



# 자전거 수요량 예측 모델링

AI6 마정흠

# 목차



1. 배경 및 소개
2. EDA
3. 전처리
4. 모델링
5. 결과 및 해석
6. 앞으로 대응

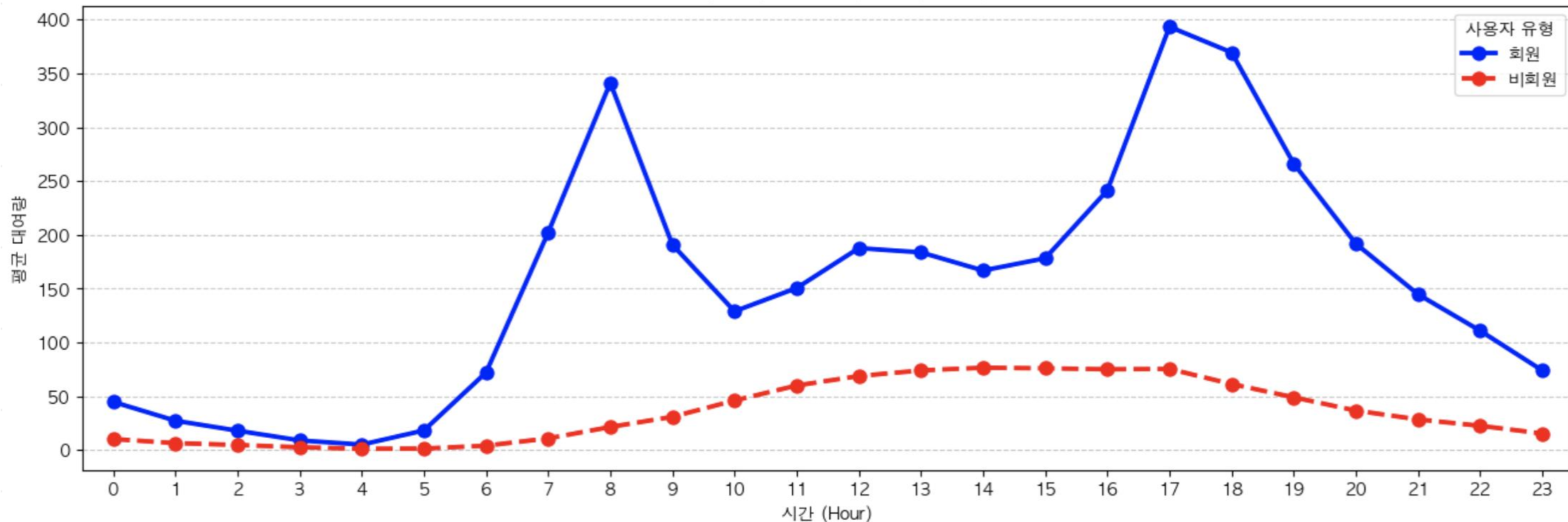
## 배경 및 소개

- 우리는 자전거 대여 시스템의 운영 담당자입니다.
- 목표는 자전거 대여 패턴을 분석하여 자전거 배치 및 운영 전략을 최적화하고, 대여 수요를 정확히 예측하는 것입니다.
- 이를 통해 대여 시스템의 효율성을 높이고 사용자 만족도를 증가시키는 방법을 찾는 것

