

# Subway Digital Poster Ad



# Data

# Seoul Metro - Passenger Boarding and Alighting Data by Station, Date, Time Slot, and Passenger Type (Lines 1 to 8, June 2022 to May 2024)

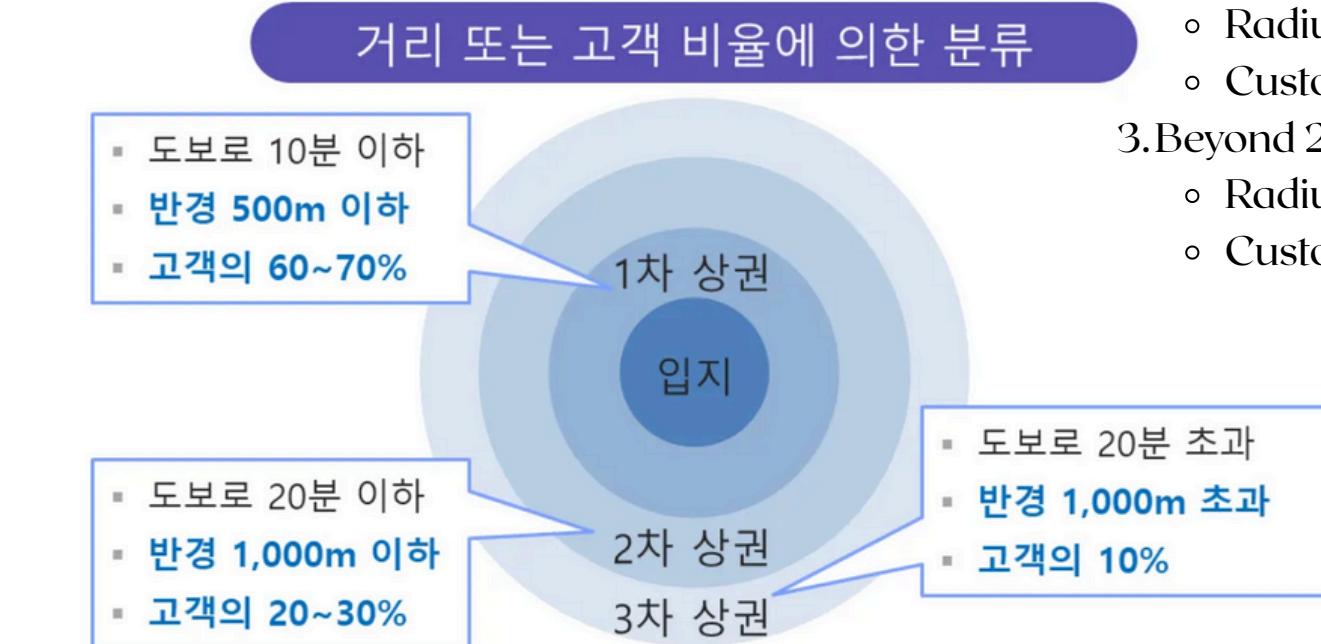
수송일자	호선명	역번호	역명	승하차구분	승객유형	06-07시간대	06-08시간대	07-09시간대	08-09시간대	09-10시간대	10-11시간대	11-12시간대	12-13시간대	13-14시간대	14-15시간대	15-16시간대	16-17시간대	17-18시간대	18-19시간대	19-20시간대	20-21시간대	21-22시간대	22-23시간대	23-24시간대	24시간대이후	
						시간대전	시간대	시간대	시간대	시간대	간대	간대														
0	2022-06-01	2.0	239	홍대입구	승차	일반	1159	632	532	703	876	936	1178	1676	1968	2347	3045	3880	4048	3889	3697	4725	5441	5282	2126	8.0
1	2022-06-01	2.0	239	홍대입구	승차	어린이	1	0	2	3	5	1	20	29	52	63	101	99	97	67	53	30	18	6	0	0.0
2	2022-06-01	2.0	239	홍대입구	승차	중고생	0	1	0	0	4	0	0	0	0	2	2	2	0	6	2	1	2	1	1	0.0

# Subway Digital Poster Ad

## Reference - Mobile Population Data for Customer Type Classification

Using Seoul Commercial Analysis Data

### 분석 리포트 Analysis Report



Criteria for Mobile Population Data in Commercial Analysis

1. Within 10 minutes walking distance
  - Radius: Less than 500m
  - Customers: 60–70%
2. Within 20 minutes walking distance
  - Radius: Less than 1,000m
  - Customers: 20–30%
3. Beyond 20 minutes walking distance
  - Radius: Over 1,000m
  - Customers: 10%

