Course 3 Assignment Report

TURTLE GAMES

Improving Overall Sales Performance

Benedict McEvoy | April 2023

INTRODUCTION

Turtle Games is a manufacturer and retailer of games with a global customer base. They are interested in improving their overall sales performance. In order to enhance sales, we have analysed sales data and customer reviews so the marketing team can formulate a more effective marketing strategy. The questions we have focused on answering are:

1. What are customers saying about Turtle Games?
2. Who should we prioritise marketing efforts for?
3. Should we have different marketing strategies for different types of customers?
4. Are there certain products we should be promoting?
5. Which regions impact global sales the most?

For this analysis we have utilised Python to analyse the customer reviews data, and R to analyse the sales data.

1. What are customers saying about Turtle Games?

Using Python, our first step was to prepare customer reviews and customer review summaries for natural language processing. This involved changing the text to lowercase, removing punctuation, and then removing duplicate reviews and summaries.

Next, we created WordCloud images displaying the most used words in the reviews and summaries. For this, we made sure the reviews and summaries only contained alphanumeric characters and went about removing English stop words. Knowing the most used words was informative and interesting, but it did not tell us about customer opinions.

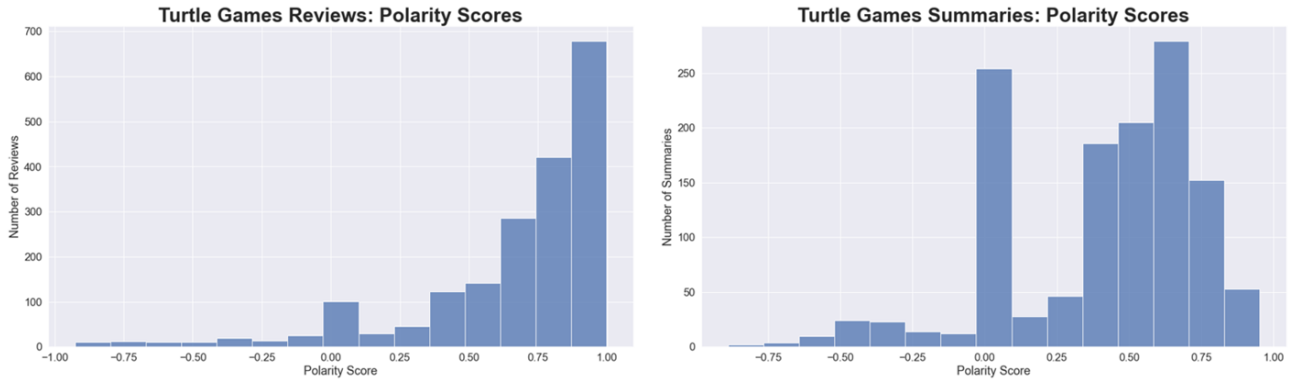

Figure 1: *Polarity Scores for Customer Reviews and Customer Review Summaries*

We then carried out sentiment analysis and determined polarity and subjectivity scores for each review and summary. The histograms shown in Figure 1 display the distribution of polarity scores. A polarity score of 1 indicates a positive sentiment, -1 indicates a negative sentiment, and 0 indicates a neutral one. Many reviews had a polarity score greater than 0 suggesting customers felt positive about Turtle Games. However, many customers displayed a neutral sentiment with a polarity score around zero. This could be improved by offering certain customers promotions etc.

2. Who should we prioritise marketing efforts for?

Turtle Games should prioritise its marketing efforts on loyal customers. These are customers who generate repeat business and increase overall sales in the long term. Being able to predict the loyalty of a customer would be a useful tool as it could identify whom to prioritise marketing efforts on.

Using loyalty points as an indicator of customer loyalty, we investigated what variables are significant in predicting the number of loyalty points a customer has. For this, we built various linear regression models and decided on the best one to predict loyalty points. The model that used spending score, remuneration, and age was the best as it explained 83.5% of the variability in the number of loyalty points a customer has. It also gave the most accurate predictions when looking at the mean absolute error values.

Checking the assumptions of linear regression for this model, we found that there were no signs of multicollinearity between the variables when looking at the variance inflation factors. Unfortunately, however, there were signs of heteroscedasticity when carrying out the Breusch-Pagan test on the residuals of the model. This is a limitation of our model.

3. Should we have different marketing strategies for different types of customers?

Having one type of marketing strategy might not be the optimal way to increase sales as this might not suit all types of customers. Therefore, identifying groups within the customer base and tailoring the marketing strategy to best suit each type would be more effective at increasing sales.

To find groups within the customer base we looked at remuneration and spending score. Plotting these variables in a scatterplot, we noticed that there seemed to be five clusters within the data. To confirm the number of clusters in the data we carried out a k-means clustering analysis. Plotting elbow and silhouette charts confirmed that five clusters made the most sense.

