



Predicting future outcomes Turtle Games

1. Background

Turtle Games makes and buys games and toy products to sell worldwide.

The problem statement I developed is:

Turtle Games wants to better understand their customers to improve their marketing and sales.

The analysis will be of data from sales as well as customer reviews. I will consider the initial questions Turtle Games has proposed and in answering them I will help them to consider what steps can be taken to meet the business objective of:

“Improving overall sales performance by utilising customer trends”.

Through utilising data to

- Best segment their customer base,
- Undertake improvements based on feedback and,
- Inform what will drive more sales.

2. Analytical approach 350

The principles I adopted were to:

- i. Make the code as reproducible as possible, so colleagues could check and use it. Mainly by using functions so the code was modular and easy to read.
- ii. Use markdown and comments to provide a helpful explanation on my thought processes and approach, whilst using PEP 8 guidelines.
- iii. Follow design principles with visuals to ensure they could be understood clearly and used in presenting insights to the business.
- iv. Minimise errors by accepting where possible.

My approach to the analysis:

- I assessed the data and prepared it for analysis, including checking for null values, reviewing summaries of the data such as types and number of values.
I also used functions to assess the quantitative data with the describe function. These checks assured a high level of data reliability, given the number of duplicates and null values.
- I examined the data for missing values. The meta data supplied indicated this was a complete data set however given the latest year is 2016 this calls into question the validity as the data from the last c.7 years is missing.
- I decided to include outliers as they represented an important part of the feedback and the narrative from my analysis.
- I mainly used two tools python and R.
- For data exploration I summarised and performed some analysis that could be used in describing the context and challenges.



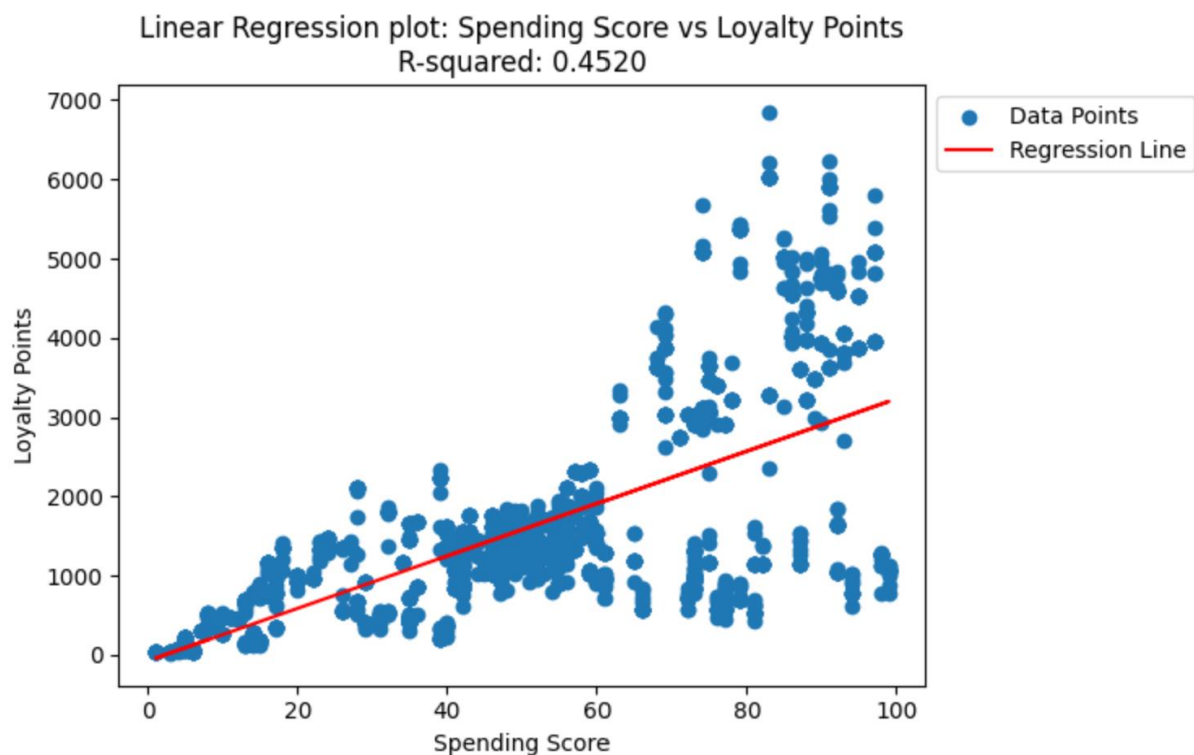
- I plotted some summary graphs to understand the distribution of data.
- For predictive modelling I used specific methods, tools and libraries to answer specific questions e.g., moments for the Shapiro-Wilk test in R and nltk for the natural language processing analysis in Python.
- I investigated the accuracy of models and tested them, for instance for machine learning I read the customer reviews to see if I agreed with the model.
- (There were reviews that I flagged to the business given the risk of some negative comments).
- For clustering I used the silhouette and elbow plus an additional library to develop, fit, and then improve the quality of the model.