# Subway Digital Poster Ad

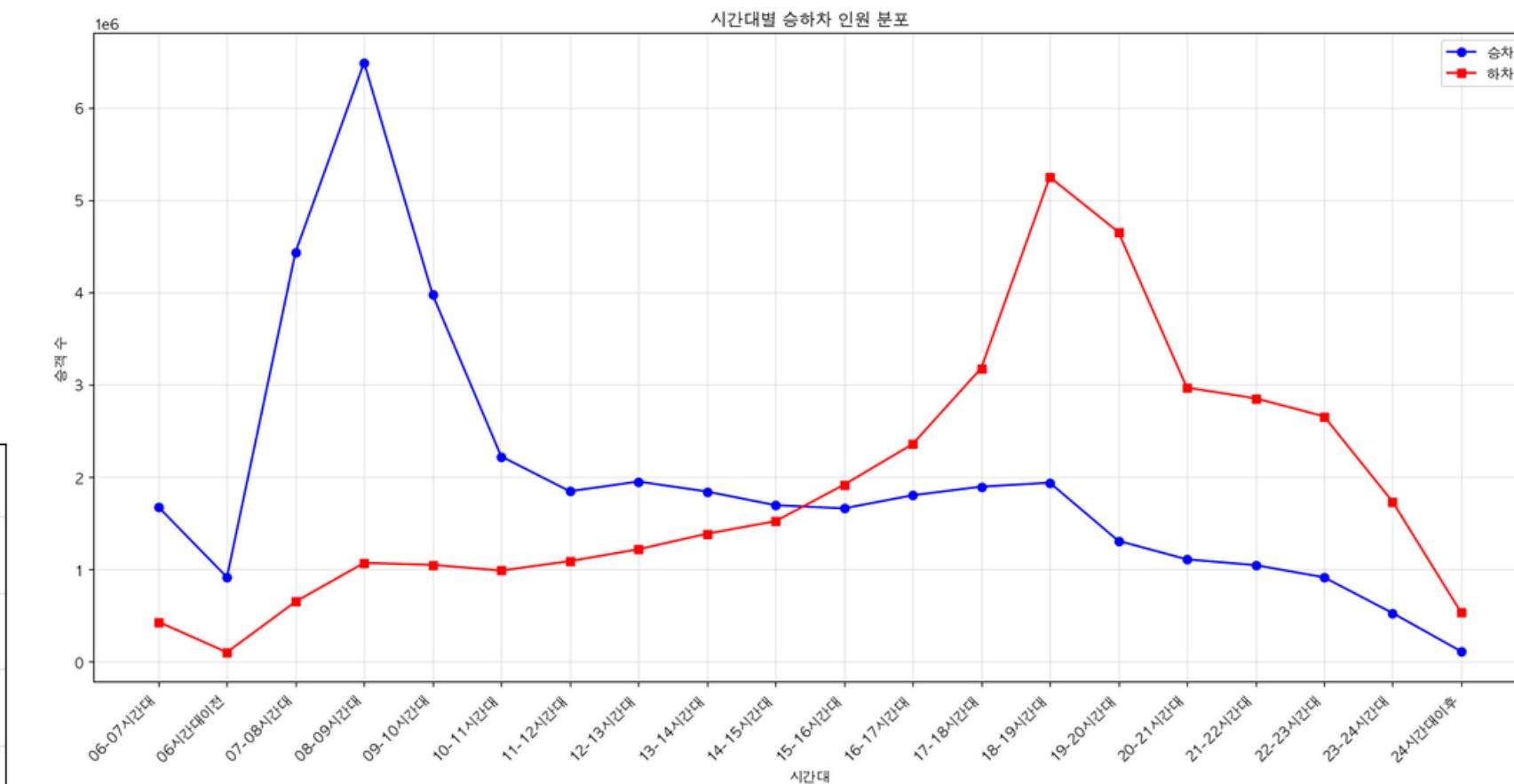
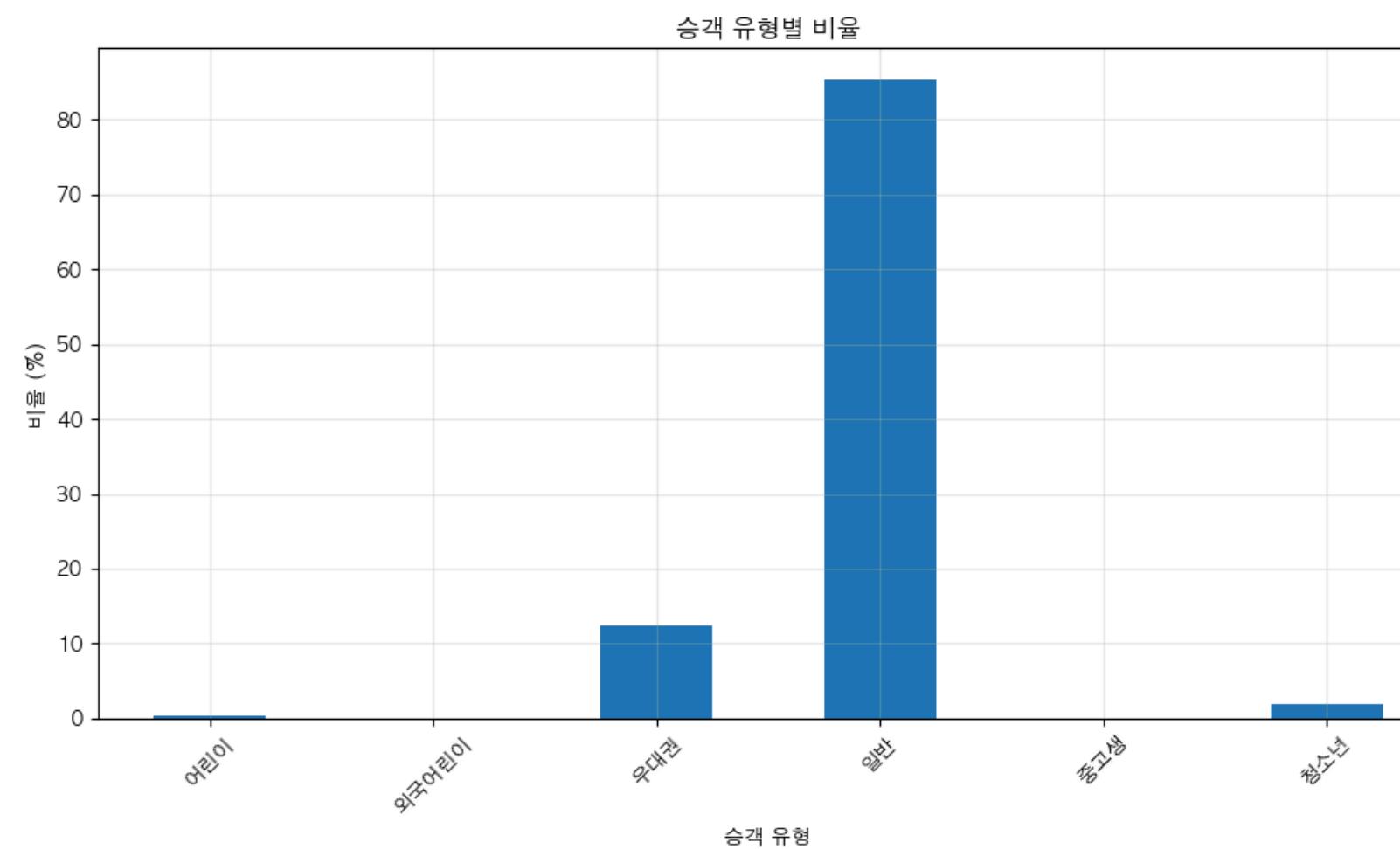
## Reference Paper

