Video Game Capstone Project Proposal

To all video game developers, thank you for your time in discussing my proposal to hopefully help further your sales in this booming industry. While they may not have started off as hot as they currently are, the video game industry is one of the most successful in the world but it does come with some issues. While I doubt a crash as bad as the one that occurred in 1983 would happen again, there is certainly a fear of oversaturation in the market. Sure, many companies have successful franchises as well as systems that keep them in the black but more is not always better and smaller companies can struggle to survive. Releasing a similar game each year can grow stale and gamers will always latch to a new game eventually. My proposal is to look into the sales of video games, especially the all-time best sellers, to see what kinds of games sell the best.

Looking into a dataset from kaggle, I can dig into all of the best-selling games from 1985 to 2017 with much more info than just the title and the amount sold. I can look at the genre, release year, sales by regions, review scores, ESRB rating and more. I could find out what types of games sell the best in different regions. Should all focus be just on those areas or can certain games that are huge in one region become big in another? And while first-person shooters may be big now, how long will they stay on top? When looking at the years where some of the all-time greats came out, we may get an idea of how long certain genres stayed big. Plus, how important are review scores? Should companies focus more on user scores and sales of a franchise instead? There is a lot to be gleaned from this dataset just on the surface alone. However, this project looks to go a little deeper than that.

As helpful as this kaggle dataset will be, I also want to compare it with another set from the University of Portsmouth that contains sales info from 2004 to 2010. The point of looking at this group is to better compare the inferences made from the best-sellers dataset and whether they are truly applicable. As useful as that data is, we are talking about some of the best franchises ever and that name recognition can have a big factor on their ability to sell well. To compare it with a different and somewhat more contemporary list will lead to better analysis as well as predictions for the future of the industry.

And that really is the goal of this project, to analyze all of the trends in the selling of video games and creating a predictive model that will help decide what a company should do moving forward. To do that, I will need to use some in-depth machine learning techniques to create a strong model. But to take this to another level, I would also use some natural language processing practices to help look further into what games sell best. These techniques would look at a kaggle dataset of video game reviews and see which words stick out. While there will always be mixed opinions on games, these reviews will help find what words are used most often when people discuss a game they like as well as dislike. This can give a gaming company more insight into what gamers prefer for either in a specific game or an entire franchise.

With everything put together, I hope this project can show deeper strategies for video game sales. Not only will I have the code uploaded to my GitHub but also write a report that breaks down all of the major steps I take as well as construct a PowerPoint presentation to help explain the biggest takeaways from this endeavor. While their will always be fads in the industry as well as new consoles that come out, gamers will persist. This industry has stayed around and grown because of its fans but that does not mean a recession cannot happen. Companies need

to stay diligent and focused in order to stay afloat in this competitive field. Hopefully this project will give them all the insight needed to be successful and also maintain the love that gamers have for their craft.