



Analysis of a Publisher's Book Quality

An evaluation of the average book ratings between Penguin Random House, Simon & Schuster, and HarperCollins



Overview and problem statement

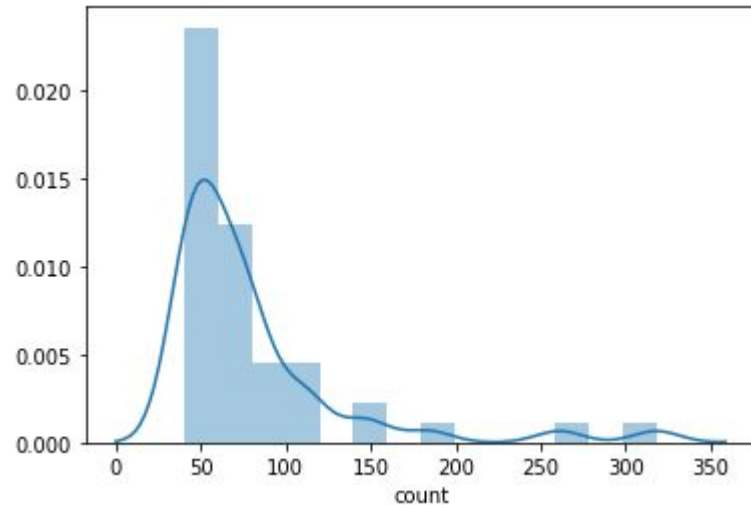
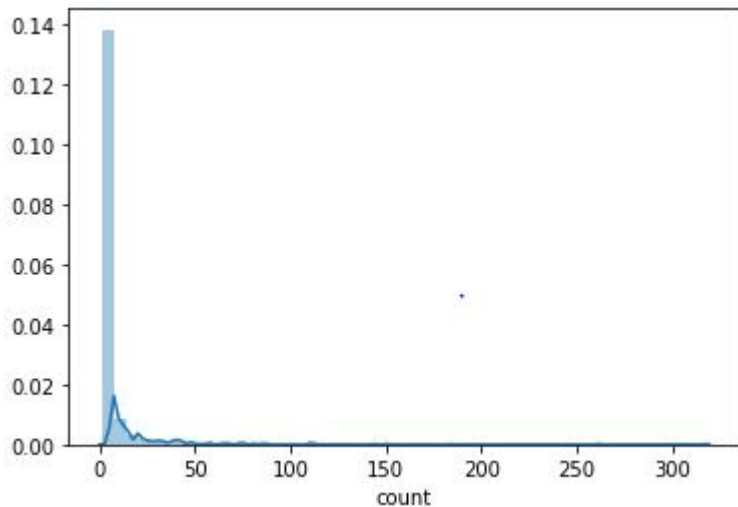
- There are thousands of books and a handful of large publishing companies
- Does the quality of books vary between these publishers?
 - If so, which publishing companies deviate from the others?
 - Which publishers have a higher average book rating?
 - Should a writer consider these differences when looking for a publisher?



The Data

- 10,352 unique book titles from Goodreads, a popular book recommendation site. Majority from Penguin Random House
- Data includes information on the book, such as
 - Author
 - Number of pages
 - Average rating on Goodreads
 - Publisher
- Link to the raw data on Kaggle: [link](#)

