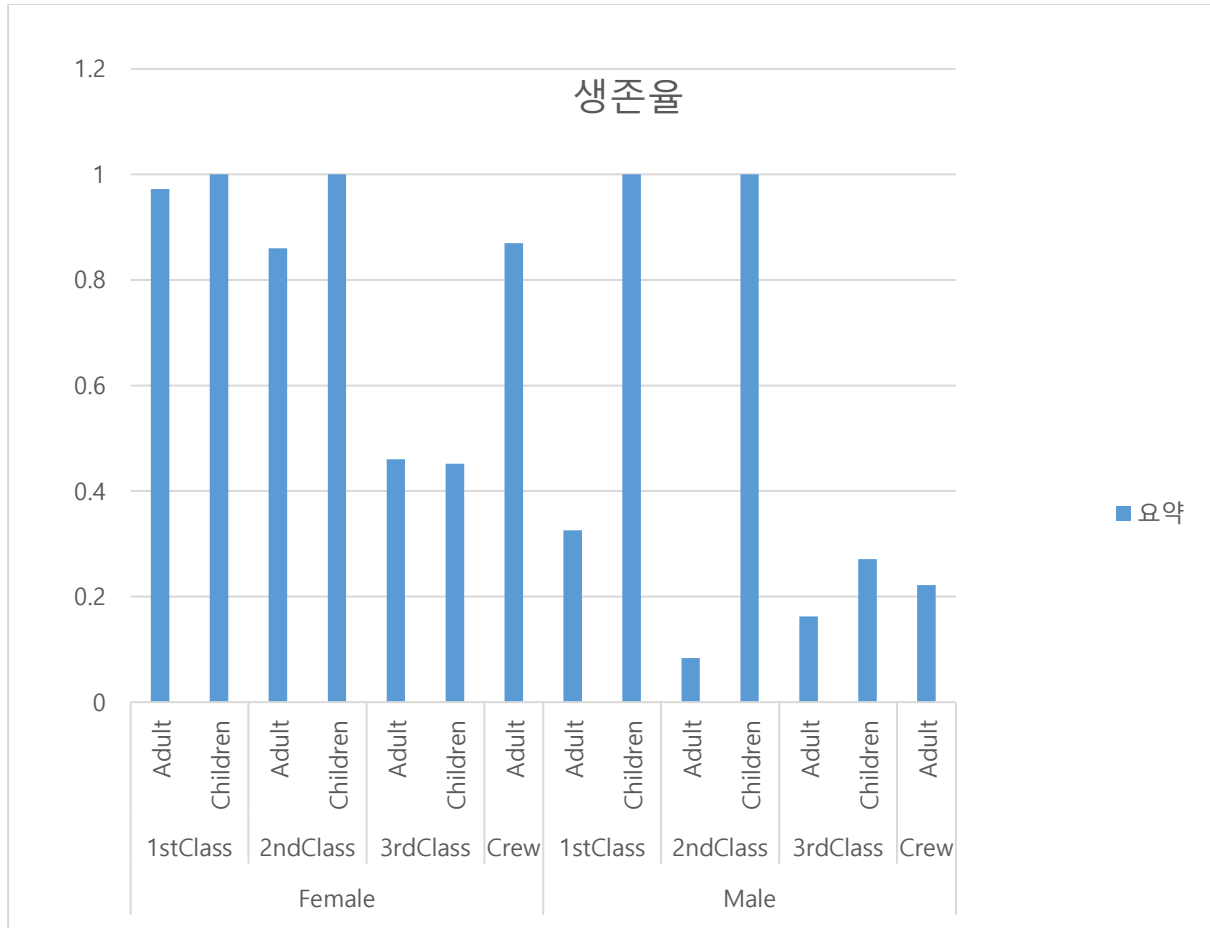


HW 3

