Amazon is a top company in online consumer revenue and throughout the global pandemic caused by COVID it is assumed Amazon's revenue surged up. Using data mining as their advantage Amazon could focus on consumer trends to determine what is more likely to be purchased during quarantine times and use such trends to suggest future purchases for other users. The purpose of this project is to analyze trends used by Amazon and analyze how such information can help bring in sales and up Amazon’s growing revenue.

COVID and the global pandemic it caused had people resort to online shopping this caused a surge in Amazon’s sales. A surge in sales could be beneficial when it comes to prediction as Amazon can use consumer shopping trends to predict potential shoppers’ future purchases. In order for such prediction to work Amazon can use shopper’s wish list/saved cart information and track other user’s similar purchases. This can help Amazon figure out trends and target a specific ‘type’ of consumer to successfully suggest products and increase sale revenue.

To achieve this, data on sales prior to covid must be used to compare how trends are changing. As the global pandemic changed how people live their everyday lives it is crucial for the model to take in global factors such as this. This will allow for the model to correctly determine what to suggest to users based on outside factors. Although this is an important part of the project, I am not entirely sure where this data can be taken from and the problem on what different sources consider a ‘pandemic’ will alter results.

\*\*Because the data is hard to come by since consumer data is often rightfully protected. I am starting with using Book reviews to determine future trends\*\*

Figure (Fig.) 1 above shows the top books from 2009-2019 based on the total reviews it received. These reviews are not a completely reliable source as we must be able to have access to each review to then determine which reviews are positive versus negative. This can greatly impact the ranking, for now I assume all reviews left are positive and similar results are given when compared against Fig. 3. Fig. 3 considers the overall rating each book received which is a score out of 5, as the top books score 4.9 max, we can compare figures to see if the assumption of reviews left being mostly positive is true.

Similarly with Fig. 2 I used the same technique as Fig. 1 but this time the recent years of the book release was considered. This is important as it will allow to pinpoint which books were positively reviewed as seen in Fig. 1 but during 2019. This can give me an idea of what people were interested in during the start of the pandemic and see what comparisons can be made.

The sole purpose of Fig. 4 is due to Fig. 3 which includes repeated name titles. At first glance this looks like a mistake but thanks to Fig. 4 we can see the same book being released multiple years. This specific graph brings up an interesting take on how successful books become based on a year they are (re)released. This is arguable my most important graph as this can lead to further insight on why books have multiyear success or why certain years had a higher number of books read.