# EDA

## 시간대별 평균대여량 확인

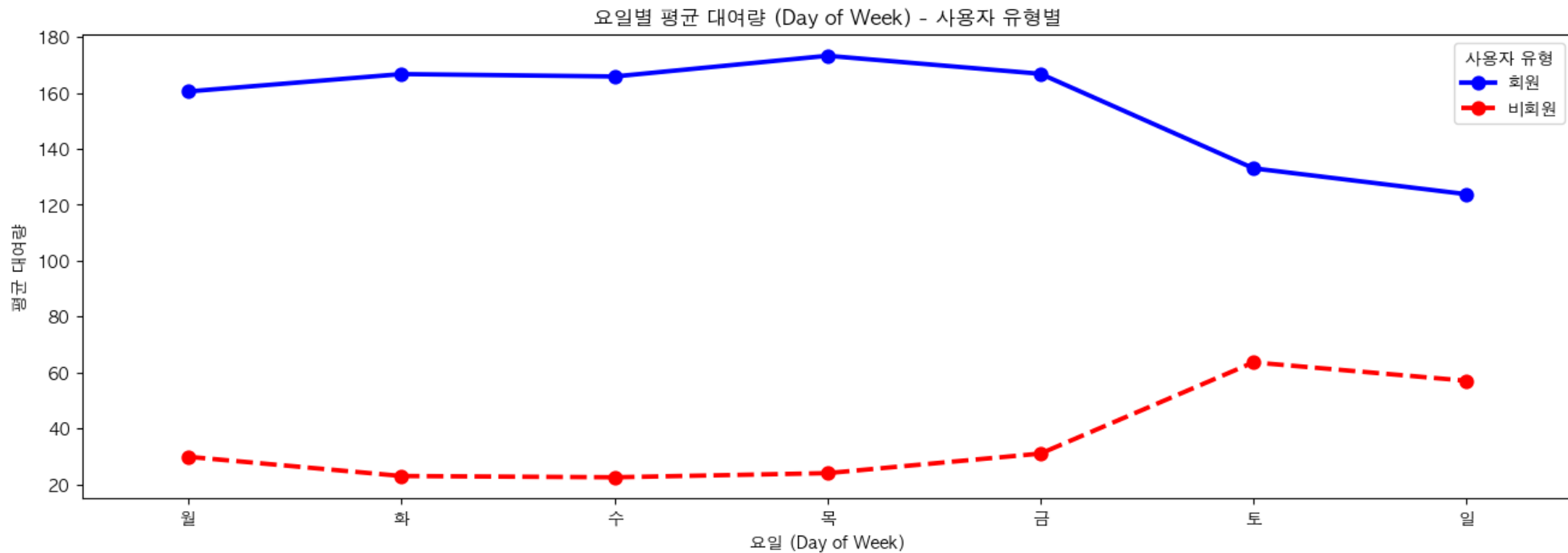
시간대별(Hour) 평균 대여량 - 사용자 유형별



- 오전 8시, 오후 17~18시 쯤 대여량이 급증한다는 것을 확인 가능
- 회사원, 학생들이 출퇴근용으로 사용시 급증하는걸로 예측
