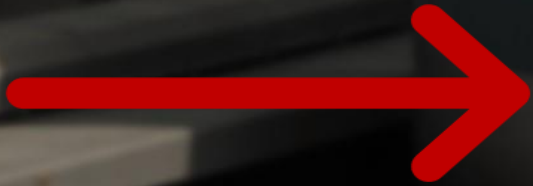


# 지하철 이용승객 EDA 프로젝트



2024.04.05.

이창민, 황유성, 이준석,  
맹의현, 진영인, 이에니

BETAGO

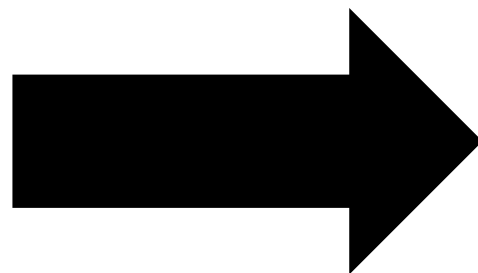
- 01 지하철 승객수가 많은 **일자**
- 02 지하철 승객수가 많은 **달**
- 03 지하철 승객수가 많은 **요일**
- 04 **월별**로 구한 **일자별** 승하차승객수
- 05 가장 승객이 많은 **승차역**
- 06 **노선별**로 구한 **역&요일별** 승차승객수
- 07 **1호선**에서 가장 **하차**를 많이 한 **역**

# Contents

# 발표 시작 전, EDA 정당성 설명

☑ 왜 숫자별 요일 매핑을 바꾸었는가

```
conv = {  
    0 : '일',  
    1 : '월',  
    2 : '화',  
    3 : '수',  
    4 : '목',  
    5 : '금',  
    6 : '토'  
}
```



```
conv = {  
    0 : '월',  
    1 : '화',  
    2 : '수',  
    3 : '목',  
    4 : '금',  
    5 : '토',  
    6 : '일'  
}
```

## 왜 숫자별 요일 매핑을 바꾸었는가

<div>오늘 &lt; 2019.01 &gt; <input type="checkbox"/> 음력 <input type="checkbox"/> 손없는날 <input checked="" type="checkbox"/> 기념일</div>						
일	월	화	수	목	금	토
30	31	1 신정	2	3	4	5

## ☑ 왜 숫자별 요일 매핑을 바꾸었는가

<div> <div>오늘 &lt; 2019.01 &gt;</div> <div> <input type="checkbox"/> 음력           <input type="checkbox"/> 손없는날           <input checked="" type="checkbox"/> 기념일         </div> </div>						
일	월	화	수	목	금	토
30	31	1 신정	2	3	4	5

### < 수정 전 >

사용일자	노선명	역명	승차총승객수	하차총승객수	등록일자	요일
2019-01-01	2호선	을지로4가	3862.0	3728.0	20190104	월

## ☑ 왜 숫자별 요일 매핑을 바꾸었는가

<div> <div>오늘 &lt; 2019.01 &gt;</div> <div> <input type="checkbox"/> 음력           <input type="checkbox"/> 손없는날           <input checked="" type="checkbox"/> 기념일         </div> </div>						
일	월	화	수	목	금	토
30	31	1 신정	2	3	4	5

### < 수정 후 >

사용일자	노선명	역명	승차총승객수	하차총승객수	등록일자	사용요일
2019-01-01	2호선	을지로4가	5268.0	5595.0	2019-01-04	화

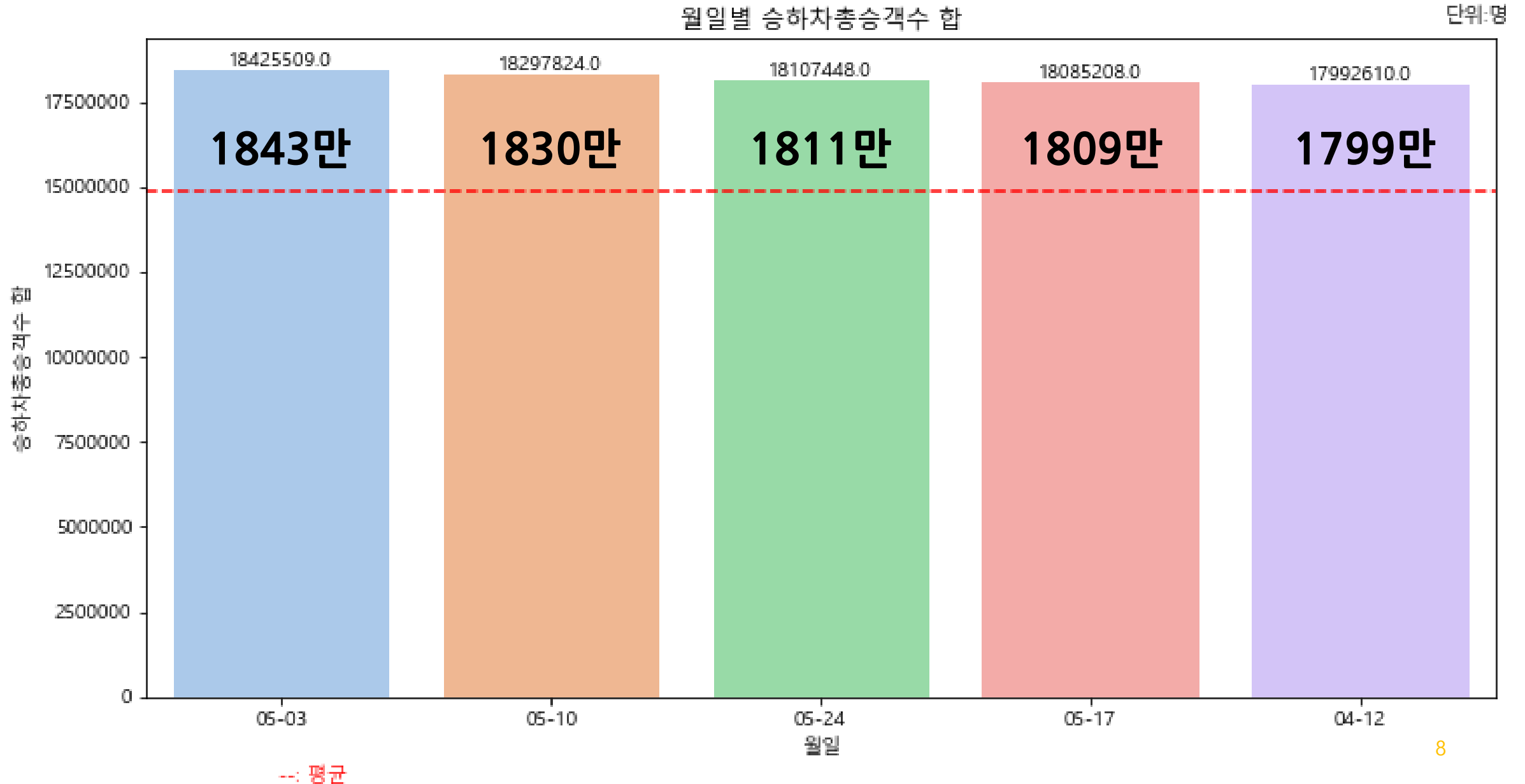
1. 2019.01~06 중에 언제 지하철을 가장 많이 이용했을까?( 기준: 승하차총승객수 )

5월 3일



1. 2019.01~06 중에 언제 지하철을 가장 많이 이용했을까?( 기준: 승하차총승객수 )

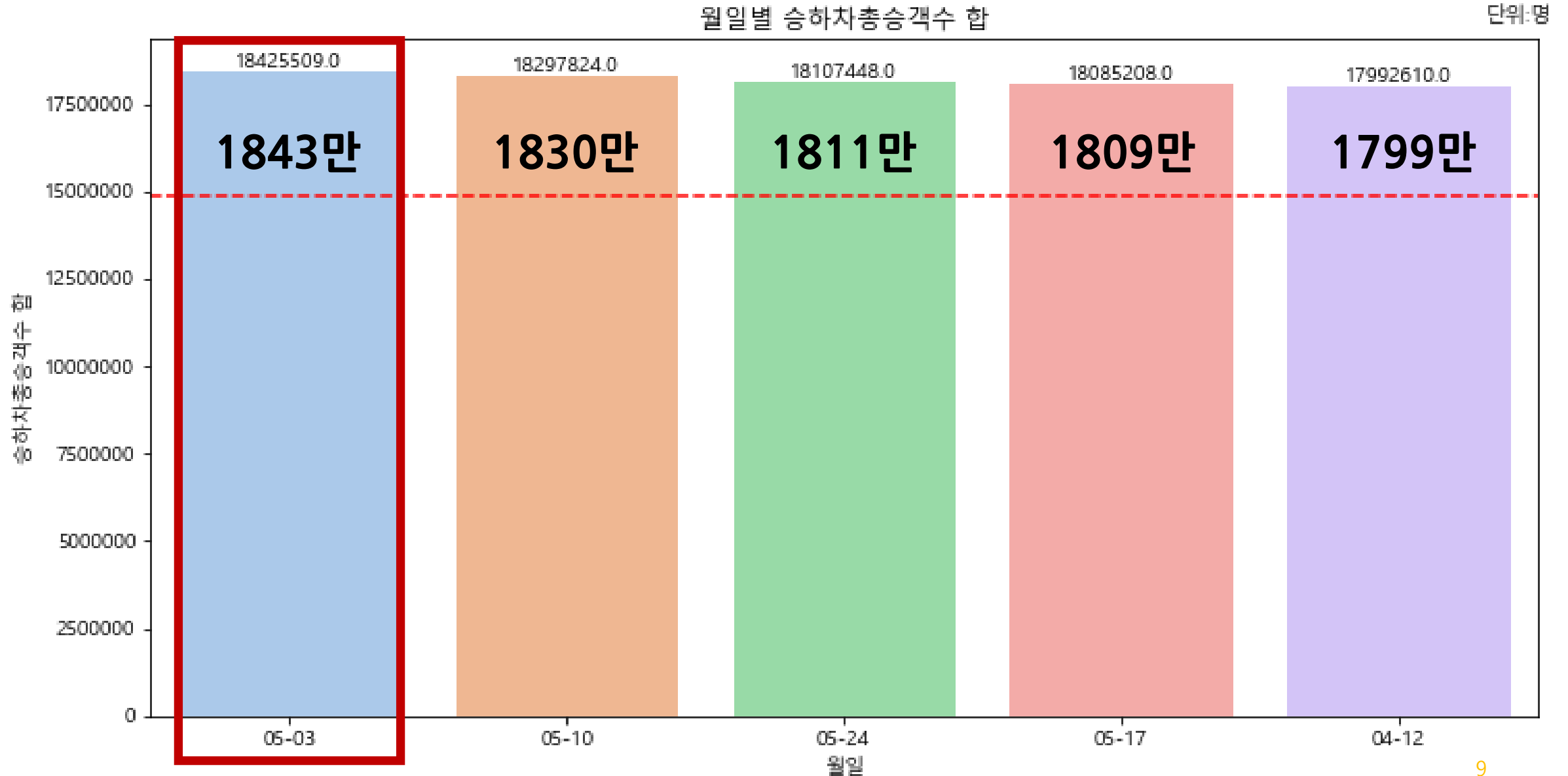
5월 3일





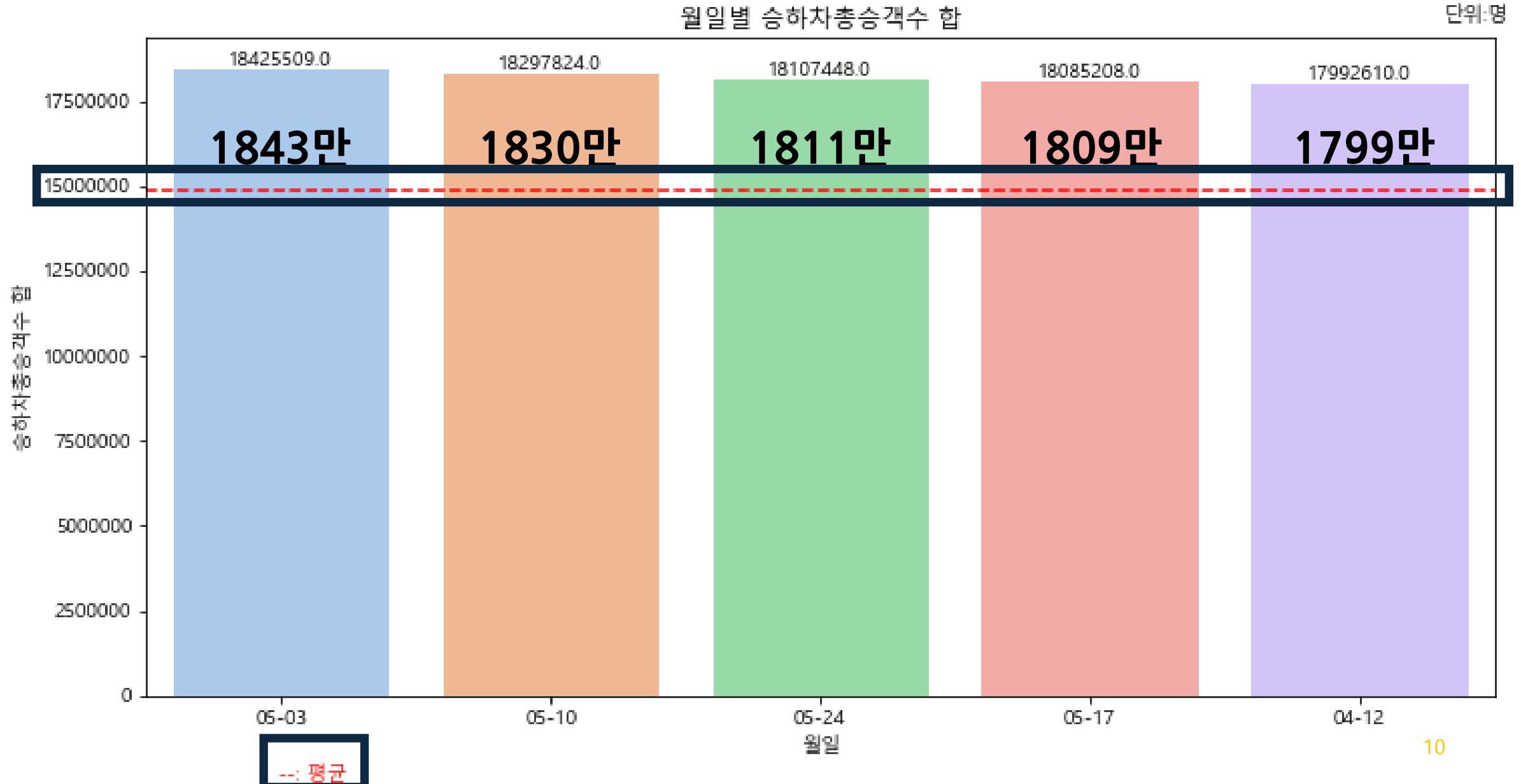
1. 2019.01~06 중에 언제 지하철을 가장 많이 이용했을까?( 기준: 승하차총승객수 )

5월 3일



1. 2019.01~06 중에 언제 지하철을 가장 많이 이용했을까?( 기준: 승하차총승객수 )

5월 3일



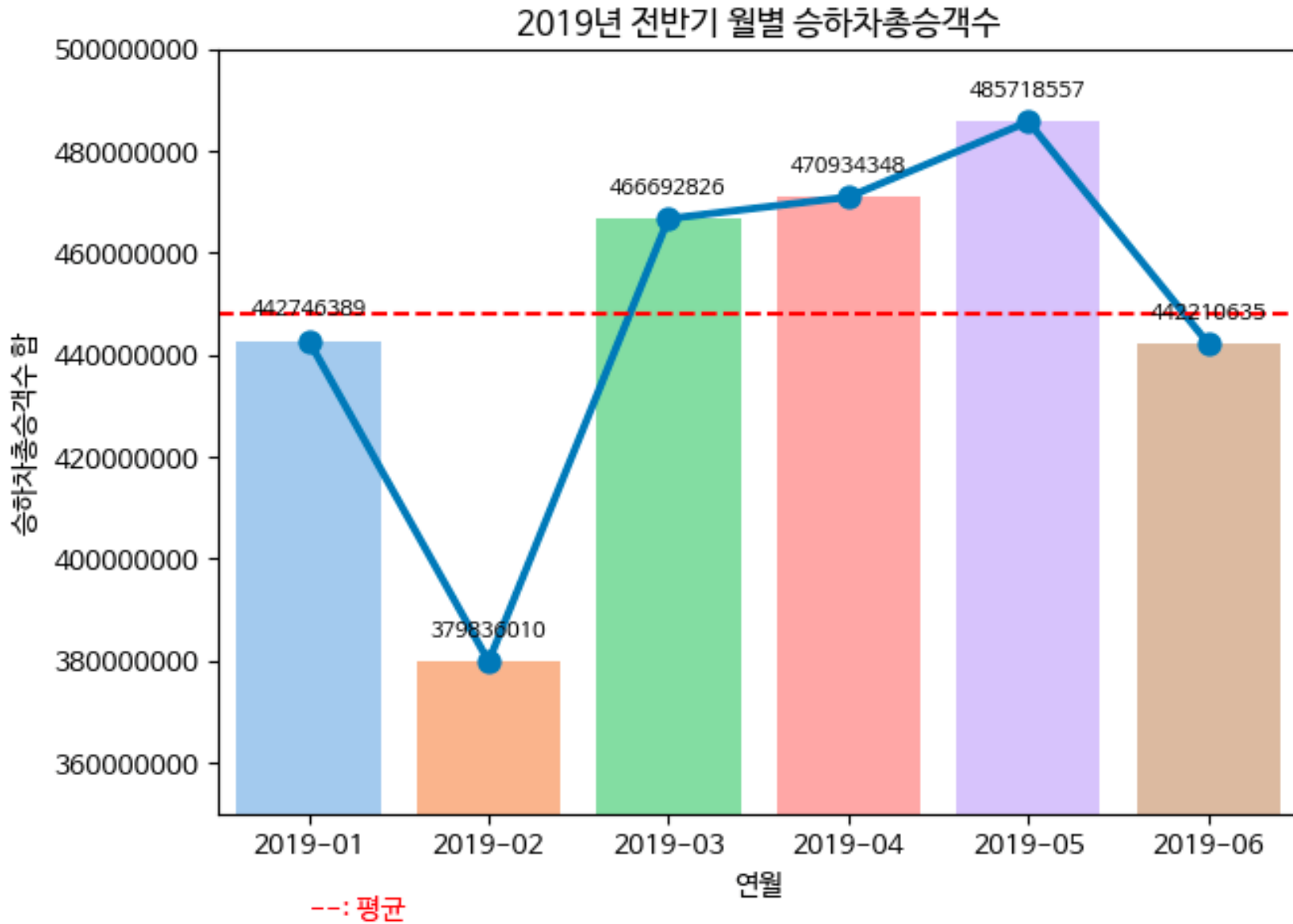
2. 1월~6월 중에 5월에 지하철 승객수가 많다? (기준: 승하차총승객수)

5월



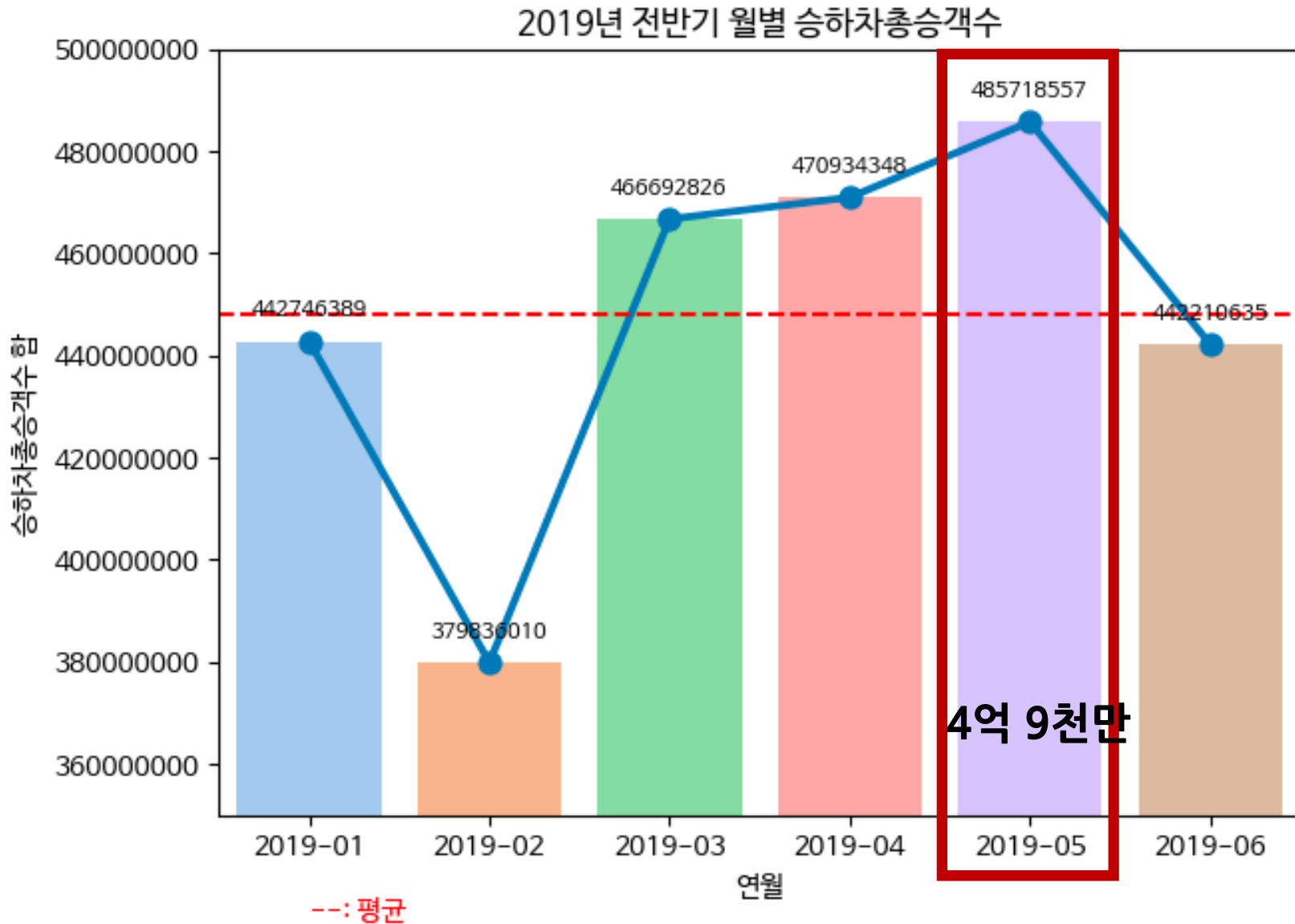
## 2. 1월~6월 중에 5월에 지하철 승객수가 많다? (기준: 승하차총승객수)

