실습2. Bike sharing demand

문제 정의

- 도시 전역의 kiosk를 통한 자전거 공유 수요를 예측하는 문제
- 회귀 문제에 해당
- 이미 경진대회는 종료됨(late submission은 가능)

평가방법

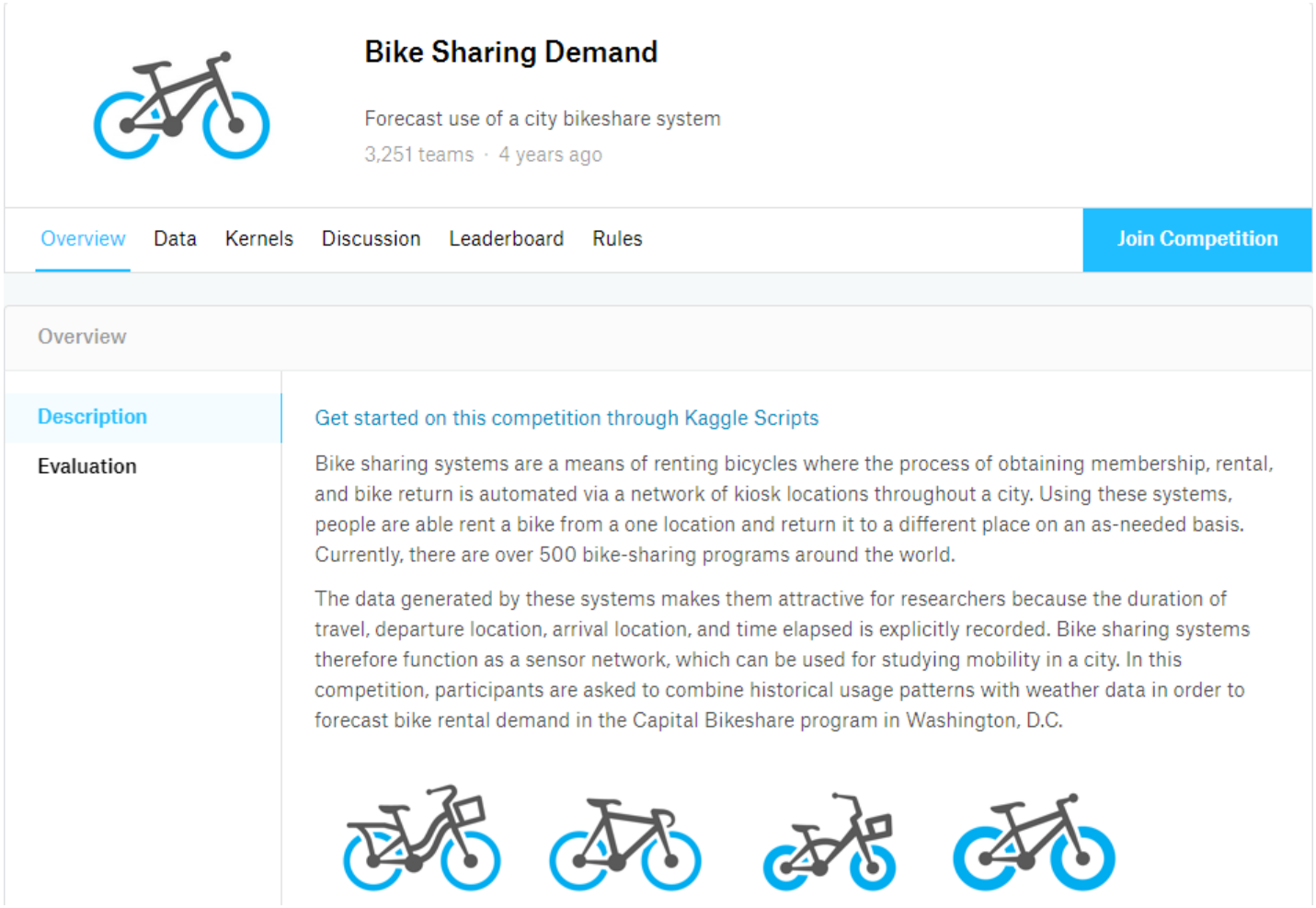
- RMSLE(Root Mean Squred Logarithmic Error) 사용

$$\text{RMSLE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\log(p_i + 1) - \log(a_i + 1))^2}$$

- n: test set의 데이터의 개수
- p_i: 예측값
- a_i: 실제값

◉ 경진대회 주소

- https://www.kaggle.com/c/bike-sharing-demand



데이터

독립변수

- datetime: 시간 정보
- season: 계절정보(봄(1), 여름(2), 가을(3), 겨울(4))
- holiday: 공휴일 여부
- workingday: working day(주말/공휴일X) 여부
- weather: 날씨
  - 1: Clear, Few clouds, Partly cloudy, Partly cloudy
  - 2: Mist + Cloudy, Mist + Broken clouds, Mist + Few clouds, Mist
  - 3: Light Snow, Light Rain + Thunderstorm + Scattered clouds, Light Rain + Scattered clouds
  - 4: Heavy Rain + Ice Pallets + Thunderstorm + Mist, Snow + Fog

- 데이터
  - 독립변수
    - temp: 온도(℃)
    - atemp: 체감온도
    - humidity: 상대습도
    - windspeed: 바람 속도
  - 종속변수
    - count: 자전거 대여량
    - casual: 미등록 사용자의 대여량
    - registered: 등록 사용자의 대여량