5/15 건축물 부분

일단 박순만 교수님 미팅 후 엄청나게 데이터를 쳐냈음. 어차피 이제 중점은 이게 아니니까 패스.

통합 부분은 거의 진행 완료됨. 발표자료에는 그냥 됐다고 해 두는게 우리 건강에 좋을 듯.

???: 아직도 데이터 전처리가 안 끝났어?

지역지구 쪽 데이터만 붙이면 완성인데 일단 그거 빼놓고는 완료. 근데 갑자기 드는 생각은 지역 지구구역을 어떻게 활용하는가? 였는데 구역 별 입점 가능 업종이 있었다는 사실이 이거 치면서 기억남.. 밤 새서라도 노가다해서 조건 다 걸 예정.

또 하나의 이슈! 층별개요정보 코드가 최신 자료에서 숫자가 한글로 바뀌었다?! 노가다를 해서 해결할 예정.. 미래의 나 화이팅!

차주부터는 모델 제작기에 들어갈 예정. 제 생각은 Logistic Regression은 Supervised Learning인데 우리는 건축물 데이터에 어떤 업종이 들어가야 가장 좋은지 알 수 없음. 즉, 문제집은 있는데 답지는 없는 상황. 그래서 상권 데이터에 업종별 linear regression model을 만들어서 건축물 데이터 쪽에 업종에 대한 column을 만들어서 그것을 바탕으로 Logistic Regression model을 만들어야 하지 않을까 싶음.

예상으로는 모델 그냥 만드는 건 복붙이 해결해주기 때문에 hyper parameter tuning이 가장 큰 문제가 될 것으로 보임.

+) 구조가 바뀌었는데 그림을 다시 그려볼까요?

+) Logistic Regression에 사용할 업종을 건축물 데이터에 주기 위한 다른 방법이 있을까요?

--------------------------------------------------------------------진행도--------------------------------------------------------------------

건축물 데이터 전처리 부분 99.9% 완료

Logistic Regression & Linear Regression: 25%

1. 데이터 전처리 (100%)

2. 업종별 예상 매출액 도출하는 Linear Regression Model (20%)

3. Linear Regression Model tuning, approve

4. 업종을 추천하는 Logistic Regression Model (20%)

5. Logistic Regression Model tuning, approve