



AMAZON BESTSELLING NOVELS

Data Analysis

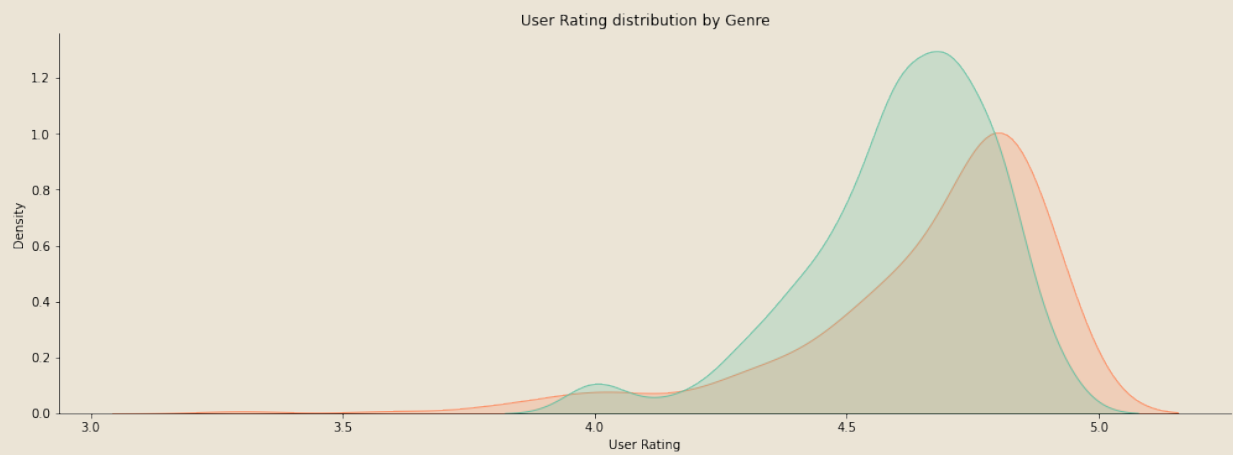
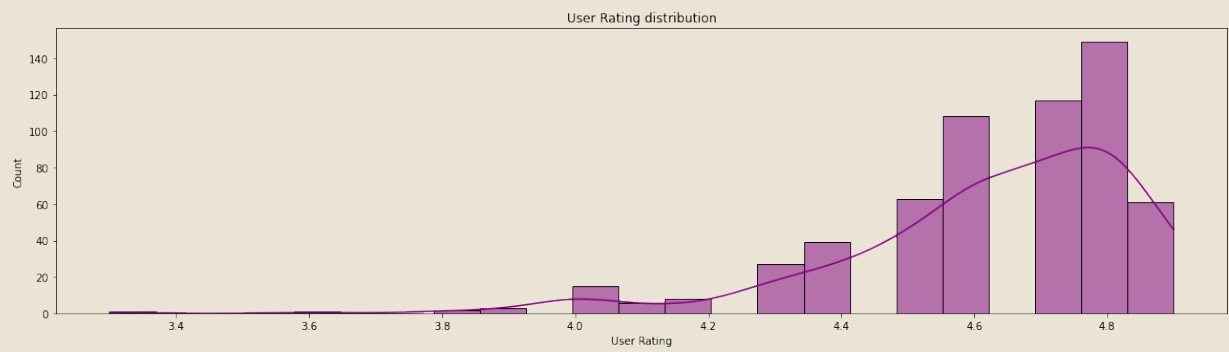
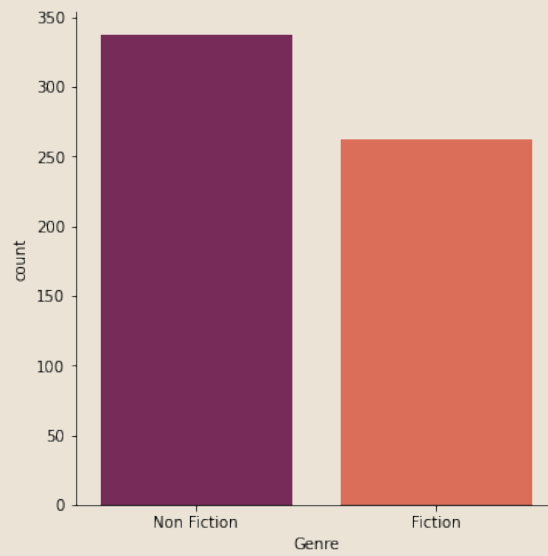
2009-2020