석사학위논문
기계학습을 이용한 서울 지하철 승하차 인원예측
Predicting the number of people getting on and off the subway using machine learning

	p-value
Day_of_the_week	0.000000
National_Holiday	0.000000
Season	0.79930
Temp(Avg)	0.60458
Temp(Min)	0.69456
Temp(Max)	0.56700
Prec	0.61691
Line_Cnt	0.000000
Bus_Terminal	0.000000
Bus_Cnt	0.000000

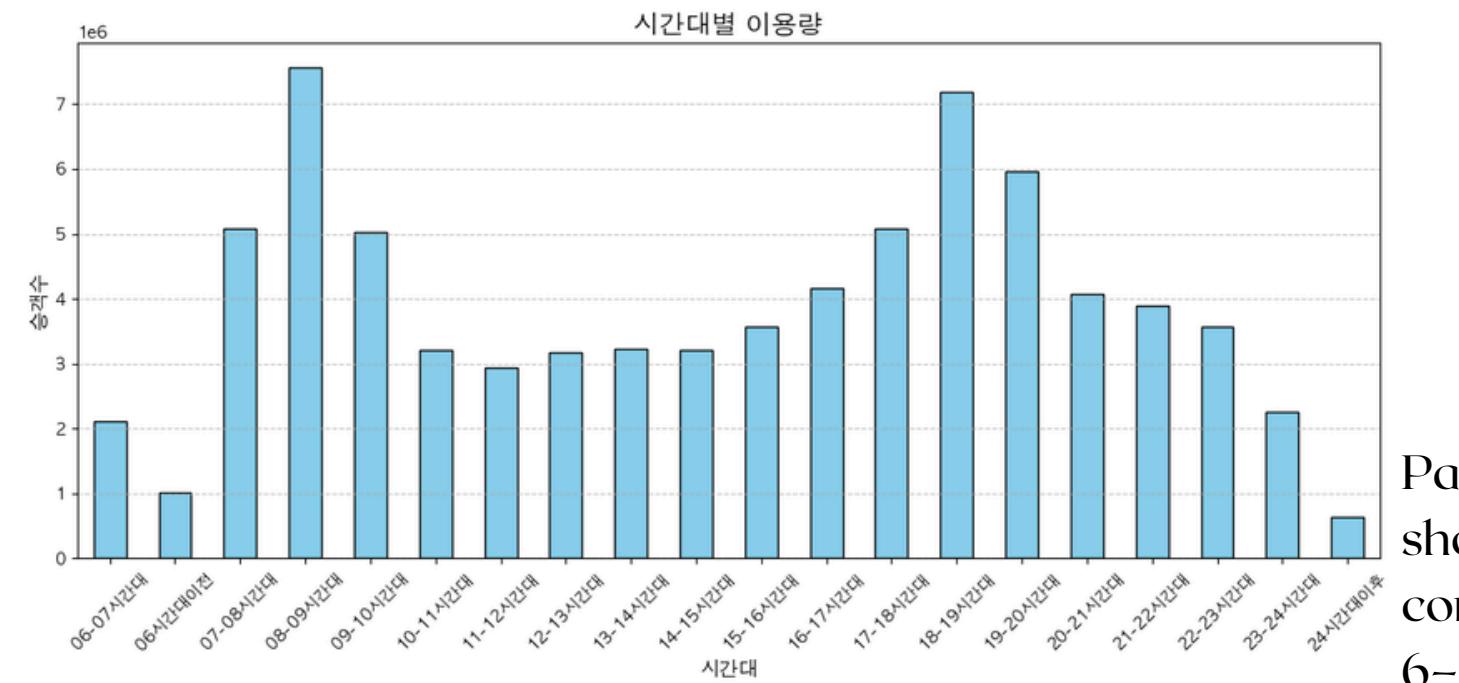
Season, temperature, and precipitation show no significant correlation, while day of the week and holiday status are correlated. (Values that do not change are excluded as the analysis is based on a single station.)

