

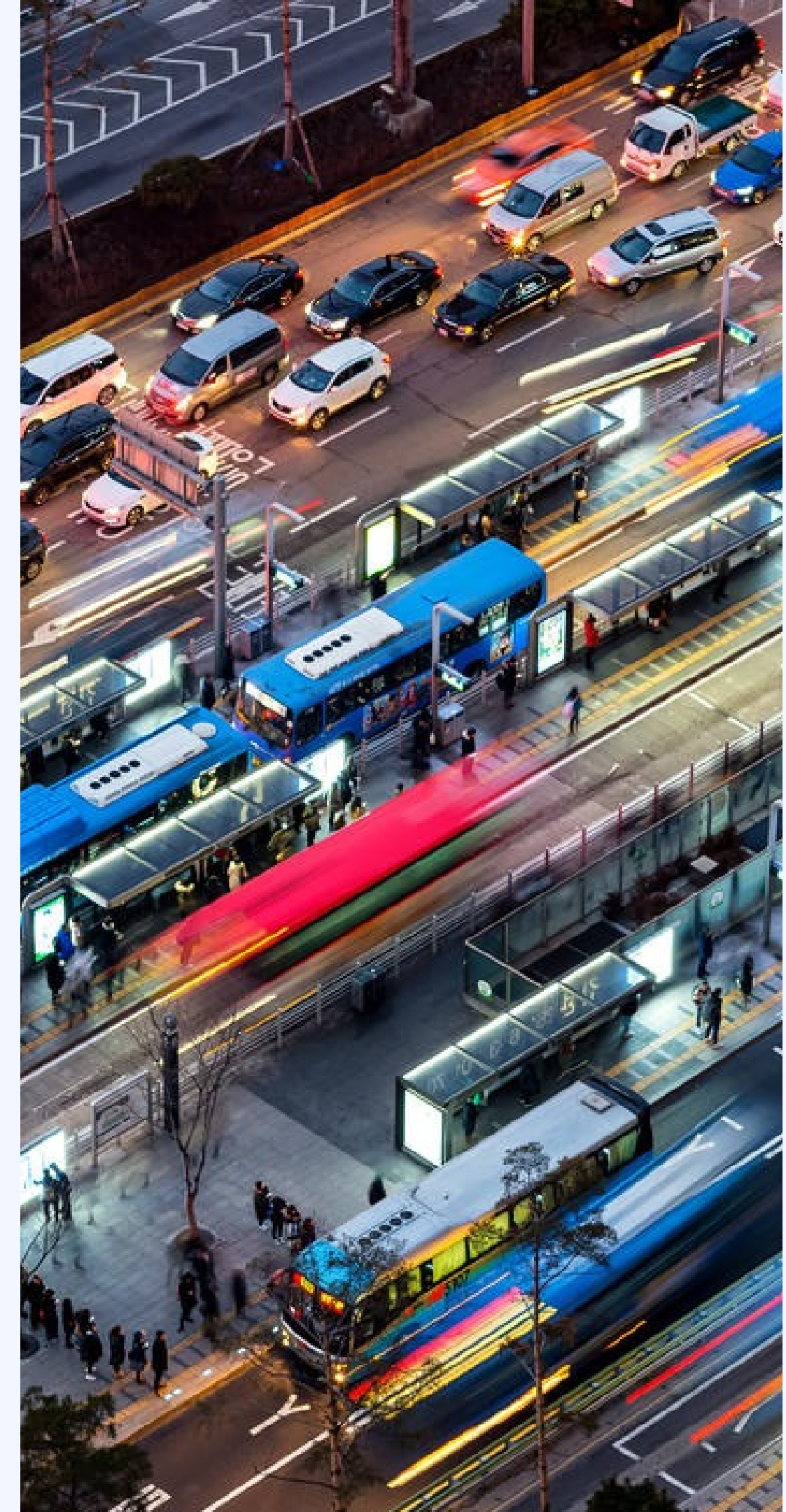
# 대전시 대중교통 통행수요 예측

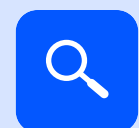
지역사회 문제해결형 빅데이터/AI활용 공모전

LOVE MAE

20202105 김수경 20222866 김승일

20222303 박성빈 20191417 박성진





# 목차

## 1. 데이터 분석

- EDA

## 2. 데이터 가공

- Data 전처리

## 3. 모델링

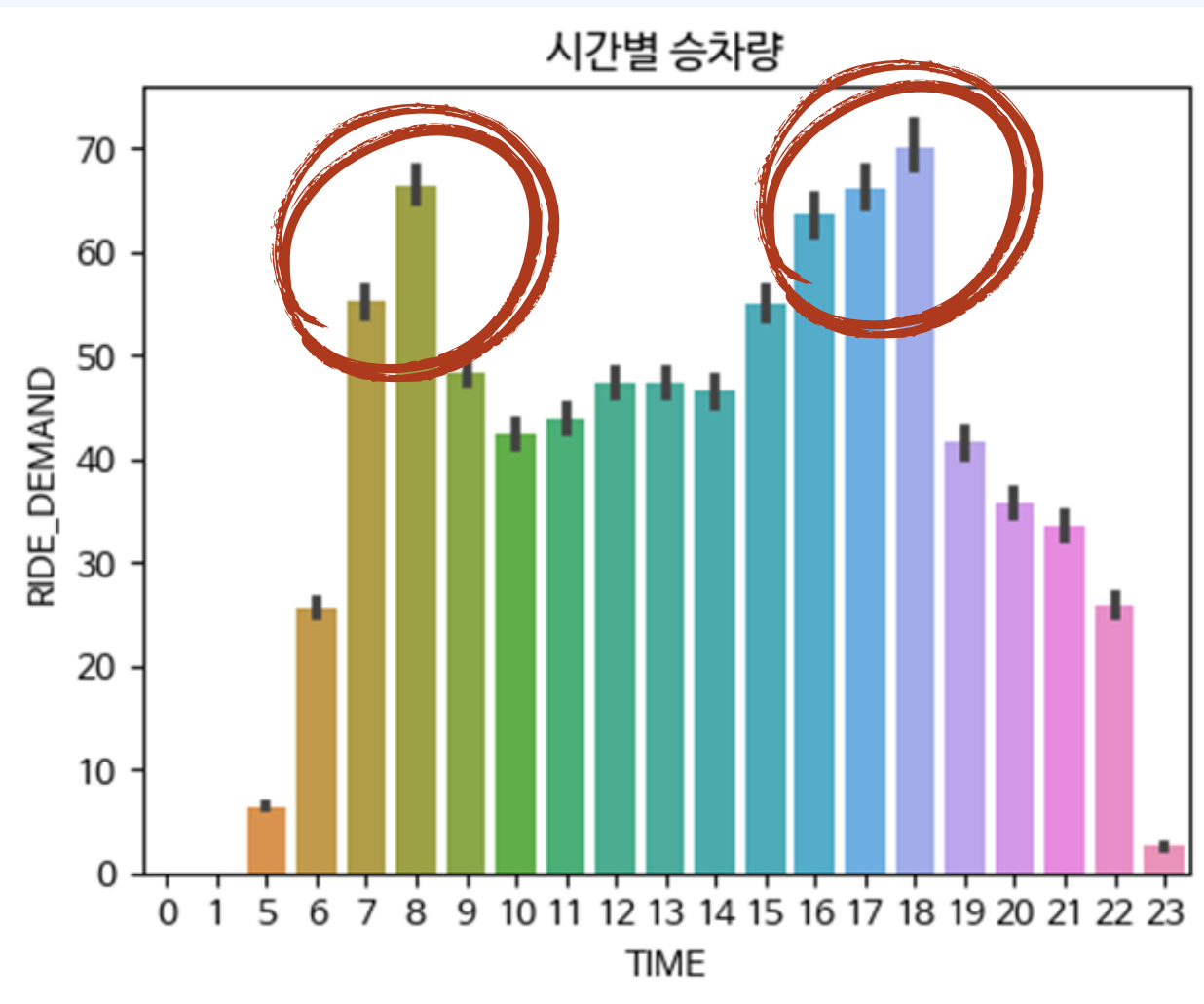
- AutoGluon

## 4. 모델링 분석 결과

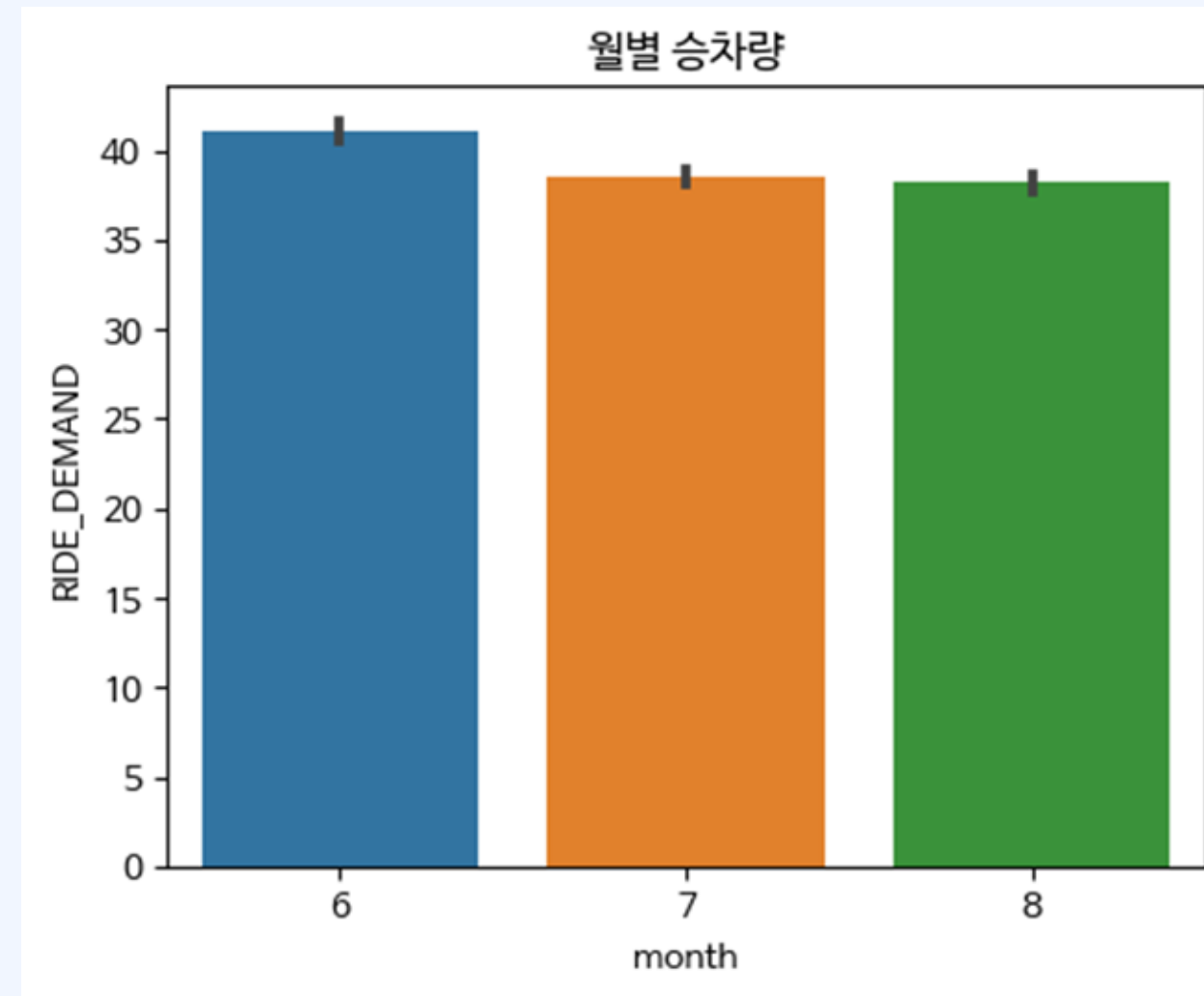
- train data, test data 비교

# 1.데이터 분석

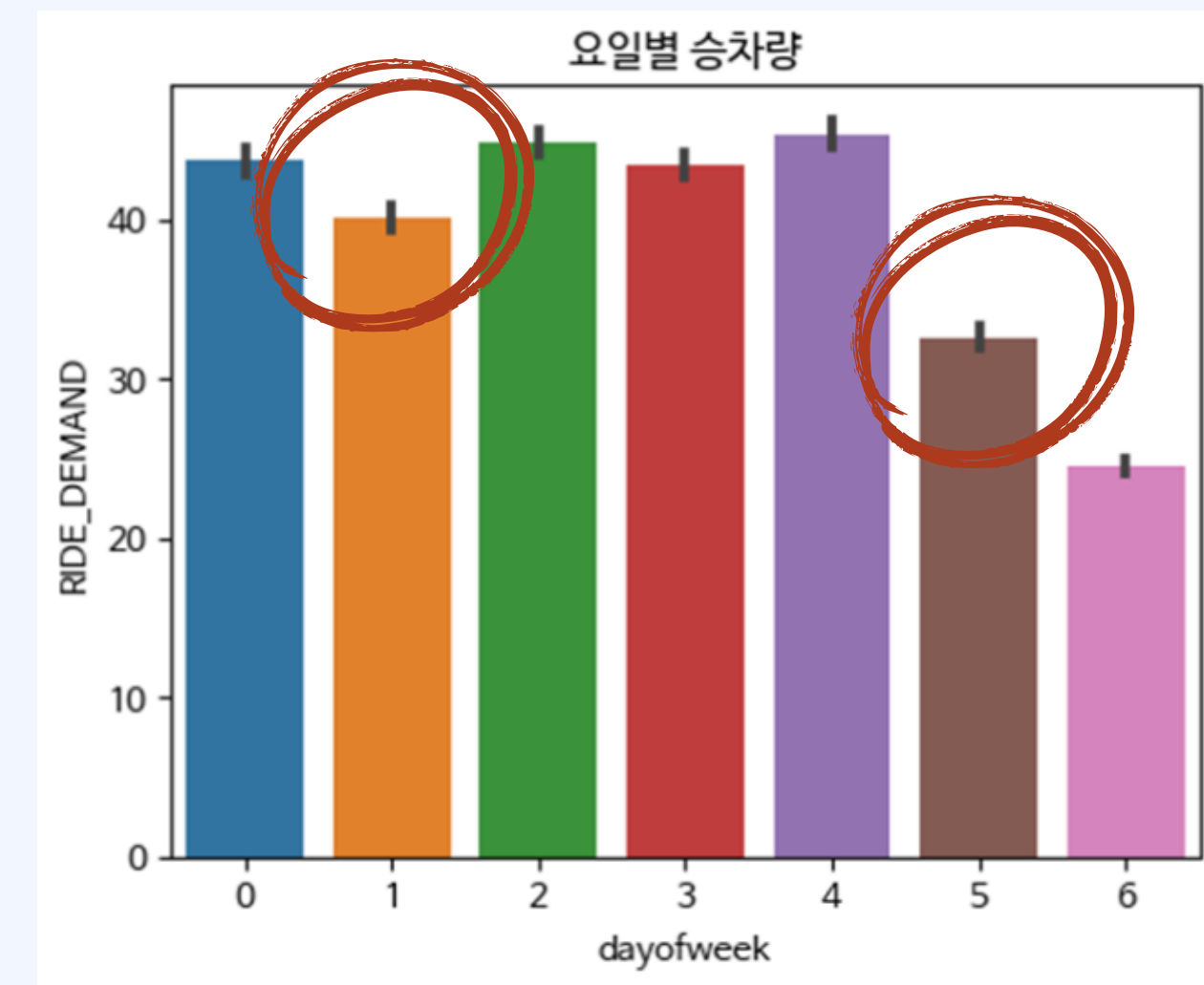
## EDA



출퇴근 시간에 따른  
승차수요 차이



