

# 의료 데이터 분석 경진대회

분석 결과 발표

2024.11.26

컴퓨터공학과 | 20214110 안세준



# CONTENTS

## 발표 순서

- 1 소개 및 배경
- 2 예약 간격 분석
- 3 SMS 수신 여부 분석
- 4 예약 시간대 분석
- 5 NO-SHOW 예측 모델 학습



1.

## 소개 및 배경

평균 노쇼 비율

```
average_noshow_rate = (df['No-show'] == 'Yes').mean()  
print(f"평균 노쇼 비율: {average_noshow_rate:.2%}")
```

✓ 0.0s

평균 노쇼 비율: 20.19%

데이터셋

df.columns

✓ 0.0s

```
Index(['PatientId', 'AppointmentID', 'Gender', 'ScheduledDay',  
      'AppointmentDay', 'Age', 'Neighbourhood', 'Scholarship', 'Hipertension',  
      'Diabetes', 'Alcoholism', 'Handcap', 'SMS_received', 'No-show',  
      'DaysDifference', 'ScheduledDay_Weekday', 'AppointmentDay_Weekday',  
      'ScheduledHour'],  
      dtype='object')
```

분석 의도

- 데이터 내 환자들은 20.19%의 확률로 병원에 나타나지 않음.
- 노쇼를 줄이기 위한 요인 식별의 필요성을 인식.
- 데이터셋의 많은 데이터가 노쇼에 영향을 미칠 가능성이 있다고 판단.
- 노쇼에 영향을 미치는 요인을 분석하고 해결 방안을 모색.

## 2. 예약간격 분석

예약 간격이란?

- ScheduledDay(예약일), AppointmentDay(진료일)
- DaysDifference(예약 간격) = 진료일 - 예약일
- 예약 간격은 진료일과 예약일 간 차이를 의미

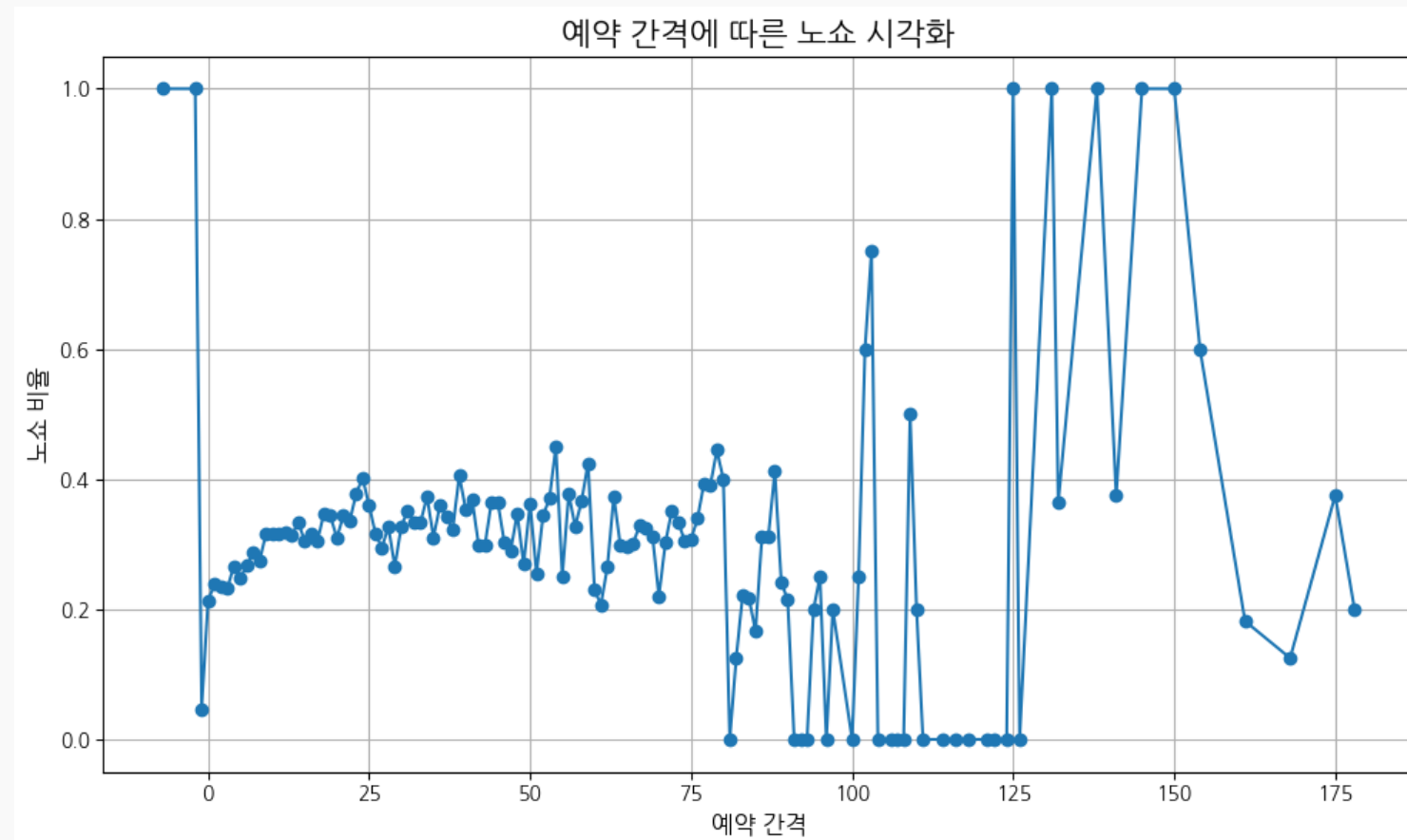
```
# datetime 타입으로 변환
df['ScheduledDay'] = pd.to_datetime(df['ScheduledDay'])
df['AppointmentDay'] = pd.to_datetime(df['AppointmentDay'])

# 날짜 차이 계산
df['DaysDifference'] = (df['AppointmentDay'] - df['ScheduledDay']).dt.days
```

