EDA - Video Game Sales

By

Shubhkirti Prasad,

Rushank Ghanshyam Sheta,

Harsh Maheshkumar Pandya

shubpras@iu.edu, rsheta@iu.edu, hapandya@iu.edu

Under the guidance of **Prof. Julia Fukuyama**

Abstract – In this study, we present an exploratory data analysis of a dataset that includes information on the sales and evaluations of video games in this study. By looking at many factors like sales, ratings, genre, platform, and area, the analysis seeks to develop an understanding of the video game market. The report details the steps taken to organize and clean the dataset, then uses visualizations and statistical analysis to find trends and connections between the variables. Important discoveries include the dominance of particular platforms and genres, the effect of ratings on sales, regional variations in video game preferences, and understanding why some publishers did better than others in this ever-evolving video game industry. The report ends with a summary of the insights and suggestions for additional research.

Keywords – Video Game Sales, Video Games Trends, Exploratory Data Analysis (EDA), Data Analysis, Univariate Analysis, Bivariate Data Analysis, Multivariate Data Analysis, and Statistical Modelling

1. Introduction –

With a projected global market value of $143 billion in 2020, the video game industry is a quickly expanding and dynamic sector of the economy. With millions of participants, the market is extremely competitive and is influenced by trends and consumer preferences. Companies in the video game industry need a thorough understanding of consumer behavior, sales trends, and new trends in order to stay one step ahead of the competition.

This report presents an exploratory data analysis

(EDA) of a dataset containing information on video game sales and ratings. The dataset, which was obtained from Kaggle, contains information on more than 16,000 video games that were released between 1980 and 2016. The report opens with a succinct overview of the video game industry and the value of data analysis in making wise business decisions. The method used in the EDA is then explained, along with the steps taken to clean and get the dataset ready for analysis. In this report, we have mainly focussed on generating insights using Sales as a metric because Game Critics or User Reviews tend to be more inaccurate or biased sometimes. An overview of the learned lessons and suggestions for additional research are provided in the report's conclusion.

2. Objectives -

The video game sales and ratings dataset's exploratory data analysis (EDA) can be helpful for a variety of purposes, including business decision-making, spotting market trends, comprehending consumer preferences, and informing marketing plans. Companies can choose which games to develop, which platforms to target, and how to market their products by gaining insights into sales patterns, ratings, genres, platforms, and regions. The EDA can also assist in spotting developing market trends and offer insightful data on consumer preferences, giving businesses an advantage over rivals. The purpose of this analysis is to explore answers to questions like What are the current market trends for video games, and how have factors like region, genre, platform, and year of release changed over time in this constantly evolving industry?

3. Contents -

4. Data Description -

The dataset provides information on video game sales from various regions, including North America, Europe, Japan, and other regions, as well as global sales. The dataset includes 18 columns and 16,598 rows where each row represents sales for a given video game for that given year and it includes the following variables:

- Name: the name of the video game.
- Platform: the gaming platform (e.g PC)
- Year_of_Release: year of game release
- Genre: the game genre (e.g action, sports)
- Publisher: the name of the publisher.
- Sales(in millions): sales in North America, Japan, Europe, Other Regions, Global. Namely NA_Sales, JP_Sales, EU_Sales, Other_Sales, Global_Sales)
- Critic_Score: aggregate score compiled by Metacritic website.
- Critic_Count: the number of critics used in coming up with the Critic_score.
- User_Score: score by Metacritic's subscribers.
- User_Count: number of users who gave the user_score.
- Rating: The ESRB ratings (Entertainment Software Rating Board).

5. Data Preprocessing -

This included pre-processing like converting the "Year_of_Release" attribute from categorical to continuous and removing null values from it. We preprocessed the "Year_of_Release" feature by removing null values and retaining only the data until 2016. We made this decision due to the fact that there were only a few observations for years beyond 2016, and thus insufficient data to make meaningful conclusions. In order to conduct a more focused analysis of trends and patterns of video game sales and reviews, the dataset was filtered to include observations with year values between 1980 and 2016.

Additionally, new data frames were generated from the filtered dataset to better structure the data for information visualization and to address specific research questions.

