



“날씨가 바뀌는 광고 전략”

딥러닝으로 예측한 지하철 혼잡도를 활용한 광고 최적화 방안



부제: 시계열 모델 기반 장기간 예측 및 사후 분석

목차

01

프로젝트 개요

02

프로젝트 팀 구성 및 역할

03

프로젝트 수행 절차 및 방법

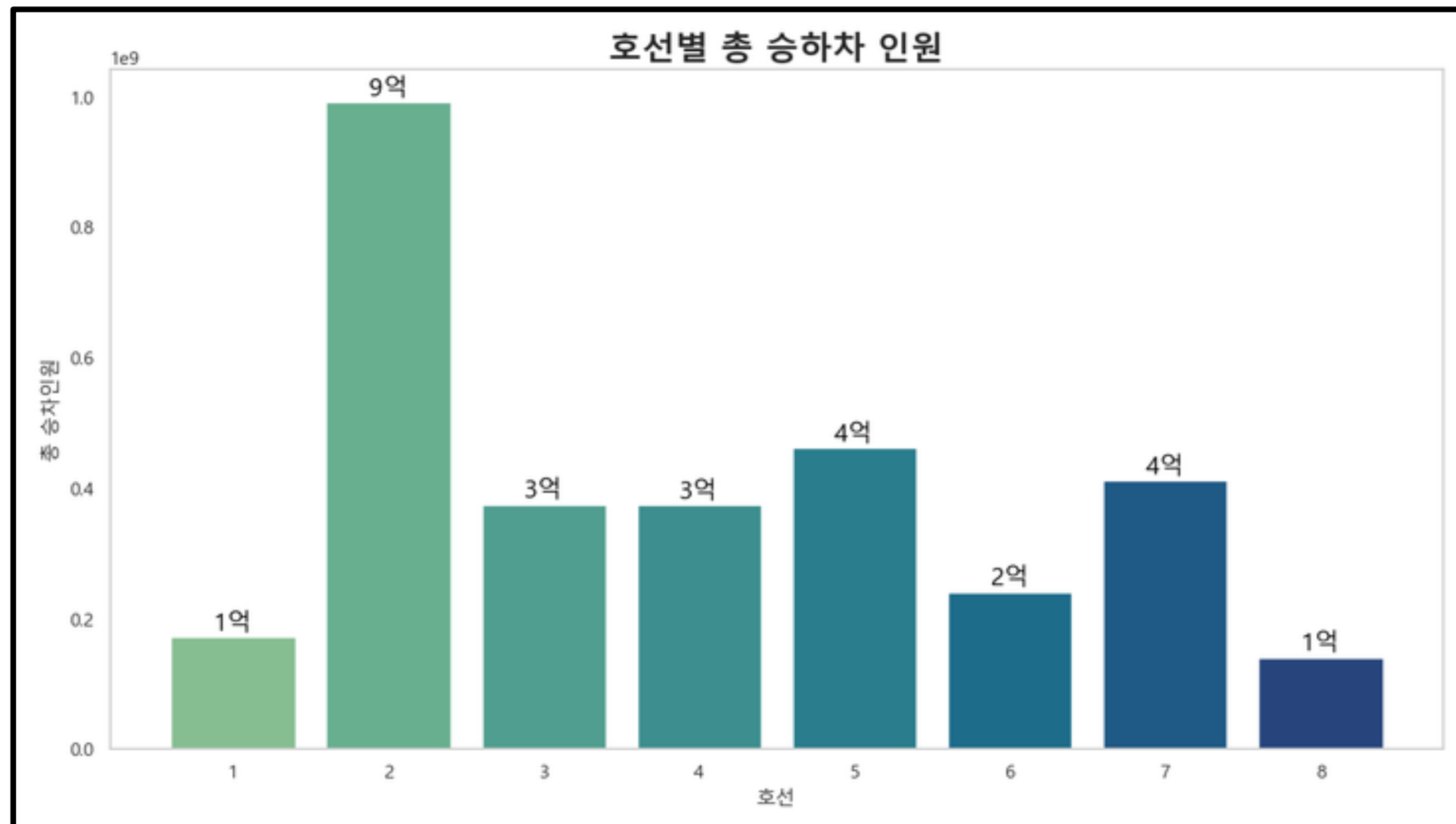
04

프로젝트 수행 결과

05

자체 평가 의견

□ 서울 지하철, 데이터로 본 광고의 기회



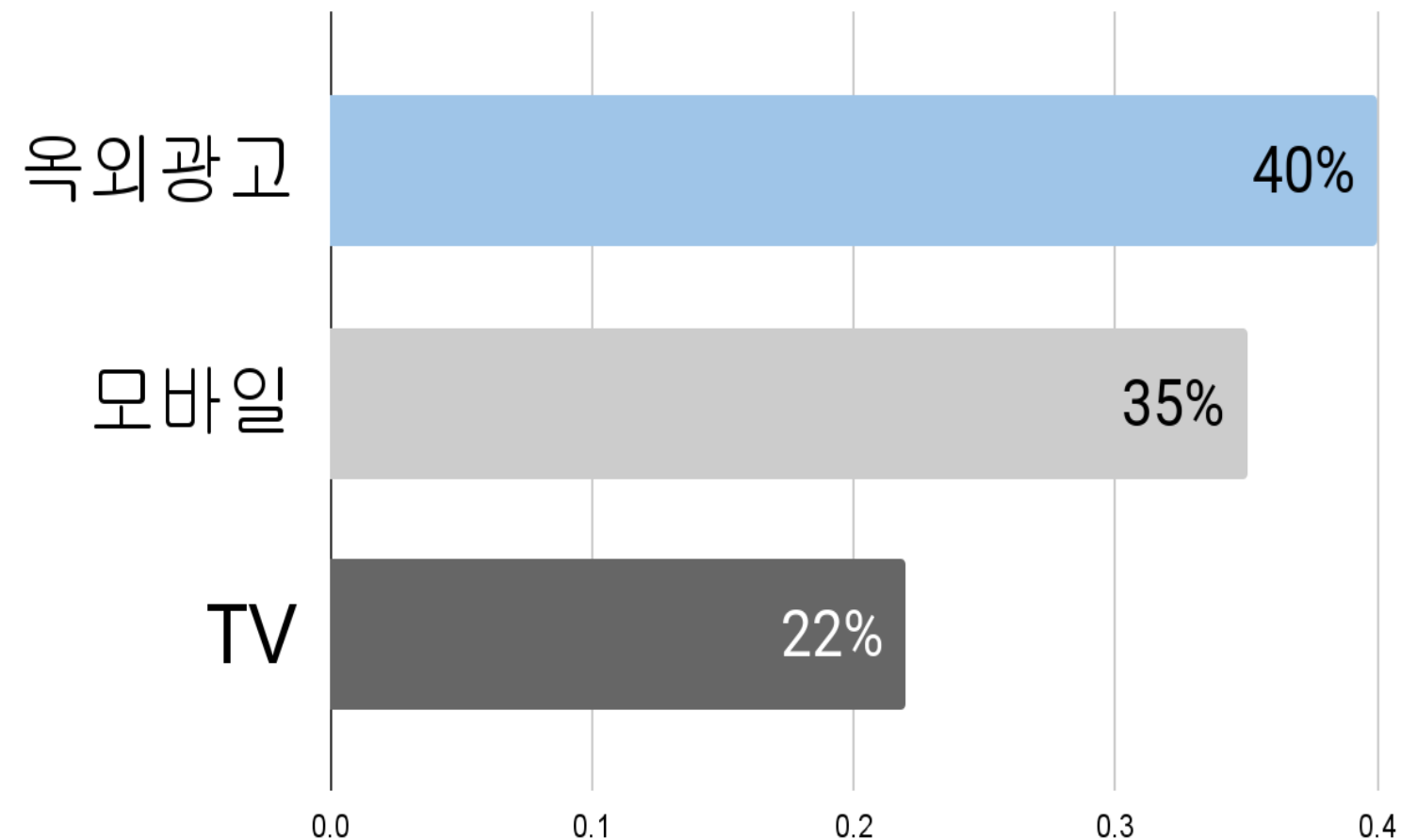
출처 : 서울교통공사_일별통행통계

2024년 서울지하철 1~8호선
총 승하차 인원
27억명

지하철은
시민들의 주요 교통수단으로
다양한 성별과 연령층이 이용하여
광고 노출에 매우 유리한 장소

□ 지하철 광고의 효과와 특성

광고 매체별 비보조 회상률(%)

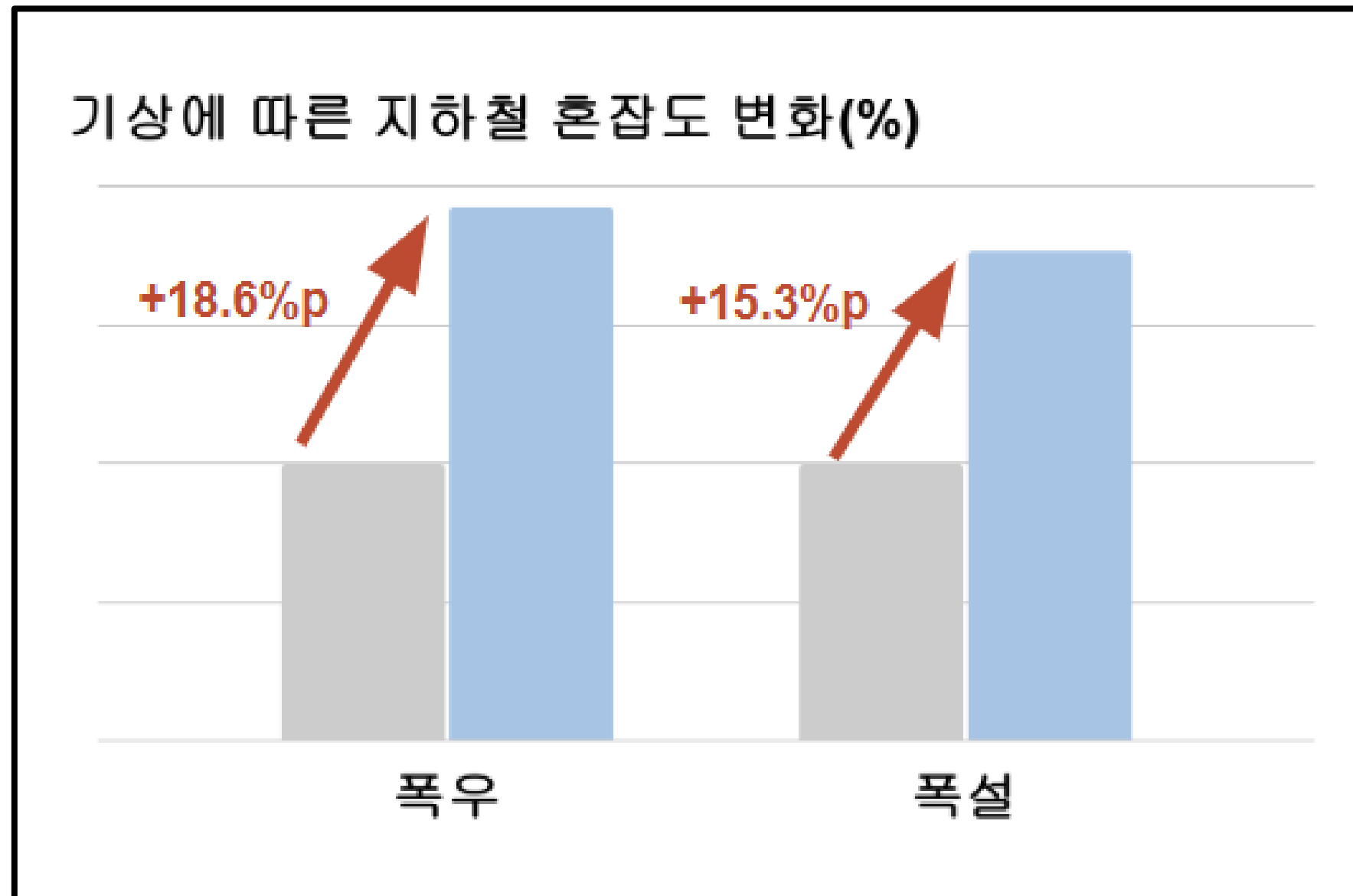


지하철 광고의 시각적 임팩트
+
회피 불가능한 노출 특성
=
광고 메시지가 더 오래 각인되는 효과

출처 : 영국 기반의 마케팅 및 광고 전문 매체, The Drum

**비보조회상률 : 브랜드나 제품을 제시하지 않고 소비자가 떠올리는 브랜드의 빈도를 측정

□ 기상에 따른 지하철 혼잡도의 증감



2021년 대한교통학회 게재된 논문 결과,
폭우, 폭설 시 지하철 이용률이 증가
→ **기상에 따라 지하철 혼잡도 증가**

출처 : 이승연, 이영인. (2021-11-10). 날씨가 대중교통 이용자의 지하철 또는 버스 선택에 미치는 영향 :스마트카드 데이터를 사용하여. 대한교통학회 학술대회지

