

Hey guys! Welcome to my second medium post!

For this post, I want to outline what I've learnt in August while in the SheCode Africa mentorship program. The journey so far has been rigorous with weekly courses, articles, blog posts and other resources to learn from and assignments to practice what I learn. Nevertheless, with the amount of progress I've made, I am grateful for the opportunity to be a part of this year's cohort!

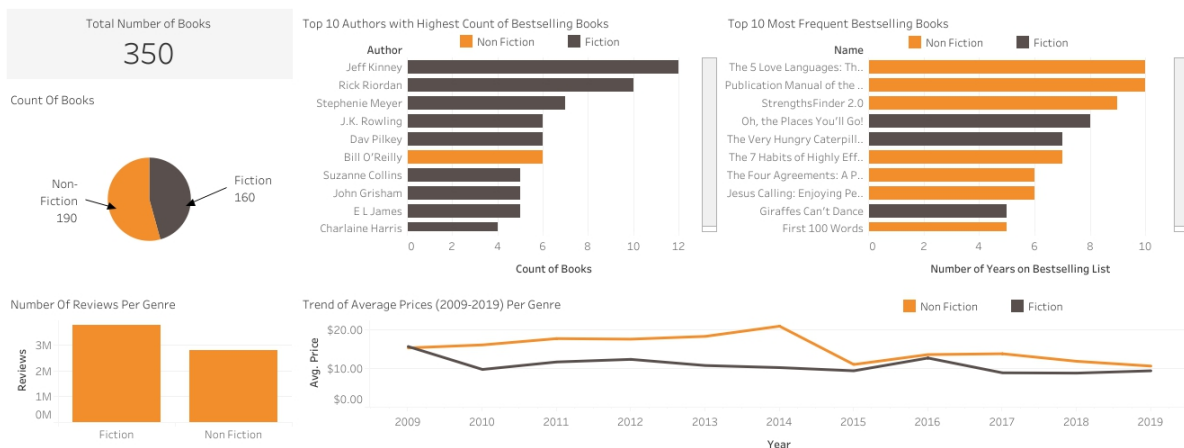
In the first week of August, which is the second month of the program, I learned some concepts in statistics, like Linear regression, probability etc. In the second week, I took a course on SQL (Structured Query Language) and completed 10 HackerRank challenges. I learnt ways to analyse data using SQL. The challenges were a good way to test my knowledge of SQL and I had fun working on them.

In the third and fourth weeks, I worked with Tableau and Power B.I., respectively. Tableau and Power B.I. are business intelligence (B.I) tools. Basically, they are used to build visualizations and dashboards, and just tell a story with data. I had been introduced to Power B.I. before starting this program, but learning it again made me understand it much better. On the other hand, Tableau was a new field for me, and I enjoyed exploring it.

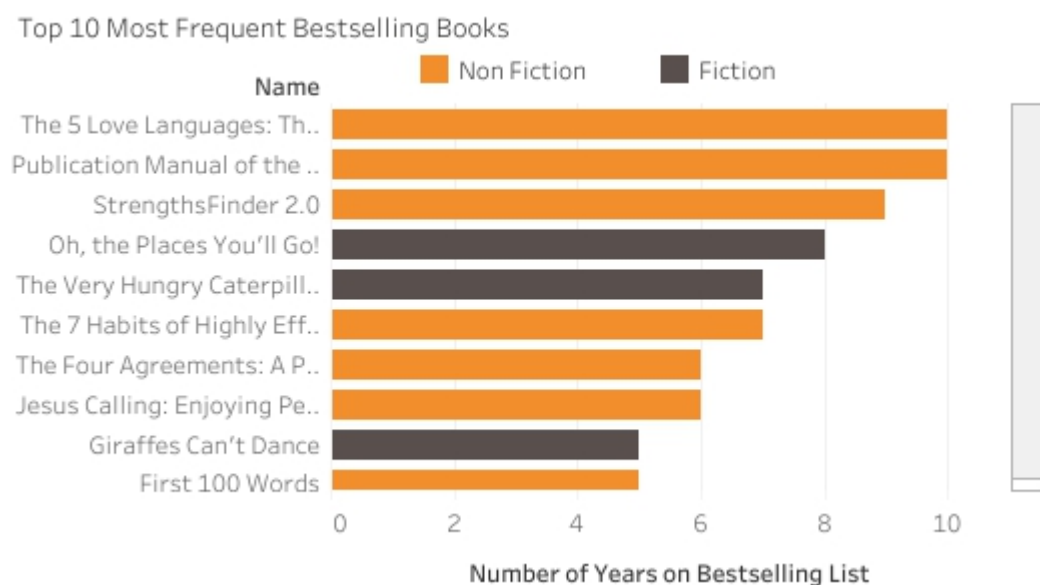
I used a dataset from Kaggle to practice. The data was on Amazon's Top 50 Bestselling books from 2009-2019. It was a clean dataset with no missing value and 550 entries (rows) and 7 features (columns). Link to dataset: [data](#)

I loaded and explored the data which involved checking for correlation between the columns and getting basic information about the data. Then I started visualization.

The first thing I wanted to know was the number of books in the dataset. Since the data was on the top 50 Bestselling books over 11 years, I expected 550 books.



However, the number of distinct books in the data was 350. This meant that some books were on the Bestselling list for multiple years. So I created another visual to show the number of years each book spent on the Amazon Bestselling list in the 11 years timeframe.



