

8월 11일 회의

파생변수를 어떻게 만들지 고민

- 손형락: 중분류 코드별로 모델을 제작. make dataset 활용. gru 활용. 0.477 나옴.
- 오승태: 모델 쪽은 건드리지 않음. 새로운 피쳐를 만들고 파생변수를 만들면 성능이 잘 올라갈 것 같음. 제품이 똑같으면 소분류가 다름. 결측값은 없었음. 이상치 확인 방법? 공휴일, 이벤트 날 확인해서 날짜에 가중치. 제품 특징을 넣어서 라벨 인코딩으로 묶기. 브랜드 관련 메타 데이터는 감이 잘 안 잡혔음. → 이상치를 빼는 건 좋지 않은듯
- 이세은: 대상을 분류하자. 이상치 탐지해서 모델에 추가하는 방향.
- 가격을 수량으로 나눠서 일별 가격 변화를 확인하자
- 주말을 토일로 묶고, 요일별로 묶는 것도 좋을듯
- 제품 판매량은, 제품이 얼마나 팔렸는지보다, 가격이 제품 판매량이 연관성이 있다. 제품 판매량과 가격이 비례할 거라고 가정하면, 월화수목금토일의 제품 판매량을 알 수 있을 것이다.
- 기간 조정을 하는 것도 생각해보자

역할 분배

1. 요일, 월, 계절 별 파생변수 -승태
2. 가격을 수량나눠서 일별 가격 변화 확인 후 파생변수 생각 - 12(보류)
3. 제품 특성 데이터 가져와서 제품별 특징 파생변수, 특정한 대상이 있는지도 가져와보기 - SYL
4. 브랜드 데이터는 전체 평균으로 가져와서 브랜드별 파생변수 - 형락
5. 이동평균 파생변수 조사해보면서 몇달치 넣을지 정하고 파생변수 넣어보기 -세은