5월

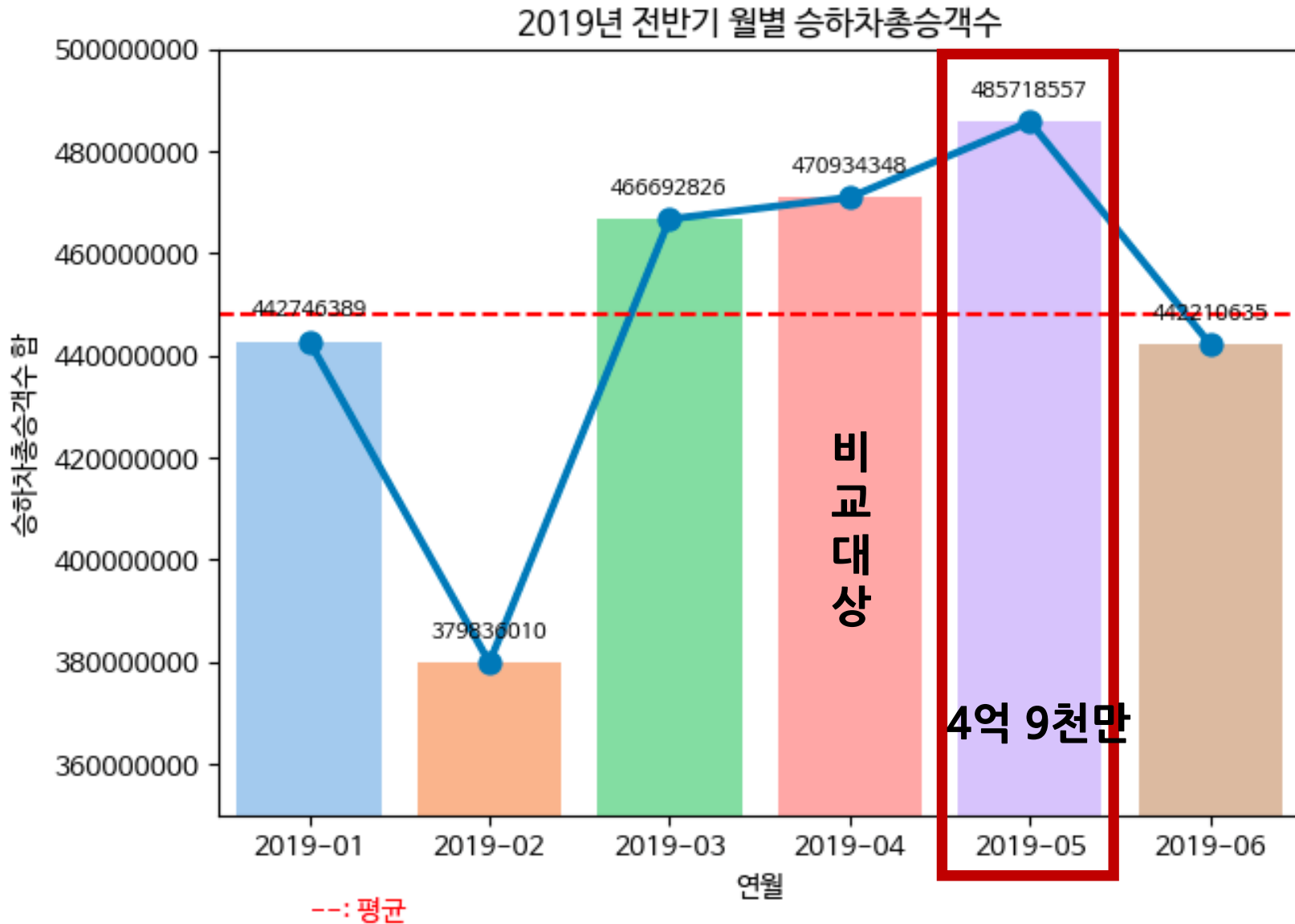


## 2. 1월~6월 중에 5월에 지하철 승객수가 많다? (기준: 승하차총승객수)

5월



2. 1월~6월 중에 5월에 지하철 승객수가 많다? (기준: 승하차총승객수)



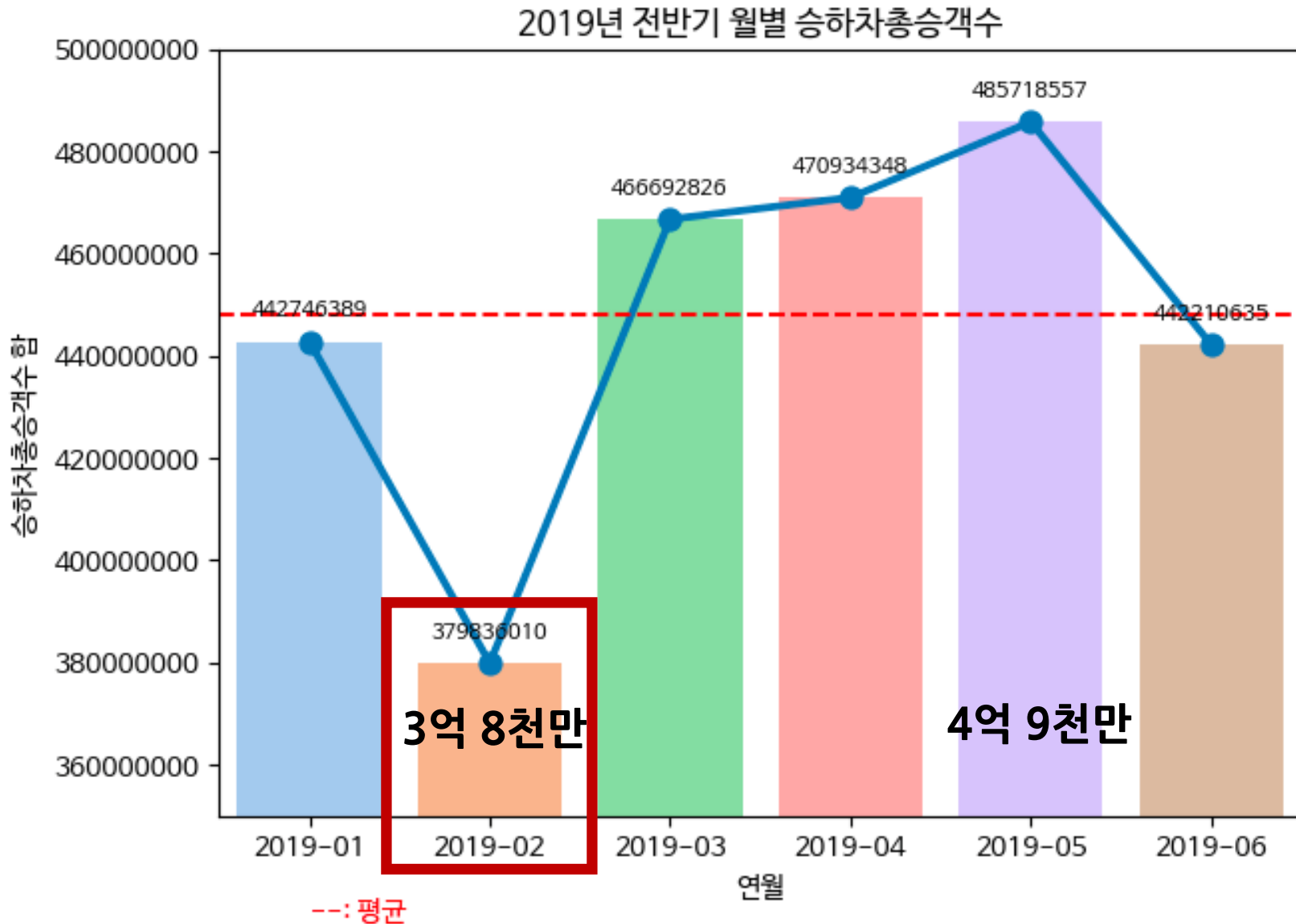
5월 승하차총승객수가  
4월 승하차총승객수보다

485,718,557 - 470,934,348

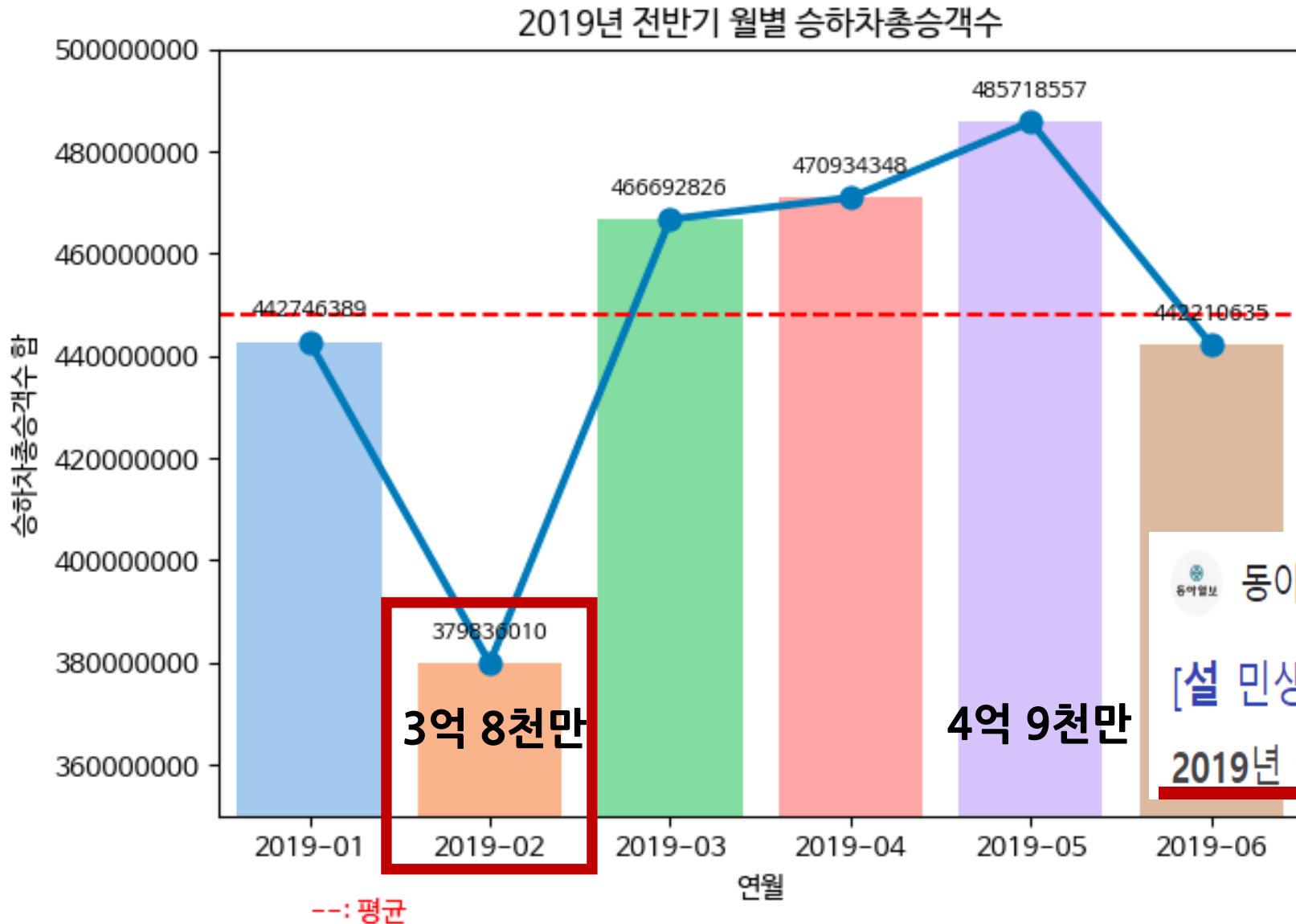
14,784,209 ↑

## 2. 1월~6월 중에 5월에 지하철 승객수가 많다? (기준: 승하차총승객수)

5월



## 2. 1월~6월 중에 5월에 지하철 승객수가 많다? (기준: 승하차총승객수)



달력		양음력변환		전역일계산		
오늘 < 2019.02 >		<input type="checkbox"/> 음력 <input type="checkbox"/> 손없는날 <input checked="" type="checkbox"/> 기념일				
일	월	화	수	목	금	토
27	28	29	30	31	1	2
3	4 입춘	5 설날	6 음 1.1	7	8	9
10	11	12	13	14 벨런타인...	15	16
17	18	19 음 1.15 우수 정월 대보름	20	21	22	23
24	25	26	27	28 2.28 민주...	1	2

동아일보 · 2019.01.22. · 네이버뉴스

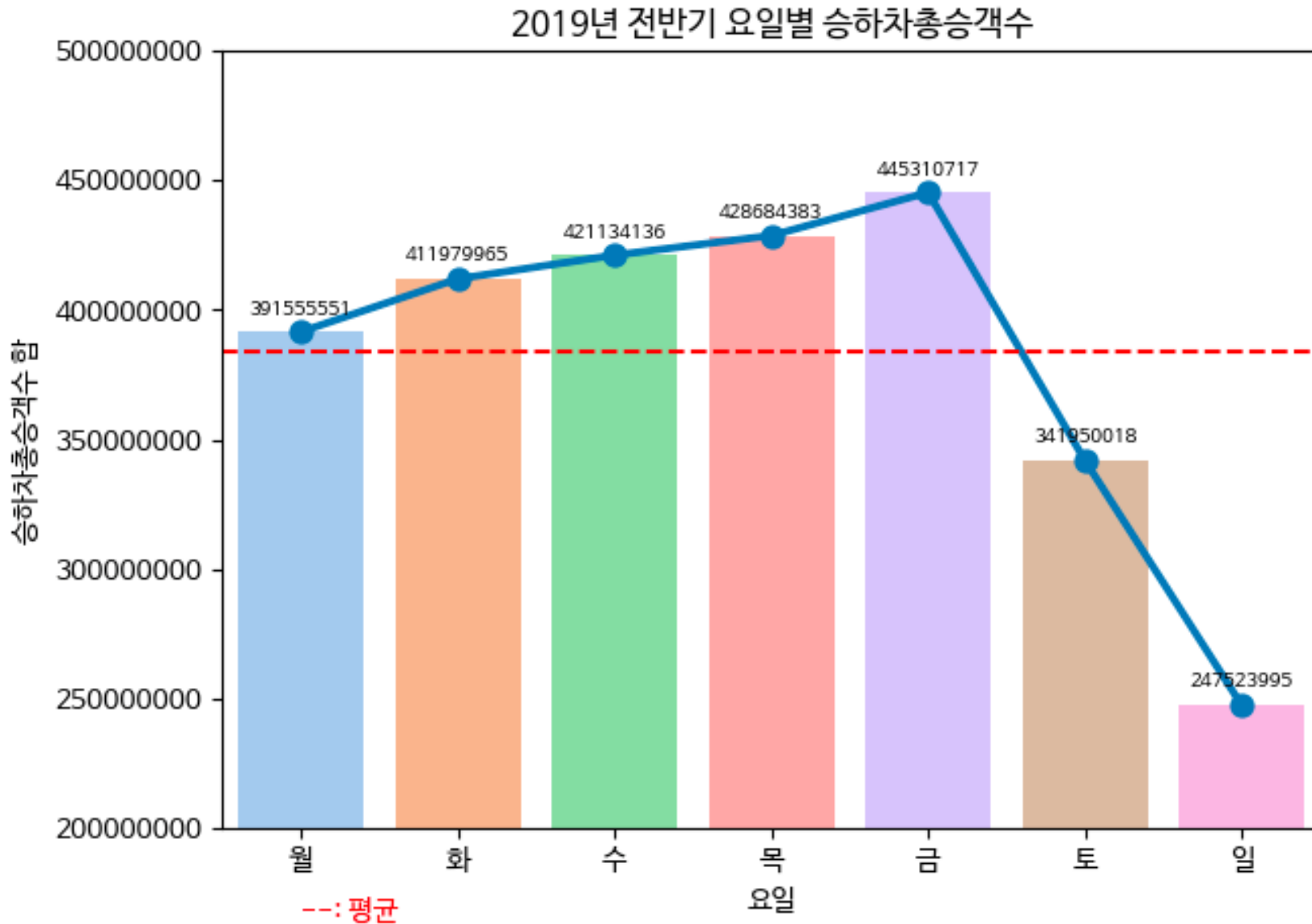
[설 민생안정대책] 고속도로 통행료 '공짜'

2019년 설 연휴기간 전국 고속도로 통행료가 면제되고,

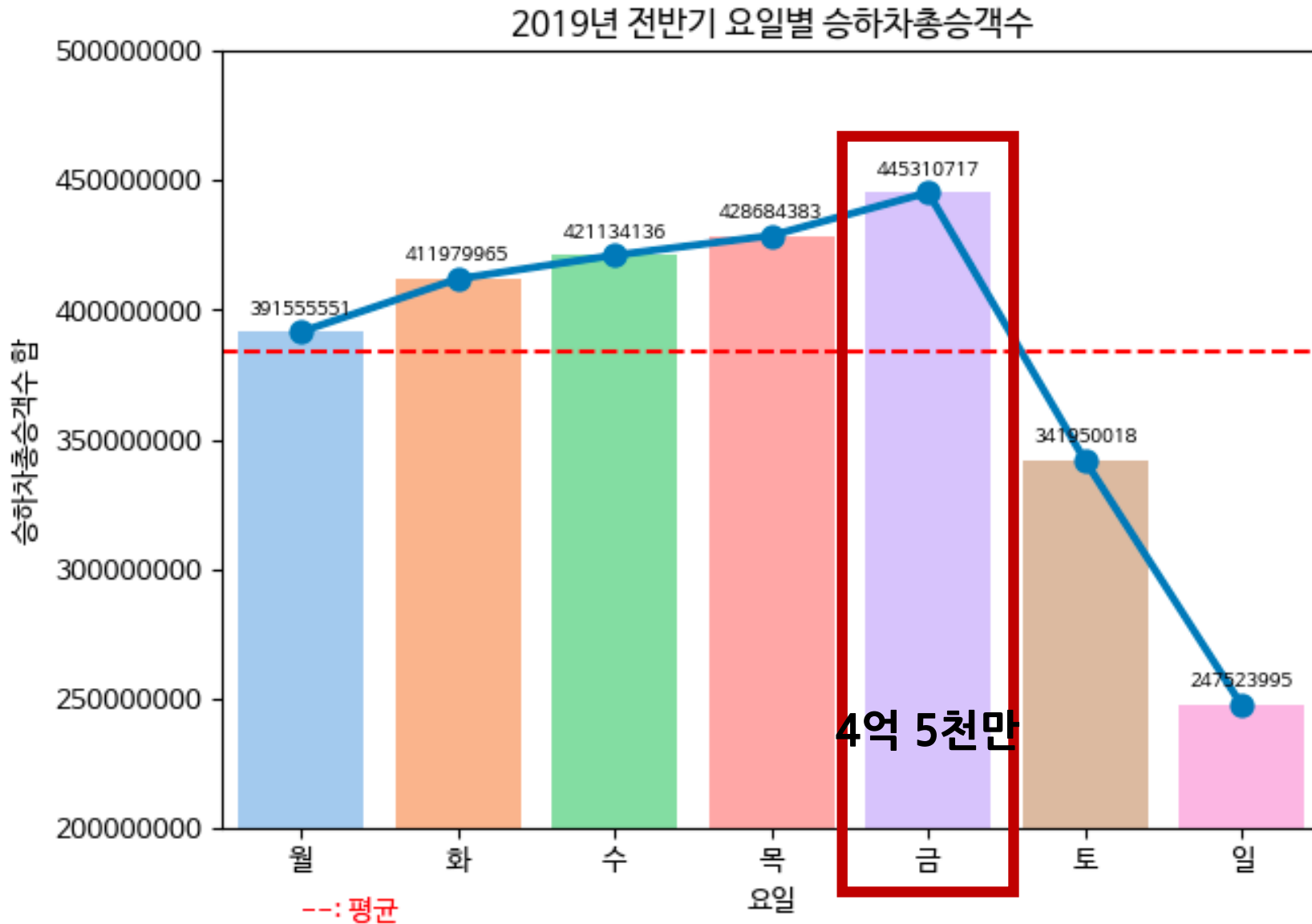


3. 요일 중에서 목요일에 지하철 승객수가 많다? (기준: 승하차총승객수)

