

Turtle Games Analysis

Author: Matt Rose

Date: Thursday 8th September 2022

Version: 1.0

Version

Author	Version	Date	Comments
Matt Rose	0.1	8 th September 2022	Initial Draft
Matt Rose	1.0	8 th September 2022	Final Version

Table of Contents

Version	2
Scenario	3
Approach & Insights	3
Accrual of Loyalty points	3
Customer Clustering	4
Social Sentiments	6
Product Effect on Sales	8
Data Reliability	10
Relationship between Sales and Location	12
Additional Exploration	12

Scenario

Turtle Games is a games manufacturer and retailer and supply customers on a global scale. They manufacturer and sell their own label games whilst also acting as an intermediary for other games manufacturers, selling products which they purchase from competitors.

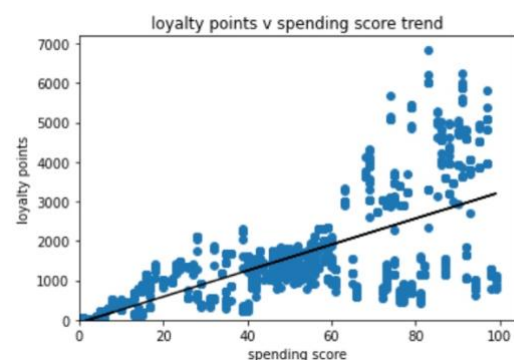
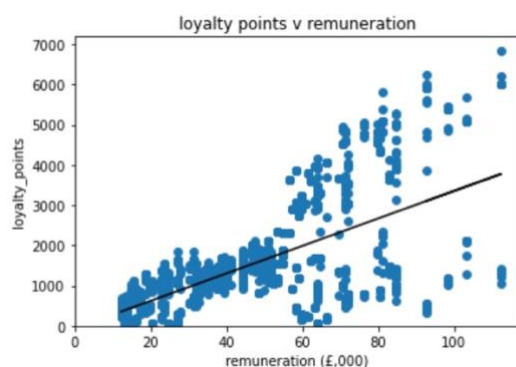
Through the analysis of available data, Turtle Games stakeholders want to understand the following:

- How customers accumulate loyalty points
- How groups within the customer base can be used to target specific market segments
- How social data can be used to inform marketing campaigns
- Impact on sales by product
- How reliable the data is
- What the relationship is (if any) between geographical locations and global sales.

Approach & Insights

Accrual of Loyalty points

Initial Analysis using linear regression techniques against the Turtle Games data shows that there is clear linear relationships between loyalty points, spending score and salary, however there is probably deeper analysis needed by implementing multiple linear regression to show even further trend analysis between the columns.



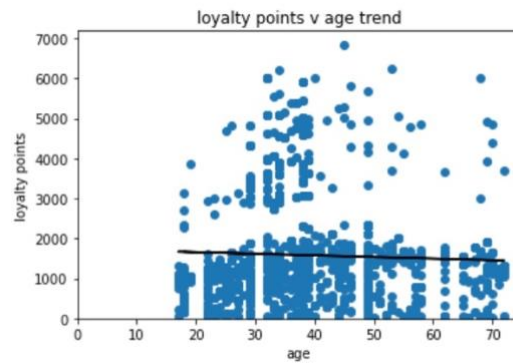


Figure 1 - Graphs to show the relationships between Loyalty Points, age, spending score and remuneration

There was no linear relationship between loyalty points and age. This would be useful to use as an independent variable in a multiple linear regression to see if it has any influence on the model when coupled with other independent variables.

When plotting the datapoints and the regression line on each model, it is interesting how the higher that the data moved in the independent variable, the more erratic the data becomes with values showing both below and above the trend line. Ideally a further data cleanse would be performed to remove any outliers outside of the IQR which may drive a more accurate model and trend line.

Customer Clustering

To establish groups within the customer base which can be targeted for specific marketing campaigns, we first use k-means clustering. This allows the ability to determine the optimal number of customer clusters.

Prior to running the predictive models, the plot of the score and remuneration data led to a scatterplot which initially implies a 5-cluster model.

