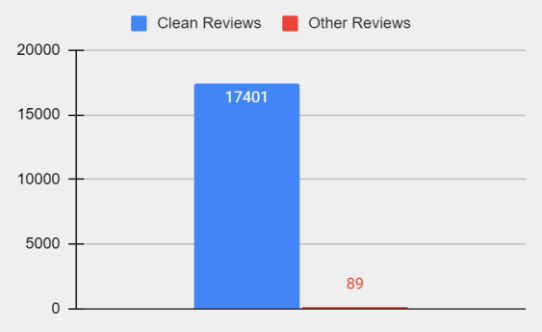
Project Methods
Parin Pandya

For this project, we have been provided with a pre-processed train and test split from Kaggle. So we already have a training and testing dataset. The training dataset and the testing dataset are almost the same, except there is one extra column in the training dataset. It is a column called user suggestion, which means whether the user would suggest or recommend this game to others or not. Using binary classification, 0 to represent no suggestion, and 1 to represent a suggestion. The testing dataset contains all other columns except for the user suggestion column. Another way the datasets differ is that they include differing games from each other.

For the analysis of this project, I will be using textual analysis techniques, because all the game reviews are of text format. There are some reviews which are special characters, or in another language other than English. There are even some empty rows in some columns. This is where the text analysis techniques are most needed. I will be using NLP techniques, like the Python pandas and NLTK libraries to help clean and filter the reviews. I will try to remove any special characters, punctuation, and even some common words. This way, the reviews will have a clear and common structure without any irrelevant characters, and each review will be shrunken down to a clear and concise form.

To evaluate my analysis, I will use a metric based on some Machine Learning techniques. Along with the Textblob library, I will be able to classify if a review is positive or negative, based on its sentiment. Each word will be measured, with a positive word being > 0, and a negative word being < 0. After this I will be able to add up the total of the words in a review, and then determine if the review is positive or negative.

1. Revised Dataset Figure:
    - Scraped 89 reviews



2. Revised Results Figure: