subway :

1. line이 여러개 지나는 환승역 주의

2. 라인 수에 따른 가중치 여부 확인 필요

3. address of law 결측치 주의

schools :

1. 초 중 고 학교별 가중치 확인

2. 운영주체 별 가중치 확인

3. 고등학교 구분별 가중치 확인(일반, 자율, 특성화, 특수목적), 결측치 많음

4. 남고 여고 공학 여부별 가중치 확인 (남고의 경우 집값 하락 가능성 있다?)

5. 설립일 효과 확인(오래된 학교, 최신 학교별 영향력 확인 필요. - 명문학교 효과?, 최신시설 효과? 등 확인 필요)

train

1. city 별 가중치 확인 (서울이냐 부산이냐 / 두 지역 모두 대도시이므로 상호간의 영향력은 적을 듯?)

2. transaction year month 영향력 있는가? (발생일과 연관지어서 영향력 확인해얄 듯)

3. transaction date 분류가 영향력 있을 것 (11~20 등으로 구분되어져 있으므로 어떻게 배정할 것인지가 관건일 듯/ 구분 방법 생각해 보기)

4. year of completion 영향력 확인 ( 최신일수록 높은가? 오래될수록 낮은가? -> 만약 상관 관계가 적다면 위치가 중요한 요인인가? 가격에 얼마나 영향 주는지 확인 필요)

5. exclusive use area 면적별 가격 확인 필요 (넓을수록 비쌀 것으로 예상되지만 확인 필요할 듯)

6. floor (고층 저층 여부의 가중치 확인)

7. 경위도로 위치 확인할 것인가 확인 필요.(법정동코드와의 관계 확인 필요할 듯/ 법정동 코드를 대분류라고 한다면 경위도로 자세한 위치 확인할 것인가, 범위로 나누어서 구분지을 것인가)

8. total parking capacity in site (주차대수가 높다면 -> 이것을 차가 많아 불편해 가격이 낮을 것으로 봐야하는가 아니면 세대 당 차가 많다는 의미로 부유층의 거주지로 봐서 가격이 높다고 볼 것인가?/ 해석 방향에 따라 다른 결과 나올듯)

9. total household count in sites (위의 주차대수와 관련된 공식 도출할 수도 있을듯, 세대별 차량 대수 등으로 거주자의 재산 정도를 파악해서 가격에 가중치를 준다거나/ 세대별 차가 적다면 편리하므로 집값이 좋다고 볼것인가/

혹은 세대 자체에 집중해서 세대가 많을수록 가격이 낮아지는가 높아지는가를 확인할 수 있을 것 -> 세대가 너무 많으면 낮을 것이란 예상?)

10. apartment building count in sites (단지 내 총 동수를 가지고 세대수와 연관시켜 세대별 사용 면적 등도 도출할 수 있을 듯. 얼마나 사람이 많은가에 따른 불편 정도 등 여러 분야로 생각해 볼 수 있을 듯.)

11. tallest building in sites, lowest building in sites (최고층과 최저층 동의 동수인데..어떤 식으로 해석할 것인가? index1번의 경우에 아파트 단지 내에 건물이 1채인데 최고층이 8이고 최저층이 4라는 게 말이 되는가? 지하를 의미? 확인 필요할 듯)

12. heat type 난방 방식에 따른 영향 (중앙난방의 경우에는 주공아파트가 많음/ 이에 민자 주공에 따른 영향 방식으로 연동해 해석할 것인가 자체적으로 구분지어 가격에 가중치 줄 것인가 생각해 보기. 난방연료와도 연관지을 수 있을듯)

13. heat fuel (난방연료에 따른 영향력도 확인해 볼 수 있을 것. 도메인 공부 필요함. 어느 부분이 좋은가/ 어느 시공사가 짓고 어느 사람들이 살고 어느 지역에 있는지에 따라 연료 방식에 차이가 있는가? 도메인 공부 필수!!!)

14. room ID 평형 ID로 단위가 평형임을 주의하자.(QnA에 따르면 외국인도 참여하기 때문에 한국의 평형 부동산 정책에 대한 이해력을 높이기 위해 넣은 feature라고 한 듯)

15. supply area (공급면적 = 전용면적 + 주거공용면적 이라는 공식 주의! 전용면적과 주거공용면적에 대한 이해 필요. 두 요소의 비율에 따른 차이가 있는지 확인 필요,)

16. total household count of area type 평형의 총세대수 (이거 무슨 말인지? 확인 필요. 특정 범위당 세대수를 의미하는건가? feature 확인 필요함.)

17. room count 방 개수 당 영향력 확인할 것.

18. bathroom count 욕실 수 (욕실과 방 개수 사이의 차이나 비율 확인을 통한 가중치 확인해볼 것.)

19. front door structure 현관구조별 영향력 확인할 것, (현관구조와 다른 구조와 관련된 feature들간의 상관 관계 확인해 볼 것. 계단식이나 복합식의 경우에는 아파트 세대수 등의 요소들과 연관 관계 잇을 듯. 복도식 아파트의 경우에는 세대수가 많을 것 같음)

20. transaction real price 실거래가 확인

글자로 된 feature들

heat\_type

heat\_fuel

front\_door\_structure

-> 요소들을 index화할 것인가 onehot 인코딩 할 것인가 생각해볼 것.

transaction의 경우에는 범주화되어 있는데 어떻게 구분할 것인가.

apartment id , address by law 등 feature값들의 범위 조절 필요할 듯. 어느 것들은 1~ 한자리수도 있고, 어느 것들은 여러 자리수이므로 조절 필요함.