퍼블릭 클라우드 개발환경

캐글 (Kaggle)

캐글이란: 데이터사이언스 경진대회 플랫폼

- 캐글 (Kaggle)
 - 가장 유명한 데이터 과학 경진대회 플랫폼
 - 2010년 예측모델 및 분석을 위한 플랫폼 서비스로 출발하여 **2017년 구글에 인수**
 - 2019년 기준 13,000여개의 데이터를 공개
 - 의료, 경제, 자연과학, 공학 등 거의 모든 분야의 데이터를 다루며 무려 190개 이상의 국가로부터 100만명 이상의 회원이 가입하여 활동 중
 - 주어진 과제에 예측모델을 만들고 학습 결과를 업로드 하면 정확도가 평가됨
 - 이를 기반으로 포인트를 획득하여 레벨을 업그레이드 할 수 있음
 - 레벨에 따라 데이터 과학자로 취업할 수 있는 기회가 주어지기도 함
 - 챌린지에서 입상을 하게 되면 다양한 범주의 상금 획득 가능
- 데이콘 (Daicon) / Al.Factory
 - 국내 최대의 데이터 사이언스 경진대회 플랫폼 (한국형 캐글)

캐글이란: 데이터사이언스 경진대회 플랫폼

- 기업에서 본인들의 문제를 공개적으로 **해결**하고 싶었다.
- 기업에서 훌륭한 데이터사이언스를 **채용**하고 싶었다.
- **정부기관/단체**에서 데이터사이언스를 **양성**하고 싶었다.
- 개인은 데이터사이언스로 성장하고 싶었다.

기업, 정부기관, 단체, 연구소, 개인

Dataset With Prize



Dataset & Prize 개발 환경(kernel) 커뮤니티(follow, discussion)

전 세계 데이터 사이언티스트

- 목표
 - 개인의 실력 향상을 위한 툴로 사용하는 것이 가장 좋음
- 캐글 내 활동 가능 분야
 - Competition: 대회 순위에 따른 메달
 - Notebook: 좋은 설명, 좋은 코드에 따른 메달
 - Dataset: 좋은 데이터 셋
 - Discussion: 댓글 및 좋은 토론
- 캐글 내 등급 (Kaggle Performance Tier)









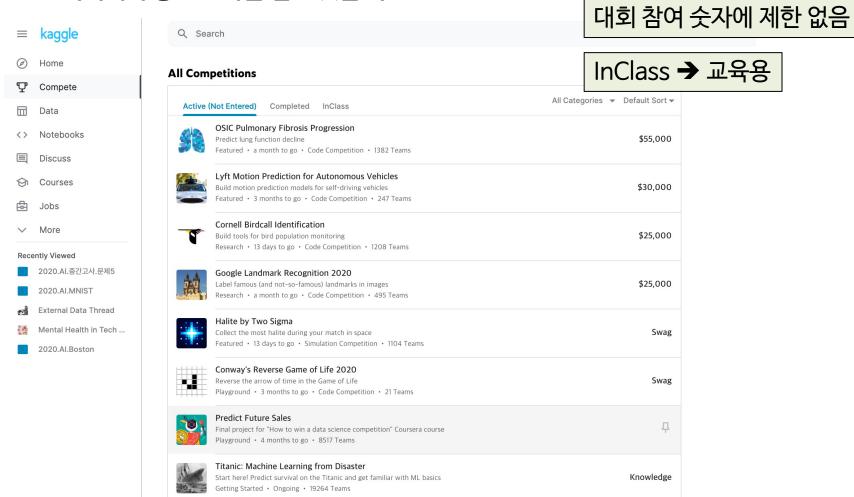


초록색(Novice) 다음은 하늘색(Contributor) 다음은 보라색(Expert) 다음은 주황색(Master) 다음은 금색(Grandmaster)