6. Univariate and Bivariate Analysis -

6.1 Distribution of Global_Sales

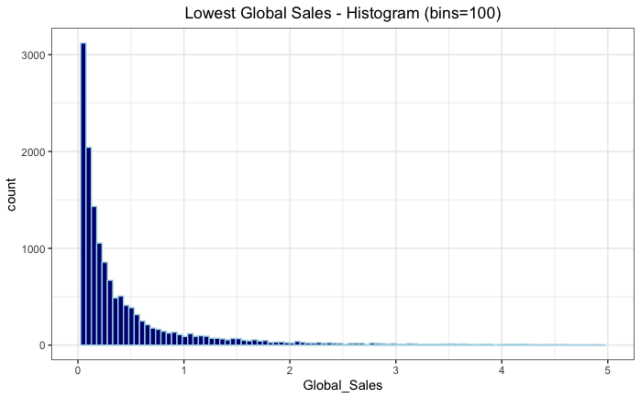

Fig 6.1

For the above histogram plot, we had to limit the x-axis because the data is highly skewed and most of the sales data had values less than a million dollars. From the above histogram, we can visualize the distribution of sales variable i.e for most of the instances (row) the sales in one year are zero to one-tenth of a million.

6.2 Distribution of Sales over time for each region

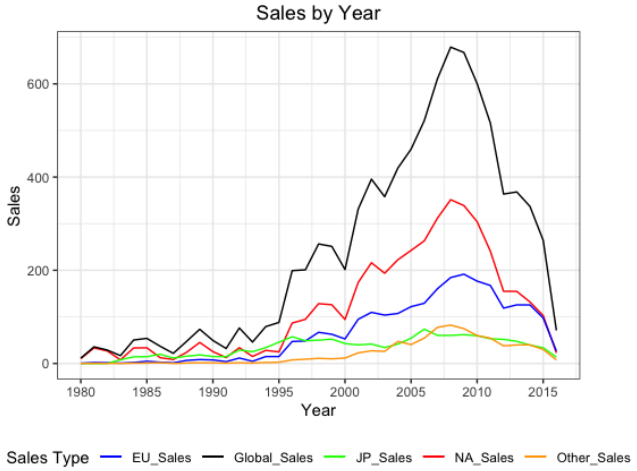

Fig 6.2

From the above plot, we can observe that there was a steady increase in video game sales globally from 1995 to 2008. One reason for this could be the increase in popularity of gaming consoles such as PlayStation, Xbox, and Wii during this time period. However, after 2008, there was a fall in sales which could be due to the global financial crisis that occurred in that year. The decrease in disposable income for many people may have led to a decrease in video game purchases. Another pattern we observe is that the North American market dominates global video game sales. This could be due to various reasons such as the large population of gamers in North America, the presence of major gaming companies in the region, and a higher disposable income for purchasing video games. The sales in North America also showed a strong correlation with the global sales trend.

Total Sales in Europe were around 200 million in 2008, while sales in other regions combined were around 100 million. It's interesting to note that Europe sales were less than half of North America sales, which were around 350 million. This could be due to cultural differences and preferences in gaming, as well as differences in marketing

strategies and availability of gaming platforms.

On the other hand, the sales in Japan didn't correlate with the global sales trend. One reason for this could be the difference in cultural preferences for gaming. Japan has a unique gaming culture with a focus on arcade games and handheld consoles, while console gaming dominates the market in other regions. Sales in other regions, which include countries in South America, Africa, and Asia, may have also been affected by factors such as economic conditions, accessibility to gaming technology, and cultural preferences.

Overall, the peak for all regions in terms of video game sales was around 2008, which may have been influenced by the release of major gaming consoles and popular games during that time period.

7. Multivariate Data Analysis and Generating Insights & Answering Questions -

7.1 What are the most popular gaming platforms?

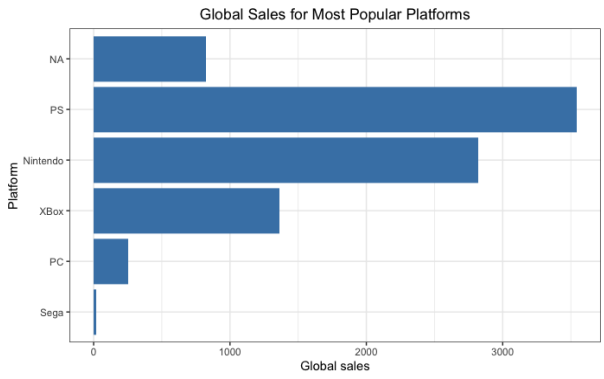

Fig 7.1

Based on the plot above, it can be observed that PS has the highest sales globally, with sales of around 3500. This is followed by Nintendo with sales of around 2800. On the other hand, Sega marks the lowest sales, with very less sales.

7.2 Which Genre has most of the sales?

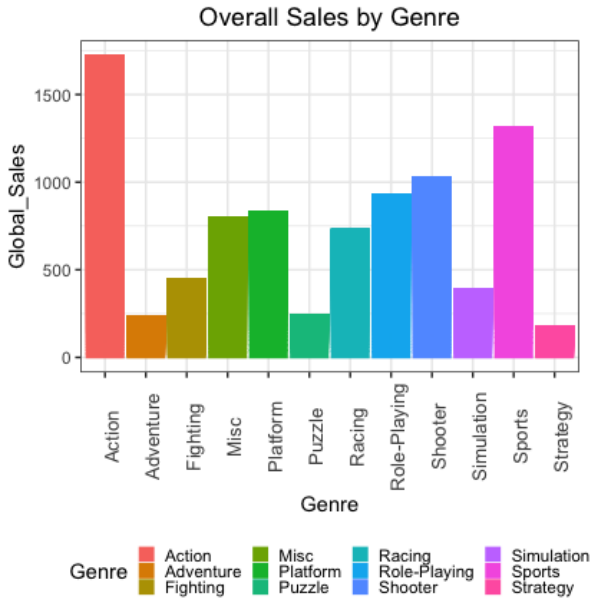

Fig 7.2

Based on the plot above, it can be observed that the Action genre has the highest sales, with a value of approximately 2000 with the Sports genre coming second, with aggregate sales of around 1250, followed by the Shooter genre with sales of about 1000. The lowest sales value can be observed for the Strategy genre, with a value of around 250. Overall, the plot indicates that the Action genre has the highest sales among all genres, followed by Sports and Shooter genres

7.3 Which Publisher has the highest Global Sales?

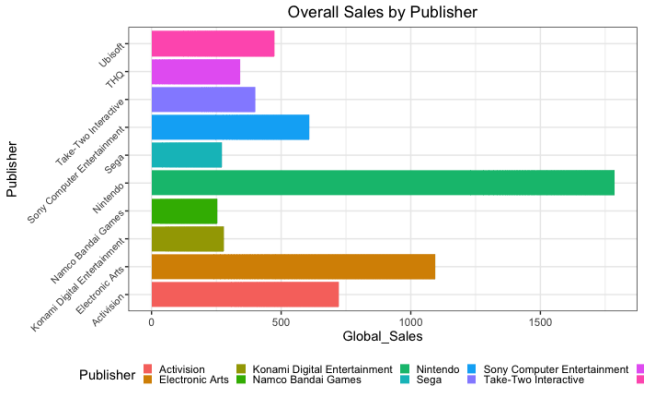

Fig 7.3

From the plot above, it appears that Nintendo has

the highest global sales with approximately 2000, followed by Electronic Arts with around 1150 and Activision with 750. Sega and Namco Bandai Games have comparable sales with around 200 sales. These numbers indicate that Nintendo has a significant lead in global sales compared to its competitors

7.4 What are the all-time best Games and Platforms in North America?

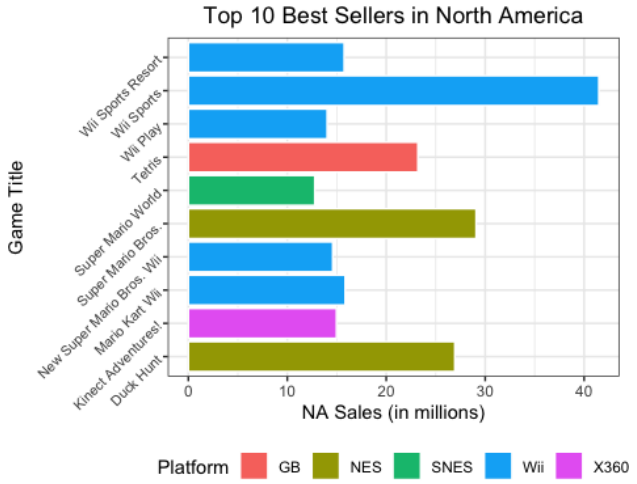

Fig 7.4

From the above plot it is observed that the North American video game market has been dominated by a few all-time classic titles. Topping the list is the sports simulation game "Wii Sports" which has sold over 40 million in the region. The second best-selling game is the iconic "Super Mario Bros." with sales of around 30 million. Following close behind is the shooting game "Duck Hunt" with sales of approximately 25 million. Another installment of the Mario franchise, "Super Mario World", is in fourth place with sales of 15 million. Overall four of the top ten games are for the Nintendo Wii platform.

