Ricky Gajeski, Cheyann Odle

Prof. Beagley

DATA-151-A

4/19/23

Video Game Sales

In 1983, after the release of E.T. the Extra-Terrestrial video game for the Atari 2600, the video game market fell into a recession as a response to the game’s (and many others) poor quality. For the next two years, the video game struggled to survive as consumers around the globe rejected video games. It was not until the release of the Nintendo Entertainment System in 1985 did consumers regain trust in the video game market and started to purchase and play them again. As with any business, it is best that an entire industry does not collapse, especially if is avoidable by your own control. Moving forward, it became important for video game companies to reflect as to why or what makes a game successful. As data analysts, we want to discover the most common variables within best-selling video games and create a model to predict what games with certain variables would be best sellers. This is an important question to answer as it would keep the video game industry thriving and overall help improve consumer reception to video games as the product would hopefully come out better and make more tailored to trends of the industry at the time. So far, the video game market has been successful in staying profitable. In fact, according to Paul Tozour of *Euronews*, the video game industry should reach a value of $326 billion by 2026, and its success during the financial crisis of 2008 has shown that the video game industry is recession-proof (*Euronews*). But Tozour also points out a potential industry problem: consumer interest is at threat if another recession occurs while the industry’s reputation stays in a negative light. With a potentially shaky future ahead, it would be important for video game companies to understand what common trends appear in the best-selling video games. Therefore, our data analysis becomes important as it would help provide video game companies with a list of trends that will help create a successful video game rather than another E.T.

A good way to begin is to start just start comparing features. “What Makes a Blockbuster Video Game?” is the title of a peer reviewed article by Joe Cox questioning what quality combination possibly make up a blockbuster video game using US sales data. Joe Cox used ordinary least squares and logistic regression models to run on his data set. He used an estimation of those models to counter the long tail of his data. The importance of a video game’s profitable success is emphasized by how much risk goes into making a videogame, by explaining how the company is in charge of all the costs but may not be able to make a profit due to the game flopping. The game flopping then bankrupts the company. This can be avoided if companies knew what qualities to focus in on. Joe Cox found, by his data, “handheld platforms are found, on average, to sell significantly fewer copies than for home platforms “ (Cox). He concluded that sales are up further with three qualities: the more major a publisher, a home platform, and good quality.

While qualities are necessary to determine relationships between features it is necessary to keep an eye on the qualities and the direction of impact or lack of impact, they have on the features you are comparing. Hoon S. Choi and coauthors wrote a peer reviewed article covering a how different qualities can correlate to the overall outcome of video game sales. “According to cue utilization theory, products consist of wide range of of quality cues such as price, brand name, packaging, and color, which indicate their potential quality to consumers” (Cho et al). This gives some examples as to what counts as a quality to include or consider comparing to overall sales. The most unexpected thing could have the deciding factor whether someone purchases something or not, especially in terms of video games. With so many games to choose from, it becomes even more critical to maximize the chances for a sale.

Can video game sales be predicted from the platform, genre and publisher?

Data and Methods

After searching for a data set, we found one from Kaggle named Global Video Game Sales. In this data set, data entries include the video game’s title, publisher, platform, release year, genre, and sales (both regional and global). With this data set in hand, we decided to use google collab for the time being to code our model, then import it to GitHub due to our familiarity with coding on Collab and ease of access for both of us to work on the code. To start, we first cleaned our data by removing entries that had null in any of the variables. We then converted each of the string variables (title, publisher, platform, and genre) to an integer in order to use those variables in future analyses. With our data now cleaned, we then created a few graphs to better visualize our data. Which look like the following:

Chart, histogram

Description automatically generatedChart, histogram

Description automatically generatedChart, histogram

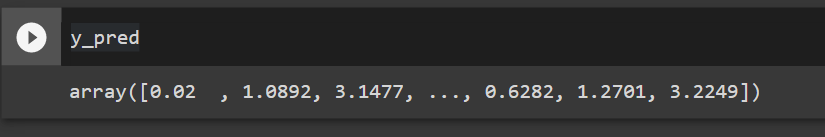
Description automatically generated Now having a good idea of how our data looked, we proceeded to create a heat map of the correlation values between the variables, which produced the following: Graphical user interface