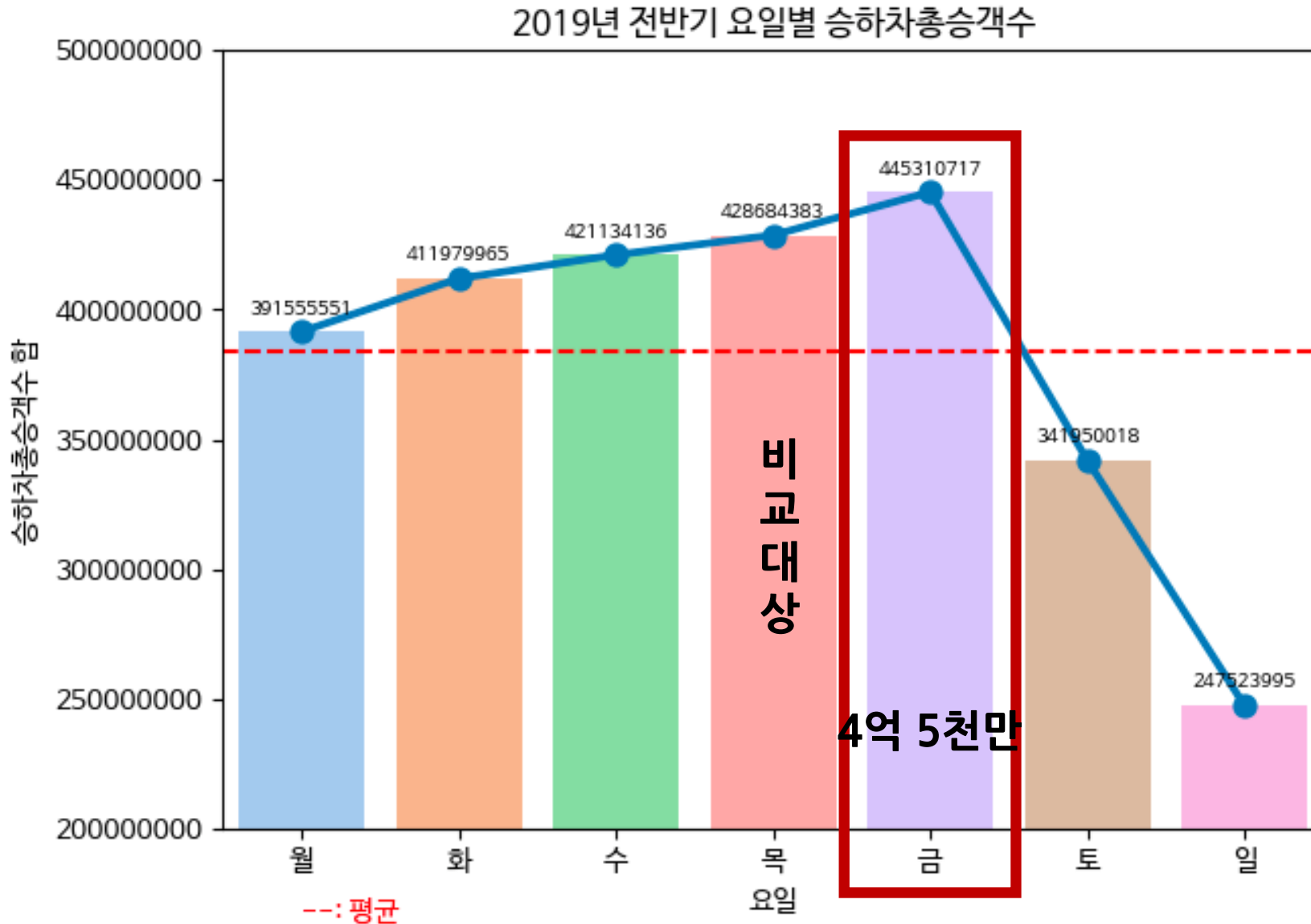
### 3. 요일 중에서 목요일에 지하철 승객수가 많다? (기준: 승하차총승객수)



### 3. 요일 중에서 목요일에 지하철 승객수가 많다? (기준: 승하차총승객수)



### 3. 요일 중에서 목요일에 지하철 승객수가 많다? (기준: 승하차총승객수)



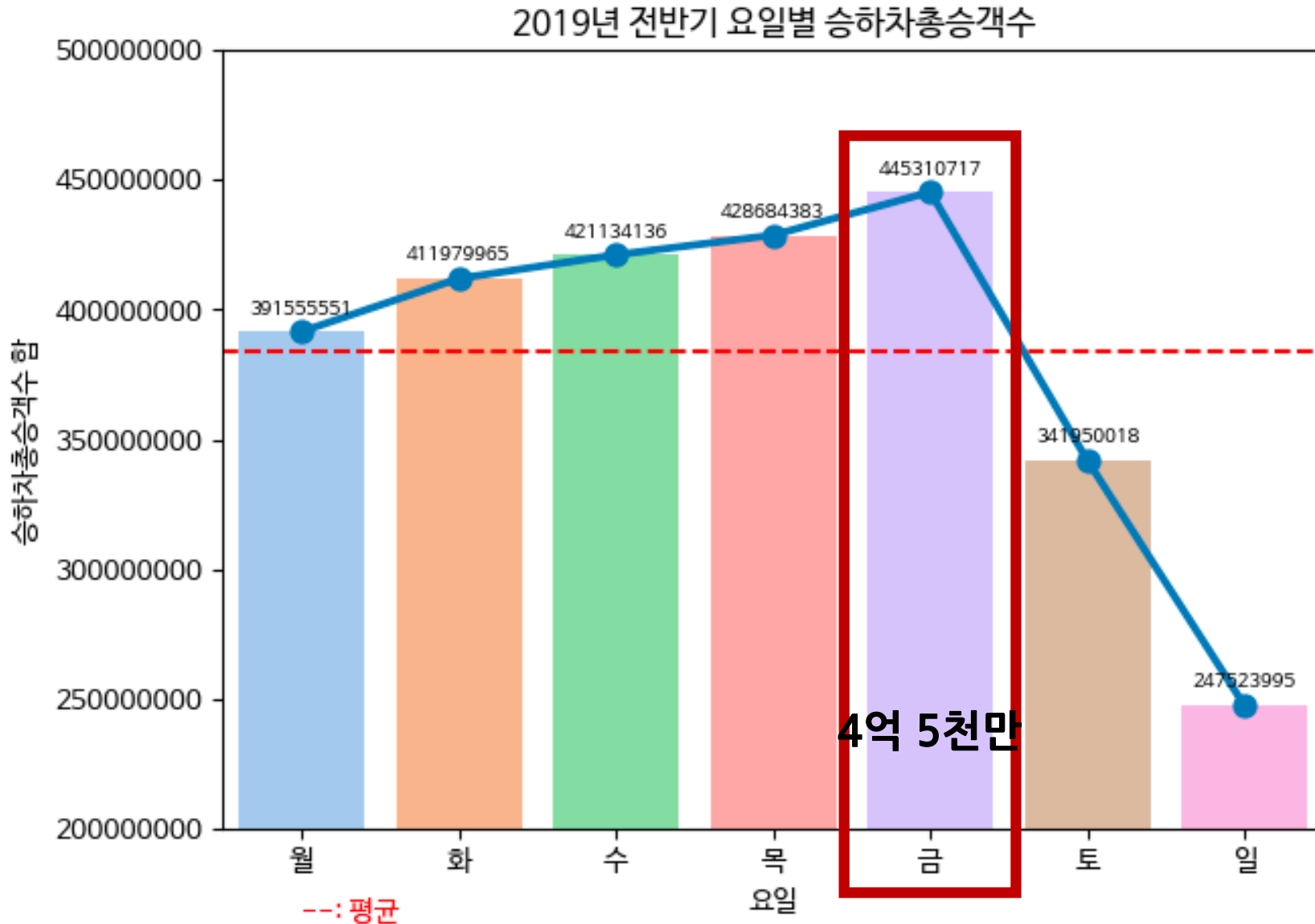
금요일 승하차총승객수가  
목요일 승하차총승객수보다

$445,310,717 - 428,684,383$

16,626,334



### 3. 요일 중에서 목요일에 지하철 승객수가 많다? (기준: 승하차총승객수)

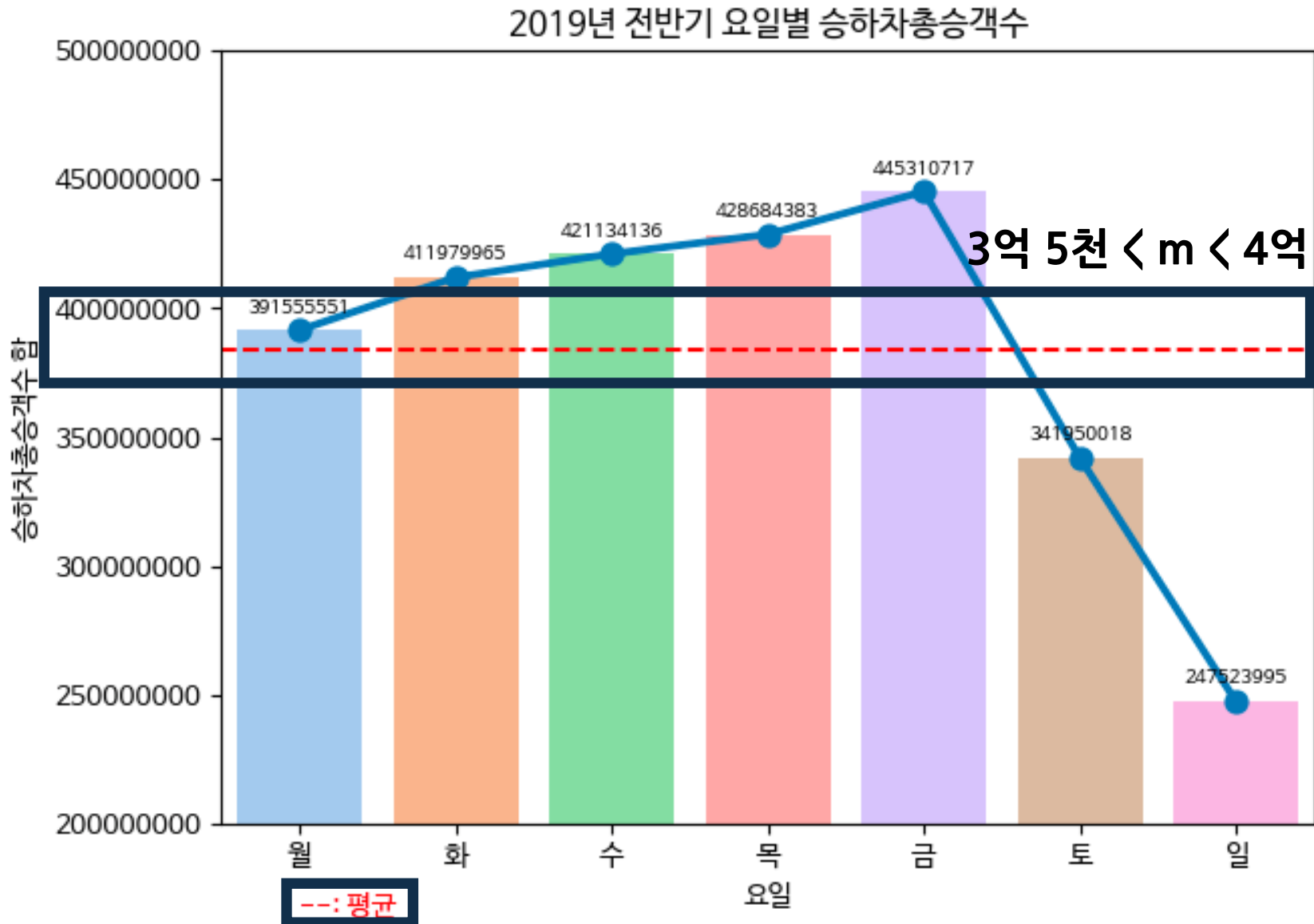


“**불**타는 **금**요일”의  
영향일 것(추측)

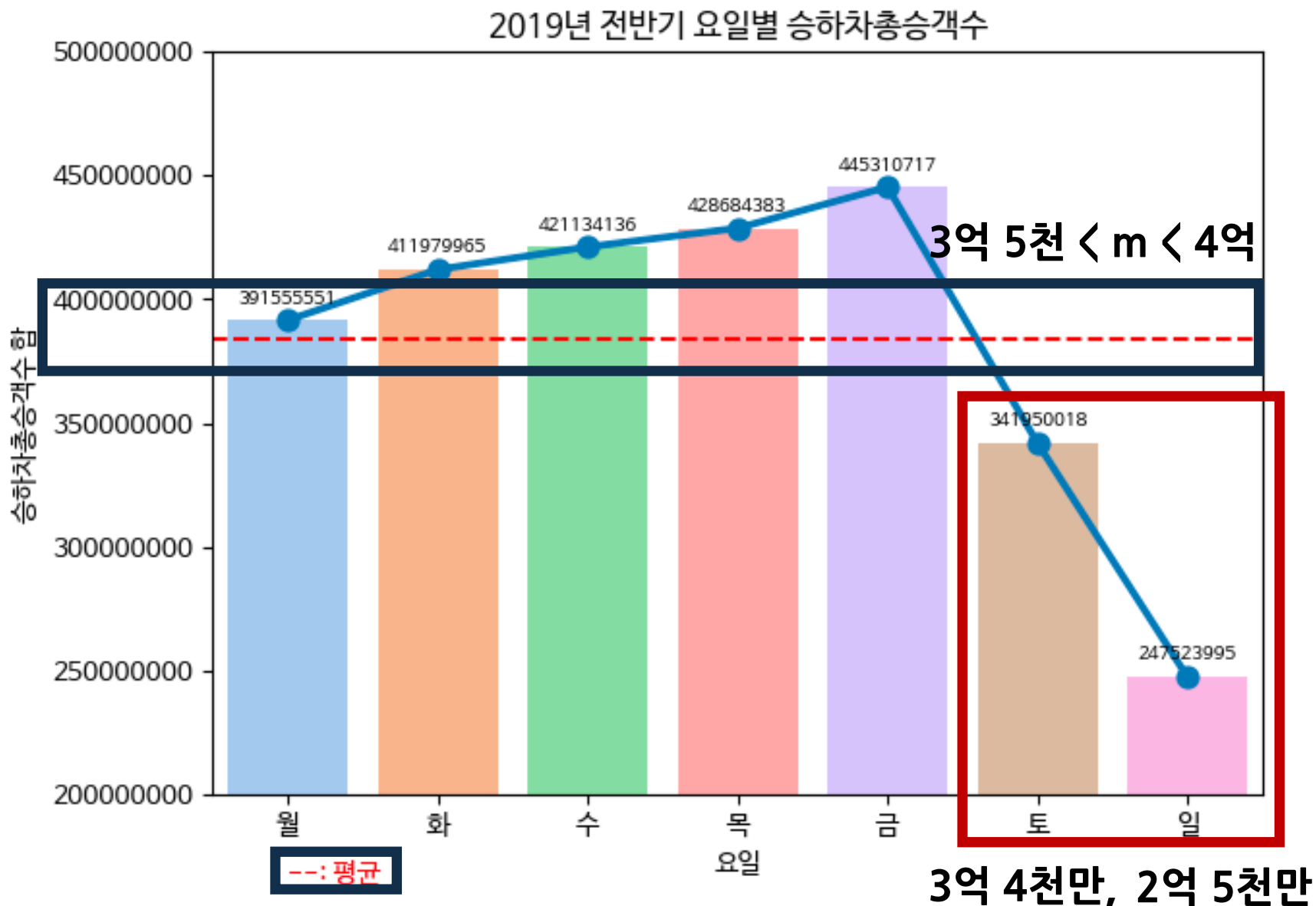


Feature 제한적  
=> 추측에 대한 불확실성  
( ( 아쉬움 ) )

### 3. 요일 중에서 목요일에 지하철 승객수가 많다? (기준: 승하차총승객수)

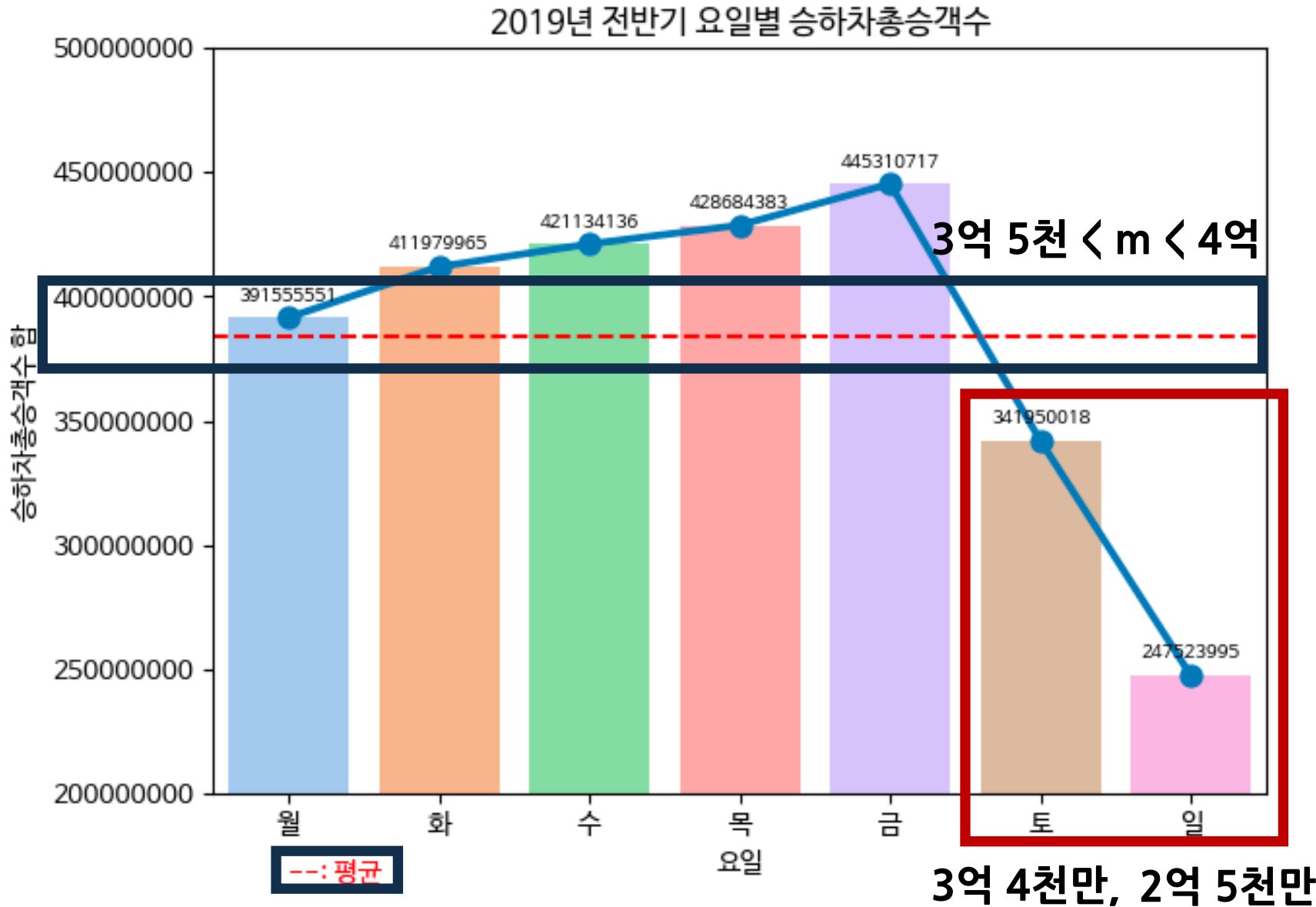


3. 요일 중에서 목요일에 지하철 승객수가 많다? (기준: 승하차총승객수 )



주말 승하차총승객수  
^  
요일 기준 승하차총승객수 평균

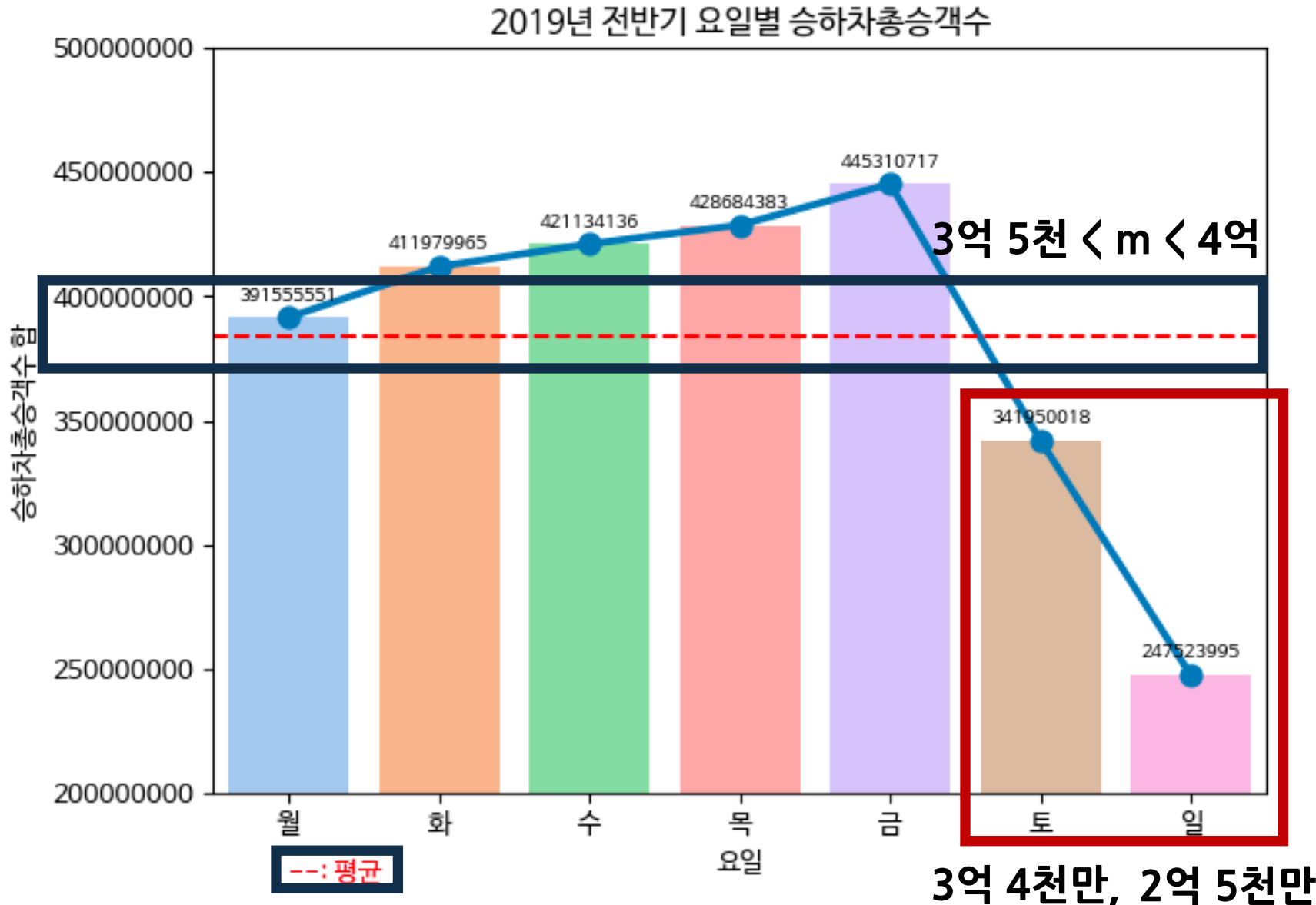
### 3. 요일 중에서 목요일에 지하철 승객수가 많다? (기준: 승하차총승객수)



주말 승하차총승객수  
 ^  
 요일 기준 승하차총승객수 평균  
 ↓  
 지하철을 이용하는  
 근로자, 학생의 영향일 것(추측)

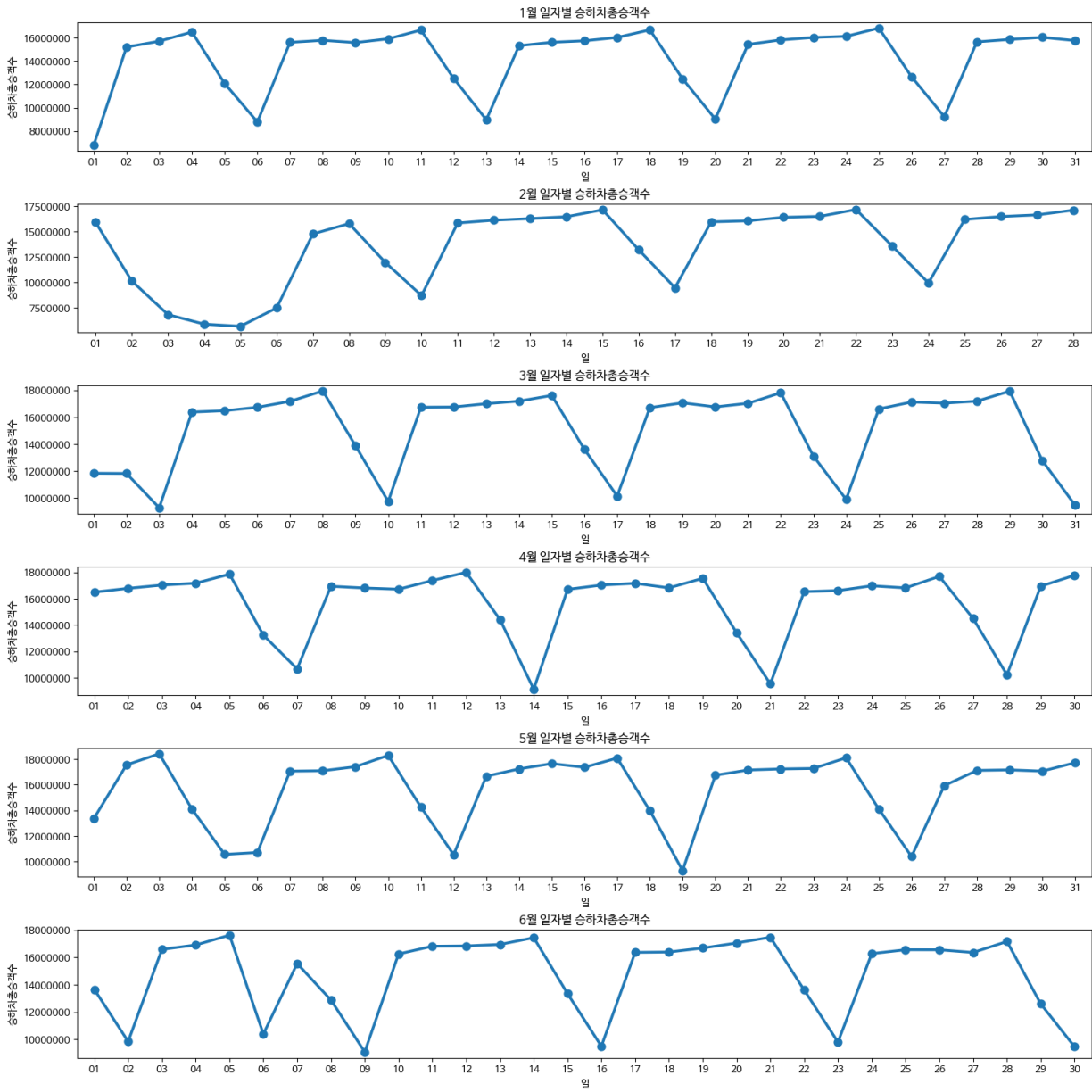


### 3. 요일 중에서 목요일에 지하철 승객수가 많다? (기준: 승하차총승객수)



주말 승하차총승객수  
 ^  
 요일 기준 승하차총승객수 평균  
 ↓  
 지하철을 이용하는  
 근로자, 학생의 영향일 것(추측)  
 ↓  
 시간 데이터 미비  
 => 추측에 대한 타당성 제로  
 ((아쉬움))

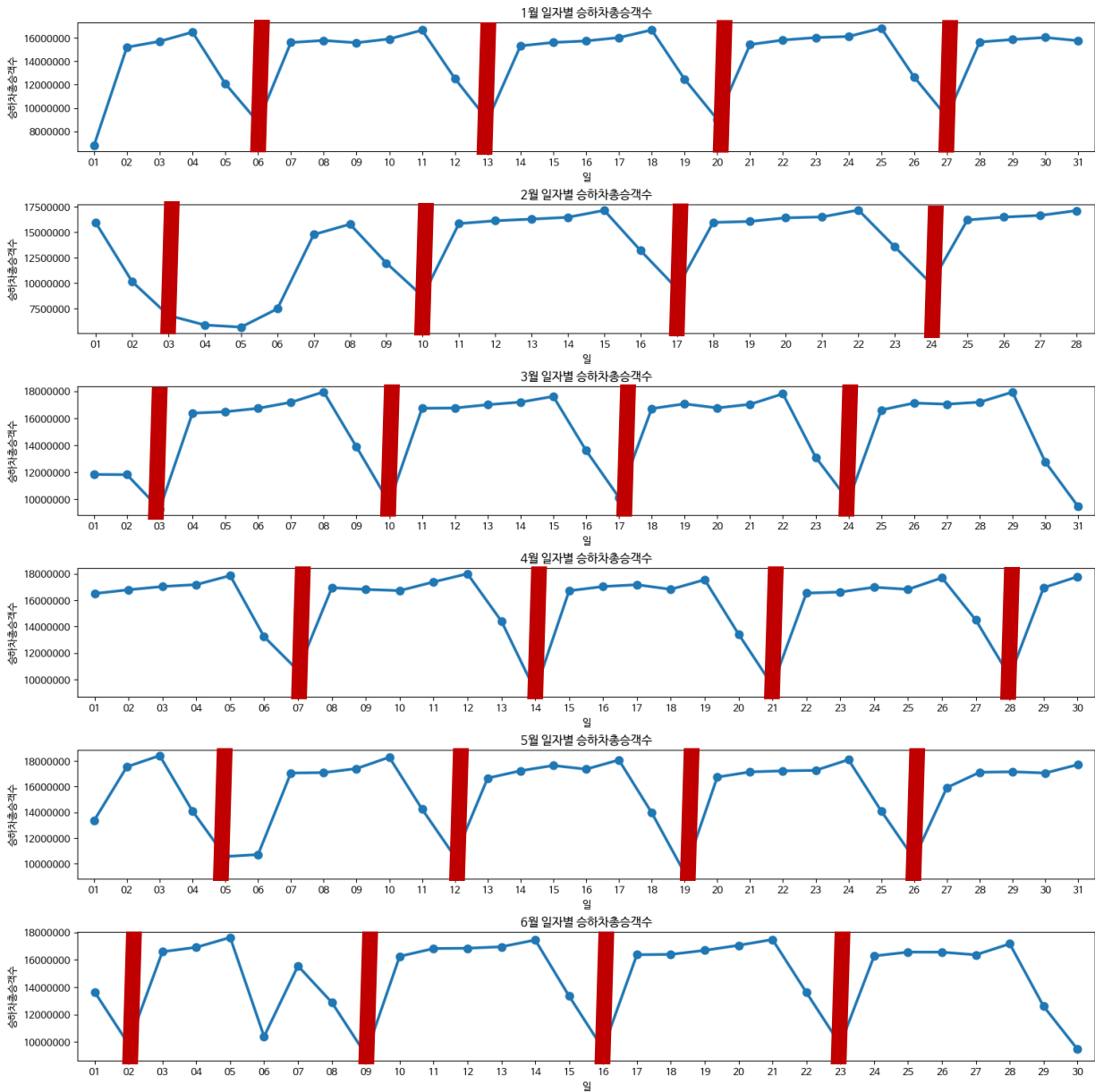
4. 연월 각각에 대해 일자별(월일별) 승하차총승객수 그래프 그려볼까요? (pointplot)



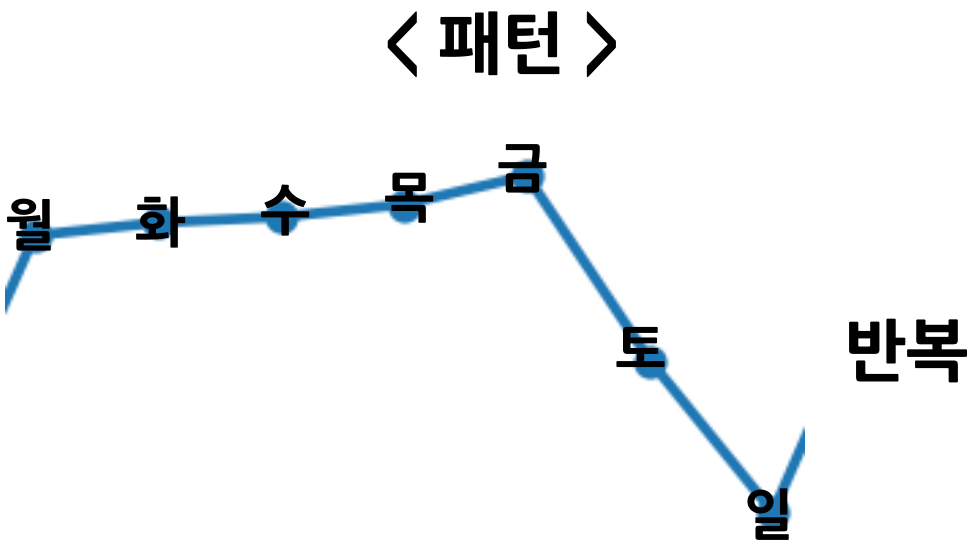
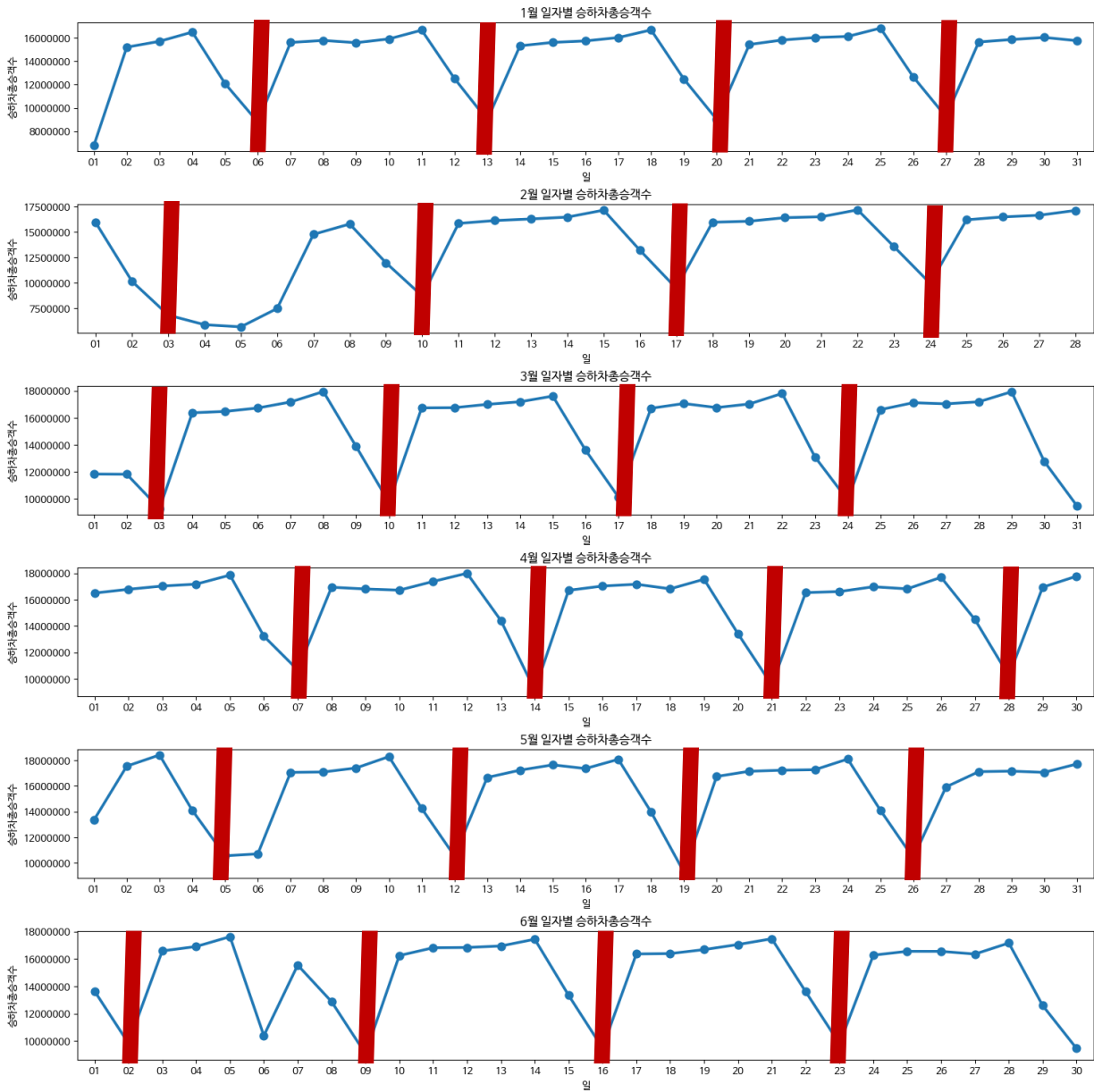
4. 연월 각각에 대해 일자별(월일별) 승하차총승객수 그래프 그려볼까요? (pointplot)



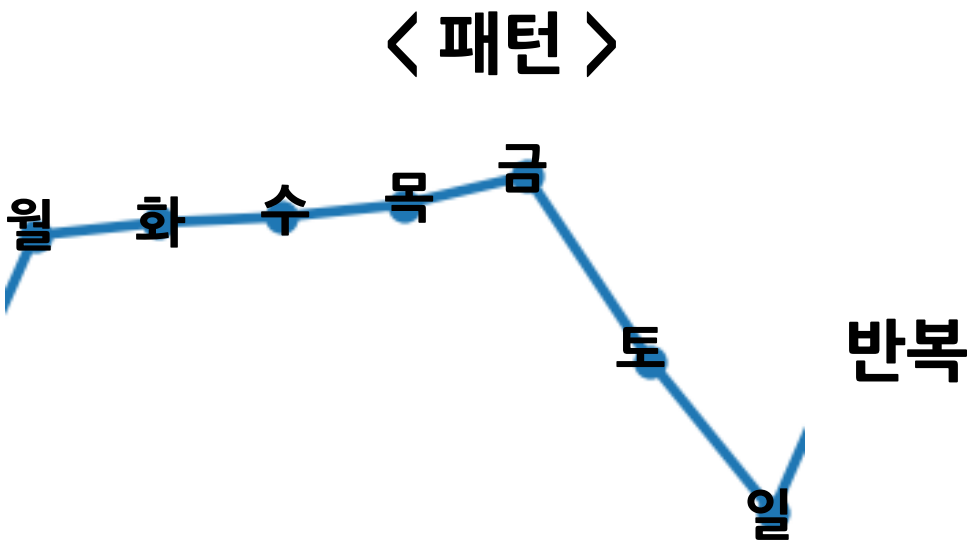
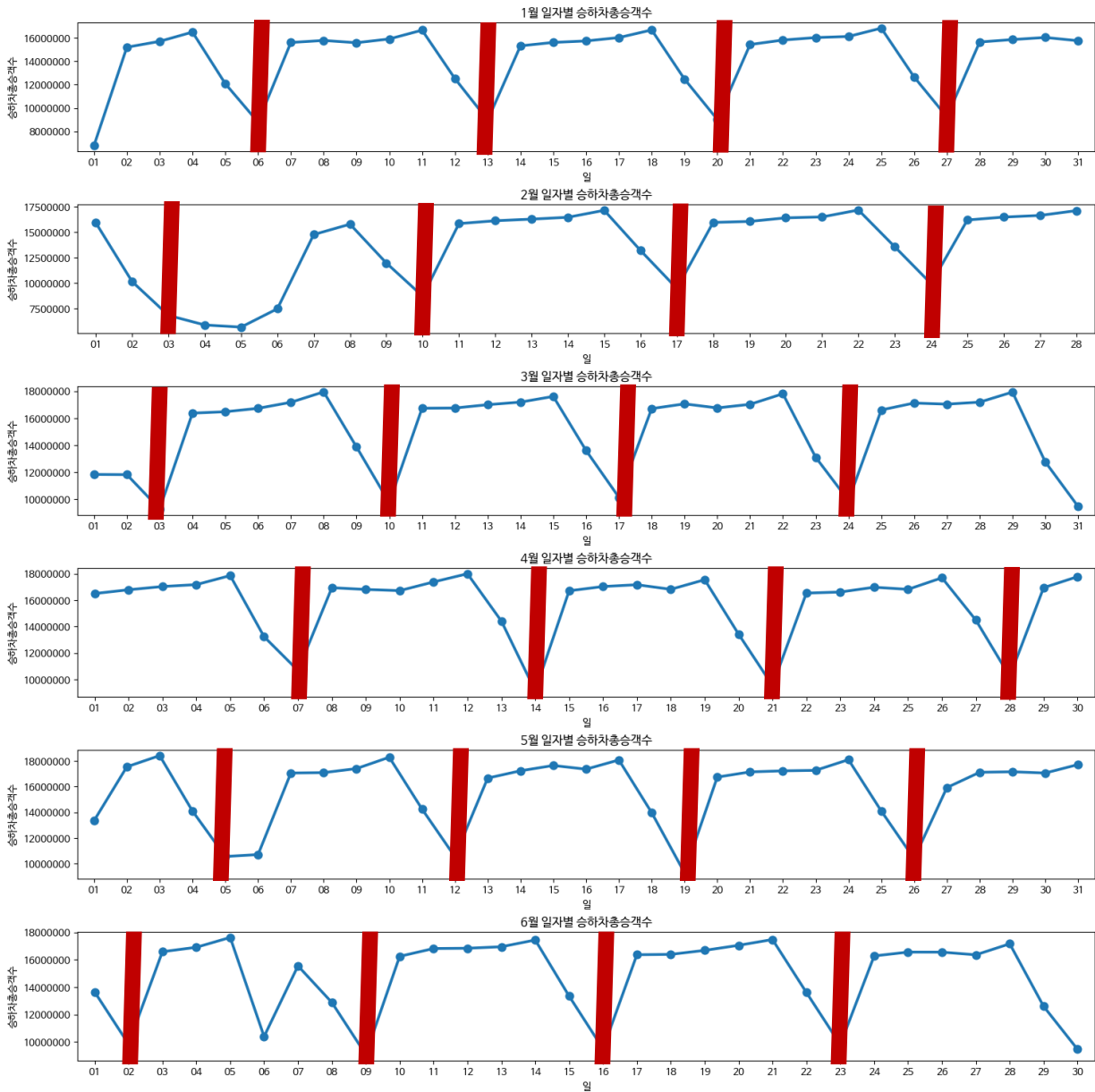
< 패턴 >



4. 연월 각각에 대해 일자별(월일별) 승하차총승객수 그래프 그려볼까요? (pointplot)