7.5 Which games had the highest sales in Europe?

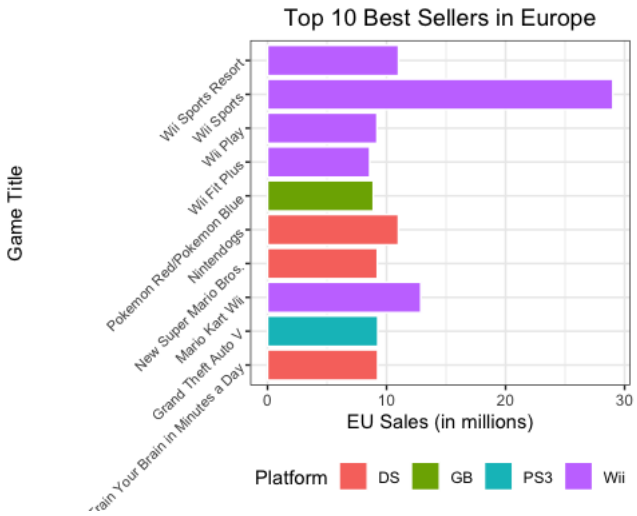

Fig 7.5

The top three best-selling games in Europe are Wii Sports, Mario Kart Wii, and Wii Sports Resort, with sales of 28, around 12, and around 10 million USD respectively. Wii Sports, Super Mario Bros, and Duck Hunt were the top three best-sellers. Similar to North America, the Wii platform dominates the top ten best-sellers chart in terms of the number of games featured. However, the sales of the other games in Europe are generally lower than in North America, with most games having sales below 10 million.

7.6 What is the most popular games and platform in Japan?

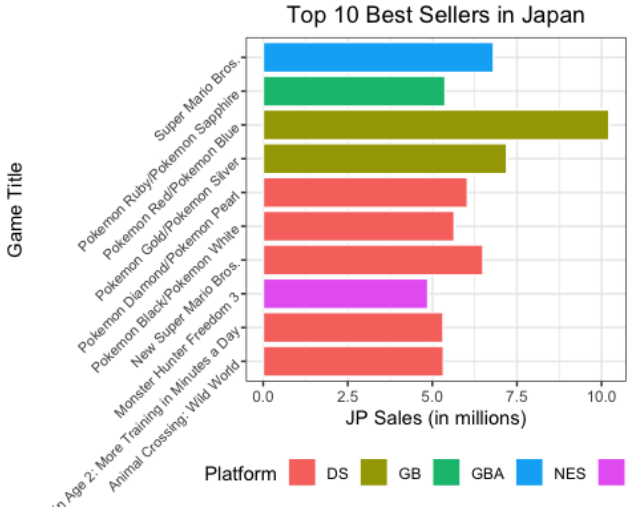

Fig 7.6

For Japan, the top-selling video game is Pokemon Red/ Pokemon Blue, which has sold around 10 million USD over the years. The next best-sellers are Pokemon Gold/ Pokemon Silver with sales of around 7.5 million, and Super Mario Bros. and New Super Mario Bros. which have comparable sales to Pokemon Gold/ Pokemon Silver.

When compared to North American sales, the sales numbers in Japan are generally lower. For example, the best-selling game in North America, Wii Sports, has sold 40 million, while the top-selling game in Japan, Pokemon Red/ Pokemon Blue, has sold only 10 million, and most of the best-selling games are released for the Nintendo DS Platform.

7.7 Who are the top publishers for each year?

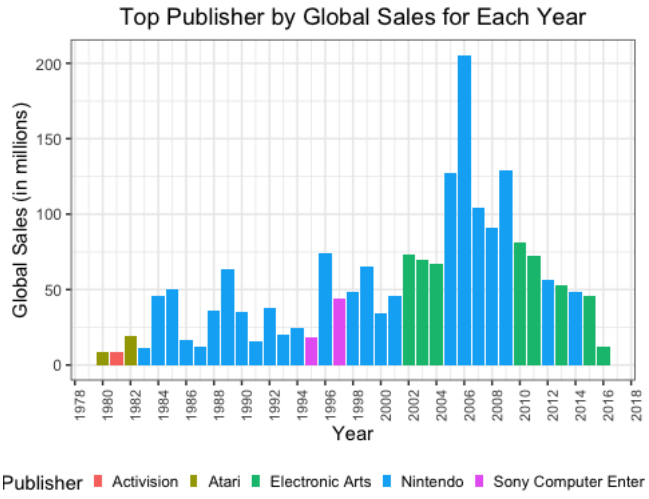

Fig 7.7

Atari and Activision were dominant players in the early years of the video game industry, with the former launching the first home video game console, the Atari 2600, in 1977. Sony Computer Entertainment has been a major player (only two times) in the video game industry since the launch of its PlayStation console in 1994. The company's dominance in 1995 and 1997 can be attributed to the success of the PlayStation. However, the company also faced challenges, particularly in the mid-2000s. Electronic Arts (EA) has also been a significant player in the video game industry, particularly in the sports genre. Nintendo had a dominant presence in the video game industry for several decades, starting from its early years in the 1980s with the launch of the Nintendo Entertainment System in 1983.

In summary, Nintendo was the top Publisher for most of the years. However, we can see a trend where Electronic Arts started selling more games than Nintendo from the year 2002 until 2005 after which again Nintendo was the top publisher. And lastly from the year 2010 Electronic Art's sales were more than Nintendo's. For further analysis we are particularly interested in analyzing data from the years 2005-2009 and 2010-2014 which will help us understand how Electronic Arts was able to beat its competitor Nintendo in terms of total global sales.