2016314726 정영준

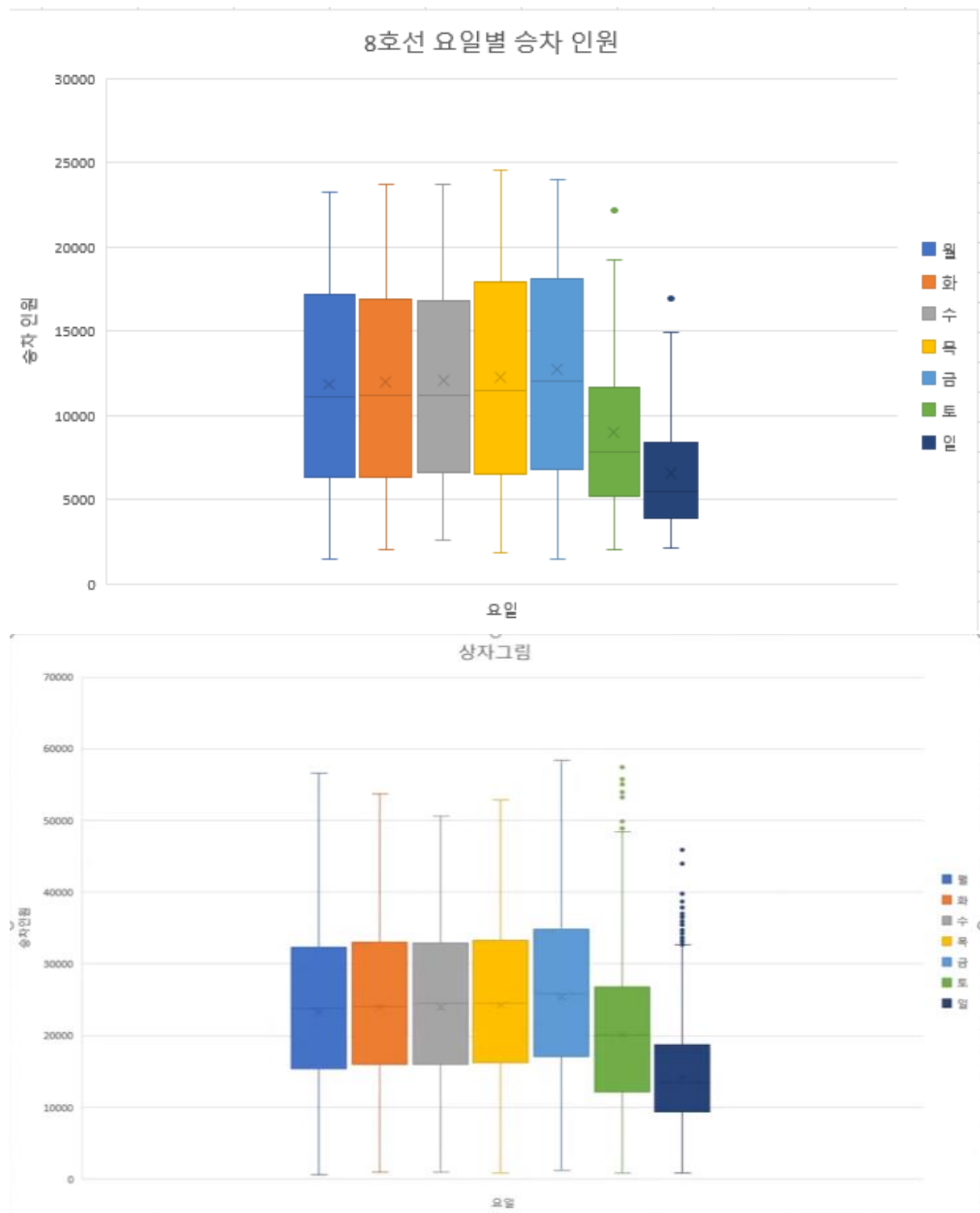
1.

