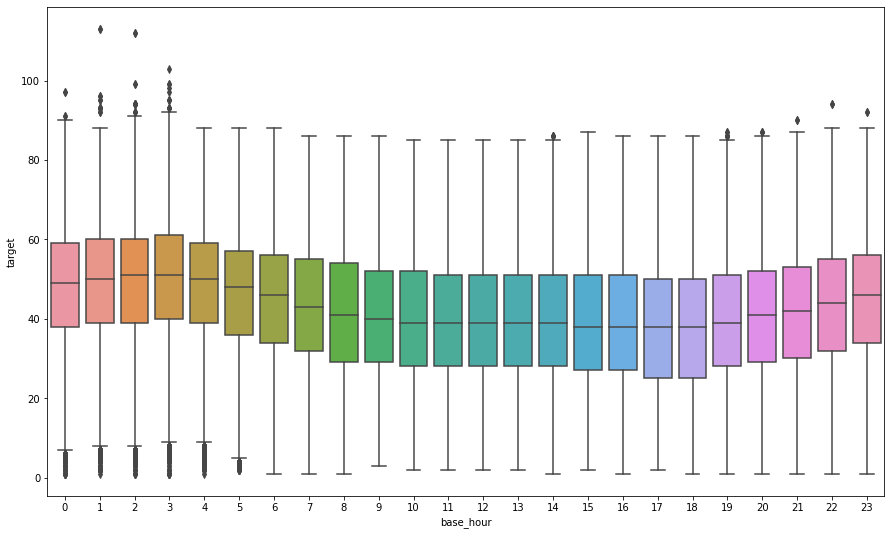
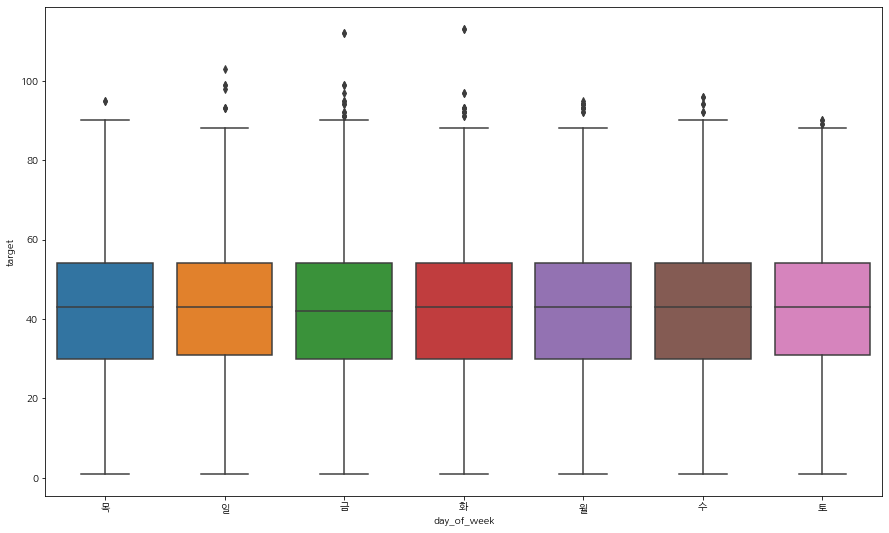
|  |
| --- |
| **Homework Assignment 3**  보고서 및 논문 윤리 서약  1. 나는 보고서 및 논문의 내용을 조작하지 않겠습니다.  2. 나는 다른 사람의 보고서 및 논문의 내용을 내 것처럼 무단으로 복사하지 않겠습니다.  3. 나는 다른 사람의 보고서 및 논문의 내용을 참고하거나 인용할 시 참고 및 인용 형식을 갖추고 출처를 반드시 밝히겠습니다.  4. 나는 보고서 및 논문을 대신하여 작성하도록 청탁하지도 청탁받지도 않겠습니다.  나는 보고서 및 논문 작성 시 위법 행위를 하지 않고, 명지인으로서 또한 공학인으로서 나의 양심과 명예를 지킬 것을 약속합니다.    학 과 : 융합소프트웨어학부 데이터테크놀로지  과 목 : 통계적 데이터 분석  담당교수 : 오민식 교수  강좌 번호: 6005  학 번 : 60211704  이 름 : 최지현 (서명) |