# Subway Digital Poster Ad

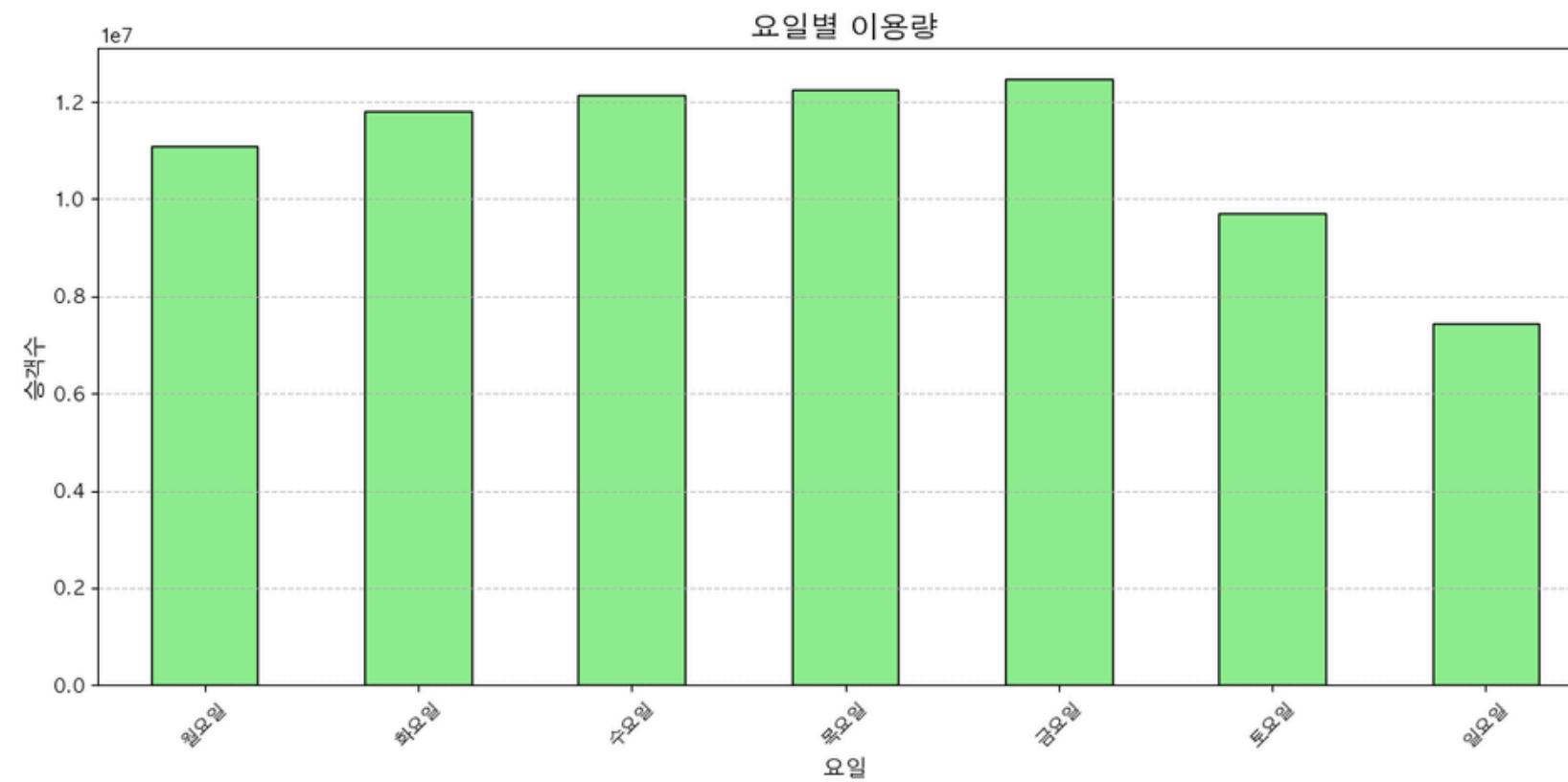


Passenger patterns by time of day show distinct peaks during commuting hours (morning and evening), with general passenger types