

**Video Game Data Regression Analysis**

Guled L. Gedi

BIA650: Data Mining for Decision Making

Dr. Oluwafunminiyi Awe

Oct. 27<sup>th</sup>, 2024

## **Abstract**

In order to understand the fall in popularity of the handheld video game device, lineaar regression was carried out to compare the two groups. Using SASStudio, the researchers were unable to provide a clear conclusion due to technological constraints. Providing input on how to improve this for later research, they shared what they had found and presented their visualizations for their shareholders.

## **Introduction**

Companies always need to find answers to 2 things: how to make products that people will buy, and how to learn from products that haven't sold as well in order to make better products in the future. The video game industry is no different to this. In the late 90s and early 2000s, handheld video games were slowly rising in popularity and were a strong part of the success of the industry. However, in the last decade or so, their popularity has declined to the extent that barely any video game companies manufacture new handheld systems or even sell handheld consoles in their retail stores. This led one of the video companies to reach out to the researcher of this current project to try and understand how the eventual obsolescence of the sale of handheld consoles occurred and seeing if home consoles are on a similar trajectory in order to change their marketing schemes for the better.

## **Methods**

The data was collected from the Canvas website under the file name 'VideoGameSales'. It contained 6984 data points, collecting data from 1985 to 2015. In order to analyze the data accordingly, after running the summary statistics, the data was filtered and separated into two datasets: Home and Handheld. The data was then uploaded to SASStudio, and multilinear regression was done on both datasets. Multilinear regression was chosen as it was determined through using scatterplots for each variable to check the linearity condition of a dataset. Out of the 18 variables in the original dataset, only 5 were selected: Global Sales (as the response) & Platform, Global Sales, Critic Score, Marketing\_Priority, and User Score as the

explanatory variables. These were chosen due to their real-world relevance to deciding the success of a video game, as well as being linear enough to be used in the data regression model.

## Results

After running the models, the following regression equations were created for each dataset:

$$\begin{aligned} \text{Global\_Sales\_Home} = & 11.864557 + 0.011972(\text{Critic\_Score}) - \\ & 0.025815(\text{User\_Score}) - 0.182279 (\text{DC}) - 0.089142 (\text{GC}) - 0.169314 (\text{PC}) + 0.503606 (\text{PS}) + \\ & 0.039599 (\text{PS2}) + 0.168081 (\text{PS3}) - 0.135579 (\text{PS4}) + 0.438952 (\text{Wii}) + 0.032208 (\text{WiiU}) + \\ & 0.160304 (\text{X360}) - 0.143102 (\text{XB}) - 12.288418 (\text{Marketing Priority 1}) - 10.812388 (\text{Marketing} \\ & \text{Priority 2}) \end{aligned}$$

$$\begin{aligned} \text{Global\_Sales\_Handheld} = & 13.069762 + 0.008782 (\text{Critic\_Score}) - \\ & 0.006579(\text{User\_Score}) + 0.040228 (3\text{DS}) - 0.174144 (\text{DS}) - 0.011028 (\text{GBA}) + 0.007691 (\text{PSP}) \\ & - 13.439790 (\text{Marketing Priority 0}) - 12.027974 (\text{Marketing Priority 1}) \end{aligned}$$

The Handheld model was quite good, with a R-Squared of 0.7387. However, while running the regression, due to the large number of entries in the Home file, SASStudio was unable to run the analysis. Therefore, the two models cannot statistically compared or be said that there is any difference between the home consoles and the handheld consoles. Looking at their distributions on a scatterplot over time (see Appendix), however, it can be seen that Home consoles sell more quantity wise and are more constant in their data spread, as opposed to Handheld devices.

## **Discussion**

While the information gathered here is a step in the right direction, due to data restraints, it remains to be seen how significantly different the two categories of gaming console are to each other. Several suggestions could be made to produce more precise results in the future. Firstly, it would be to include data after 2015, as with the whole picture of the wane of the handheld devices, it will be more accurate to draw conclusions and then to apply those findings to home devices. Secondly, it would be to include more consoles in the Handheld section, as seen above, since in the data provided, there were 11 home consoles and only 4 handheld consoles. That might lead to misrepresentation of the data. One final point would be to include more variables that would help paint a bigger picture of the video game industry, since some of the variables are severely more significant to the Global\_Sales than others.