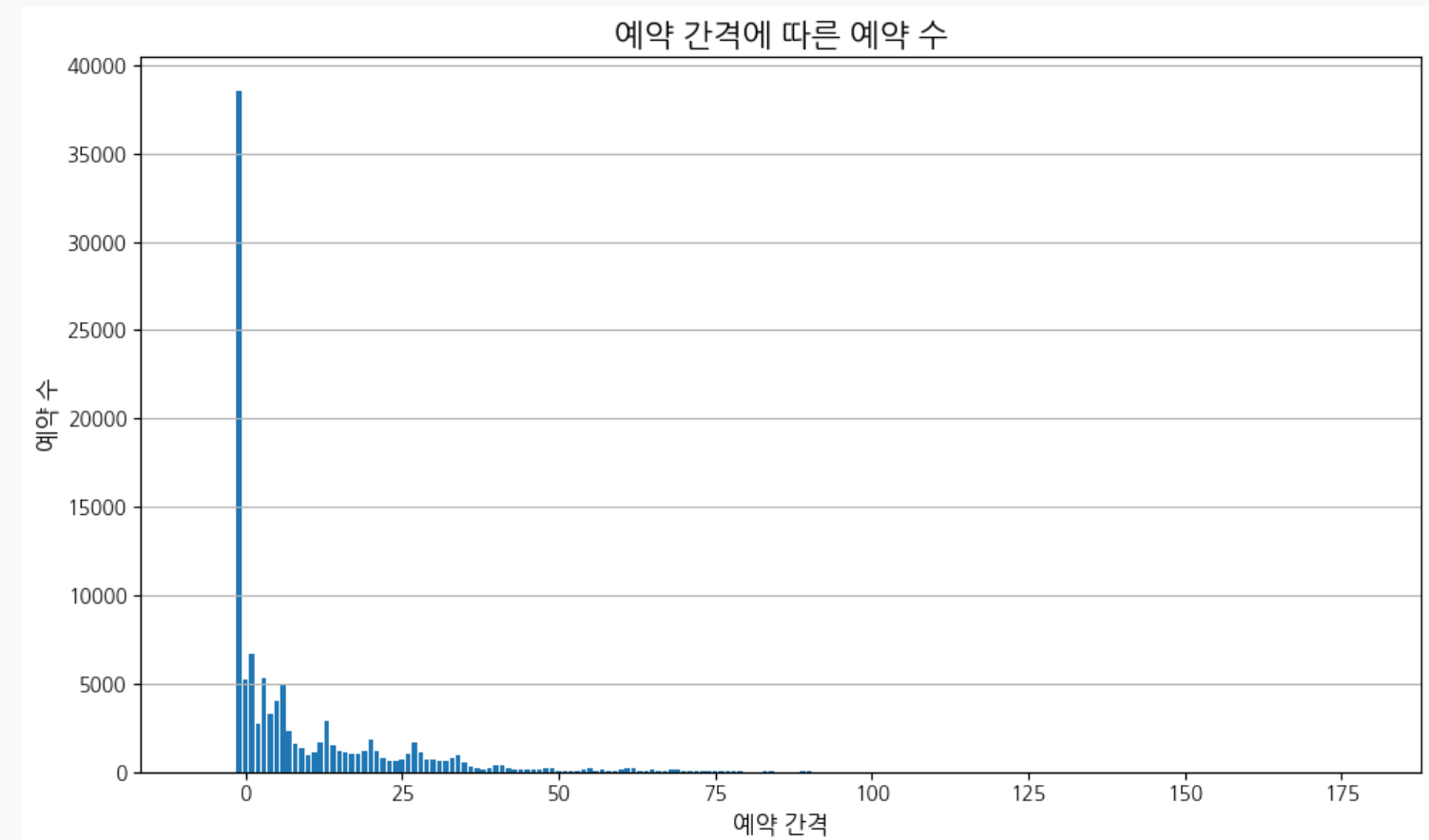


## 2. 예약간격 분석

### | 예약 간격에 따른 노쇼 시각화 |



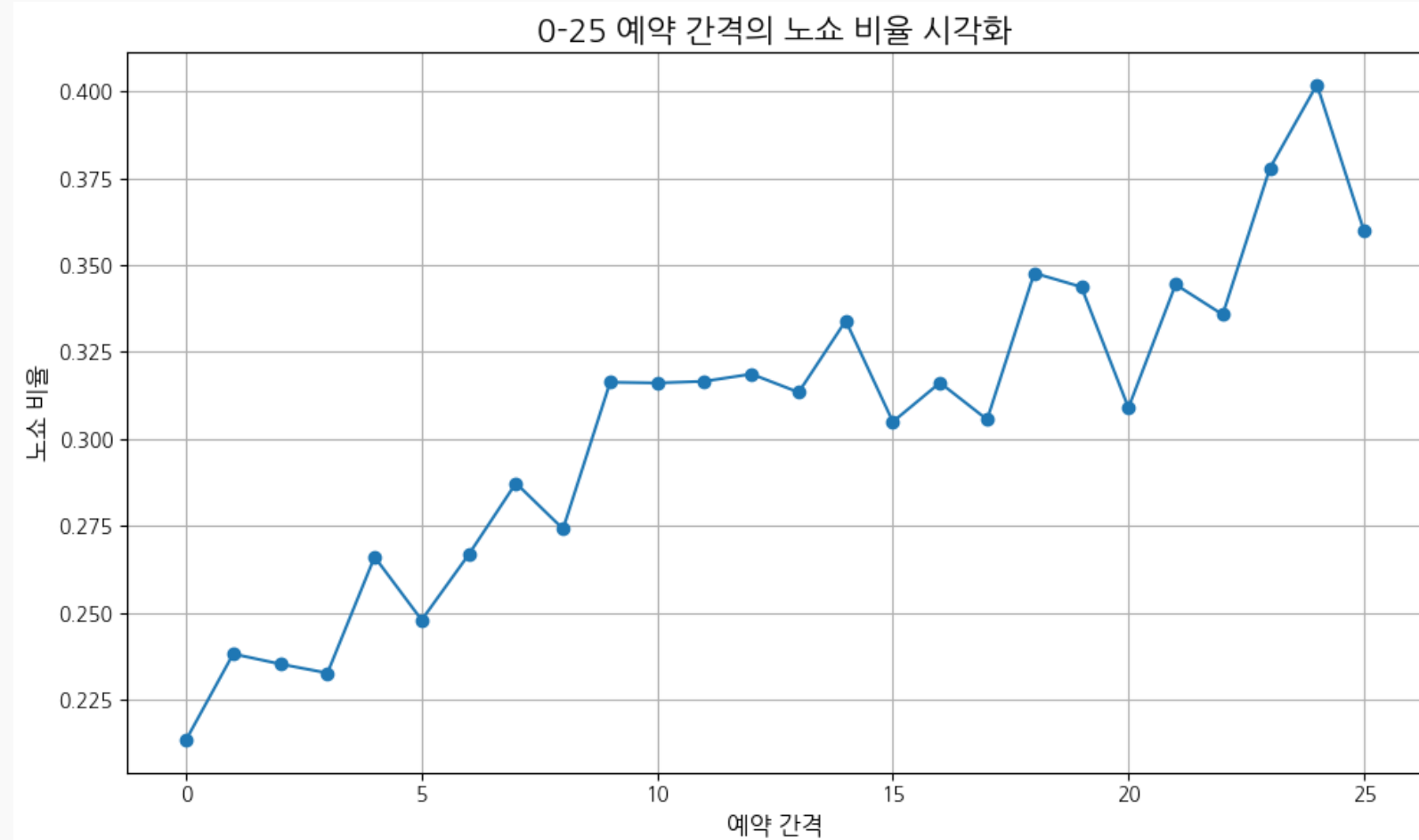
### | 예약 간격에 따른 예약 수 시각화 |





## 2. 