being overwhelmingly dominant

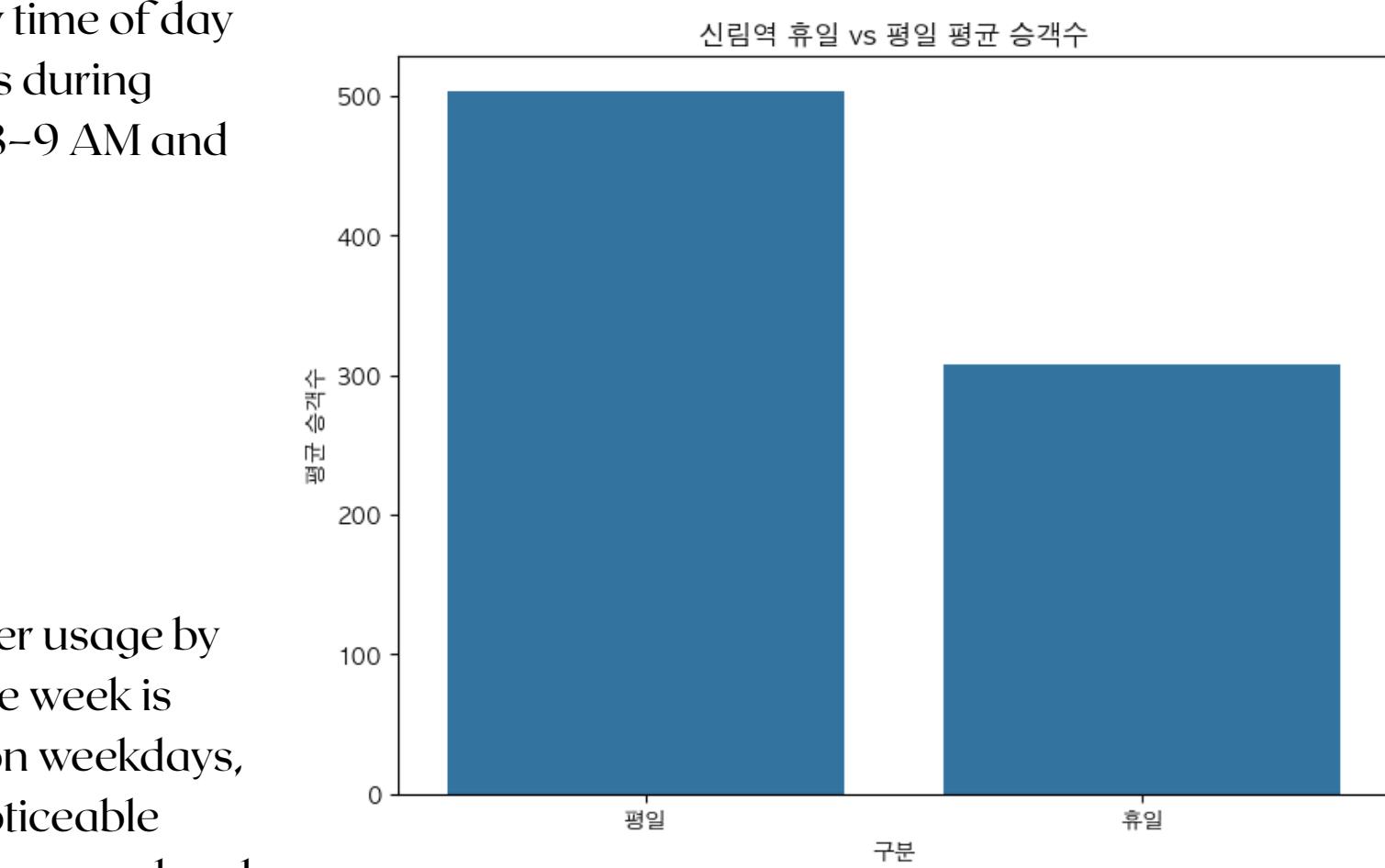
# Subway Digital Poster Ad



Passenger usage by time of day shows distinct peaks during commuting hours (8–9 AM and 6–7 PM)