□ 지하철 광고료 책정의 한계점

*지하철 차내광고

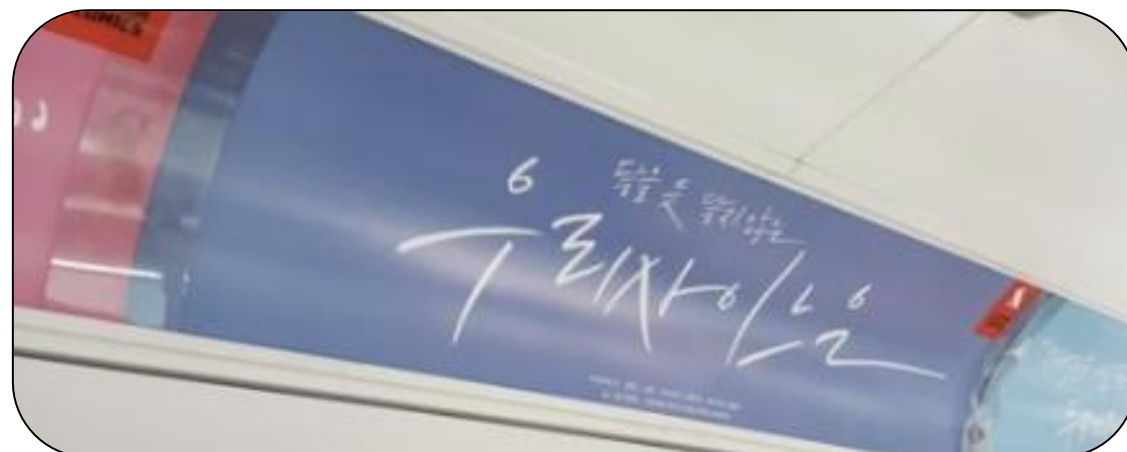
[출처 : 광고대행사
좋은광고연구소]

월 광고료(원)	액자형	출입문 조명	래핑광고(1량)
	18,000~30,000	20,000~40,000	6,000,000~10,080,000

- 호선·매체 구성·인쇄·출력비 기준 **월정액화**로 기상에 따른 **지하철 혼잡도 반영**x
- **날씨 데이터를 기반**으로 지하철 혼잡도 정량화하여 **광고비 차등화가 필요**
 - 높은 혼잡도 예상 > 시야 좁음 > 액자형, 천장 모서리형 광고료 상향 조정
 - 낮은 혼잡도 예상 > 시야 넓음 > 래핑 광고 광고료 상향 조정

□ 지하철 혼잡도에 따른 열차 내 광고 위치 선정 전략

[모서리광고&출입문 조명광고]



[래핑광고]



□ 지하철 광고의 효과 측정

객관적인 광고 효과 지표
부재로 광고 효과 입증이
어려움

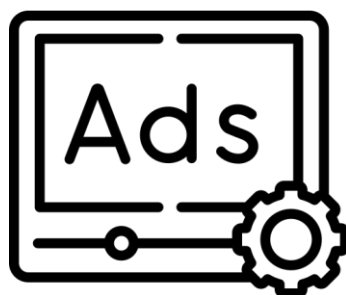
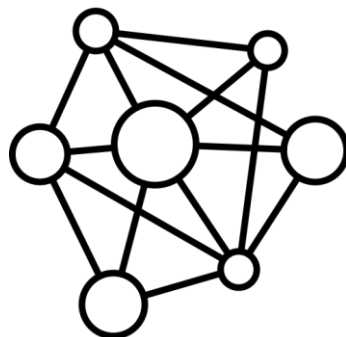
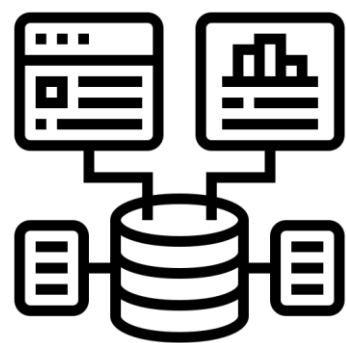


신뢰할 수 있는 지표
개발과 표준화 필요



한국지방재정공제회에서
필요성을 인식 후,
<옥외광고 효과지표
표준화 모델>개발 진행 중

- OOH (Out of home 옥외광고물) :공중에게 항상 또는 일정기간 계속 노출되어 공중이 자유로이 통행하는 장소에서 볼 수 있는것으로서 간판 그 밖에 이와 유사한 것



서울 지하철의 1시간 단위 혼잡도 데이터와
기상 및 날씨이슈(폭염·황사·태풍·한파) 데이터를 결합하여 분석

다양한 전처리 기법으로 시계열 모델 구조를 실험하여
기상 영향을 반영한 지하철 혼잡도를 예측

예측된 장기 혼잡도 추세를 지하철 광고 전략 수립에 활용
iM뱅크를 예시로 마케팅 전략 제시



박민정

조장
EDA
데이터 전처리
모델 성능
개선



서정원

EDA
데이터 전처리
모델 성능
개선



이우태

EDA
데이터 전처리
모델 구축
모델 성능
개선



전은진

EDA
데이터 전처리
모델 성능
개선



최다은

EDA
데이터 전처리
모델 성능
개선

구분	기간	활동	비고
주제 기획	2025.5.19-2025.5.20	주제 기획, 기획서 작성	
데이터 수집 및 전처리	2025.5.21-2025.5.22	필요 데이터 및 수집 절차 정의 외부 데이터 수집 데이터 정제 및 정규화	기초모델링 병행
데이터 기초 분석	2025.5.22	EDA	기초모델링 병행
모델링	2025.5.23-2025.5.26	모델 구축 및 튜닝	
PPT 작성	2025.5.27-2025.5.28	발표용 PPT 제작	

□ 수집 데이터

- 데이터 기간: 2021년 ~ 2023년 (3년)
- 지하철 혼잡도 데이터
 - ⊙ 서울 지하철 1시간 단위 혼잡도 정보(1~8호선, 총 300여개 지하철 역)
 - ⊙ 서울통공사_일별통행통계
- 기상 데이터
 - ⊙ 지하철역 인근 AWS, ASOS 지점의 1시간 단위 관측 및 객관분석 자료, 미세먼지
 - ⊙ 기상이슈별 데이터 - 태풍, 한파, 황사

□ 모델 활용변수(1)

출처	데이터명	컬럼명	설명
서울교통 공사	지하철 혼잡도 데이터	시간	날짜 및 시각 05시~01시 (지하철 운행시간 동안만 제공)
		호선	지하철 호선
		역번호	역 번호
		역명	역 명
		상하구분	지하철 상행 하행 구분
		AWS 지점 코드	AWS(기상관측장비) 지점 코드
		혼잡도	열차 내 혼잡도(%)

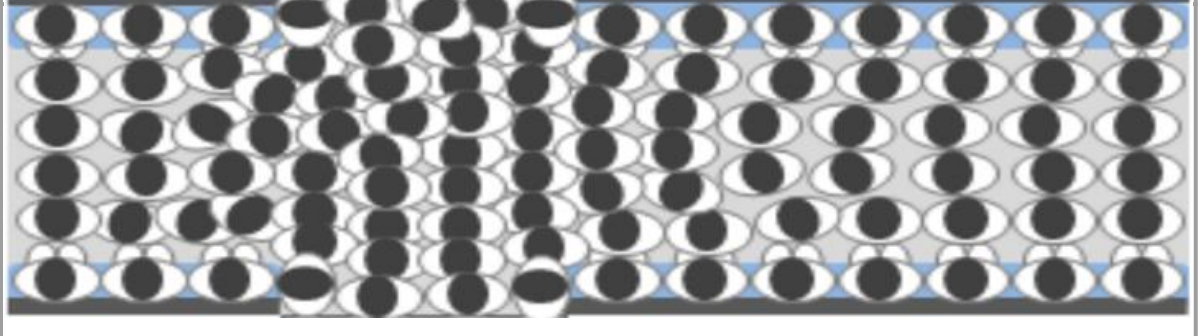
□ 모델 활용변수(2)

출처	데이터명	컬럼명	설명
기상청	기상 데이터	기온	정시 기온
		체감온도	습도나 바람 등을 고려해 사람이 체감적으로 느낀다고 추상하는 온도
		풍향	정시 10분 평균 동향(degree)
		풍속	정시 10분 평균 풍속(m/s)
		일 강수량	0~24시까지의 일 강수량(mm)
		시간 강수량	1시간 강수량(mm)
		상대습도	정시 상대 습도(%)
		일사량	단위 시간 동안 단위 면적에 도달하는 태양 복사 에너지의 양
	미세먼지	미세먼지 농도(PM10)	서울특별시 일평균 미세먼지 농도(pm10)

□ EDA 활용변수

출처	데이터명	컬럼명	설명
기상청	황사 한파 태풍	날짜	해당 기상이슈가 발생한 일자
		황사여부	황사특보 발생 여부
		한파여부	한파특보 발생 여부
		태풍여부	태풍 발생일자 ~ 소멸일자에 해당 여부
공공데이터 포털	서울교통 공사_일별 통행통계	수송일자	날짜
		호선	서울 지하철 1~8호선 구분
		역명	역 이름
		승하차구분	승차/하차 구분
		승객유형	일반, 노인, 어린이, 청소년으로 분류
		시간대 별 승객 수	1시간 단위로 집계된 승객 수

$$\text{열차 혼잡도(\%)} = \frac{\text{실제 탑승 인원(인)}}{\text{열차 탑승 기준 인원(인)}} \times 100$$

혼잡도	재차인원	차내상태 설명	혼잡 상태도
50%	80명	좌석에 모두 착석하고 간간이 서 있음	
100%	160명	여유롭게 서 있음 앞에 사람들이 있어 시야가 다소 가려짐	
200%	320명	출입문 주변이 매우 혼잡함 서로 몸과 얼굴이 밀착되어 숨이 막힘	

수식 출처 : 서울교통공사

혼잡도 사진 출처 : 서울연구원 정책리포트 208

1. 결측치 처리

- 측정불가해서 -99로 기재된 기상데이터 존재
 - 5% 미만(기온, 풍향, 풍속, 일강수량, 시간강수량, 체감온도), 10% 미만(상대습도), 37% (일사량)

번호	방법	결정계수	RMSE
1	<ul style="list-style-type: none"> 5% 미만의 -99값 제거 (그 외) 10% 미만의 -99값은 선형 보간(상대습도) 10% 이상의 -99값은 이진 변수 생성(일사량) 	0.9624	4.5909
2	<ul style="list-style-type: none"> 10% 미만의 -99값은 선형 보간 10% 이상의 -99값은 이진 변수 생성(일사량) 	0.9610	4.6741
3	<ul style="list-style-type: none"> 일사량이 없는 시간은 0으로 대체 -99값이 있는 모든 변수는 선형 보간 	0.9623	4.5965
4	<ul style="list-style-type: none"> 일사량이 없는 시간은 0으로 대체 5% 미만의 -99값 제거 10% 미만의 -99값은 선형 보간(상대습도, 일사량) 	0.9575	4.8816

- 선형보간법 : 두 알려진 점 사이의 값을 직선으로 연결해 추정
- 일사량 : 지구 대기에 먼저 도달한 후 지표면에 도달하는 복사열

2. 이상치 처리

- 풍향 음수값 0.36%
 - 풍향은 바람이 불어오는 방향으로 $0^{\circ} \sim 360^{\circ}$ 값을 가짐 > 인스턴스 제거

3. 범주형 변수 원-핫 인코딩 및 임베딩

- 역명,호선,상하구분,AWS지점코드

4. 정규화

- 독립 변수의 스케일 차이를 제거하기 위해 **MinMax scaling** 적용
- 종속 변수를 정규분포 형태(평균 0, 표준편차 1)로 만들기 위해 **Standard scaling** 적용.

