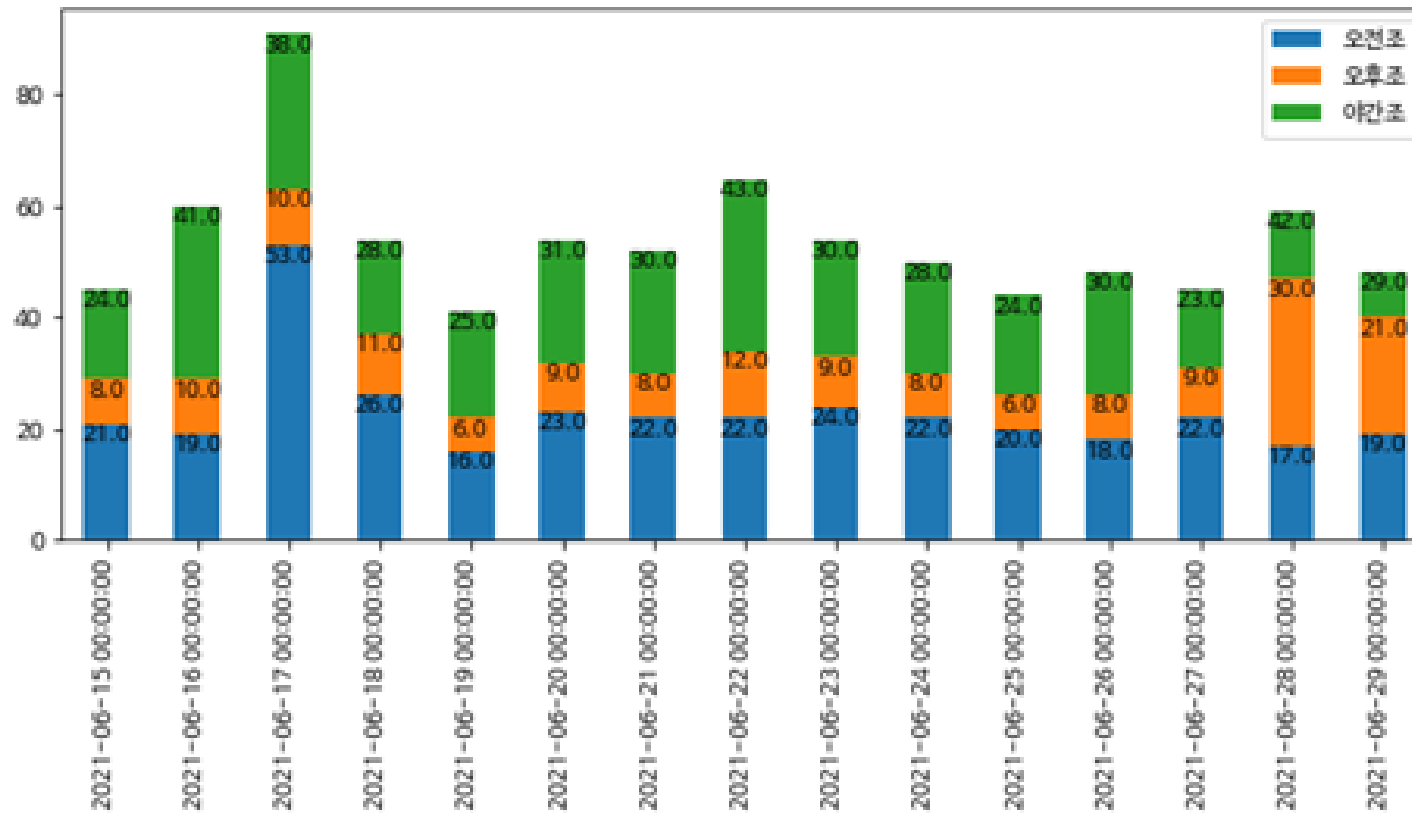
$$y_{\text{야간}} = \frac{\sum_{i=23}^{\text{익일}7} x_i}{100 \times 8} + \frac{\sum_{i=23}^{\text{익일}7} x_i}{100 \times 8} + \frac{x \sum_{i=23}^{\text{익일}7} x_i}{200 \times 8}$$

- 보완점

1. UPH → 작업자의 평균 UPH
2. 업무자 담당 물건 주문시간 → CJ대한통운에 적합한 시간

활용방안 및 기대효과

TEST SET 예측값을 사용해 필요 인력 계산



- 비용 절감
- 고객 신뢰 강화
- 경쟁력 강화



