LSE Data Analytics Career Accelerator

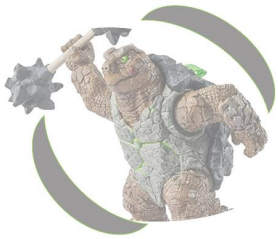
Course 3 Assignment

Turtle Games: Insights and Analysis

By Carl Szabo

September 2022

Contents

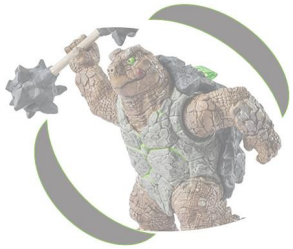


Background and Key Questions

Turtle Games is an international game manufacturer and retailer with a global customer base. Its product range includes books, board games, video games, and toys. Turtle Games has a business objective of improving overall sales performance by utilising customer trends.

Key questions this report will address:

- How do customers accumulate loyalty points?
- How can groups within the customer base be used to target specific market segments?
- How can social data (e.g. customer reviews) be used to inform marketing campaigns?
- What impact does each product have on sales?
- How reliable is the data?
- What are the relationships between North American, European, and Global Sales?



Analytical Approach: Data Cleaning

Turtle Games provided two files to clean and analyse:
- Customer Reviews
- Customer Sales

Data Cleaning:
After importing the data in Jupyter using the python package pandas, steps were taken to ensure data completeness and consistency.
- Checked the size of the dataframe
- Look for missing values (none)
- Ensure no duplicates (none)

The next step involved streamlining the dataset and focussing on the most relevant datapoints
- Removed unnecessary columns to focus on sales by product and platform
- Renamed renumeration and spending score columns for easier use

Analytical Approach: Python Analysis

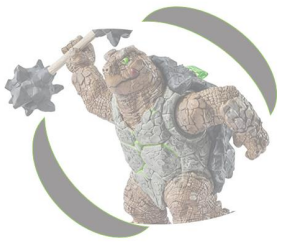
Python was used primarily to help the business address questions 1 to 3:

- How customers accumulate loyalty points?
- Identifying customer groups
- Using social data to inform marketing campaigns

How customers accumulate loyalty points?
To understand this better 3 models were used to compare relationships between:

- **Loyalty points vs spending**
- **Loyalty points vs income**
- **Loyalty points vs age**

Linear regression was used to conduct this analysis with the ols package from statsmodels



Analytical Approach: Python Analysis

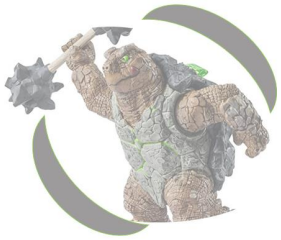
How customers accumulate loyalty points (cont)

- The purpose of this process is to identify variables that are meaningful for the next stage of analysis
- During this process we found that **income and spending were significant predictors of loyalty points, but age was not**

Identifying customer groups

The marketing department wants to better understand usefulness of income and spending scores. To do this analysis we needed to:

