Turtle Games

Leveraging Customer Trends for Improved Sales Performance

Matteo Natalini

July 2023

Introduction

Welcome to the Turtle Games report, a leading global game manufacturer and retailer. This report will cover key objectives that Turtle Games wants to explore, including customer loyalty point accumulation, market segmentation through customer groups, utilization of customer reviews for marketing campaigns, the impact each product has on sales, and the relationship between North American, European, and global sales.

Analytical approach

Turtle Games provided 2 datasets: one containing customers' details and reviews and the other with details of video games sales.

Using Python's *statsmodels* library, we analysed the correlation between customers' loyalty points, age, remuneration, and spending score. Our findings showed that loyalty points have a strong positive correlation with spending score and remuneration, but a slight negative correlation with age.

We created two OLS models, one for each variable. Both models had a good fit as over 40% of the observed variation can be explained by the inputs.

To enhance our model, we employed k-means clustering to classify customers into groups based on their spending score and remuneration. The analysis identified 5 distinct clusters, which were validated by the Elbow and Silhouette methods.

We used natural language processing to analyse customer reviews and summaries. After removing duplicates, we had 1351 unique entries to analyse the 15 most frequent words and overall sentiment. For each review, we utilized the tokenization approach to generate a list of lowercase words.

In order to eliminate insignificant words, we utilized a list of stop-words to filter our values. After careful consideration, we added "game" to the list as it appeared frequently and was irrelevant to the analysis. Then, we determined the frequency of the most commonly used words in the reviews.

Using the *textblob* function, we could assign polarity and subjectivity scores to the words used. This allowed us to compile a list of the most negative and positive reviews.

We used R Studio to analyse the sales data. Using *tidyverse,* we cleaned and verified the data frame. We removed unnecessary columns and grouped the data by product ID to determine the total sales. To gain insights into the impact of products on sales, we created advanced visualizations using the *ggplot* library.

Our focus then turned to the correlation between European, North American, and global sales. We built a multiple linear regression model with the *psych* library to predict global sales based on the other two variables. The data were not normally distributed, as the QQ plots and the Shapiro-Wilk test confirmed, so further investigation is recommended.

Visualisations and insights

We used linear regression to predict loyalty points accumulation with *remuneration* and *spending score*.

Looking at the scatterplot, we can observe that the red line representing the predicted values closely aligns with the actual values in the first half, particularly when the spending score ranges from 40 to 60. However, after a score of 60, the correlation becomes almost nonexistent (fig.1).

Figure 1

A similar scenario is observed for *remuneration*, with a statistically significant positive relationship between the two variables for remuneration below £58,000 (fig.2)

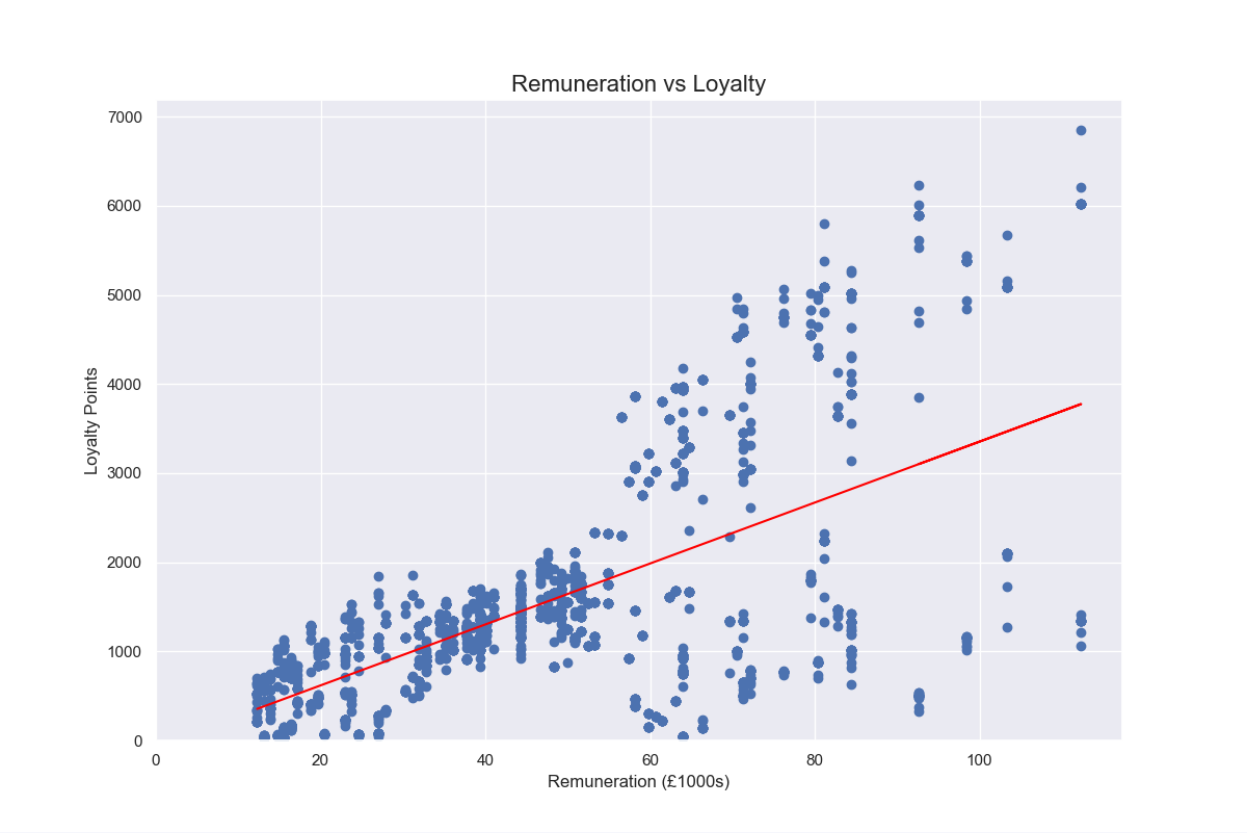

Figure 2

Our models have shown that using *remuneration* and *spending scores* can be useful in predicting the accumulation of *loyalty points*. However, this prediction can only be accurate if the customer belongs to a certain group. Clustering can assist us in identifying these groups.

Figure 3 shows 5 groups. There are two separate clusters formed by customers who earn less than 30k or over 60k. One group has a spending score of no more than 40, while the other ranges from 60 to 100.

