

## Part 2. R 통계분석 (데이터 분석 전문가 양성과정)

### X.1

# 미니 프로젝트

경북대학교 배준현 교수  
(joonion@knu.ac.kr)



## X.1. 미니 프로젝트

### ■ 미니 프로젝트:

- 프로젝트 주제: “R을 활용한 데이터 통계분석”
- 프로젝트 절차:
  - 다음 페이지를 보고 통계 분석 기법을 각자 선정할 것
  - 해당하는 통계분석 기법을 잘 적용할 수 있는 데이터셋을 선정할 것
    - R 내장 데이터셋은 분석대상으로 선정할 수 없음
    - 외부 패키지의 데이터셋이라 하더라도 CSV 파일로 저장하여 로드할 것
  - 선정한 데이터셋에 대한 통계분석을 실시할 것
    - 선정한 데이터셋에 대한 EDA를 먼저 실행할 것
    - 선정한 통계분석 기법을 해당 데이터셋에 적용해 볼 것
    - 기타 해당 데이터셋과 통계분석 기법에 대해서 다양한 연구를 해 볼 것.



## X.1. 미니 프로젝트

### ■ 통계분석 기법:

#### • 집단간 차이검정:

- 평균검정: 독립표본 평균검정, 대응표본 평균 검정 등
- 교차분석: 카이제곱검정, 적합도 검정, 독립성 검정, 신뢰성 검정 등
- 분산분석(ANOVA): 일원, 이원, 반복측정, 공분산 등

#### • 선형 회귀분석:

- 선형회귀: 단순, 다항, 다중 선형회귀
- 선형회귀 모델의 해석, 평가, 예측 등

#### • 일반 회귀분석:

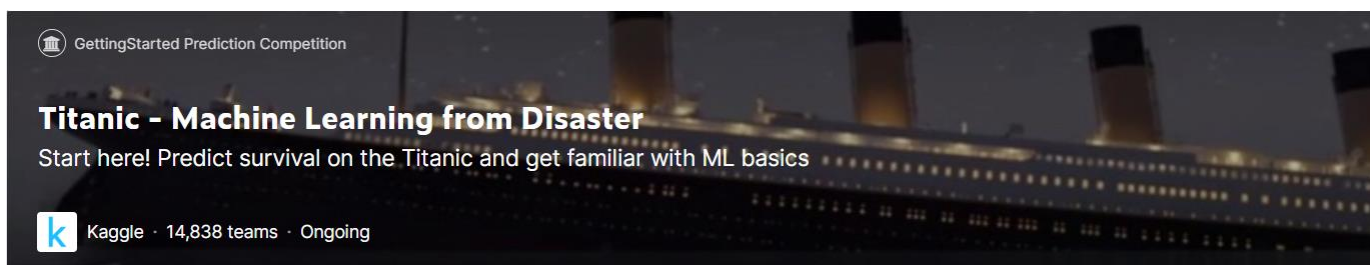
- 로지스틱 회귀와 분류 모델 적용



# X.1. 미니 프로젝트

## ■ 대상 데이터셋: 참고용 (가급적 다른 데이터셋 선정을 권장함)

<https://www.kaggle.com/competitions/titanic>



<https://www.kaggle.com/competitions/bike-sharing-demand/data>

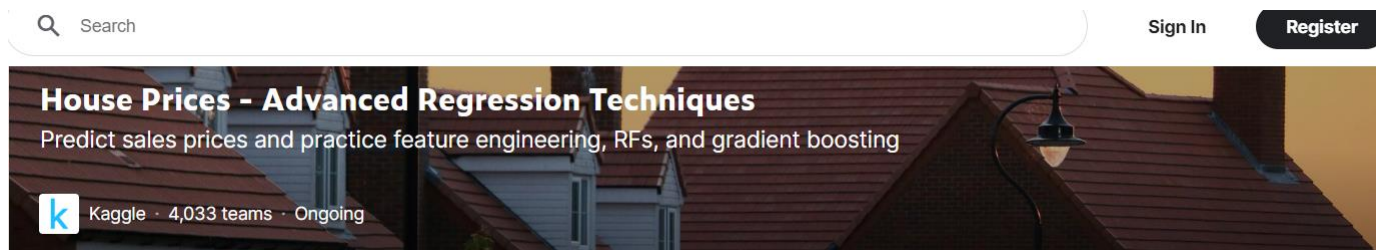


### • 선형 회귀 모델



# X.1. 미니 프로젝트

<https://www.kaggle.com/competitions/house-prices-advanced-regression-techniques/data>



<https://www.kaggle.com/datasets/uciml/pima-indians-diabetes-database>

## Pima Indians Diabetes Database

Predict the onset of diabetes based on diagnostic measures



- 이항 로지스틱 회귀



## X.1. 미니 프로젝트

- 프로젝트 수행 방법 및 제출결과:
  - 팀원은 상기한 통계 분석을 각각 하나씩 맡아서 분석할 것
    - 분석 결과물은 R Markdown으로 작성하여 github으로 제출할 것
  - 팀장은 미니 프로젝트를 위한 github 리포지토리를 만들 것
    - 팀원의 개별 분석 결과물을 github에서 관리할 것
    - 전체 프로젝트 소개를 하는 발표자료를 만들 것 (팀원별 역할 소개 포함)
  - 최종 제출 결과물:
    - 상기한 결과물이 포함된 미니프로젝트 github 리포지터리 주소



## X.1. 미니 프로젝트

### ■ 평가 항목:

#### • 팀원:

- 담당한 통계분석 기법에 적절한 데이터셋을 선정했는가?
- 선정한 데이터셋의 특성을 적절하게 파악하고 시각화했는가?
- 선정한 데이터셋에 통계분석 기법을 제대로 적용했는가?
- 통계분석 결과를 R 마크다운으로 충실하게 작성했는가?
- 통계분석 결과를 간결하게, 명확하게, 자신있게 발표했는가?

#### • 팀장:

- 전체 프로젝트의 내용이 github 리포지터리에 잘 정리되었는가?
- 전체 프로젝트의 내용이 수준 차이가 크게 나지 않도록 잘 조율되었는가?
- 전체 프로젝트에 대한 정리 내용을 간결하게, 명확하게, 자신있게 발표했는가?

*Any Questions?*

