Pre-processing:

I started by removing any columns (variables) I find unnecessary to predicting the ESRB rating of a game. I started by removing the columns id, slug, name, released, tba, updated, website, rating, rating\_top, playtime, reviews\_count, platforms, developers, publishers, added\_status\_yet, added\_status\_owned, added\_status\_beaten, added\_status\_toplay, added\_status\_dropped, added\_status\_playing id, as they wouldn’t be needed due to them either being identifiers or not related to the ESRB rating e.g. genre is good variable to keep as the genre of a game may affect the ESRB rating, such as a horror game.