Video Game Sales Prediction Analysis

**Abstract**

With video games becoming a more present pass time and even a more audience-based competition, it is key for video publishers to determine what types of games the general population of video game players are pursuing and which genres they will continue to enjoy in the future. Therefore, a predictive analysis is to be performed on video games sales across North America, Europe, Japan, and across the Globe to determine which video game genre will produce the most revenue in each geographical location and globally overall. To orchestrate this analysis, historical video game data will be cleaned and structured to allow for models to be built to predict future video game sales for each respective location. With these predictions, it will show how much each genre is gaining in sales by the millions.

**Intro/background of the problem**

In the beginning, video games were primarily played at locations where people of all ages would come to play and take turns competing with each other. However, as the years progressed, platforms were being developed allowing for blooming gamers to now play more conveniently at home. Since then, more and more video game consoles have been produced and integrated into our society whether it be a PlayStation, an Xbox, or a Switch from Nintendo containing games that spread across various genres such as racing, shooter, role-playing, or platforms. Because of the continuous development of video games, a new console around every 7-8 years is released allowing for video game publishers, those that create the video games, to spread their content. But, with the game community constantly expanding, how do publishers determine which genre of video games to focus their efforts on for the present or upcoming consoles?

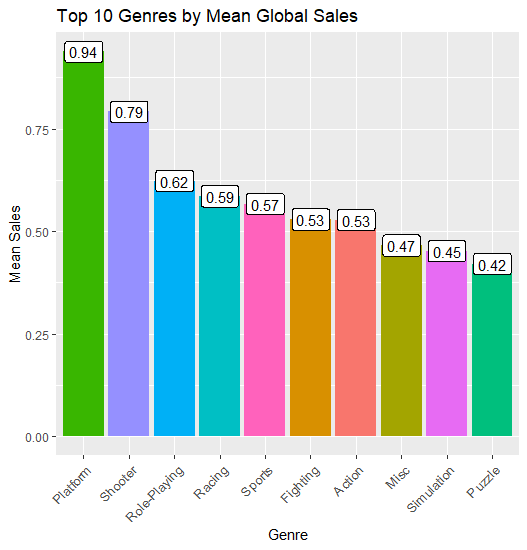
**Methods**

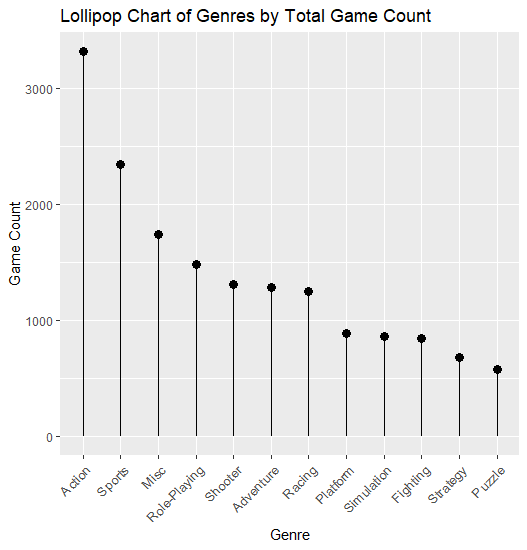
**Data Preparation/Cleaning**

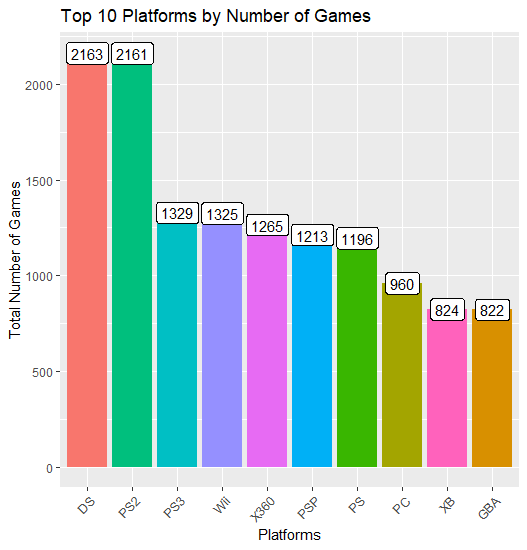
For the data preparation aspect of this analysis, I first looked into the data set to see if empty or NULL values existed within it. If they did, I removed the record of the video game that contained them. Luckily, in this data set, it did not contain any empty or NULL values. The genre and year variables had their data types changed to character and numeric, respectively.

**Exploration Data Analysis**

In investigating the data that was present, I looked into the top genres across all of the geographical locations by using the mean sales of each area. At the same time, I wanted to gain perspective on how many games had been released within each genre to determine if the sales of the genre were higher simply because more games were released for the genre. Lastly, I explored how many games were released on each platform in the data set.







**Feature Selection (Still in Progress…)**

In determining which features to select, I will not use all of the features possible, but plan on using the features of genre, year released, sales from each geographic location, and the platforms that they were released on since many of the consoles come from the same developers such as Sony, Microsoft, and Nintendo. My primary focus is to find the prediction of the sales for North America therefore it will be used for the instances of the model in the future.

**Modeling (Still in Progress…)**

For my predictive model, the current plan is to employ LSTM, Long Short-term Memory method. Since I am seeking video game genre sales for the future, I will be using this time-series forecasting model to determine how much of a genre will produce across all of the locations.

**Results**

From looking at the exploratory analysis that I performed on the data, I would make an educated guess that the sales for the platform genres will continue to increase; however, video games were more centered on platforms for a long time. Therefore, I believe that the shooter genre will see an increase in sales since it has been present for many years and is still producing more content. On the other hand, looking at the other locations, this is not the case. For example, the Japan sales for shooter did not reach into the top ten genres for this area, but it had role-playing genres considerably higher than most of its other genres.

**Discussion – What’s next?**

Even though my main focus is to look at the North American sales for genres, I truly want to explore the same scenario for all the locations as well as globally since many video game developer’s audience reach all over the globe. Along with determining the genres, I would be interested in looking into the sales forecast for the video game publishers to determine if they are continuing or drastically changing.

**Conclusion**

At this time, I do not have a conclusion of the analysis since the modeling and results are currently still under development.

**Acknowledgement**

I appreciate the support and growth of the video game community for all types of people that play the games as well as the video game publishers who constantly acknowledge those that play their games. At the same time, I thank everyone for any support and or criticism that they may have on the analysis. I am still learning and hope to do my best.

**References:**

Poh, M. (2020, June 23). Evolution of Home Video Game Consoles: 1967 - 2011. Retrieved June 28, 2020, from <https://www.hongkiat.com/blog/evolution-of-home-video-game-consoles-1967-2011/>

History Channel. (2017, September 01). Video Game History. Retrieved June 28, 2020, from <https://www.history.com/topics/inventions/history-of-video-games>

Data Source:

<https://www.kaggle.com/gregorut/videogamesales>