|  |  |
| --- | --- |
| the BestSelling Books    Programming Language: Phyton |  |

**📝 PROJECT SUMMARY:**

Dataset on Amazon's Top 50 bestselling books from 2009 to 2019 📖. This project contains 550 books, data has been categorized into fiction and non-fiction.

The goal of this project is to analyze data on bestselling books to identify market trends, successful authors, and other factors that influence sales.

**🎯DATASET:**

I downloaded the dataset from Kaggle.

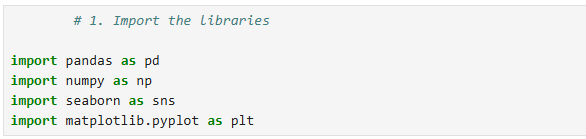
Here is the link: <https://www.kaggle.com/datasets/sootersaalu/amazon-top-50-bestselling-books-2009-2019?resource=download>

**Columns informations:**

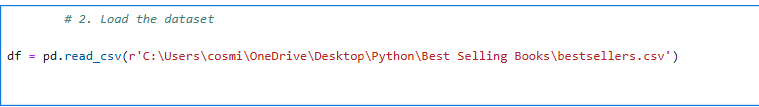
|  |
| --- |
| Name: Book name |
| Author: Book author |
| User Rating: Amazon user rating (0.0 - 5.0) |
| Reviews: Number of user reviews |
| Price: Book price |
| Year: The year(s) it ranked |
| Genre: Fiction or non-fiction |

**🔹 STEPS AND ANALYSIS:**

1. Import the libraries

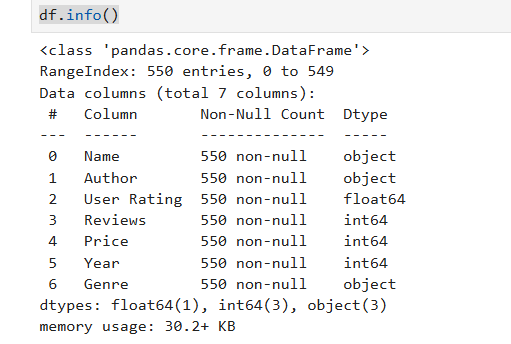


1. Load the dataset into Python program.

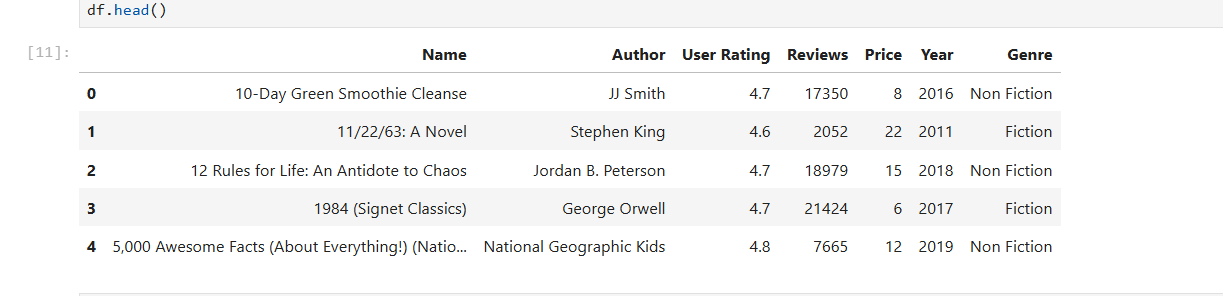


1. Explore the data 🔍

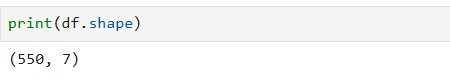
* Review the type of data



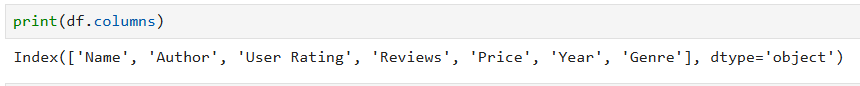
* The first 5 rows of the spreadsheet



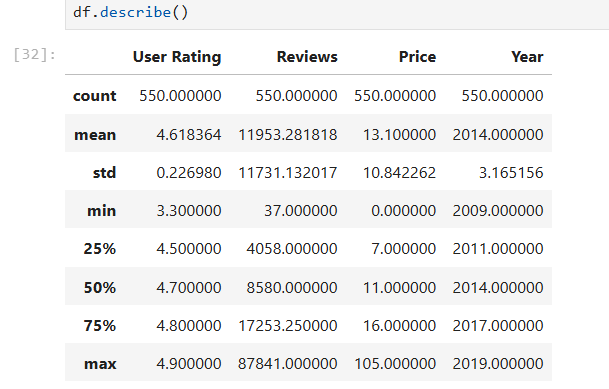
* The shape of the spreadsheet



* The column names of the spreadsheet



* Get summary statistics for each column

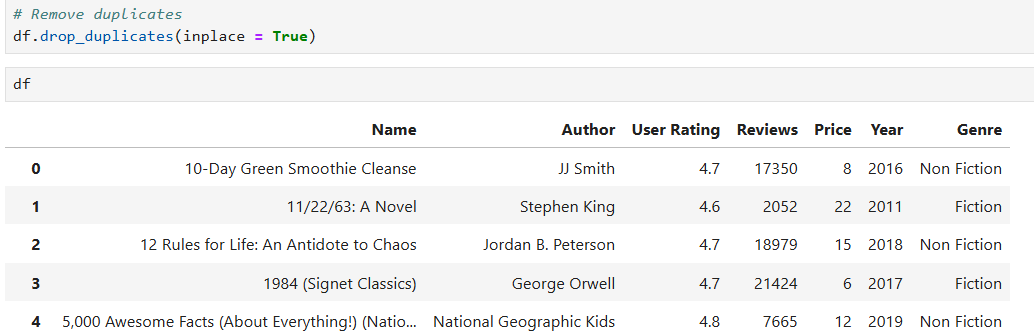


These functions will give a good overview of the data, including:

* The size of the data
* The column names
* Summary statistics for each column

1. Clean the data 🛠

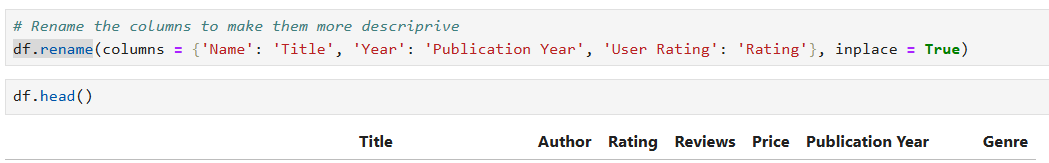
* Remove duplicates



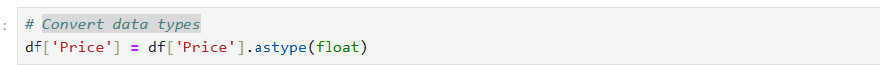
* Check if there are null values in a Dataframe



