Potential Datasets:

* Sources:
  + Kaggle
  + <https://datasetsearch.research.google.com/>
  + <https://www.kaggle.com/marlesson/myanimelist-dataset-animes-profiles-reviews?select=reviews.csv>

MyAnimeList

We will look at a dataset that shows an Anime’s ratings & reviews. We will try to train the AI to look at the reviews and make a prediction of the rating based on it, using Sentiment Analysis, which is a form of machine learning referring to the use of natural language processing. It uses text analysis and classification to evaluate qualitative data typically expressed as text.

MyAnimeList uses a review system that allows users to rate various aspects of an Anime series out of 10 and then give an overall score, as well as a text-based review of the Anime. By applying Sentiment Analysis, we should be able to make a prediction of the rating score of a given anime based on its written reviews.

* Predicted issue:
  + Omitting embedded scores from written review (All written reviews in this data-set include the scores for various aspects of the anime, we do not want our AI to base its prediction on these scores as this defeats the purpose of Sentiment Analysis).