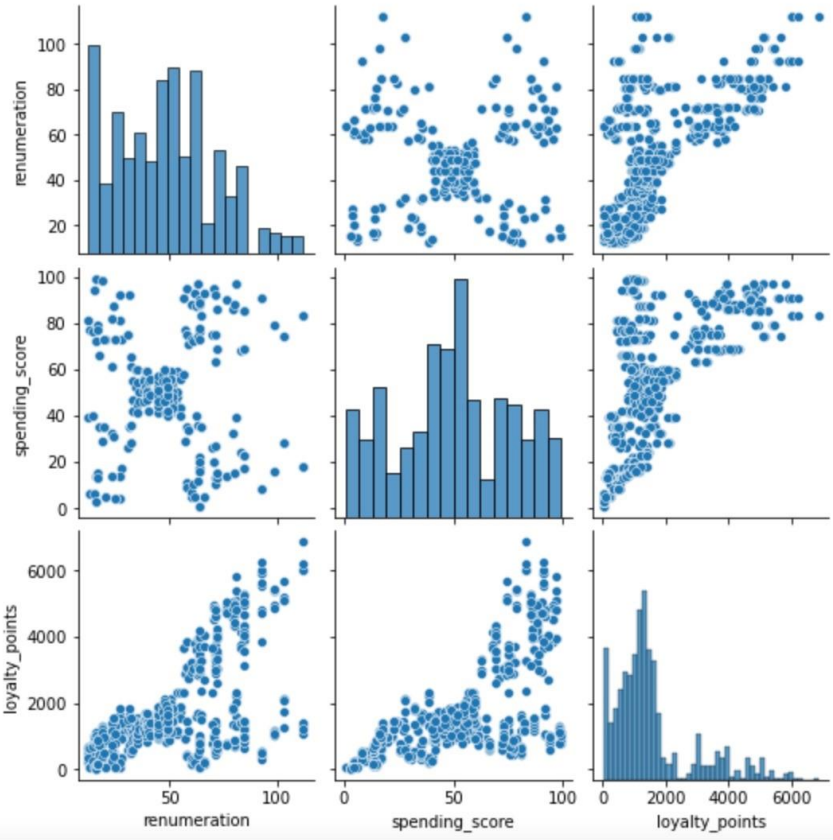
- Create a new dataframe holding just the 'loyalty points', 'renumeration' and 'spending_score' columns.

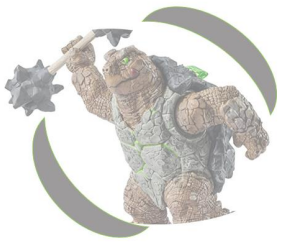


Analytical Approach: Python Analysis

Identifying customer groups (cont.)

- Next, we plotted the relationships between the '**spending_score**' and '**renumeration**' using matplotlib and seaborn to visually understand how the two might interact.

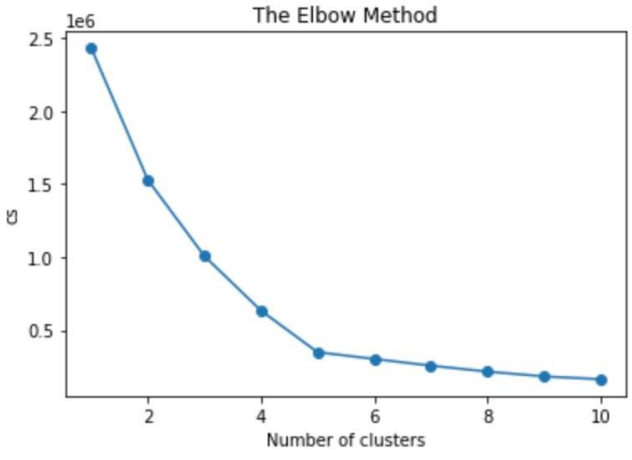


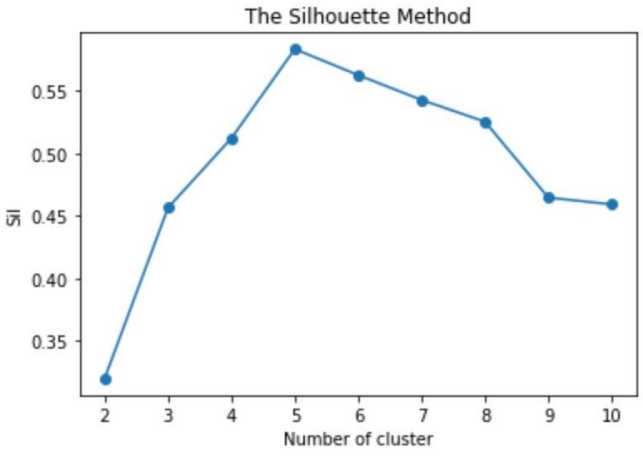


Analytical Approach: Python Analysis

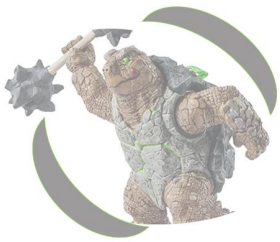
Identifying customer groups (cont.)