* Rename the columns to make them more descriptive

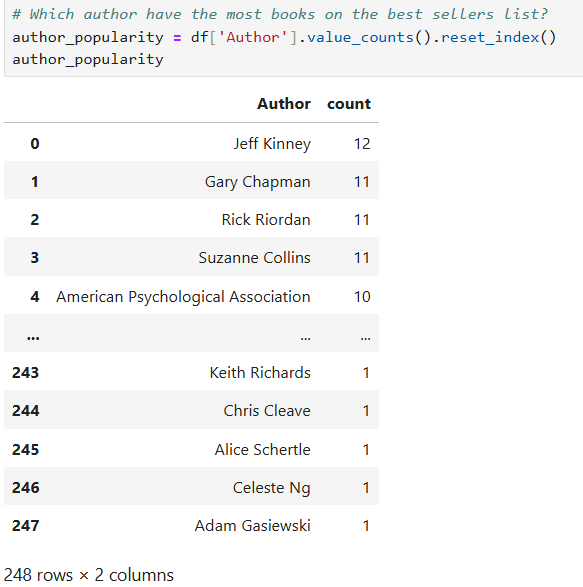


* Convert data types

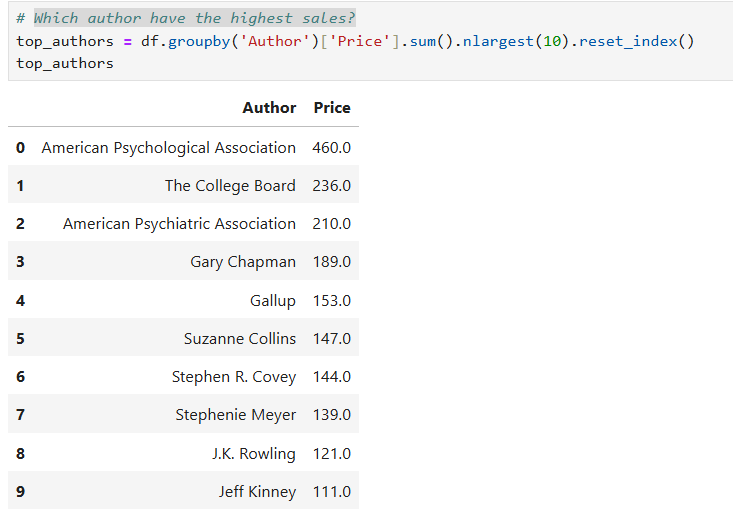


1. Exploratory data analysis **📊**

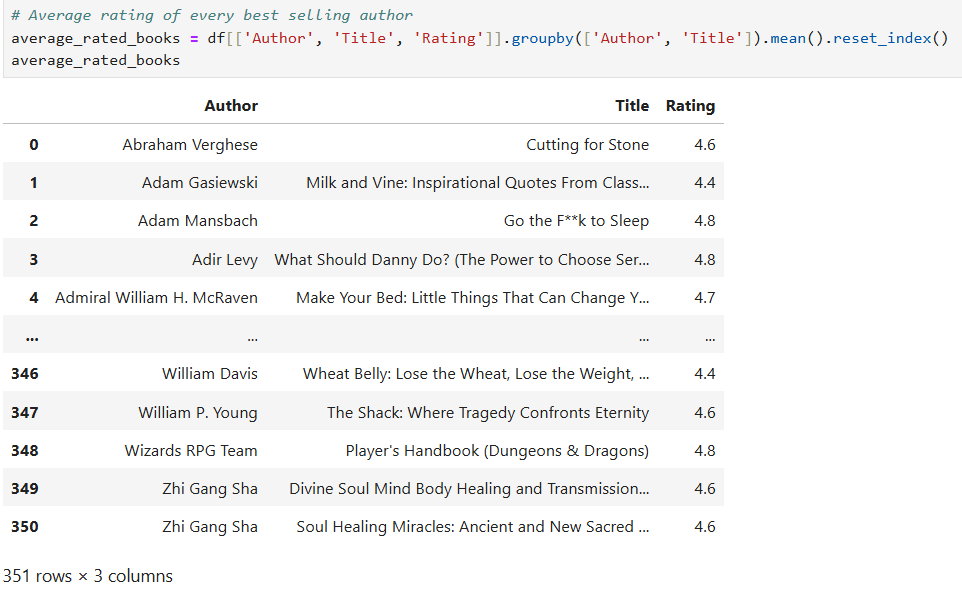
**📌** Which author have the most books on the best sellers list?



**📌** Which author have the highest sales?



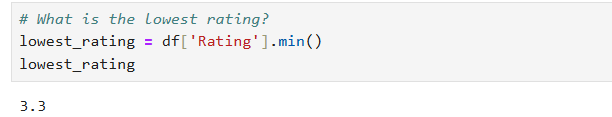
**📌** Average rating of every bestselling author



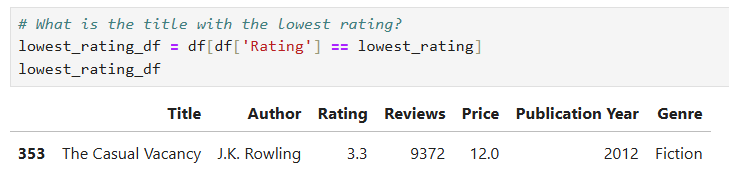
**📌** Display the top 10 rated books



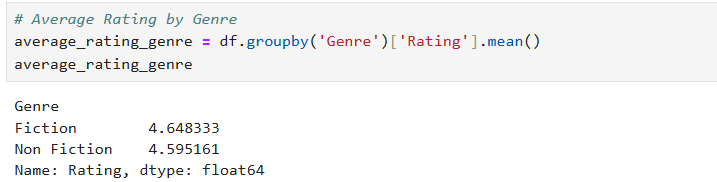
**📌** What is the lowest rating?



