






Assignment 3: Turtle Games




Predicting Future Outcomes




Github Repository:


https://github.com/Harsh46/Gulati_Harshdeep_DA301_Assignment

 **Gulati_Harshdeep_DA301_Assignment** Public


 Pin  Unwatch **1**  Fork **0**  Star **0**


 main  1 branch  0 tags


 Go to file  Add file  Code


About 


No description, website, or topics provided.

 Readme

 Activity

 0 stars

 1 watching

 0 forks

Releases

No releases published

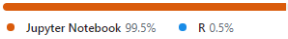
[Create a new release](#)

Packages


No packages published

[Publish your first package](#)

Languages





Language	Percentage
Jupyter Notebook	99.5%
R	0.5%

 Harsh46 Add files via upload

630751e now 38 commits

Data	Delete This is the data used in the assignment	yesterday
Figures and Graphs	Delete Figures and Graphs used	8 hours ago
Gulati_Harshdeep_DA301_Assignme...	Final version of jupyter notebook	8 hours ago
Gulati_Harshdeep_DA301_Assignme...	Add files via upload	now
Gulati_Harshdeep_DA301_Assignme...	Update and rename Gulati_Harshdeep_DA301_Assignment_Rscript.R.R...	9 minutes ago
Harshdeep_Gulati_DA301_Assignme...	Add files via upload	6 minutes ago
README.md	Update README.md	8 hours ago

 README.md 

Gulati_Harshdeep_DA301_Assignment

LSE DA201: fictional project Turtle Games improve overall sales performance by utilising customer trends.

Assignment 1 insights and summaries:

Background/Context of the Business

Turtle Games, a game manufacturer and retailer with a global customer base, is looking to improve overall sales performance by utilising customer trends. This data analytics project aims to help Turtle games understand:

1. Customer accumulation of loyalty points
2. Their different target market segments
3. How to form marketing campaigns using customer reviews
4. The sales based on different products
5. The reliability of their data
6. The relationships between sales in NA, in EU, and Globally

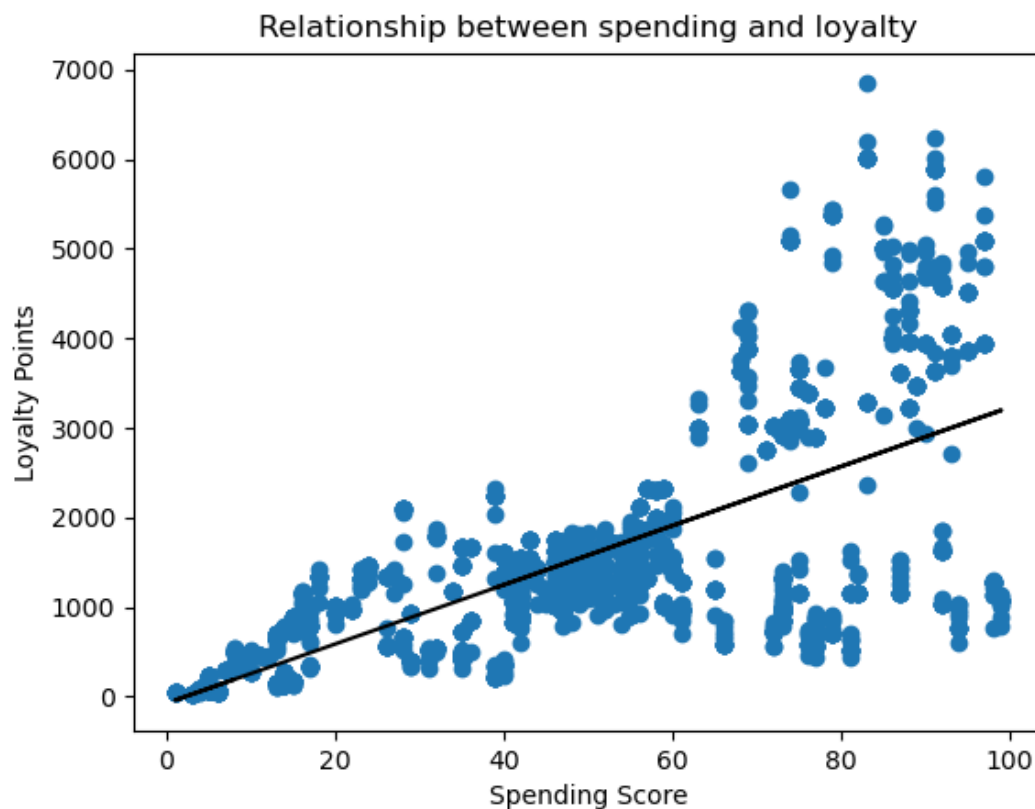
General Analytical Approach

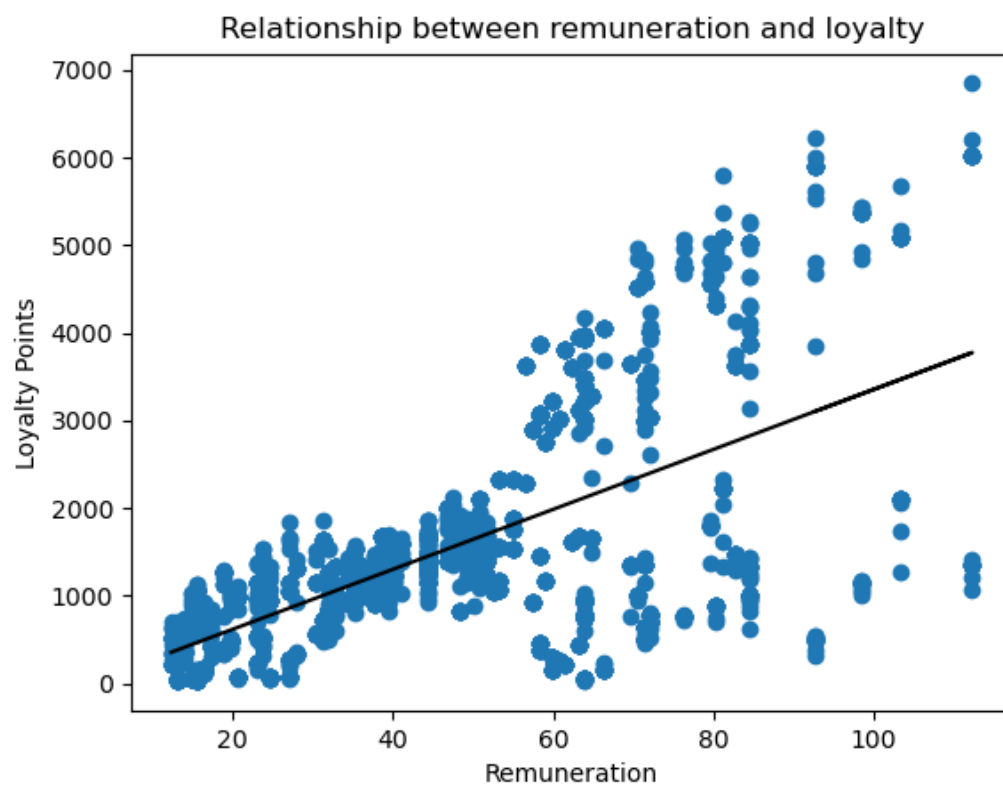
Data was first sense-checked using the `.info` and `.describe` functions in python and the `view` and `summary` functions in R. Missing data was also inspected, however there were none in any of the datasets used. Appropriate subsets were also created. Libraries used in Python were `numpy`, `matplotlib`, `seaborn`, `statsmodels`, `sklearn`, `scipy`, `wordcloud`, `nltk`, `os`, `collections`, and `textblob`. Packages used in R were `tidyverse`, `moments`, and `psych`. Further analytical approaches differed depending on task and shall be discussed in the sections to follow.