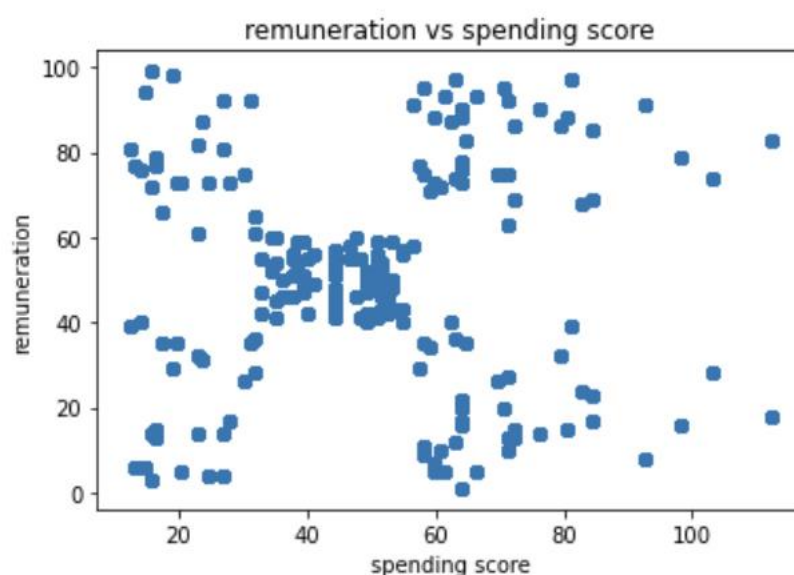


Figure 2 - Scatterplot of remuneration and score

Continuing the analysis to determine usefulness, the use of the Elbow and Silhouette methods shows that we should consider 4, 5 and 6 clusters as this is where the line curves before trailing.

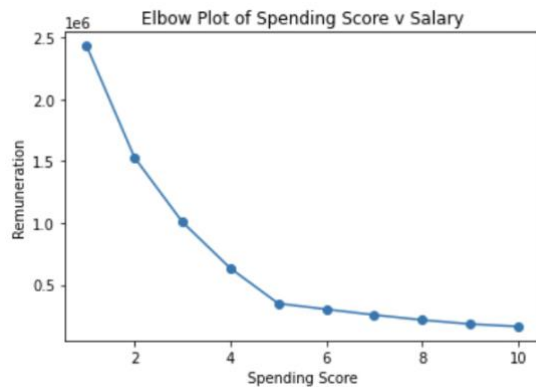


Figure 3 - Elbow Method

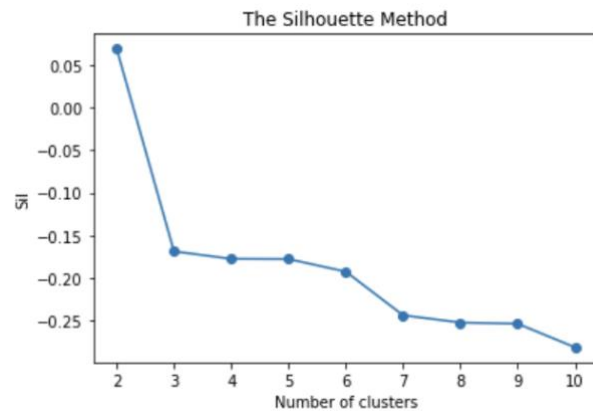


Figure 4 - Silhouette Method

Following the analysis of each of the different numbers of clusters outlined from the above figures, the resulting optimal cluster is deemed as 5 and is shown below. This clearly indicates the various clusters of customer salary vs score meaning that Turtle Games could specifically market certain products to specific demographics of customer.