- 회원/ 비회원 대여량 차이가 있기에 사용자별 극명한 패턴 존재

# EDA

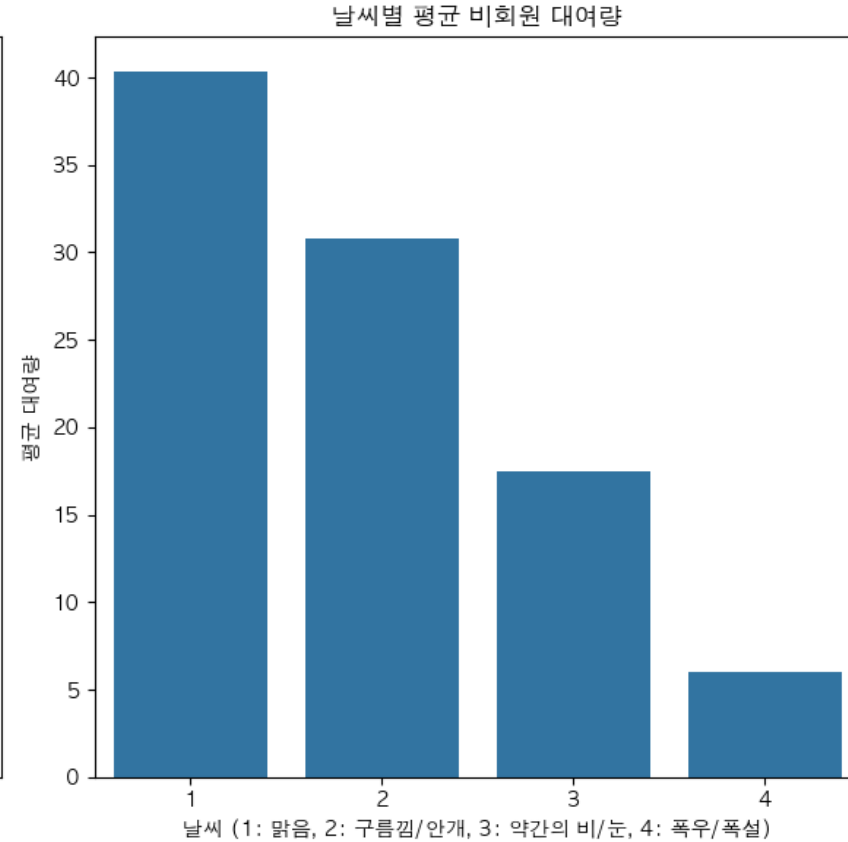
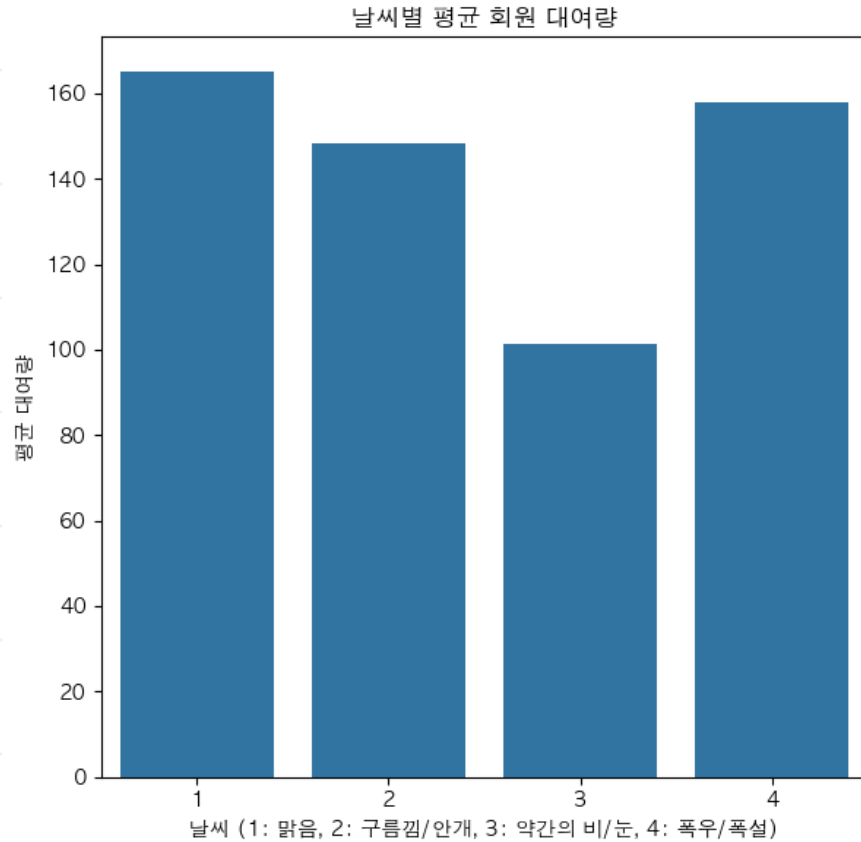
## 요일별 대여량 확인