Passenger usage by day of the week is highest on weekdays, with a noticeable decrease on weekends



At Sinlim Station, the average number of passengers on holidays is significantly lower than on weekdays

# Subway Digital Poster Ad

```
# 연령대 추정 함수
import random

def estimate_age_group_probabilities(hour):
    probabilities = {
        '20대~50대 직장인': 0,
        '20~30대 일반인': 0,
        '20대~60대 일반인 및 주부': 0,
        '20~30대 일반인 및 직장인': 0,
        '기타': 0
    }

    if 7 <= hour < 10: # 7시부터 10시 전까지 (출근 시간 확장)
        probabilities['20대~50대 직장인'] = 0.7
        probabilities['20~30대 일반인'] = 0.09
        probabilities['20대~60대 일반인 및 주부'] = 0.05
        probabilities['20~30대 일반인 및 직장인'] = 0.11
        probabilities['기타'] = 0.05
    elif 18 <= hour < 20: # 18시부터 20시 전까지 (퇴근 시간)
        probabilities['20대~50대 직장인'] = 0.7
        probabilities['20~30대 일반인'] = 0.09
        probabilities['20대~60대 일반인 및 주부'] = 0.05
        probabilities['20~30대 일반인 및 직장인'] = 0.11
        probabilities['기타'] = 0.05
    elif hour >= 21 or hour < 5:
        probabilities['20~30대 일반인'] = 0.6
        probabilities['20대~50대 직장인'] = 0.2
        probabilities['20~30대 일반인 및 직장인'] = 0.1
        probabilities['기타'] = 0.1
    elif 10 <= hour < 16: # 10시부터 16시 전까지 (주간 시간대 조정)
        probabilities['20대~60대 일반인 및 주부'] = 0.5
        probabilities['20대~50대 직장인'] = 0.2
        probabilities['20~30대 일반인'] = 0.2
        probabilities['기타'] = 0.1
    elif 16 <= hour < 21:
        probabilities['20~30대 일반인 및 직장인'] = 0.4
        probabilities['20대~50대 직장인'] = 0.3
        probabilities['20~30대 일반인'] = 0.2
        probabilities['20대~60대 일반인 및 주부'] = 0.05
        probabilities['기타'] = 0.05
    else:
        probabilities['기타'] = 1.0

    return probabilities
```

Left Image (Python Function)

- “This Python function estimates the probabilities of different passenger age groups based on the time of day. It categorizes passengers into groups such as workers, general adults, students, and others.”

Right Image (Pie Chart)

- “This pie chart illustrates the proportion of categorized passenger types, including workers aged 20–50, general adults aged 20–60, students, priority passengers, and others.”



# Subway Digital Poster Ad

	수송일자	호선명	역번호	역명	승하차구분	승객유형	월	년	시간대	승객수	일	휴일여부	요일	요일_문자	시간	추정연령대
0	2022-06-01	2.0	230	신림	승차	일반	6	2022	06시간대이전	801.0	1	False	2	수	6	기타
1	2022-06-01	2.0	230	신림	승차	어린이	6	2022	06시간대이전	0.0	1	False	2	수	6	기타
2	2022-06-01	2.0	230	신림	승차	중고생	6	2022	06시간대이전	0.0	1	False	2	수	6	기타
3	2022-06-01	2.0	230	신림	승차	청소년	6	2022	06시간대이전	9.0	1	False	2	수	6	기타
4	2022-06-01	2.0	230	신림	승차	우대권	6	2022	06시간대이전	246.0	1	False	2	수	6	기타
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
155515	2024-05-31	2.0	230	신림	하차	어린이	5	2024	24시간대이후	0.0	31	False	4	금	24	20~30대 일반인
155516	2024-05-31	2.0	230	신림	하차	중고생	5	2024	24시간대이후	0.0	31	False	4	금	24	20~30대 일반인
155517	2024-05-31	2.0	230	신림	하차	청소년	5	2024	24시간대이후	16.0	31	False	4	금	24	20~30대 일반인
155518	2024-05-31	2.0	230	신림	하차	우대권	5	2024	24시간대이후	24.0	31	False	4	금	24	20~30대 일반인
155519	2024-05-31	2.0	230	신림	하차	외국어린이	5	2024	24시간대이후	0.0	31	False	4	금	24	20~30대 일반인

155520 rows × 16 columns

```
# 기존 승객유형별 나이대 매핑
age_group_mapping = {
    '중고생': '10대',
    '어린이': '10대',
    '청소년': '10대',
    '우대권': '50대 이상',
    '외국어린이': '10대'
}
```

```
# 일반 승객 나이대 비율
base_age_group_ratios = {
    '20대': 0.25,
    '30대': 0.35,
    '40대': 0.25,
    '50대 이상': 0.15
}
```

```
# 나이대별 승객 수 계산 및 데이터프레임에 추가
age_group_columns = ['10대', '20대', '30대', '40대', '50대 이상']
```

## Top Table (Passenger Data Issue)

- The table shows passenger data, but anomalies were detected in the age group column (highlighted in red).
- This issue requires further investigation to ensure data accuracy.