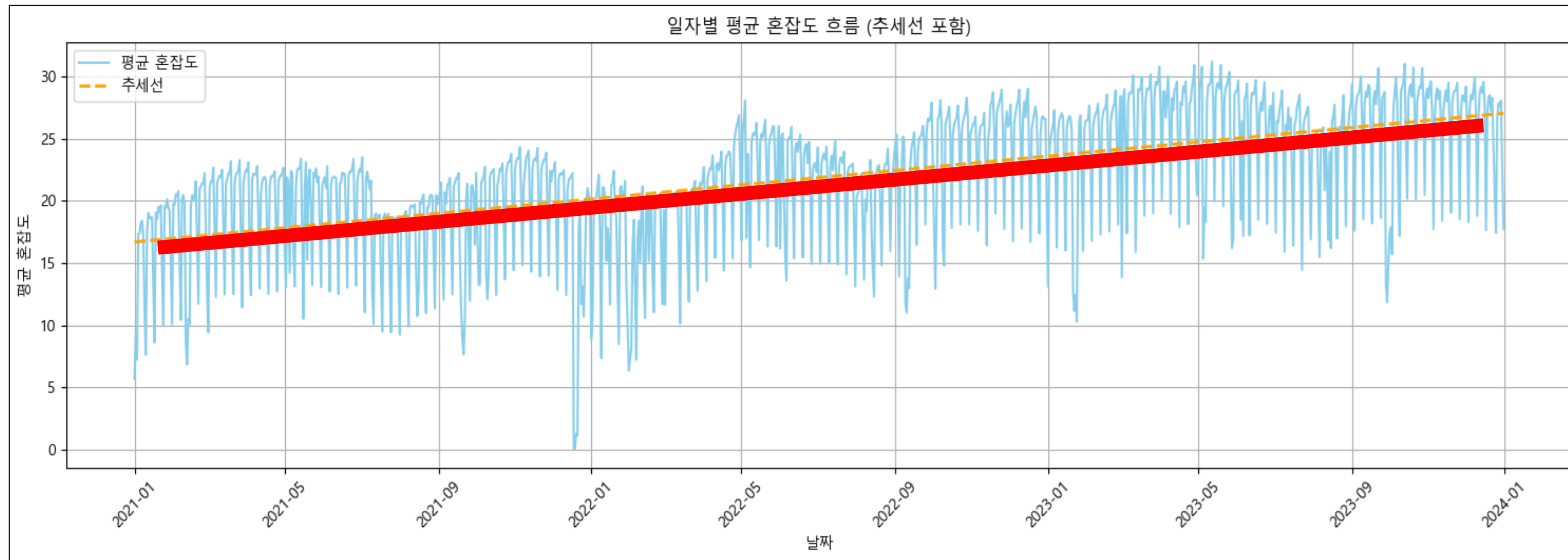
6. 파생 변수 생성

파생 변수명	방법	비고
요일, 년, 월, 일	주기성을 가질만한 시간 데이터 추출	
휴일여부	평일(0), 주말/공휴일(1)	라이브러리 holidayskr 활용 → 공휴일을 추출
시_sin, 시_cos 시간_sin, 시간_cos	<pre>df1['시_sin'] = np.sin(2 * np.pi * df1['시'] / 24) df1['시_cos'] = np.cos(2 * np.pi * df1['시'] / 24) df1['요일_sin'] = np.sin(2 * np.pi * df1['요일'] / 7) df1['요일_cos'] = np.cos(2 * np.pi * df1['요일'] / 7)</pre>	시계열 변수의 주기성 처리 (Sin, Cos 변환)
미세먼지_등급	미세먼지 >=151 : 매우나쁨 미세먼지 >=81 : 나쁨 미세먼지 >=31 : 보통 미세먼지 >=0 : 좋음	수도권대기환경청에 근거

04

데이터 기초특성(EDA)

□ 일자별 평균 혼잡도 흐름

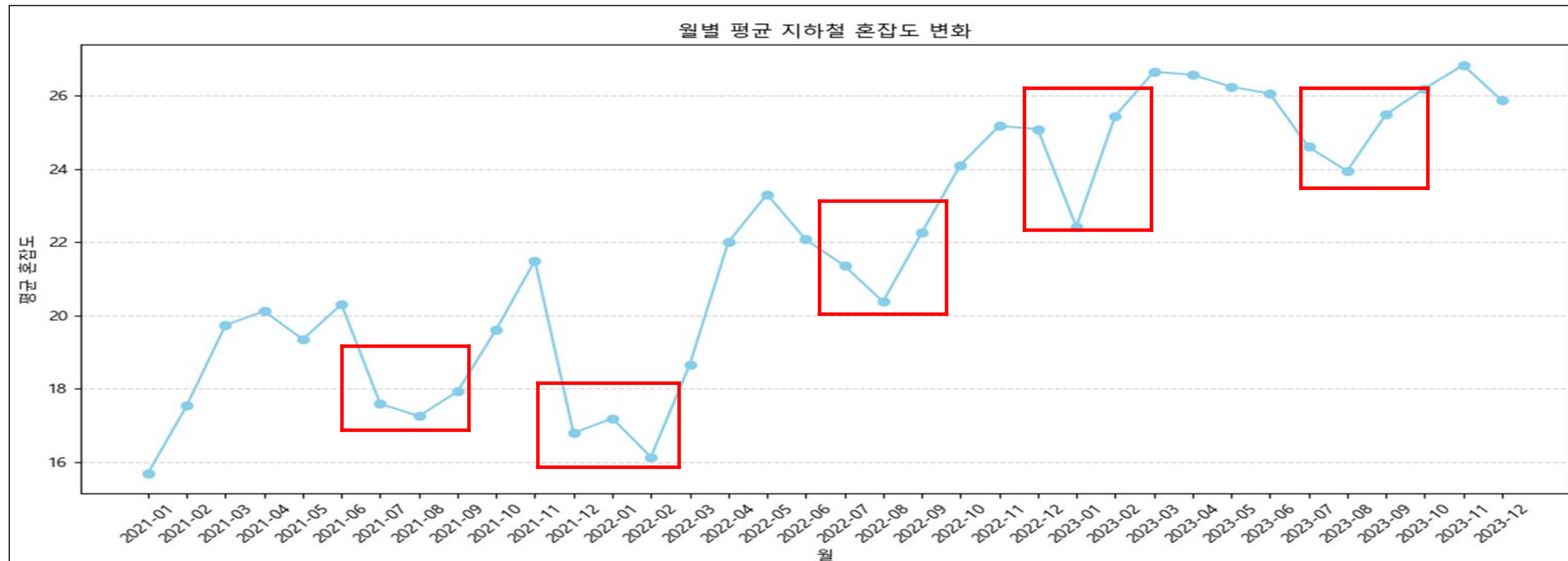


지하철의 혼잡도 **상승** 추세
코로나19 이후 사회적 거리두기 완화

04

데이터 기초특성(EDA)

□ 월별 평균 혼잡도 흐름

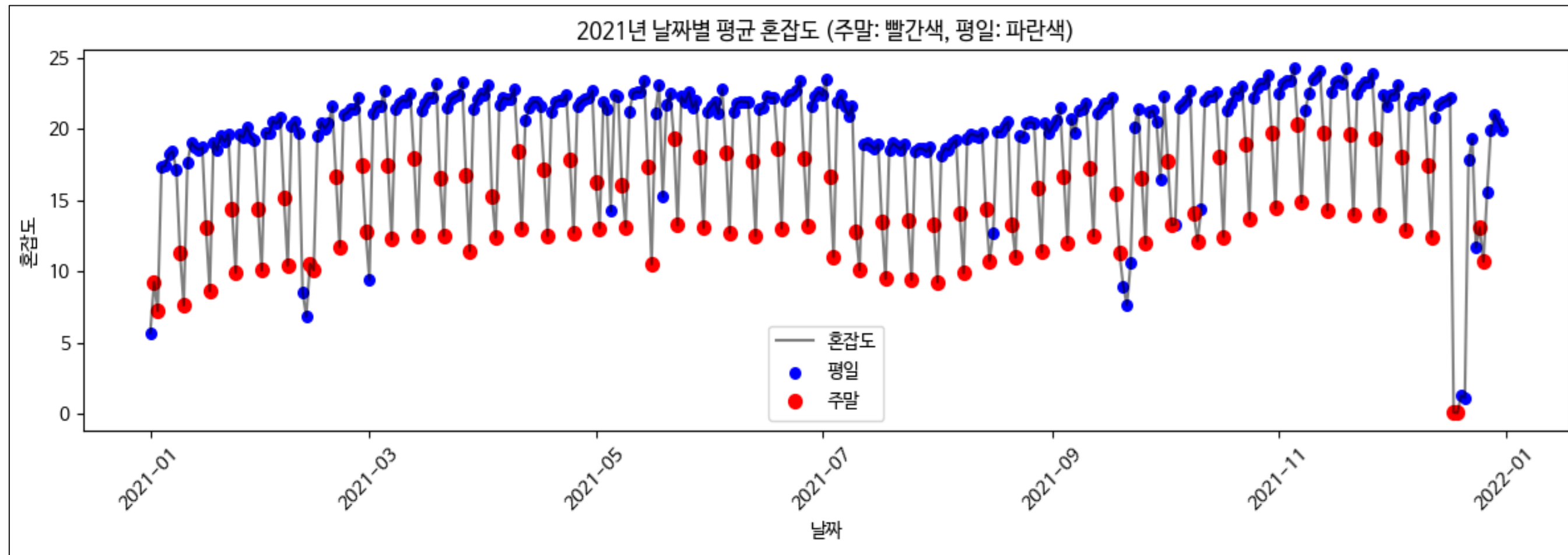


여름방학(6-8월), 겨울방학(12-2월)
시기에 혼잡도가 급격하게 낮아지는 양상

04

데이터 기초특성(EDA)

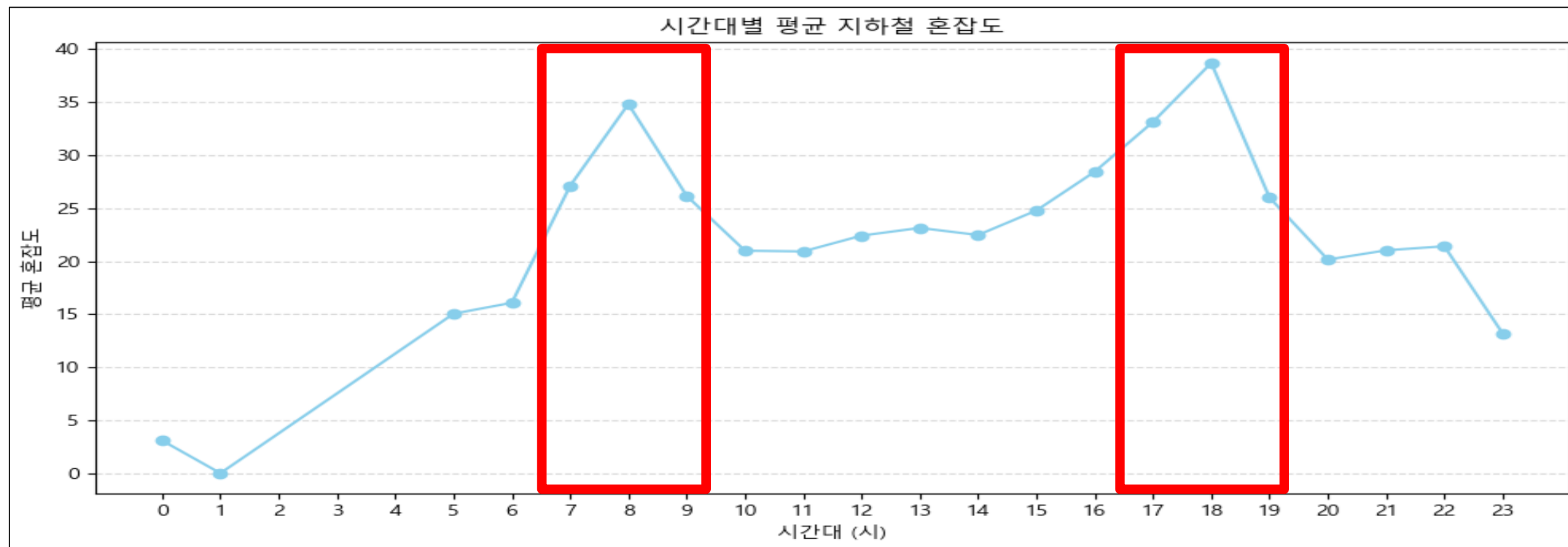
□ 평일/토요일/일요일 혼잡도 비교



04

데이터 기초특성(EDA)