예약간격 분석

### | 예약 간격에 따른 노쇼 시각화 |



### | 발견된 이슈 |

- 분석내용
- 표본이 많은 0~25까지 예약간격 구간 한정
  - 예약 간격이 클 수록 노쇼 비율이 꾸준히 증가

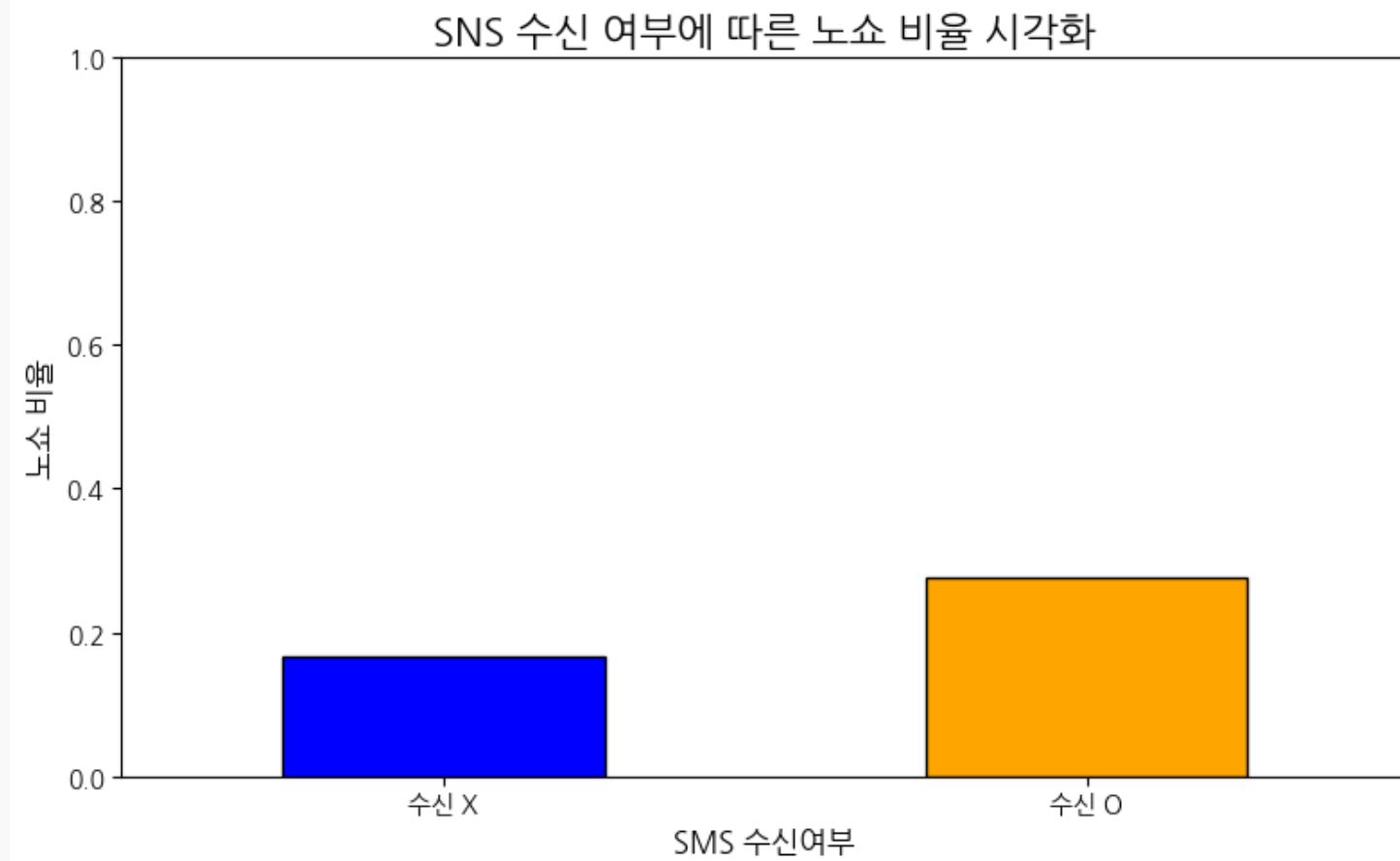