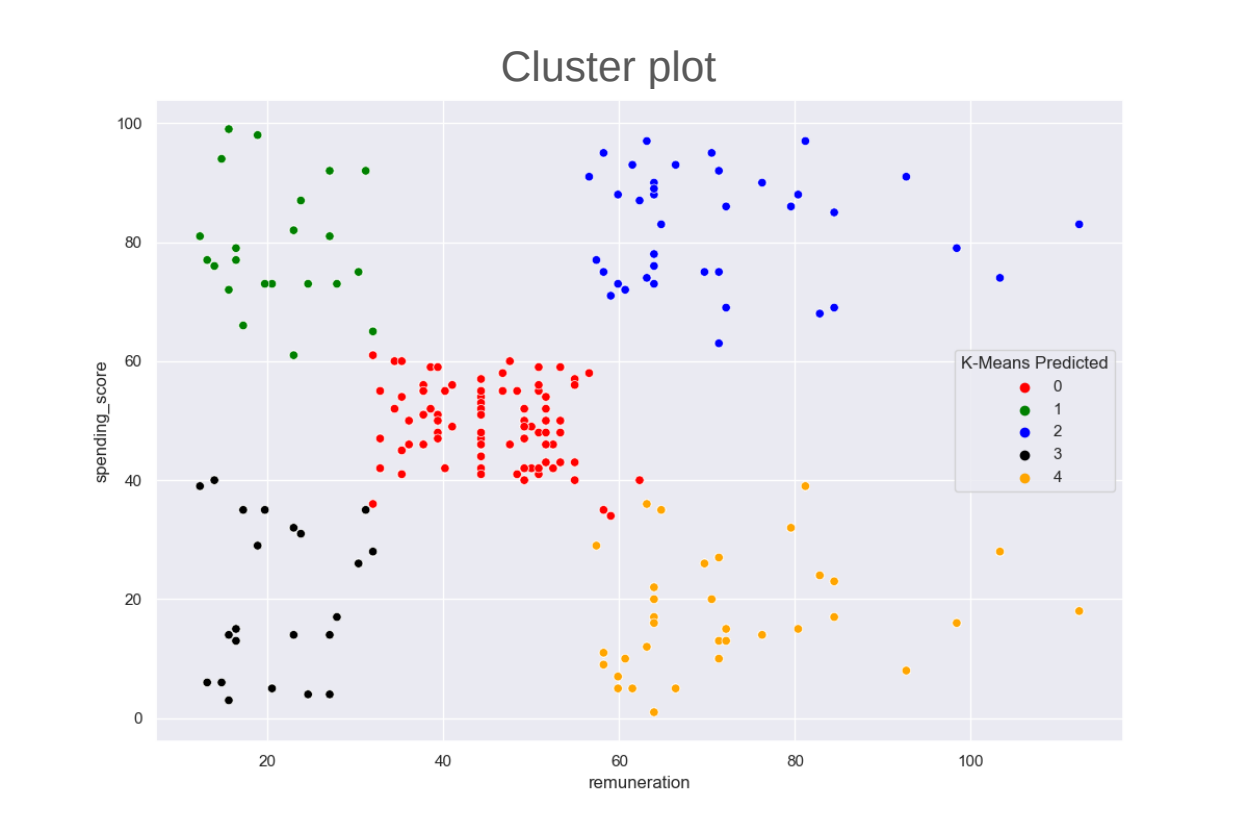

Figure 3

Our task was to uncover how customer reviews can influence marketing campaigns. By utilizing *NLP*, we have compiled a collection of frequently used words and determined whether they convey a positive or negative sentiment. We found that the top 15 words used frequently[1] were mostly positivity (fig.4, fig.5).

Figure 8 shows the sentiment score for all reviews and highlights that there is a large number of neutral or mildly positive comments. but overall the distribution is more oriented toward positivity.

[1] We opted to utilise the most frequently used words in the summaries as they held more significance.

