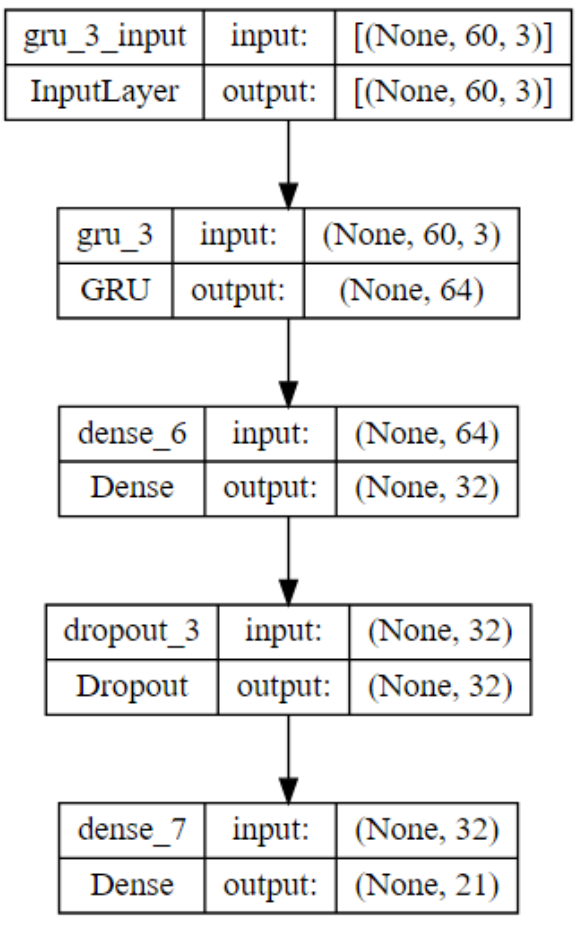
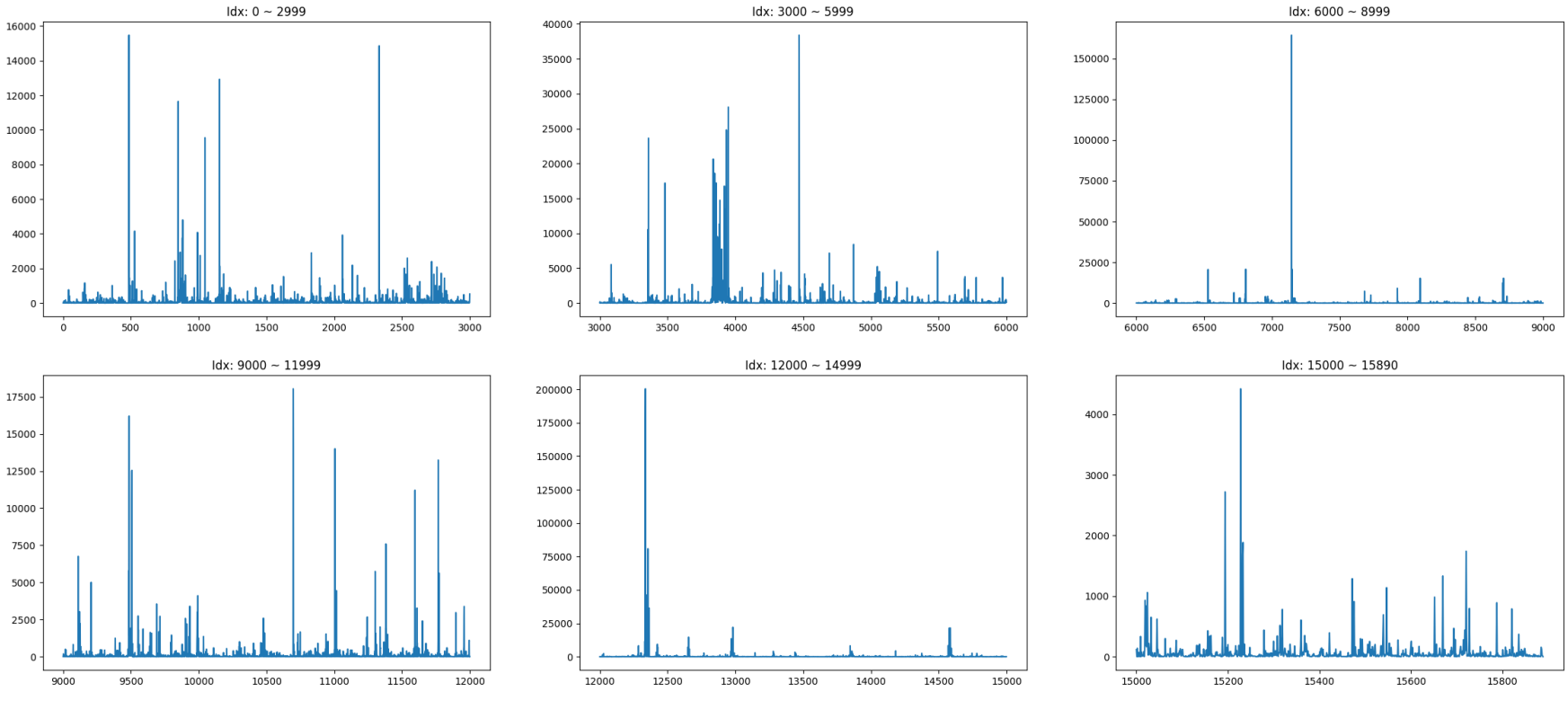
브랜드 언급량 전체 평균 추가