- K-means clustering was used to find the correct number of customer groups, leveraging both the Elbow and Silhouette methods





- **Both methods seem to indicate 5 is optimal** but in an attempt to be cautious 4 and 6 clusters were also analysed before concluding.



Analytical Approach: Python Analysis

Using social data to inform marketing campaigns

- Natural language processing was used to glean insights from the customer reviews of Turtle Games' existing products.
- For the analysis we used numerous packages including NLTK, WordCloud and TextBlob. The following was done to prepare the data:

- Dropping unnecessary columns

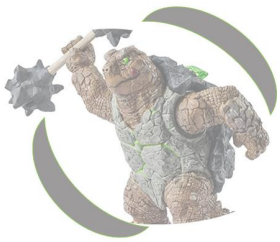
- Checking for missing values (none)

- Change to lower case and remove punctuation

- Dropping duplicates

- Apply tokenisation using word_tokenize via NLTK

- Create wordclouds using WordCloud



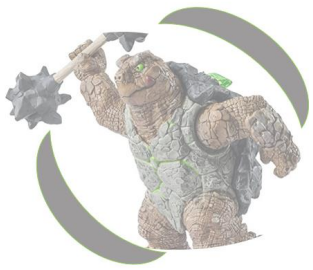
Analytical Approach: R Analysis

R was used primarily to help the business address questions 1 to 3:

- What impact does each product have on sales?
- How reliable is the data?
- What are the relationships between North American, European, and Global Sales?

What impact does each product have on sales?
The sales department wishes for the analysis to be conducted in R. For the next stage we will be looking at the customer sales data provided by Turtle games.

To get better insight into the data the tidyverse package was utilised, redundant columns were removed, and scatterplots and boxplots were created to better understand the drivers of sales for different products.



Analytical Approach: R Analysis

How reliable is the data?

- To test for reliability the data was grouped by product id.
- Then a Shapiro-Wilk test was performed to explore the normality of the data. This can help us understand if additional tests assuming normality would be beneficial.
- Also, skewness and kurtosis were examined using the Moments R package.

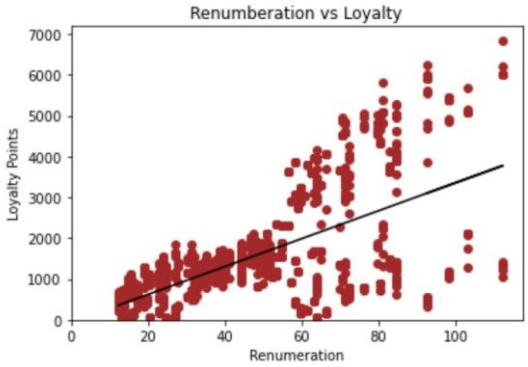
What are the relationships between North American, European, and Global Sales?

- Linear regression (single and multi-variate) was used to analyse the relationship between sales in different regions.
- Correlation analysis was also performed to look at how much of the variability in global sales can be explained by EU or NA sales figures.

Visualisation and Insights

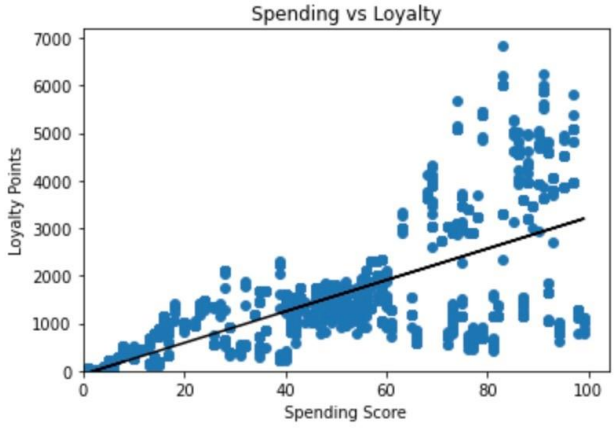
How do customers accumulate loyalty points?

- The Sales team wants explore the relationship between Loyalty points, Age, Renumeration and Spending



The chart to the left and the line of best fit indicate a strong positive relationship between renumeration and loyalty points.
The R2 of 0.38 indicates 38% of the variability in loyalty points can be explained by renumeration, so customers who earn more may make for good targets for the sales team.

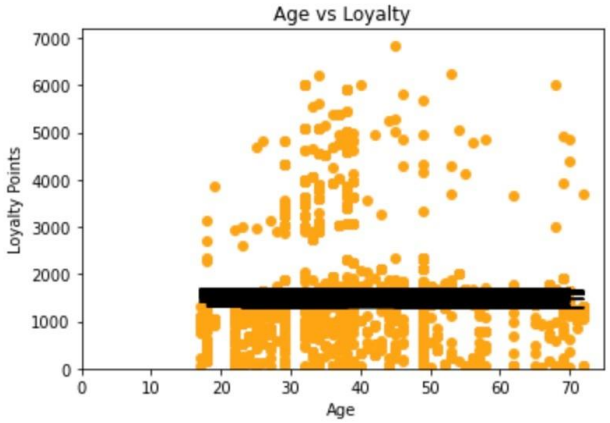
The chart to the left and the line of best fit indicate a strong positive relationship between spending and loyalty points.
The R2 of 0.45 indicates 45% of the variability in loyalty points can be explained by spending score. This may make customers who spend more good targets for future sales efforts.

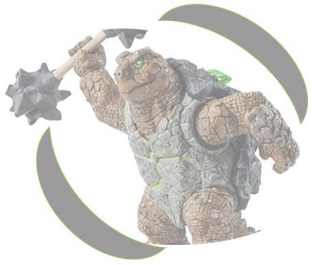


Visualisation and Insights

How do customers accumulate loyalty points (cont.) ?

When analysing ‘age’ vs loyalty points no clear and obvious relationship appears. The R2 of 0.02 confirms, and the sales team might wish to ignore ‘age’ as an important consideration.

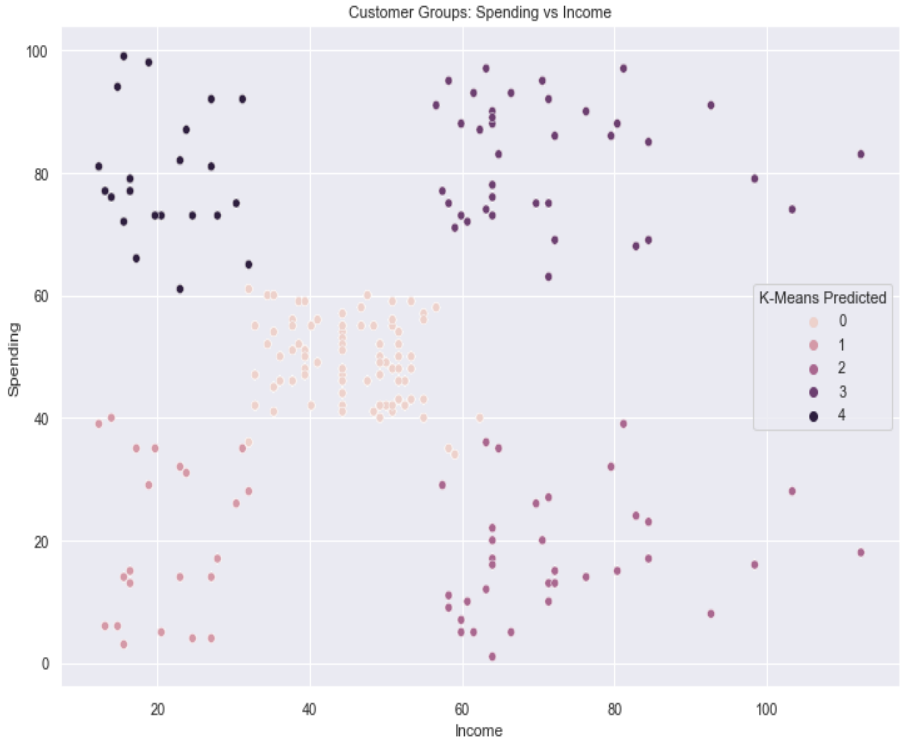




Visualisation and Insights

How can groups within the customer base be used to target specific market segments?

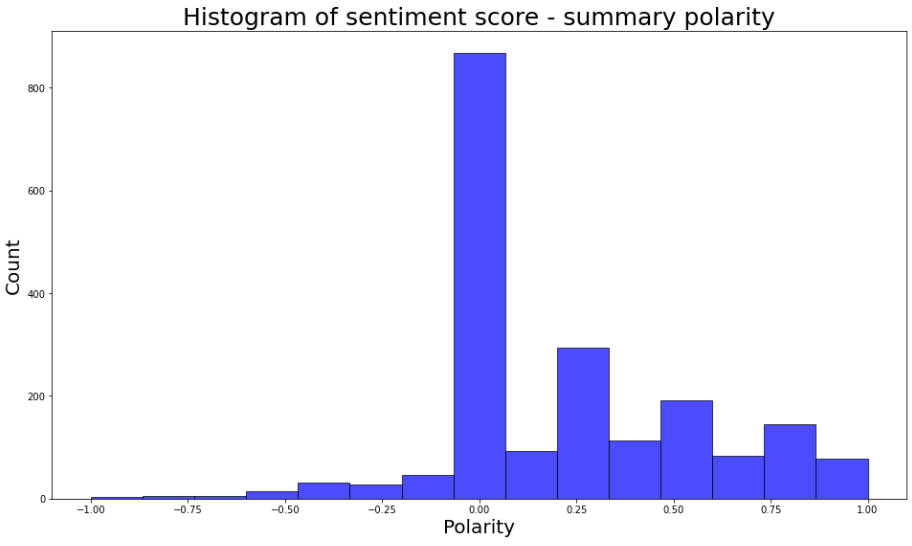
- There appear to be 5 main customer ‘types’ for Turtle Games. Middle income, middle spenders are most common
- **Group 2 is one of the most interesting**, a high potential group with disposable income but a low spending score.



Visualisation and Insights

How can social data (e.g. customer reviews) be used to inform marketing campaigns?

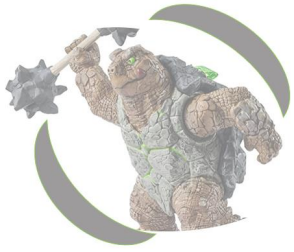
-



Word	Frequency
stars	420
five	338
game	319
great	294
fun	218
love	92
good	91
four	56
like	54
expansion	52
kids	50
cute	45
book	43
one	38
awesome	36

Reviews of Turtle Games products skew positive (above), which is a good indication customers are generally happy with their purchases.

Some of the most common from the top 15 words list include '**great**', '**fun**', '**awesome**' and '**love**'.

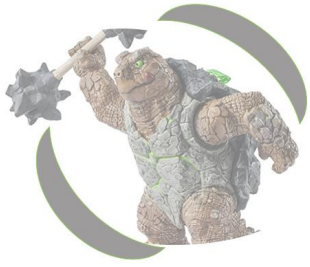


Visualisation and Insights

How can social data (e.g. customer reviews) be used to inform marketing campaigns? (Cont.)

