Alex Barkovitch CMP 262 Final Project: Data Analysis Summary

For my project, I decided to perform analysis on a dataset containing information about video games. I used the following dataset from Kaggle: https://www.kaggle.com/datasets/jummyegg/rawg-game-dataset/data

I completed this project in a Jupyter Notebook, and the packages used were pandas, numpy, matplotlib, matplotlib.pyplot, and seaborn.

I've always loved games, so I thought this would be an interesting dataset to gain insights from. I asked the following questions: Which genres have the highest rated games? Which platforms have the highest rated games? Which ESRB ratings have the highest rated games, and what are the top 5 games in each ESRB category? Which games have the highest number of recommendations? How do the Metacritic ratings compare to the RAWG ratings? What games have the highest completion rate and which games were dropped the most?

I made a few charts to visualize the results of each question. When looking at the maximum value for ratings, RPG, Adventure, and Shooter games were the top 3 genres, but when looking at medians, it was Platformer, RPG, and Fighting.

When looking at which platforms have the highest rated games, the differences between max and median was even more clear. When sorting by median value, the top 3 platforms are SNES, Nintendo 64, and SEGA Saturn. It's interesting because all 3 of those platforms are very old and haven't been making new games recently, but these games have a specific style that developers could see success trying to emulate. Looking at max ratings however, it's PlayStation 4, PC, and Nintendo Switch that take the lead. It's interesting that PlayStation 4 has the highest max rating but has the 3rd from the lowest median rating.

I also gained a few insights from the analysis on ESRB ratings. Mature games had the highest max rating and the highest average, but also one of the lowest min ratings. Games rated Everyone had the lowest maximum rating compared to other ESRB ratings, but the highest median value. I also looked at the top 10 games with highest ratings in each category. Persona 5 Royal was the overall highest rated game in the list, and is rated Mature.

When looking at which games had the most recommendations, the top 3 most suggested games are Call of Duty: Black Ops III, The Walking Dead: Season 1, and Mass Effect 3. The game with the lowest number of suggestions is Undertale.

I also compared the ratings of games from different websites, Metacritic and RAWG. It was interesting how different these results were. The top 3 Metacritic rated games are Grand Theft Auto V, Super Mario Odyssey, and The Legend of Zelda: Breath of the Wild. The top 3 RAWG rated games are The Last Of Us Remastered, The Witcher 3: Wild Hunt, and Portal 2. It's also interesting that in the RAWG ratings list, Call of Duty: Black Ops III is rated second from the bottom, but was the game that had the most recommendations.

When looking at completion status of different games, I added up the values for amount dropped and amount completed and looked at the 100 games with the highest total sum. Uncharted 2: Among Thieves had the highest completion rate average, and Dota 2 had the lowest; only a 20% completion rate, with 80% having dropped the game. When just looking at the numbers, Grand Theft Auto V has been completed more than other games on the list.

If I were to present this analysis to a video game developer, I would tell them that platformer's and RPGs are the most popular genres right now. If it has a mature ESRB rating and is available on Xbox, PlayStation, and/or PC, the game is likely to be successful, according to these results.

Regarding future work, I think web scraping other video game review sites to get a wider variety of data would be very insightful. Comparing results from multiple different gaming websites is helpful to make results of the analysis more accurate. As we can see from comparing just the game ratings of Metacritic and RAWG, I would likely get some different results for my other questions based on using ratings from other sites. Combining the results from these different sources would make recommendations to developers or game sellers more valuable.