3. Visualisations and insights into the six questions

1.

Rationale: To show the strongest correlation an R^2 figure of 0.45, a moderate level, was for spending score and loyalty points. This indicates that there is predictive power, but other variables should be considered.

Interpretation: Reasonable weight should be given to the fact loyalty points are accumulated by spend.



2.

Rationale: Identified customer groups within the customer base that can be targeted for marketing purposes. Clustering is unsupervised so accuracy checks could not be undertaken.



Interpretation: I developed profiles of the clusters based on their characteristics. As I suggest these can be used for tailored campaigns to their profile.



3.

Rationale: I developed word clouds from reviews and identified the top repeating words as well as undertaking sentiment analysis.

Interpretation: There is a need to monitor this as there is a reputational risk for TurtleGames. Plus, a positive review like "Great game!"⁷²³ could be a testimonial.

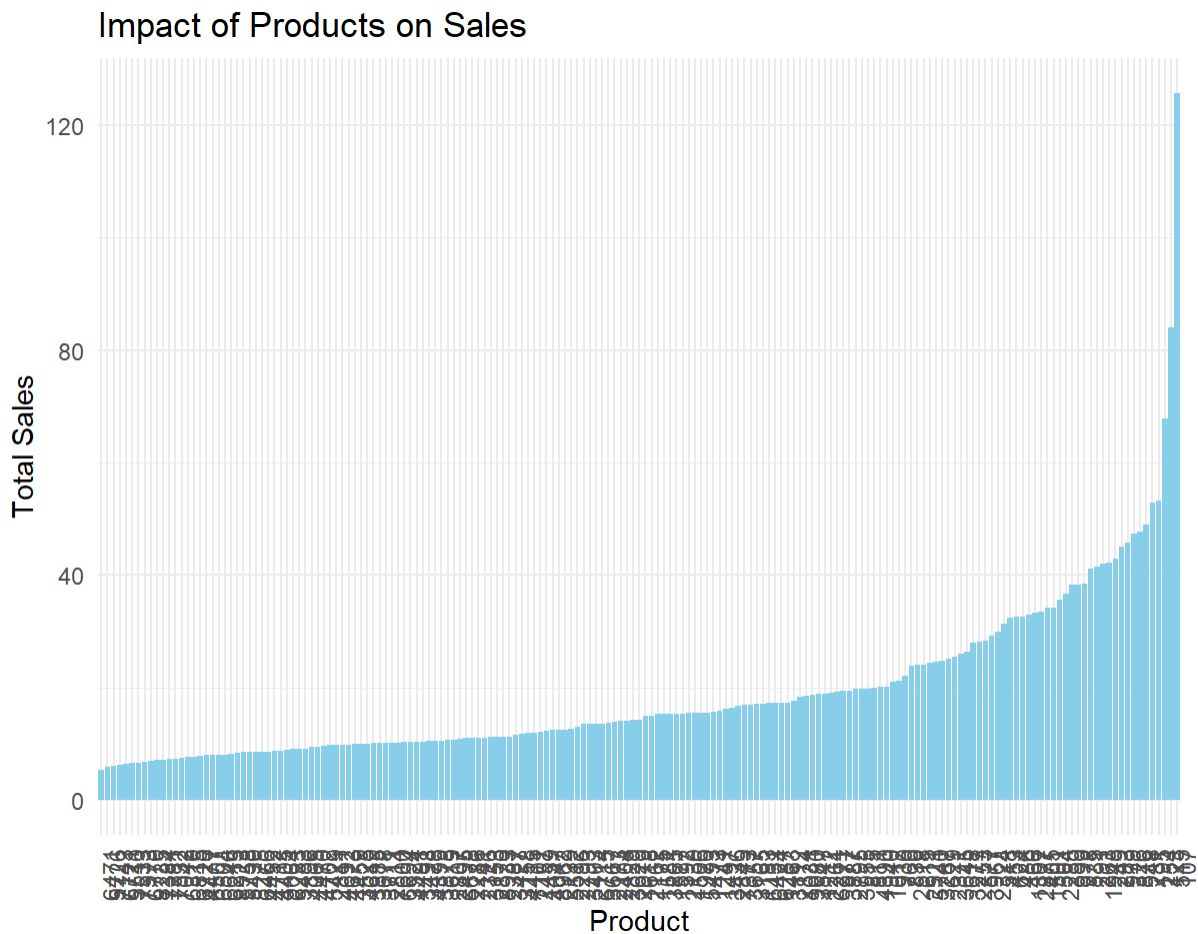


4.



