1. 올바른 값 넣기

- 임대보증금, 임대료에서 '-' 라는 값이 존재
  - 이를 Null로 변경하고, dtype을 object에서 float로 변경해준다.

2. NULL 처리하기

- Train 에서 Null을 포함하는 feature
  - 임대보증금, 임대료, 지하철, 버스
- test에서 Null을 포함하는 feature
  - 자격유형, 임대보증금, 임대료, 지하철
- Null data 처리
  - 임대보증금, 임대료 (공통)
    - 임대보증금이 null이면 임대료도 null이다.
    - '-'인 경우도 모두 null로 만들어 주었으므로, **null을 모두 0으로 만들기**
  - 지하철, 버스
    - (최영민) Null 값을 모두 0으로 만들기
    - (이채원) 버스는 Null->-1, 지하철은 지역별로 코드 다르게 만들기.
      - 충청남도-> -1
        대전광역시-> -2
        경상남도-> -3
    - Null이 들어있던 example들 분석
      - 지하철
        - Train
          - Null로 되어있던 example의 단지코드 가지 수 : 20
          - Null로 되어있던 example의 수 : 211
        - Test
          - Null로 되어있던 example의 단지코드 가지 수 : 5

        - ◆ Null로 되어있던 example의 수 : 42
    - 버스
        - ■ Null로 되어있던 example의 단지코드 가지 수 : 1
        - ■ Null로 되어있던 example 수 : 4
- ■ 자격유형 (Test)
    - ◆ Null인 example이 단 2개 (단지코드 : C2411, C2253)
    - ◆ C2411
        - A만 존재 -> **A로 채우자**
    - ◆ C2253
        - 임대보증금, 임대료가 존재하면 C, 없으면 D로 설정 되어있다.
        - 해당 example은 존재하므로 **C로 채우자**

3. 중복 확인

- train에서는 320개, test에서는 73개의 중복 데이터가 존재했다.
    - ■ 제거하는 것이 좋아 보인다.
- 단지 코드별로 feature별 중복 양상도 체크 해봐야할 것 같다.

4. 지역명 숫자로 매핑

5. 전용 면적 값 단순화

- 5의 배수로 변경
    - ■ 5로 나눈 몫에 5 곱하기
- 상/하한선 적용
    - ■ 상한 : 100
    - ■ 하한 : 15
    - ■ 이 때의 unique 값 (train : 15, test : 14)

6. 단지코드를 기준으로 데이터 취합

- Ex) 같은 전용 면적을 가지는 세대 수를 계산해서 하나의 feature로 만들기

7. baseline에서 feature 사용 변화

- 공급유형, 임대료, 임대건물구분, 임대보증금, 자격유형