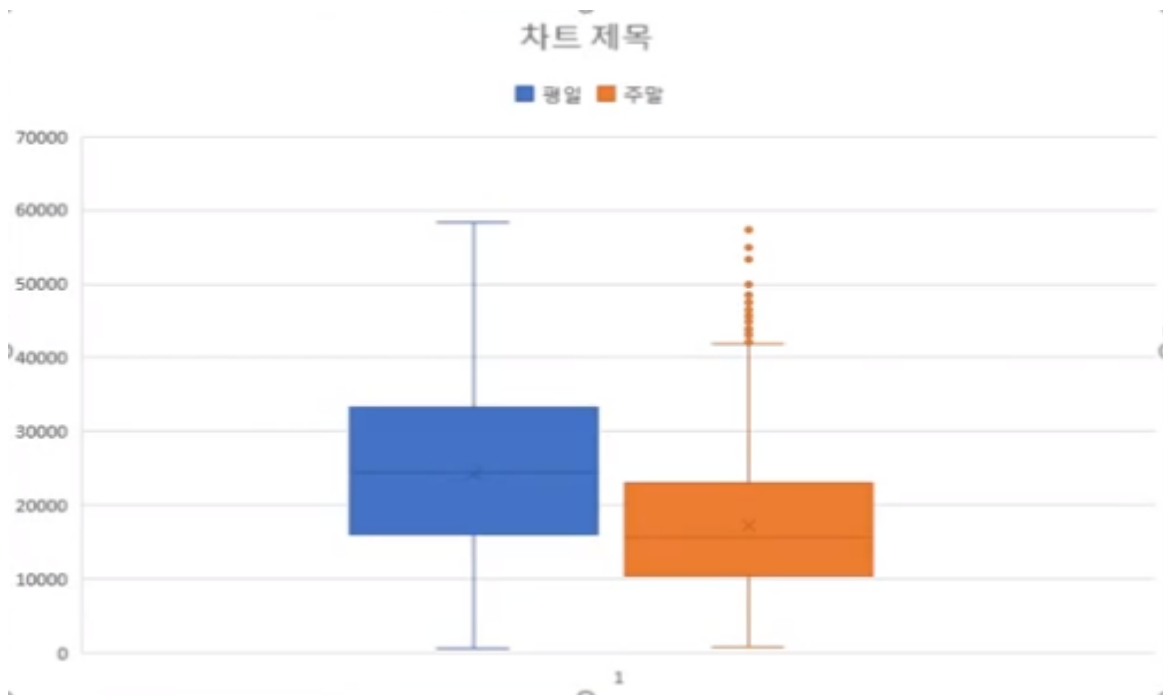
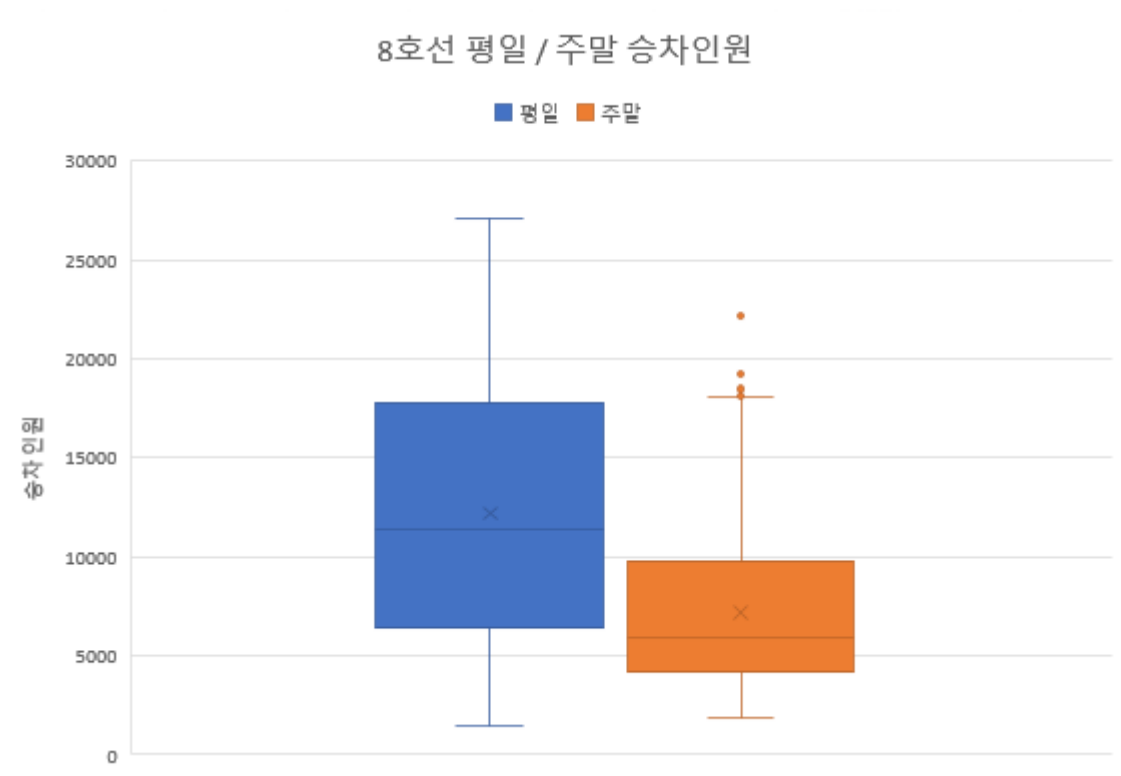


Gender에 따른 Class별 생존율의 차이를 추가로 Adult와 Children까지 파악하기 위해 작성한 그래프입니다. 먼저 Female과 Male의 동일 Class에서 생존율의 차이를 보면 1st Class에서는 어린이는 모두 생존하였고 성인은 1st Class의 여성의 경우 1st Class의 남성보다 훨씬 높은 생존율을 기록했습니다. 2nd Class에서도 동일한 결과가 관찰되었습니다. 3rd Class부터 생존율이 급격하게 감소합니다. 어린이들도 50%이하의 생존율을 기록하였으며 80%가 넘는 여성 성인의 생존율도 40%대로 급락합니다. 선원의 경우 여성 선원은 남성 선원보다 높은 생존율을 기록하였으며 3rd Class 동일 성별 승객보다 높은 생존율을 기록하였습니다. 1st Class와 2nd Class의 여성 또는 어린이의 생존율은 모두 80% 이상으로 높았으며 이는 높은 Class의 여성과 어린이의 구조가 먼저 이루어졌다고 생각할 수 있습니다. 남성 성인의 경우 모든 선실등급에서 낮은 생존율을 기록했는데 이는 여성과 어린이

의 구조를 도왔다고 생각해볼 수 있습니다.

2 - 1.





각각 먼저 나온 상자 그림이 8호선의 그래프이고 후에 나온 그래프가 4호선의 상자 그림입니다. 먼저 두 그래프에서 모두 승차인원이 4호선이 8호선에 비해 약 2배 가량 많은 것을 알 수 있었습니다. 요일별 승차인원 상자 그림을 보면 둘다 주말에 승차인원이 평일에 비해 적어지는 비슷한 분포를 보였습니다. 이를 자세히 보기위해 평일 주말로 나누어 이를 상자그림으로 그려보면 두 호선 모두 평일의 승차인원이 더 많고 주말의 승차

인원이 적었으며 주말 승차인원에서 outlier가 관측된 것을 볼 수 있었습니다. 이는 평일에 출근, 퇴근, 통학 등의 이유로 지하철을 이용하는 인원이 많기 때문이라고 해석이 가능합니다. 주말의 outlier는 특정 축제나 공휴일 등이 영향을 끼쳤다고 생각할 수 있습니다.