1. Customer accumulation of loyalty points

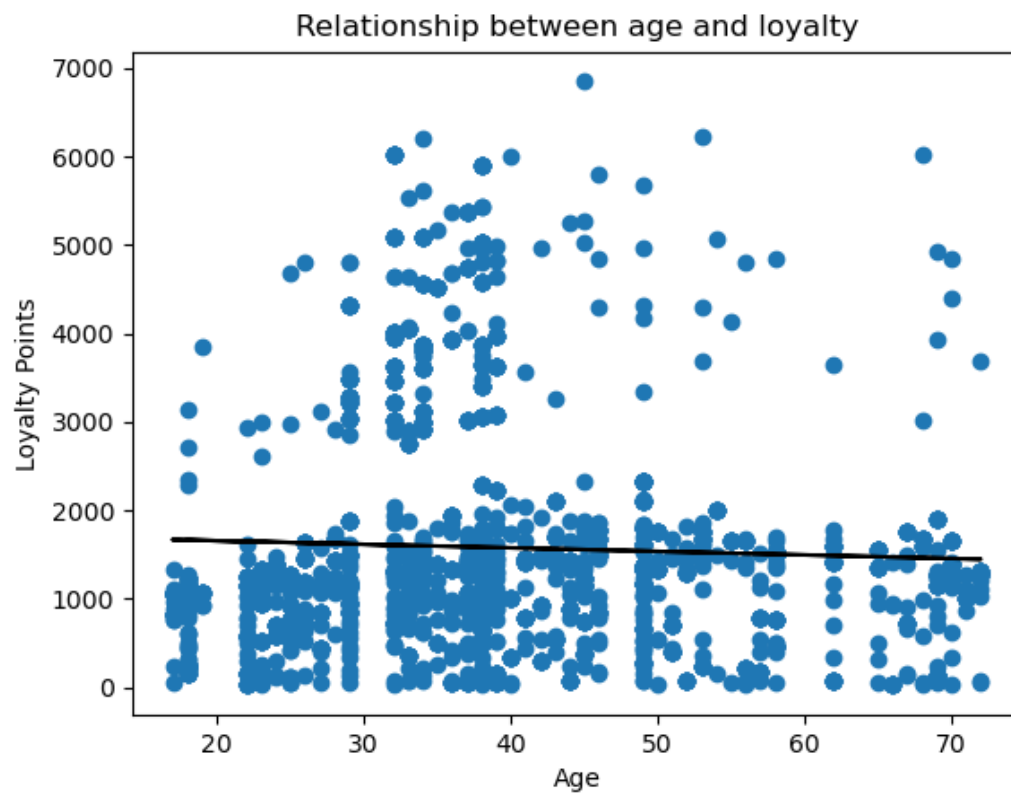
Simple linear and Multiple Linear Regression (MLR) models were made to investigate the relationship between loyalty points and the variables: spending, remuneration, and age.

The analysis suggests that there is a significant ($p < 0.05$, Confidence Intervals [CI] did not have 0) relationship between spending and loyalty points as well as remuneration and loyalty points. The simple linear regression model for spending had an R-Squared value of 45.2% and a coefficient value of 33.06. From this, we can predict that as spending score increases by 1, the predicted value of loyalty points increases by 33.06. For remuneration, the R-Squared value is 38.0% and the coefficient value is 34.19.





