

정 규 세 셴   1 주 차

ToBig's   11기 유기윤

# 1주차 EDA 및 전처리 : 과제 2

## Unit 04 | 과제 2 (2 시간~)

- 아파트가 데이터셋을 이용해서 다음을 해보세요:

## 전처리

1. 결측치가 있는 변수를 찾아시고, 그 중 하나를 선택해서 결측치를 창의롭게 처리해보세요. 단, 단순히 변수를 지우는 것은 금지입니다.
2. 본 데이터셋은 범주형 변수가 많은 데이터입니다. 범주형 변수 7개를 찾으신 뒤, 알맞은 방법으로 인코딩 해주세요. Nominal (명목형) 과 ordinal (순서형)을 구분한 뒤 알맞게 인코딩 해주세요.

힌트 : `pd.DataFrame.replace()` 를 사용하시면 편리합니다.

## EDA (모든 시각화 결과에 대한 1-2줄 설명 필수입니다)

3. 연속형 변수들 간 상관관계를 나타내는 correlation matrix 를 만드신뒤, heat map 으로 시각해보세요.
4. 연속형 변수 5 가지를 선정해 타겟 변수와 관계를 나타내는 그래프로 시각화 해주세요. (타겟 변수는 Hammer Price 입니다)
5. 3차원 정보 (변수 3 가지) 를 담고있는 도표를 3개 그려주세요. (타겟 변수가 포함 안되도 괜찮습니다.)

## Unit 04 | 과제 2 (2 시간~)

**\*\* 주의 사항 \*\***

## 전처리

- 결측치 처리 방법, 인코딩 처리 방법을 자세하게 주석을 통해 설명해주세요. 여러분의 생각을 알고싶어요!

## EDA

- 시각화 패키지를 활용한 커널 / 튜토리얼은 많습니다. 구글링을 통해 원하는 코드를 찾는 것도 필요한 과정입니다. 다만, 참고하셨다면 꼭 출처를 밝혀주세요!
- 모든 EDA 결과를 통해 얻은 인사이트 / 결과 설명은 필수입니다. 특별한 경향이 보이지 않았다고 괜찮으니 자유롭게 서술해주세요.

부족한 강의 들어주셔서 감사합니다. 한 주 동안 힘내주세요! ☺