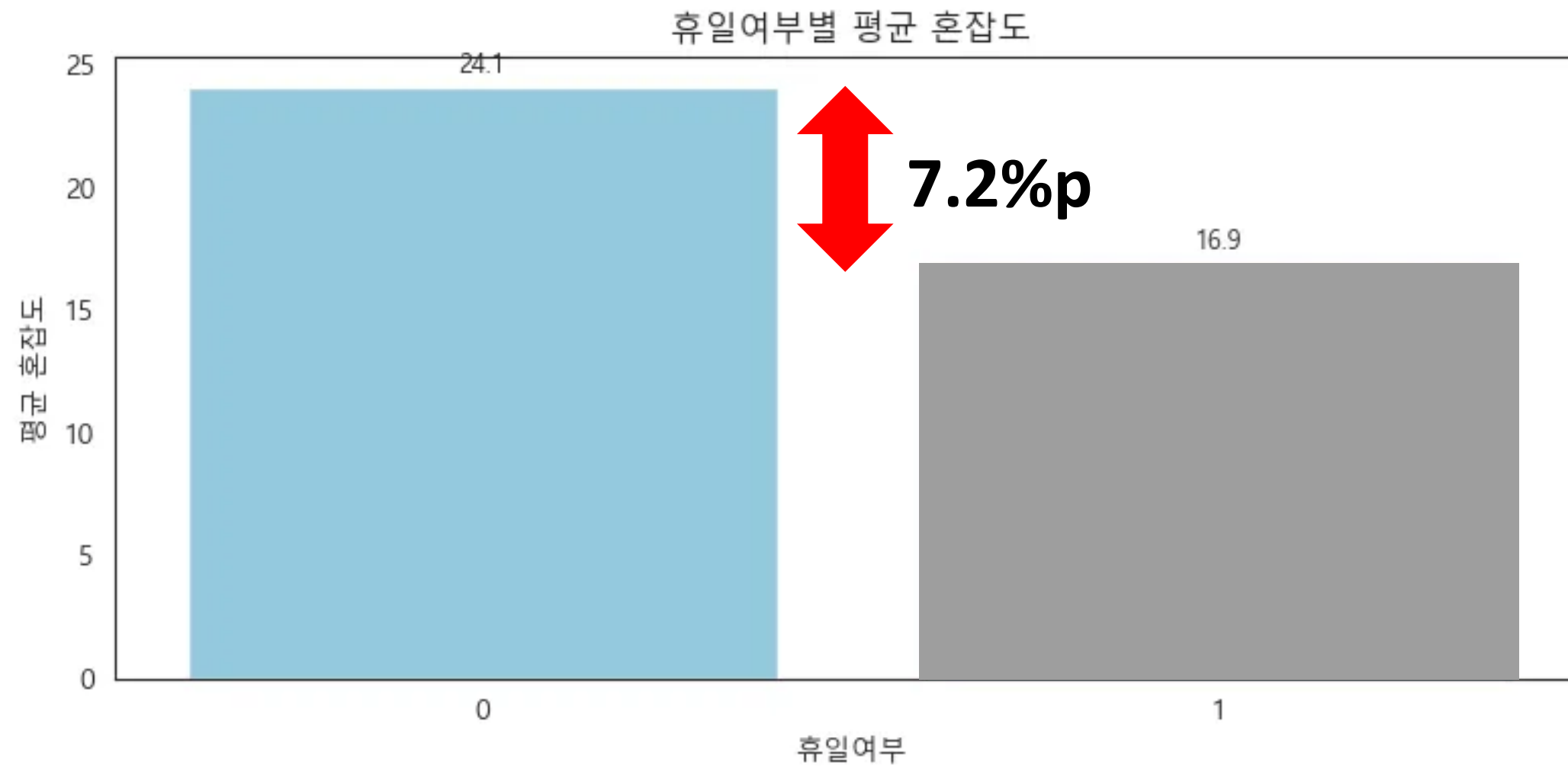
□ 시간대별 평균 혼잡도 비교



통근 수요 집중으로 인해

출근 시간대(7시~9시), 퇴근 시간대(17시~19시) 혼잡도가 가장 높음

□ 휴일여부별 평균 혼잡도 비교

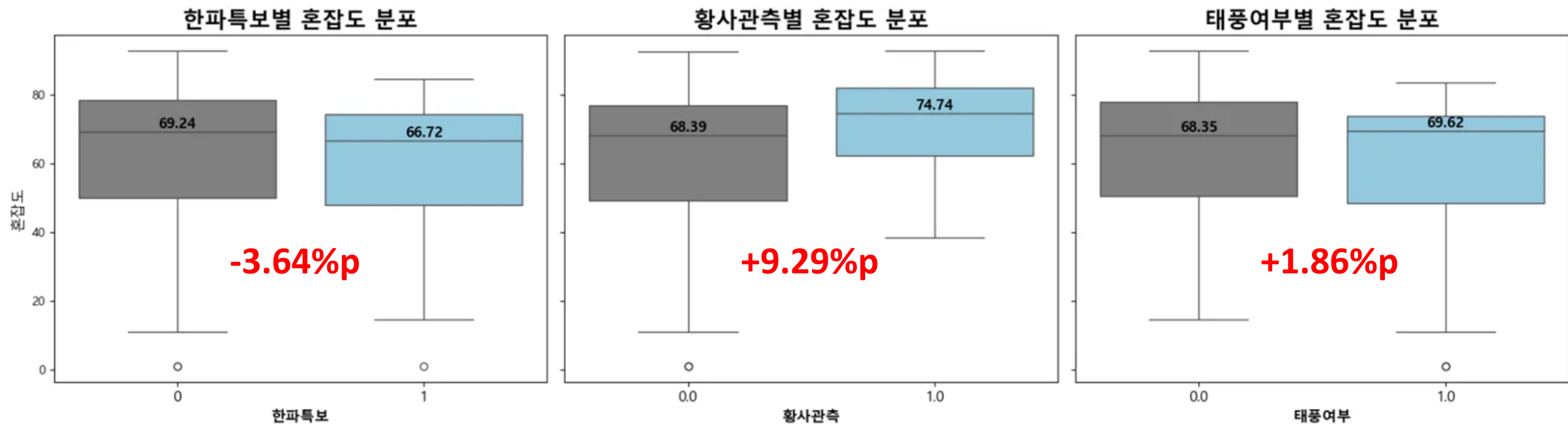


주말/공휴일 혼잡도 < 평일 혼잡도

04

데이터 기초특성(EDA)

□ 기상 이슈별 혼잡도 변화

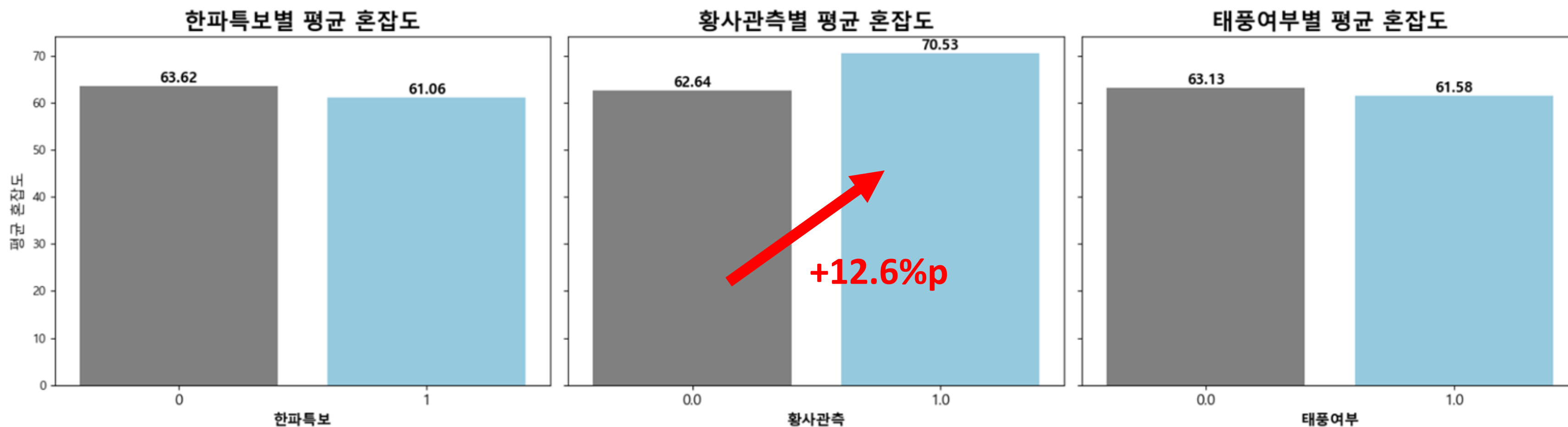


황사 발생 시 혼잡도가 약 9.29% 증가 → 외출이 **지속**되는 경향
한파·태풍은 혼잡도 변화가 크지 않음 → 외출 자제 효과가 **제한**적

04

데이터 기초특성(EDA)

□ 자연 재해별 혼잡도 변화



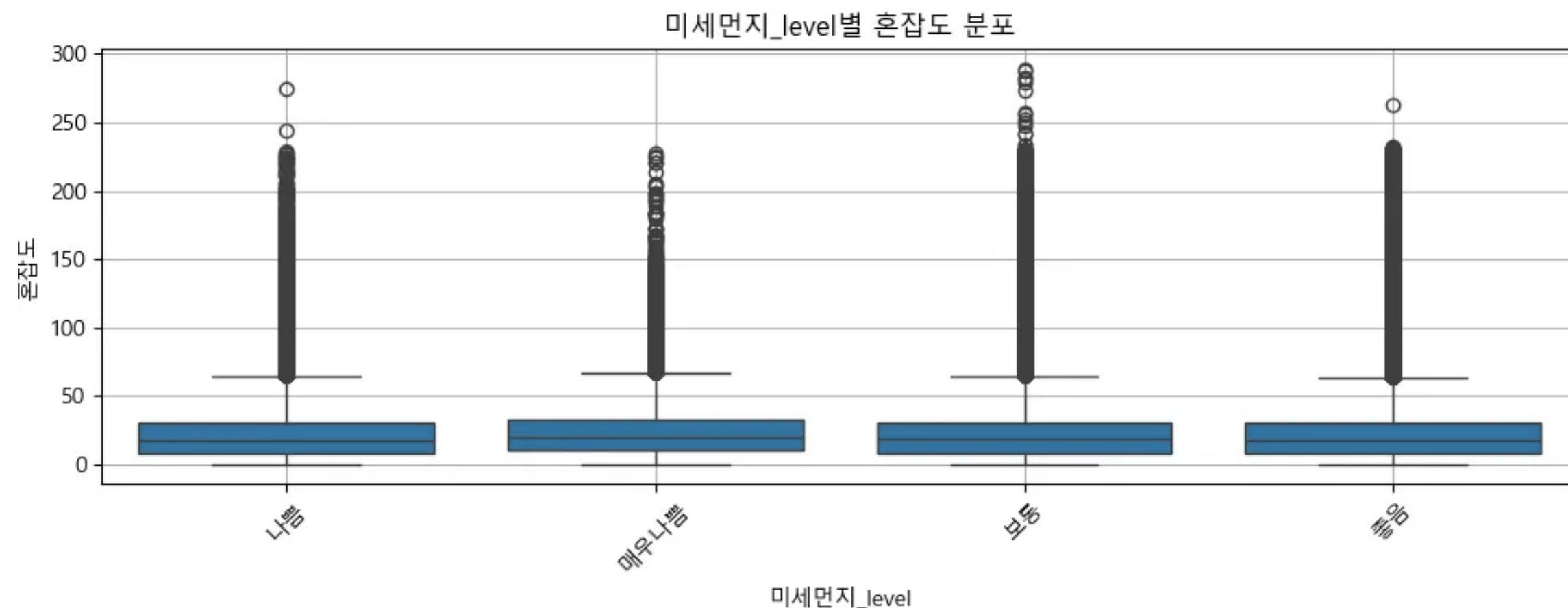
한파·태풍 발생 시 혼잡도는 감소하지만, 황사 발생 시 증가하는 경향을 보임
황사는 혼잡도와 양의 상관관계를 나타냄

04

데이터 기초특성(EDA)