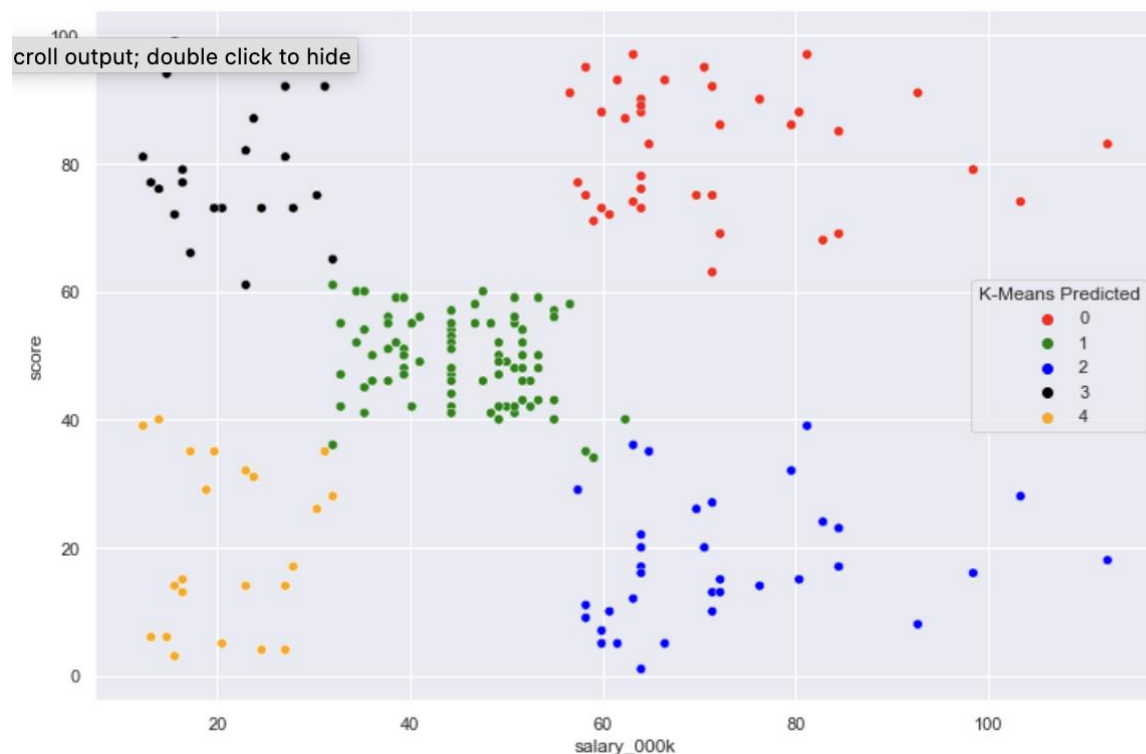


Figure 5 - scatter plot of 5 cluster model

Social Sentiments

Consumer data and opinion has never been more powerful than it is today. Using Python, we can get to the heart of customer sentiments and provide the ability to analyse how Turtle Games is seen as well as how the products are reviewed. This is important data to allow Turtle Games to react, pre-empt and identify improvement processes in their business to deliver as high a customer satisfaction as possible – thus leading to repeat business.

The analysis has been performed over a series of review and summary data, stripping out common 'non-useful' words to deliver a word cloud of the most frequently used words that Turtle Games customers are using. As is shown below, most of the words used are seemingly positive, with words like 'game', 'great' and 'fun' being used in a lot of the reviews. Whilst this is a good start initially as it identifies the most frequent words, further analysis on the more negative words would be very helpful in determining what the biggest challenges that customers face.



Figure 6 - Word cloud from customer reviews

In addition to the word cloud, we have been able to determine the sentiment scores as well as identify the top and bottom 20 reviews for Turtle Games. Plotting the sentiment scores on a histogram shows that we have majority of positive or neutral sentiment, with only a few showing as negative sentiment. This shows that Turtle games is, for the most part, doing a good job with its customers, but further analysis can be performed on the negative sentiments to identify improvement areas.