### | 분석결과 |

- 결론
- 예약간격 관리가 노쇼 감소에 중요

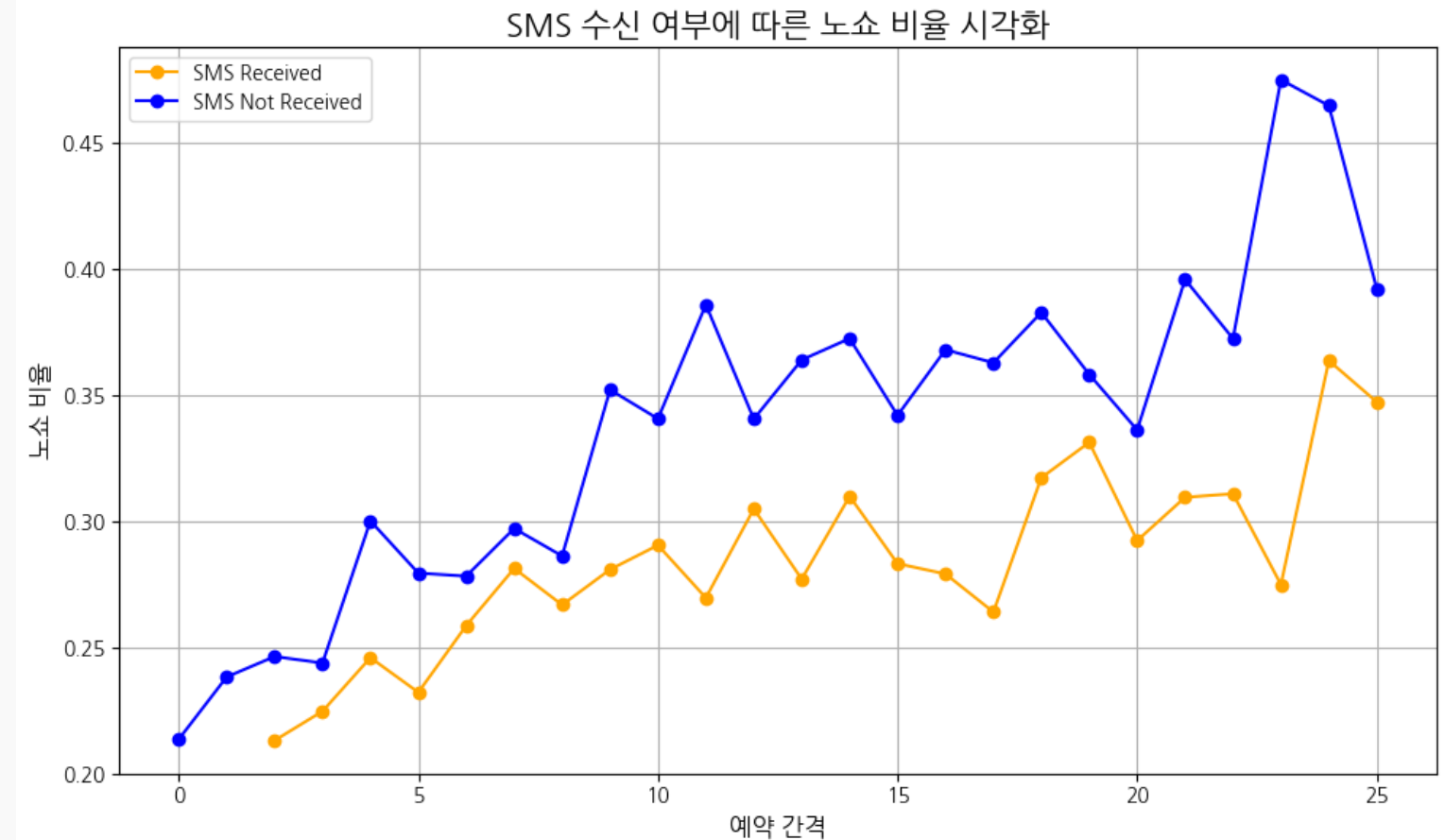


### 3. SMS 수신여부 분석

#### | SMS 수신여부에 따른 노쇼 비율 시각화 |