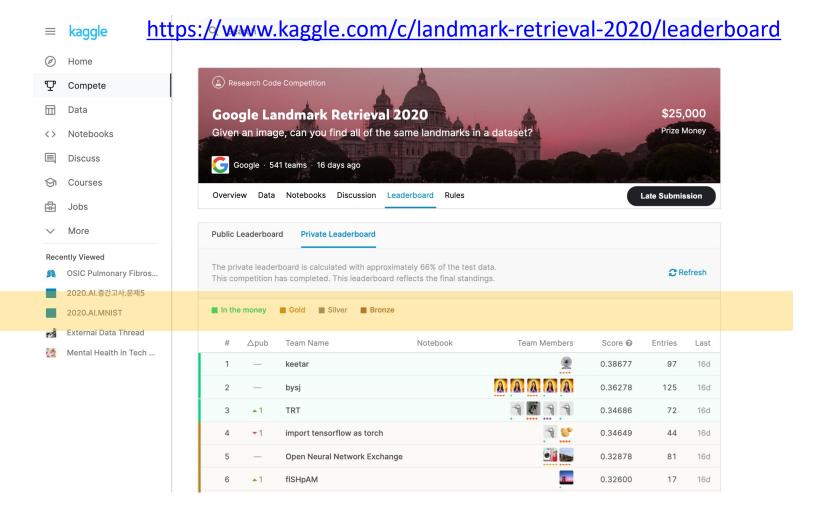
기업 인턴십 조건

Competition

■ 대회에서 좋은 결과를 얻는 것을 목표로 함

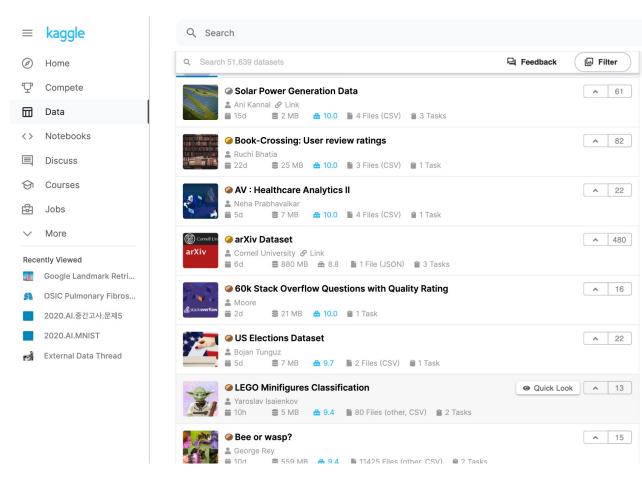


- Competition
 - 대회에서 좋은 결과를 얻는 것을 목표로 함 → 메달 획득

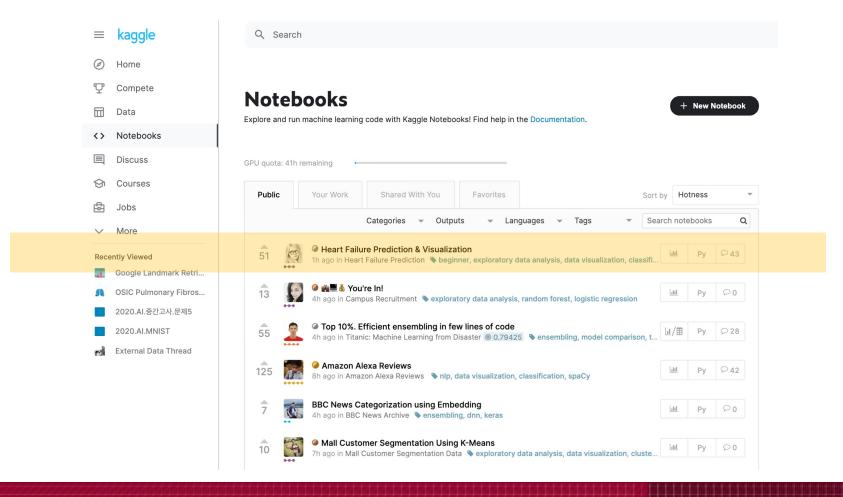


Dataset

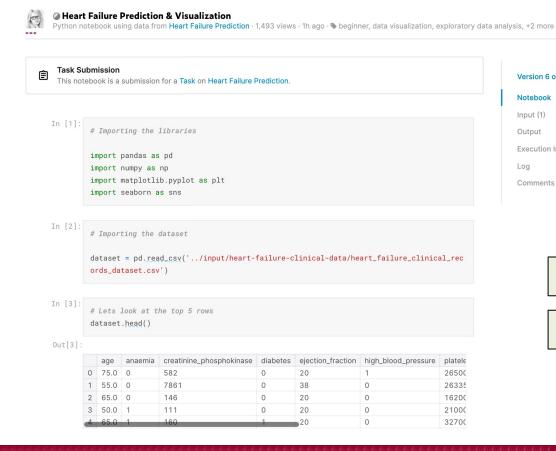
- 개인/단체/회사의 데이터 셋 공유, 가치 있는 데이터 셋 공개 및 가공
- 데이터 셋을 통한 커뮤니티 기여



- Notebook
 - 커뮤니티 내 소통의 창구, 설명과 시각화에 노력
 - Jupyter Notebook의 캐글 판



- Notebook
 - 커뮤니티 내 소통의 창구
 - Jupyter Notebook의 캐글 판 https://www.kaggle.com/sanchitakarmakar/heart-failure-prediction-visualization