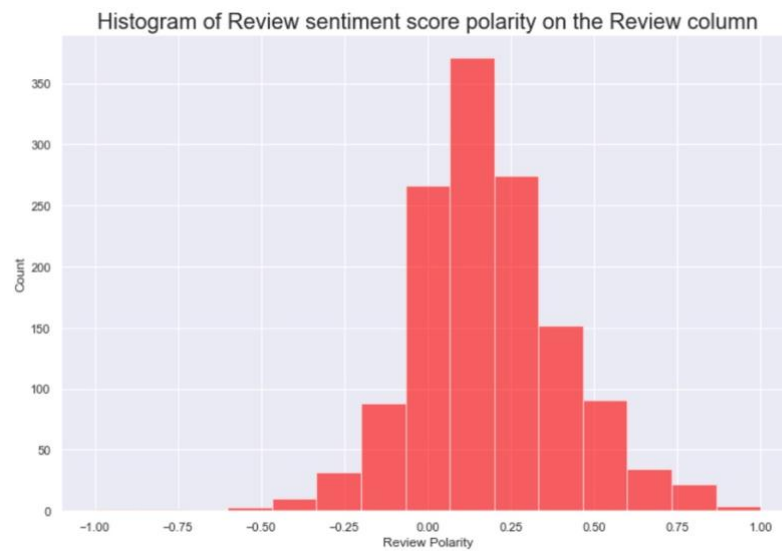


Figure 7 - Review sentiment histogram

	review	review_polarity	review_subjectivity
	perfect	1.000000	1.000000
	my daughter loves her stickers awesome seller thank you	1.000000	1.000000
	perfect for tutoring my grandson in spelling	1.000000	1.000000
the best part i see is the box what a wonderfully diverse and rounded set for the cost i am so happy and as the dm you know that if i am happy my players are happy		0.880000	0.860000
great quality very cute and perfect for my toddler		0.816667	0.916667
the pictures are great ive done one and gave it to a friend of mine who likes dragons		0.800000	0.750000
great seller happy with my purchase 5 starrrr		0.800000	0.875000
great easter gift for kids		0.800000	0.750000
these are great		0.800000	0.750000
bought this because i wanted it all these dd games are great		0.800000	0.750000
husband seems happy with it		0.800000	1.000000
great accessory to use with the playing mat		0.800000	0.750000
great price arrived on time with no damage will be a great addition to my collection		0.800000	0.750000
this is a great accessory to the starter set i would recommend this to anyone who owns the starter set		0.800000	0.750000
my granddaughter loves these so happy to find peppa pig items for her		0.800000	1.000000
great doll to go with the book animals cant wait to read book with the doll to the grandkids		0.800000	0.750000
a great creation tool it helps me concentrate		0.800000	0.750000
prompt service and a great product		0.800000	0.750000
this is a great tool to have at hand when playing quiddler		0.800000	0.750000
this is a great product i use it as a therapeutic tool and it has been very effective		0.790000	0.875000

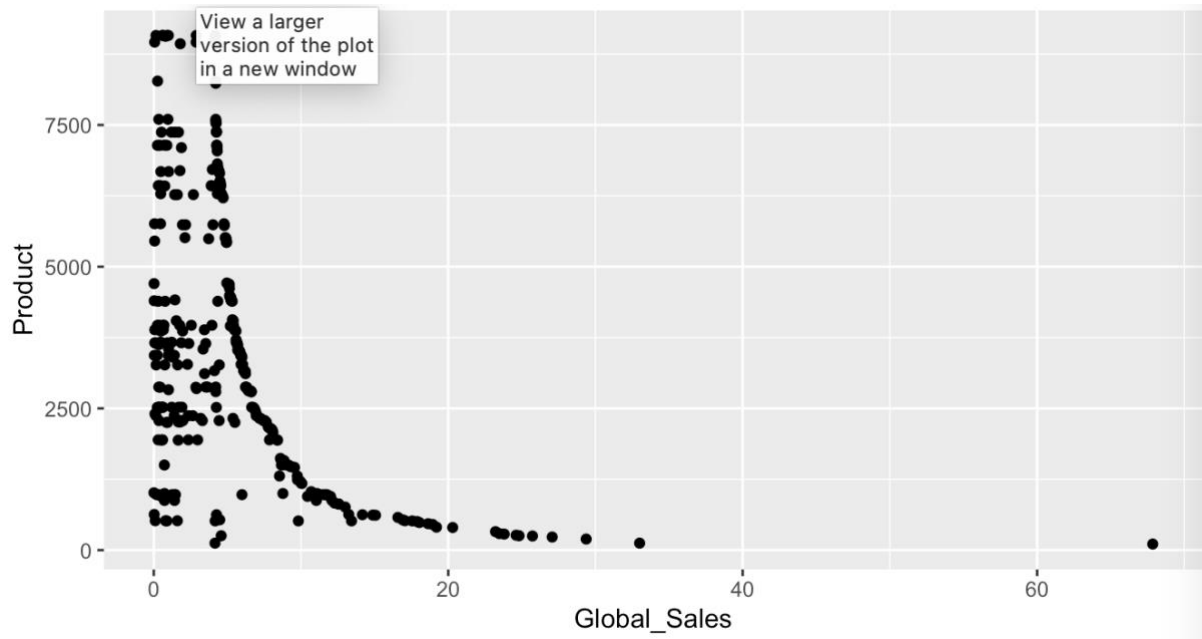
Figure 8 - Top 20 reviews based on sentiment

	review	review_polarity	review_subjectivity
booo unless 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	1.000000
incomplete kit very disappointing		-0.780000	0.910000
one of my staff will be using this game soon so i dont know how well it works as yet but after looking at the cards i believe it will be helpful in getting a conversation started regarding anger and what to do to control it		-0.550000	0.300000
i bought this as a christmas gift for my grandson its a sticker book so how can i go wrong with this gift		-0.500000	0.900000
i sent this product to my granddaughter the pompom maker comes in two parts and is supposed to snap together to create the pompoms however both parts were the same making it unusable if you cant make the pompoms the kit is useless since this was sent as a gift i do not have it to return very disappointed		-0.491667	0.433333
my 8 yearold granddaughter and i were very frustrated and discouraged attempting this craft it is definitely not for a young child i too had difficulty understanding the directions we were very disappointed		-0.446250	0.533750
i purchased this on the recommendation of two therapists working with my adopted children the children found it boring and put it down half way through		-0.440741	0.485185
this game although it appears to be like uno and have an easier play method it was still too time consuming and wordy for my children with learning disabilities		-0.400000	0.400000
my son loves playing this game it was recommended by a counselor at school that works with him		-0.400000	0.400000
if you like me used to play dd but now you and your friends grewed up and cant be together because all the responsibilities and bla bla bla this game is for you come to the dungeon		-0.400000	0.400000
you can play the expansions one at a time or add then both in for a longer game if your into lords of waterdeep this is a must have		-0.400000	0.400000
if you play dungeons and dragons then you will find this board game to be dumb and boring stick with the real thing		-0.393750	0.550000
i was a bit disappointed in the quality of the cardboard pieceholders and the fact that they changed the names of some hotels otherwise i mean its a terrific game		-0.365625	0.709375
very fun game to use with kids working on handling anger you play like uno but have to answer questions about anger		-0.352500	0.265000
i really like this game it helps kids recognize anger and talk about difficult emotions		-0.350000	0.450000
i am a therapist for children and this game is so valuable to bring out insight and solutions to deal with and identify feelings of anger i use it frequently		-0.333333	0.300000
confusing instructions and its not for 6 year olds its boring too its asking the same question but each question is worded differently		-0.325000	0.531250
as my review of gf9s previous screens these were completely unnecessary and nearly useless skip them this is the definition of a waste of money		-0.316667	0.316667
the adventures are tough but you can get through them it all comes down to the die roll just like any dd game		-0.314815	0.507407
a crappy cardboard ghost of the original hard to believe they did this but they did shame on hasbro disgusting		-0.305556	0.763889

