왜 이 데이터 셋을 골랐는지

변수 이름과 설명

Baseline 정하기

Horse Colic 데이터 셋은 missing data가 많은 데이터 셋이므로, NaN을 먼저 처리해 주어야한다. Nan의 data type이 String인 경우에는 mode를, Numeric인 경우에는 median을 이용하여 NaN을 채워주었다.