Rationale: Due to a small sample size I focused on relative comparisons.

Interpretation: 19 products with the most impact, sales above £27m which account for 5.7% of products and 25% of sales, a mean impact of £44m.



5.

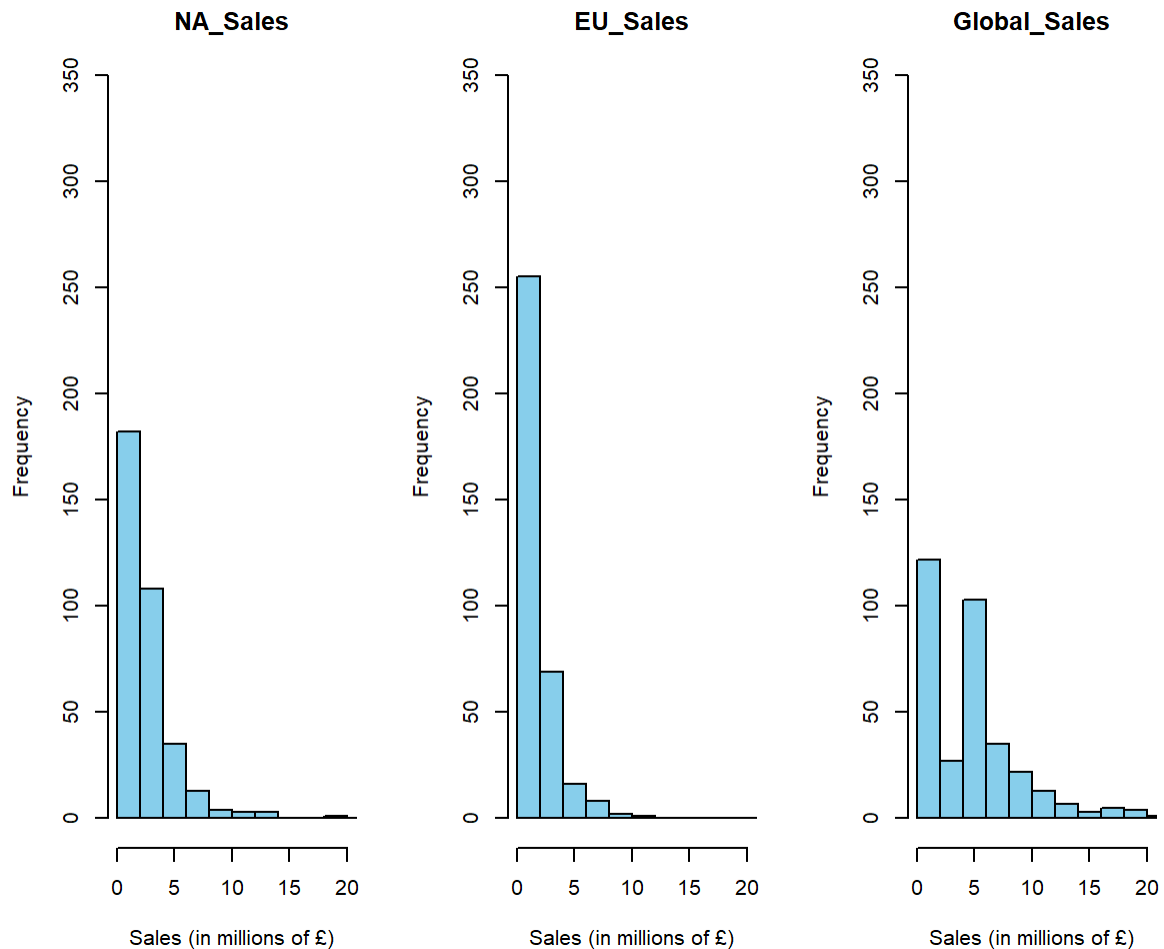
Rationale: Analysis of kurtosis and skewness values, performed the Shapiro Wilk test, created a linear regression model and assessed the correlation coefficient.

Interpretation: There are more outliers compared to normal distribution and the data set is positively skewed, a long tail on the right. The distribution is leptokurtic.

6.

Rationale: Show correlation and sales under £20m have a higher frequency. A strong positive correlation suggests American, European sales closely relate to global sales; knowing they both contribute to global performance.