미세먼지와 혼잡도의 상관관계

<미세먼지 & 혼잡도 상관관계 heatmap>

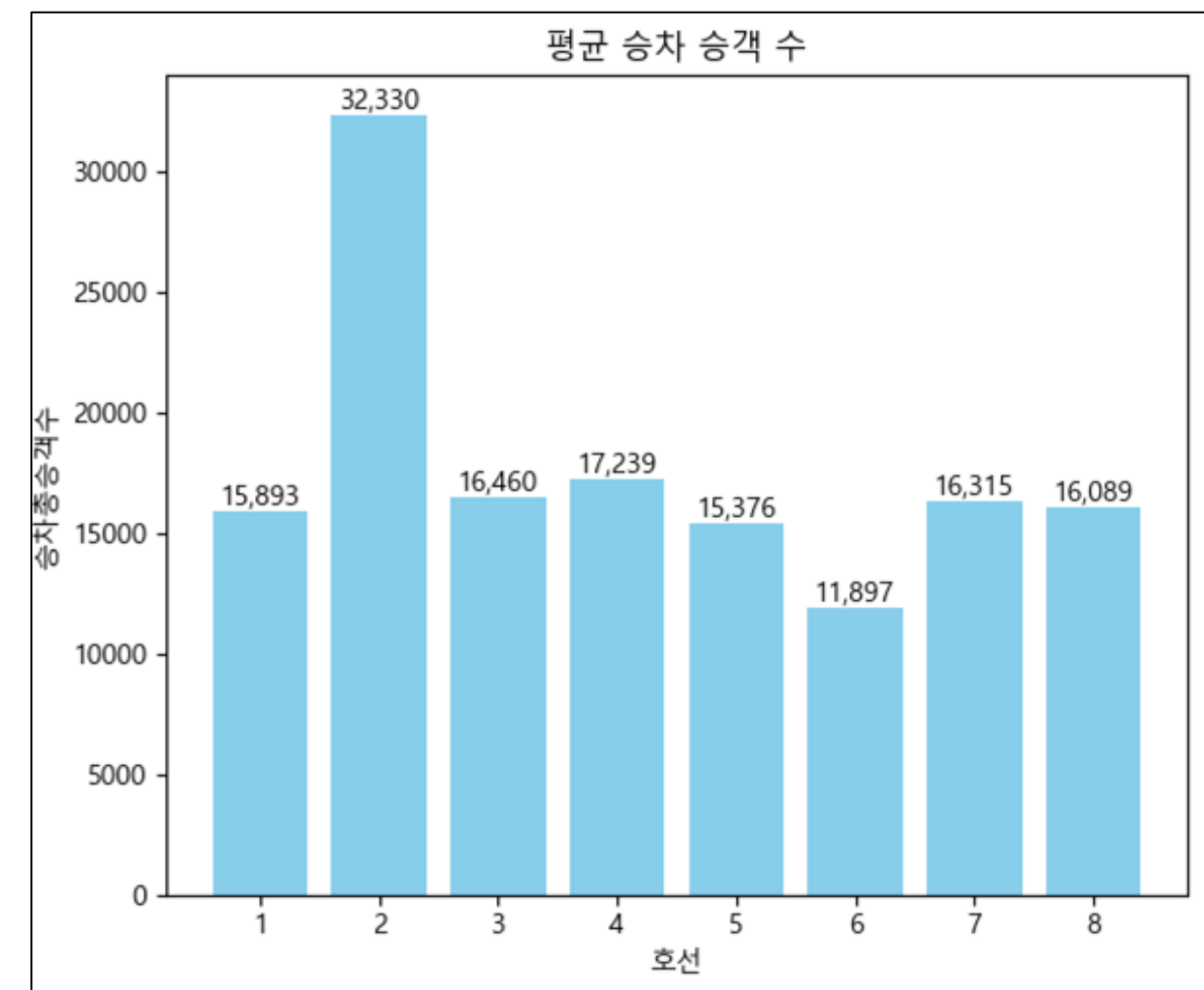
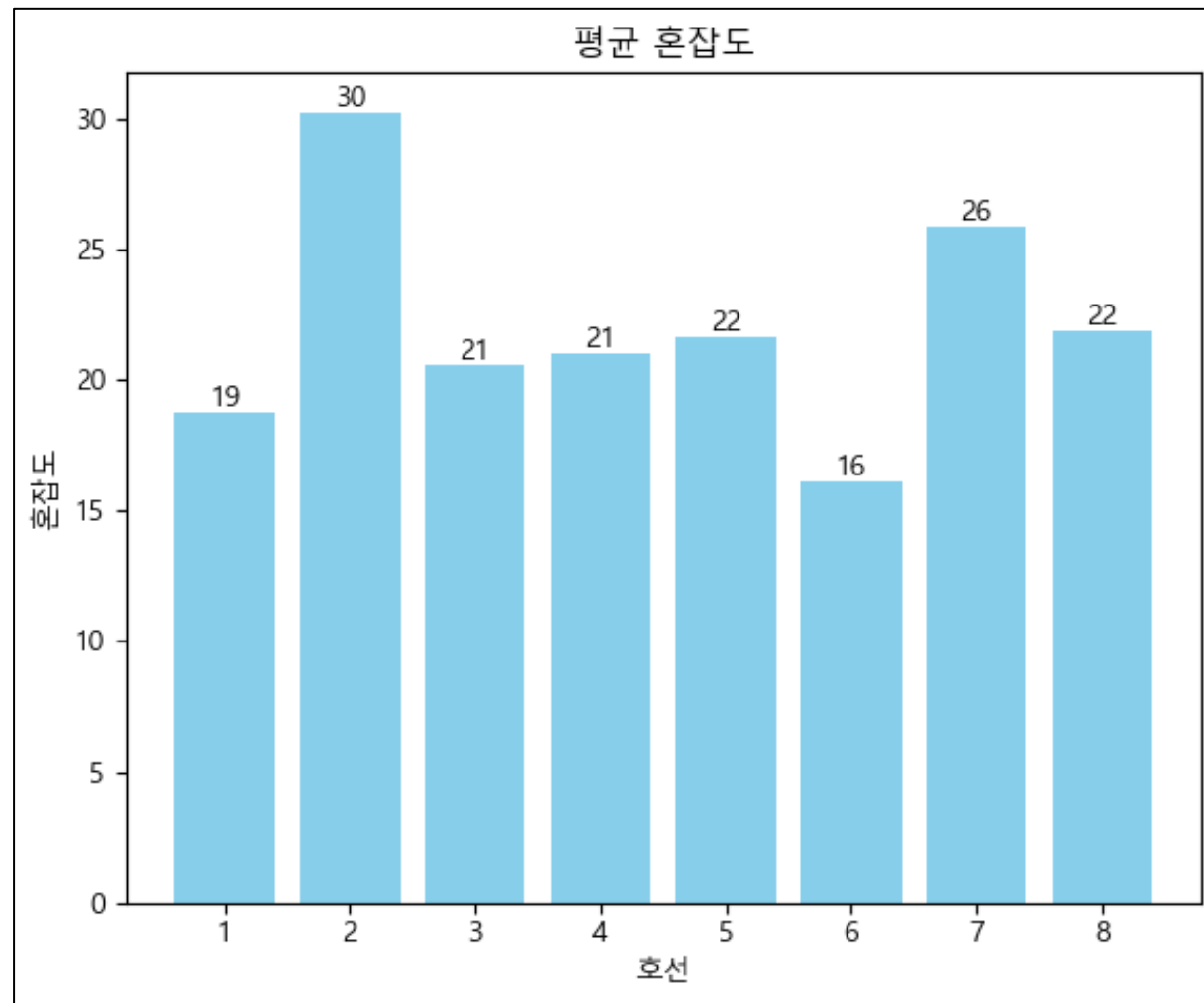


황사와의 연관성을 고려해 미세먼지 데이터 수집
미세먼지와 혼잡도 간 상관계수는 0.02로, 유의미한 **상관관계없음**

O4

데이터 기초특성(EDA)

□ 호선별 혼잡도의 차이



호선별 평균 혼잡도, 승차 승객 수의 차이 존재
→ 호선별/전체호선 2가지로 모델링 진행

1. 데이터 분할

전체 데이터의 80%를 모델 훈련용, 20%를 테스트용으로 분할

- 훈련 데이터 : 2021년 1월 1일 0시 ~ 2023년 7월 24일 23시
- 테스트 데이터 : 2023년 7월 25일 0시 ~ 2023년 12월 31일 23시
- 입력 데이터 x : (정규화된 수치형 변수)
- 타깃 변수 y : (정규화된 혼잡도)

2. 시계열 학습 데이터 구성

- 과거 21시간 연속 데이터를 기반으로 이후 1시간 혼잡도 예측
- Keras의 `TimeseriesGenerator` 활용하여 LSTM 입력 형태 생성

4. 학습 설정 및 콜백

- Optimizer : Adam(학습률=0.001)
- Loss function: 평균제곱오차(MSE)
- Callback
 - a. ModelCheckpoint : 최적 모델 가중치 저장
 - b. EarlyStopping : 검증 성능 3회 이상 개선 없으면 조기 종료(오차 그래프로 추가 학습 결정)
 - c. ReduceLROnPlateau : 검증 성능 정체 시 학습률 감소

5. 학습 실행

- 기본 epoch 설정 : 50
- 검증 데이터 성능 기준으로 학습 횟수 조절

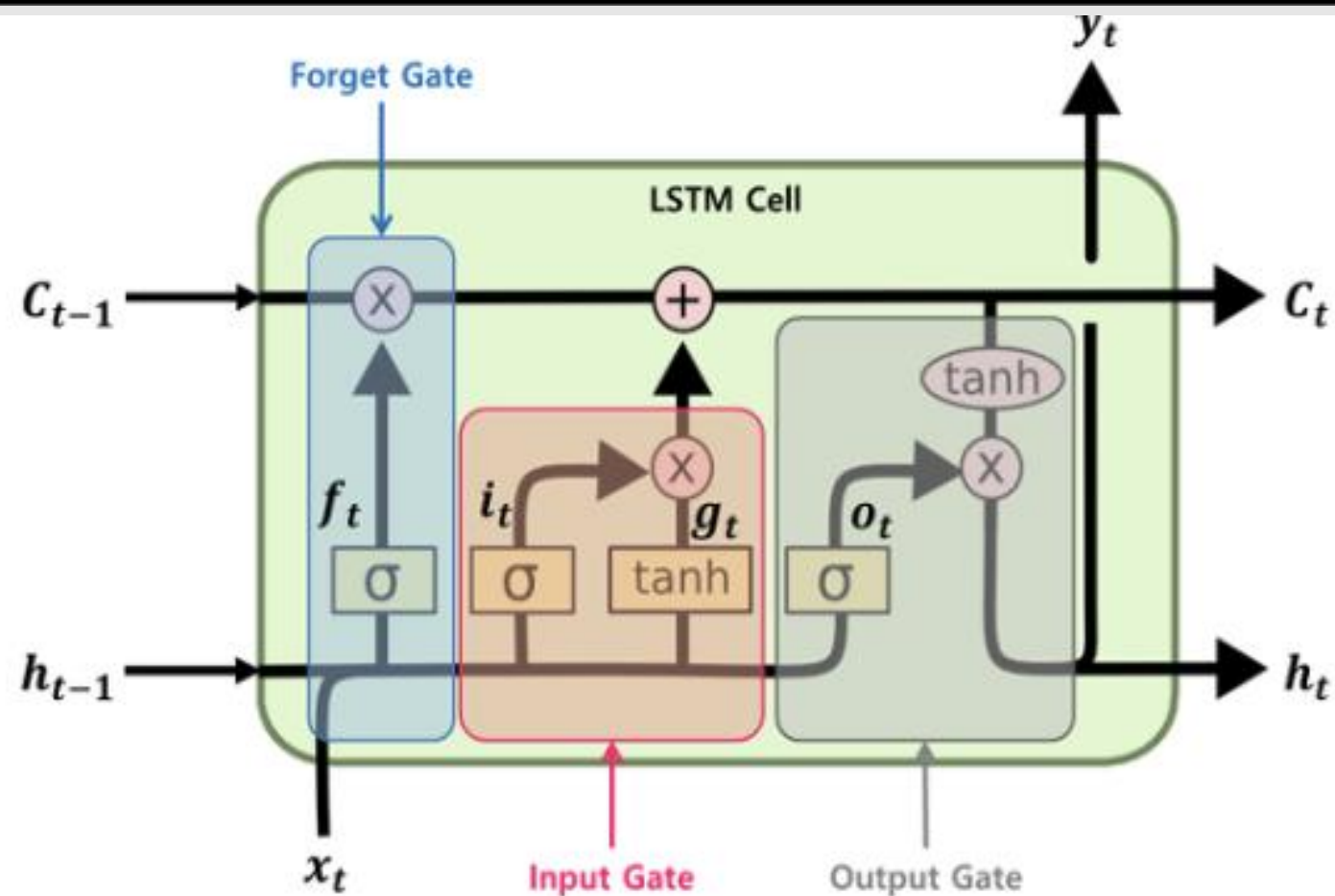
04

분석 모델 개요

- 게이트라는 매커니즘을 이용해 RNN의 장기 의존성 문제를 완화한 LSTM, GRU 모델 사용

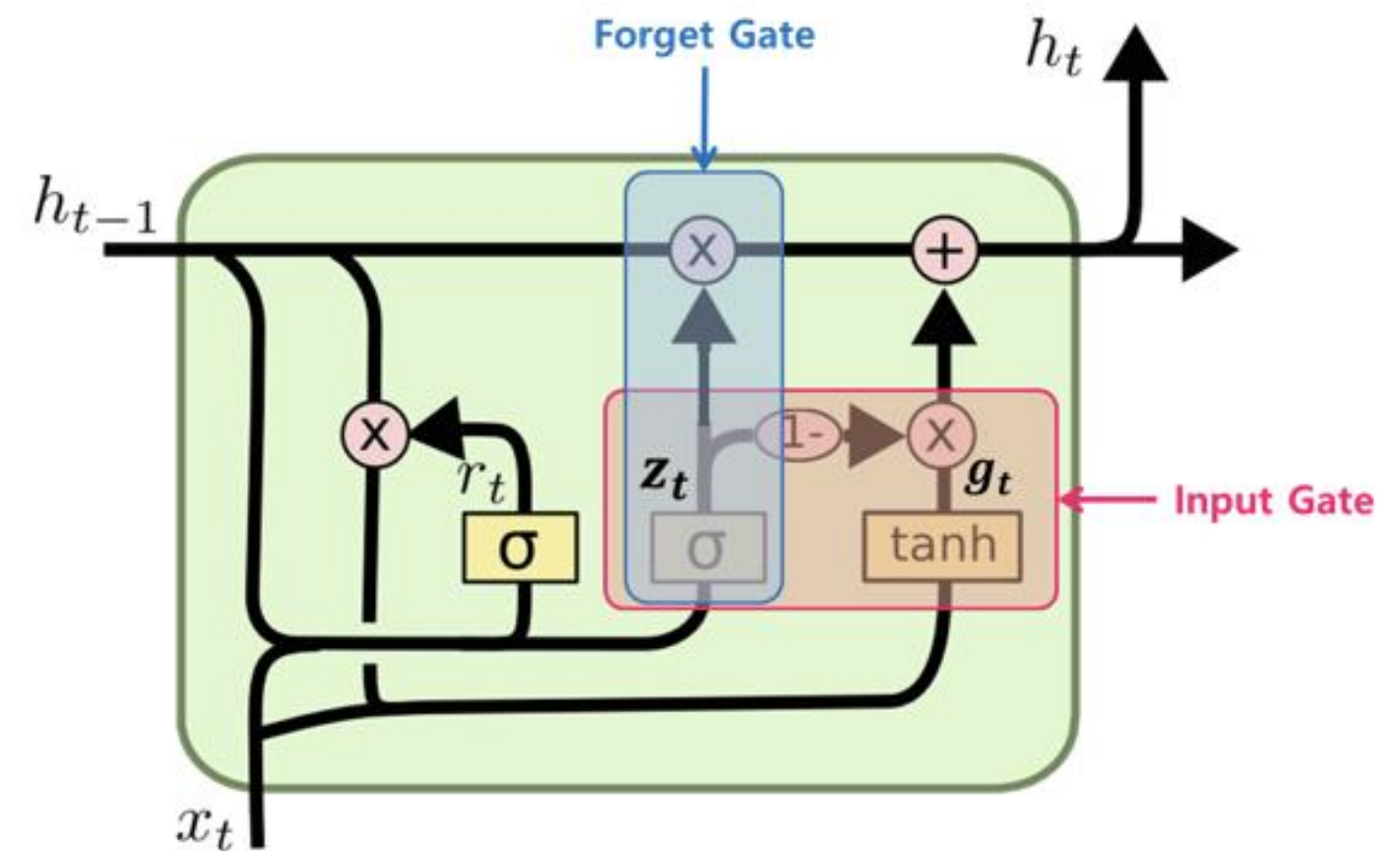
LSTM

- 3개의 게이트 입력, 망각, 출력 사용
- 중요한 정보는 장기적으로 기억할 수 있는 구조**
- 장기 의존성을 더욱 정교한 정보 조절 가능



GRU

- 2개의 게이트 리셋, 업데이트만 사용
- 매개변수의 양이 적고, 연산량이 적어 학습 속도가 빠르다고 알려짐
- 장기 의존성을 단순하지만 효과적으로 처리



