Analysis on Customer Base and Sales Performance

Sophia Velicer

Turtle Games

Introduction

Turtle Games is an international game manufacturer and retailer, selling both their own brand products and other brand product. The products include a range of platforms such as books, board games, video games and toys. To help Turtle Games optimize sales performance our team seeks to better understanding their customer base, integrating customer feedback into the marketing process and understanding current sales performance and patterns through data analysis. We will initially answer the following questions:

- How do customers accumulate loyalty points?
- How can groups within the customer based be used to target specific market segment?
- What is the current customer opinion and how can it be used to inform marketing strategies?
- How does each product perform in sales?
- How reliable is the data?
- Can we make prediction on global sales?

Approach

The data was good, but needed some tidying before analysis could begin. For both the sales and review data sets I ensured there were no duplicate values, made sure that column names were descriptive, checked for outliers, and created subsets of the data set that excluded unnecessary information. There was one large outlier in the sales data (product #107), however the entry made sense and followed the general trend of the rest of the data, so it was included in the analysis. A number of different processes were used in analyzing the data. An overview on the approach for each topic is below:

Loyalty Points

- The variance inflation factor was calculated, which indicated a high level of multicollinearity between variable.
- A regression model was fitted to a training set of data and then tested with a test data set.
- Different variables were used to make a linear regression model with loyalty points and then the R-squared values were compared.
- Spending score was the best indicator with an R squared value of 0.43.

Customer Grouping

- I used the sklearn.cluster package to perform a K-means clustering analysis to group customers based on their spending score and remuneration.
- Five distinct clusters were formed with the majority of customers falling in the middle of both categories.

Customer Review Analysis

- I used TextBlob and nltk to make a word cloud, determine the frequency of certain words, and analyse the polarity of reviews and summaries of the reviews made about Turtle Games.
- When plotting the polarity of reviews and review summaries, most were neutral or positive.
- A discrepancy appeared between the polarity of the reviews and the review summaries with more summaries appearing to be just neutral.

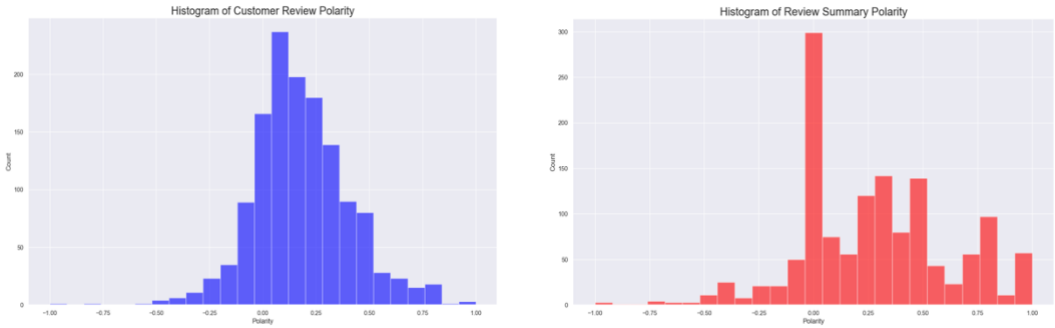

Figure 1 Comparison of polarity between customer review and review summary.

Product Sales Performance

- The data was first subsetted to only include the sales data to streamline analysis.
- Explorative plots were made using the quick plot function.
- A scatterplot was best at describing the sales of individual products.
- Product 107 was a notable outlier in all regions, but was a valid entry.
- To compare sales between North America, Europe and the globe I grouped the data by product and then reshaped the data using the melt function before plotting.