More could be learned from the unhappy customers however. For example, the most negative review by polarity score:

review	review_polarity
booo unles you are patient know how to measure i didnt have the patience neither did my daughter boring unless you are a craft person which i am not	-1.000000

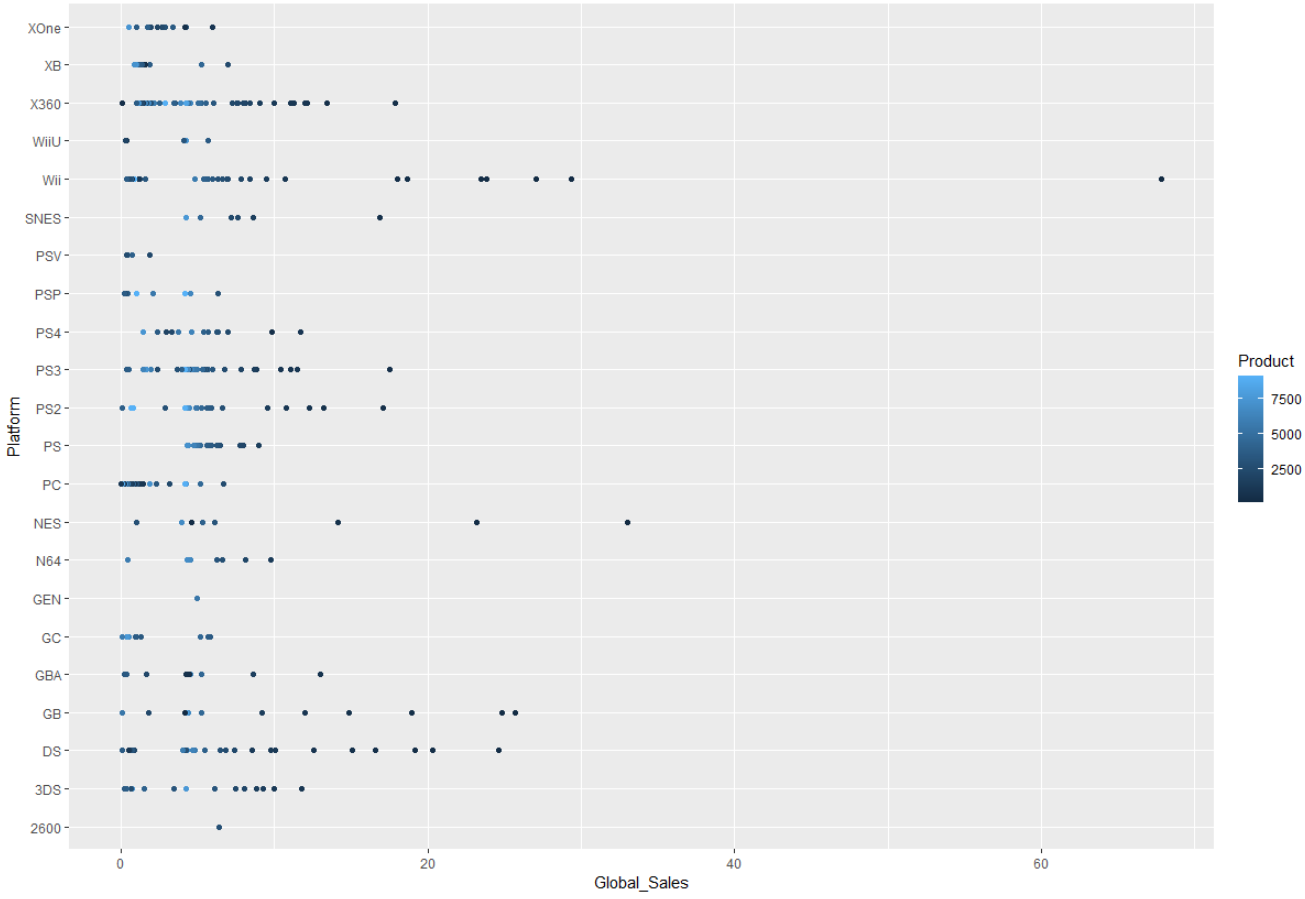


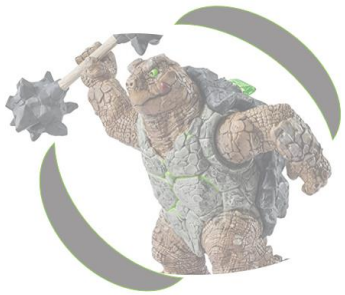
Are there common negative reviews around certain products or types of products (e.g. arts and crafts)?