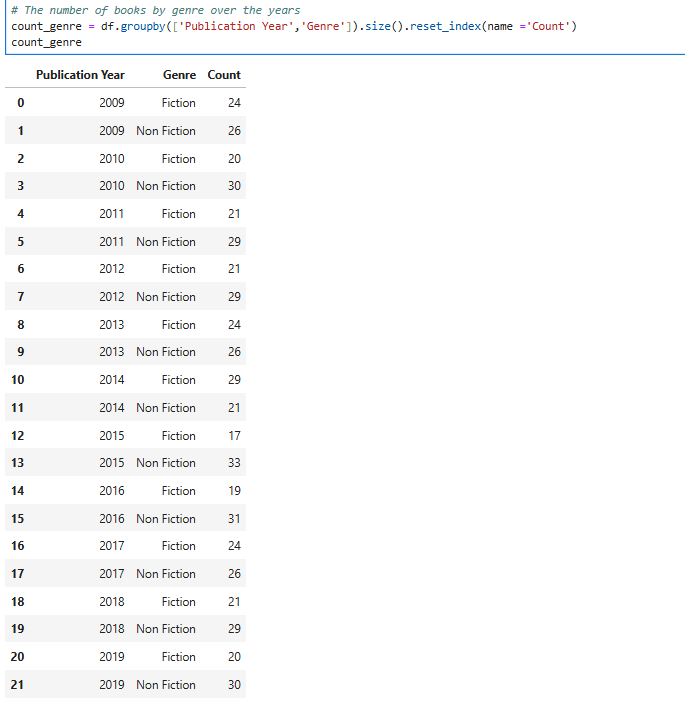
**📌** What is the title with the lowest rating?



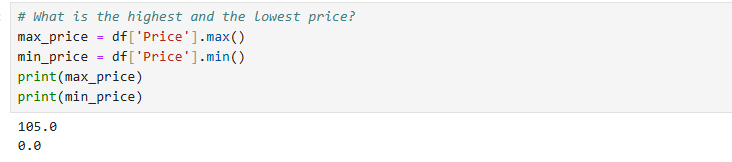
**📌** Show the average Rating by Genre



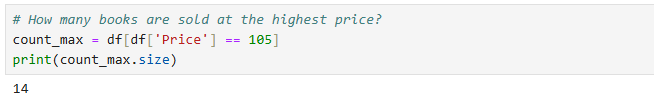
**📌** The number of books by genre over the years

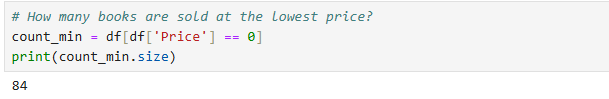


**📌** What is the highest and the lowest price?

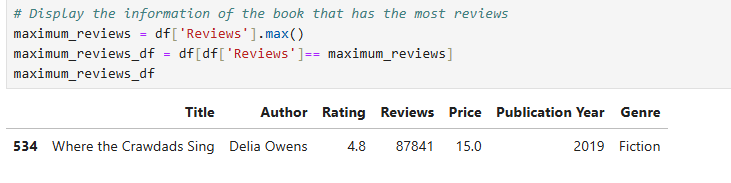


**📌** How many books are sold at the highest price?



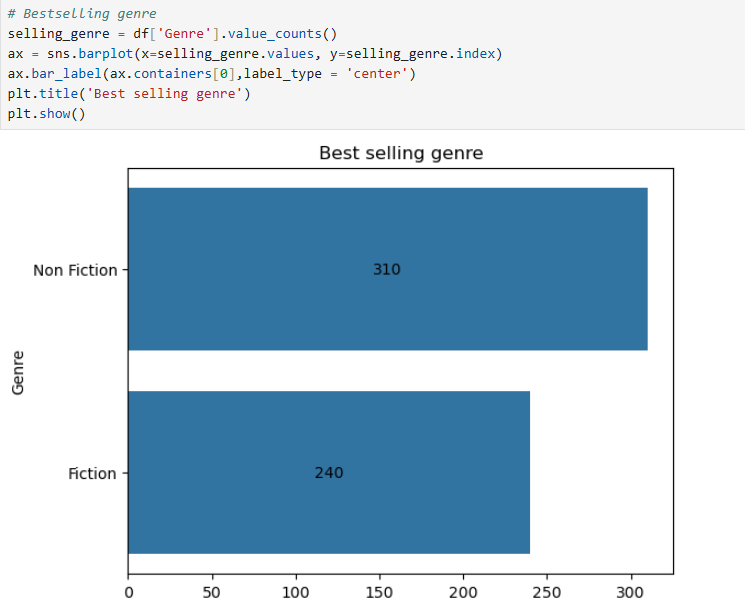
**📌** How many books are sold at the lowest price?  