**Q1) base\_hour (시간)에 따른 교통량 분포를 알고 싶다. 시간에 따른 (교통량=target) 분포를 확인하기 위한 box plot을 그리고 시간에 따라 교통량이 차이가 있는지 본인 의견 작성**



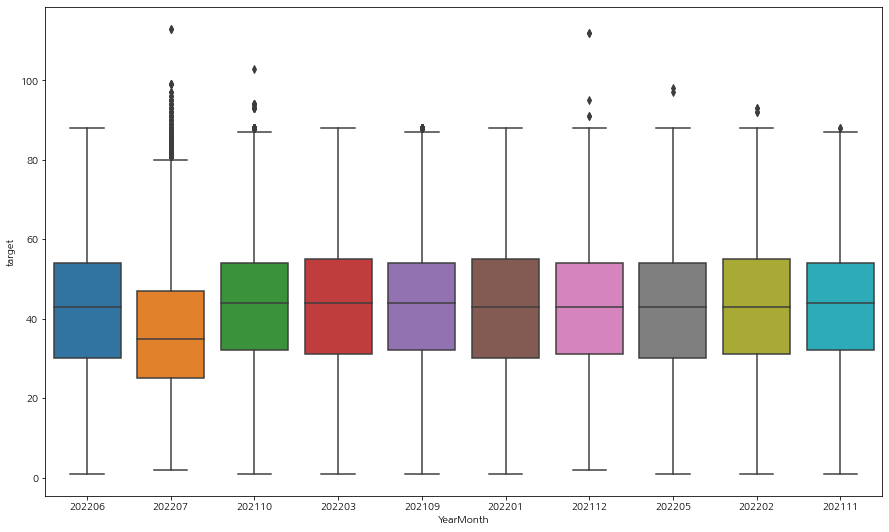
시간에 따라 교통량의 차이가 나타나 보인다. 00~03시에 평균적으로 높은 교통량을 보이고, 16~20시에 평균적으로 낮은 교통량을 보인다.

**Q2) 요일별 교통량 분포를 box plot으로 그리고 요일별로 차이가 있는지 의견 작성**



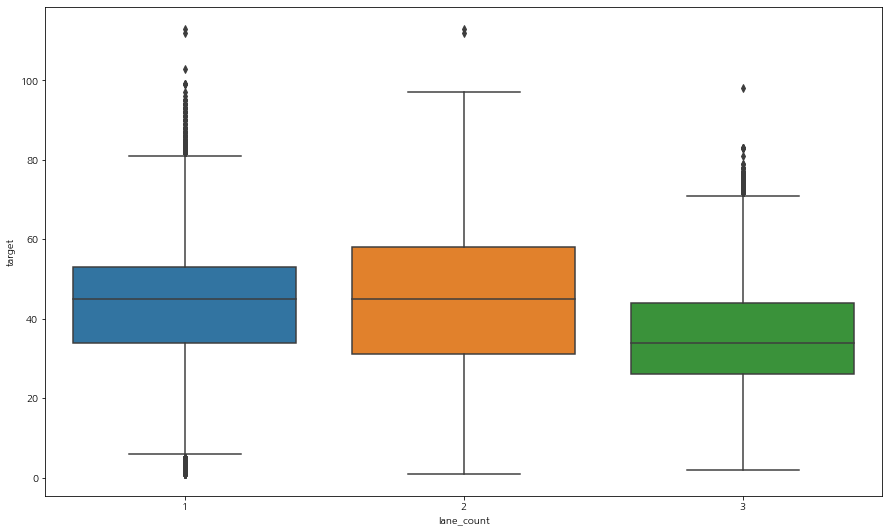
요일별로 교통량 분포는 거의 차이가 없어 보인다. 각각의 box plot의 위치가 거의 동일하게 나타난다.

