Video Game Sales Prediction Analysis

**Executive Summary**

Video games have become a multi-million form of entertainment, technology, and art all wrapped into one experience for someone of almost any age to experience on various platforms. As technology in our world continually evolves and allows for us to dive deeper into it, so do the systems that allow us to play all different types of video games. With these new systems, or consoles, being distributed around the world every seven to nine years, companies are interested to determine the future of video games sales as they release the newest platforms which have the possibility of acting faster, providing sharper and more realistic graphics, and creating whole new worlds within games. Therefore, this project focuses on the business inquiry of predicting future video game sales upon the release of the newest consoles allowing for these games to be played by the video game community.

To be able to achieve this goal of being able to predict future sales, it will involve historical information about all video games that have been sold from the past forty years for the use of consoles. Because of the use of consoles, it will not involve arcade games and mobile games that have been released within the mentioned time frame. At the same time, since video games have become a global moment that has reach nearly every country allowing for users to be able to play and communicate with others all around world, it is key for companies to understand how video game sales are progressing especially since the two major consoles will be open to the public by the end of the year which include numerous new games for each respective console.

**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 future sales and 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 location is gaining in sales by the millions.

**Introduction**

**Background**

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.

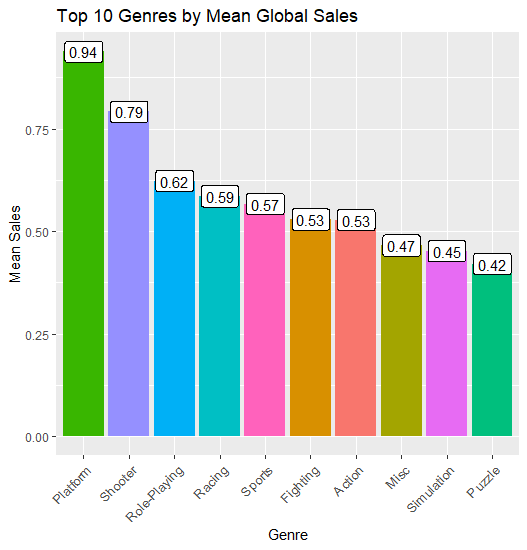
**Methods**

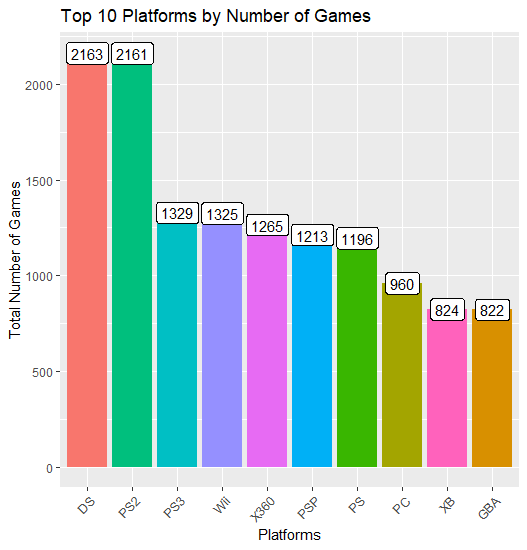
**Data Preparation**

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. For the model that is being employed for this analysis, I have also converted the time series data to supervise to be able to use features for the model.

**Exploration Data Analysis**

In investigating the data that was present, one of the goals was to determine the mean amount of sale across the North American, European, Japan, and Global regions for videos games that have been released for consoles. With the new consoles being released primarily by Sony and Microsoft, it was important to determine how their past consoles have done in the scope of how many games were released for their consoles. Even though there are more consoles within the data set such as the ones developed by Nintendo, these two main console developers are aiming to have their consoles released within a short time of each other. Below, are bar graphs giving a glimpse into the global sales of the various genres as well as the count of video games to console. With the second visual, it is important to remember that there are various video games which are nearly identical that were released for multiple consoles.





**Feature Selection**

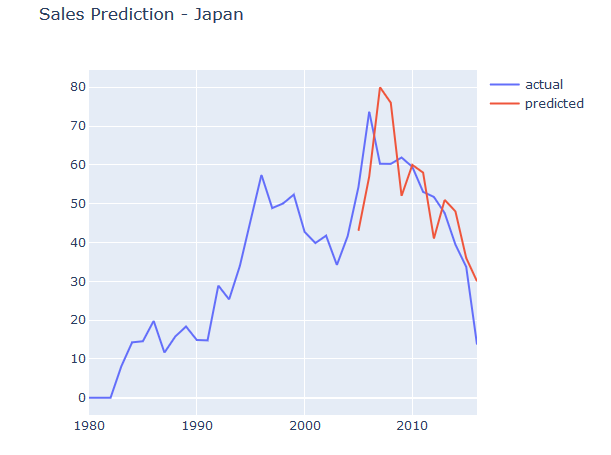
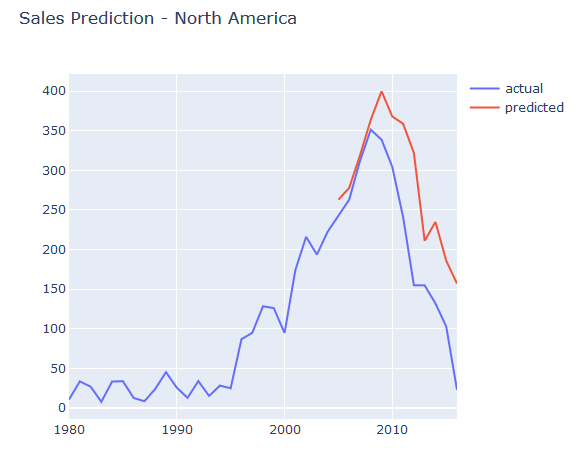
In determining which features to select, the use of previous years of video games sales data will be utilized to forecast the upcoming years. For my model, I have chosen to go back twenty years which is about halfway back into the data set as well as the fact that gaming consoles began to increase highly in popularity towards the end of the 90s and the beginning of the 2000s. Depending on how many years you want to have for your model, this value can change between different models.