**📌** Display the information of the book that has the most reviews



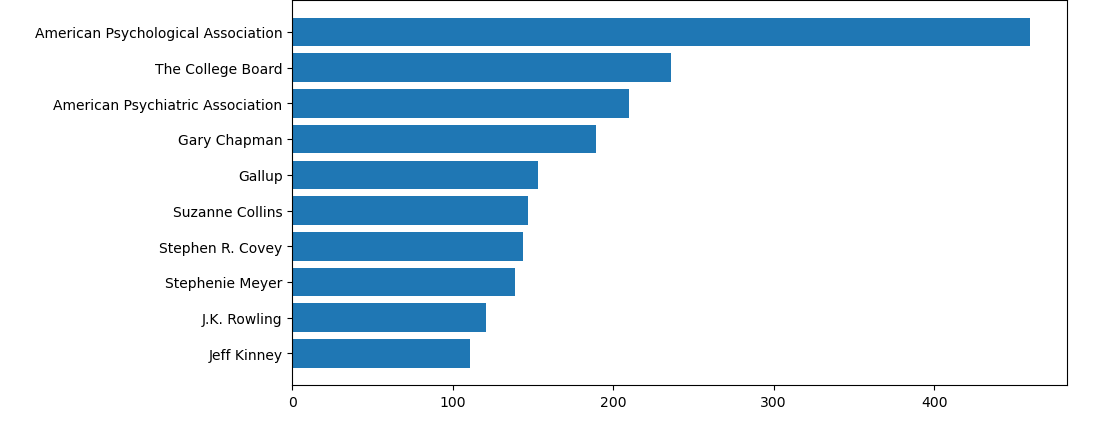
1. Data visualization **📊**

**📌** Bestselling genre



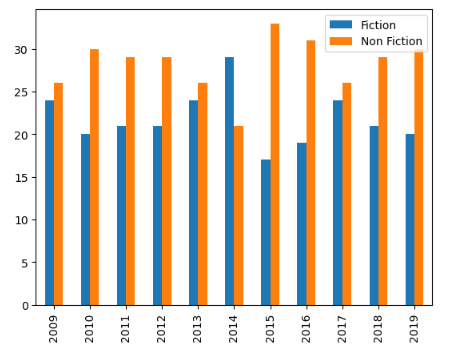
It seems that non-fiction is at the top of the list of preferences.

**📌** Top 10 authors with the highest sales

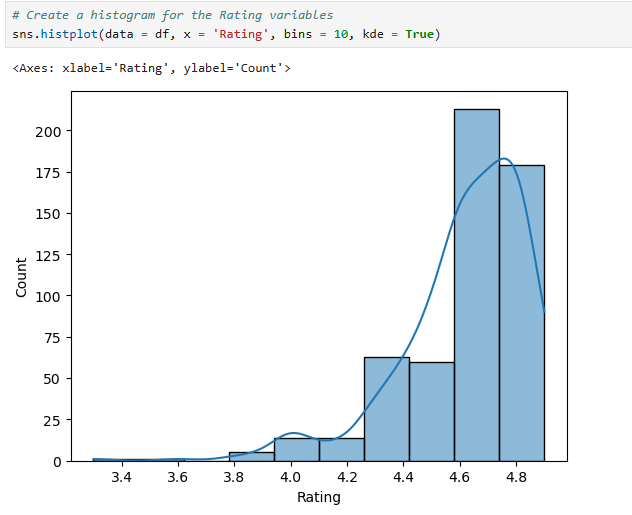


American Psychological Association is ranked 1st in the sales top with a considerable difference from the other positions.

**📌** The number of books by genre over the years



**📌** Create a histogram for the Rating variables



Rating variable has a left-asymmetry

🔍 **KEY FINDINGS**:

* There are 84 books for which price is 0. It may be because of those books being available in electronic format.
* The average rating received on a bestselling book is 4.6.
* For all the years except 2014, the number of fiction best sellers have been greater than nonfiction best sellers.
* The most books have high ratings, that may indicate a general positive tendency among users or a possible influence of recommendations.