



119 신고 건수 예측을 위한 CatBoost 모델링과 활용방안 제안

기상 데이터

강연주 방지현 서동혁
송수현 장가람 조하늘

Summary

최근 기상 이변으로 인해 119 신고 건수가 급증함에 따라 소방 인프라의 부담이 가중되고 있다. 본 연구는 소방 데이터와 날씨 빅데이터를 융합하여 부산시의 119 신고 건수를 예측하는 모델을 개발하고, 이를 실무에 활용할 방안을 제시한다. 데이터 분석 과정에서 COVID-19, 지역별 지형 특성 등 외부 요인을 반영한 파생 변수를 생성하였고, 여러 트리 기반 모델의 성능을 비교 분석한 결과 CatBoost 모델이 가장 우수한 예측력을 보였다. 특히, 카운트 데이터의 특성을 고려하여 Poisson 손실 함수를 적용했을 때 모델의 성능이 더욱 향상되었다. SHAP 분석을 통해 '상세 주소(sub_address)'와 '자연재해 여부(disaster_day)'가 예측의 핵심 변수임을 확인하였으며, 이를 바탕으로 실시간 예측 및 시각화가 가능한 웹 대시보드를 개발했다. 최종적으로 본 연구는 예측 모델을 활용한 스마트 알림 시스템, 대화형 챗봇, 그리고 효율적인 소방 자원 배치 방안을 제안하여 도시 안전 관리 체계의 디지털 전환에 기여하고자 한다.

Background

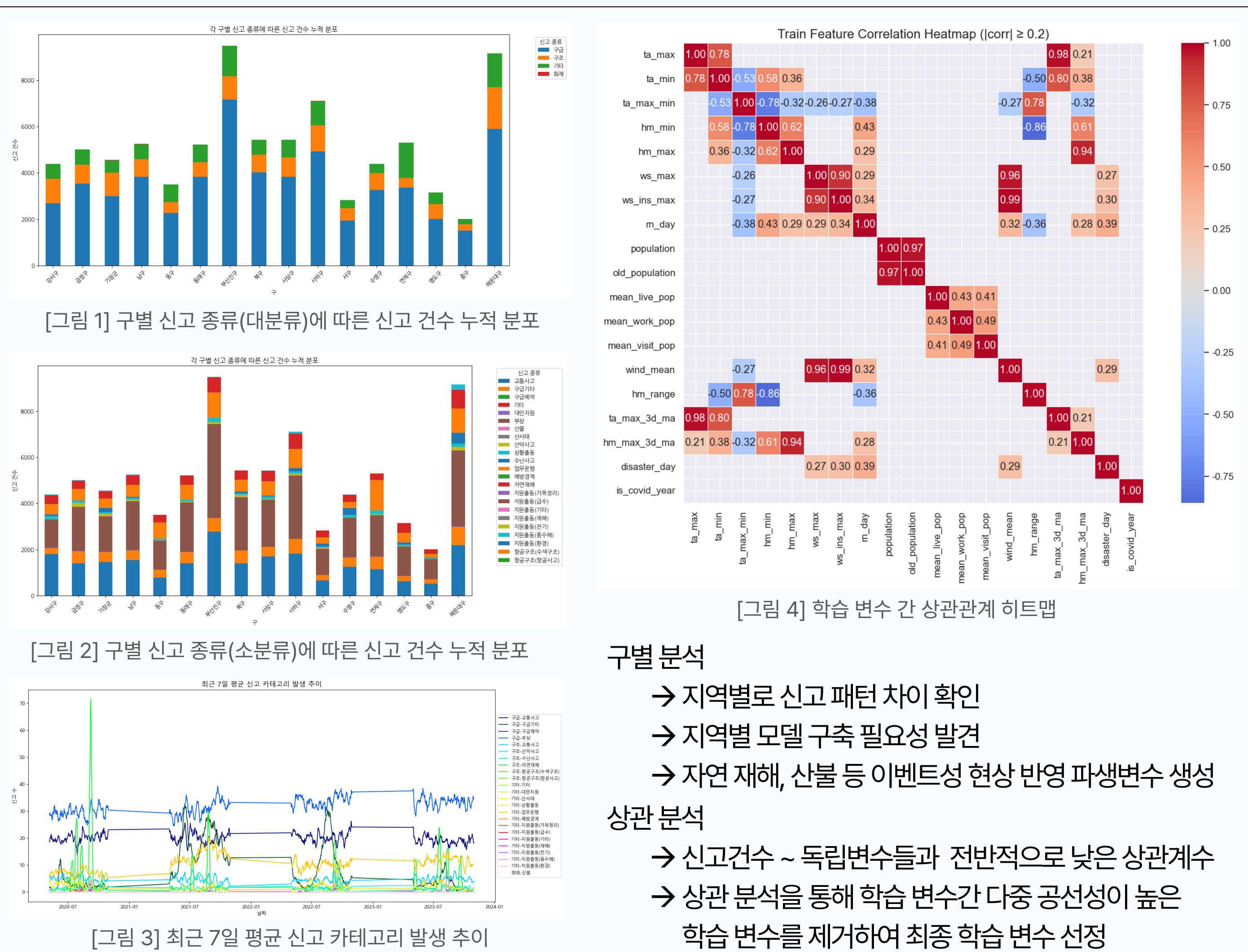
최근 소방청 통계에 따르면 전국 119 신고 건수는 2017년 1,155만여 건에서 2022년 1,255만여 건으로 5년 새 100만 건 이상 늘어났다. 특히 2022년에는 전년 대비 약 47만 건(↑3.9%)이 증가하며 '하루 평균 3만 4천 건, 3초당 1건' 수준에 근접했다. 부산만 보더라도 2023년 신고가 75만 건을 넘어 '42초에 한 번' 콜로 접수되는 등 대도시 소방 인프라가 이미 한계치에 도달하고 있다.