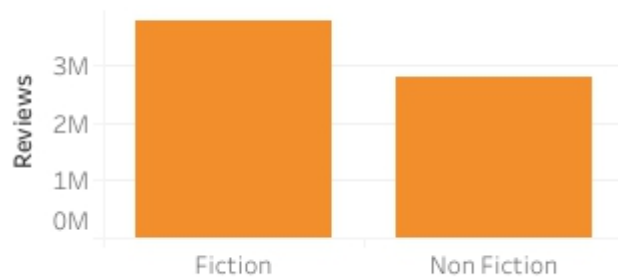
I noticed that most re-occurring Bestselling books were non-fiction. This could mean that people buy non-fiction books more than fiction books.

Another intriguing thing to note is that though non-fiction books have a higher count than fiction books, they accumulated a lower number of reviews than fiction books. This suggests that people tend to review fiction books more than non-fiction.

Count Of Books

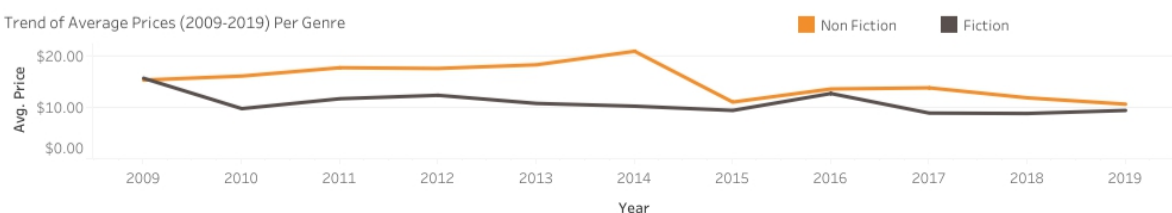


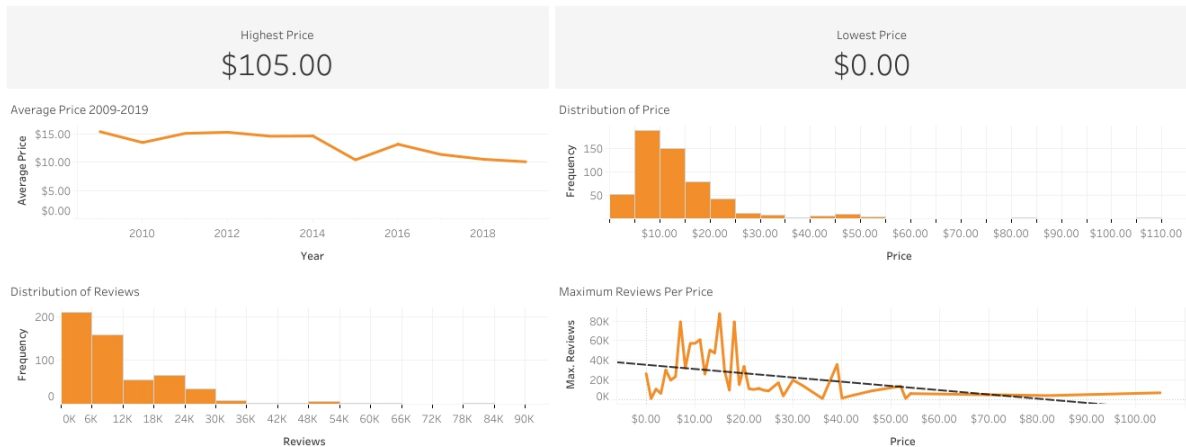
Number Of Reviews Per Genre



A reason for this could be the relative prices of the two classes of books, which might affect people's access to the books. When comparing fiction and non-fiction books, non-fiction generally costs more than fiction. However, the trend of book prices shows a decrease over the 11 years.

Trend of Average Prices (2009-2019) Per Genre

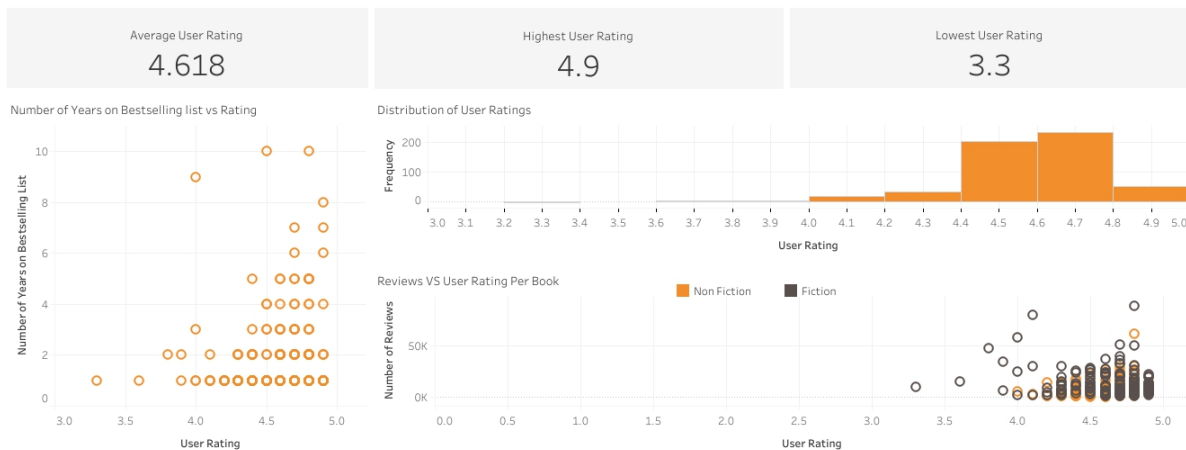




As the price increases, the number of reviews decreases.



Lastly, I saw that most bestselling books were rated between 4.6 and 4.8 with the average rating being 4.618.



Most of the books with high ratings (> 4.5) spent less than 6 years on the Bestselling list, while books rated 4.0 and 4.5 spent 9 and 10 years, respectively, on the list.

Link to project: [Amazon project](#)

What other insight can you spot in the data? Drop a comment below.