7.8 Why did Electronic Art's, and Activision's sales increase and Nintendo's sales decrease?

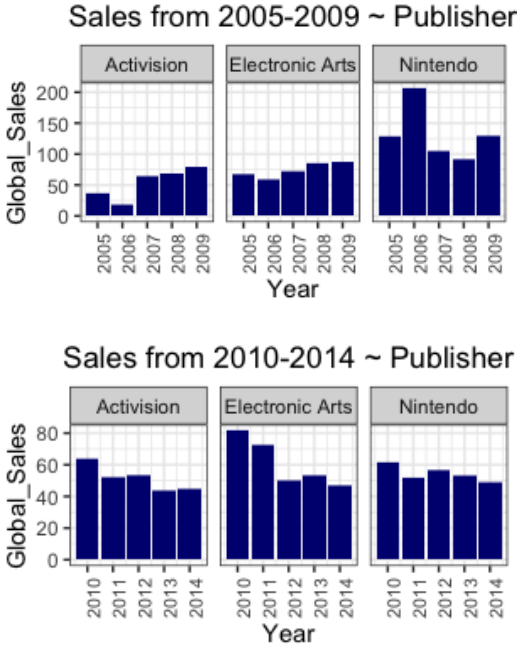

Fig 7.8

During the years 2005 to 2009 Nintendo, Electronic Arts(EA), and Activision had the best sales. Nintendo's sales were all-time high of about 200 million and averaged around 130 million over this period of 5 years, while EA had average sales of 75 million followed by Activision with 60 million in sales. However, from the year 2010 to 2014 Nintendo's sales dropped to an average of 55 million and EA's average sales were highest at around 60 million followed by Activision with an average of 50 million. In summary, we observed that Nintendo was having highest sales in the mid-5 years while in the last-5 years, EA crossed Nintendo in average sales and Activision was also able to significantly increase sales compared to other publishers. Hence we are interested in finding the probable reason why Nintendo's sales dropped by around 50% while sales for EA and Activision increased from the years 2005 to 2009 and 2010 to 2014 for these publishers.

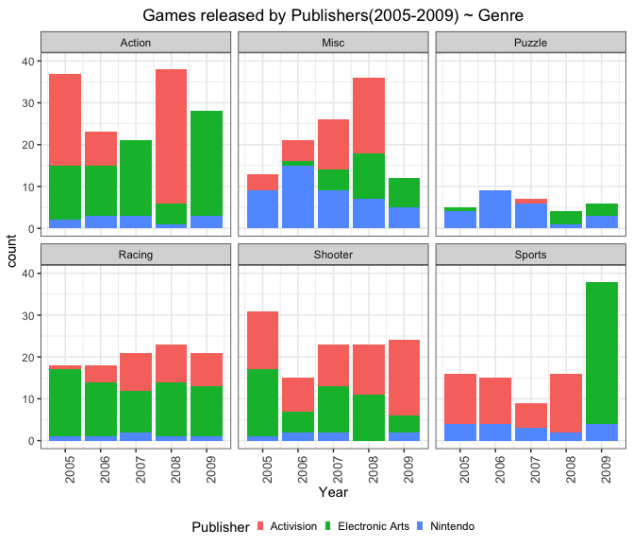

Fig 7.9

From the above-stacked bar plot represents the number of games released(y-axis) by the publishers Electronic Arts, Activision, and Nintendo for Genres like Action, Miscellaneous, Puzzle, Racing, Shooter, and Sports from the years 2005 to 2009.

Nintendo was releasing comparatively more games in Miscelenous and Puzzle Genres while Activision was releasing more games in Action, Shooter, and Sports. EA was also releasing comparatively more games in Action, Racing, and Shooter categories. In the year 2009, we can see a huge number of Sports games released by EA and Activision not releasing any games for sports.

Generally, we can see a trend that suggests Activision and EA releasing more games in categories like Action, Shooter, and Sports when compared to Nintendo which had little to no game releases in these categories. This might be the reason why EA sales increased more than Nintendo for some of the upcoming years.

Let's inspect the trends of Game releases and Genres for the upcoming five years.