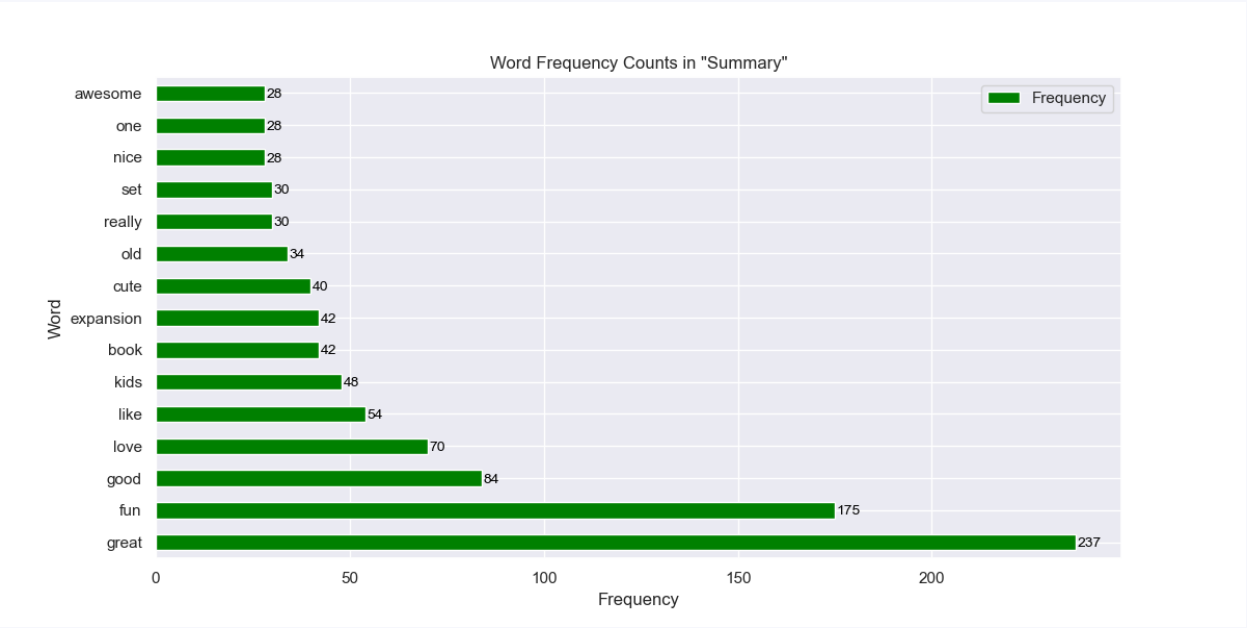

Figure 4

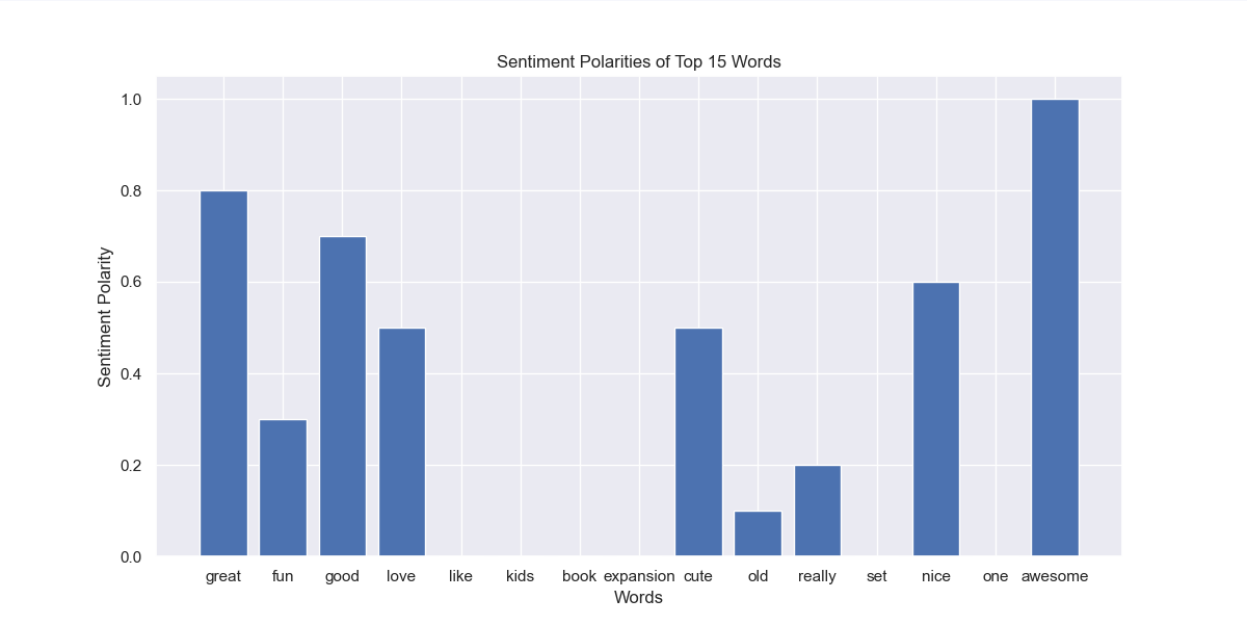

Figure 5

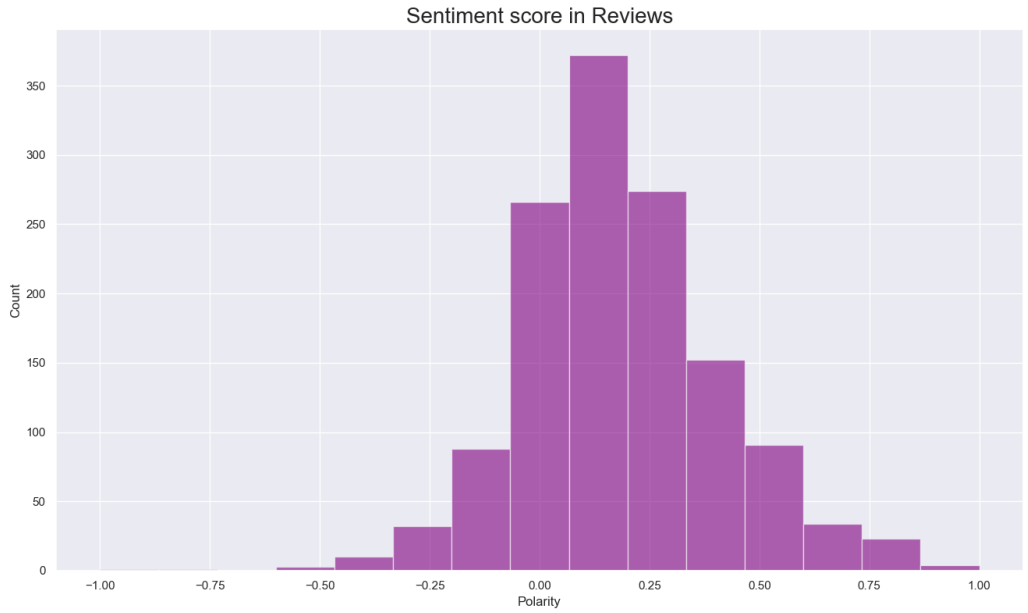

Figure 6

Let's take a closer look at our *sales data* and see how products impact global sales. In Figure 7, we have a boxplot that displays the distribution of sales across platforms worldwide. We noticed an outlier in the Wii box plot, and upon further investigation, we discovered that it's a sports game from 2006 that has