- 비회원이 주말에 3~4배정도 증가하는 이유는 간단한 여행, 여가활동으로 이용을 많이 하는걸 예측합니다.
- 회원은 회사원, 학생들이 출퇴근용으로 사용시 급증하는걸로 예측
- 평일과 주말 데이터를 나눠 데이터를 모델링 하는건 긍정적인 요소

# EDA

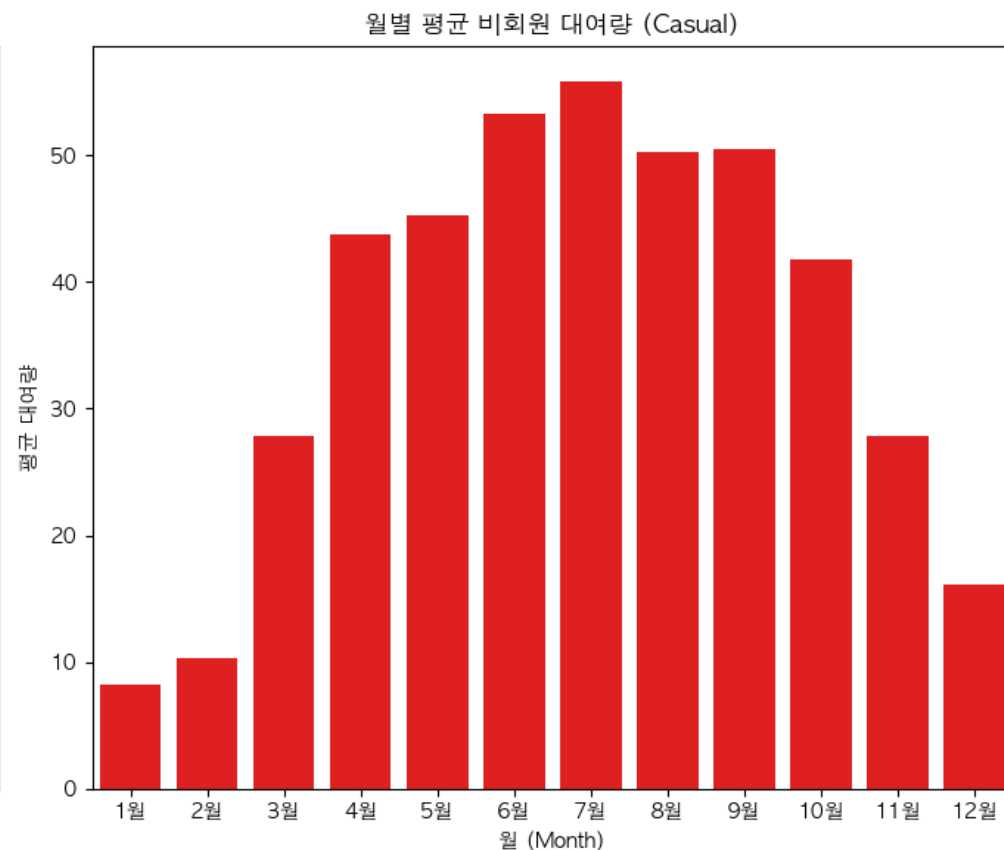
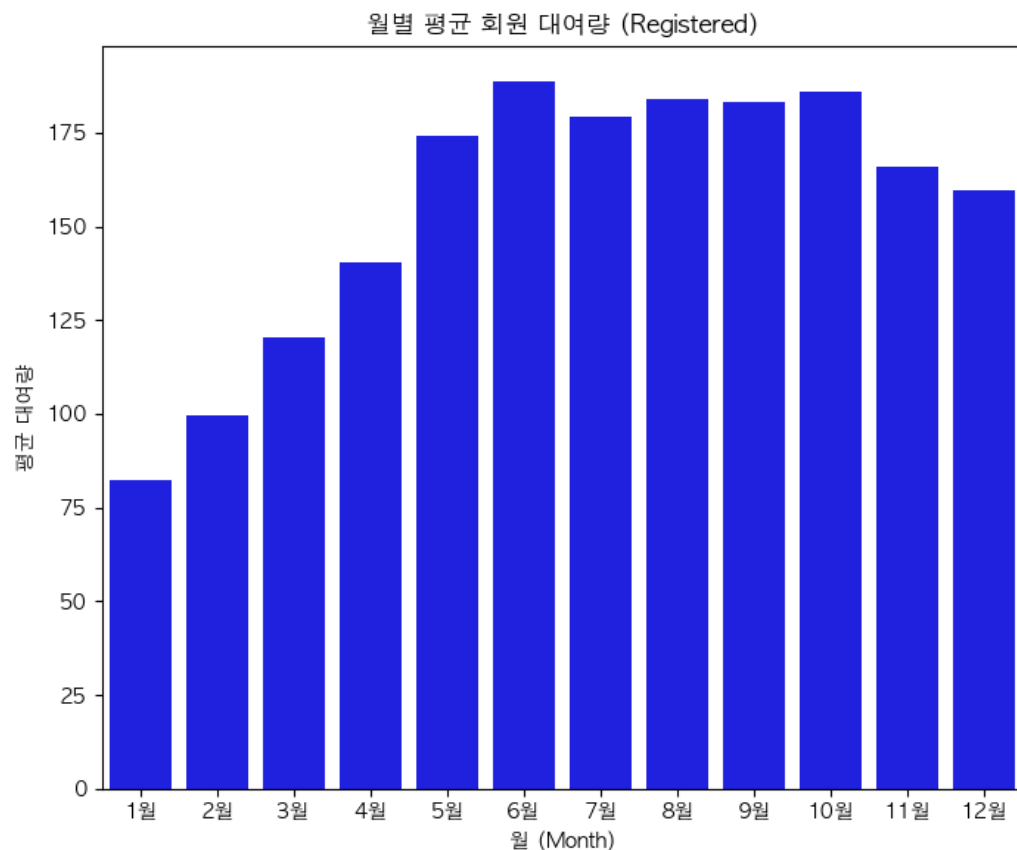
## 날씨별 평균 대여량