데이터 분석 및 시각화

Version 6 of 6

Notebook Input (1)

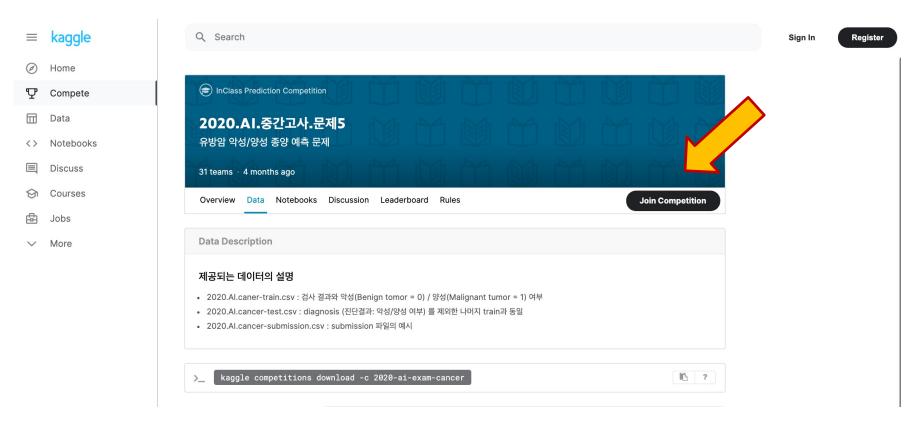
Output **Execution Info**

Log

Comments (43)

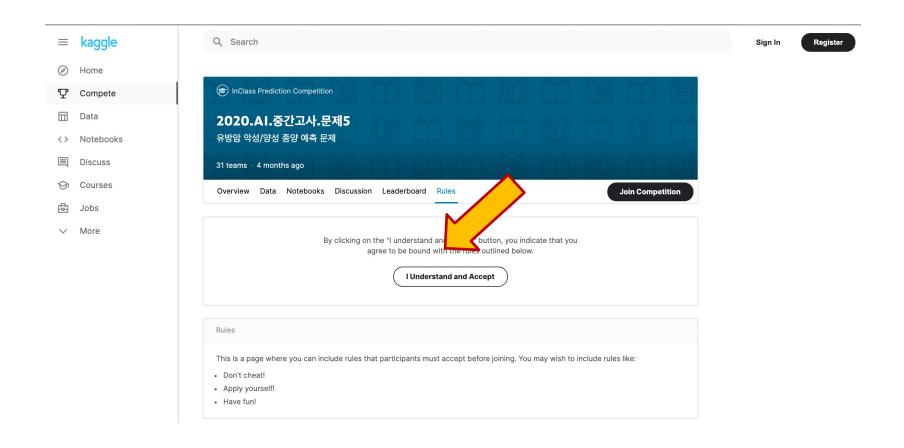
참가 방법과 리더보드 사용 방법

- 캐글 사용법 예시: 인공지능 중간고사 문제
 - https://www.kaggle.com/c/2020-ai-exam-cancer/
- 참가를 위해 [Join Competition] 을 클릭



참가 방법과 리더보드 사용 방법

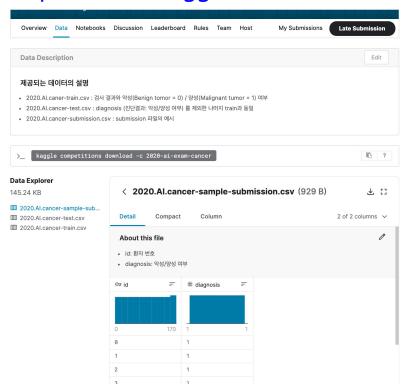
■ 참가를 위해 [I Understand and Accept] 클릭



Data 사용법

- Data 탭에는 문제 해결을 위한 <u>학습/테스트</u> 데이터 그리고 <u>정답 제출 템플릿</u> 파일이 있음
- Description에는 제공되는 제공된 <u>데이터의 설명</u>이 있음
- 데이터 분석 후 정답 제출 템플릿에 정답을 적어 파일을 리더보드에 제출

https://www.kaggle.com/c/2020-ai-exam-cancer/data (둘러보기)



Notebook 사용법

- 일반 Jupyter Notebook과 동일
- Notebook을 통해 리더 보드에 결과 파일 제출 연동
- 학습용 GPU 사용 가능, 주 40시간 (추사 사용시 구글 클라우드 결재 필요)

https://www.kaggle.com/hyeongjun0117/kernel301998fa4c (둘러보기)