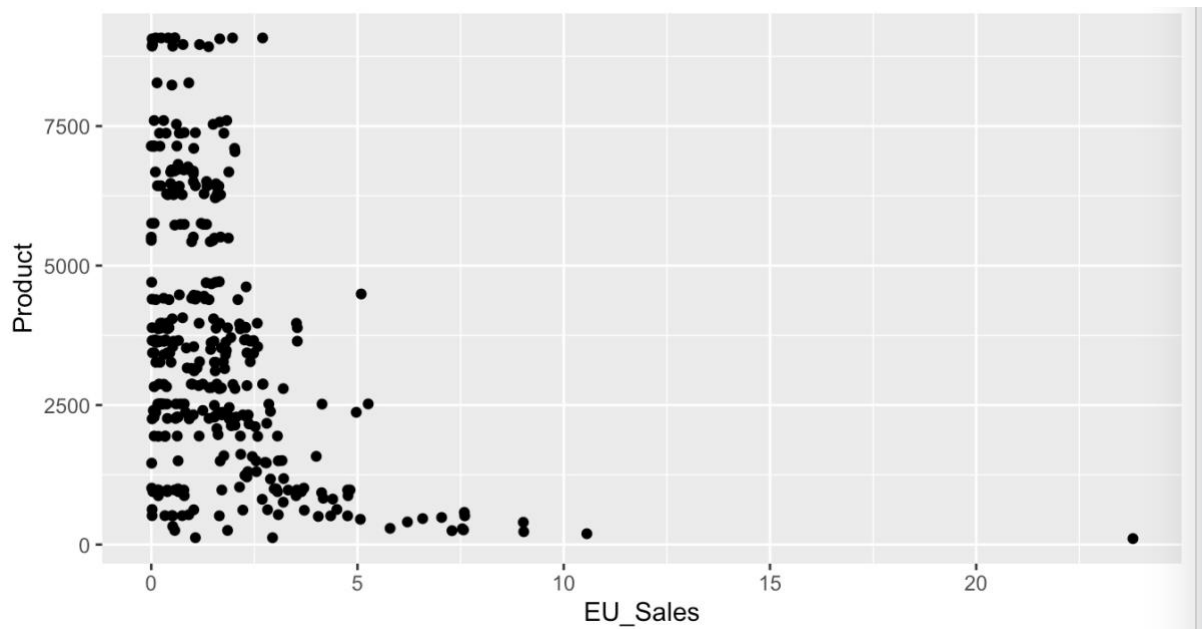
Figure 9 - Bottom 20 reviews based on sentiment

Product Effect on Sales

To undertake the analysis of the effects of Products on Sales, there was a switch from Python to R to make the most of the statistical libraries that R can access. The initial plot of Products against Global Sales show that the lower, more original products with lower product IDs contribute far more to Global Sales than newer products with higher Product ID's. **This is predicated on the assumption that lower product IDs are the original product ranges of Turtle Games.**

*Figure 10 - Product vs Global Sales*

In addition, we can see the relationship where the products have the biggest impact in EU and NA Sales, which follow the same trajectory as the Global Sales. This indicates that both EU and NA Sales follow the global pattern and that there are consistent products that Turtle Games sells which are globally adopted.