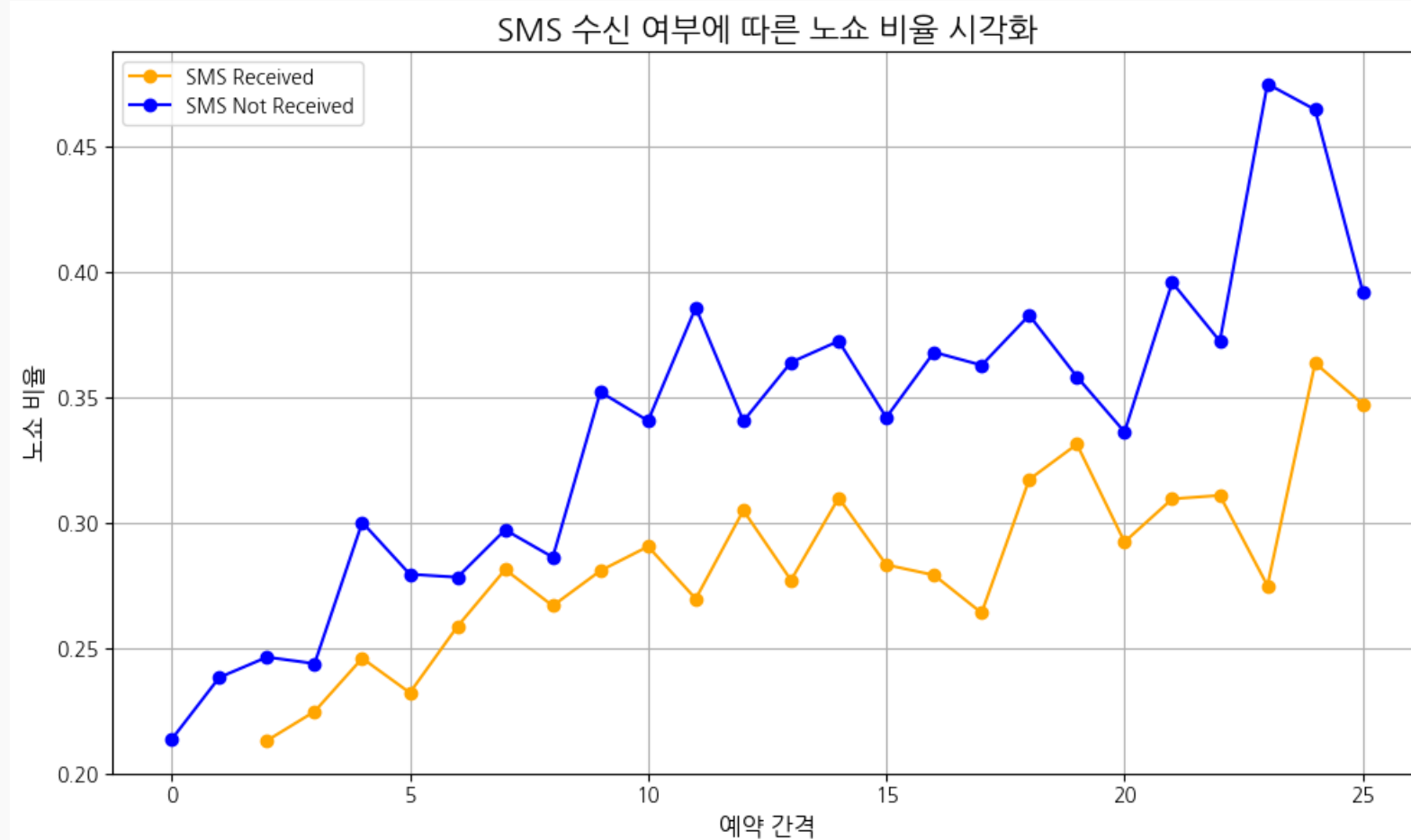
#### | 예약 간격에 따른 노쇼 비율 시각화 |





### 3. SMS 수신여부 분석

#### | SMS 수신 여부에 따른 예약 수 시각화 |



#### | 발견된 이슈 |

- 분석내용