sold exceptionally well compared to other games. That's definitely a bestseller!

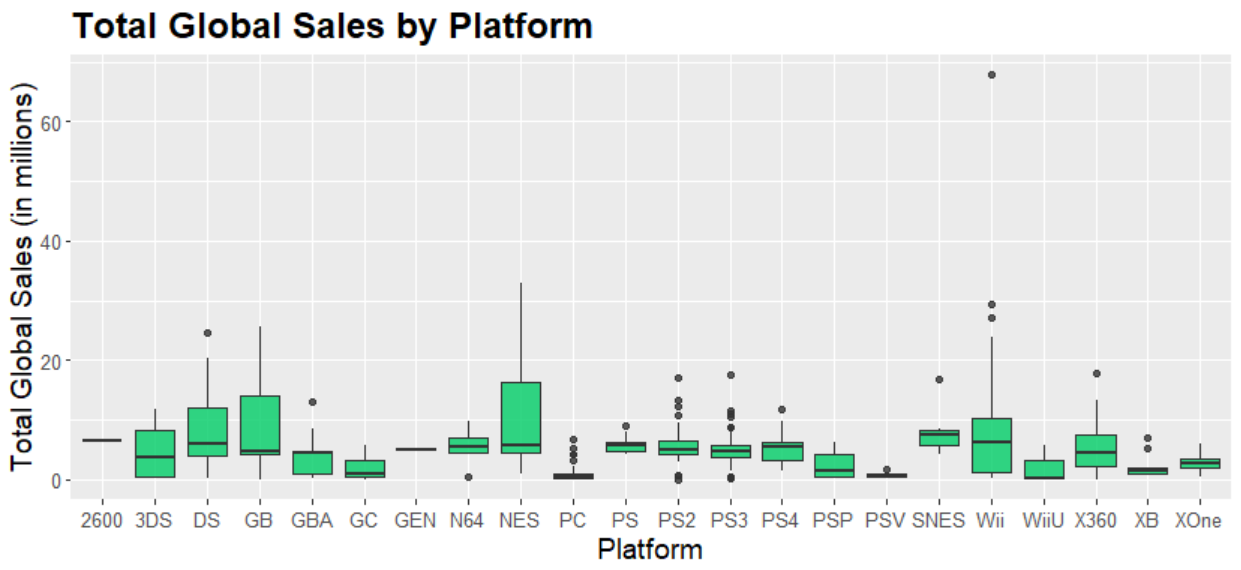

Figure 7

Data 1982 - 2016

Figure 8 displays global sales grouped by genre, with each platform represented by a different colour. It is evident that certain genres are more popular on specific platforms. For instance, PS platforms are preferred for action games, and equally preferred with X360 in shooter games, while the Wii dominates in sports.