Visualisation and Insights

What impact does each product have on sales?

- Turtle Games sells many different products across different platforms. Looking at the data we can quickly spot outliers.



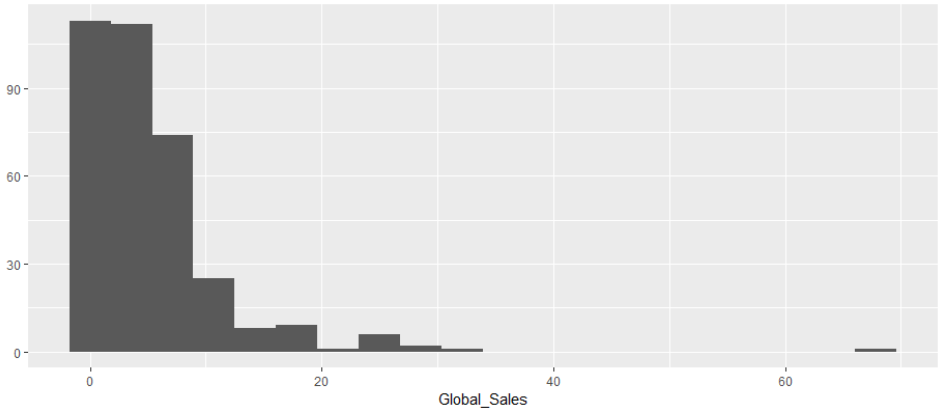


What is it that drives the success of certain products Turtle Games sells? Does platform matter?

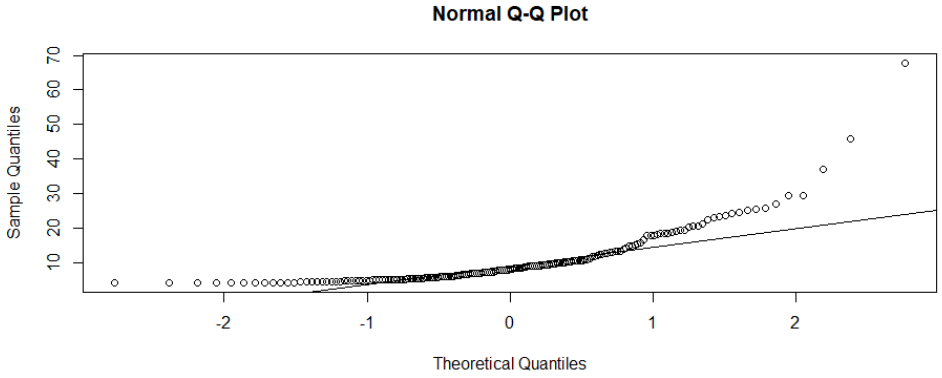
Visualisation and Insights

How reliable is the data?

Looking at global sales data we can see it is not normally distributed.



The data is clustered toward the left side of the graph and the right tail extends out quite far to the right (outliers).



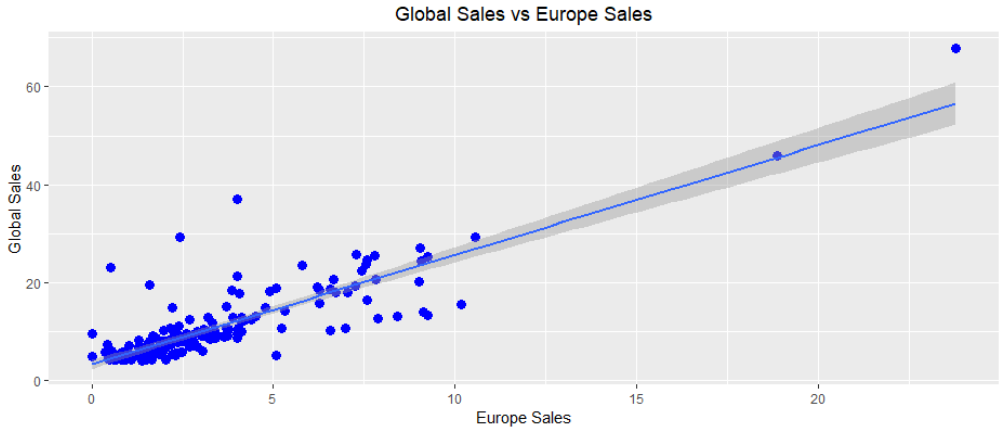
A Q-Q plot confirms the non-normal distribution as the dots do not follow the trend line at extremes.