The Data - Cleaning Up

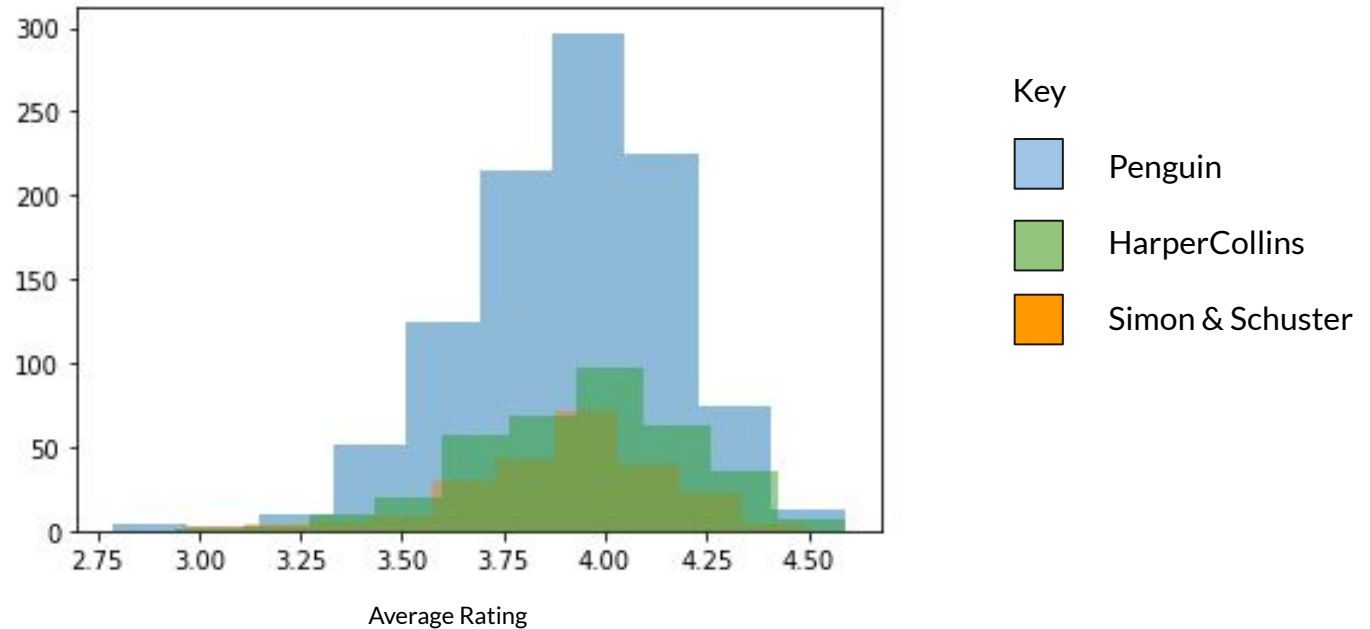


Impact of Publishers:

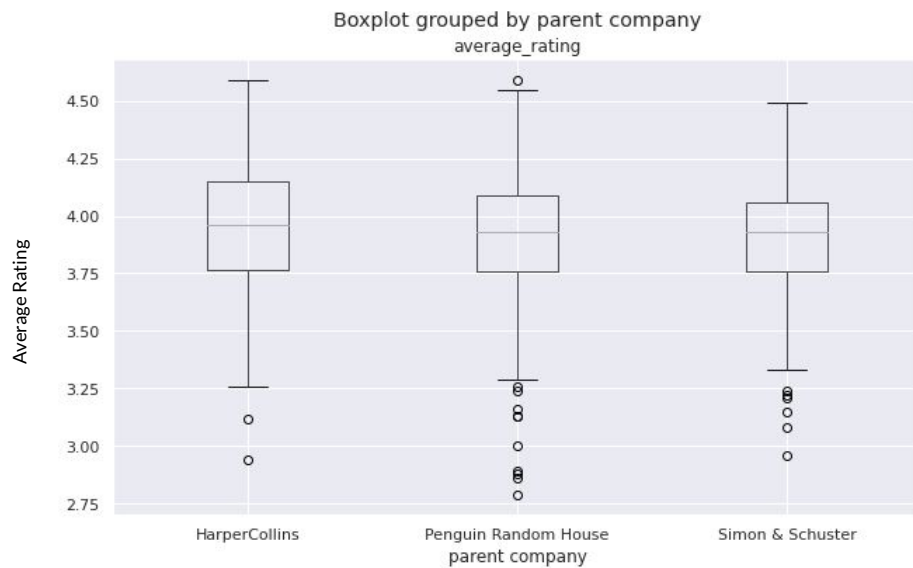
Is there a difference in book quality between different publishers?



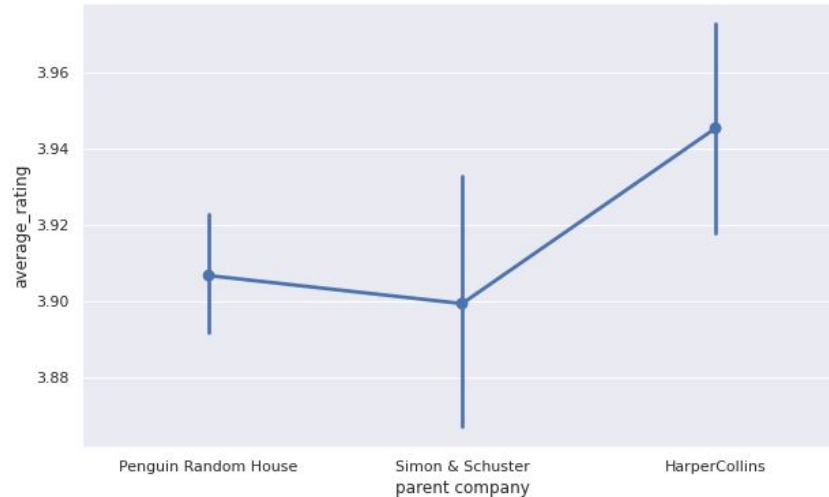
Distribution Visualization - Histogram



Distribution Visualization - Box Plot



Statistically significant difference between HarperCollins and the others





Test Results

Group 1	Group 2	p-value	Significant Difference
HarperCollins	Penguin Random House	0.0366	Yes
HarperCollins	Simon & Schuster	0.0826	No
Penguin Random House	Simon & Schuster	0.9	No



Recommendations

How do these differences affect people?

- Readers and Consumers
 - Books by HarperCollins seem to be favorably rated compared to other publishers
 - New readers without a specific preference will find better quality with HarperCollins
- Authors and Writers
 - HarperCollins reflects well on publications

Penguin
Random
House



SIMON &
SCHUSTER



HarperCollinsPublishers



Questions?