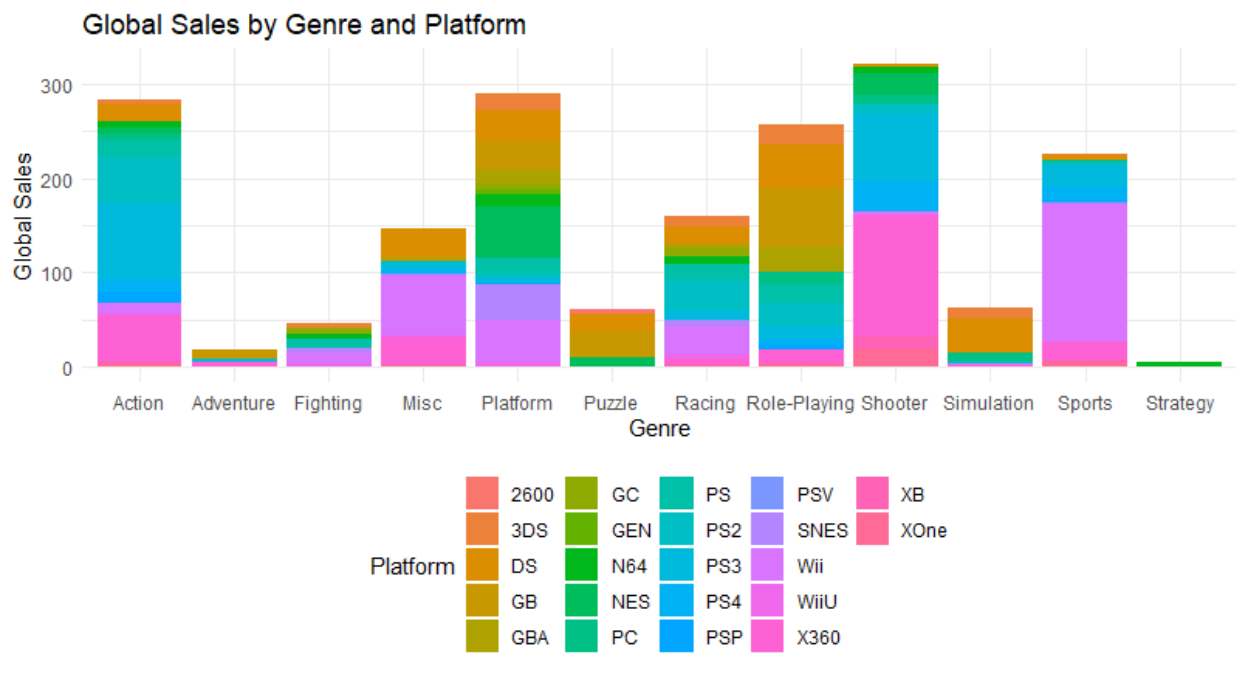

Figure 8

In Figure 9, we can see a comparison of platform sales in different markets. It is noteworthy that the Wii is the best-seller globally and in Europe, while in North America, it is X360 that takes the lead. However, it is not as popular in Europe as it is in North America.