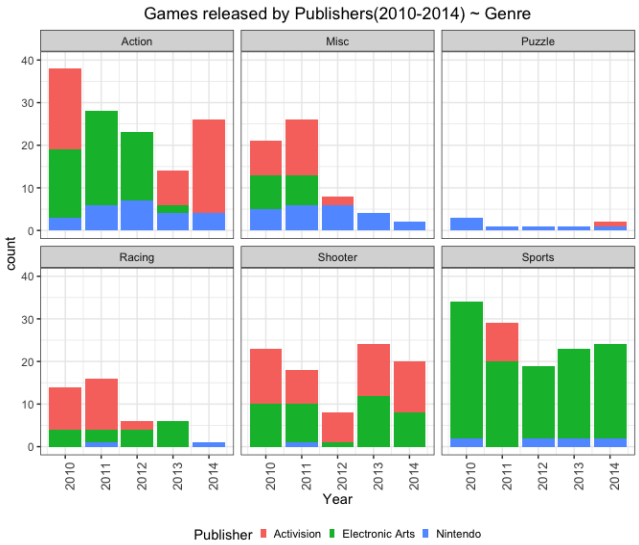

Fig 7.10

From the years 2010 to 2014, we can see that Electronic Arts kept on releasing most of the

games in the Sports and Action category and was mainly dominating the Sports category with not much competition from EA and Nintendo as they were not releasing many games in the Sports Genre. This might be one of the reasons for EA being the top-game seller for years 2010, 2011, 2013, and 2015 and also suggests that Sports was a booming game category at that time. Also, Activision was releasing video games in Action, Shooter and it stopped releasing games in Miscelenous and Racing from the year 2013 until 2014. However, the sales for Activision in these years increased compared to the prior year which suggests that Genres like Action, Shooter are becoming more popular with higher sales. Nintendo was not focusing on these hypothesized popular Genres like Action, Sports, Shooter, and Racing while its sales decreased from previous years, this further supports the hypothesis.

To answer the question of why Nintendo's sales dropped by fifty percent in the later years when EA and Activision were able to increase their total sales, we believe that EA and Activision understood the future trend of video-game Genres which suggested that Genres like Shooter, Action, Sports and Racing be some of the most popular categories and focussing more on these Genres by releasing most of the games in these categories while Nintendo did not follow this trend and released comparatively more games in Miscellaneous and Puzzle, which indicates these are not that popular Genres. Even though Nintendo released some games in Action and Sports it wasn't able to sell more games than Electronic Arts in some of the later years.

8. Modeling:

The trends in Japan seem to defy the Global trends of Video game sales. Our motive behind modeling to was to create a comparative study between the regional trends, namely Japan Vs rest of the world. For the same we developed a 2-pronged approach, first, to create a Loess plot that portrays the behavior of video game sales in Japan as compared to North America faceted by different levels of Global sales. Second, a loess plot faceted by North American sales to see the performance to see how the Global trends affected the sales of video games in Japan.

8.1 What is the trend of Genres over the years?

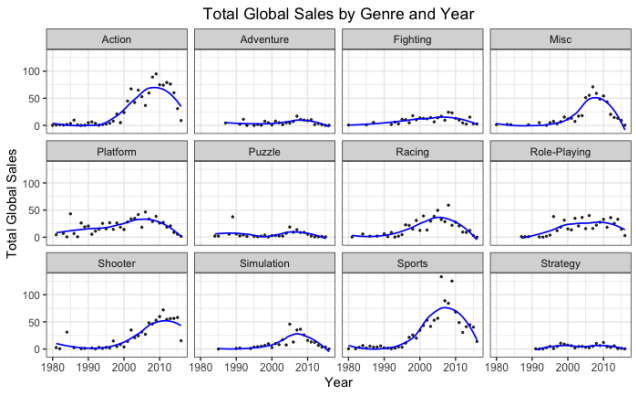

Fig 8.1

From the above Scatter plot with Global Sales on Y-axis and Years on X-axis faceted by Genre we can see the overall trend of sales for video games in each Genre. We have also fitted a loess model to better understand how sales change over time. Overall we can see that the sales at the ending years are decreasing for all of the genres which could be because of many reasons such as fewer data collected for later years, sales data not being made open source, etc. However, we can see the relative change in the sales for each Genre like Action, Misc, Racing, Shooter, Sports, and

Simulation had the highest sales around the year 2009. Action and Sports were the genres with the highest change while adventure, strategy, and puzzle had the lowest increase of them all. Sales for shooter games increased and did not drop much like Action and Adventure while Miscelenous was the genre where a huge drop in sales can be seen. In summary, the insights highly correlate with the inference from 7.10 which is Shooter is the most popular genre followed by Action and Sports, while the least popular genres are Puzzle, Strategy, Adventure, and maybe Miscellaneous.