- 회원은 날씨가 좋지 않아도 출퇴근 및 일상적인 이동을 위해 지전거를 이용한다고 예측되는데 다만 폭우 폭설때 데이터량이 줄지 않는 이유는 예측에 벗어난 결과 입니다.
- 폭우,폭설때 오히려 택시나 대중교통 이용이 불가능한 경우라 최후의 수단으로 이용되는 것 일수도...?
- 반면 비회원은 날씨가 좋을때 급증하고 날씨가 안 좋을때 급락 하는걸 보면 비회원의 간단한 여행, 여가활동으로 인한 데이터예측과 일치 합니다.

# EDA

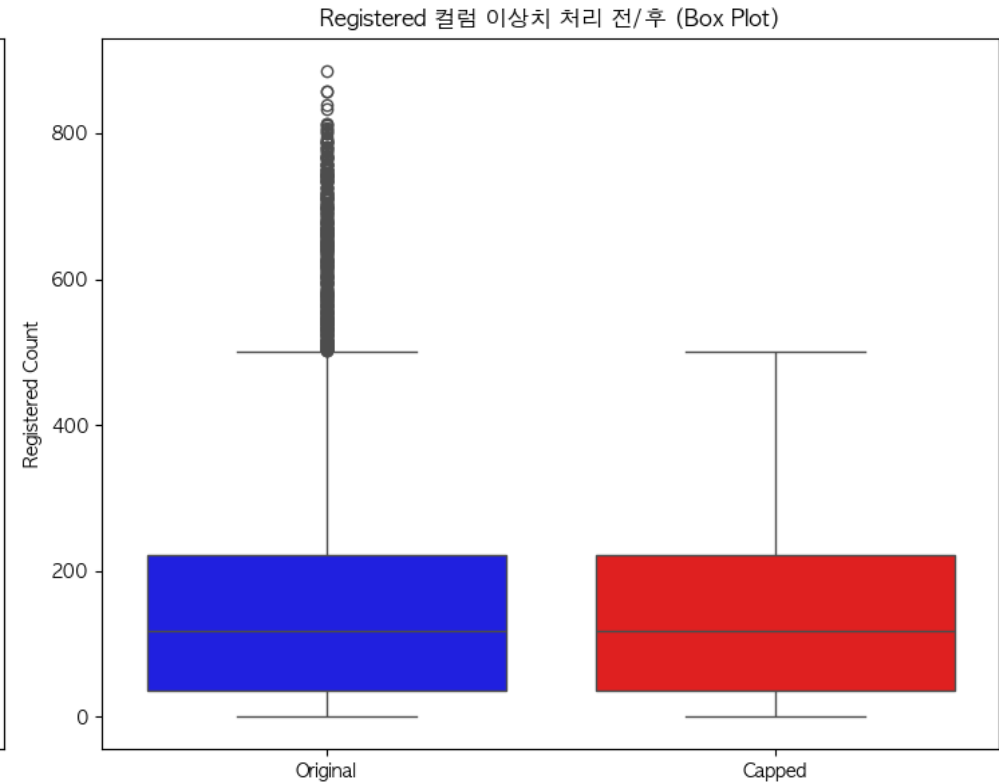
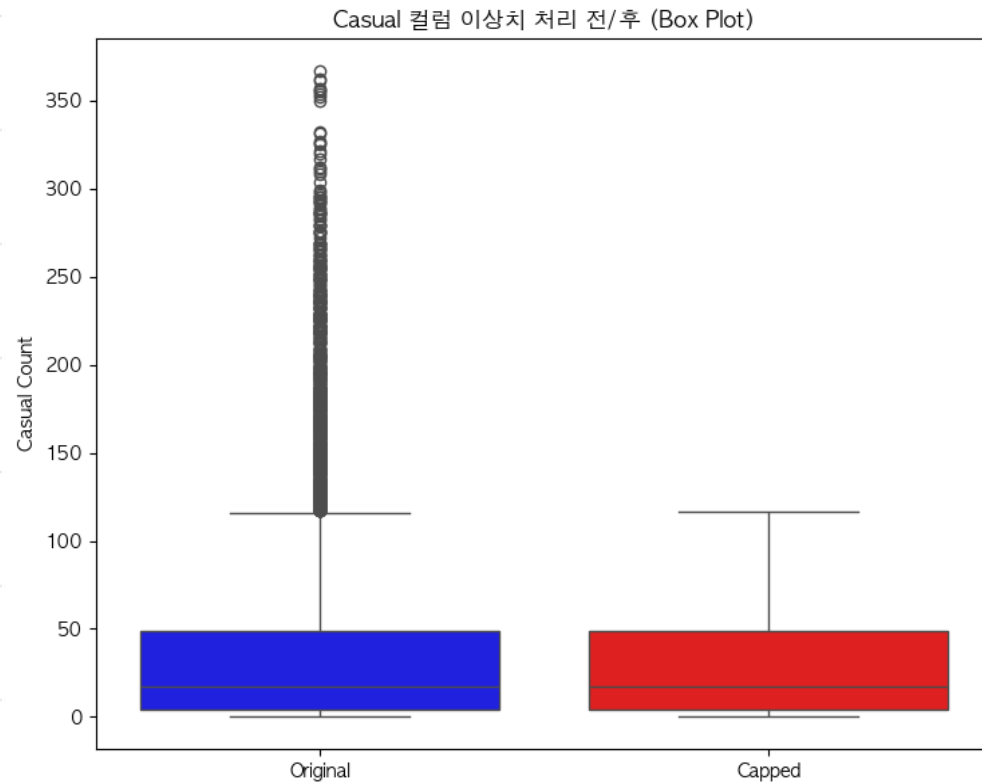
## 월별 대여량



- 회원은 연중 꾸준히 이용하는 경우가 많음, 낮은 온도인해 일상적인 통근 수요 감소, 혹은 방학
- 5월부터 ~10월까지의 예측 가능하고 안정적인 수요를 보입니다.
- 비회원은 계절에 민감한 이용층, 6월~9월에 야외활동이 많은 기간에 수요가 몰립니다. 이기간 주말 및 공휴일에 수요가 폭증할거로 예상