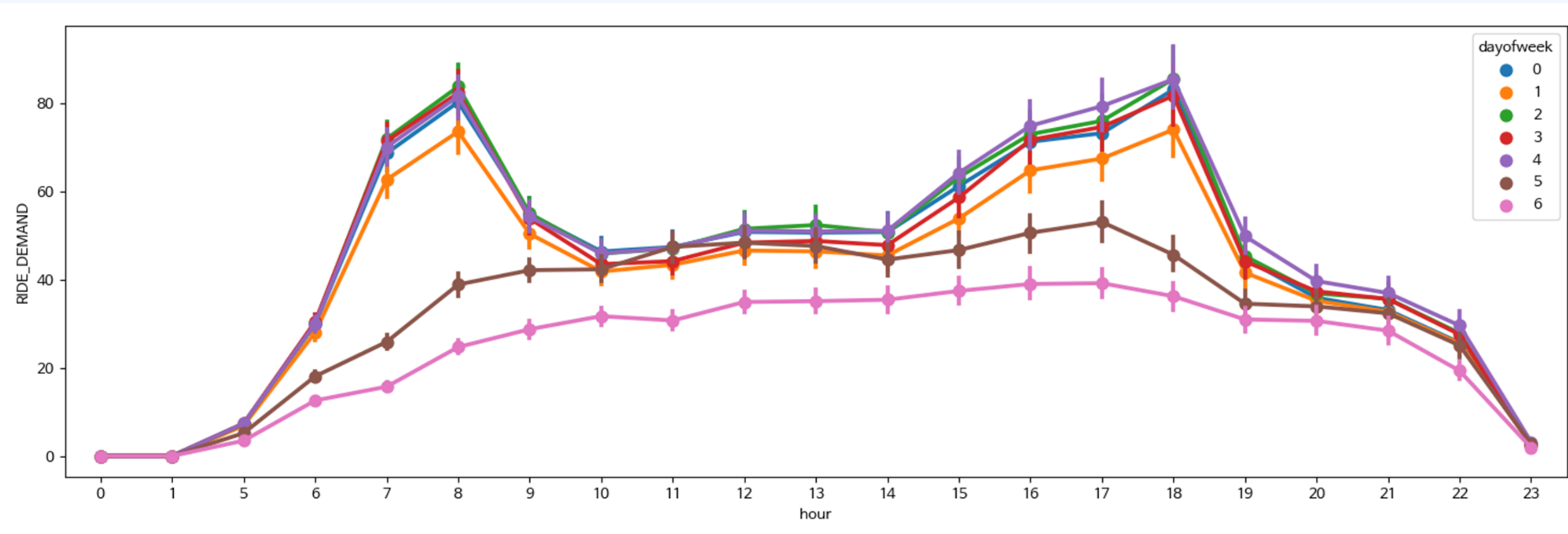
월별 승차수요 차이 X



- 평일의 승차수요 차이
- 주말의 승차수요 차이

# 1.데이터 분석

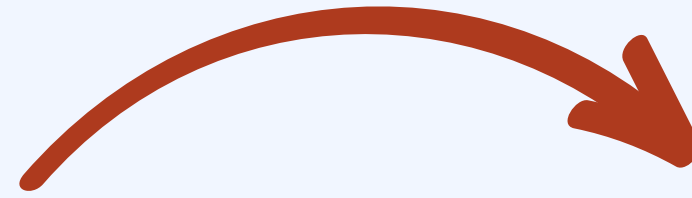
EDA



요일에 따른 시간별 승차량

## 2. 데이터 가공

제거



	gid	DATE	TIME	RIDE_DEMAND	ALIGHT_DEMAND
0	다마9599	2023-06-01	5	1	0
1	다마9599	2023-06-01	6	2	5
2	다마9599	2023-06-01	7	2	7
3	다마9599	2023-06-01	8	3	12
4	다마9599	2023-06-01	9	2	31
...	...	...	...	...	...
539485	라바0421	2023-08-24	21	0	1
539486	라바0421	2023-08-24	22	0	0
539487	라바0421	2023-08-24	23	0	0
539488	라바0421	2023-08-24	0	0	0
