**1. Housing Data**

1-1) EDA 및 데이터 전처리 (시각화 및 통계량) (5점)

1-2) Train / Valid / Test set으로 데이터 분할하고 시각화

1-3) 2차 교호작용항까지 고려한 회귀분석 수행 및 변수 선택 과정

1-4) 벌점, 앙상블을 포함해서 그 외 모형에 적합한 기계학습 모델을 3가지 생성하고 모델 최적화 과정 (MSE,MAPE, R2를 제시해야함)

**2. Corona Data**

2-1) 인구대비 코로나 확진자 비율이 가장 높은 국가 5개 제시하고 일일확진자, 누적확진자, 일일 사망자, 누적 사망자 추이를 각각 1장씩의 시계열 그래프로 시각화하기

2-2) 코로나 위험지수를 개발하고 위험지수가 높은 국가 10개를 추려내서 막대그래프로 시각화하기

2-4) 시계열 모델링 및 비선형 모델링을 진행하라.

**3. Survey Data**

3-1) 항목별, 그룹별 만족도 응답의 평균, 표준편차, 왜도, 첨도를 구하라. (10점)

3-2) 응답항목별 차이가 있는지 분석? (10점)

3-3) 탐색적 요인분석을 수행해라 (15점)

3-4) 신뢰성 지수를 개발 하는 문제 항목별 신뢰성 지수 = ( n\*r\_mean)/(1+(n-1)\*r\_mean) (15점)