Description automatically generated

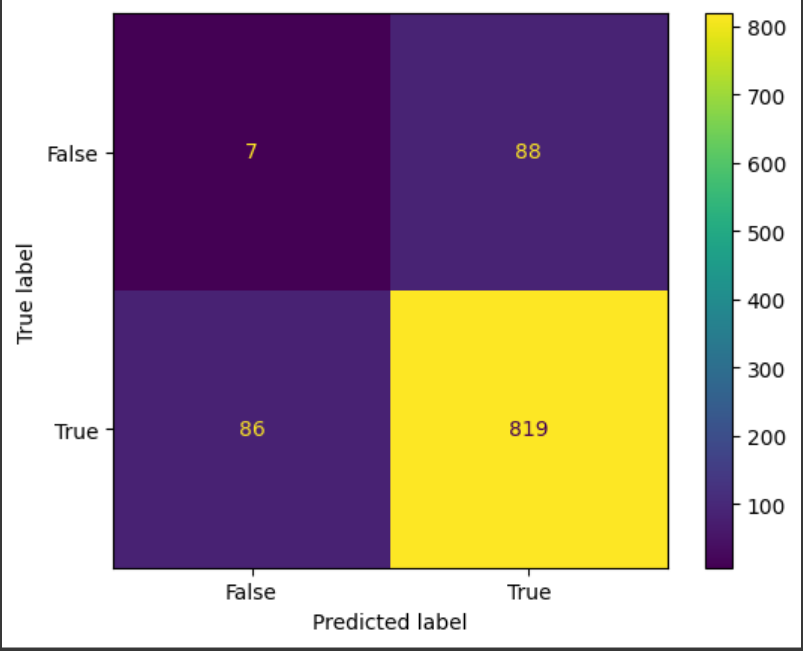
From this map it was noted that the regional sales had a high correlation with our target of the Global Sales. To make sure they were viable to use as qualities, the sum of all the regional sales were equalized and compared to the global sales. The values were equal, so the regional sale columns were dropped along with video game rank. With our data visualized, we proceeded onward with creating our model to show which variable would best predict a video game’s success. A linear model and random forest regressor were chosen. The linear model was chosen as a basis for the data as linear regressors are commonly used as a predictor. A random forest regressor was chosen as it can handle larger groups of data and get a more accurate answer, however this model can easily overfit.

Results

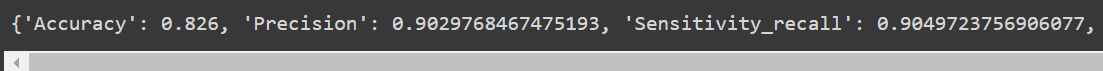
Prediction From The Random Forest For Global Sales