04

베이스 모델 성능 비교

GRU

vs



LSTM

배치사이즈	RMSE	결정계수	Epoch
32	4.5340	0.9633	21
64	4.9402	0.9565	16
128	4.6978	0.9606	40
256	4.6810	0.9609	30

배치사이즈	RMSE	결정계수	Epoch
32	4.0428	0.9709	23
64	4.6141	0.9620	15
128	4.8596	0.9579	19
256	4.6932	0.9607	29

Input → LSTM/GRU(64) → Dropout → LSTM/GRU(32) → Dropout → Output

04

LSTM 모델 구조 결정 과정

□ 전체 호선 학습 모델 vs 전이학습 모델 vs 호선별 모델 학습

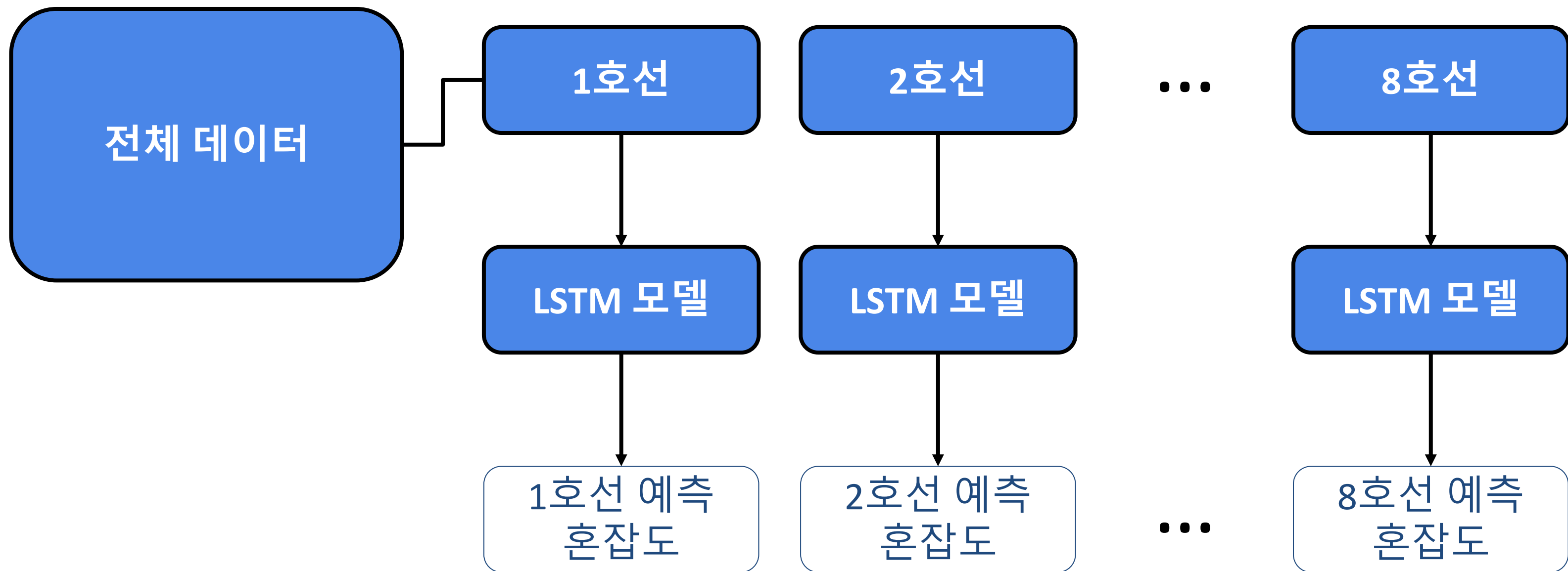
구조	RMSE	결정계수
전체 호선 학습	12.1630	0.6589
전체 호선 학습에 대해 호선별 순서대로 전이학습	13.5906	0.5741
호선별 모델 학습	3.14 ~ 4.93	0.9646 ~ 0.978

호선별 분석을 사용하였을 때, 전체 호선 학습한 것보다 결과가 좋다...?

04

분석전략

지하철 노선마다 이용 목적의 차이
→ 호선별 특성을 반영한 모델 구축을 통해 **예측 정확도 개선**



04

LSTM 모델 구조 결정 과정

□ 모델 구조 튜닝

구조	RMSE	결정계수
단방향 LSTM	4.2428	0.09679
양방향 LSTM	4.3729	0.9649

구조	RMSE	결정계수
LSTM 레이어 2단	4.2428	0.09679
LSTM 레이어 3단	5.0555	0.9544



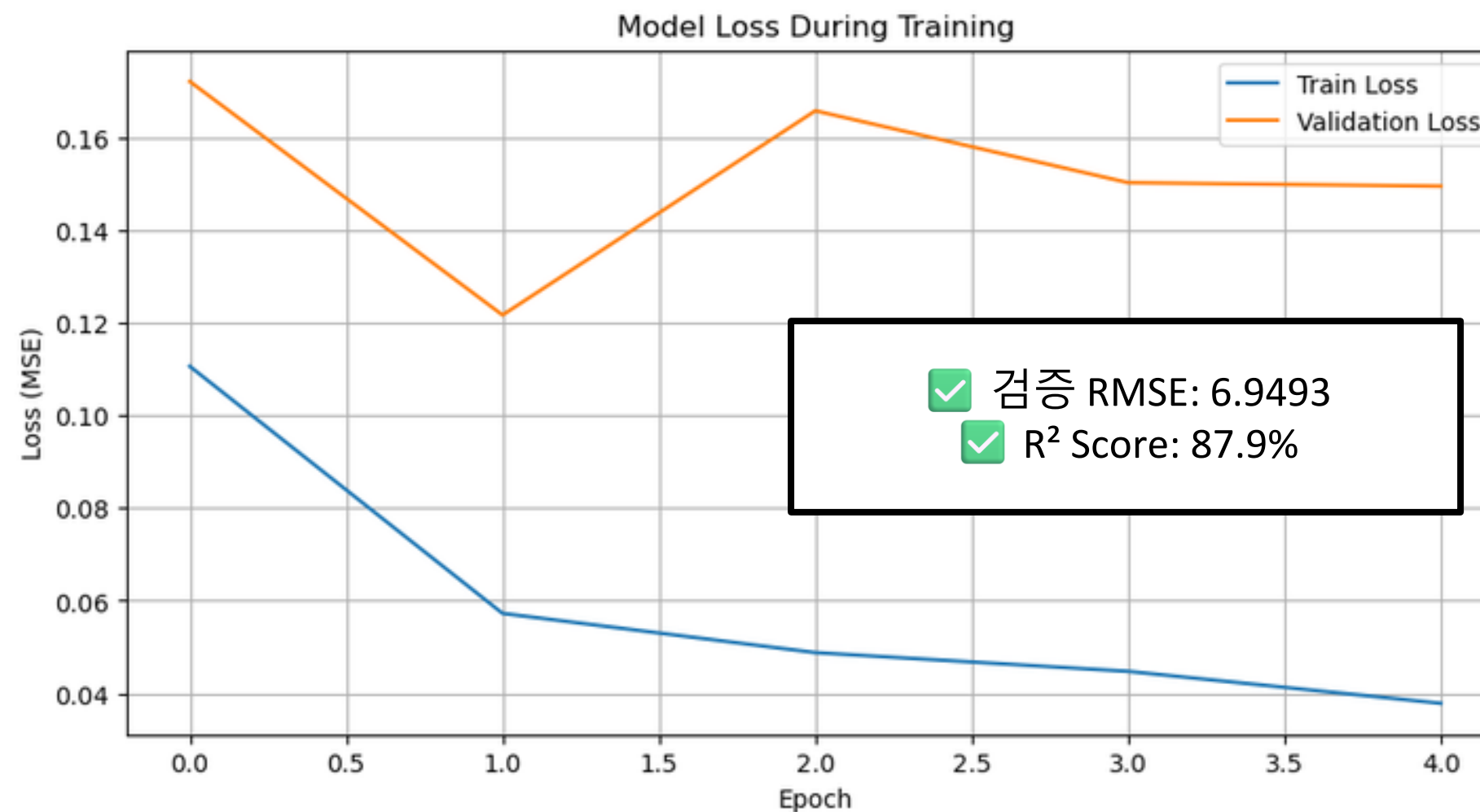
최종 LSTM 구조:

2층 단방향 LSTM

04

LSTM 모델 성능 최적화

□ 4호선 모델 성능 최적화

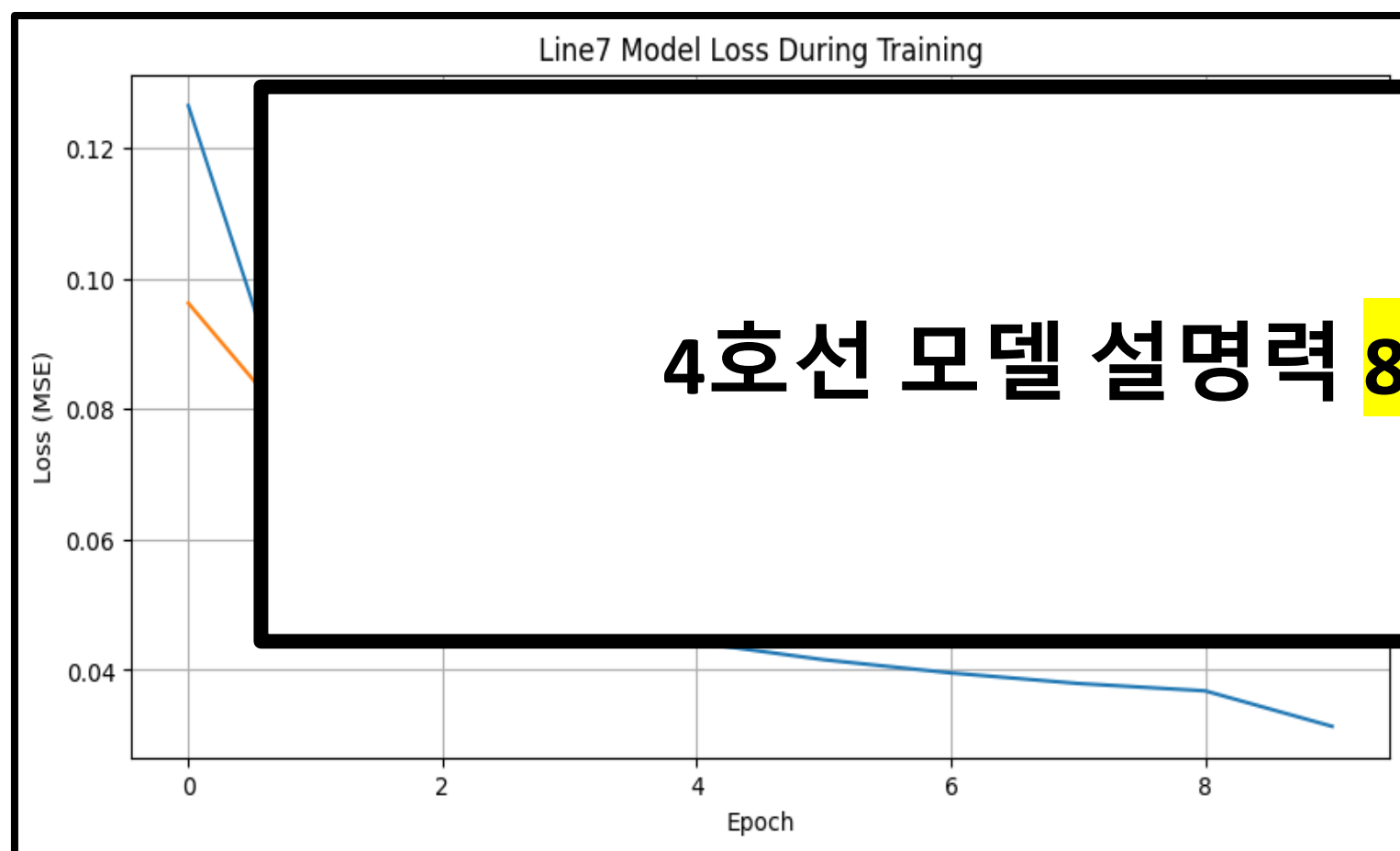


모든 호선 동일한 구조의 모델을 사용했음에도 **4호선은 과적합·성능부족** 문제 발생

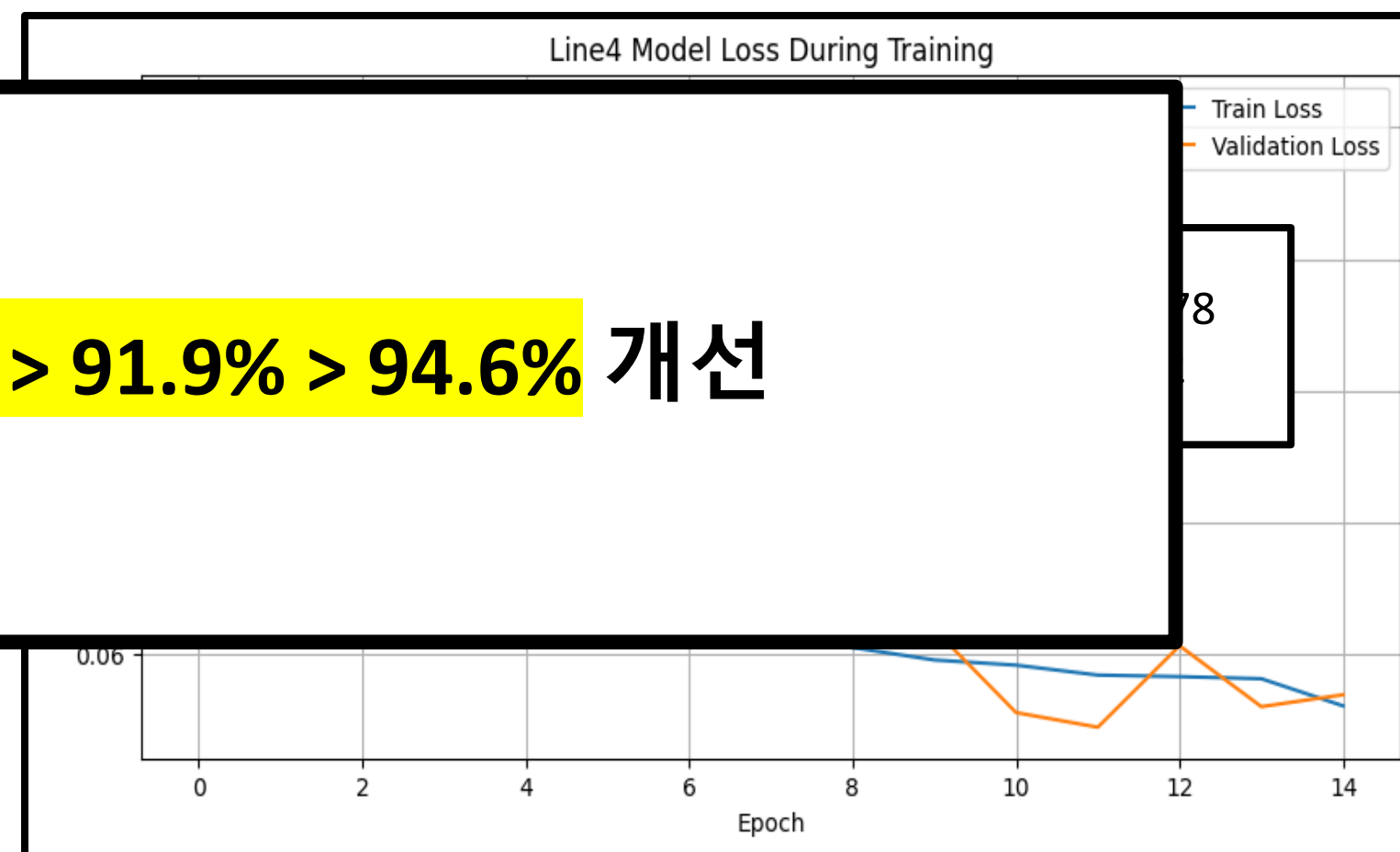
04

LSTM 모델 성능 최적화

□ 4호선 모델 성능 최적화



10회 학습시, 추가 개선 가능성 확인 됨



25회까지 추가 학습 진행

04

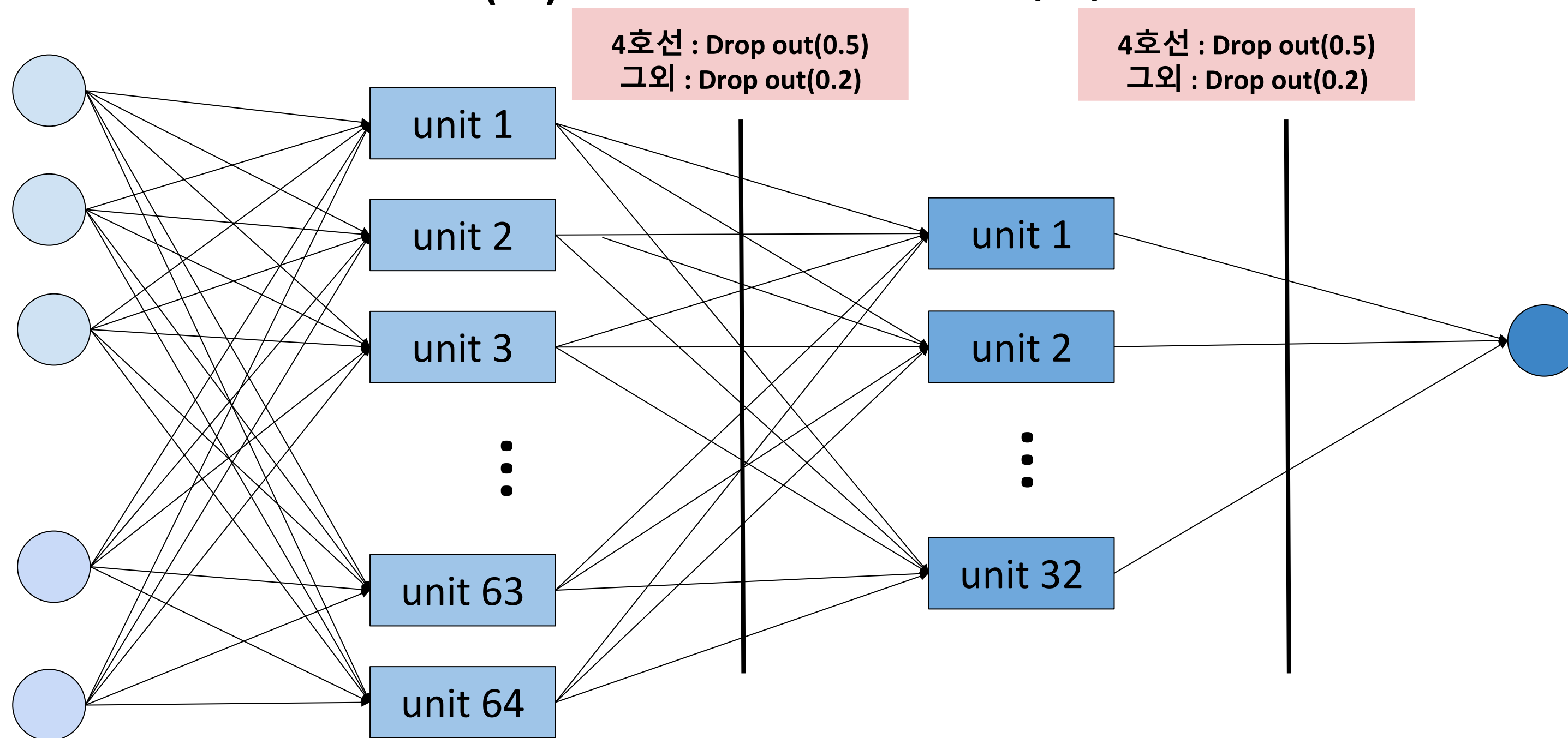
최종 모델 구조

Input layer

LSTM(64)

LSTM(32)

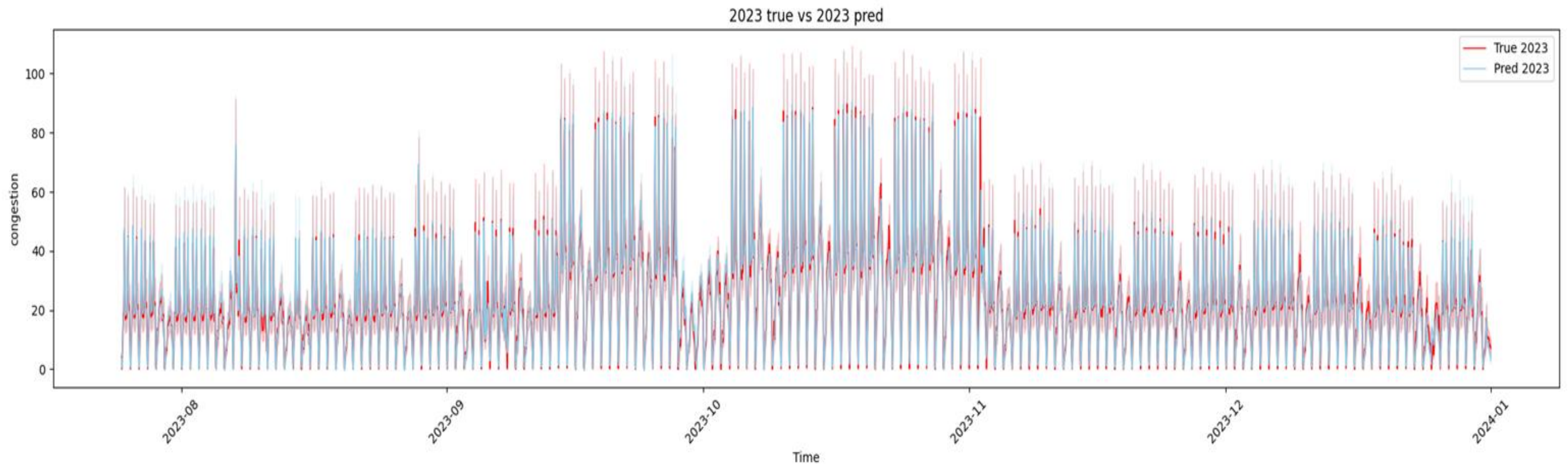
Output layer



04

8호선 혼잡도 예측 시각화

- 빨간색 : 예측
- 파란색 : 실제
- 기간 : 2023/7/25 00:00:00 ~ 2023/12/31 23:00:00 (1시간 간격)



04

호선별 최종 모델 성능

모델 설명력 **94.6% ~ 97.8%**, RMSE **3.14 ~ 4.93** 도달

호선	배치사이즈	드롭아웃	RMSE	결정계수
1	128	0.2	4.9396	0.9558
2	32	0.2	4.2428	0.9679
3	128	0.2	3.3828	0.9506
4	32	0.5	4.6478	0.9462
5	32	0.2	3.9438	0.9584
6	128	0.2	4.6183	0.9617
7	128	0.2	4.6183	0.9617
8	32	0.2	3.1449	0.9776

혼잡도 = 지하철 광고 효과를 측정할 객관적 평가 지표

기상 정보를 반영한 지하철 혼잡도를 예상하여 옥외광고 노출도 정량적 추정

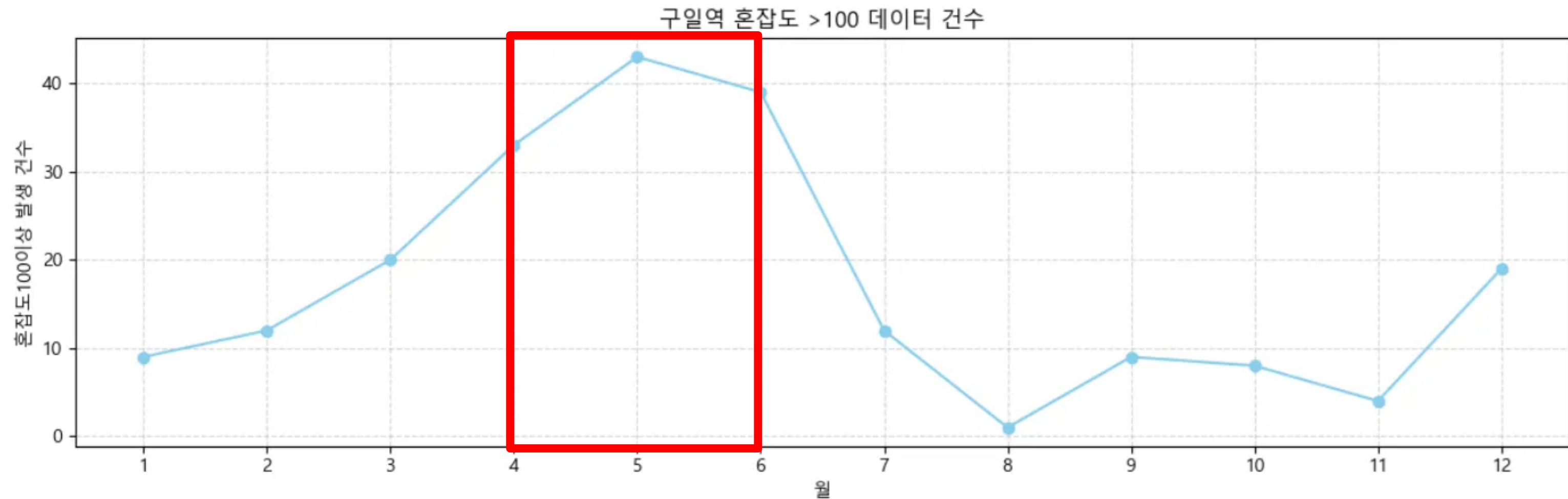
예측 모델을 통해 적절한 광고 장소 및 시기 선정

예측되는 지하철 혼잡도에 따라 월별 광고비 차등화로 광고효과 극대화

04

비즈니스 활용

야구팬 타킷



구일역은 야구 구장 “고척돔”과 대학가 위치
 평일 22시 이후, 혼잡도가 100이상으로 야구 직관 후, 귀가하는 것으로 추정



야구관련 상품을 혼잡도가 높은
4~6월 1호선 구일역에 홍보 진행

광고를 게시할 기간의 혼잡도를 예측하여
정량적인 광고 노출도 추정

프로젝트 결과물에 대한 완성도 평가

기상과 지하철 혼잡도 데이터를
이용한 예측 모델이 안정적으로 구현됨

추후 개선점이나 보완할 점

추가적인 데이터(승객의 연령, 성별 등)를 통해
광고 시점·장소 기획의 정밀도 향상 필요함

<잘한 점>

1. 모델 성능 개선을 위해 다양한 방법을 시도함
2. 옥외광고 효과를 정량화할 수 있는 모델이 없는 상황에서, 이를 직접 설계한 점은 실용성이 큼

<아쉬운 점>

연령대 정보가 있었다면, 타깃팅의 정밀도를 높일 수 있었음

느낀 점이나 경험한 성과

1. 기상 이슈가 혼잡도에 영향을 준다는 점을 확인
2. 예측 모델의 광고 전략 활용 가능성을 확인함

감사합니다!



