

Part 1. R 프로그래밍 (데이터 분석 전문가 양성과정)

X.1

미니 프로젝트

경북대학교 배준현 교수
(joonion@knu.ac.kr)



X.1. 미니 프로젝트

■ 미니 프로젝트:

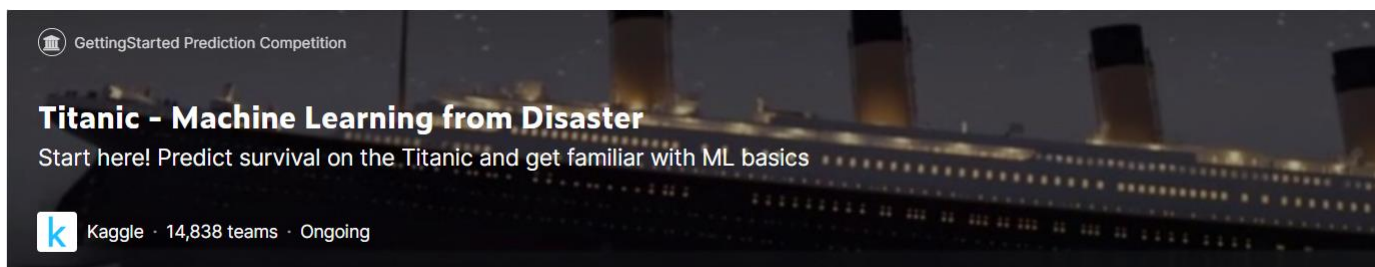
- 프로젝트 주제: “R을 활용한 데이터 탐색과 시각화”
- 프로젝트 절차:
 - 다음 페이지를 보고 대상 데이터를 선정할 것
 - 변수의 유형과 특성 요약
 - 수치형 변수 데이터 탐색과 시각화
 - 범주형 변수 데이터 탐색과 시각화
 - 데이터 전처리와 다루기:
 - 결측치 처리, 이상치 확인, 선택, 분리, 결합, 집계, 정렬, 샘플링 등
 - R로 할 수 있는 다양한 데이터 탐색과 시각화를 시도해 볼 것.
 - 기타 R 함수를 이용한 다양한 데이터 탐색과 시각화를 시도해 볼 것.



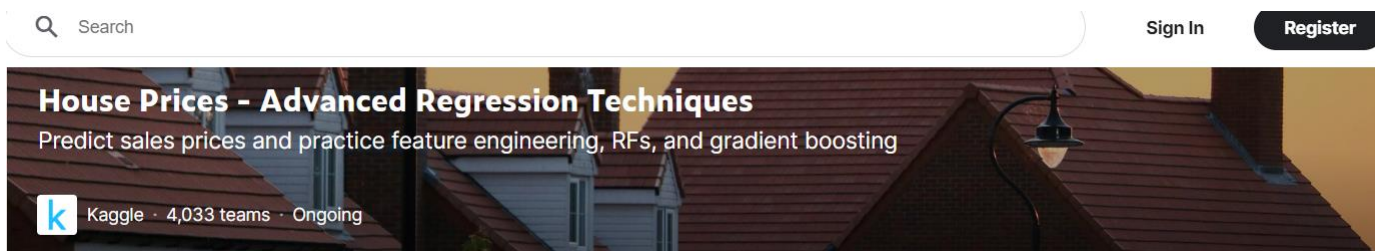
X.1. 미니 프로젝트

■ 대상 데이터셋:

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



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





X.1. 미니 프로젝트

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



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

Pima Indians Diabetes Database

Predict the onset of diabetes based on diagnostic measures



- 팀원의 수가 많거나, 특별히 관심있는 다른 데이터셋이 있으면 해당 데이터셋 선정 가능함.
- 단, 반드시 kaggle.com에 등재되어 있는 데이터셋이어야 함.



X.1. 미니 프로젝트

- 프로젝트 수행 방법 및 제출결과:
 - 팀원은 상기한 데이터셋 각각 하나씩을 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 데이터셋에 대한 충분한 사전조사가 이뤄졌는가?
- 각 개별 변수에 대한 데이터 탐색을 충실하게 실행했는가?
- 각 개별 변수에 대한 데이터 시각화를 충실하게 실행했는가?
- 분석 결과를 R 마크다운으로 충실하게 작성했는가?
- 분석 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?