- 표본이 많은 0~25까지 예약간격 구간 한정
  - SMS 수신자가 더 낮은 노쇼 비율을 나타냄

#### | 분석결과 |

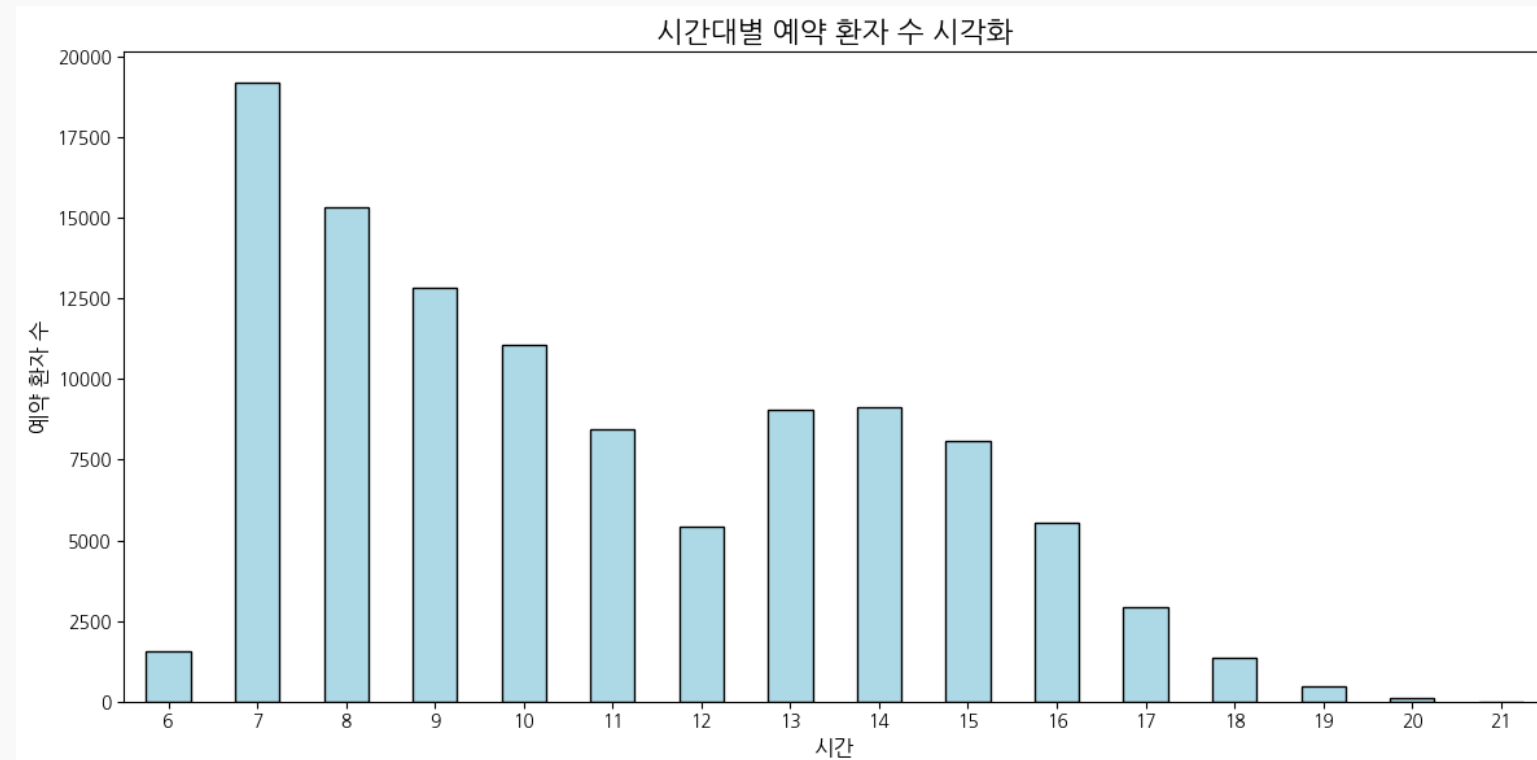
- 결론
- SMS 수신이 노쇼 비율을 낮추는데 기여
  - SMS 알림 활용 확대 방안 필요