Interpretation: My hypothesis is that the higher frequency in the EU market likely has a fragmented market due to a higher number of languages.



4. Patterns and predictions into the six questions

1. There is not only one way loyalty points happens, TurtleGames should observe and act on the basis that other variables like remuneration (R^2 of 0.38) are connected to how customers accumulate loyalty points. To increase sales through targeting customer loyalty A/B testing could be used to assess effectiveness of initiatives.
2. There are clear clusters identified and I reported profiles which should be used to personalise marketing that will resonate and inform marketing approaches. This should include revisiting the analysis as behaviours and preferences will change over time.
3. The word clouds and sentiment analysis showed a much higher frequency of positive sentiment. This should be monitored given the risks to reputation which if negative may impact sales accordingly.
4. There is a group of small but high impact products, prioritising these products will help to optimise marketing efforts. Opportunities to build and expand on them in other regions may help increase sales.



5. There are deviations from normality with some significant outliers which are generally the high impact products. They should be recognised and analysed separately given their significance to impact sales.

6. The strong correlation should be recognised (AppPlot1) and marketing efforts should be based on tailoring to a regions' characteristics which will help increase sales. The consideration of platform weighting across sales may benefit sales (AppPlot2&3)

5. Recommendations

There were limitations in this process, notably, there was a limited amount of time series sales data which would help focus and drill into detailed predictions. This value would provide more meaningful and current recommendations on how to improve and better utilise customer trends.

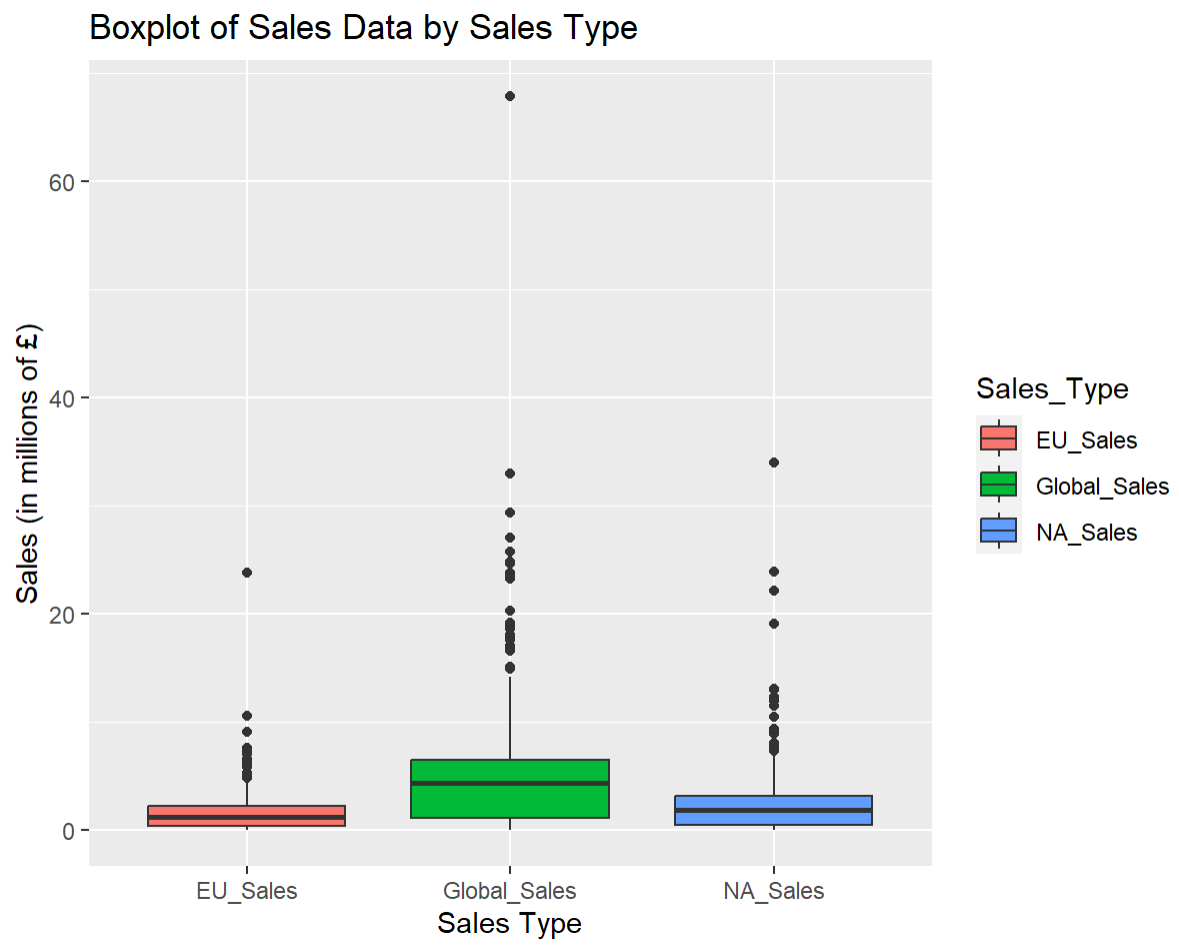
Given more time I would undertake further cross-validation and be cautious about generalisation on this data set which has no products from the last 7 years.