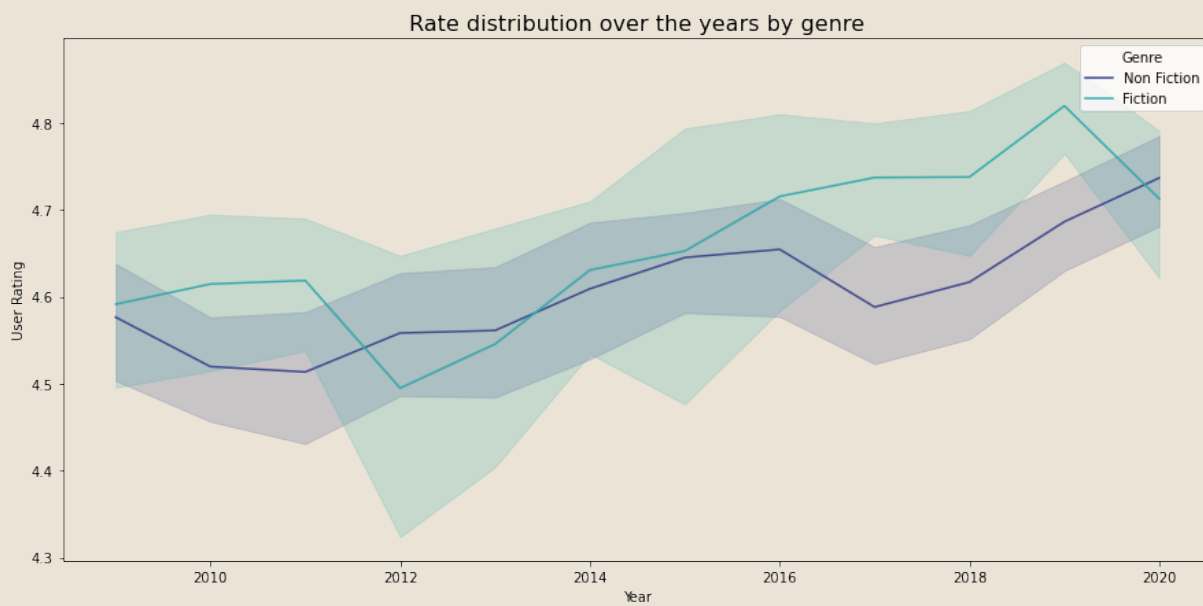
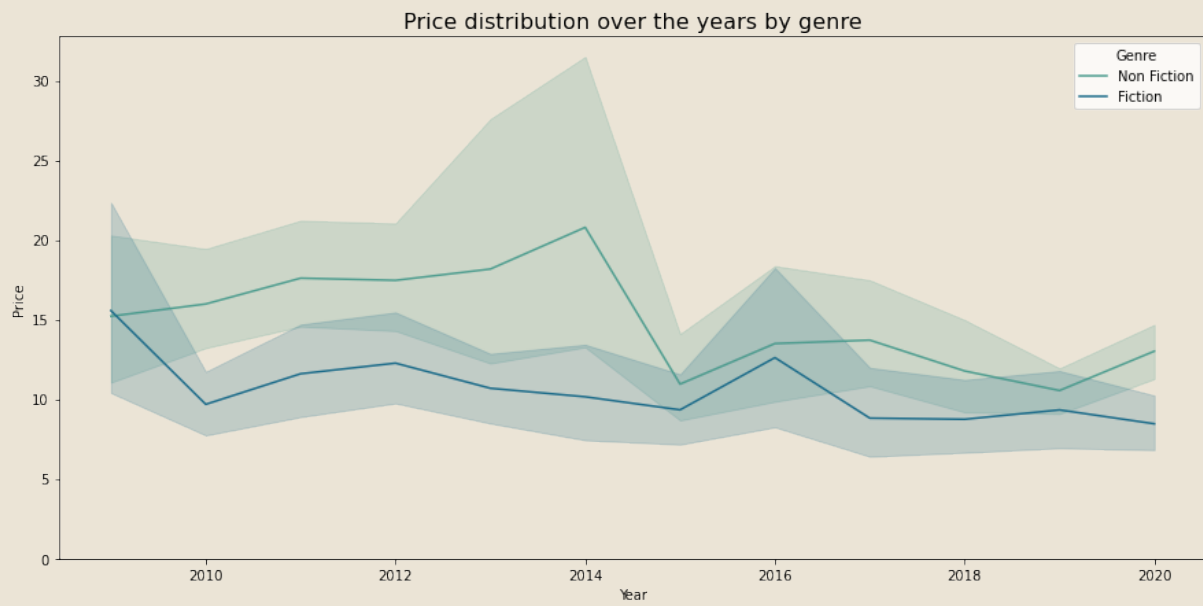
Amazon Bestselling Novels