2-2

8호선 시간대 별 평균 승차인원

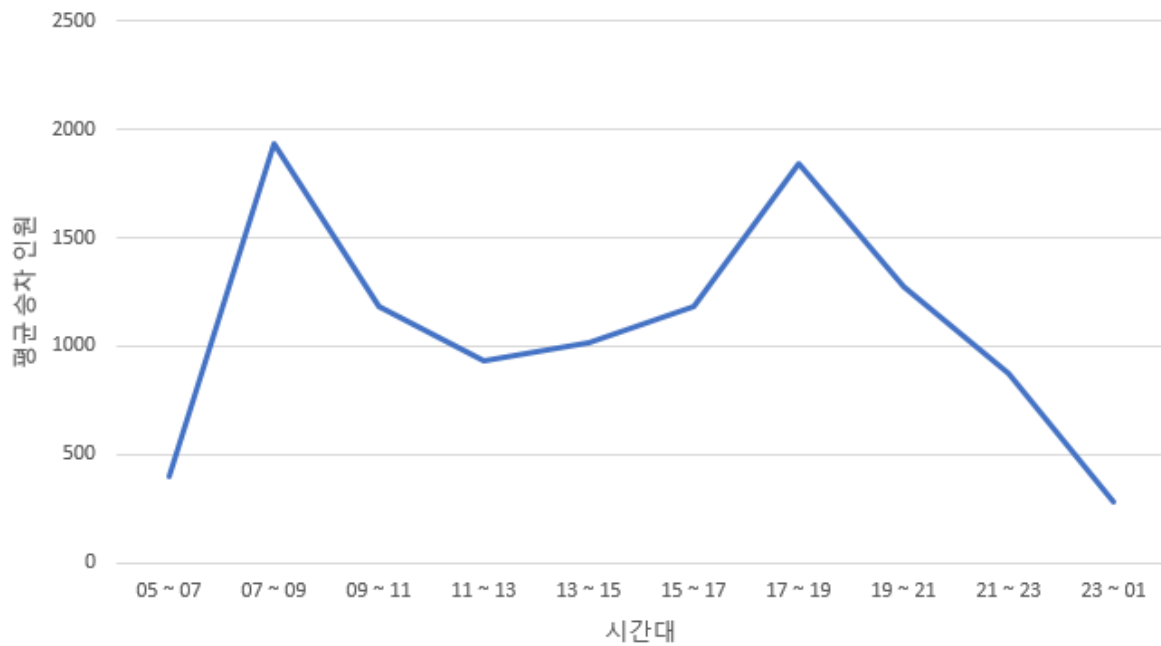


차트 제목



2시간씩 묶어서 시간대를 설정하고 이를 x축으로 하고 y축은 해당 호선의 전체 역의 승차 인원의 합을 날짜별로 평균을 낸 값을 나타냅니다. 위의 꺾은선 그래프가 8호선의 그래프이고 아래의 그래프는 4호선의 그래프입니다. 상자 그림으로 이를 나타내 보았는데 outlier가 많이 관측이 되었고 시계열 데이터이기 때문에 이를 꺾은선 그래프로 나타내는 것이 더욱 좋다고 생각하여 꺾은선 그래프로 나타냈습니다. 전체적인 분포는 비슷하나 8호선의 경우 7~9시의 평균 승차 인원이 가장 많은 반면 4호선에는 17~19시의 평균 승차 인원이 가장 많았습니다. 이 둘은 출퇴근 시간으로 4호선이 8호선보다 출퇴근 중 퇴근의 비율이 높고 8호선은 출근하는 인구의 비율이 높다고 생각해 볼 수 있습니다. 4호선에 위치한 역들이 8호선에 위치한 역보다 더 직장이 많다면 이런 결과가 나올 수 있다고 추측해 볼 수 있습니다.

3-1

	00시-02시	02시-04시	04시-06시	06시-08시	08시-10시	10시-12시	12시-14시	14시-16시	16시-18시	18시-20시	20시-22시	22시-24시
최소값	1,006	608	834	1,055	1,411	1,991	2,481	2,861	2,938	3,075	2,595	1,744
제 1 사분위	1308	753	854	1745	2993.5	2916	2899.5	3270	3773.5	3997	2939	2226
중앙값	1,449	776	875	1,994	3,659	2,990	3,002	3,401	3,927	4,388	2,981	2,373
제 3 사분위	1606	982	946	2030.5	3726.5	3157.5	3189.5	3464.5	4079.5	4423.5	3093	2526.5
최대값	2,124	1,354	1,268	2,154	3,789	3,242	3,543	3,599	4,423	4,910	3,578	2,951