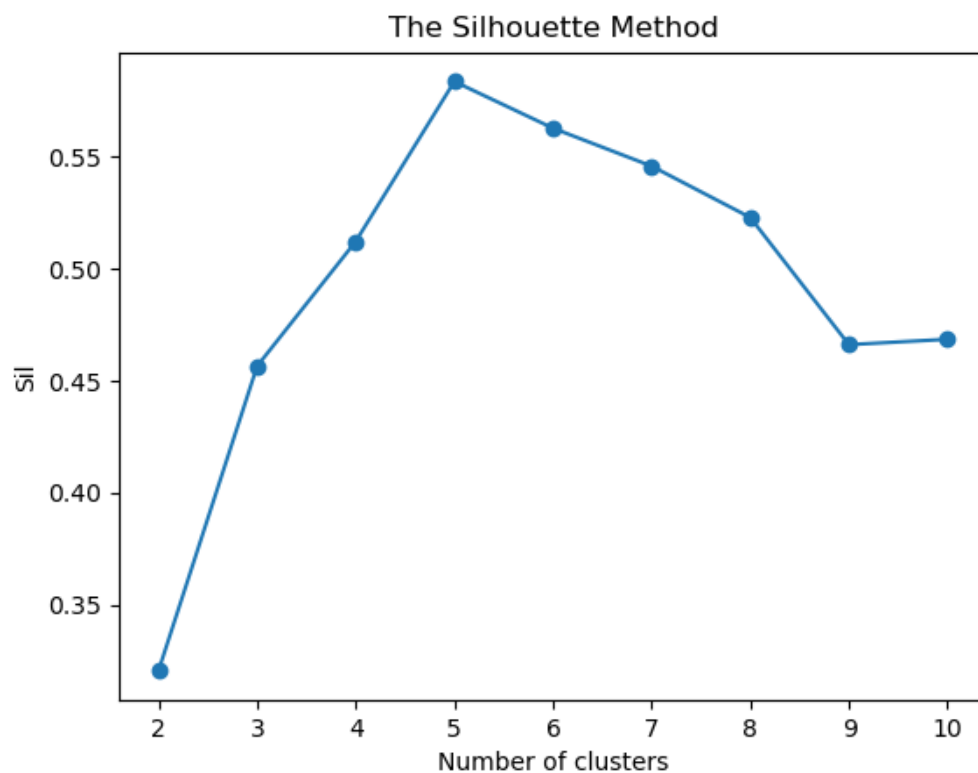
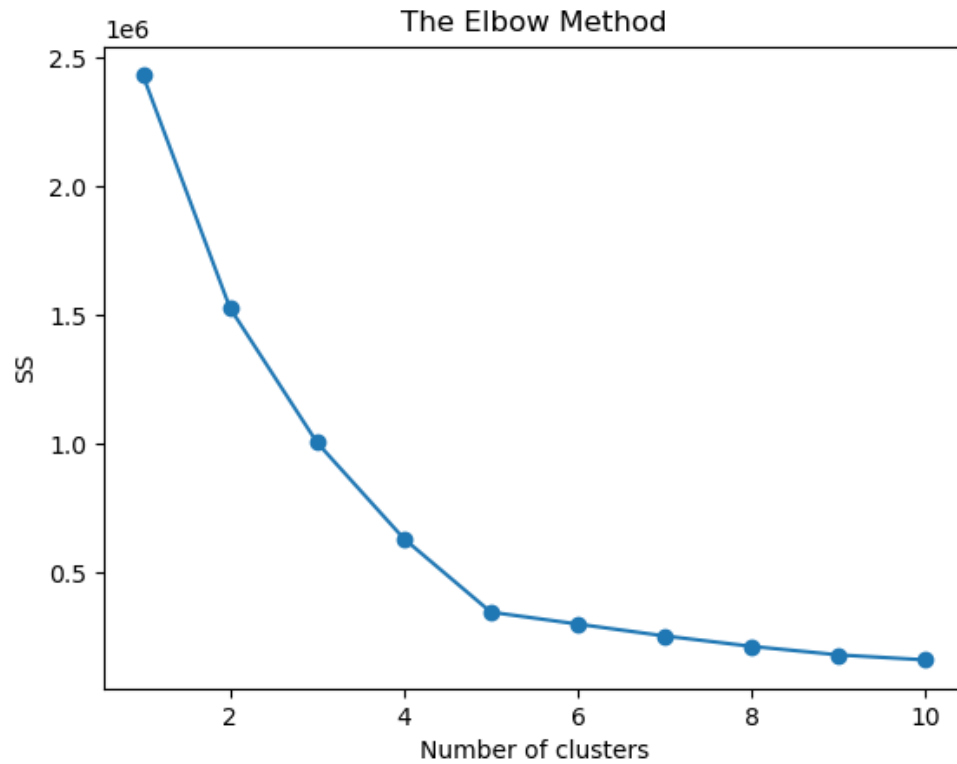
There is insignificant evidence to suggest relationship between age and loyalty ($p > 0.05$), CI had 0.



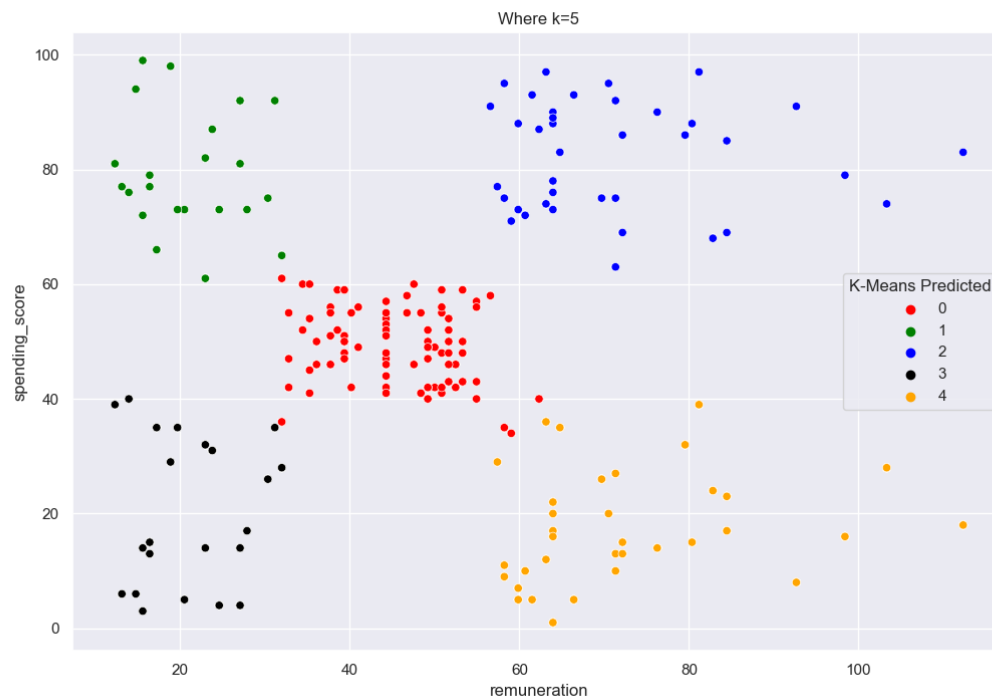
An MLR model with spending score and remuneration was created and tested using a split, train and test dataset. The regression test showed significant evidence to suggest relationship between spending, remuneration and loyalty ($p < 0.05$), CI did not have 0. Adjusted R-Squared tells us the model explains 82.3% of the variance. There was insignificant evidence to suggest multicollinearity. Given the high adjusted R-Squared value, the MLR model can be seen as the best to predict the future customer accumulation of loyalty points.

2. Different Target market segments

Before k-means clustering can be used, the number of clusters must be determined, this is done using the elbow and silhouette methods, both of which suggested a value of k between 4 and 6.



After testing all three, $k = 5$ yielded the most distinct and clusters. This suggests that customers can largely be placed in 5 groups depending on their remuneration and spending score.



Using the k -means model where $k = 5$, cluster 0 appears to be the largest as it can be seen as the typical customer group that has an average level of spending and remuneration. All other clusters are peripheral groups with similar counts of customers, it can be noted that those with a higher remuneration score appear to have a slightly larger number of observations.

Hence, it may be best to focus mostly on customers that fit within cluster 0 and those who tend to have a higher remuneration score if the goal is to target the highest numbers of customers.

Exploring further, it may be worth investigating if there was a 3rd factor moderating this clustering that could explain it.

3. Analysing customer reviews to inform marketing campaigns

Using Natural Language Processing (NLP) customer reviews were analysed. First, the data was prepared by changing to lower case, replacing punctuation, tokenisation, and removing stopwords. From this, we can see the word clouds for reviews and summaries.

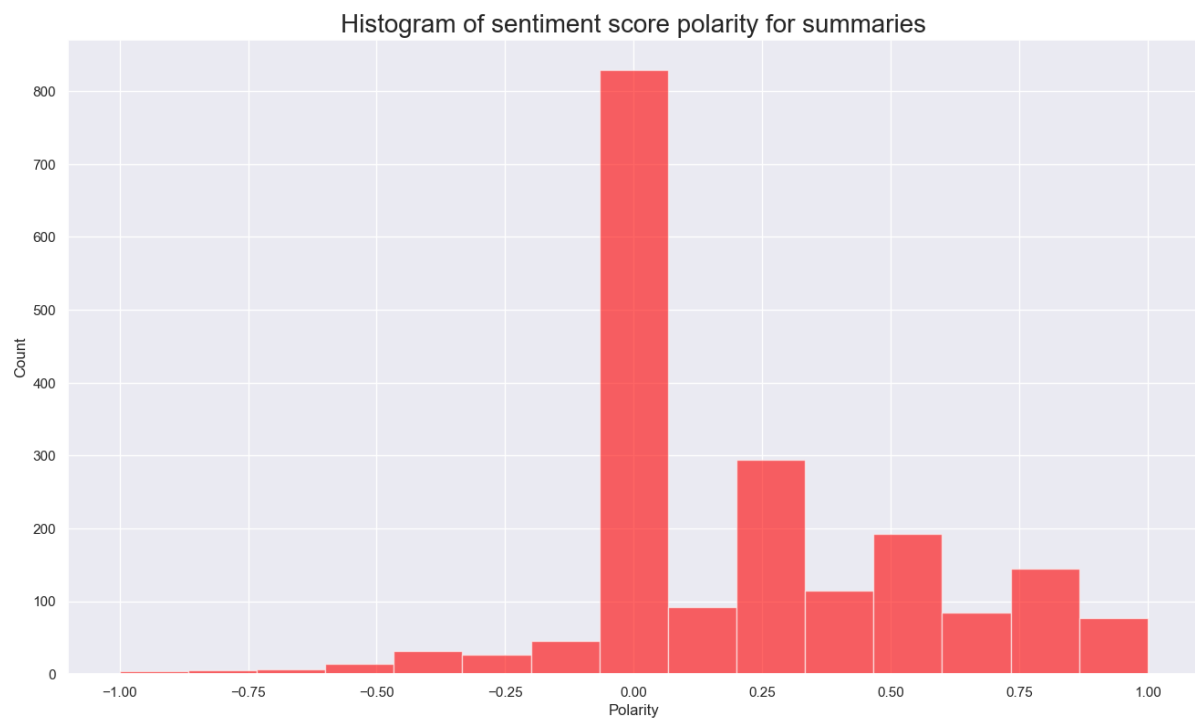
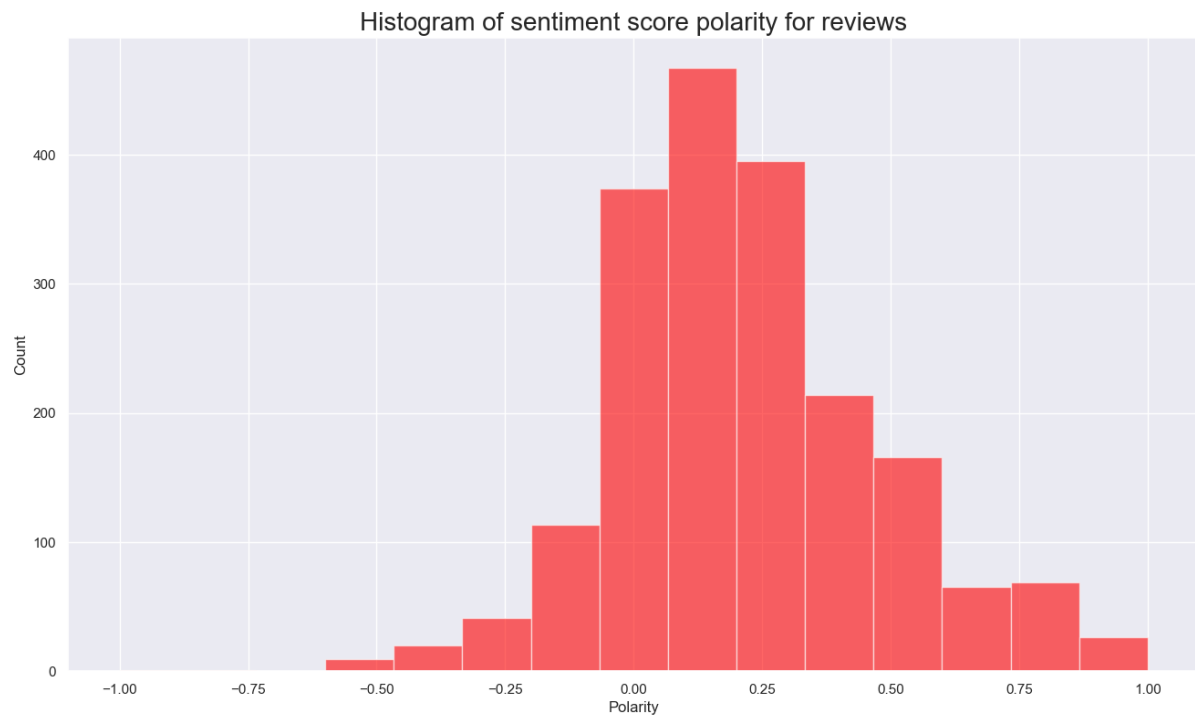
Wordcloud for Reviews



Wordcloud for Summaries



The histograms show a generally positive sentiment for both reviews and summaries with reviews especially having a higher frequency of results scoring positively for polarity. Summaries were also positive however a much larger number were neutral in sentiment.



The frequency of the most common words and the results of sentiment analysis suggest that games sold by Turtle Games are especially appealing to children and make great gifts. However, the negative sentiments suggest that the games tend to be "disappointing" or "boring" for some players.