539489	라바0421	2023-08-24	1	0	0
539490 rows × 5 columns					

	DATE	TIME	RIDE_DEMAND	ALIGHT_DEMAND
0	2023-06-01	5	1	0
1	2023-06-01	6	2	5
2	2023-06-01	7	2	7
3	2023-06-01	8	3	12
4	2023-06-01	9	2	31
...	...	...	...	...
539485	2023-08-24	21	0	1
539486	2023-08-24	22	0	0
539487	2023-08-24	23	0	0
539488	2023-08-24	0	0	0
539489	2023-08-24	1	0	0
539490 rows × 4 columns				

- 'gid'는 오버피팅의 우려 및 관련성 없음으로 **제거**

## 2. 데이터 가공

### 제거

	DATE	TIME	RIDE_DEMAND	ALIGHT_DEMAND
0	2023-06-01	5	1	0
1	2023-06-01	6	2	5
2	2023-06-01	7	2	7
3	2023-06-01	8	3	12
4	2023-06-01	9	2	31
...	...	...	...	...
539485	2023-08-24	21	0	1
539486	2023-08-24	22	0	0
539487	2023-08-24	23	0	0
539488	2023-08-24	0	0	0
539489	2023-08-24	1	0	0

539490 rows x 4 columns

	TIME	RIDE_DEMAND	ALIGHT_DEMAND	YEAR	MONTH	DAY
0	5	1	0	2023	6	1
1	6	2	5	2023	6	1
2	7	2	7	2023	6	1
3	8	3	12	2023	6	1
4	9	2	31	2023	6	1
...	...	...	...	...	...	...
539485	21	0	1	2023	8	24
539486	22	0	0	2023	8	24
539487	23	0	0	2023	8	24
539488	0	0	0	2023	8	24
539489	1	0	0	2023	8	24

