

7/20 팀 진행상황 공유

- 1. Train에서 공급유형이 장기전세, 공공분양, 공공임대(5년)인 example들의 단지코드 가 차지하는 비율 살펴보기

이 공급 유형들이 Test에 나타나지 않는다는 점과 전체 example 중 0.7%, 특히 이들을 제거해도 전체 단지코드 중 0.34%의 단지코드만이 사라진다는 점을 감안하면, 이들을 제거하여 성능 평가를 해보는 것도 좋을 것 같다.

- 2. 하나의 단지코드내에 임대건물구분 값 존재 양상(상가/아파트/상가&아파트)에 따른 등록차량수 살펴보기

주상복합 일 때 주차 수요가 더 높을 것이라는 예상은 옳지 못하다.

'아파트'일 때와 '상가&아파트'일 때 2가지 경우로 LabelEncoding하는 것이 더 좋아보인다.

단, '아파트'일 때의 주차 수요가 더 높았으므로 '아파트' -> 1, '상가&아파트' -> 0 으로 매핑해보자!

- 3. 자격유형 별로 임대보증금 과 임대료 의 평균값들을 계산하여 소득수준과 관련한 판단 지표가 될 수 있는지 확인하기.

'자격유형별 평균 임대료'를 feature로 이용하여 성능평가를 해보는 것도 좋아보인다.

- Age_gender_data 활용하기

- 지역 별로 가장 많이 거주하는 상위 3그룹을 나타낼 수 있도록 구성

	1st	1gender	2nd	2gender	3rd	3gender
지역						
경상북도	50	0	40	0	60	0
경상남도	60	0	50	0	40	0

- 이를 train과 test data에 지역을 기준으로 concat할 예정!