Word	Frequency
stars	427
five	342
game	319
great	295
fun	218
love	93
good	92
four	58
like	54
expansion	52
kids	50
cute	45
book	43
one	38
awesome	36

Top 20 Negative Reviews

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
incomplete kit very disappointing	-0.78
im sorry i just find this product to be boring and to be frank juvenile	- 0.5833 3
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	-0.55
i bought this as a christmas gift for my grandson its a sticker book so how can i go wrong with this gift	-0.5
this was a gift for my daughter i found it difficult to use	-0.5
i found the directions difficult	-0.5
instructions are complicated to follow	-0.5
difficult	-0.5
expensive for what you get	-0.5
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	- 0.4916 7
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	- 0.4462 5

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	-0.44074
very hard complicated to make these	-0.43958
kids i work with like this game	-0.4
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	-0.4
my son loves playing this game it was recommended by a counselor at school that works with him	-0.4
this game is a blast	-0.4
i bought this for my son he loves this game	-0.4
was a gift for my son he loves the game	-0.4

Top 20 Negative Summaries

Summary	Polarity
the worst value ive ever seen	-1
boring unless you are a craft person which I am	-1
boring	-1
before this I hated running any rpg campaign dealing with towns because it	-0.9
another worthless dungeon masters screen from galeforce9	-0.8
disappointed	-0.75
disappointed	-0.75
disappointed	-0.75
disappointed	-0.75
promotes anger instead of teaching calming methods	-0.7
too bad this is not what I was expecting	-0.7
bad qualityall made of paper	-0.7
at age 31 I found these very difficult to make	-0.65
small and boring	-0.625
mad dragon	-0.625
disappointing	-0.6
disappointing	-0.6
disappointing	-0.6
disappointing	-0.6
then you will find this board game to be dumb and boring	-0.59167

Top 20 Positive Reviews

Review	Polarity
came in perfect condition	1
awesome book	1

awesome gift	1
excellent activity for teaching selfmanagement skills	1
perfect just what i ordered	1
wonderful product	1
delightful product	1
wonderful for my grandson to learn the resurrection story	1
perfect	1
awesome	1
awesome set	1
best set buy 2 if you have the means	1
awesome addition to my rpg gm system	1
its awesome	1
one of the best board games i played in along time	1
my daughter loves her stickers awesome seller thank you	1
this was perfect to go with the 7 bean bags i just wish they were not separate orders	1
awesome toy	1
it is the best thing to play with and also mind blowing in some ways	1
excellent toy to simulate thought	1

Top 20 Positive Summaries:

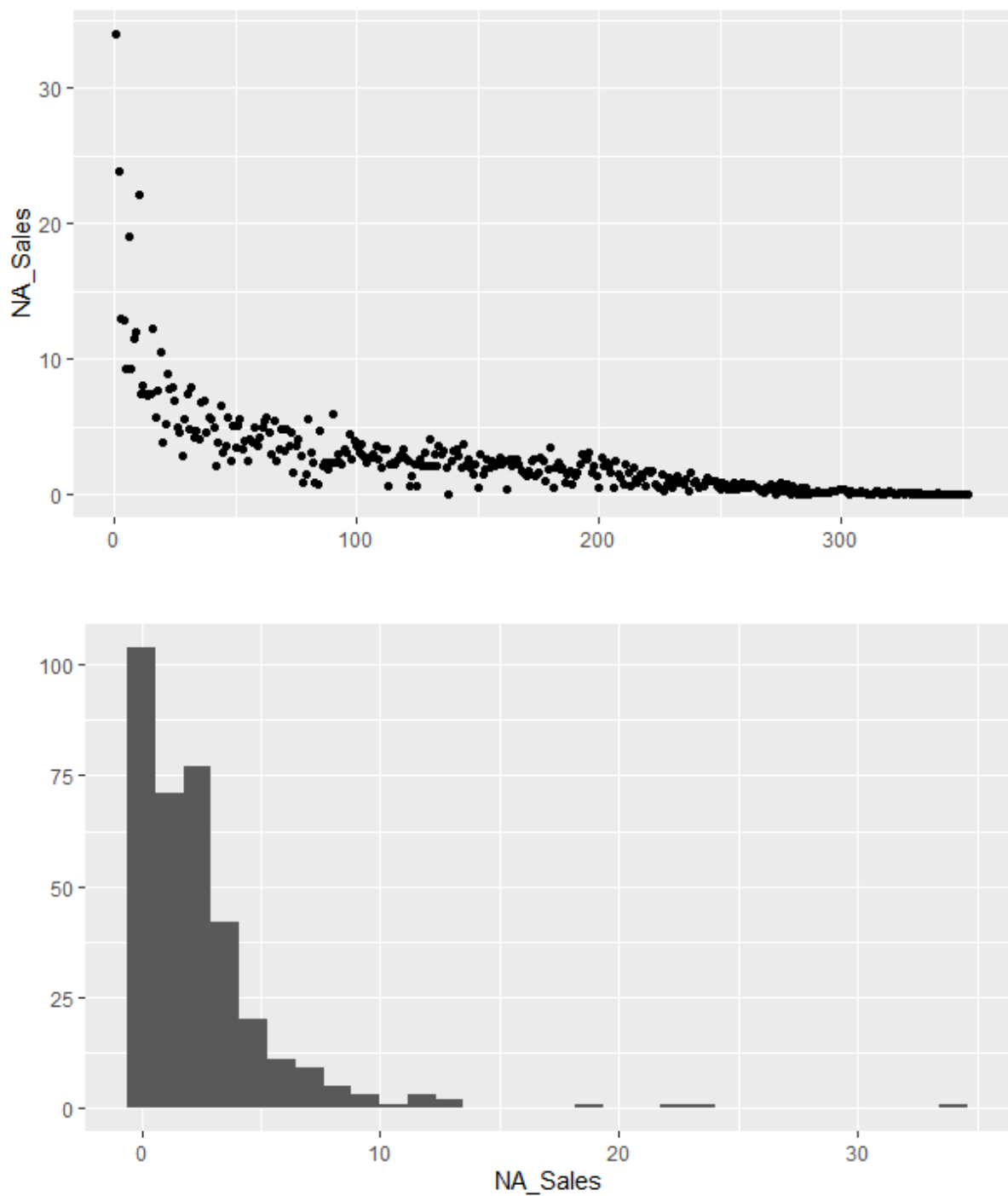
Summary	Polarity
best gm screen ever	1
wonderful designs	1
perfect	1
theyre the perfect size to keep in the car or a diaper	1
perfect for preschooler	1
awesome sticker activity for the price	1
awesome book	1
he was very happy with his gift	1
awesome	1
awesome and welldesigned for 9 year olds	1
perfect	1
excellent	1
excellent	1
excellent therapy tool	1
the pigeon is the perfect addition to a school library	1
best easter teaching tool	1
wonderful	1
all f the mudpuppy toys are wonderful	1
awesome puzzle	1
not the best quality	1

However, accuracy of these qualitative results can vary. Therefore, the results of this analysis are potentially more suited as supplementary results to support other results from more numerical methods.