모델 관련 용어 정리

- 망각게이트 : 이전 셀 상태 중 어느 정보(특징)를 버릴지 결정
입력 게이트 : 새로운 후보 정보 중 어느 만큼을 셀 상태에 추가할지 결정
- 출력 게이트 : 다음 은닉 상태로 셀 상태 중 어느 정보를 꺼낼지 결정
- 업데이트 게이트 : 과거 정보를 얼마나 유지할지 결정
- 리셋 게이트 : 과거 정보로부터 얼마나 잊을지(리셋할지) 결정
- **LSTM**: 입력·망각·출력 3개 게이트 + 별도의 셀 상태
- **GRU**: 업데이트(입력+망각 기능 통합)·리셋 2개 게이트, 셀/은닉 상태 통합
- GRU가 구조가 단순, 파라미터 수가 적음, 작은 데이터나 빠른 학습이 필요할 때 유리
- LSTM은 복잡한 기억·망각 제어가 가능해 더 긴 시퀀스 의존성을 학습하기 좋음
- 선형보간법을 적용한 근거는?
 - 일별 데이터가 약 15000건으로 결측치가 연속하는 최대 구간 길이가 1436행으로 2시간에 해당하는 데이터.
 - 결측치가 연속으로 존재하더라도 2시간에 불과한 데이터로 선형보간법으로 대체하였음

호선별 특징

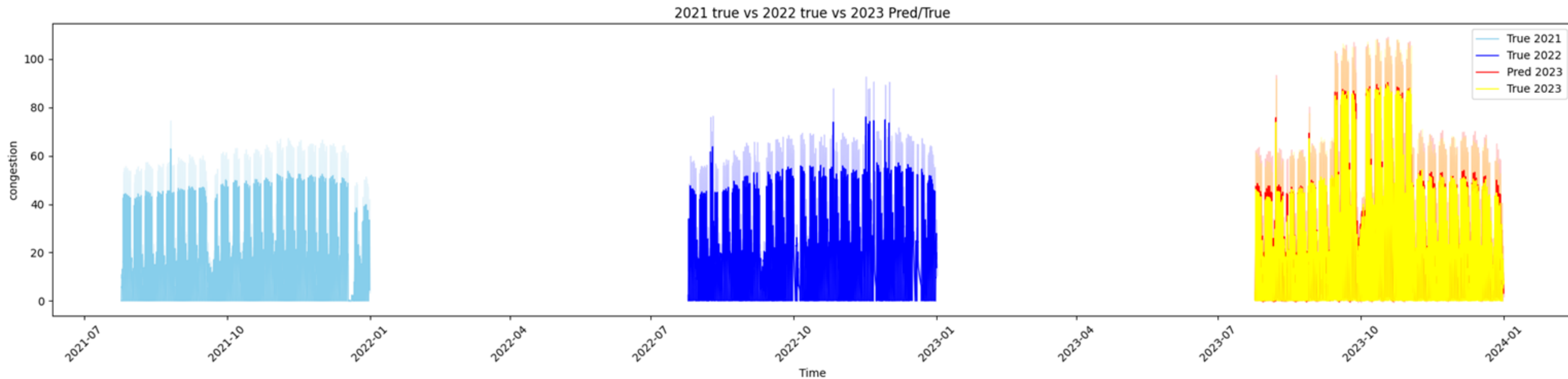
- 1호선 : 수도권 연결 노선, 출퇴근 많이 이용
- 2호선 : 순환노선 + 대학교밀집 + 서울 주요 업무지구
- 3호선 : 강남, 경복궁, 일산 연결 핵심 노선 -> 업무 + 관광 목적
- 4호선 :
- 5호선 : 김포공항~강동 구간, 공항/출근용으로 이용
- 6호선 : 홍대, 망원 등 문화/생활 동선
- 7호선 : 동서 연결 + 부천 ~ 강남 목적지로 많이 이용
- 8호선 : 송파 암사 천호 등 주거지 연결

□ 8호선 7/25 00시 ~ 12/31 23시

2021 실제값 vs 2022 실제값 vs 2023 실제값/2023 예측값

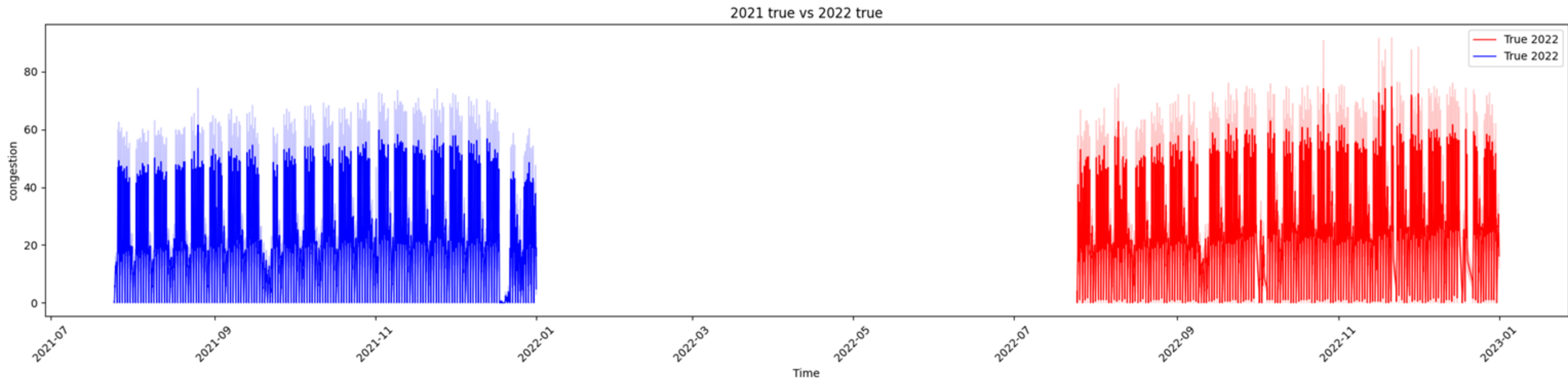
-> 2021, 2022는 유사 패턴

-> 2023은 다른 패턴이 관찰되지만 예측 모델이 정확히 예측



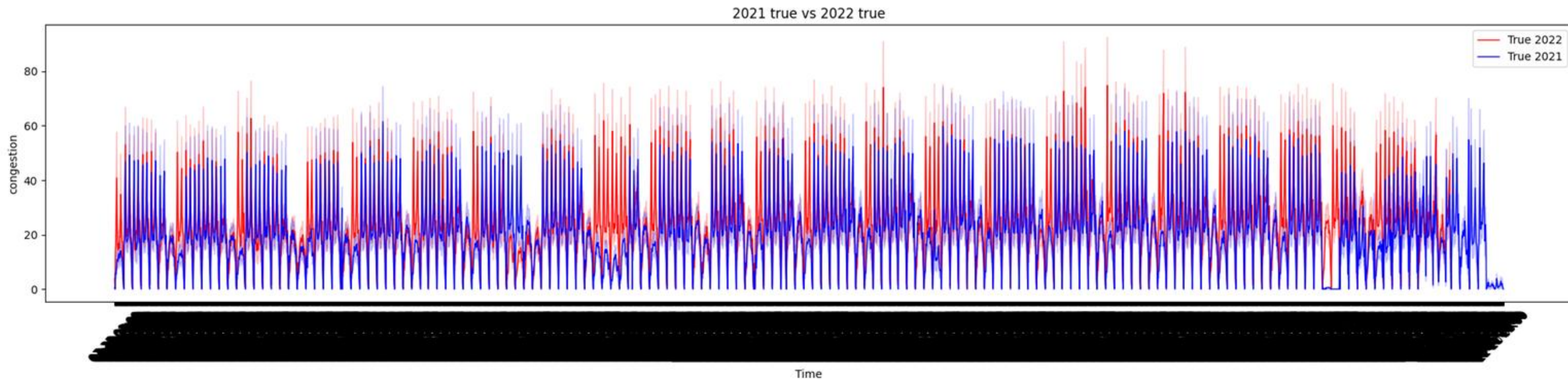
2021 실제값 vs 2022 실제값 -> 유사 패턴

□ 8호선 7/25 00시 ~ 12/31 23시



2021 실제값 vs 2022 실제값 -> 유사 패턴

□ 8호선 7/25 00시 ~ 12/31 23시

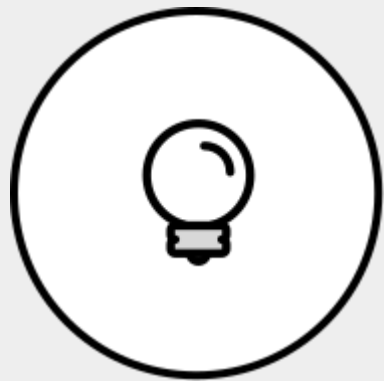


04

비즈니스 활용

[광고 효과와 혼잡도 간의 관계 도출]

STEP 1



광고 설치

구일역 QR 코드 광고판
설치



STEP 2

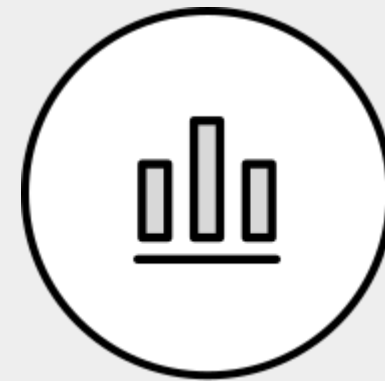


클릭 로그 추적

QR 코드 클릭 수, 시간,
상품 가입 전환율을
추적함



STEP 3

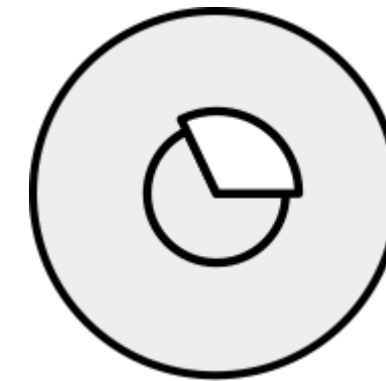


혼잡도 예측

광고 기간 예측 혼잡도
도출



STEP 4



활용

클릭 수가 가장 많은
최적의 혼잡도 구간

야구 시즌 동안 구일역에 QR 코드 광고
→ 가입 클릭 로그 추적

혼잡도+클릭 로그+ 가입 전환율 분석을 통해
최적의 혼잡도 구간을 탐색

= **광고 예산 효율화**

고객들이 지하철 광고에 가장 관심을 가지는 최상의 혼잡도를 찾아내자.

04

비즈니스 활용방안

혼잡도 예측 결과를 바탕으로 전략에 맞는 시간·노선에 광고를 배치

→ iM뱅크의 수도권 가입자 유입

[야구 관련 상품 마케팅]

구일역

- ▶ 고척 야구장 인접: 야구 시즌(4~10월) 동안 유동 인구 증가
- ▶ 야구 관심 고객 유입 多 → 야구 연계 금융 상품 홍보에 적합한 입지

→ 야구 관련 상품 마케팅 장소로 적절할 것으로 예상

참고 문헌

이승연, 이영인. (2021-11-10). 날씨가 대중교통 이용자의 지하철 또는 버스 선택에 미치는 영향 :스마트카드 데이터를 사용하여. 대한교통학회 학술대회지, 제주.

Jenni Baker(2021.4.19). Programmatic DOOH isn't just future ready, it's ready now.The Drum.<https://www.thedrum.com/news/2021/04/19/programmatic-doooh-isn-t-just-future-ready-it-s-ready-now>

김승준. (2016-02-29). 서울시 지하철의 혼잡비용 산정과 정책적 활용방안. 서울연구원 정책리포트 208

박형재. (2024-02-06). 한국공제보험신문.방송 시청률처럼...옥외광고 효과측정지표 만든다.<http://www.kongje.or.kr/news/articleView.html?idxno=3433>