**Q3) 연도+월별 교통량 분포를 box plot으로 그리고 시간이 흐름에 따라 교통량 차이가 있는지 의견 작성**



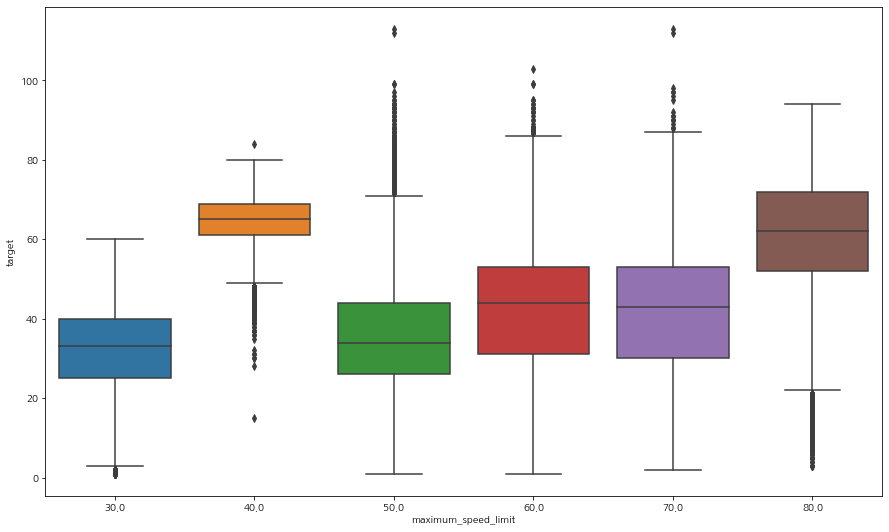
차이가 있다. 202207의 경우 가장 낮은 중앙값을 가지고 각각의 박스플랏들의 모습이 연도, 월별에 따라 차이가 있다.

**Q4) lane\_count별 교통량 분포를 box plot으로 그리고 lane\_count의 차이가 교통량 차이와 관련이 있을지 의견 작성**



Lane\_count의 차이가 교통량의 차이와 관련이 있다고 생각한다. Lane\_count에 따라 각 교통량의 IQR이 다르게 나타나기 때문이다.

**Q5) maximum speed limit 별 교통량 분포를 확인하면 상식과 위배되는 구간이 하나 나오는데 그게 어떤 구간인지 작성 – 왜 그럴까? (추가 점수 10%)**

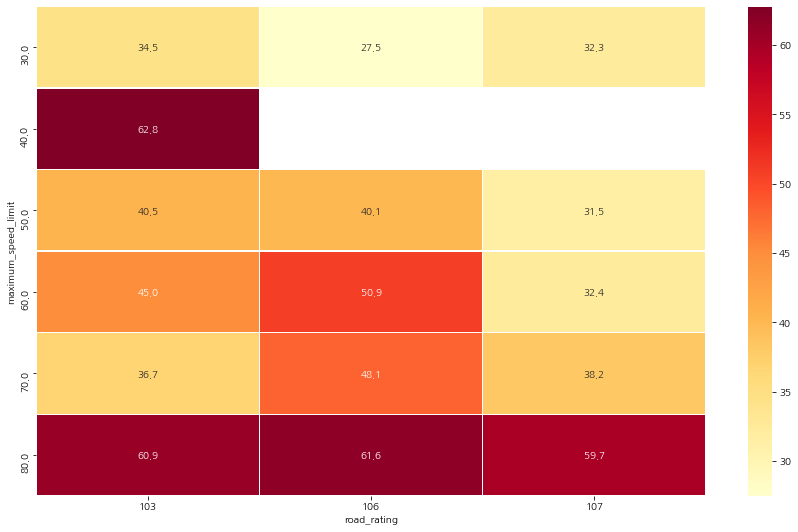


Target 값이 80~90인 구간이다. 왜냐하면 도로의 등급이나 다른 요인에 따라 교통량 분포가 많아질 수 있기 때문이다.

**Q6) box plot으로 분포 확인이 불가능한 변수가 있나? 있나/없나 에 대한 이유를 작성**

Box plot으로 분포확인이 불가능한 변수가 있다고 생각한다. 시작/도착지점의 위/경도의 경우는 너무 폭 넓은 데이터로 데이터의 분포를 확인하기가 어렵기 때문이다.

**Q7) Q6에서 가능한 변수들 중 시각화를 통해 교통량과 상관이 제일 큰 것 같은 두 변수를 선택하고 두 변수에 따른 교통량 분포를 알기 위한 heatmap 그리고 어떤 특성이 있는지 작성 (추가 점수 10%)**



최대 속도 제한이 큰 도로일수록 교통량 분포가 높게 나타나고 도로등급이 106인 도로에서 대체적으로 교통량이 높으며 도로등급 103, 최대속도 40인 도로도 교통량이 높게 나타난다.