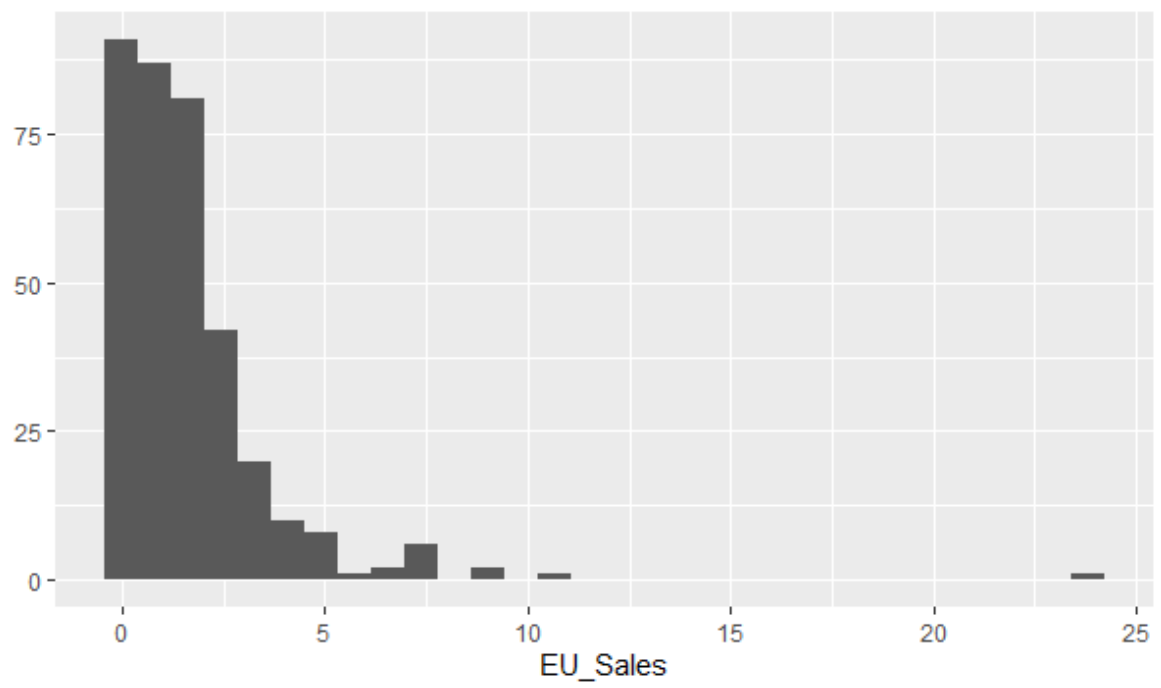
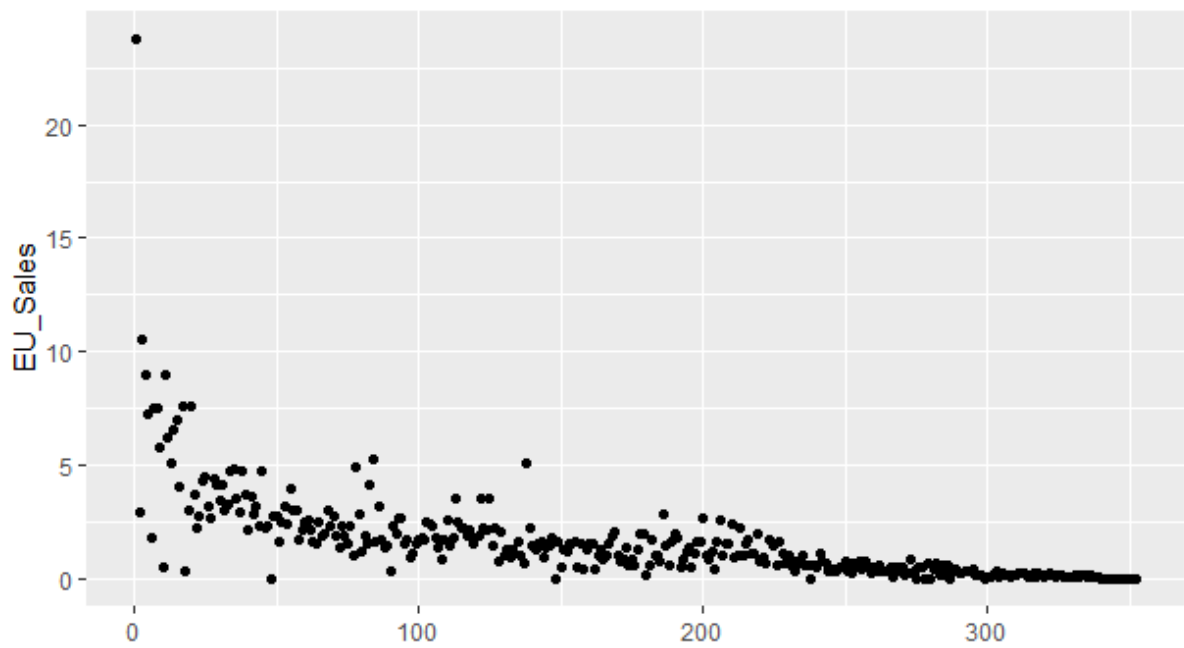
4. The sales based on different products

Scatterplots and histograms were used to investigate the sales of each product as well as their distribution. These show that most games tend to bring between USD 0 and 5 million in sales. NA tends to be a stronger performer than EU in terms of sales.