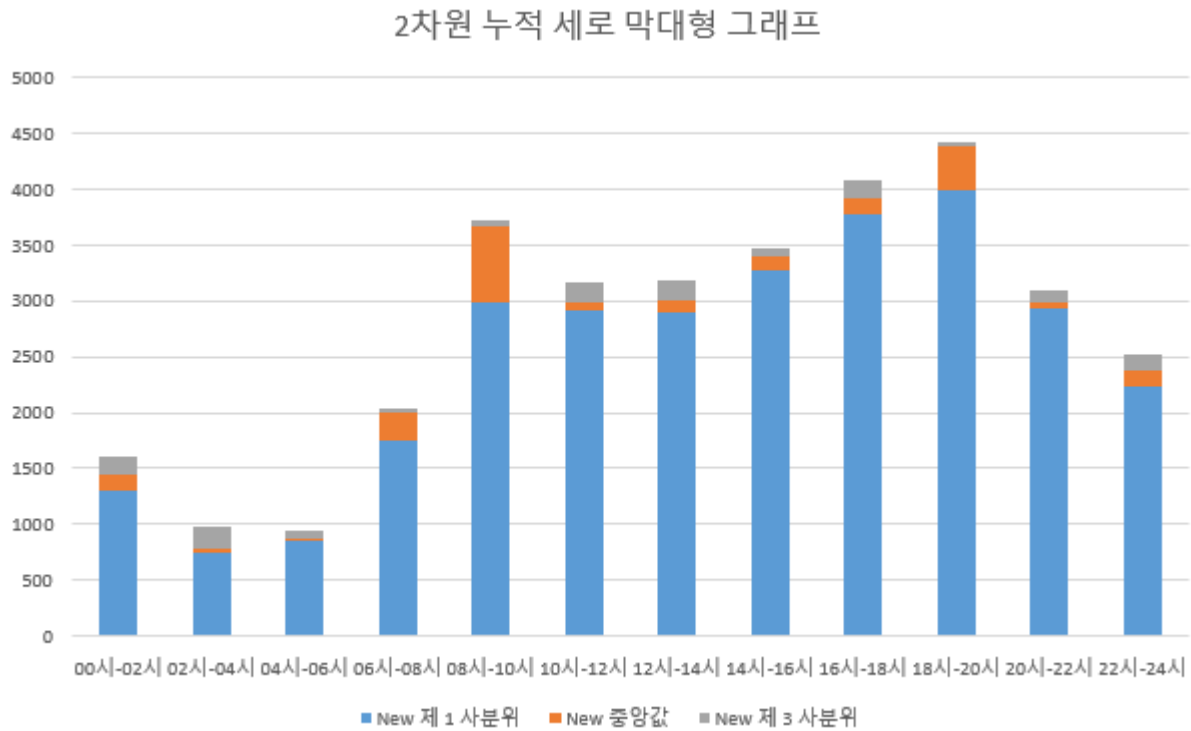
3-2

	00시-02시	02시-04시	04시-06시	06시-08시	08시-10시	10시-12시	12시-14시	14시-16시	16시-18시	18시-20시	20시-22시	22시-24시
New 제 1 사분위	1308	753	854	1745	2993.5	2916	2899.5	3270	3773.5	3997	2939	2226
New 중앙값	141	23	21	249	666	74	103	131	154	391	42	147
New 제 3 사분위	157	206	71	37	68	168	188	64	153	36	112	154

3-3

	00시-02시	02시-04시	04시-06시	06시-08시	08시-10시	10시-12시	12시-14시	14시-16시	16시-18시	18시-20시	20시-22시	22시-24시
New 최댓값	816	601	414	409	796	326	644	329	650	913	639	725
New 최솟값	302	145	20	690	1,583	925	419	409	836	922	344	482

3-4



3-5

