Shao Ru Zhang
CIS 3120- Programming for Analytics
Homework #3
25th May 2021

I chose to use the New York Times API to request 10 different query values. For the query parameters, I chose different formats of books which includes ebook fiction, ebook nonfiction, hardcover fiction, hardcover nonfiction, trade fiction paperback, paperback nonfiction, manga, paperback graphic books, series books, and young adult. I also included a code for readers to see all the list of formats available and to choose their own. Then, for each format, I requested and parsed the top best-sellers list and converted it into a dataframe for easy readability. Within each format, I chose to look at title, author, isbn13, weeks on list, ranks, ranks last week, book image height, and book image width of the book.

According to the statistical summaries below, the average time a best-seller book has been on the list is around 47 weeks, roughly a whole year. These best-sellers are an overall top seller and are not all from a specific year. It was surprising to see how some books did not make it to even a week on the list, yet it is still considered one of the top best sellers of all time. The rank of the list does not really tell us anything overall, it is just the rank of the book in each format. The rank last week is smaller than overall ranking, this makes sense since a lot of these best-sellers were not recent, so their last week rankings will deviate a lot from the overall best-sellers ranking. For the book image height, it is interesting to note that the mean of 466.5 is very close to the max of 500, which shows that there is not a large variation in average book image height compared to the maximum book image height. Book image width, however, has a smaller standard deviation meaning that the difference between each width to the average is less than the book image height's.

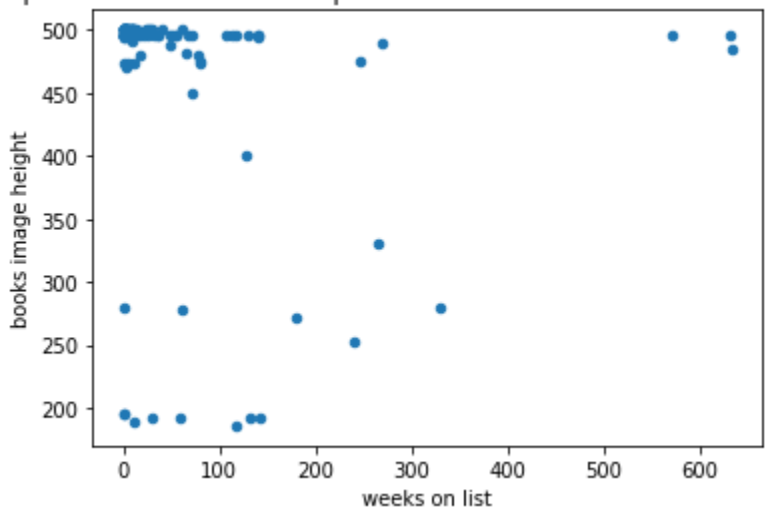
According to both the scatterplots below, it seems like most of the books are clustered around a book image height of 500 and book image width of 325, with around 0-50 weeks on the best-seller' list. There is no strong relationship, but we can generally predict the height and width of a typical best-seller book and how many weeks it typically remains on the list.

Using the pandas dataframe, statistical summary, and graphs, people can utilize this information to determine which best-seller books they may want to read during their free time. They can utilize this information to make a decision based on their ranking on their respective list, how many weeks on the ranking list, ranking last week, and the size of the book image. Authors may also use this information to keep in mind the dimensions of their upcoming books, although it does not guarantee that it will make it onto the best-sellers list.

Summary Statistics:

	weeks on list	ranks	ranks last week	books image height	books image width
count	135.000000	135.000000	135.000000	135.000000	135.000000
mean	47.392593	7.444444	3.140741	466.518519	310.785185
std	105.312265	4.190727	3.630639	82.712490	54.959467
min	0.000000	1.000000	0.000000	186.000000	128.000000
25%	1.000000	4.000000	0.000000	494.000000	321.500000
50%	6.000000	7.000000	2.000000	495.000000	329.000000
75%	48.000000	11.000000	6.000000	500.000000	330.000000
max	634.000000	15.000000	14.000000	500.000000	401.000000

Scatterplot: is there a relationship between weeks on list vs books image height?



Scatterplot: is there a relationship between weeks on list vs books image width?