# 전처리

## 이상치 처리

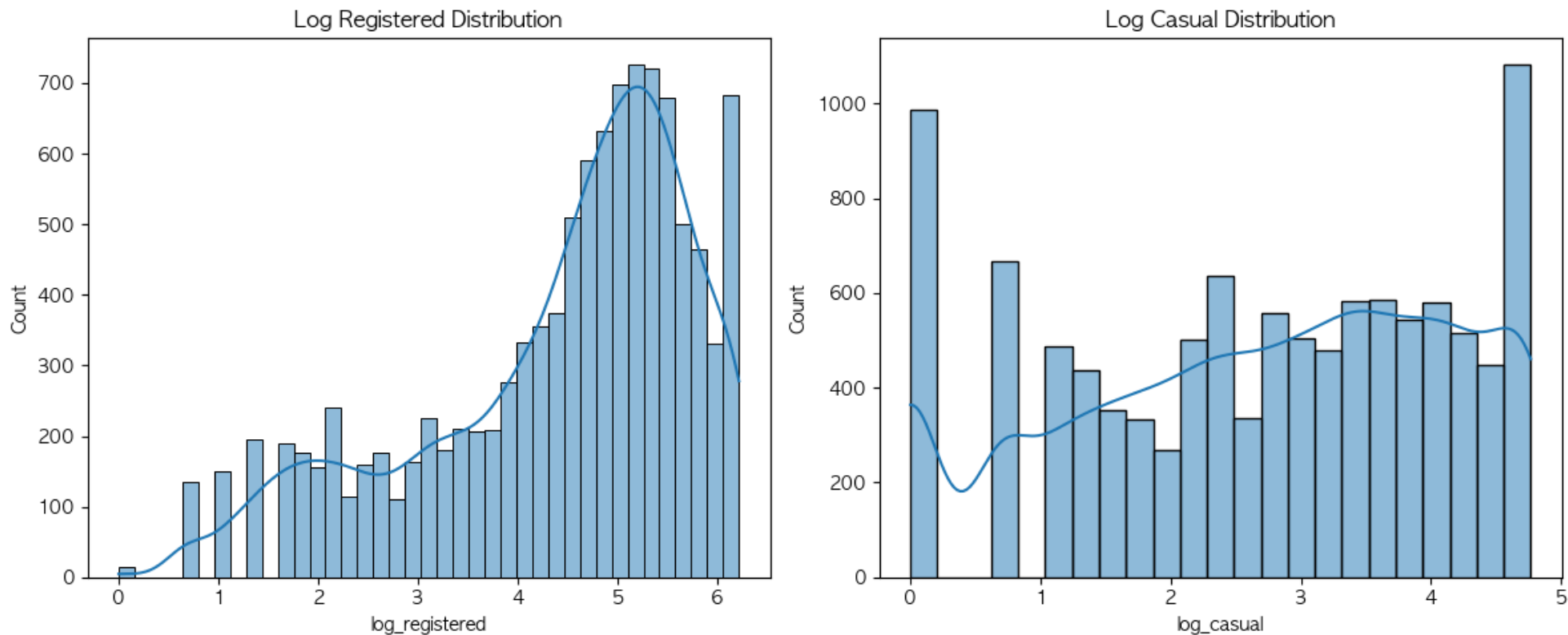


- 전처리 과정에서 비회원 회원 이상치가 를 발견하여 iqr기준에 벗어나는 데이터는 모두 제거하였습니다.



# 모델링

RMSLE를 목표로 하는 미션에서, 타겟 변수(registered, casual)의 심하게 치우친 분포를 해결하기 위해 로그 변환.



- EDA에서 비회원(casual)의 대여량이 특정 시점에 폭발적으로 증가하여 분산이 크다는 것을 확인했습니다. 이러한 이상치는 모델의 예측 성능을 저해할 수 있으므로, 처리.
- 가장 일반적인 이상치 식별 기준인 IQR 방식을 사용하여 casual과 registered 대여량의 상한선을 설정하겠습니다. 극단적인 이상치는 제거하는 대신, 해당 상한값으로 대체(Capping)하는 것이 데이터 손실을 최소화하는 좋은 방법입니다.