539490 rows x 6 columns

- 'DATE' 컬럼을 연, 월, 일로 분리 후 'DATE' 제거

# 3. 모델링

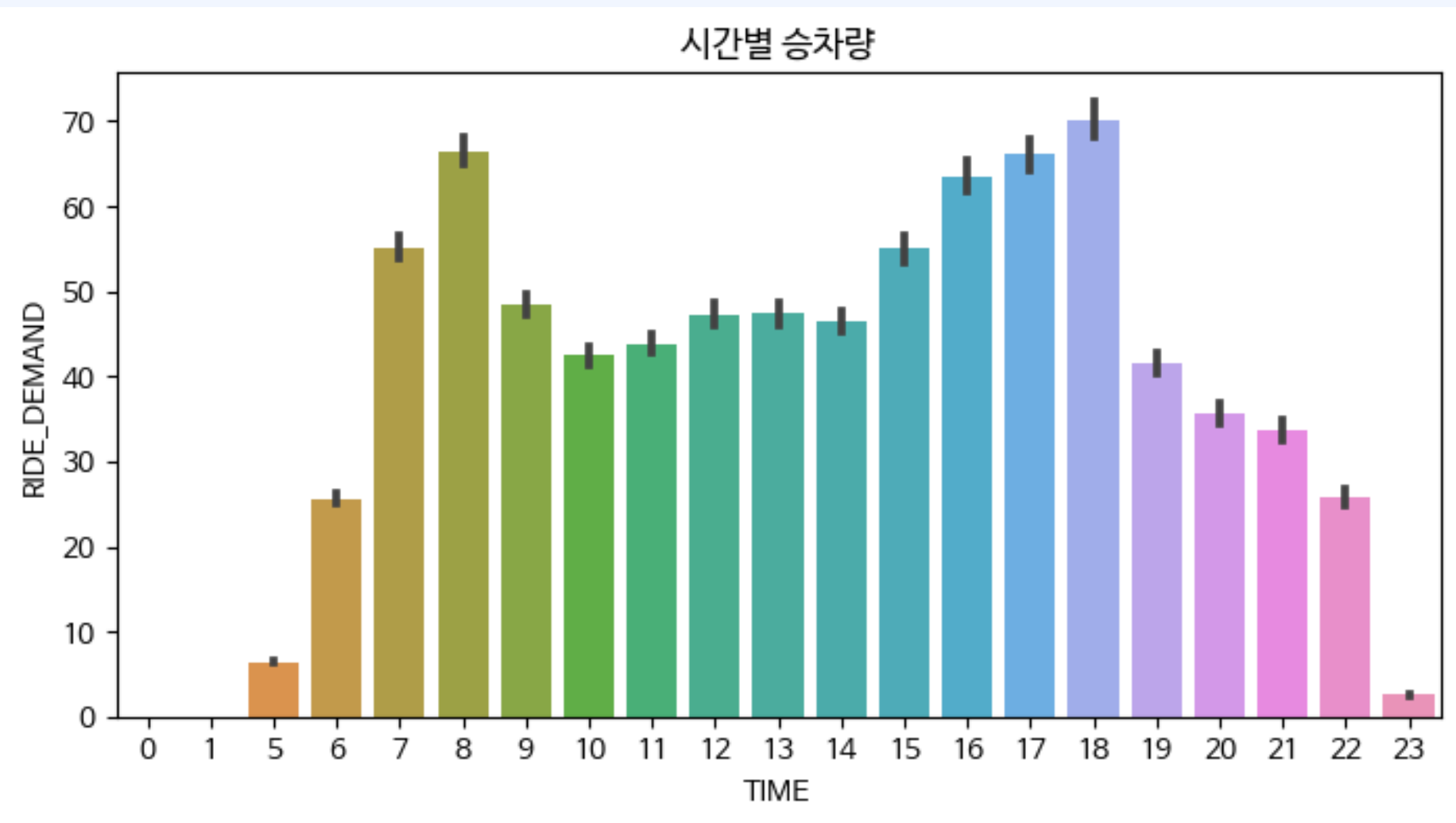