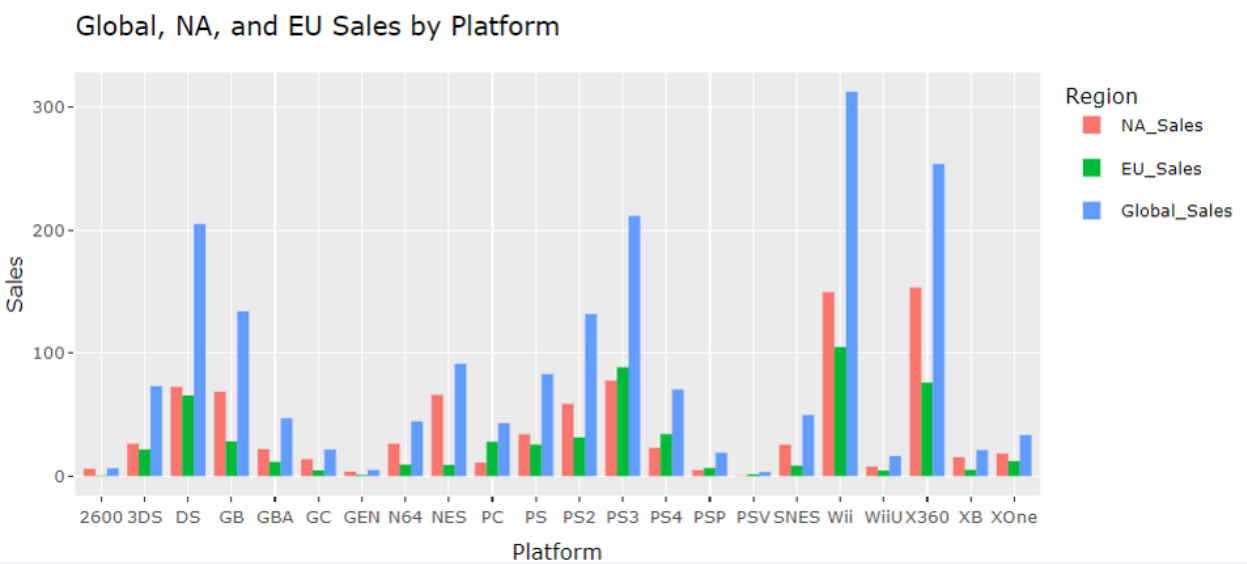

Figure 9

In conclusion, we examined the regression models created using *global sales*, in correlation with *European* and *North American sales*. Both graphs indicate a strong, positive correlation between these variables, indicating that they can be utilized to construct a robust model for predicting future global sales (Fig. 10 and 11).