*Figure 11 - Products sold vs EU Sales*

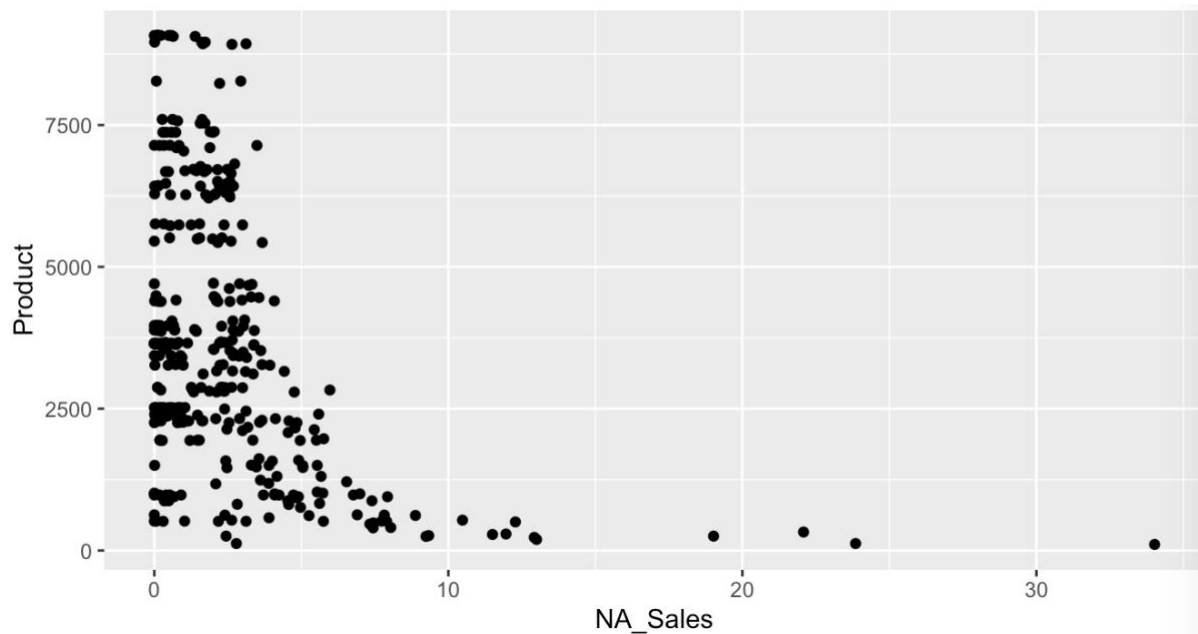
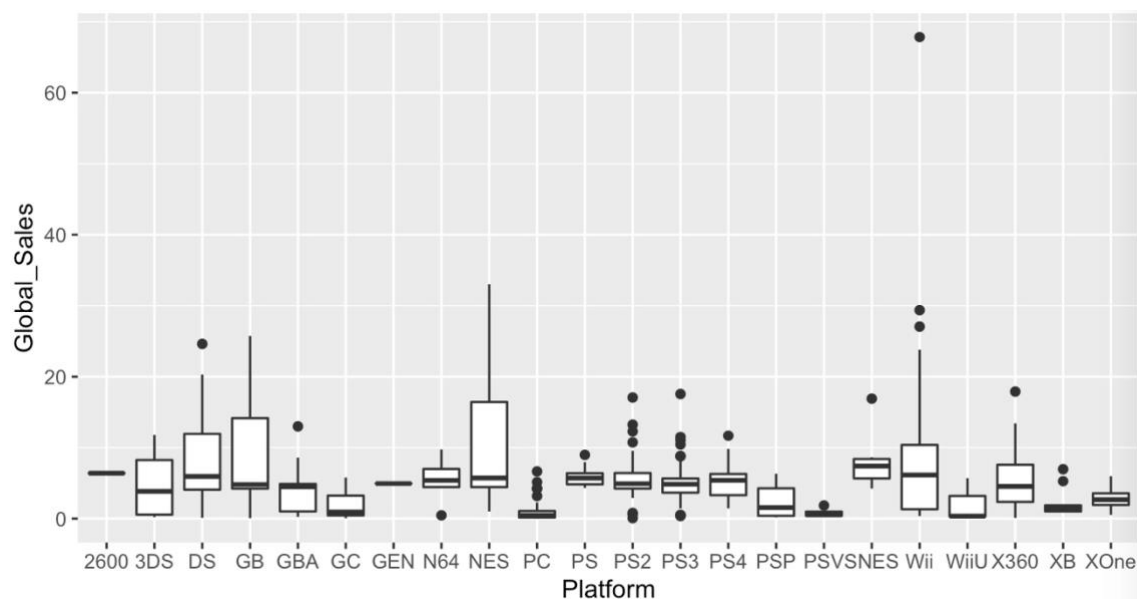


Figure 12 - Products sold vs NA Sales

When continuing with the analysis we can also see which platforms also contribute the most to the global sales with also displaying outliers – in this instance NES, GameBoy and Wii have been the most impactful on Global Sales.



Data Reliability

Fundamentally, the key question that needs to be answered is “how reliable is the data?” In order to provide the answer to this question, we turn again to R and identify data normality and skewness using a variety of statistical methods. We focused primarily on QQ (quartile-quartile), Shapiro-Wilk, skewness and kurtosis tests to identify how normalised the data is and whether it can be trusted for the purpose of the analysis we are performing.

The QQ plot shows that when we look at the EU, NA and Global Sales they all show that the data doesn't align to any sort of normality due to the fact that as the sales grow, the deviation from the trend line becomes much larger suggestion of the bell curve being after the data points that we have plotted and therefore a move away from normalised data. For normalised data, we would expect to see a bell curve on the plot.