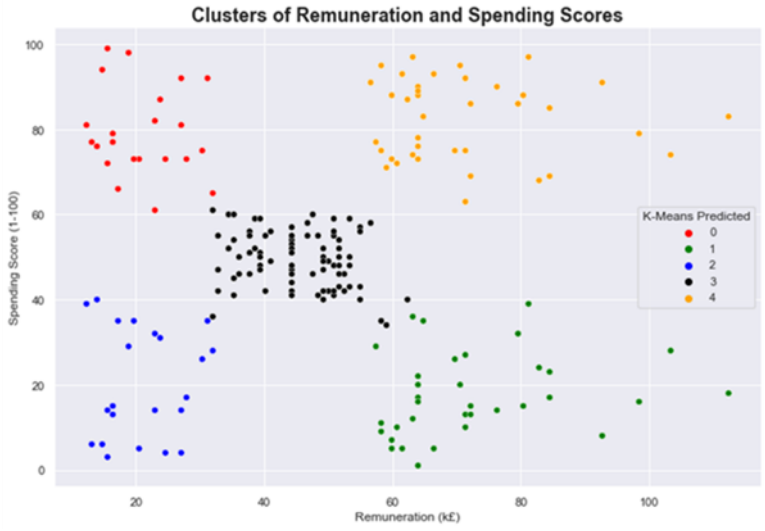

Figure 2: *Groups within the Customer Base*

These five groups can be seen in Figure 2. They correspond to customer groups with different levels of remuneration and spending scores. An example of applying a tailored marketing strategy to one of these groups could be marketing more expensive products to group 4 who are customers with higher remuneration and spending scores.

4. Are there certain products we should be promoting?

To analyse sales data, we utilised R. There were 175 products in the dataset. We grouped the dataset by product code and summed the sales for each one. The dataset contained sales information for North America and Europe, as well as global sales.

Using the ggplot package, we created bar charts to display the top six products that generated the most sales for each region. These can be seen in Figure 3. The bars that are highlighted in green are the products that feature in the top six for all three regions. They are product codes 107 and 515. I would suggest focusing marketing efforts on promoting these products to see an increase in overall sales.

Figure 3: *Top Selling Products*

It should be noted, however, that there is variation in the top-selling products for the different regions. Investigating these and making a marketing strategy more suited to each individual area, could be an option for future analysis.

5. Which regions impact global sales the most?

To answer this question, we created linear regression models to predict global sales. One of these models used North American sales to do this, while the other used European sales. To create them, we first removed the outliers from North American, European, and global sales. We then checked the linearity of the relationships between these variables by plotting scatterplots. This was generally confirmed. We then built the models which can be seen in Figure 4.

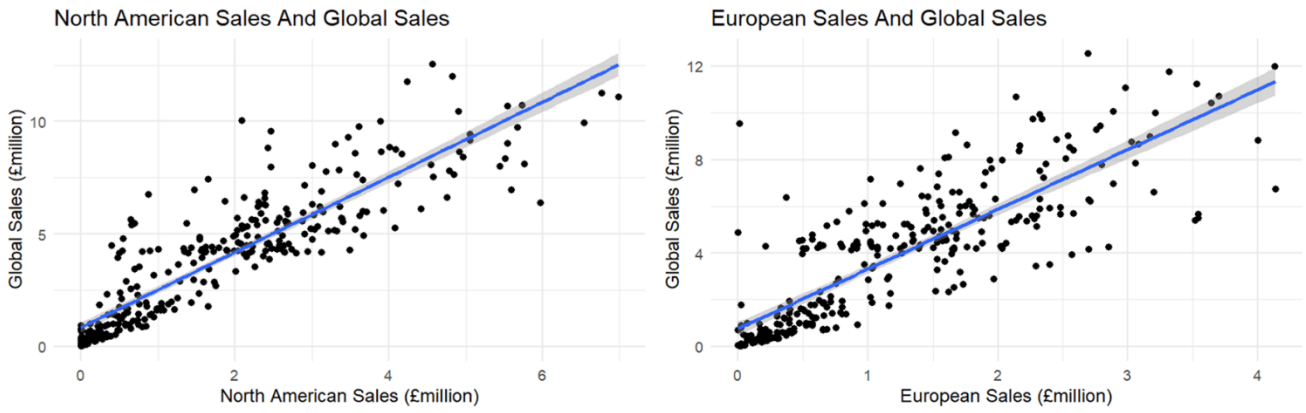

Figure 4: *Linear Regression Models to Predict Global Sales*

In both models, the variables were highly significant in predicting global sales. The model using North American sales was able to explain 79.4% of the variability in global sales whereas the model using European sales was able to explain 69%. Therefore, North American sales are better at predicting global sales. Having said this, according to the models, an increase in European sales could increase global sales more dramatically than North American sales could. It should also be noted that both of these models presented limitations in meeting the assumptions of linear regression, which hinders their reliability.

In my opinion, I would prioritise increasing sales in North America to result in a more reliable increase in overall sales. I would like to investigate other regions to get a better idea of this though.

CONCLUSION

To increase overall sales, the marketing team should consider the following:

- Many customers showed a positive sentiment towards Turtle Games. However, many showed a neutral one. This could be improved by reaching out to the right customers with effective marketing, promotional offers, and so on.

- Gaining and keeping loyal customers should be a priority for Turtle Games. A model has been built where spending score, remuneration and age can predict the number of loyalty points a customer will have, and therefore their potential loyalty. Identifying and nurturing these customers could improve overall sales in the long term.

- There are five groups within the customer base in terms of remuneration and spending score. Tailoring a marketing strategy for these different groups would be more effective than a one-size-fits-all approach.

- Product codes 107 and 515 stood out as generating the most sales. Focus marketing efforts on promoting these products.

- North American sales could explain 79.4% of the variability in global sales. Increasing sales in North America could result in an increase in overall sales. Focusing marketing efforts on this region could be a good option.