8.2 Japan VS North American Sales, given global sales:

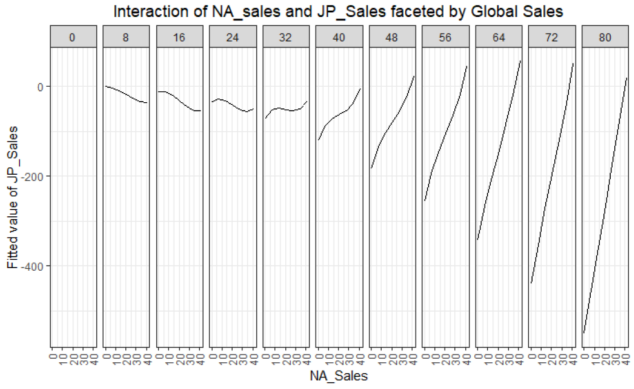

Fig 8.2

The plot shows the relationship between `NA_Sales` and the predicted values of `JP_Sales` based on the fitted loess model, with facets for different values of `Global_Sales`. The plot portrays the varying contribution of Japanese sales to global sales Vs North American sales to the global sales. The slope changes from a negative value to a positive value indicating that initially, the given games were more popular in Japan as compared to North America.

This plot when combined with the plot below shows that the intersecting lines would suggest that the relationship between "JP_Sales" and "Global_Sales" is different at different levels of "NA_Sales", and vice versa. This indicates that the relationship between "JP_Sales" and both "Global_Sales" and "NA_Sales" is moderated by the other variable.

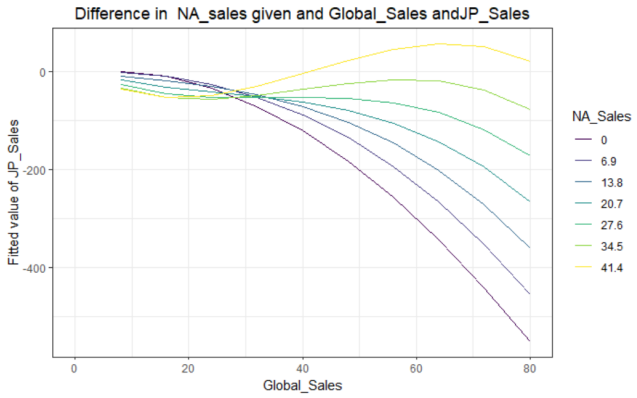

Fig 8.3

8.3 Japan VS Global Sales, given NA sales

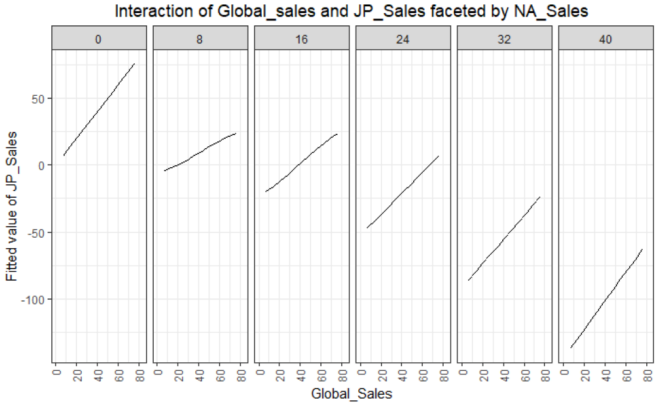

Fig 8.4

The plot shows a downward shift of sales in Japan, indicating lesser and lesser contribution to the Global sales over time. This plot can further be augmented with the plot below showing that the fitted values of "JP_Sales" are not changing systematically with respect to "NA_Sales" for different levels of "Global_Sales".

The parallel lines would suggest that the

relationship between "JP_Sales" and "NA_Sales" is consistent across different levels of "Global_Sales". This may indicate that the "Global_Sales" variable does not have a significant effect on the relationship between "JP_Sales" and "NA_Sales", and the relationship between "JP_Sales" and "NA_Sales" is not moderated by "Global_Sales".

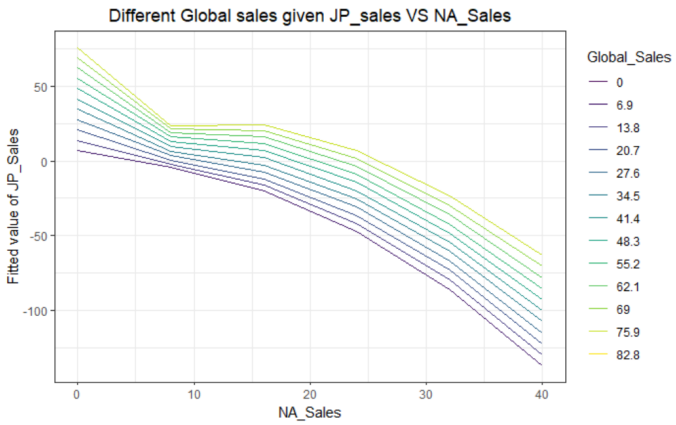

Fig 8.5

8.4 Impact of Critic Score, User Score on Sales

In order to examine the impact of critic score and user score on the sales of video games, we used a linear model.

