**Exploratory Data Analysis - Inferential Statistics**

The variables that are particularly significant in terms of explaining the answer to my project question: given all the information for a book on Goodreads, can we predict its average rating. are num\_reviews and num\_ratings columns. These two columns show how popular a book is to the readers, to the public. The number of reviews and number of ratings give us an idea how many people most likely have read the books and are willing to spend their valuable time to write a review for the book.

I will use bootstrap inferential statistics to resample the sample data we have from Goodreads. I will run 10,000 res-sampling to find the correlation between num\_review and num\_ratings columns with the book’s average number of rating column. In addition to these two columns, I will also use its author’s rating count as well as its author’s average rating columns to help me with the statistics modeling.