Autogluon 사용



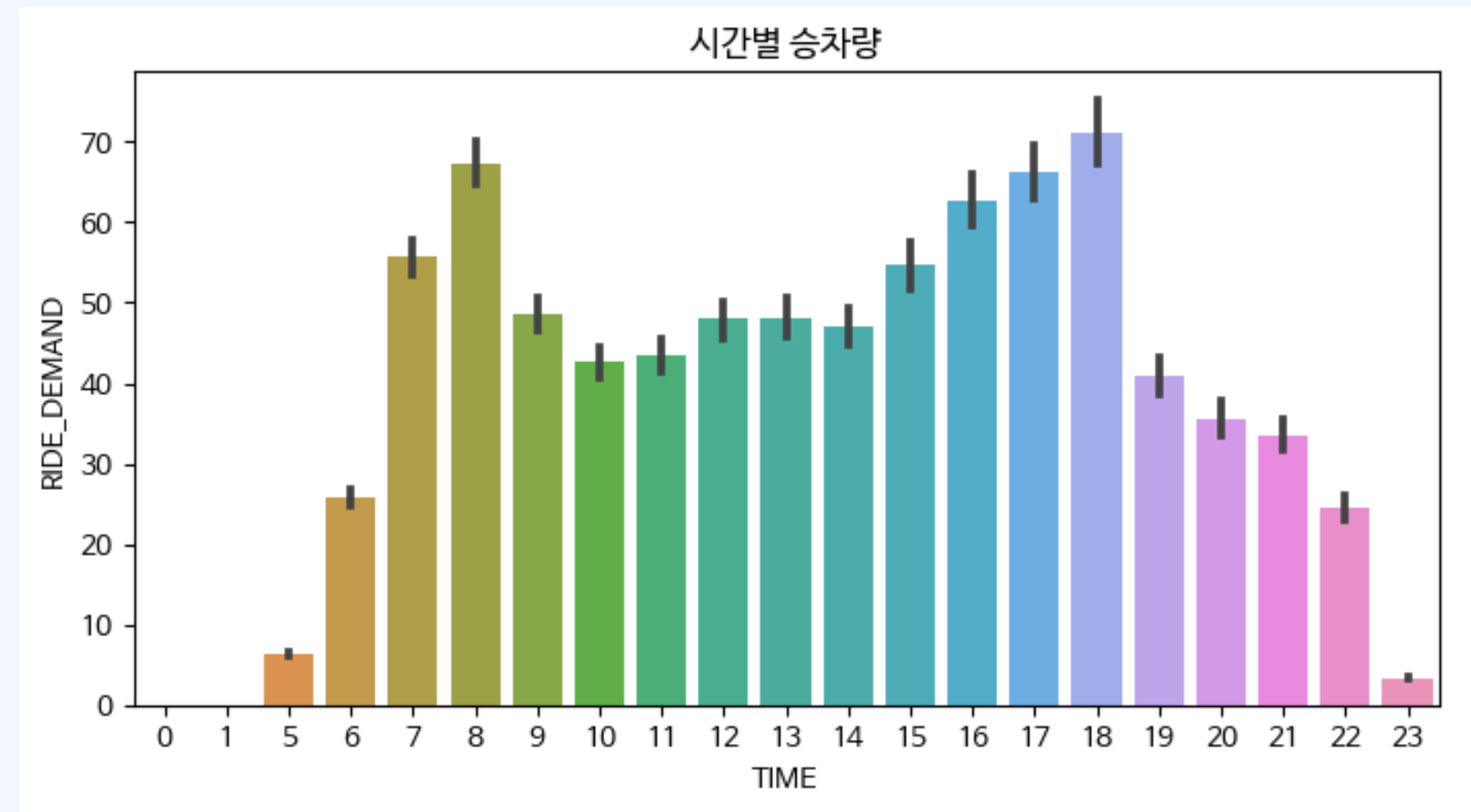
- 머신러닝 모델을 구축하고 최적화하는 기능을 제공
- 다양한 머신러닝 작업에 대해 최적의 모델을 자동으로 선택하고 튜닝