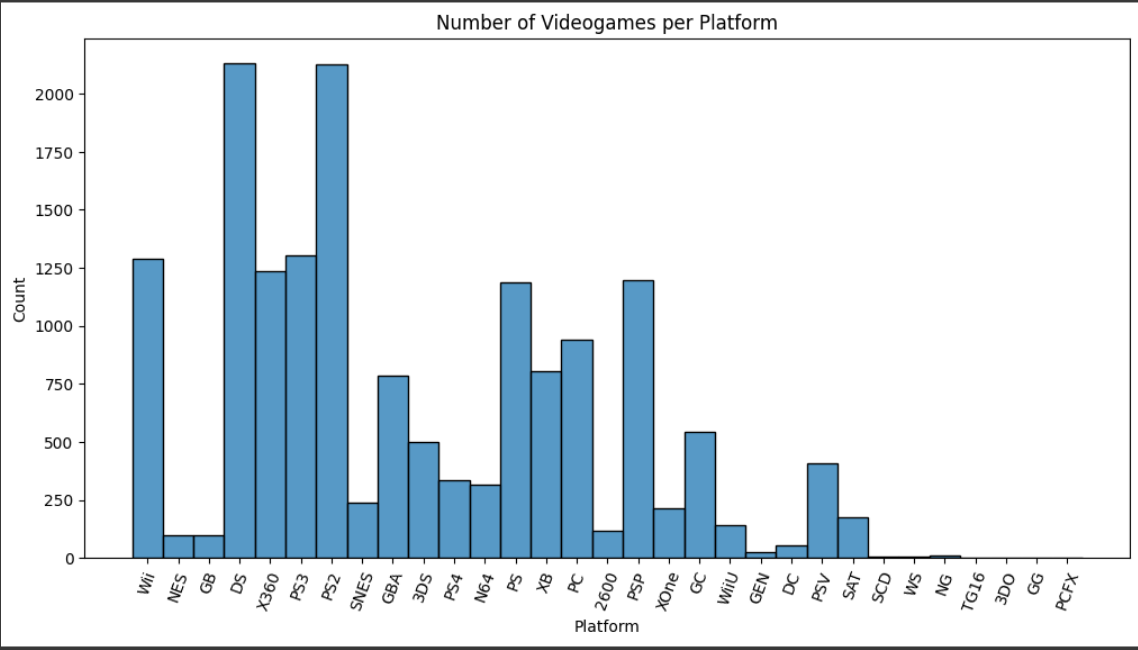
Confusion Matrix

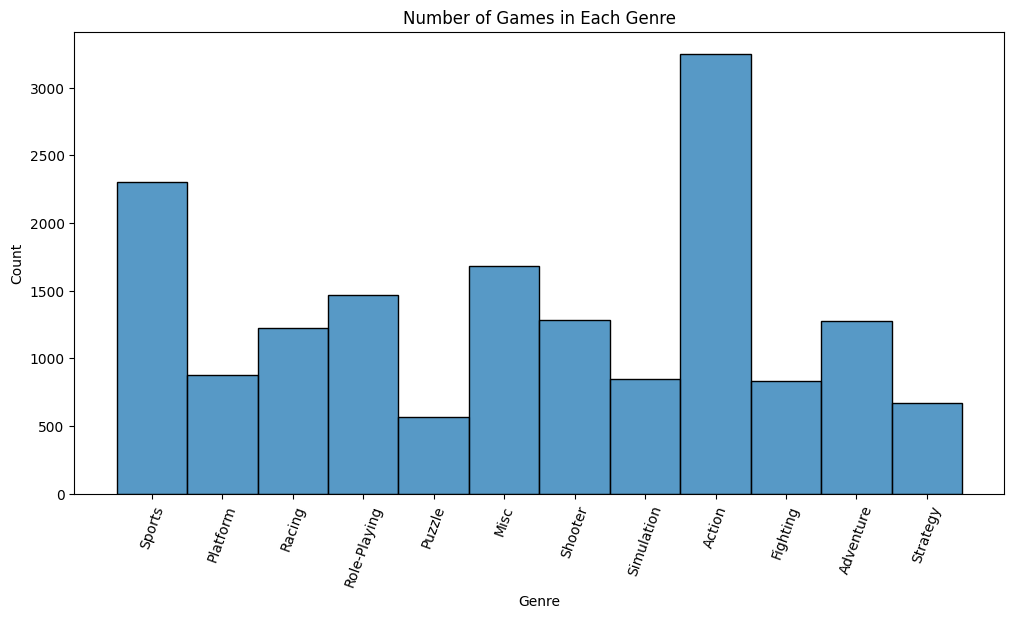


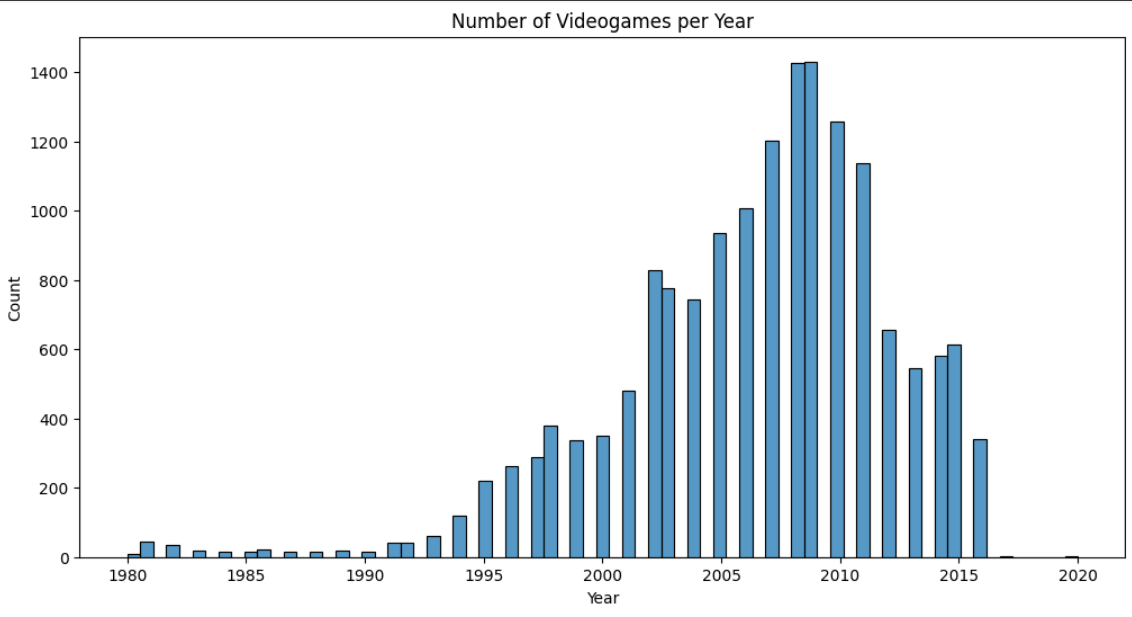
The Reading For The Confusion Matrix

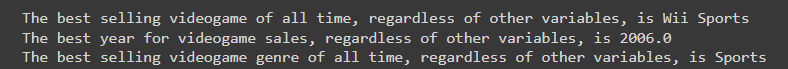






([0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11], [Text(0, 0, 'Sports'), Text(1, 0, 'Platform'), Text(2, 0, 'Racing'), Text(3, 0, 'Role-Playing'), Text(4, 0, 'Puzzle'), Text(5, 0, 'Misc'), Text(6, 0, 'Shooter'), Text(7, 0, 'Simulation'), Text(8, 0, 'Action'), Text(9, 0, 'Fighting'), Text(10, 0, 'Adventure'), Text(11, 0, 'Strategy')])





Discussion

PS2 and DS have the most in count. Action has the most games in terms of genre.

Conclusion

Currently the F score says the model is good, but the confusion matrix says the true positive readings are high.

Work Cited

Choi, Hoon S., et al. “The Effect of Intrinsic and Extrinsic Quality Cues of Digital Video Games on Sales: An Empirical Investigation.” *Decision Support Systems*, vol. 106, 2018, pp. 86–96., <https://doi.org/10.1016/j.dss.2017.12.005>. Accessed 21 Apr. 2023.

Cox, Joe. “What Makes a Blockbuster Video Game? An Empirical Analysis of US Sales Data.” *Managerial and Decision Economics*, vol. 35, no. 3, 2014, pp. 189–98. *JSTOR*, <https://www.jstor.org/stable/26607770>. Accessed 20 Apr. 2023.

Tozour, Paul. “Video Game Industry Is Incredibly Profitable. It’s Also Highly Toxic.” *Euronews*, 5 Apr. 2023, www.euronews.com/2023/04/05/video-game-industry-is-so-profitable-its-recession-proof-its-also-highly-toxic#:~:text=Video%20game%20industry%20is%20so%20profitable%20it. Accessed 11 Apr. 2023.