# 모델링


피쳐 엔지니어링

- 시간 관련 피쳐 생성 및 추출
- 범주형 변수 변환
- 상호작용 피쳐 생성
- 데이터분리
- 학습 및 검증 RMSLE 수치: 0.2894

# 모델링

## 하이퍼 파라미터 튜닝

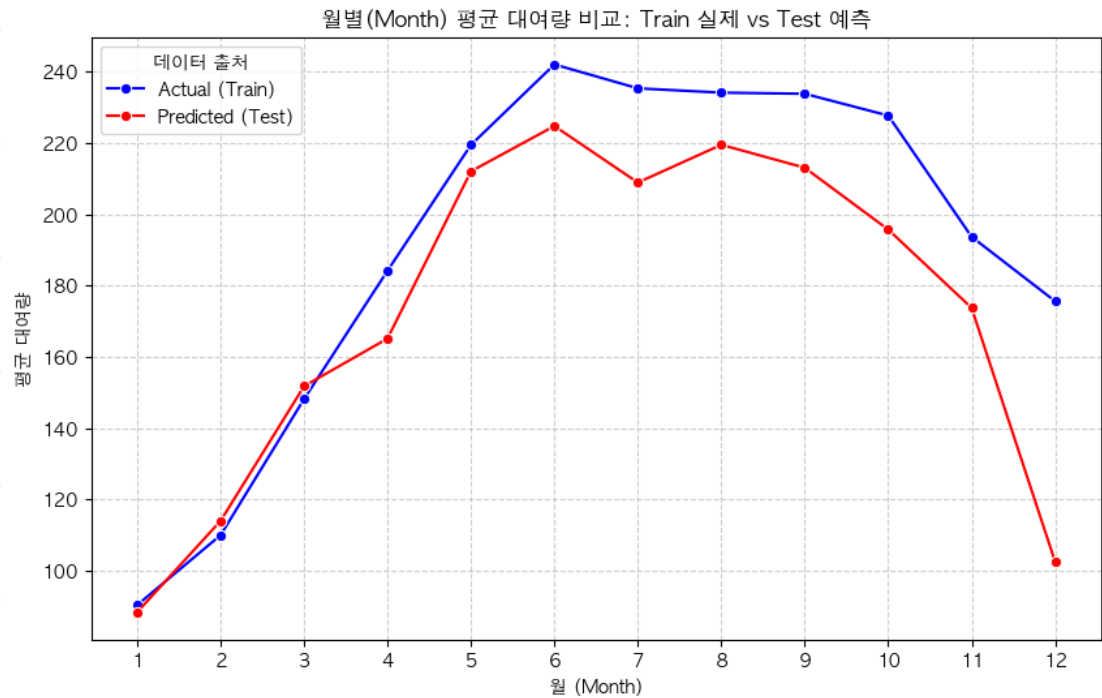
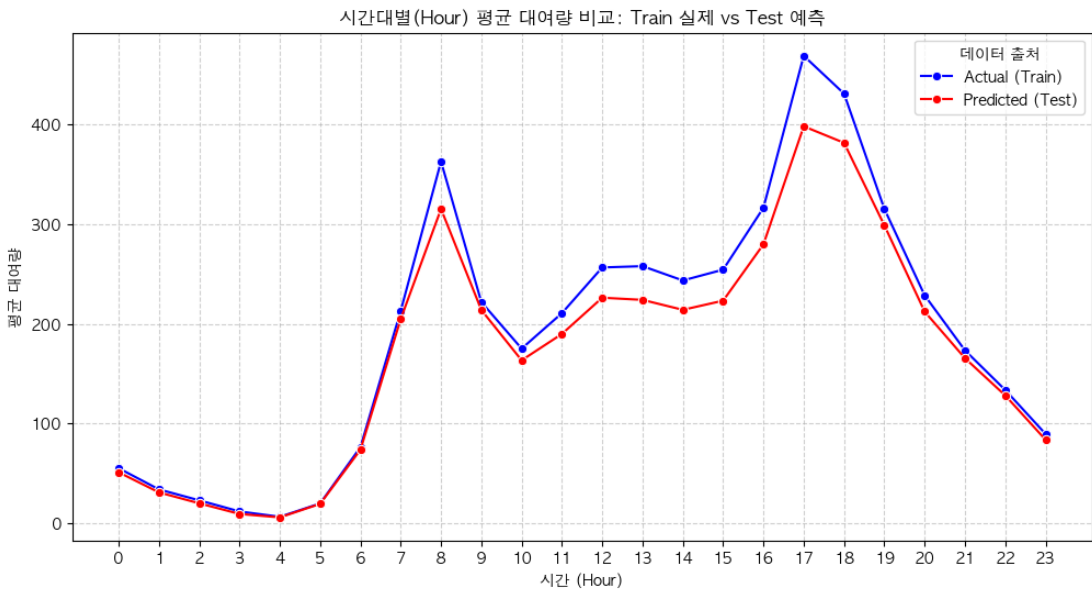
- **모델:** 두 개의 독립적인 **LightGBM** 모델을 사용했습니다.
- **튜닝:** RandomizedSearchCV 과  $\mathbf{3}$ 겹 교차 검증  $\mathbf{Fold}$ 를 사용하여 각 모델의 **하이퍼파라미터**를 최적화했습니다.
- **최종 성능:** 0.2889

All		Successful	Selected	Errors	Recent ▾		
Submission and Description		Private Score ⓘ		Public Score ⓘ		Selected	
	submission_final_lgbm.csv	0.38996		0.38996		<input type="checkbox"/>	
	Complete (after deadline) · 2m ago						