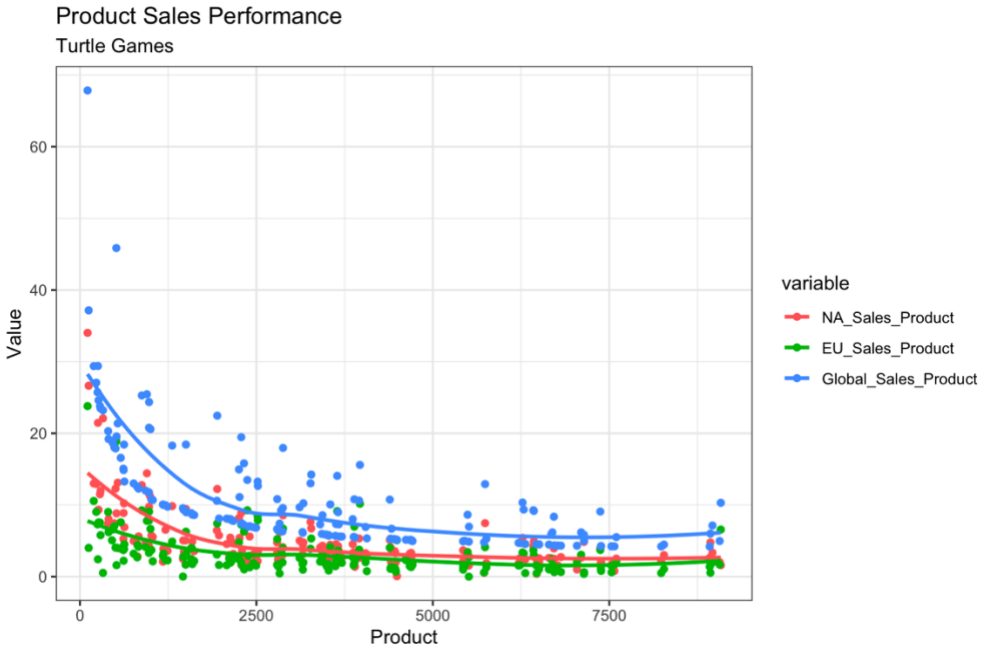

Figure 2 Product sales performance. All regions were plotted to see the different in relationship between value and product number.

Data Reliability

- The normality was tested with a qqnorm plot and a qqline plot and then confirmed with a Shapiro-Wilks test for each region (North America, Europe and Global)
- None of the regions showed a normal distribution.
- Kurtosis and skewness were then measured for each region.

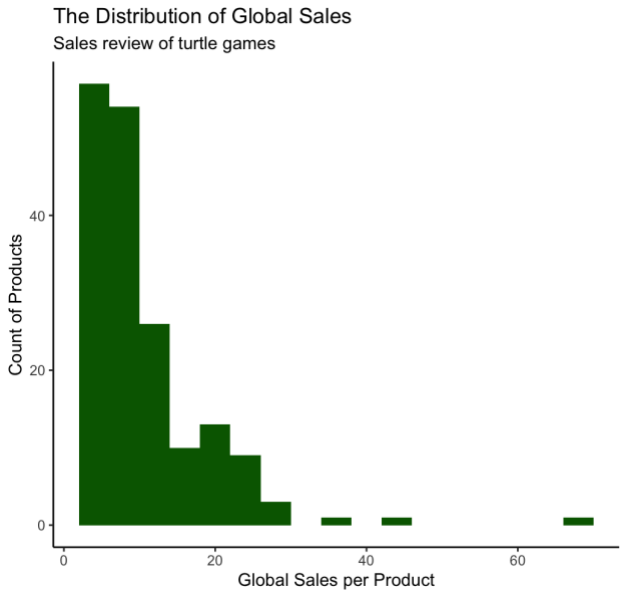

Figure 3 The distribution of global sales with a visibly high kurtosis and positive skew.

Predicting Global Sales

- Created several linear regression models using the lm function to fit a model to predict global sales.
- Chose the one that had the best adjusted R-squared value as our prediction model.

Patterns and Gaps

First, we tried to understand the customer base. While spending score was the best indicator of loyalty points, the R-squared value indicated a very inaccurate model. There is no definitive factor for loyalty points and a multi regression model would have been a better fit. Further data would also have been helpful such as recency of purchase and purchase history. Based on segmenting the groups based on income, customers were more likely to spend more or less if they either had lower or higher remuneration. Those with a middling income spend a middling amount at Turtle Games. The overall sentiment for Turtle Games is mildly positive, however the summaries reflected a different distribution, though still positively skewed.

Next, we analysed the current sales performances. There is a relationship of exponential decay between product number and sales, so the higher the number, generally the lower the sales. This may be due to lower numbered products coming to market earlier and having more time on the shelves. Knowing when this data was collected would have been helpful in the analysis. The analysis also showed that Wii, X360, P3 and DS were the preferred platforms. For all the regions, the data is not normally distributed, has a relatively high kurtosis and is positively skewed. This shows that many products perform in the lower range, which a few reaching exceptional sales.

To predict global sales, we determined the correlation between the regions which is very high. Both the North American and European sales are very significant in determining global sales and out of the models that we tested having both variable feed into a prediction model produced a model with an adjusted R-squared value of 0.96 (very accurate).

Recommendations

We would recommend Turtle Games to conduct a survey to investigate the motivation behind their customers and how much of a barrier price is to further understand the customer base. I believe further investigation into negative or neutral reviews of the company would help us understand where we can improve. Turtle Games should focus on the platforms that are most popular with its users and develop games that are compatible with them. Due the strong predictive power of North American and European sales it is possible for Turtle Games to beta test their games in either region before releasing it to the rest of the market.