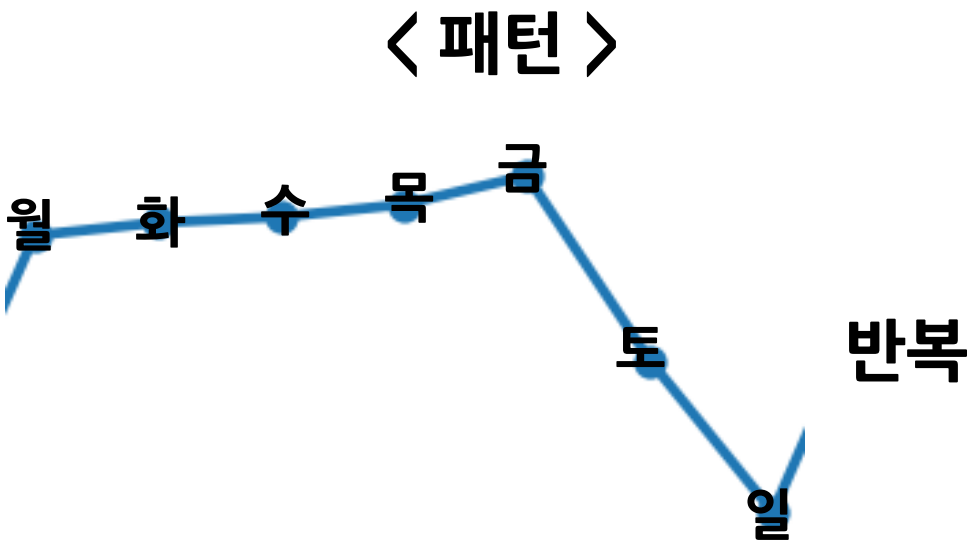
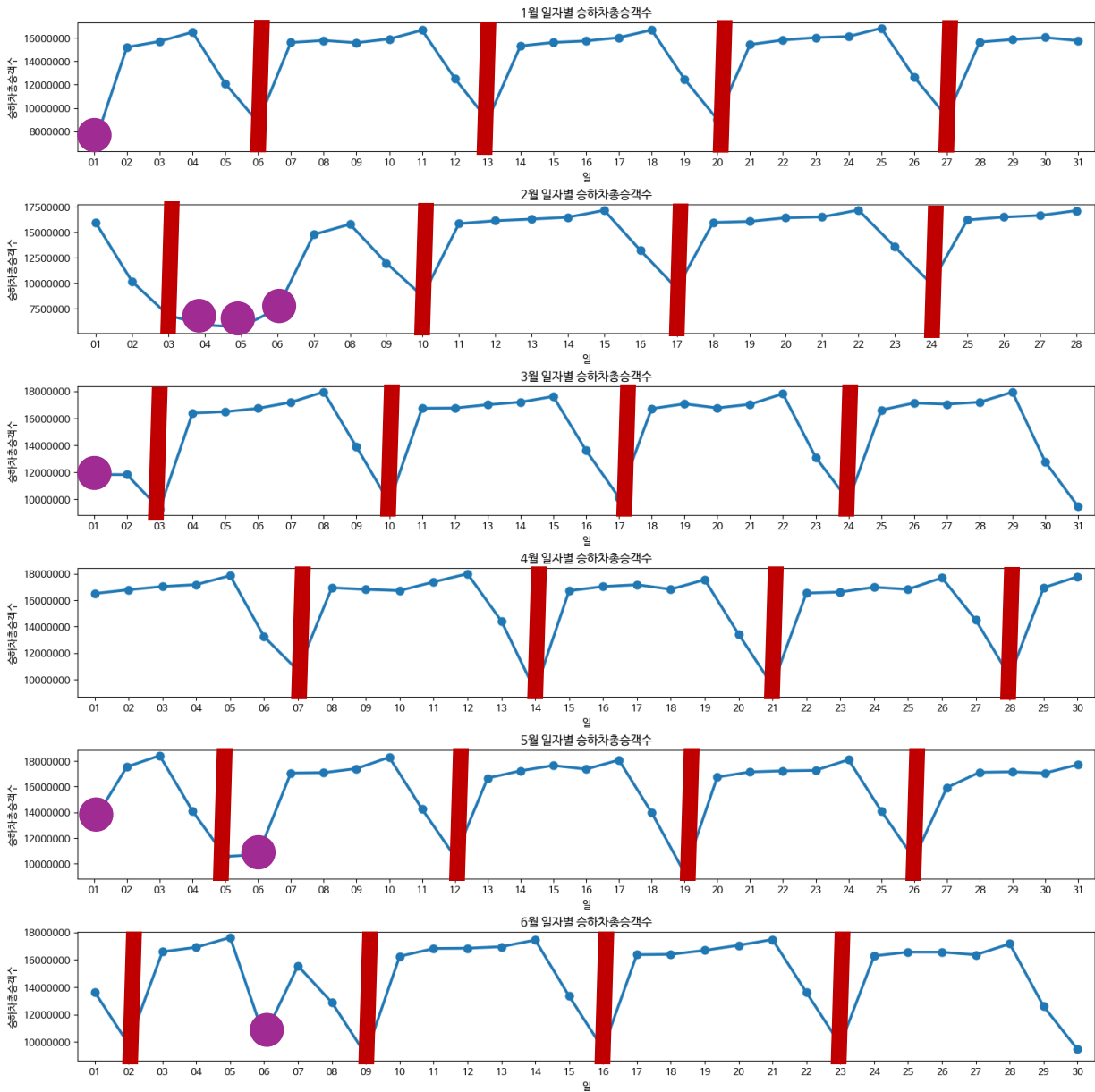
4. 연월 각각에 대해 일자별(월일별) 승하차총승객수 그래프 그려볼까요? (pointplot)



< 패턴에서 벗어나는 일자 >

● 으로 표시

4. 연월 각각에 대해 일자별(월일별) 승하차총승객수 그래프 그려볼까요? (pointplot)

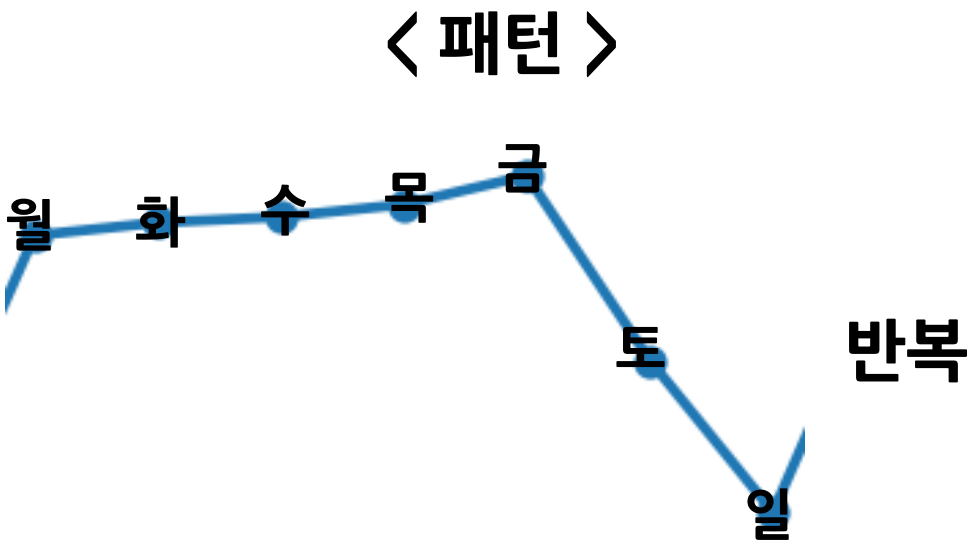
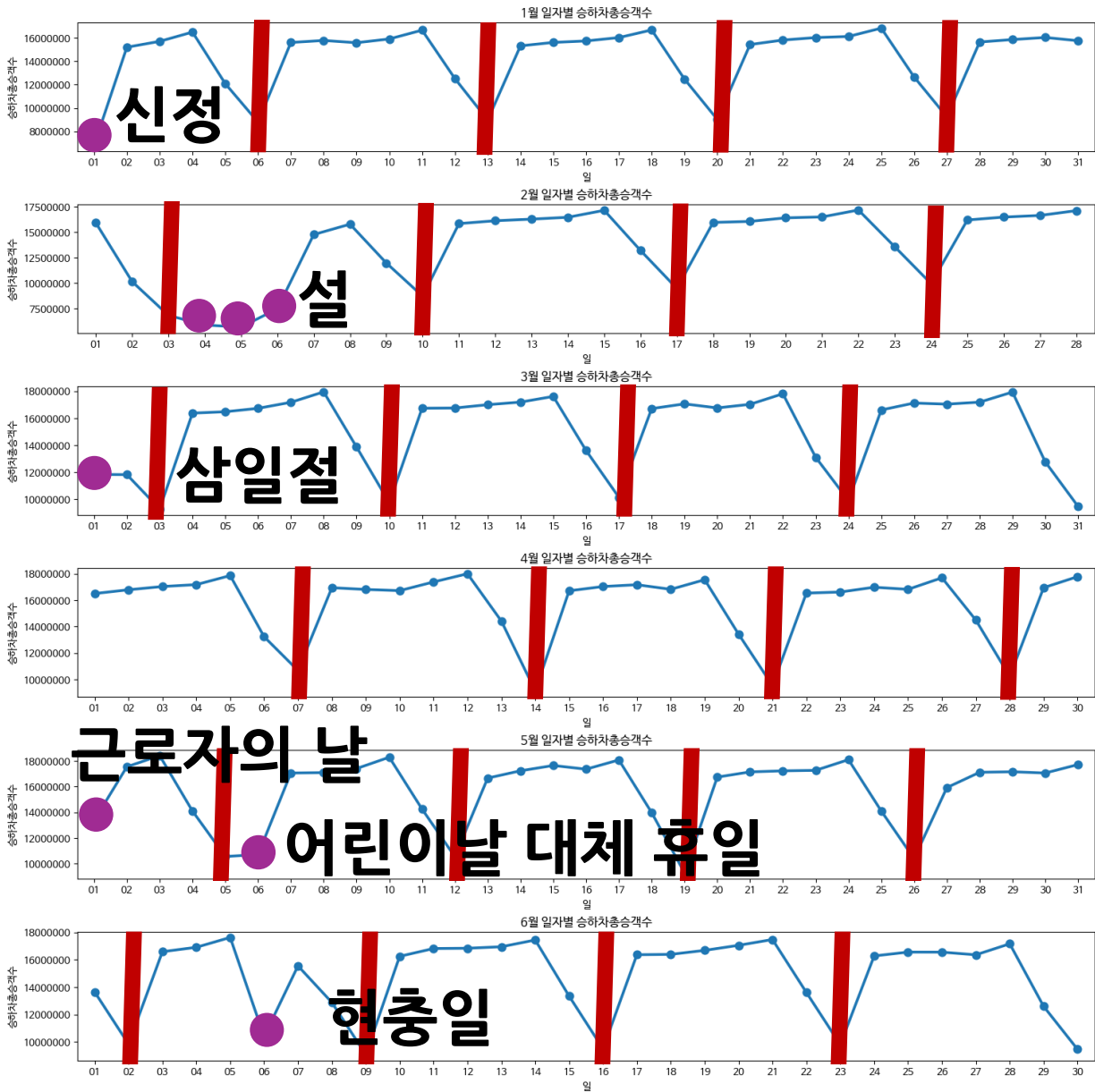


< 패턴에서 벗어나는 일자 >

● 으로 표시



4. 연월 각각에 대해 일자별(월일별) 승하차총승객수 그래프 그려볼까요? (pointplot)



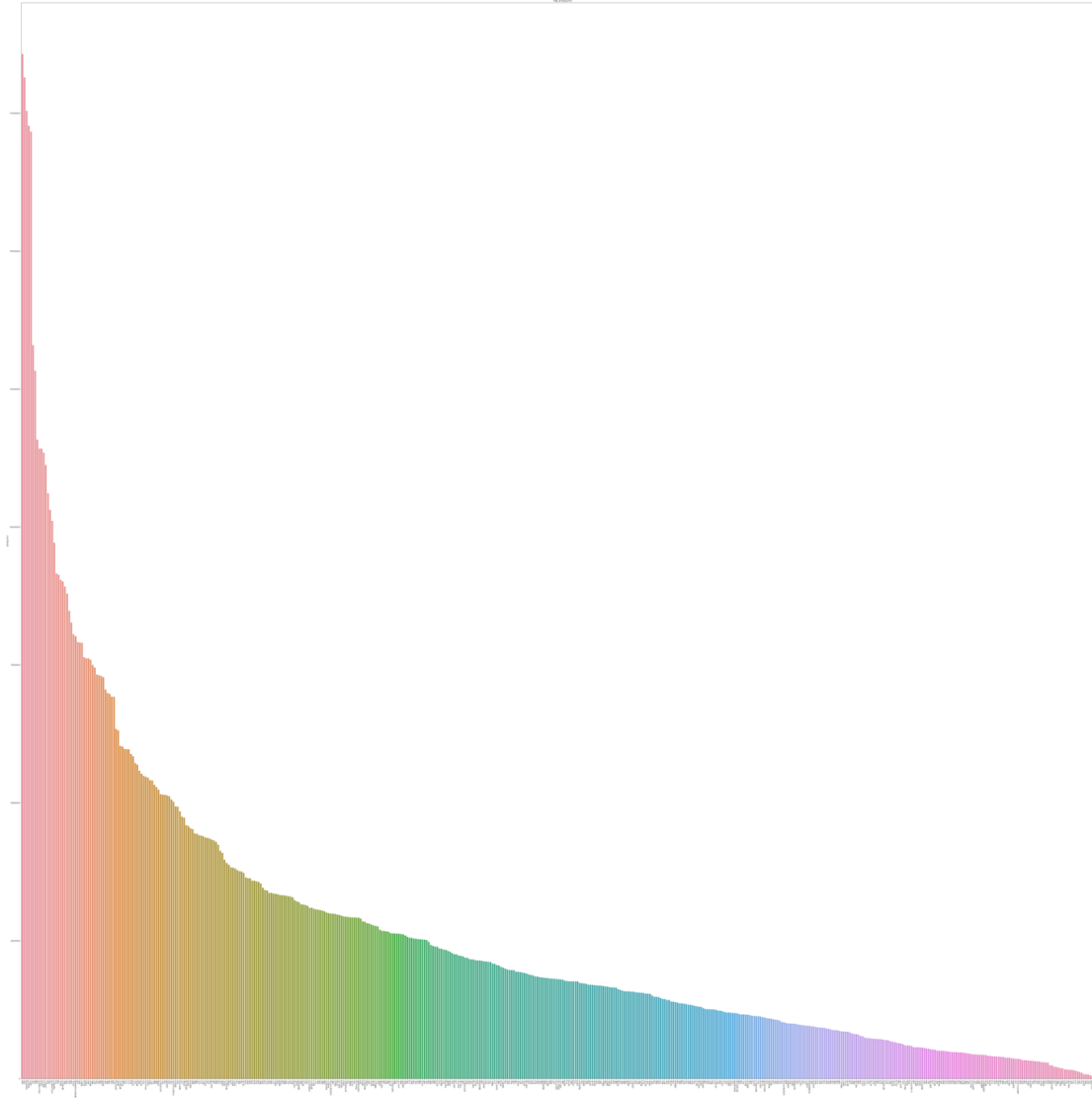
< 패턴에서 벗어나는 일자 >

● 으로 표시

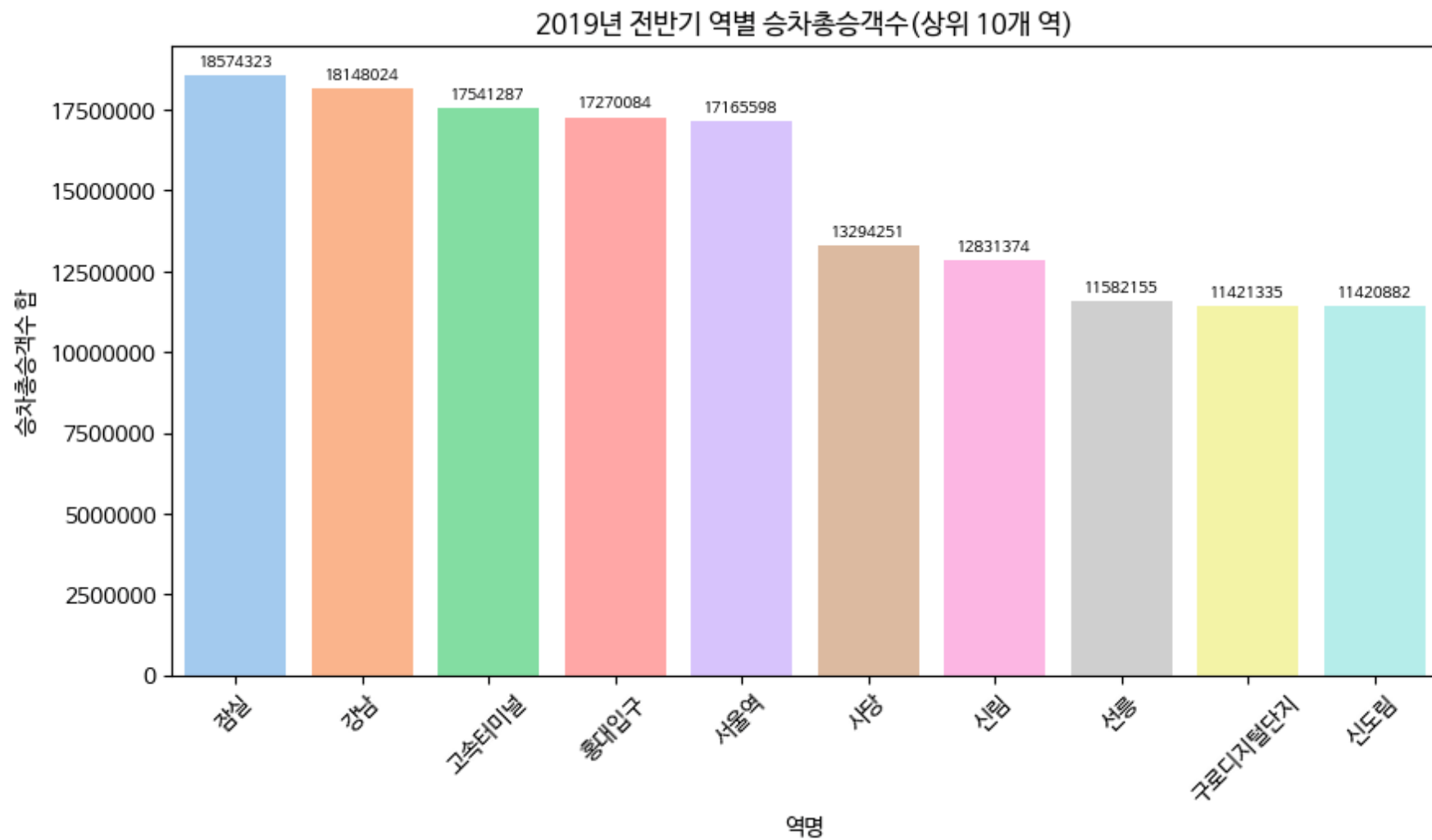
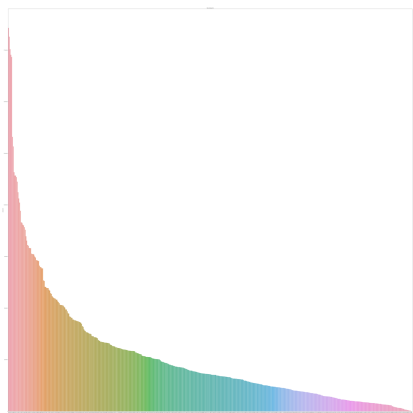
5. 가장 승객이 많이 타는 승차역은?



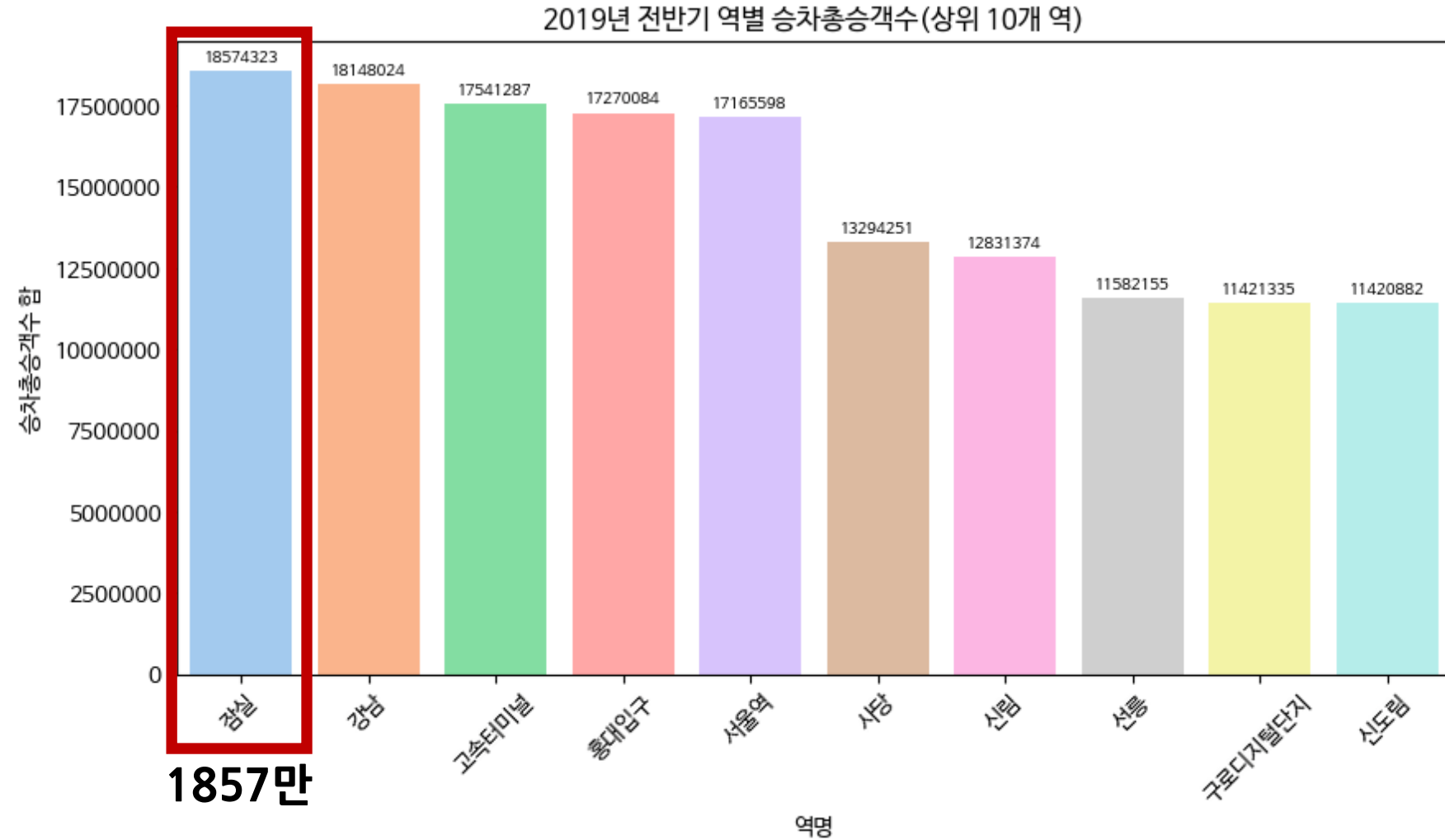
## 역별 승차총승객수 합



## 5. 가장 승객이 많이 타는 승차역은?



## 5. 가장 승객이 많이 타는 승차역은?

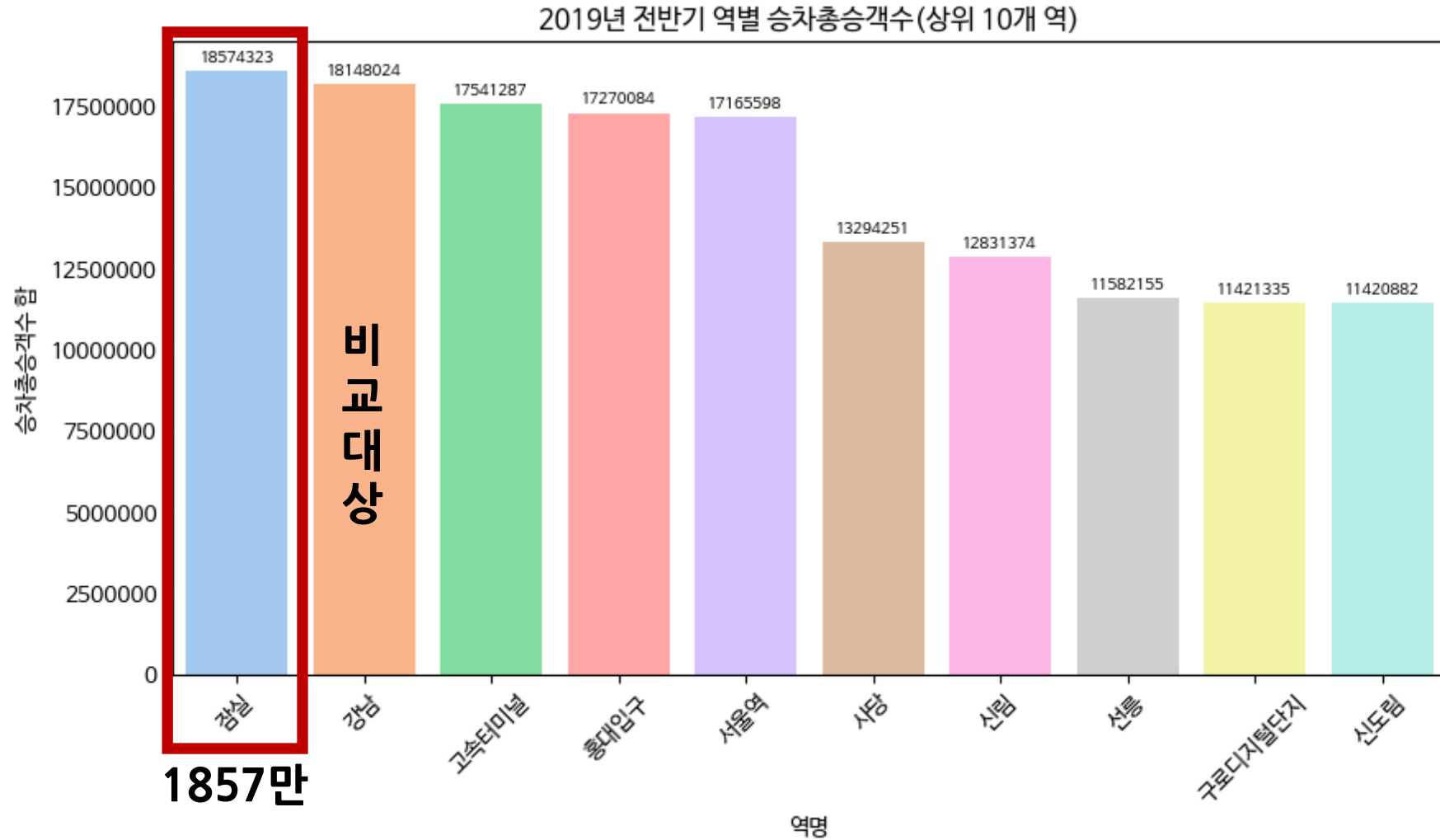


## 5. 가장 승객이 많이 타는 승차역은?

**잠실역** 승차총승객수가  
**강남역** 승차총승객수보다

18,574,323 - 18,148,024

426,299



## 6. 노선별로 역별/요일별 승차승객수를 비교해 볼 수 있을까? (1~9호선, 역별/요일별 heatmap)



노선별/요일별 승차총승객수 합



## 6. 노선별로 역별/요일별 승차승객수를 비교해 볼 수 있을까? (1~9호선, 역별/요일별 heatmap)



노선별/요일별 승차총승객수 합



〈 전반적인 특성 〉

노선명 中  
**2호선** 승차승객수가  
가장 많다

6. 노선별로 역별/요일별 승차승객수를 비교해 볼 수 있을까? (1~9호선, 역별/요일별 heatmap)



노선별/요일별 승차총승객수 합



< 전반적인 특성 >

노선명 中  
**2호선** 승차승객수가  
가장 많다

요일 中  
**금요일** 승차승객수가  
가장 많다

## 6. 노선별로 역별/요일별 승차승객수를 비교해 볼 수 있을까? (1~9호선, 역별/요일별 heatmap)



노선별/요일별 승차총승객수 합



### < 전반적인 특성 >

노선명 中  
**2호선** 승차승객수가  
가장 많다

요일 中  
**금요일** 승차승객수가  
가장 많다



6. 노선별로 역별/요일별 승차승객수를 비교해 볼 수 있을까? (1~9호선, 역별/요일별 heatmap)



노선별/요일별 승차총승객수 합



< 전반적인 특성 >

노선명 中  
**8호선** 승차승객수가  
가장 적다



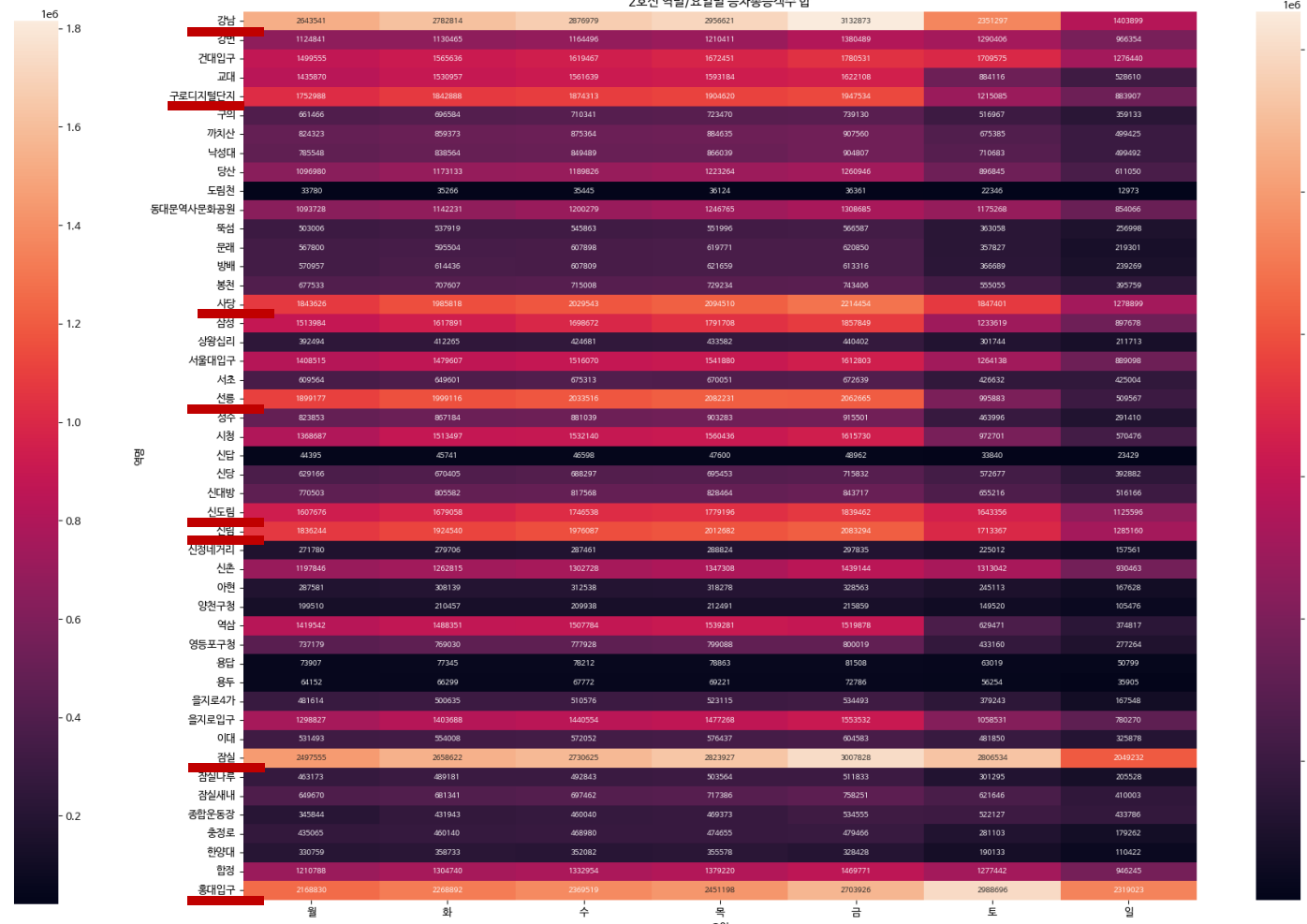
6. 노선별로 역별/요일별 승차승객수를 비교해 볼 수 있을까? (1~9호선, 역별/요일별 heatmap)



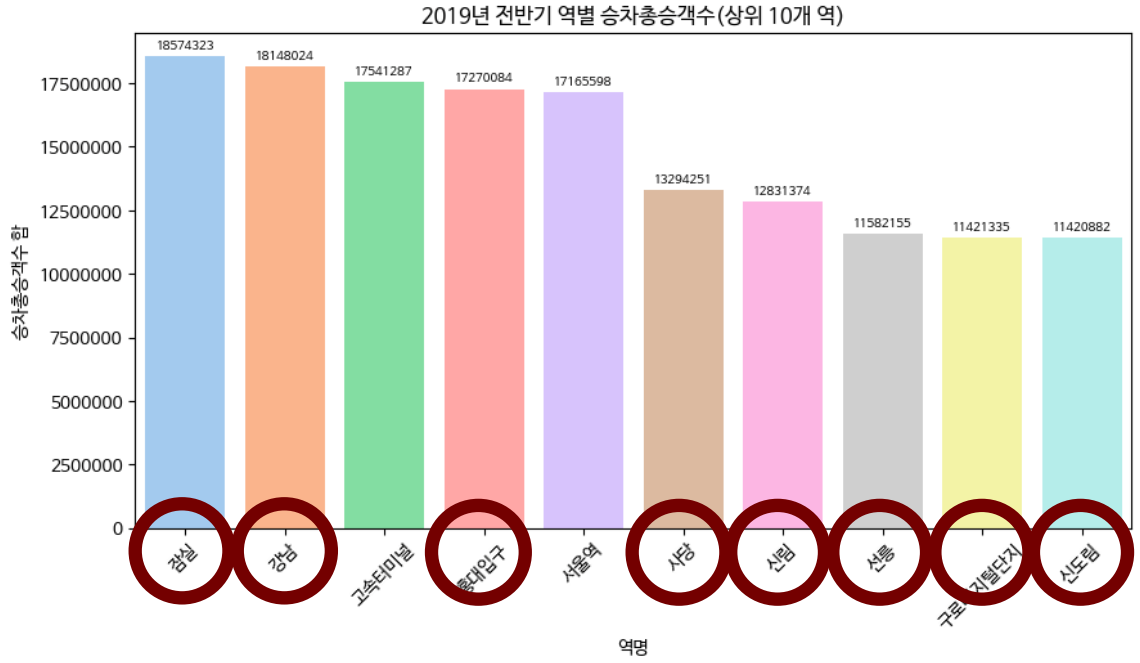
< 1호선 역별/요일별 승차총승객수 합 >

1호선 역별/요일별 승차총승객수 합									
가능	190014	201352	204871	205501	210822	171506	130598		
간석	1701356	1705336	1778368	1810008	1815630	702145	587274		
광안대	180042	187924	190023	193590	196207	149620	108544		
관악	686662	691604	697603	699997	707829	900186	632110		
구림	226010	240166	241250	246655	249150	205029	170350		
금천구청	77413	78249	79982	81646	84887	70548	62752		
노원	271041	282025	289387	291233	291222	201720	144883		
노원	556205	576837	588829	591296	596077	477382	365162		
노원	229164	230000	230027	230090	230090	164030	145909		
노원	220968	218881	222241	223196	223196	145417	101188		
노원	338955	334222	338536	341399	347596	262987	204287		
노원	304891	322426	323887	311817	320057	208033	157420		
노원	301044	311866	318118	317816	317816	111236	712344		
노원	130005	140979	145218	145187	151113	119449	90047		
노원	109553	107888	107888	103771	110862	105713	103664		
노원	234559	217662	217163	211528	160345	103664	103664		
노원	402005	418056	426224	435635	446420	314668	226285		
노원	62261	62064	63480	63118	63118	51052	8582		
노원	187902	190330	190330	190330	200281	103331	115307		
노원	172297	181311	184821	186996	188391	141592	110004		
노원	406993	406993	406993	406993	406993	406993	406993		
노원	104412	112486	114509	114509	114509	91548	69456		
노원	114164	119666	122248	121785	125677	97123	71769		
노원	406993	406993	406993	406993	406993	406993	406993		
노원	63332	65747	67780	67775	69552	40810	30789		
노원	102716	106074	106121	106654	109528	96156	76834		
노원	402051	399645	397331	399200	405388	309381	209381		
노원	437998	433836	433935	437738	439487	306283	346013		
노원	156483	166221	171263	175719	181447	155000	124621		
노원	221135	227528	227528	227528	227528	192000	125435		
노원	289154	306486	305580	310011	310011	188933	122023		
노원	276431	287576	293436	296046	296996	238035	188998		
노원	27167	28532	28888	29453	29528	21115	22517		
노원	256640	264650	271023	275022	280022	203741	151718		
노원	446988	468420	471971	476372	481412	401161	290357		
노원	402051	404420	404420	404420	404420	404420	404420		
노원	38335	39553	39731	39889	39624	28281	20964		
노원	263087	275853	277762	282202	287761	219005	164367		
노원	3843394	3873346	3873346	3873346	3873346	3873346	3873346		
노원	952889	971371	1007900	1019277	1029831	1058005	822733		
노원	47833	50531	50724	51580	54007	42808	31947		
노원	141921	146905	150486	151132	154486	138923	102599		
노원	524777	559042	569886	579991	591258	537703	384841		
노원	318515	332631	336300	341502	347220	252334	184896		
노원	434335	435663	433792	437800	441174	353207	262878		
노원	133725	146468	150024	147869	103221	131599	78812		
노원	90182	104794	103773	107404	116119	106474	82033		
노원	78616	82957	85507	89059	92914	72125	57235		
노원	301042	317126	316450	321420	323062	285689	215675		
노원	93547	96746	94300	96986	103408	146874	135185		
노원	886310	928270	943504	959537	975048	898707	490701		
노원	126470	127690	132901	131721	145988	146379	115783		
노원	1383387	1448192	1474506	1493149	1593469	1569123	1540536		
노원	309605	303812	305044	305044	305044	286209	286209		
노원	505170	612373	622274	622228	623734	407474	407474		
노원	252639	263853	267162	270982	275794	211445	157876		
노원	26574	26997	26998	27411	281572	42681	23255		
노원	72018	74703	76733	81320	78220	57523	40400		
노원	86218	86999	86745	86935	102017	86320	65071		
노원	79367	79368	79368	79368	79368	79368	79368		
노원	243837	253837	256245	261310	267904	230339	172863		
노원	877018	897546	910143	921213	937410	702017	492069		
노원	1448190	1416819	1341810	1270173	1201025	1026009	1026009		
노원	334846	347537	359333	362245	368083	299727	240295		
노원	270846	279304	282157	280770	309319	288822	204880		
노원	93184	101820	106465	103066	101119	62970	42505		
노원	519216	544641	560224	569994	551998	354426	274453		
노원	125183	129565	125713	131041	140750	158826	136399		
노원	362664	368877	362644	359241	344724	244724	178120		
노원	874268	1007644	1040790	1076677	1126648	1068488	964795		
노원	191630	204813	207501	204233	202824	142294	104508		
노원	242728	246775	241729	243112	241166	180746	134869		
노원	532018	548515	576644	575216	622930	564726	564726		
노원	130081	139762	142448	141302	155687	148401	124491		
노원	545439	561691	561691	561691	561691	561691	561691		
노원	316151	332808	335583	336004	341792	247776	176903		
노원	1169962	1207443	1203688	1210588	1304855	941534	572584		
노원	713309	701657	701657	701657	701657	701657	701657		
노원	625047	625788	625788	625788	625788	625788	625788		
노원	274545	284906	286621	292742	295562	227257	162130		
노원	60531	62545	60431	60211	60488	60223	39034		
노원	197721	203694	207788	208113	188217	141350	104350		
노원	23403	23972	24085	23633	27322	24814	20695		
노원	192038	202906	207124	207522	207522	207522	207522		
노원	324205	324205	324205	324205	324205	324205	324205		
노원	222285	223416	223416	223416	223416	223416	223416		
노원	773456	814847	831262	831262	731742	418802	322111		
노원	201127	210001	209681	209195	209195	131500	231100		
월									
화									
수									
목									
금									
토									
일									

< 2호선 역별/요일별 승차총승객수 합 >

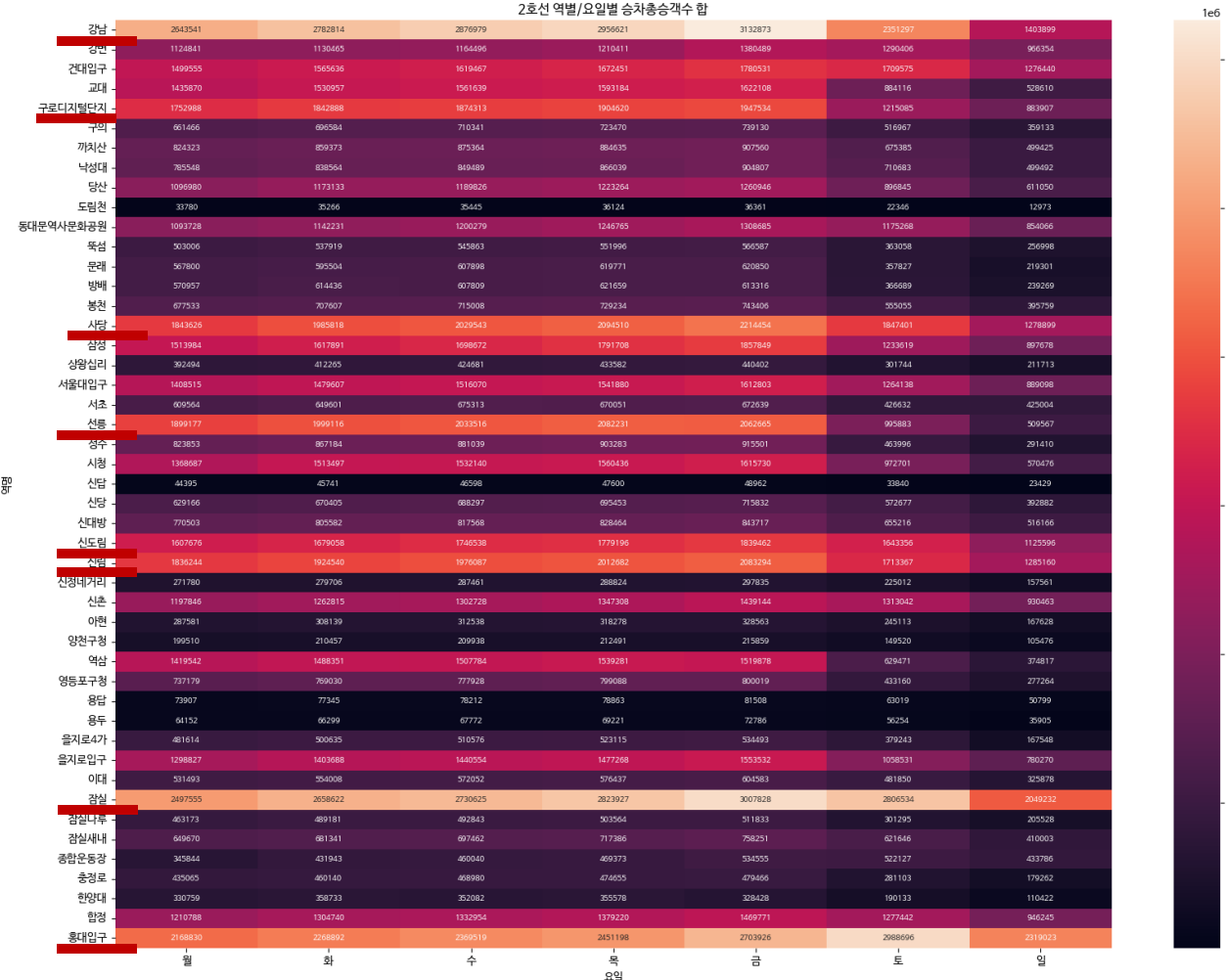


6. 노선별로 역별/요일별 승차승객수를 비교해 볼 수 있을까? (1~9호선, 역별/요일별 heatmap)