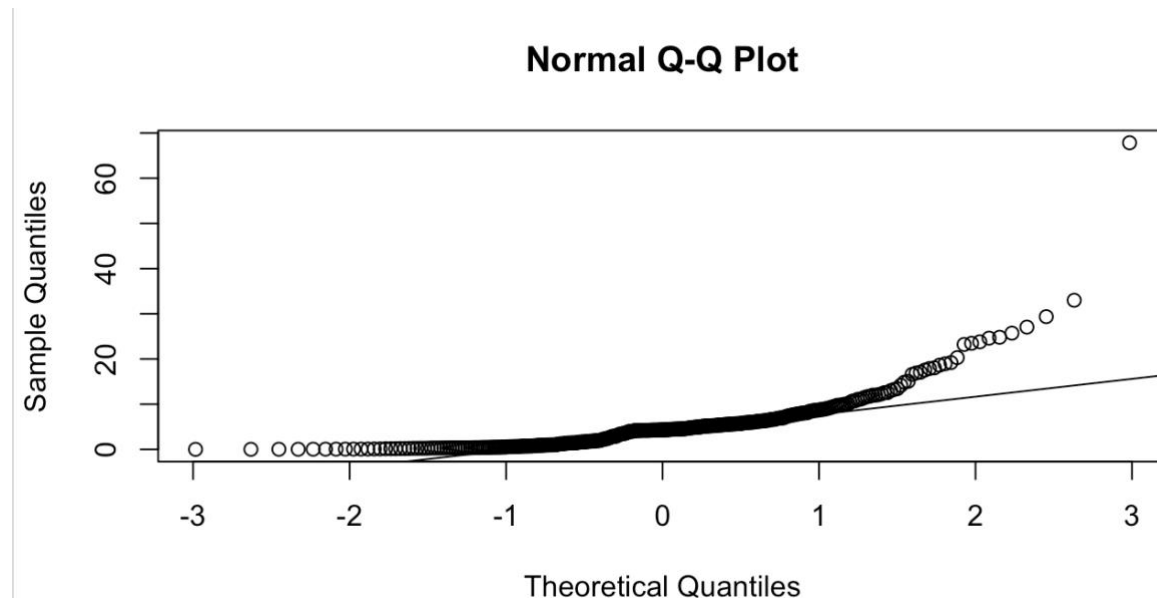


Figure 13 - Global Sales QQ Plot for normalisation

The Shapiro-Wilk test however, gives us an interesting outcome. The p value is showing as significantly less than 0.05 which means that the data is not normally distributed for Global Sales.

Shapiro-Wilk normality test

```
data: turtle.subset$Global_Sales
W = 0.6818, p-value < 2.2e-16
```

Skewness and Kurtosis values are shown as follows for NA, EU and Global Sales, this indicates a data set which is highly peaked, which moves away from normalised data and shows there is a right-skewed output of this data. Further work on the data set should be undertaken to bring normalisation and skewedness back towards the 0-centre line.

```
> skewness(turtle.subset$NA_Sales)
[1] 4.30921
> kurtosis(turtle.subset$NA_Sales)
[1] 31.36852
> skewness(turtle.subset$EU_Sales)
[1] 4.818688
> kurtosis(turtle.subset$EU_Sales)
[1] 44.68924
> skewness(turtle.subset$Global_Sales)
[1] 4.045582
> kurtosis(turtle.subset$Global_Sales)
[1] 32.63966
```

Relationship between Sales and Location

To determine the relationship between NA and EU sales to Global Sales, we ran several linear and multiple regressions to show that there is a high correlation between the NA and EU sales and Global Sales.

Having run the linear models for EU and NA Sales against Global Sales we see that there is an R-squared value of 77.01% and 87.4%. When running a multiple linear regression with a combination of both NA and EU sales, this R-squared value jumps to 96.8%. This means that a significant number of NA and EU sales can be used to explain the Global Sales, so the combination of both geographical regions make up most of the Global Sales.

It would be interesting to see how the EU and NA sales break down further into countries/states.

Additional Exploration

- Multiple linear regression to be performed against the loyalty points to see if there is a combination which more accurately predicts the combination that will enhance loyalty point acquisition.
- Breakdown of EU and NA states/countries to have a more granular view of where the sales make up is and where to target campaigns
- Cleanse the data further to get a better, more normalised data set.
- Run a multiple linear regression of loyalty points, using score, age and remuneration to see if the age has any additional influence over loyalty points.
- Identify further negative sentiment to link to specific products or determine ways of improving manufacture or business process.