This is likely due to the correlation between EU sales and North American sales (62%)

By Carl Szabo

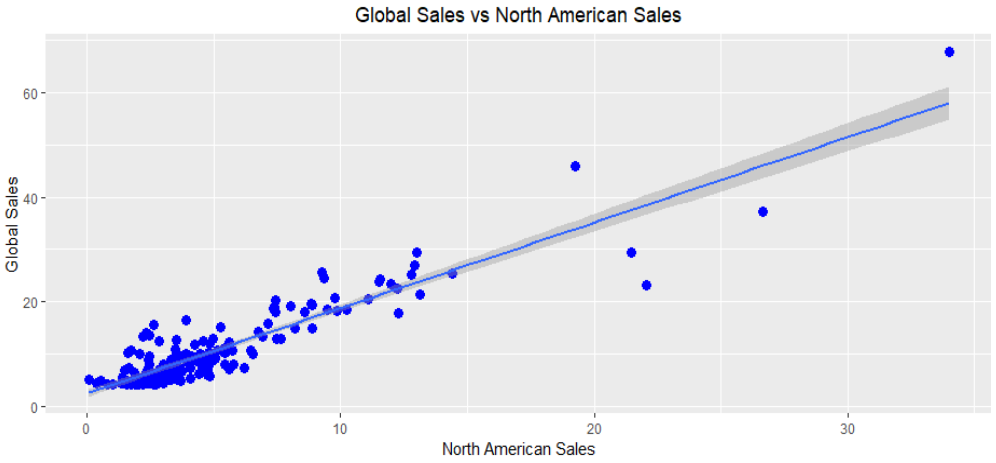
Visualisation and Insights

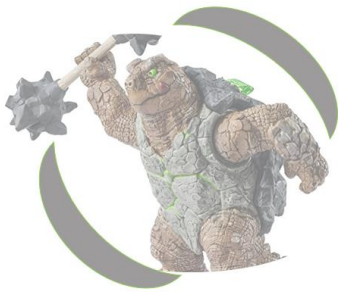
What are the relationships, if any, between North American, European, and Global Sales?



There is strong positive correlation between European sales and Global Sales.

And a strong positive relationship between North American Sales and Global Sales.





European and North American sales explain a high percentage of the variance in Global Sales, together accounting for 95%+ of the variation. Business success is closely tied to the health of these two markets.

Patterns and Predictions

Pattern 1: Customers love Turtle Games products

- Review data skews positive so that means Turtle Games is doing a great job of serving it's customers.
- How can this be leveraged to increase sales. Perhaps bonus loyalty points for friends and family referrals, or a similar idea, could build upon the good vibes from its client base.

Pattern 2: Target the right customer segments to spur future growth

- Middle Earning, Middle spending customers could be encouraged to move higher on the spending scale. Engaging with these customers now could also be beneficial as and when these customers increase their spending power.
- High Earning, Low spending customers present great opportunity. Perhaps a targeted campaign based around products that High Earning, High Spending customers lover would work well to engage with this group?

Pattern 3: Consider expanding to new markets beyond EU and North America

- Currently, the overall sales of the business are closely tied to the health of the EU and American markets. Can expansion to new markets help diversify Turtle Games' revenue streams and better ensure business stability?