케글 사이트 점수: 0.38996

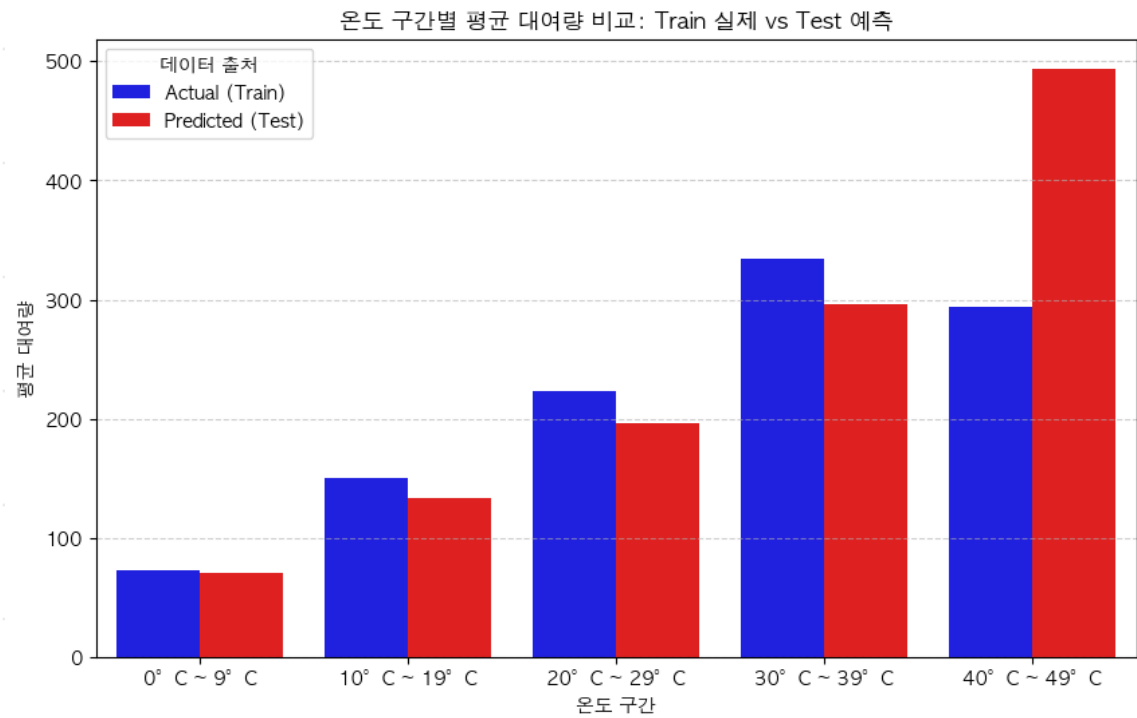
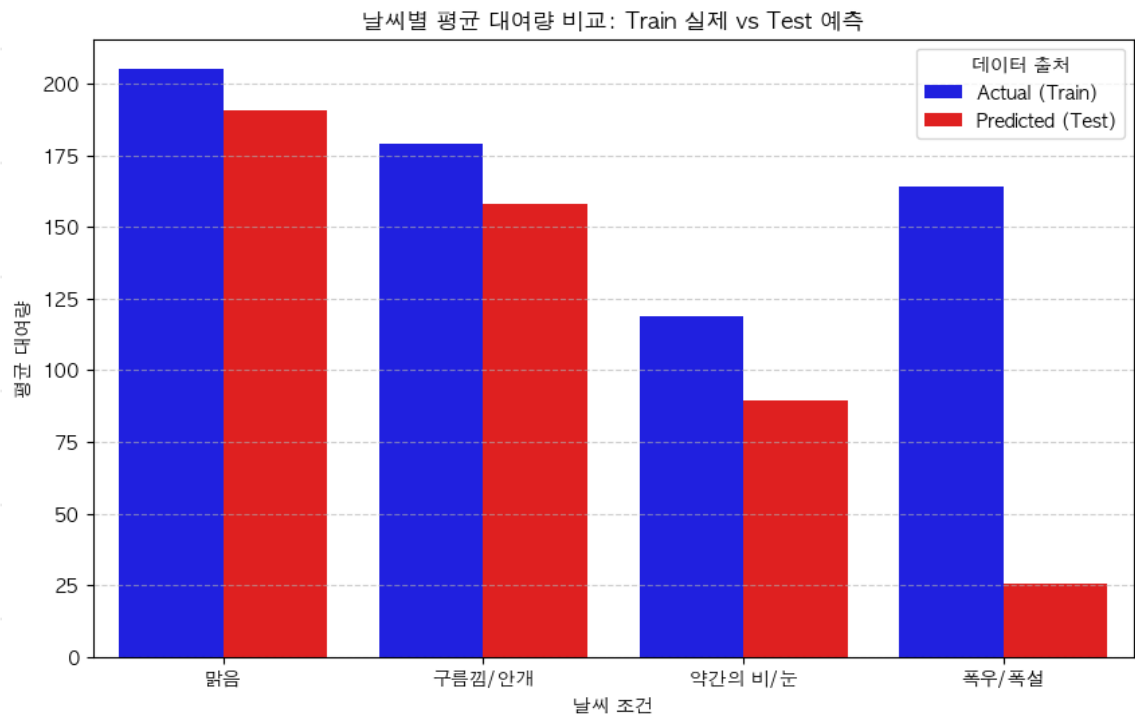
# 결과 및 해석

## 실제데이터와 test예측 데이터 비교 결과 시각화



# 결과 및 해석


실제데이터와 test예측 데이터 비교 결과 시각화



# 결과 및 해석

## 해석

- 회원/비회원 대여량이 날씨, 온도, 시간대, 요일 등의 환경 요인에 따라 매우 다른 패턴을 보인다는 것을 확인했습니다.
- 모델링시 EDA자료를 통해 회원/비회원을 나누어 모델링을 하여 0.2889수치가 나온걸로 보아 탁월한 선택이었습니니다.
- 하지만 폭우,폭설시 or 온도가 높을시 예측력은 우리 모델이 예측을 실제데이터처럼 정확히 하지 못했습니다. 특수한 상황에서의 예측력을 키우기위해 모델의 개선이 필요합니다.

All		Successful	Selected	Errors	Recent ▼		
Submission and Description		Private Score ⓘ		Public Score ⓘ	Selected		
	submission_final_lgbm.csv	0.38996		0.38996	<input type="checkbox"/>		
	Complete (after deadline) · 2m ago						

## 앞으로 대응

1. 우리 모델은 시간대별, 요일별 월별 수요 패턴을 정확히 예측합니다.

- 시간대별 재배치 :오전 8시~9시와 오후 5시~7시의 피크 시간대에 자전거가 부족하지 않도록 인력이 해당 스테이션에 집중 배치되어 재분배 작업을 수행해야 합니다.

2.비회원 고객 유치를 위한 마케팅 및 프로모션

-온도가 20도 이상으로 예측되는 주말/휴일 시간대에 비회원 대상의 시간당 할인 쿠폰이나 첫 30분 무료 등의 프로모션을 집중적으로 시행하여 수요를 극대화해야 합니다.

3. 날씨대응

- 날씨가 안좋은 구간에는 대여 수요가 급감합니다. 재배치 인력을 줄이고, 자전거 정비 등 날씨에 영향을 받지않는 인력으로 배치 합니다.



감사합니다.