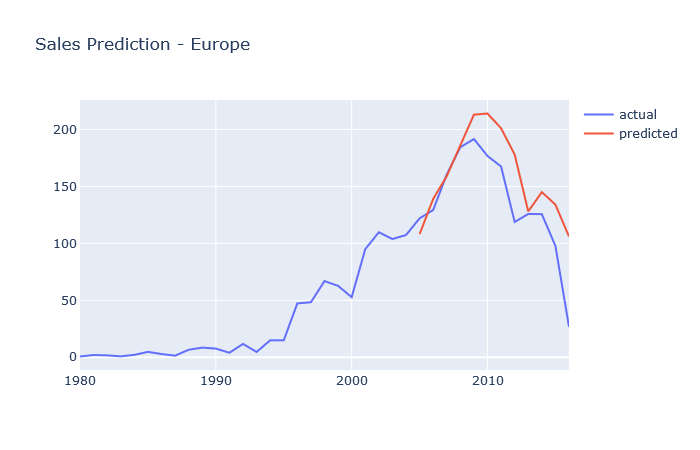
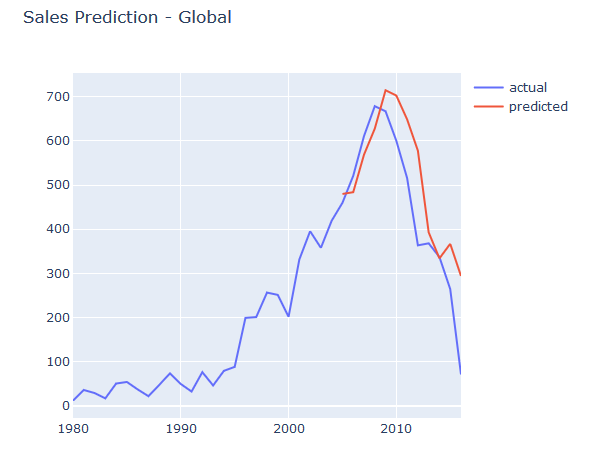
**Modeling**

Once you have decided the range of years for your model, it is important to test how well you selected said features for your model. To do so, you will fit a linear regression of ordinary least squares and calculate the adjusted R-square. Upon scaling your data for a high accuracy, it allows you to be in a strong position to construct the model for predictions. Before starting the model, the data will be split for training and testing which will contain the last seven years of sales. For the predictive model, the plan is to employ LSTM, Long Short-term Memory method. In creating this type of model, it will need to include feature and label sets from scaled datasets. With fitting the LSTM model, it will improve upon itself and attempt to reduce the error in each one of its one instances to continually learn from itself. After the model has had the opportunity to train with the historical data, the next step is to call a prediction with the test data that was split off from the data set. After reshaping the prediction results from the model, we can then compare the results with the actual data from the set. From there, we will be able to see the comparisons from our predictions and the actual sales for all of the geographical locations that have been mentioned.

**Results**

Upon building the models for each of the geographical locations as well as on a global scale, I was able to see the actual results against the sales that were predicted from the model. For each of the models, it was able to see the climb in sales around 2006 as well as the decrease in results around 2012. However, because of the data source that was used for this analysis, I do not believe that it included all of the games sold for consoles after the year 2015 since the video game sales drop very drastically. From previous knowledge of the video game community, I know that the community is growing and more games are being released considering all of the remaster editions that are being transferred from pervious consoles or video games now becoming more available for other consoles than they were originally released for at the time. Overall, when it comes to the deep learning predictive model that I utilized for this project, it was able to fairly predict the sales forecast for the years based on the historical data that was used to train the model in the beginning. It did well in determining the trend that video games were taking in each geographical location as well as stay in the price range for each region.





**Discussion & Conclusion**

When it comes to video games as a whole, it no longer can be centered on one area of the world and be treated the same in each geographical location. For companies, being able to determine and optimize their focus on future sales of video games within each area would determine which geographical locations are purchasing video games which could then be correlated to which areas will be the next genre of consoles or simply the trend of the video game community as a whole. My deep learning predictive model allowed my project to reach its goal in that it was able to fairly predict the trends based on historical data. In the future, if someone were to carry on my project, I believe there is the possibility for it be more accurate and possibly predict further and change feature set.

**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 especially Professor Iranitalab. I am continuing to learn 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>