폭염 일수는 과거(1980~2000년) 대비 최근(2001~2023년) 평균적으로 증가하고, 발생 기간도 6~9월로 확대되는 추세이다. 이처럼 기상 이변·기후 급변은 '전력·보건·안전' 전 분야에 부담을 주고 있으며, 소방서비스 역시 예외가 아니다. 따라서 소방데이터와 날씨 빅데이터를 융합한 119 신고 건수 예측 연구의 필요성이 증가하게 되었다.

Pipeline



EDA · Correlation Analysis

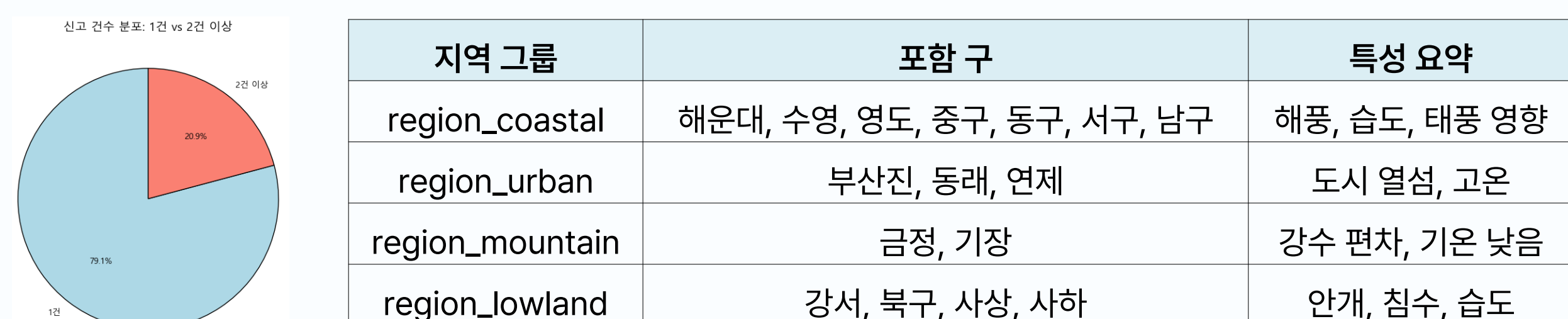


References

유성원, 2024, 2024년 소방청 통계연보, 소방청
이현지, 2024, (보도자료) 부산소방재난본부, 2023년 하루 42초마다 전화벨, 부산광역시 소방재난본부 종합상황실
기상청, 2024, (보도자료) 2024년 여름철 기후특성, 기상청 본청 기후변화감시과

Modeling

클래스 불균형 & 지역 기반 예측 모델



[그림 5] 신고 건수 1건과 2건 이상

[표 1] 지역별 변수 특성

- 신고 1건이 **79.1%**를 차지
→ 모델이 다수 클래스(1건)에 과적합되고, 고신고일 예측력 저하 발생
- 지역별 기상·지형 특성을 반영하기 위해 해안, 도심, 고지대, 저지대 **4개 그룹**으로 분할 학습

트리 기반 모델

모델종류	MAE	RMSE	R2
XGBoost	0.46	1.28	0.05
lightGBM	0.44	1.03	0.38
CatBoost(RMSE)	0.43	0.99	0.42
CatBoost(Poisson)	0.43	0.97	0.45