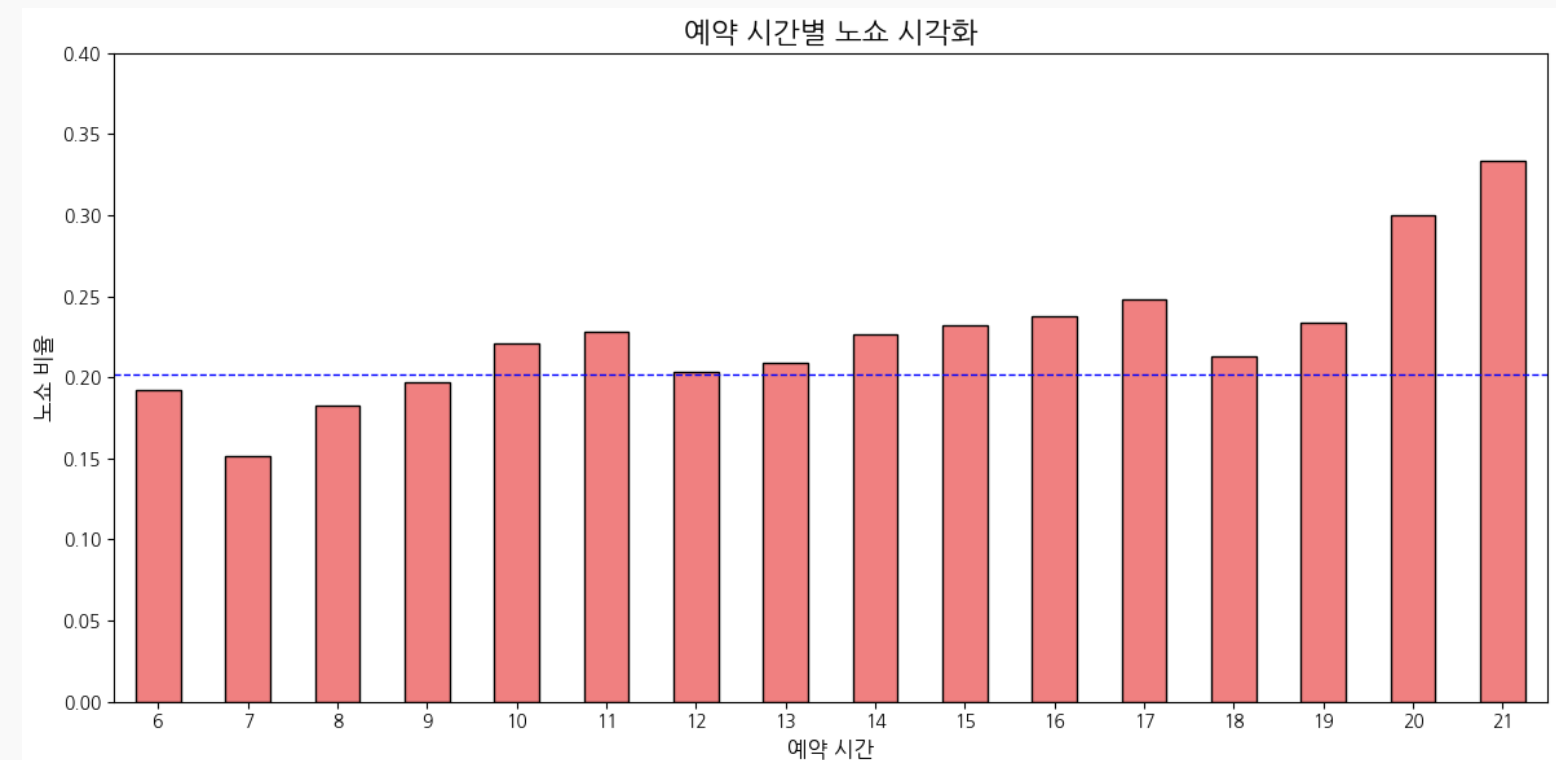


## 4. 예약시간대 분석

### | 예약 시간에 따른 예약자 수 시각화 |



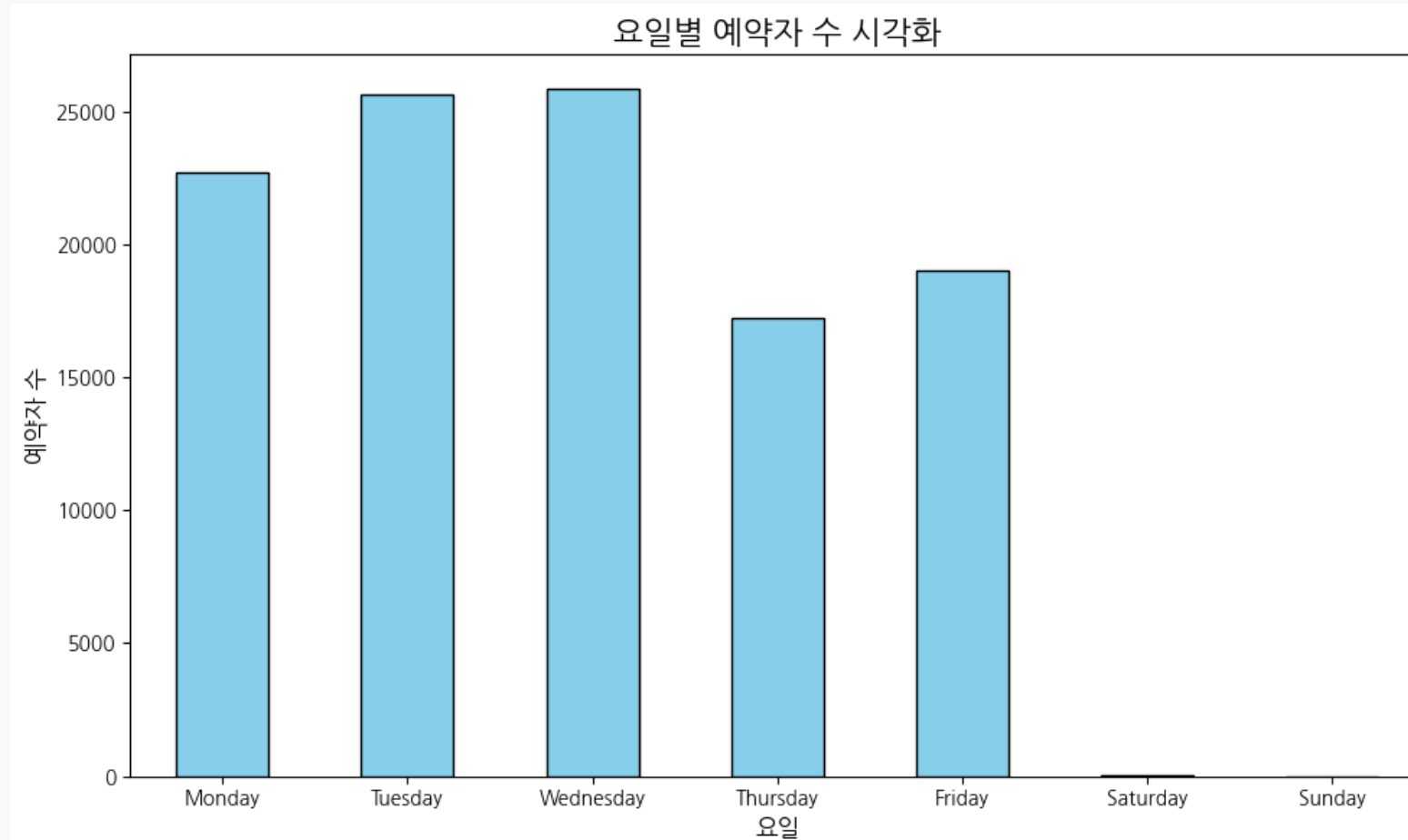
### | 예약 시간에 따른 노쇼 비율 시각화 |





## 4. 예약 시간대 분석

### | 요일별 예약 수 시각화 |



### | 발견된 이슈 |

- 분석내용
- 7~8시에 가장 많은 환자들이 예약
  - 그러나 해당 시간에 가장 적은 노쇼 비율을 보임

### | 분석결과 |

- 결론
- 7~8시 예약 확대
  - 주말에도 특정 시간대 예약 검토



## 5. NO-SHOW 예측 모델 학습

### 사용 모델

- Random Forest 모델 사용
- No-Show 데이터 불균형 해소를 위해 언더 샘플링 기법 사용
- 나이, 성별, 지역, 지병여부 등 데이터를 학습

### 학습 평가

- ROC-AUC 지표를 평가 기준으로 사용
- 모델의 AOC-AUC 점수는 0.7320
- 뛰어나진 않지만 준수한 수준의 성능을 지님

ROC-AUC: 0.7320



의료 데이터 분석 경진대회



들어주셔서 감사합니다.

THANK YOU.