## Appendix

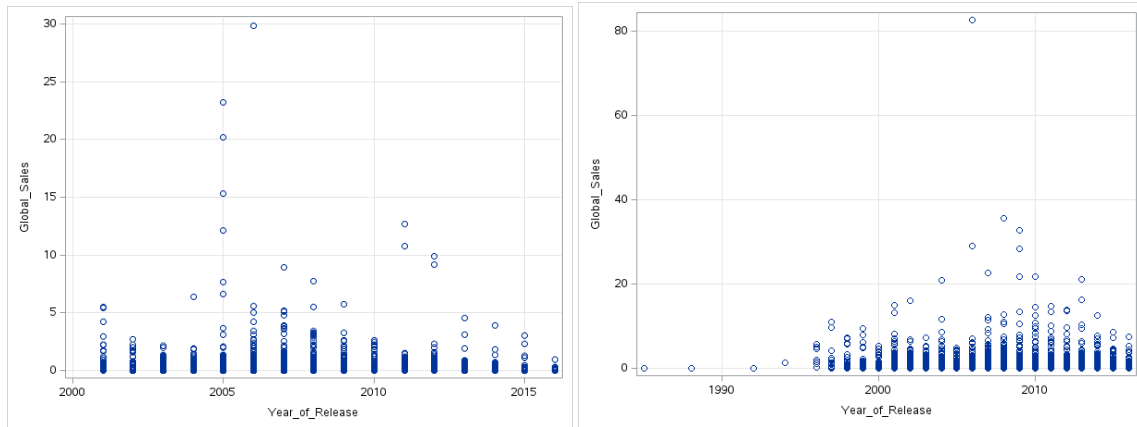


Fig. 1: Scatterplots representing the spread of sales of handheld (left) and home (right) consoles

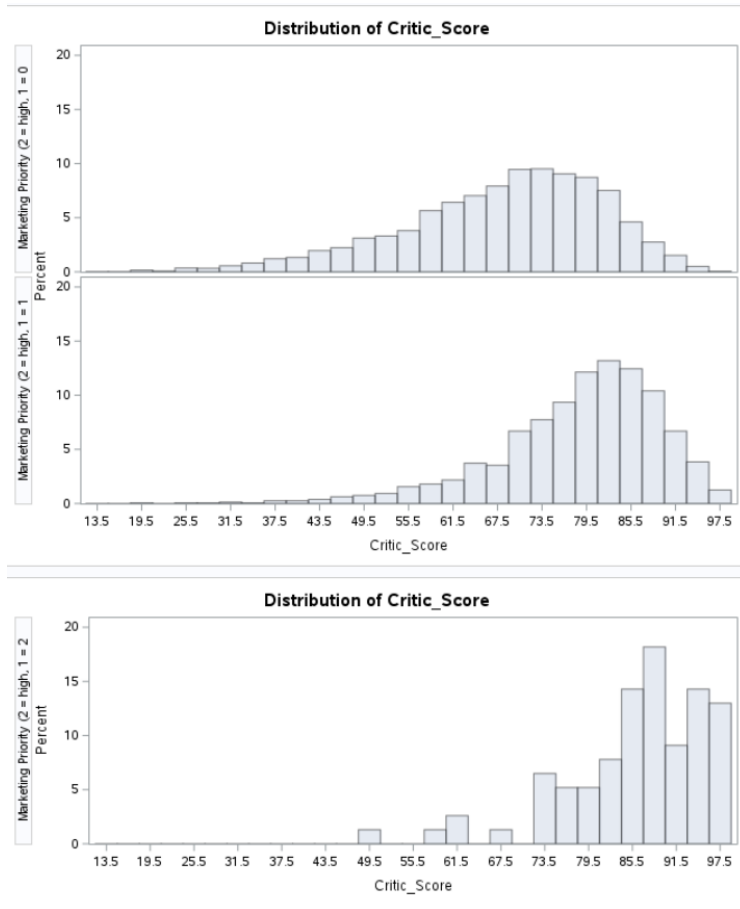


Fig. 2: Histogram displaying Critic\_Scores per Marketing Priority. As seen here, all three histograms appear to be left-skewed, which indicates that the mean is less than the median. It also indicates that the higher marketing priority a game is given, the higher score it shall receive, which is indicated in the following figure. This leads to expecting greater sales as well.