# 4. 모델링 분석 결과

모델링한 결과 train data와 test data의 시간별 승차량 차이 확인



train data

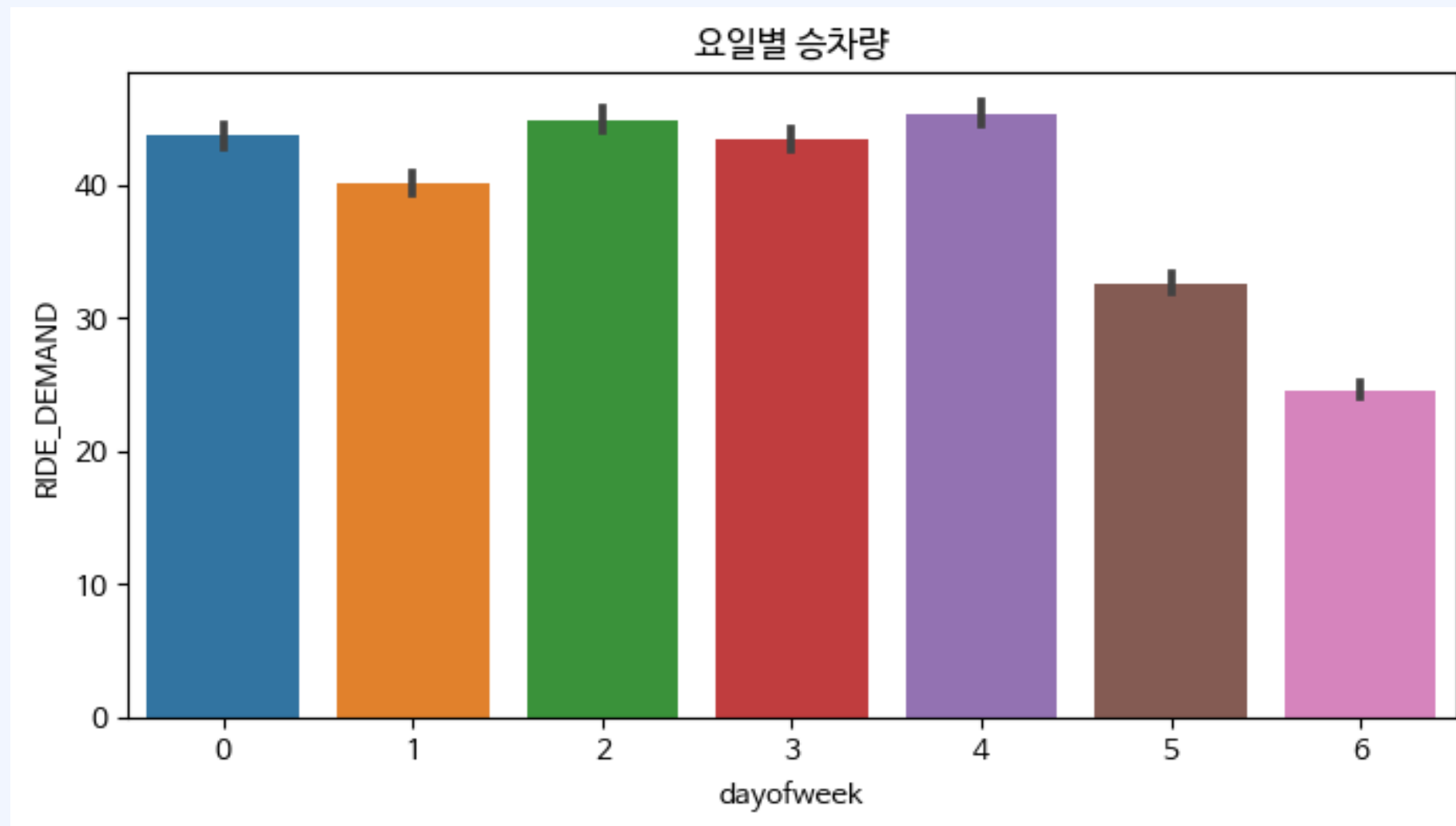


test data

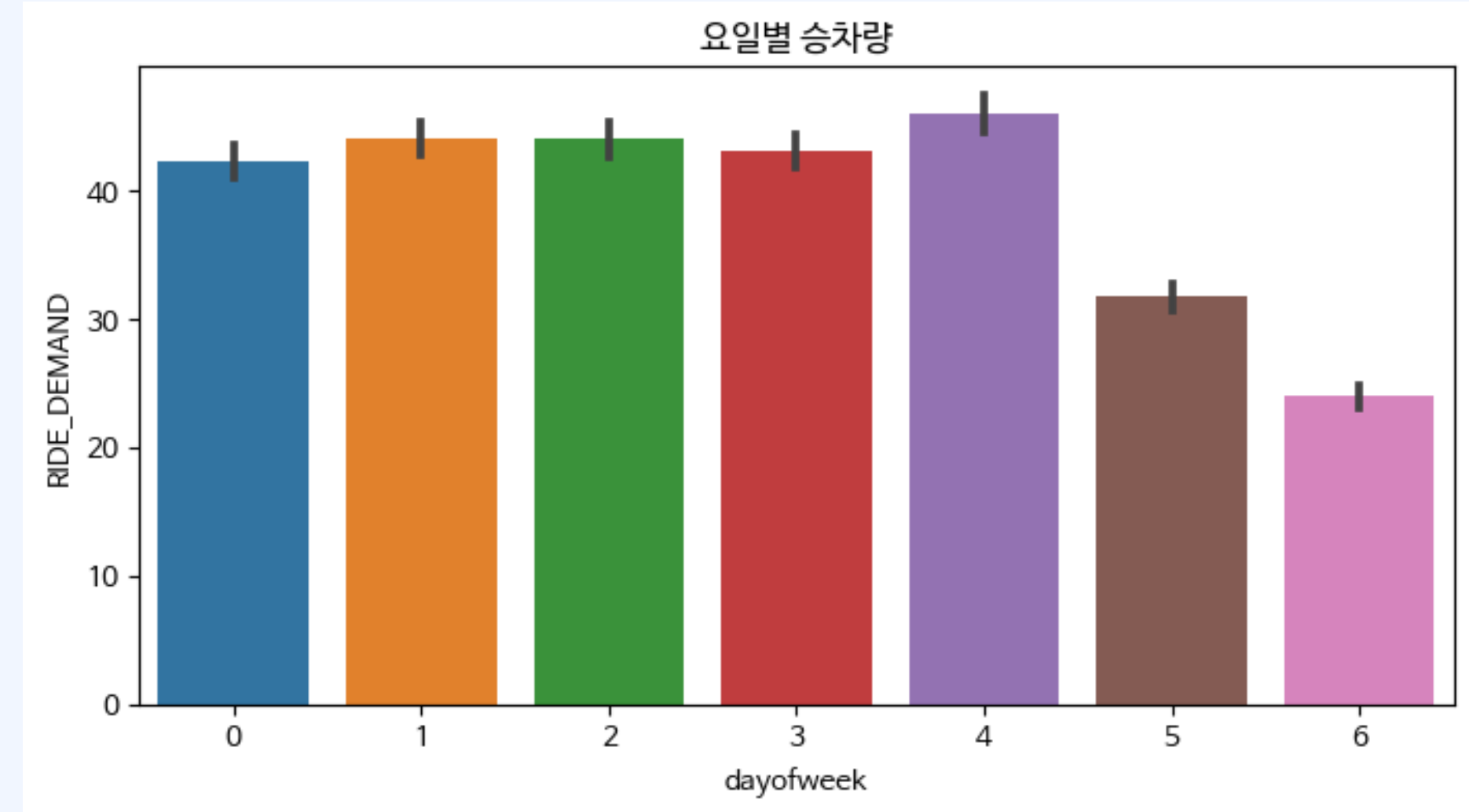


# 4. 모델링 분석 결과

모델링한 결과 train data와 test data의 요일별 승차량 차이 확인



train data



test data

# 프로젝트를 마무리하며

## 아쉬운 점

- 기존 모델링의 오버피팅 확인 X
- 기존 모델링 안에서 보완 X
- 예측된 결과에서 화요일의 승차량이 높아짐의 이유 확인 X

감사합니다

# 조원별 역할

## 전반적인 과정은 **다같이** 진행

- 김수경: 팀장, 방향성 제시, ppt 제작
- 김승일: 모델링 방법 조사
- 박성빈: EDA, 발표
- 박성진: 전처리 및 모델링