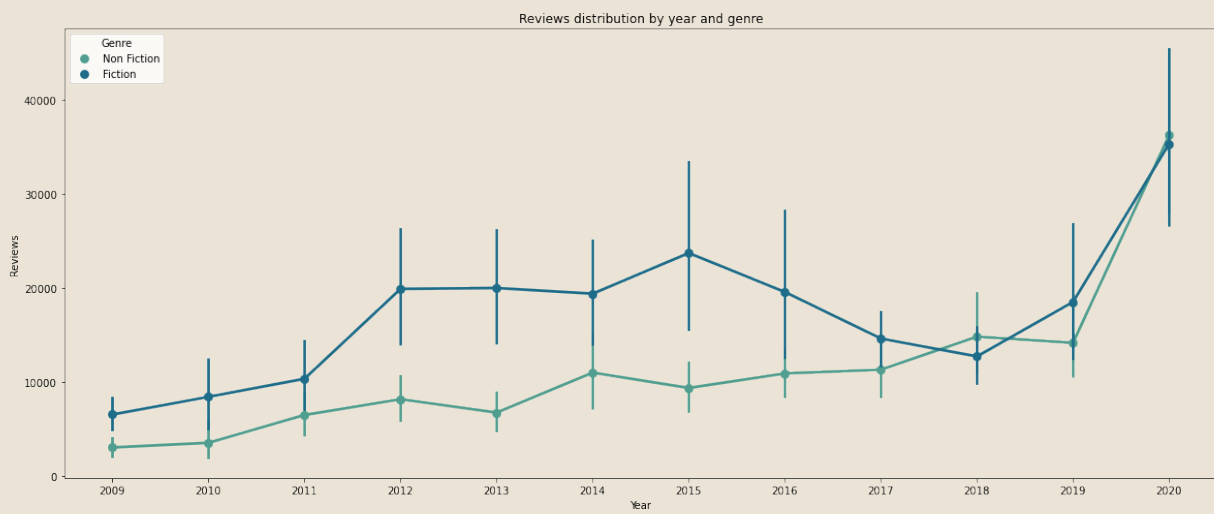
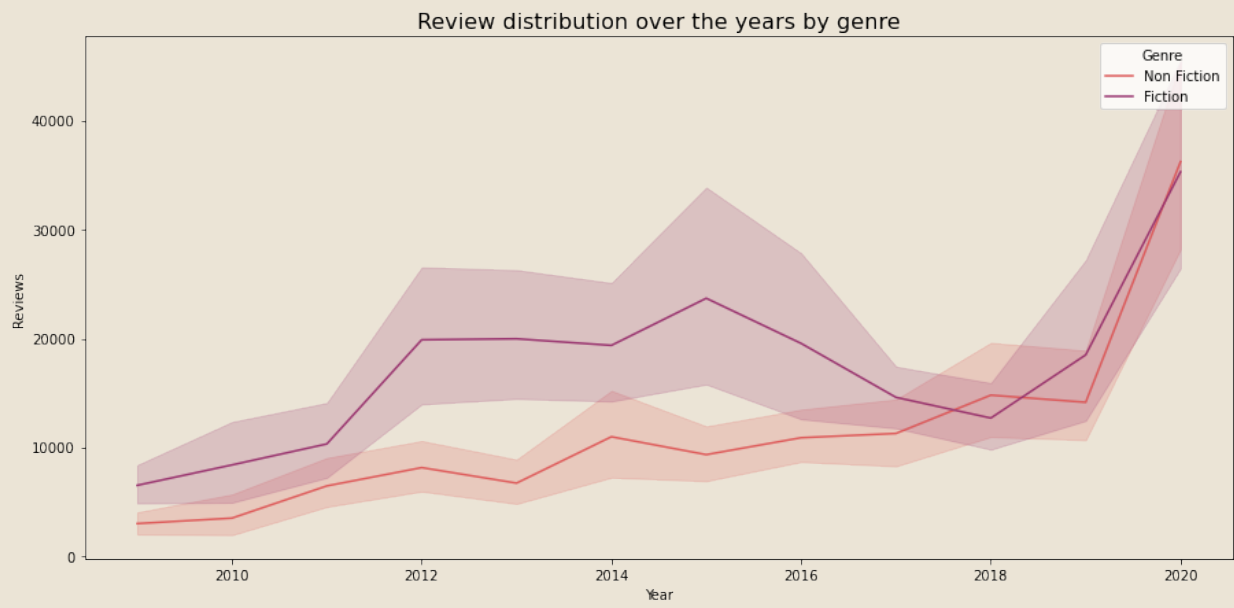
This file contains data on top 50 bestselling novels on Amazon each year from 2009 to 2020. The data was collected from Kaggle.

I'll be answering the following questions along the way:

1. Is there any correlation between the variables?
2. What is the genre distribution?
3. What is the user rating distribution?
4. What is the user rating distribution by genre?
5. What is the price distribution by genre over the years?
6. What is the rate distribution by genre over the years?
7. What is the review distribution by year and genre?







Conclusion

By the analysis of the charts I noticed that, there are no direct correlations visible between any of the variables. Which means that the User Rating is not influenced by the genre, price or year of release.

On the first chart, we can see that there is a larger amount of Non Fiction Genre than Fiction Genre

On the second chart, it is clear that the distribution User Rating is concentrated on the higher rating

On the next chart, the Fiction books seem to be doing better than non-fiction. Non-fiction however seems to be on the rise in 2020 to possibly surpass fiction books with higher ratings.

Moving to the next chart we can see that the book prices are decreasing in both genres over the years, with Non Fiction having a spike around 2014.

As it was previously showed, the lineplot rate distribution chart proves again that the rate is increasing over the years. However, on this chart you can see that both genres were on the uptrending.

On the review distribution chart, the uptrend is not clear until 2018. From 2019 we can see that both genres had a strong uptrend

On the last chart, we have the same review distribution and you can see better the spikes in 2015, 2019 and 2020.