[표 2] 트리 기반 앙상블 모델 비교

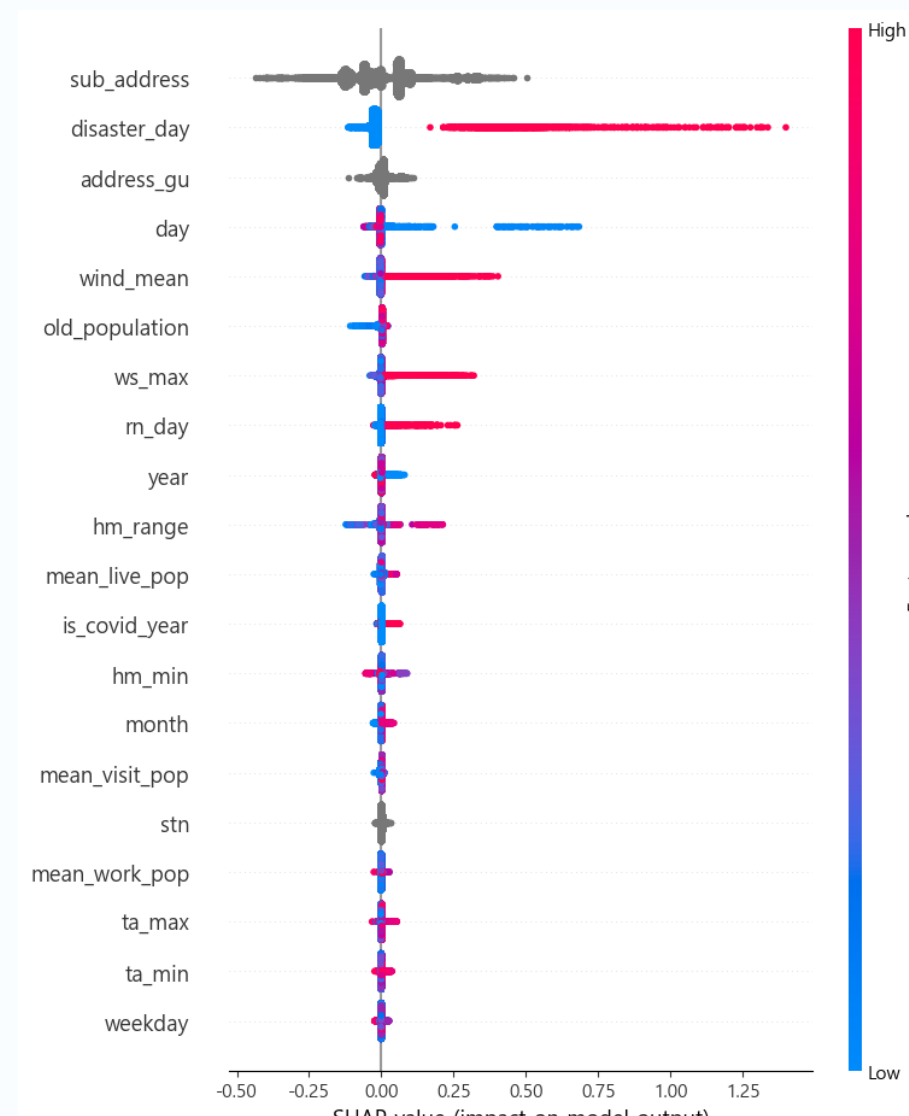
최종 모델 선정

CatBoost + Poisson 손실 함수

- 신고 건수의 0 이상의 카운트 데이터 특성 반영 → 카운트형 예측 문제에 적합한 전략
- RMSE(0.97)·R²(0.45) 개선 → 극단적 오차 억제 및 설명력 향상

Results

1) SHAP 분석



[그림 6] SHAP Value 활용 변수 중요도 분석

변수명	영향도 및 해석
sub_address	지역별 신고 패턴 차이
disaster_day	자연재해 발생 시 신고 급증 경향
wind_mean	강풍이 불면 신고 증가 가능성
ws_max	최대 풍속이 높은 날 신고 급증
rn_day	강수량이 많을수록 신고 증가

[표 3] 변수 중요도

2) 활용 방안

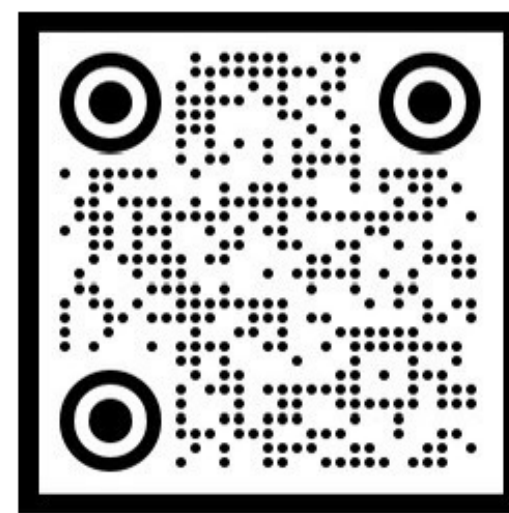
- Streamlit 기반 대시보드로 119 신고 + 기상 데이터 통합 분석
 > 구 단위 상세 조회·예측 모듈
 > 부산시 전체 종합 시각화 모듈
- 정책·재난 대응 지원 및 디지털 안전 관리 플랫폼 확장 가능



[그림 7] Streamlit을 활용한 웹 기반 대시보드 시스템

3) 아이디어 제안

- 119 신고 건수 & 기상 정보를 활용한 스마트 알림 시스템 도입
- LLM 기반 대화형 챗봇 개발
- 위험 예상 지역 기반 인력 배치를 통한 자원 운영 최적화



방법론 보고서

Contribution

Tobig's 22th	강연주 상관분석	Tobig's 22th	송수현 데이터 분석
Tobig's 22th	방지현 데이터 전처리	Tobig's 23th	장가람 데이터 분석 · 시각화
Tobig's 23th	서동혁 데이터 수집 · 분석	Tobig's 22th	조하늘 데이터 분석 · 모델링