Turtle Games should:

1. Focus on the building sales by learning from the success of high impact products.
 - They can measure this by products that earn more than £27m in sales
 - First undertake detailed analysis on these products that have outperformed others
2. Build on the engaged customer base to improve marketing.
 - Continue to assess and refine the customer profiles and engagement by reviews
 - Firstly undertake A/B testing on how to make the most of customer loyalty
2. Monitor risks that might damage reputation and prevent sales.
 - Measure the sentiment by fine tuning models with custom lexicons
 - Use positive reviews and identify recurring themes such as quality issues to act on

6. Appendix

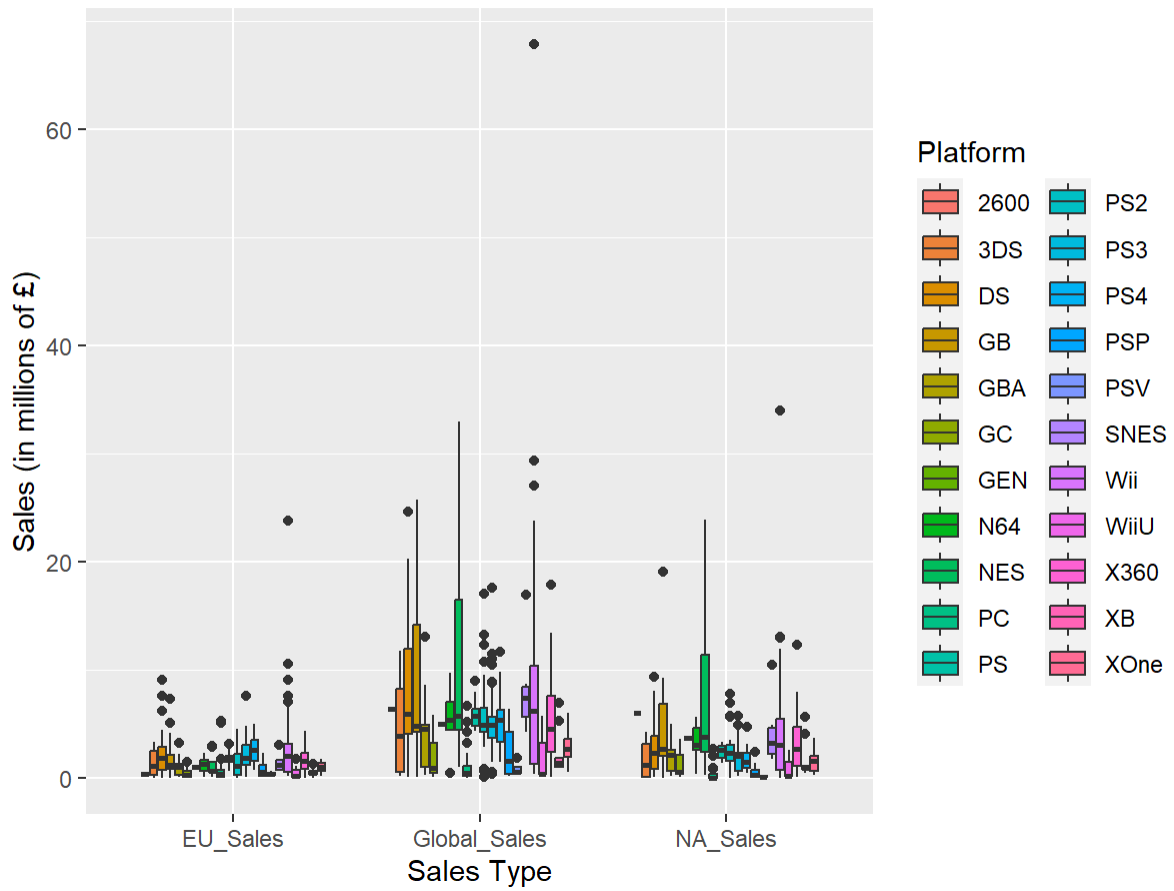
Plot 1.





Plot 2.

Boxplot of Sales Data by Platform





Plot 3.