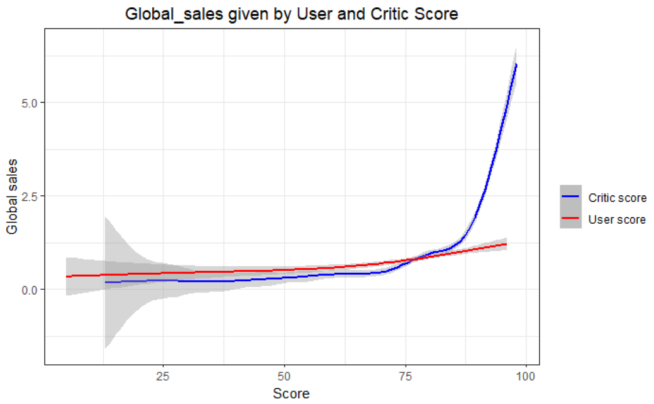

Fig 8.6

Here it can be seen that the global sales are more affected by the Critic score than the user score.

Some reasons for the same could be:

1. Credibility and reputation: Game critics are often seen as experts in their field, and their opinions may carry more weight and credibility than the opinions of the general public.
2. Reach: Critics often have a wider reach than individual users. They may write for popular publications or websites, or have a large social media following, which means their opinions can be seen by a much larger audience than any individual user review.
3. Consistency: Critics often use standardized criteria to evaluate games, and their evaluations are generally more consistent than those of individual users, making the reviews more trustworthy.

However it is interesting to note that while the other regions follow a similar pattern, the Japanese market puts almost an equal amount of emphasis on both, the User score and the Critic score as portrayed by the following graph.

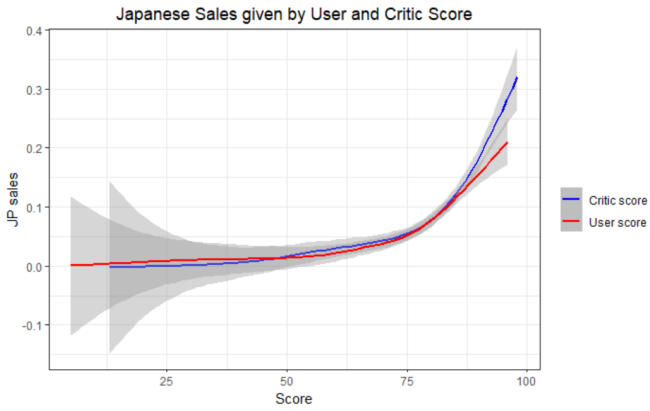

Fig 8.7

Some of the modeling conclusions are as follows: Japan contributed consistently to global sales over the years and was least affected by the sales trends in the other regions. Europe, North

America, and all the other regions followed very similar trends in-game sales over the years. The sales in all the other regions were highly affected by the 'Critic-score' values, however, in Japan. the sales were equally affected by the user as well as the critic score.

10. Conclusion and Future Scope:

We were able to propose a theory for why some Publishers like Nintendo saw a decline in sales while others managed to increase video game sales and become market leaders like Electronic Arts. We found that the shooter genre was the most popular, followed by action and sports, while puzzle, strategy, and adventure were the least popular. Additionally, we observed that global video game sales have been decreasing in recent years, which could be due to several factors, such as fewer data points being collected for later years or sales data not being made open source.

We also investigated the impact of regional sales on global sales and found that Japan contributed consistently to global sales over the years and was least affected by the sales trends in the other regions. On the other hand, Europe, North America, and all the other regions followed very similar trends in-game sales over the years. We discovered that critic scores have a more significant impact on global video game sales than user scores, which could be due to several reasons, such as the credibility and reputation of critics and the wider reach of their opinions.

To sum up, our examination of the video game sales dataset has yielded numerous valuable insights that can guide companies in the industry to make informed decisions based on data. By comprehending these trends and patterns, businesses can stay ahead of the competition and prosper in this rapidly growing market.

For future work, one potential direction could be to analyze the impact of platforms and publishers on video game sales. Another interesting avenue would be to investigate the relationship between video game sales and other factors such as marketing budget, release date, and game ratings. Additionally, further research on recent data could explore how the video game industry is evolving with new technologies such as virtual reality and augmented reality.

11. References –

1. https://www.kaggle.com/datasets/rush4ratio/video-game-sales-with-ratings
2. https://medium.com/@abhijitbose.djf/report-on-video-game-sales-4f774e6f3d89
3. https://jfukuyama.github.io/teaching/stat670/notes/
4. https://www.rdocumentation.org/packages/ggplot2/versions/3.4.2