\* 2호선 역 ○으로 표시

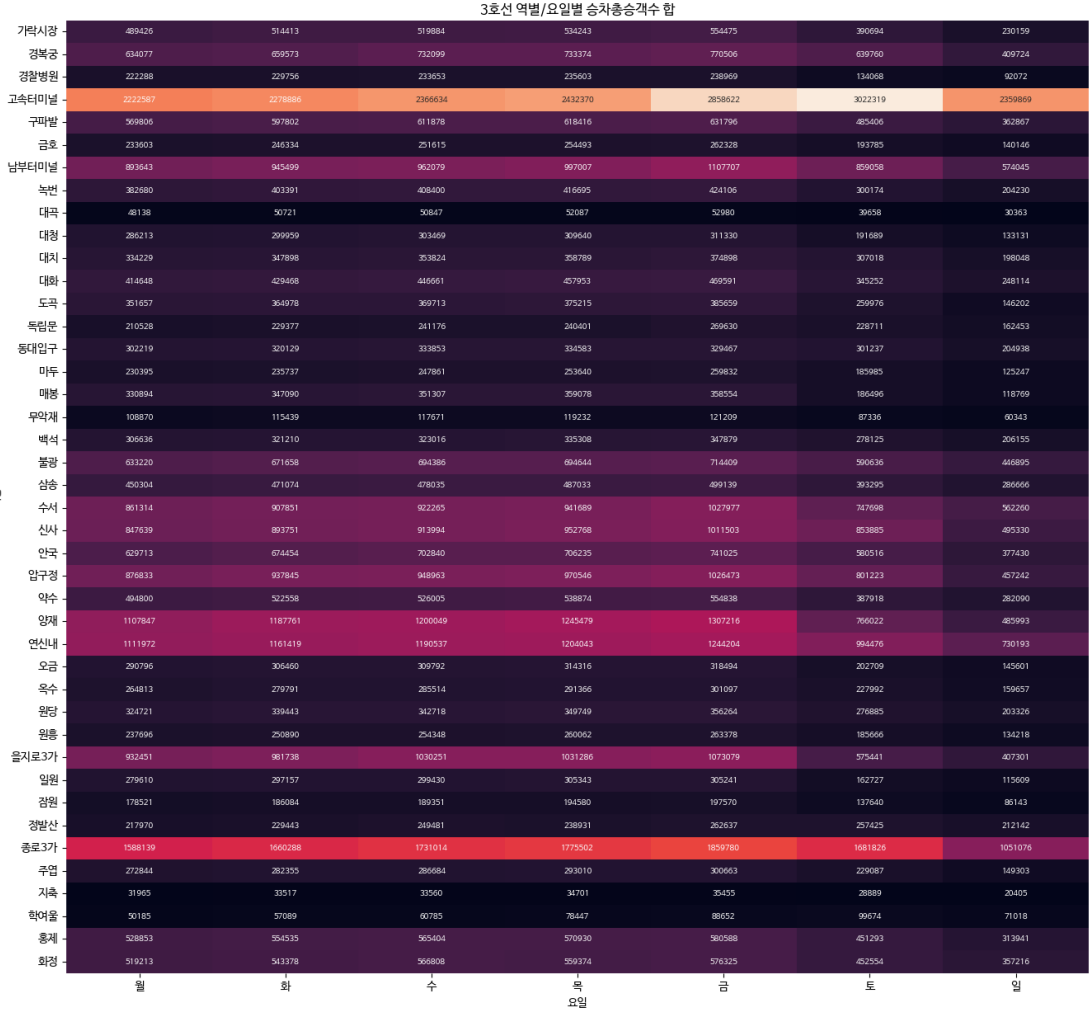
< 2호선 역별/요일별 승차총승객수 합 >



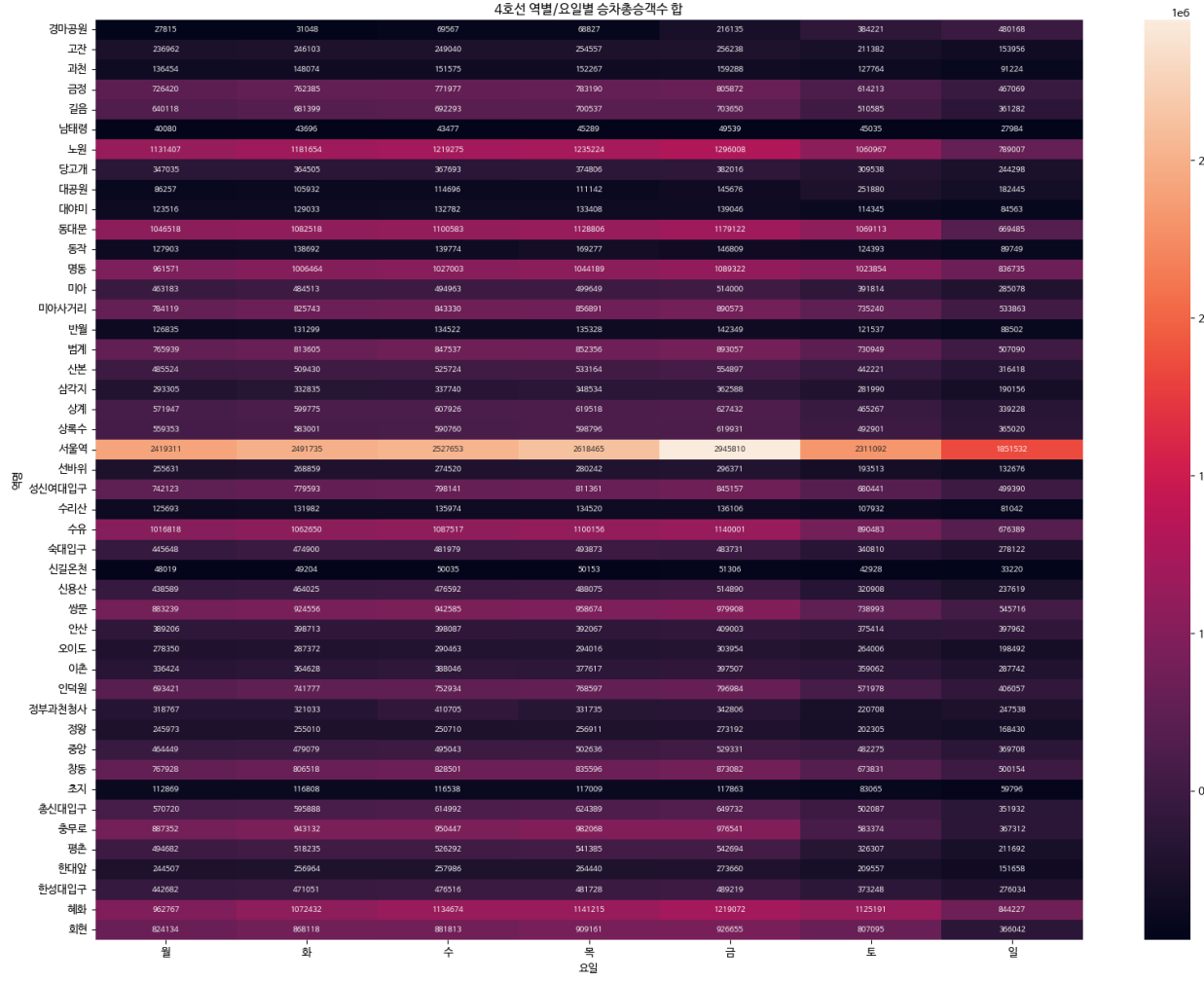
6. 노선별로 역별/요일별 승차승객수를 비교해 볼 수 있을까? (1~9호선, 역별/요일별 heatmap)



< 3호선 역별/요일별 승차총승객수 합 >



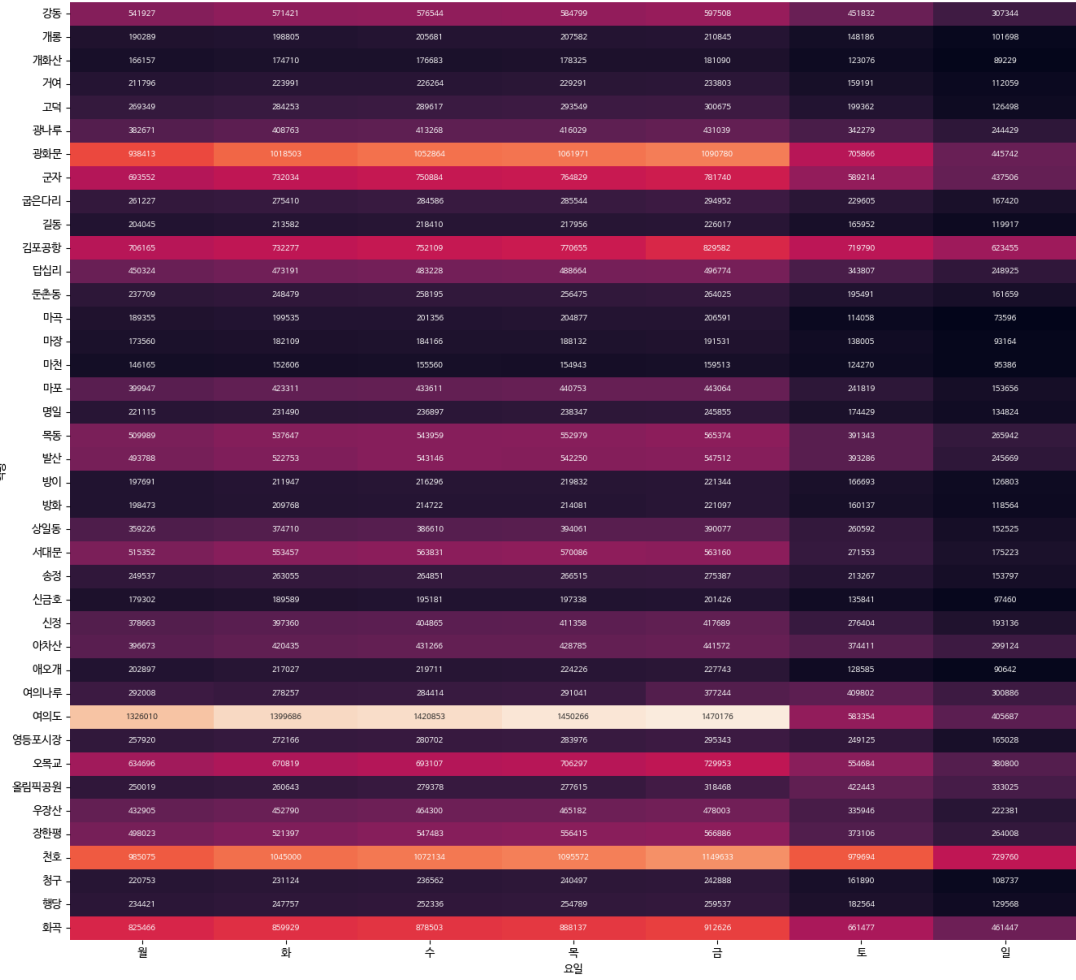
< 4호선 역별/요일별 승차총승객수 합 >



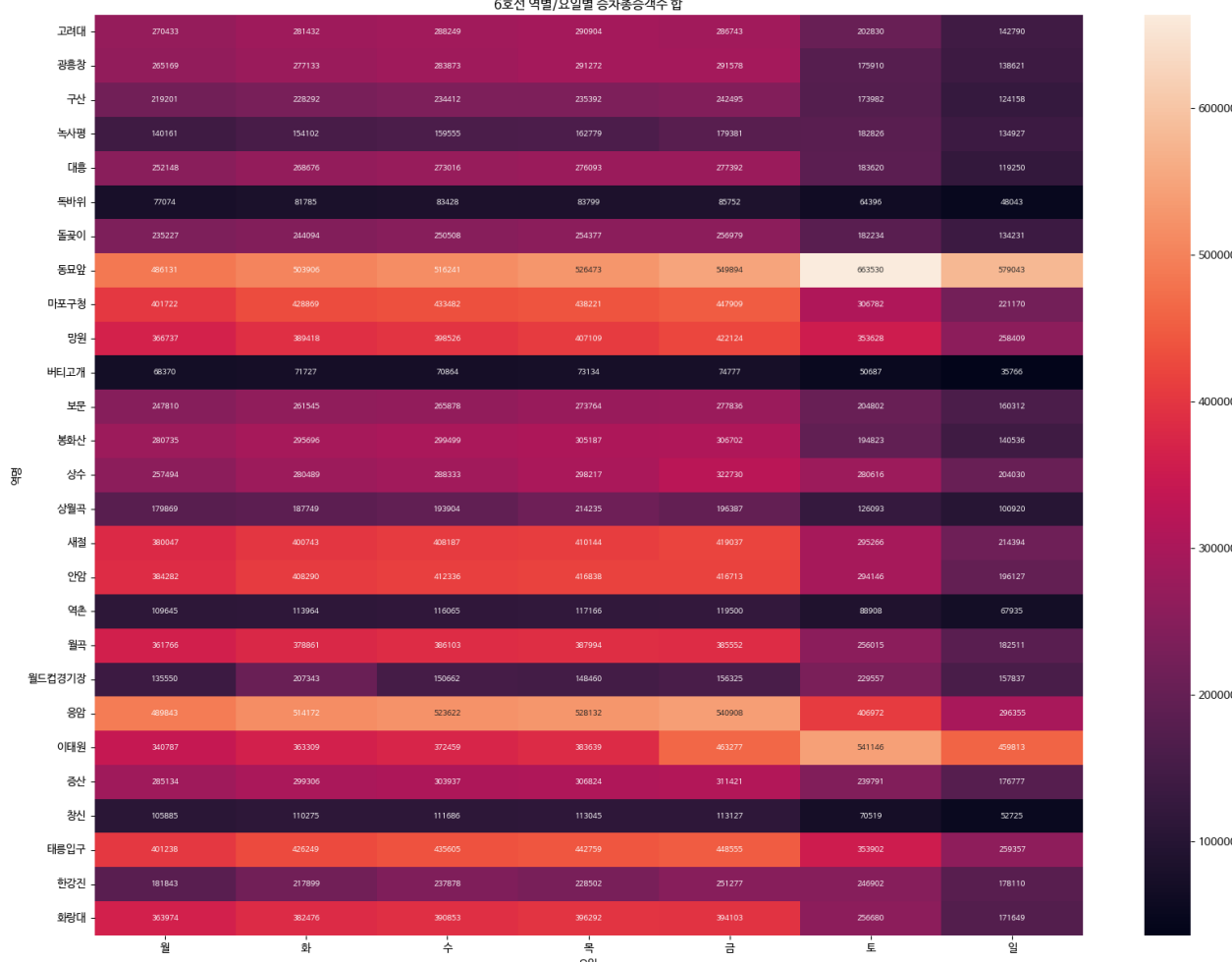
6. 노선별로 역별/요일별 승차승객수를 비교해 볼 수 있을까? (1~9호선, 역별/요일별 heatmap)



< 5호선 역별/요일별 승차총승객수 합 >



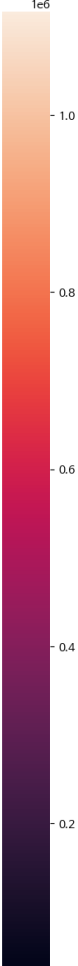
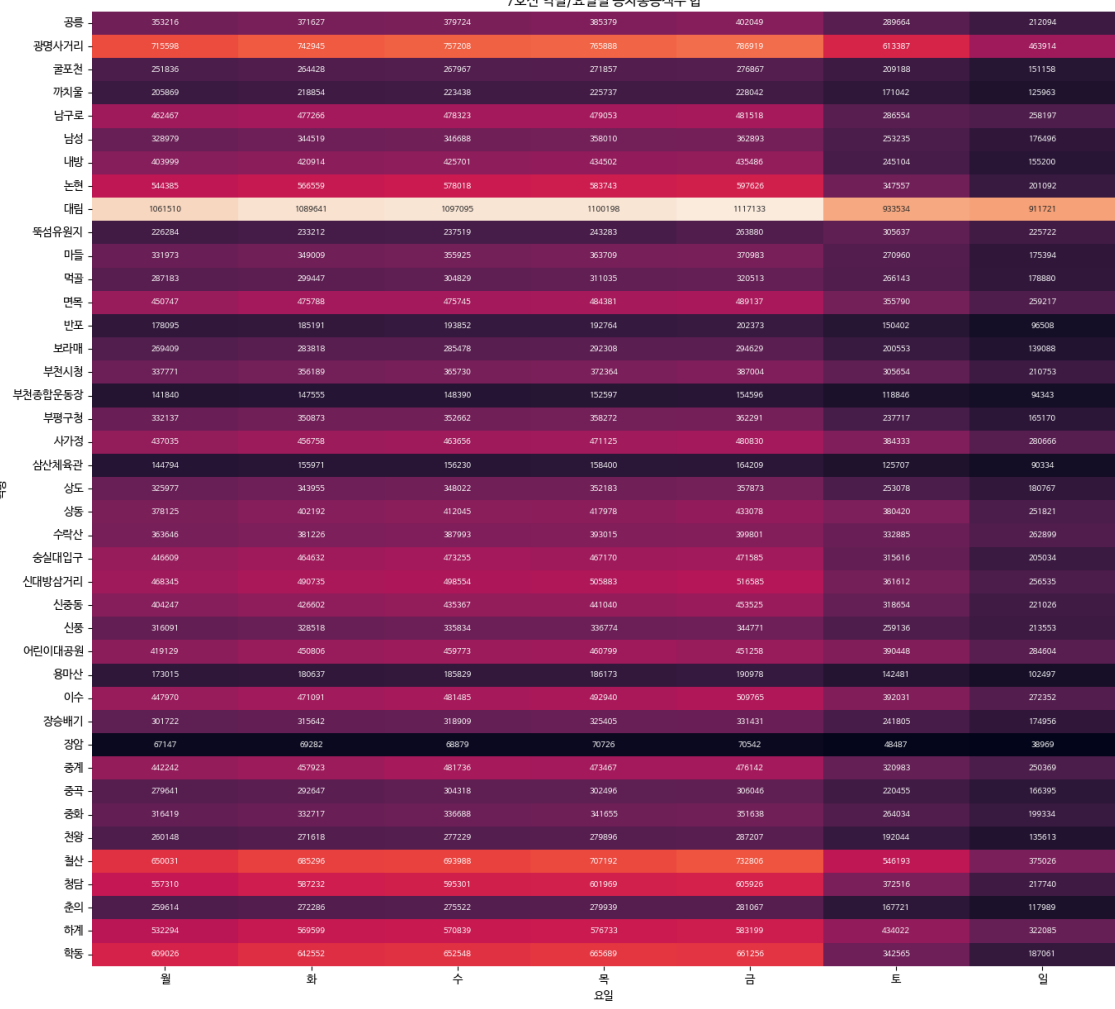
< 6호선 역별/요일별 승차총승객수 합 >



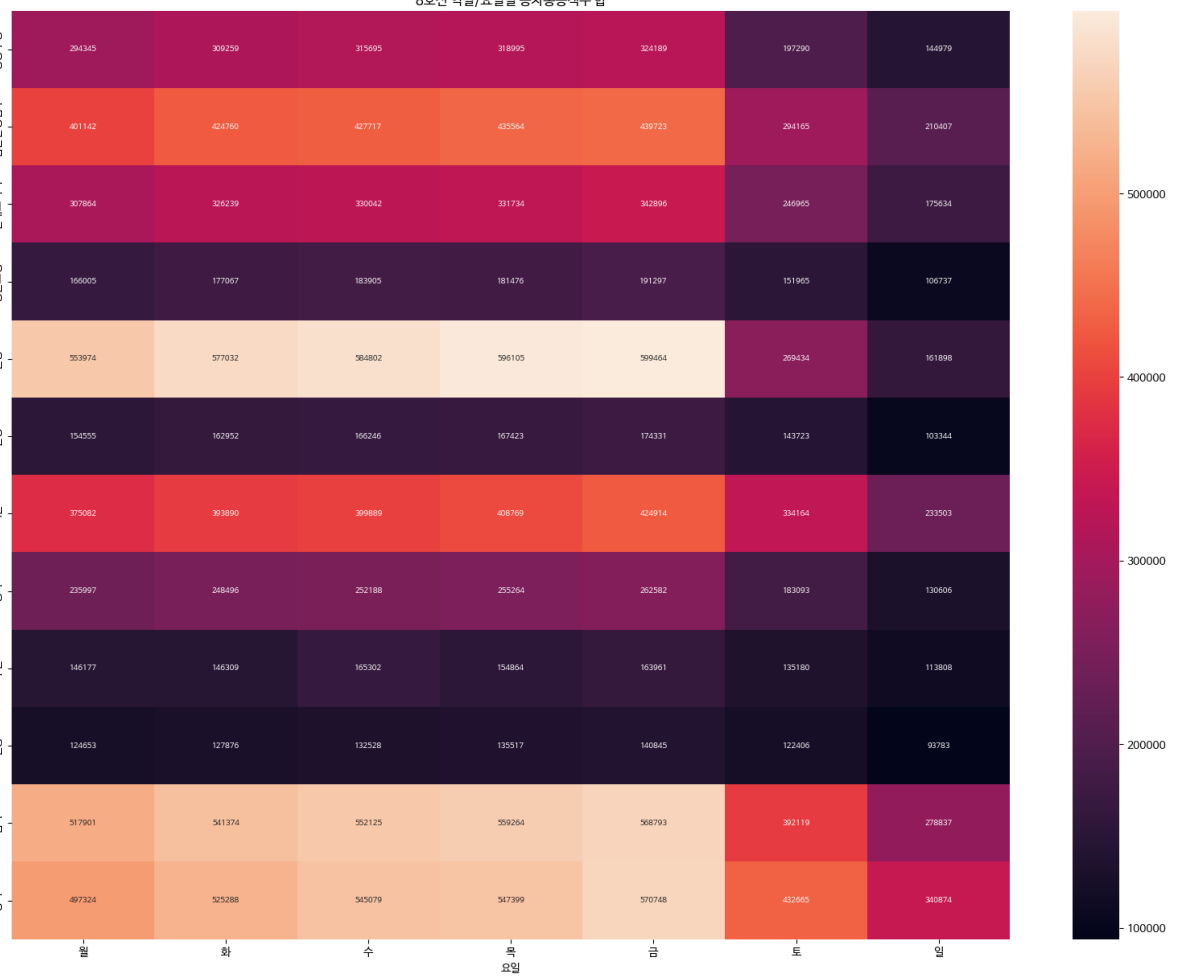
6. 노선별로 역별/요일별 승차승객수를 비교해 볼 수 있을까? (1~9호선, 역별/요일별 heatmap)