## Age Group Mapping (Left Python Code)

- The Python code maps passenger types to specific age groups, such as teenagers, children, and priority passengers.
- Mapping ensures consistency in age group classification.

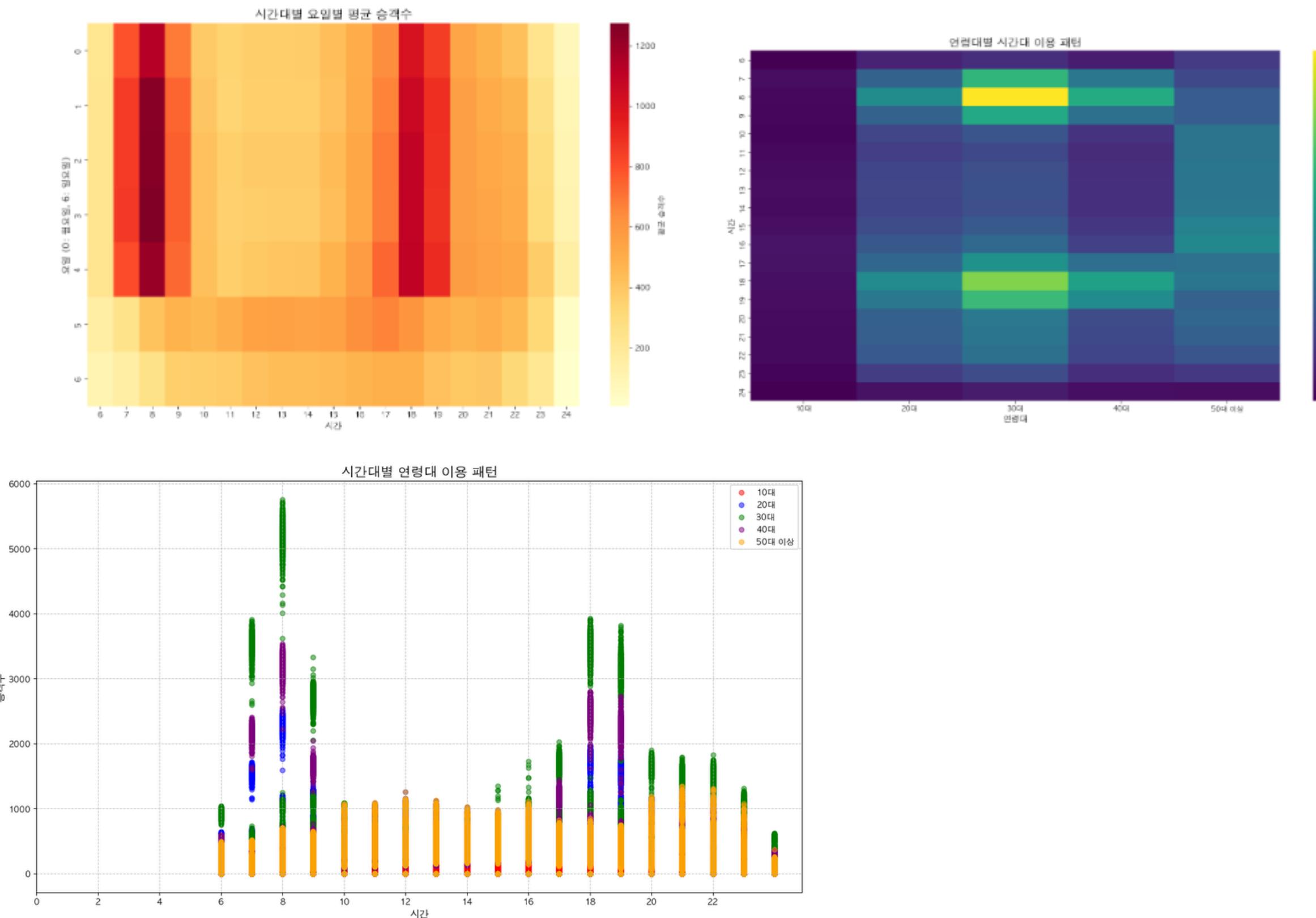
## Base Age Group Ratios (Middle Python Code)

- This code defines the ratio of general passengers by age group, with 30s accounting for the largest proportion (35%).

## Age Group Columns (Right Python Code)

- Age group columns are added to the data frame for calculating passenger counts by age group
-

# Subway Digital Poster Ad



Left Heatmap (Average Passenger Count by Hour and Day)

- “This heatmap shows the average number of passengers by hour and day of the week. Peak usage occurs during commuting hours (morning and evening).”

