

Best Selling Books Exploratory Data Analysis

Objective: To explore the statistics of the best-selling books of all time to see what popular books I should be reading

Data Set: <https://www.kaggle.com/datasets/drahulsingh/best-selling-books>

Summary

This was my first time using the Python libraries I have learned on my own & and in my current Computer Science course. I struggled the most trying to produce the pie chart in the last visualization. Many of the letters overlapped, including the percentages. Because of this struggle, I learned how to add a legend & increase the size of the pie chart. The best-selling author of all time is J.K Rowling. I somehow have never read a single Harry Potter book nor have I ever seen the movies. This is a sign to start reading the Harry Potter series. The book with the most sales was 'A Tale of Two Cities' by Charles Dickens. This was oddly surprising to me because I rarely hear anyone talk about this book or Charles Dickens. The oldest best-selling book was The Divine Comedy (La Divina Commedia), which was also very surprising. I was expecting the Bible to be in this dataset as the 'oldest best-selling book.' About 75% of all best-selling books are in English with French being the runner-up at just 4.9%! When trying to analyze data, I always want to know what I am working with before starting to code statistics. I like to import all the libraries I may use, just to get it out of the way. The approach I took was looking at the first 3 sets of data and the last 3 sets of data to get a feel of what I may be working with. I then check for any duplicate data & any null values. In this case, there were 56 null values in the 'Genre' section, but I decided to ignore those in this analysis (I obviously won't ignore null values when something is at stake). I then used Matplotlib and Seaborn to visualize the statistics I found. Two skills I have demonstrated are Data Visualization and Statistical Analysis. Overall, this was a fun starter project to get a good feel of my current knowledge of Python.