< 7호선 역별/요일별 승차총승객수 합 >



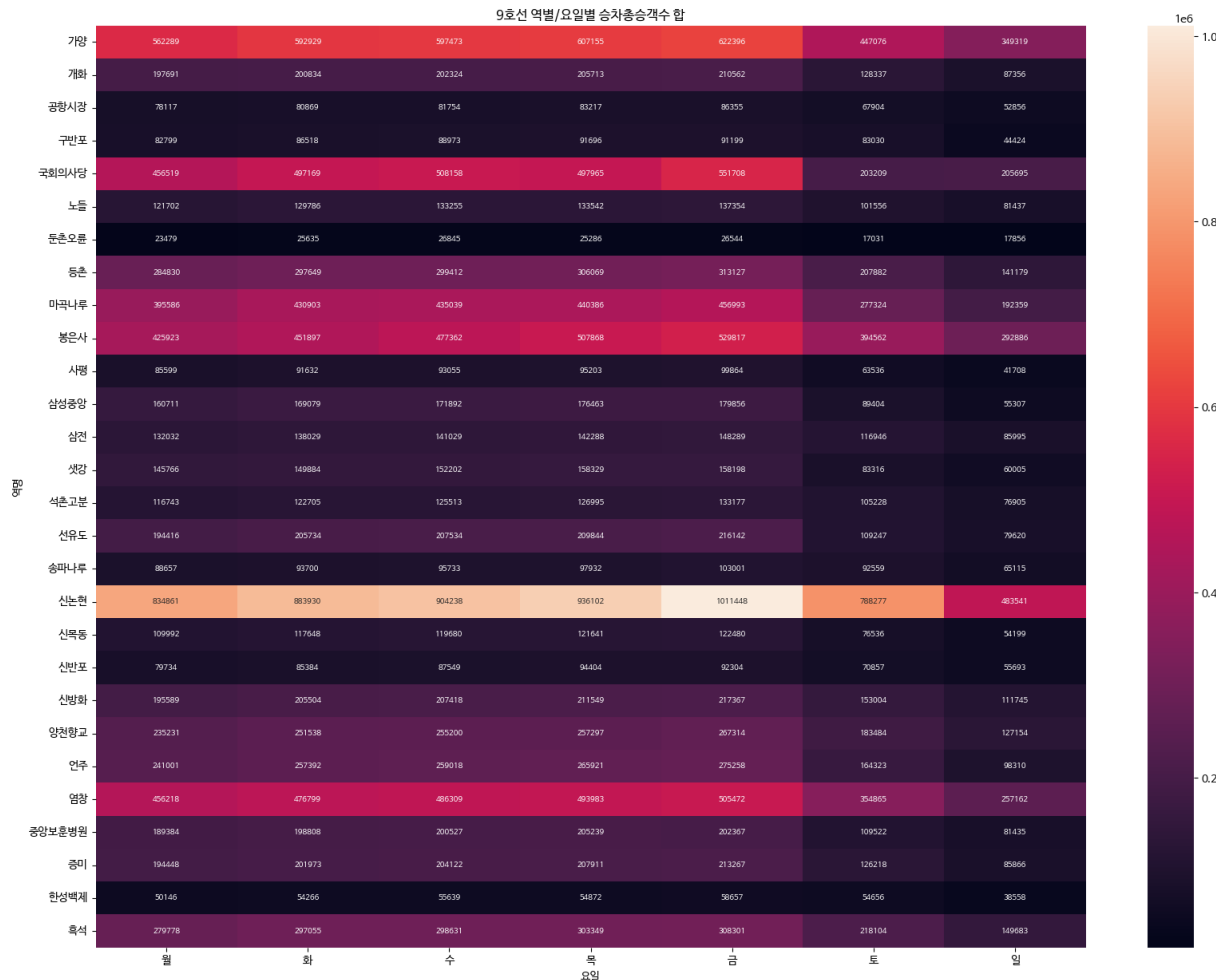
< 8호선 역별/요일별 승차총승객수 합 >



6. 노선별로 역별/요일별 승차승객수를 비교해 볼 수 있을까? (1~9호선, 역별/요일별 heatmap)



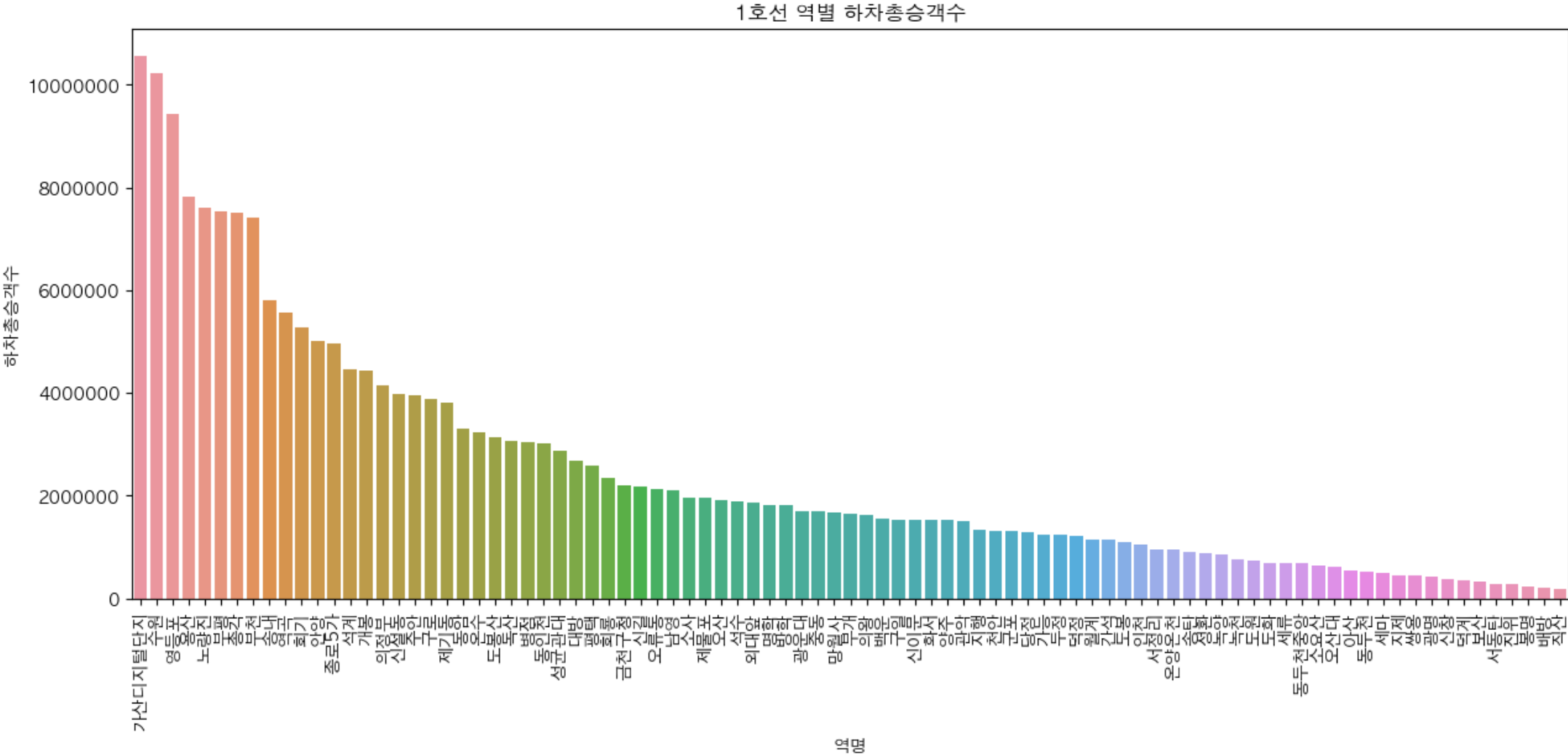
< 9호선 역별/요일별 승차총승객수 합 >



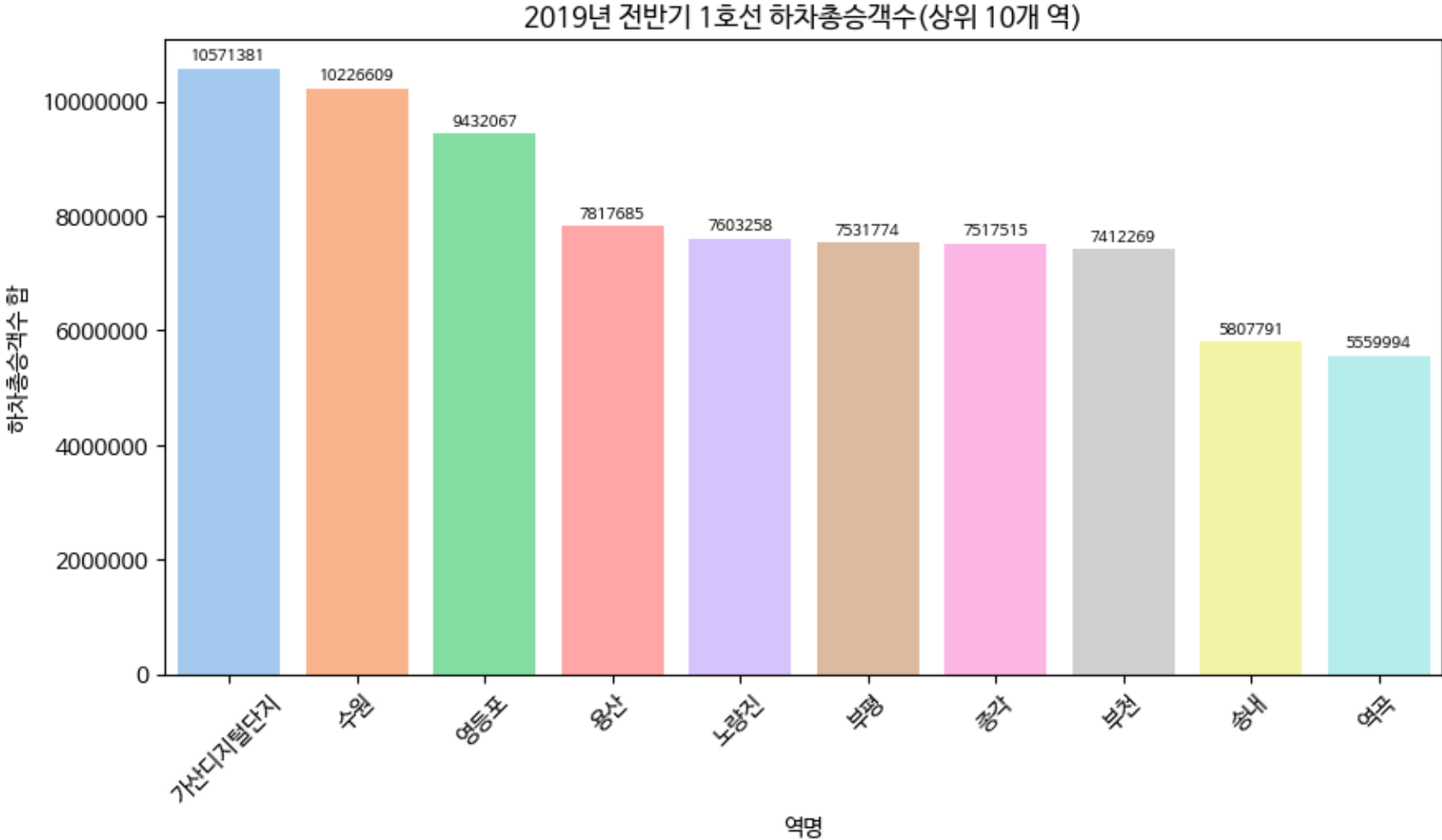
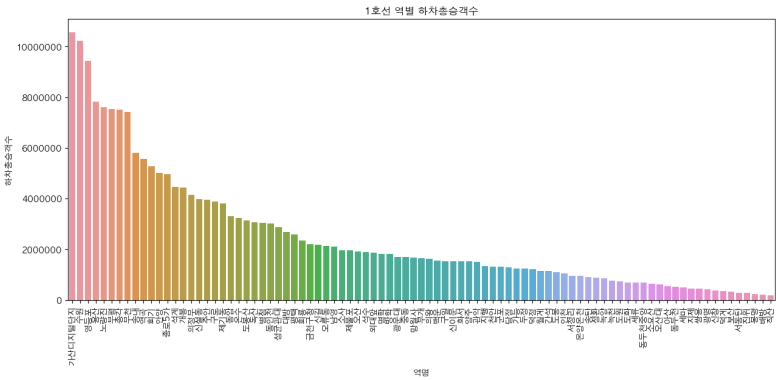
7. 1호선에서 가장 하차를 많이 하는 역은? (groupby)



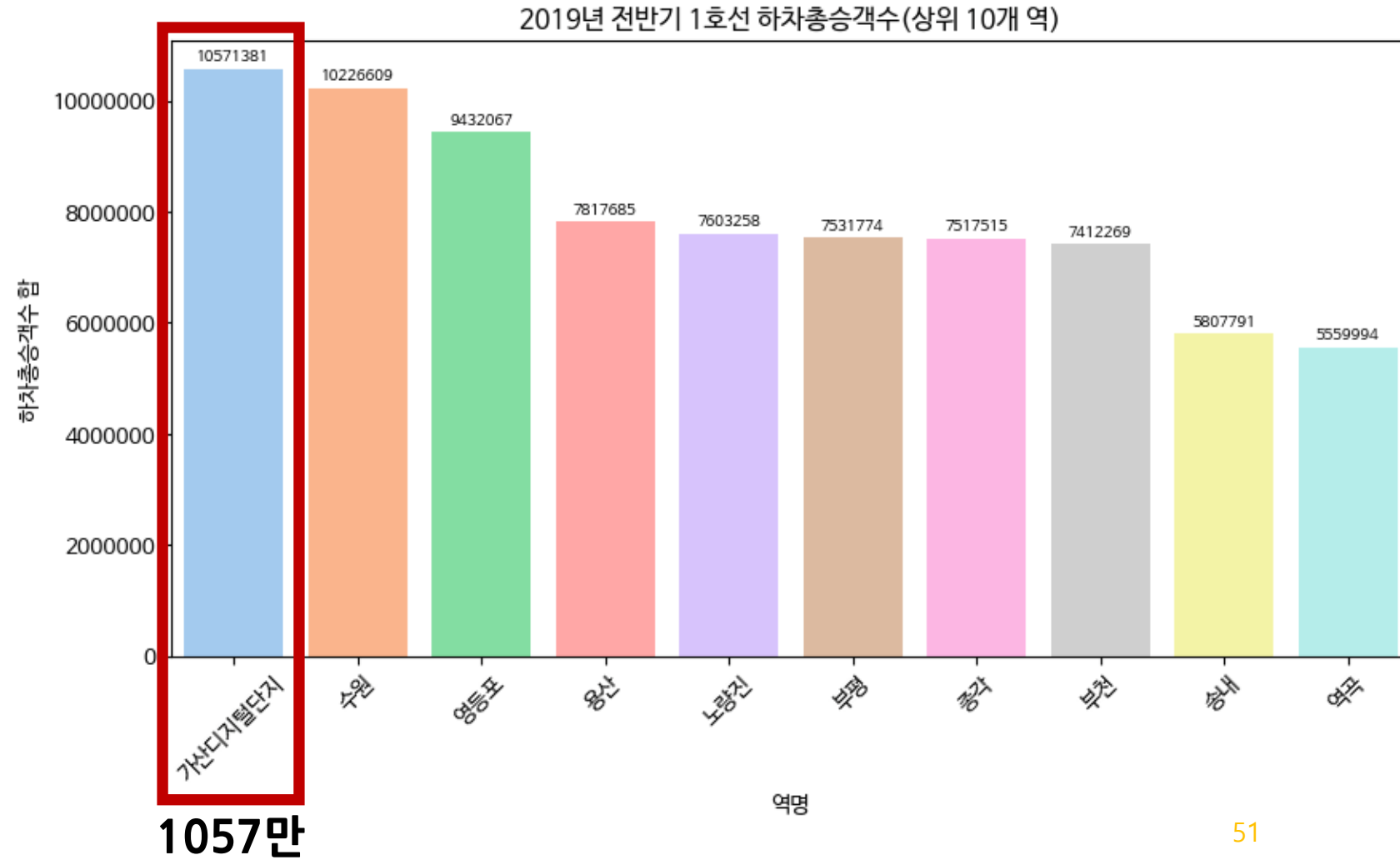
7. 1호선에서 가장 하차를 많이 하는 역은? (groupby)



7. 1호선에서 가장 하차를 많이 하는 역은? (groupby)



## 7. 1호선에서 가장 하차를 많이 하는 역은? (groupby)

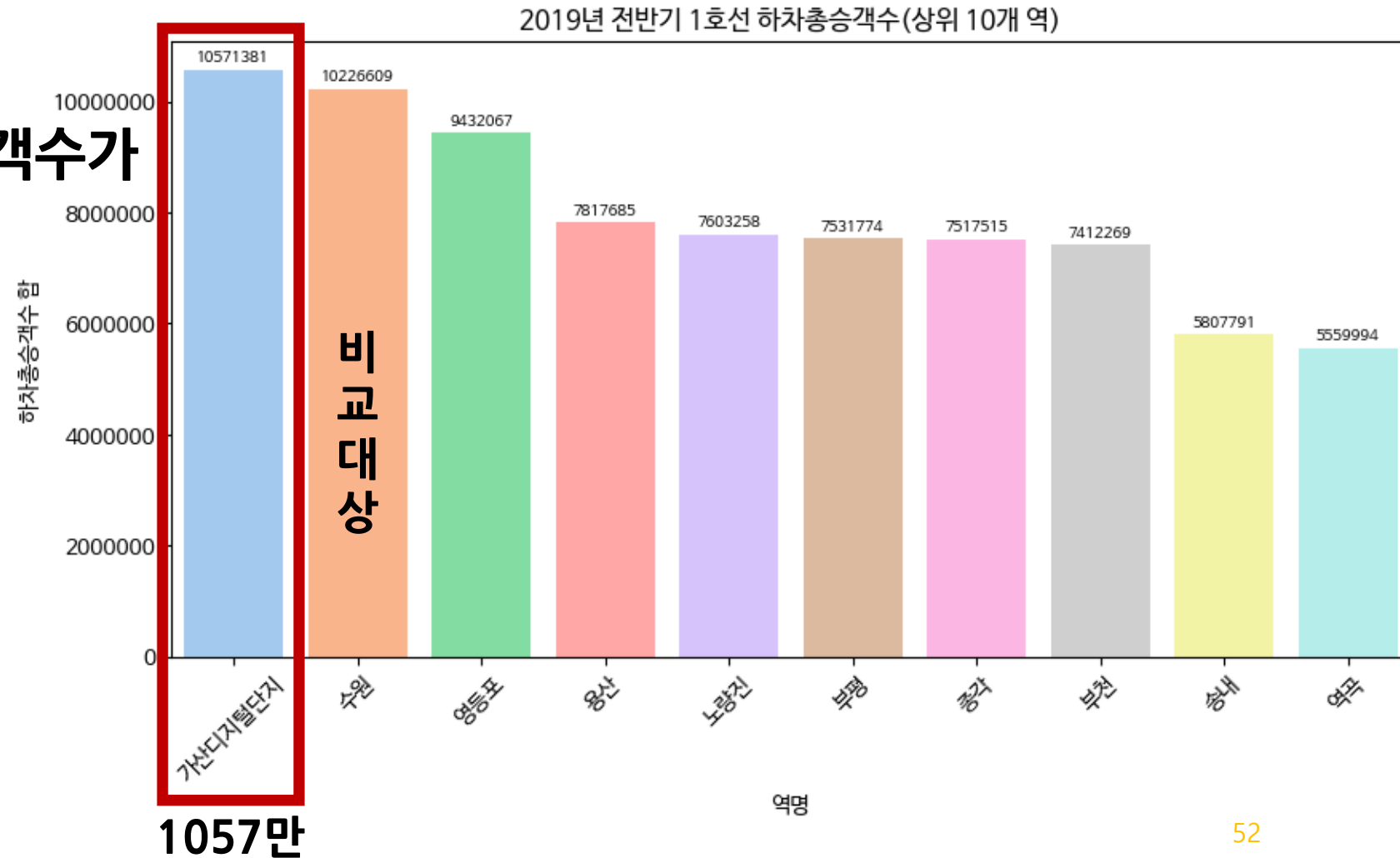


7. 1호선에서 가장 하차를 많이 하는 역은? (groupby)

가산디지털단지역 하차총승객수가  
수원역 하차총승객수보다

10,571,381 - 10,226,609

344,772



# EDA를 하면서.. 아쉬웠던 점

## ★역에 따른 행 개수 차이

- 의문: 서울역은 4개의 호선이 겹침.

1 4 공항 경의중앙

< 시청 1 서울역 남영 >

하지만, 사용일자별 서울역 포함 행의 개수는 5개.

	사용일자	노선명	역명	승차총승객수	하차총승객수	등록일자	역지역	위도	경도
26211	2019-01-01	4호선	서울역	5689.0	13266.0	2019-01-04	수도권	37.553316	126.969738
26212	2019-01-01	4호선	서울역	5983.0	731.0	2019-01-04	수도권	37.553316	126.969738
26213	2019-01-01	4호선	서울역	39420.0	31121.0	2019-01-04	수도권	37.553316	126.969738
26214	2019-01-01	4호선	서울역	1278.0	919.0	2019-01-04	수도권	37.553316	126.969738
26215	2019-01-01	4호선	서울역	14043.0	9955.0	2019-01-04	수도권	37.553316	126.969738

# Thank you

이창민, 황유성, 이준석, 맹의현, 진영인, 이예니

BETAGO