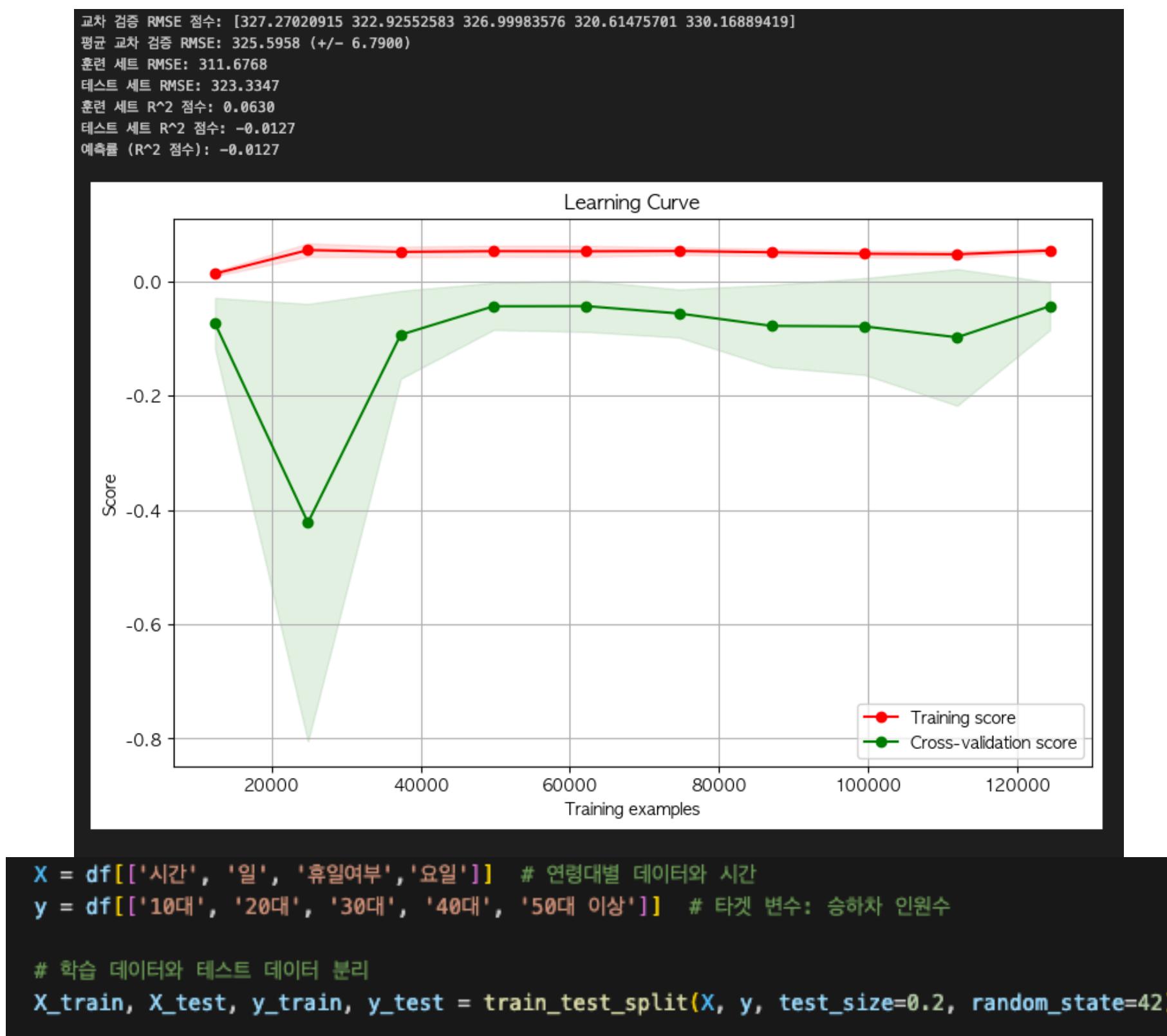
Right Heatmap (Passenger Usage by Age Group and Time Slot)

- “This heatmap illustrates passenger usage patterns by age group and time slot. The 30s age group is concentrated during commuting hours.”

Bottom Scatter Plot (Passenger Patterns by Age Group and Time)

- “This scatter plot visualizes passenger patterns by age group and time of day, with distinct peaks for the 30s age group during morning and evening commuting hours.”

# Subway Digital Poster Ad



## Analysis Results

- “The analysis results indicate that Sinlim Station lacks distinct features in its data, leading to the decision to exclude it from the service.”

## Learning Curve (Graph)

- “The learning curve shows training and cross-validation scores over increasing training examples. The model achieves stable performance but exhibits slight overfitting.”

## Dataset and Variables (Python Code)

- “Features include time, day, holiday status, and weekday, with target variables segmented by age groups (10s, 20s, 30s, 40s, 50s+). The split divides data into training and test sets with a 20% test size.”

## Business Characteristics (Table)

- “Business data for Sinlim Station includes 457 Korean restaurants, with 20 new openings, 29 closures, and an average operating period of 119 months.”