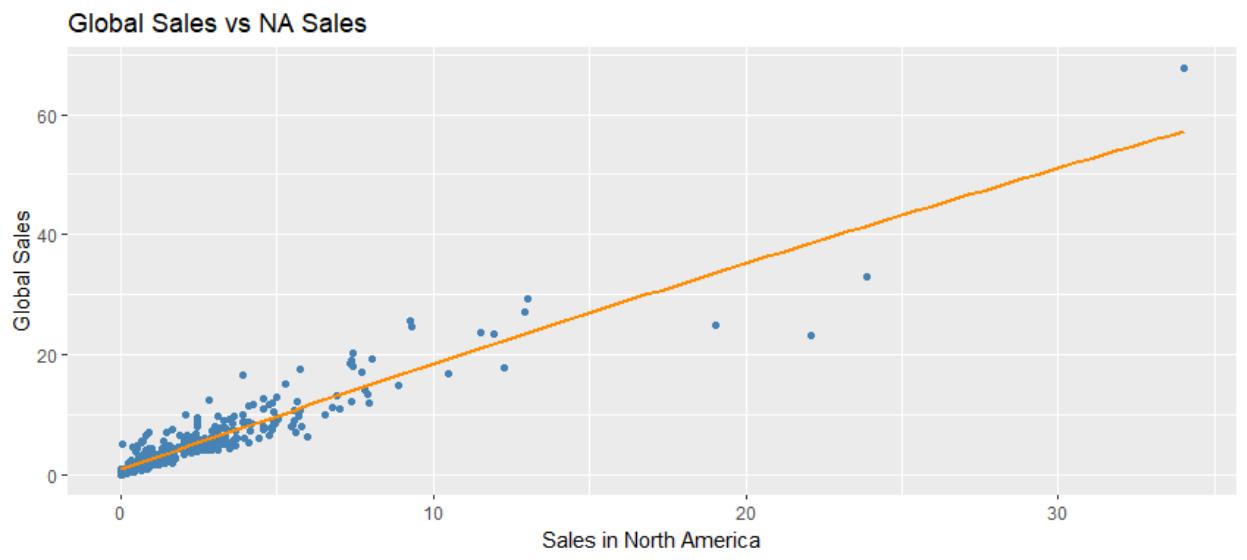

Figure 10

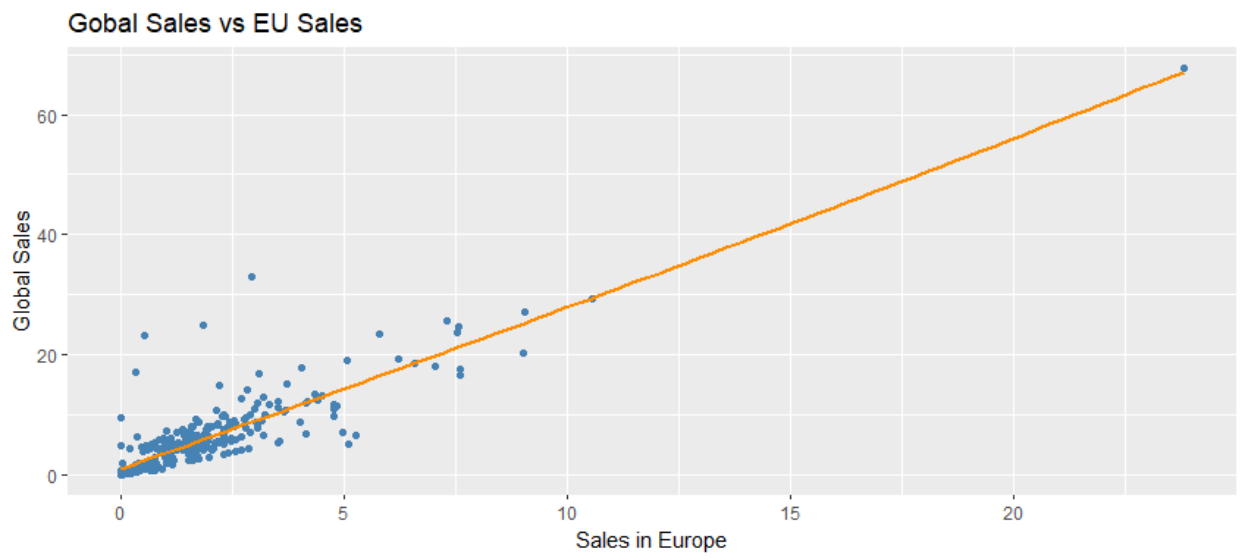
Gobal Sales vs EU Sales
Global Sales
Sales in Europe
0
20
40
60
0
5
10
15
20

Figure 11

Patterns and predictions

We found a correlation between customers' loyalty points, remuneration, and spending score. We identified various clusters based on these characteristics, allowing for the creation of a stronger model to predict future loyalty point accumulation. Further research can help refine this model.

Based on our analysis of commonly used words and their polarity scores, we have discovered some significant features that can be leveraged in future marketing campaigns. The frequent use of "fun" indicates the importance of highlighting the entertainment value of games. "Kids" and "cute" imply that the games are popular among families and children. Emphasizing the child-friendly aspects of games can attract parents who seek enjoyable and educational activities for their kids. "Expansion" suggests that customers are interested in building on their existing collections. Providing expansions for popular games can increase revenue by enticing customers to purchase more. Additionally, highlighting classic games or offering retro-themed merchandise can appeal to customers seeking a nostalgic experience, as indicated by the word "old".

After analysing sales data, we have found that the Wii Sport is the best-selling product compared to other items on the market. We have also determined which platform is most popular for the top genres and have identified the best-selling platforms in different regional markets. This information can guide marketing decisions in various geographical areas.

We have discovered a strong correlation between sales in Europe, North America, and globally. Additionally, we have developed and successfully tested a robust model that enables us to predict future global sales based on input from the other two regional sales.