- 현재까지 제출한 성능 중 가장 좋음.
  - **Validation PSFA: 0.5855, val_loss: 0.173, Dacon Score: 0.5252(14일 15:00시 기준 72위)**
  - 새로운 파생변수 도입시 PSFA점수보다 높아지고, val_loss가 낮아지면 더 점수가 높아질 것으로 기대
  - 모델링 시 마지막 Epoch에서 PSFA Score를 계산하도록 Callback 함수 구성
    - 각 Epoch이 끝난 후 구해봤는데 RAM 문제로 인해 Session이 날아감.
- **사용 변수**
  - 제품 소분류 코드: Label Encoding
  - brand_mean (Min_Max_Scaling 후 전체 평균 취함): 7일, 30일, 60일 등 여러 일자 평균을 시도해 봤으나 전체 평균을 취하는 것이 가장 좋았음.
- **사용 모델 (GRU)**

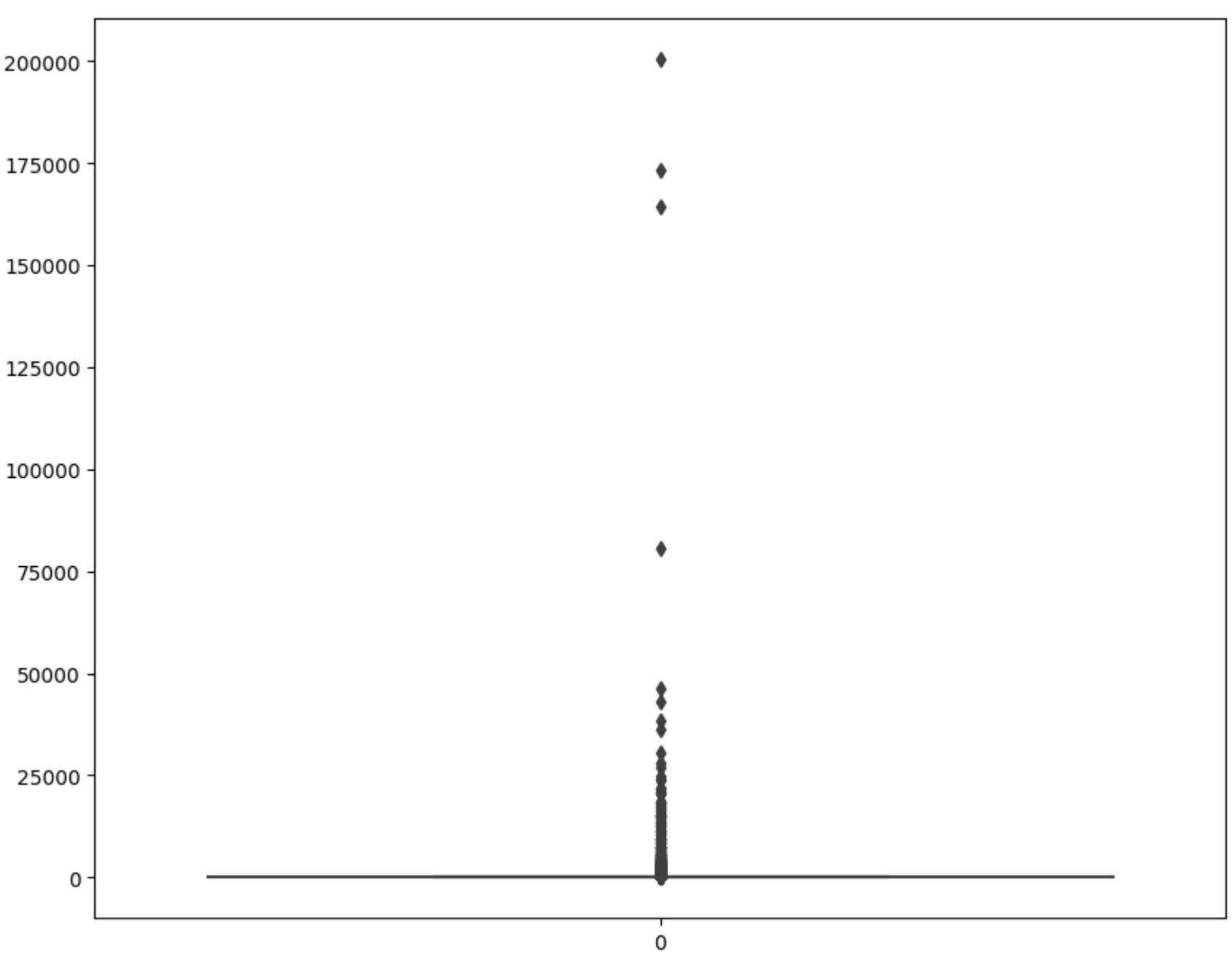



```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 gru (GRU)                   (None, 64)                13248

 dense (Dense)               (None, 32)                2080

 dropout (Dropout)           (None, 32)                0

 dense_1 (Dense)             (None, 21)                693

=================================================================
Total params: 16,021
Trainable params: 16,021
Non-trainable params: 0
_________________________________________________________________
<keras.engine.sequential.Sequential at 0x7e890fef5e70>
```

- **시도해봤으나 좋지 않았던 방법**
  1. 일별 판매량(sales.csv)를 Min_Max_Scaling하여 전체 평균으로 도입하였으나 성능이 떨어지는 경향이 있음.
  2. 각 제품의 판매량 중 이상치(Outlier: Q3 + 3 * IQR = 236.0) 이상인 판매량을 모두 236.0으로 대치 후 진행 → **Validation PSFA: 0.61 val_loss: 0.201, Dacon Score: 0.4027**



Picture 1. 각 제품 별 MAX 판매량 값



Picture 2, 제품 별 최대값의 Boxplot

- Target 값 원본을 수정하는 방법으로 너무 많은 데이터가 영향을 받음. → 성능이 상당히 좋지 않음.
- 특정 기준(판매량 5000?)을 넘는 최댓값들을 두번째 최댓값으로 대치 한 후 모델링 해보는 방법?
- 다른 추가적인 방법이 있을지?

3. 추가적인 방법이 있을지 고민 중