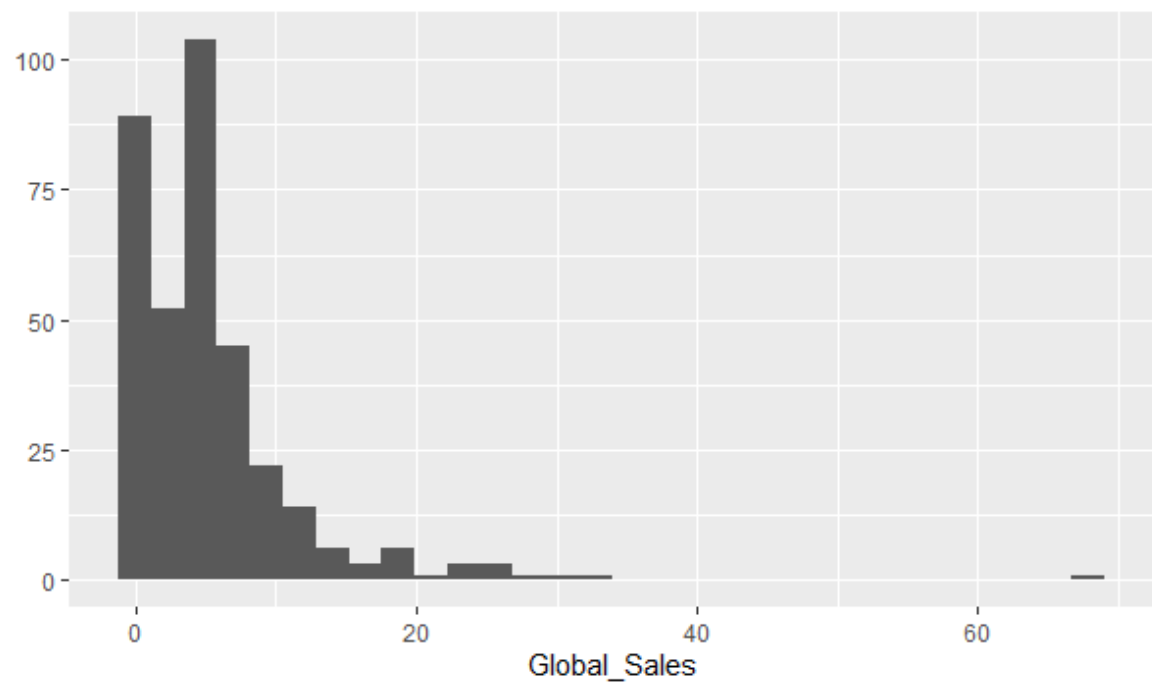
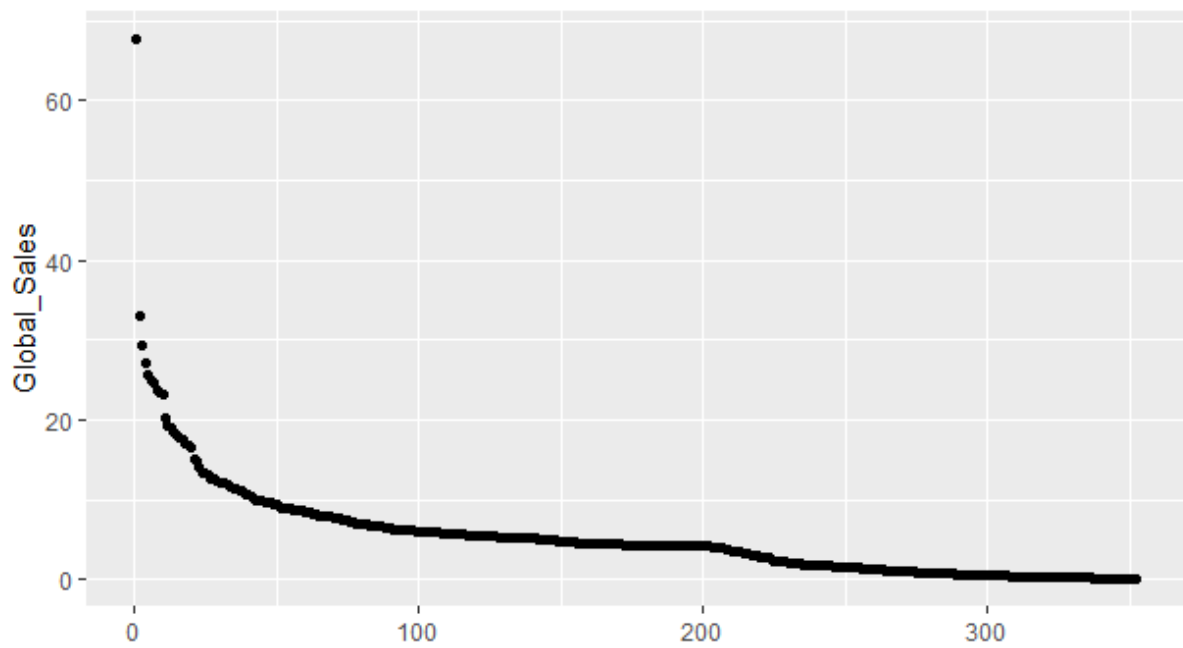
NA Sales:



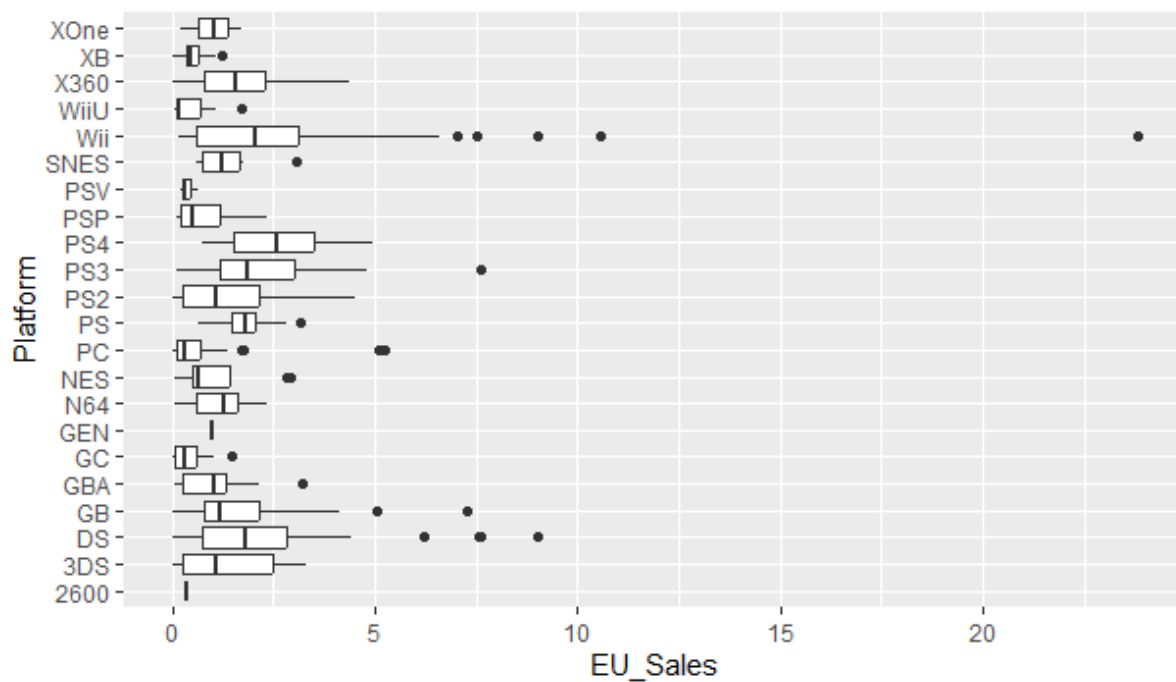
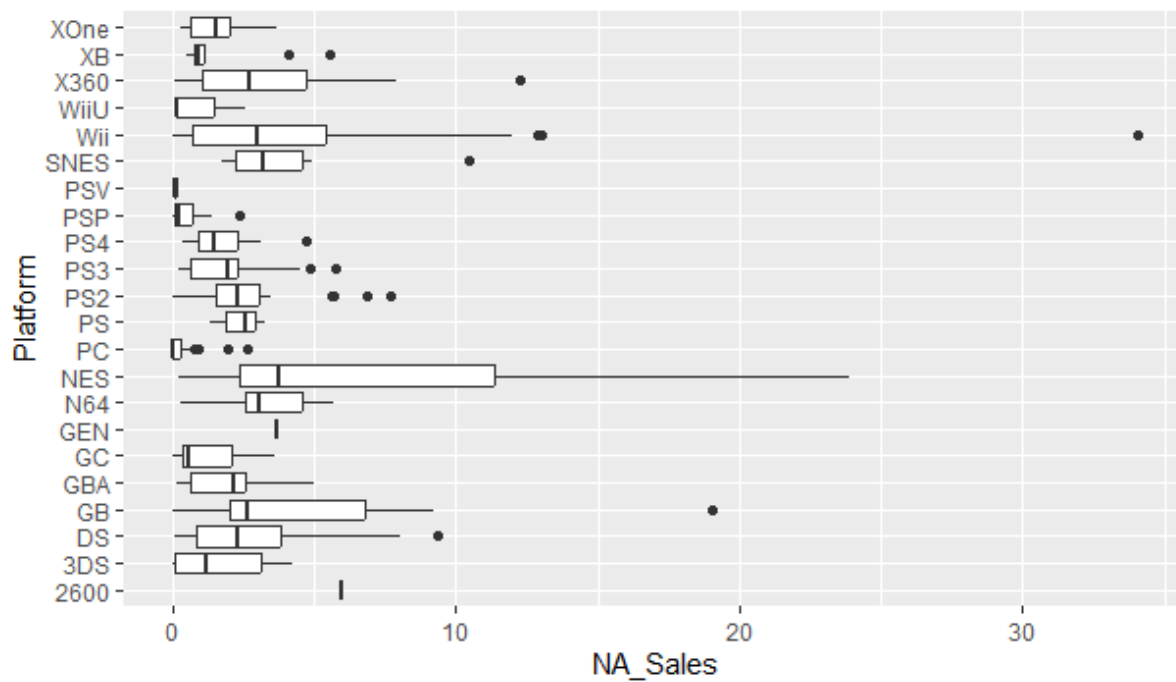
EU Sales:



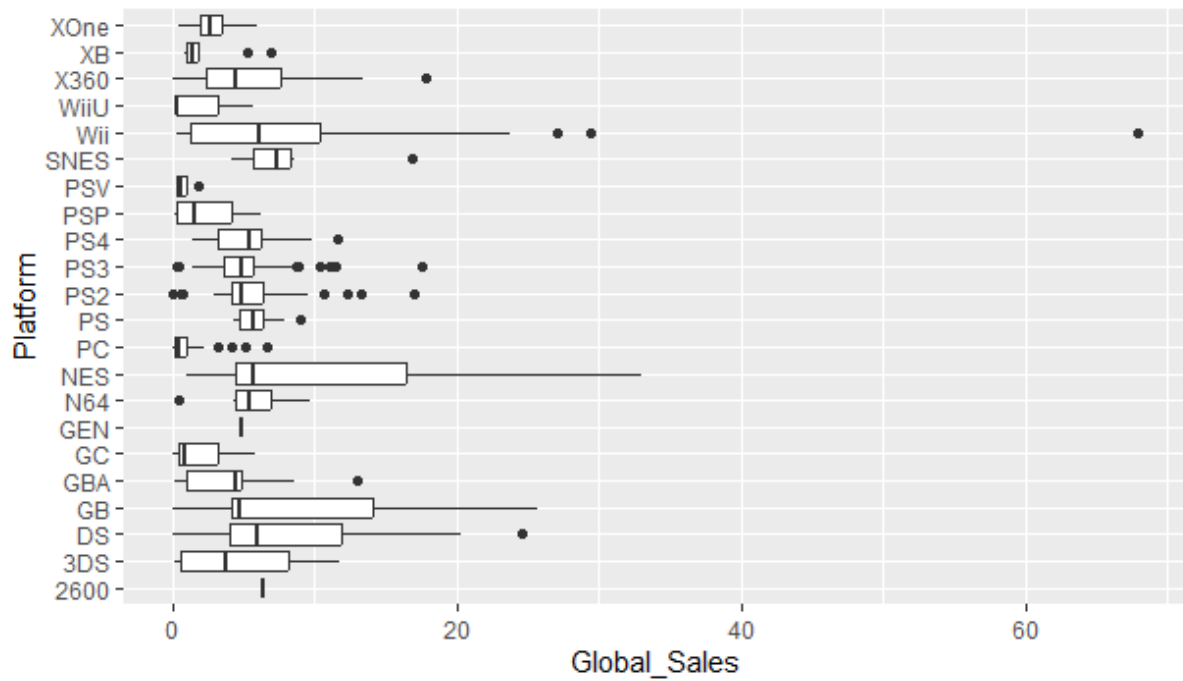
Global Sales:



Boxplots were created to find out the difference in sales for products depending on their console.



For NA, sales seem to show a preference for older consoles like the NES and GB as compared to the EU sales. Both show strong sales for Wii games.



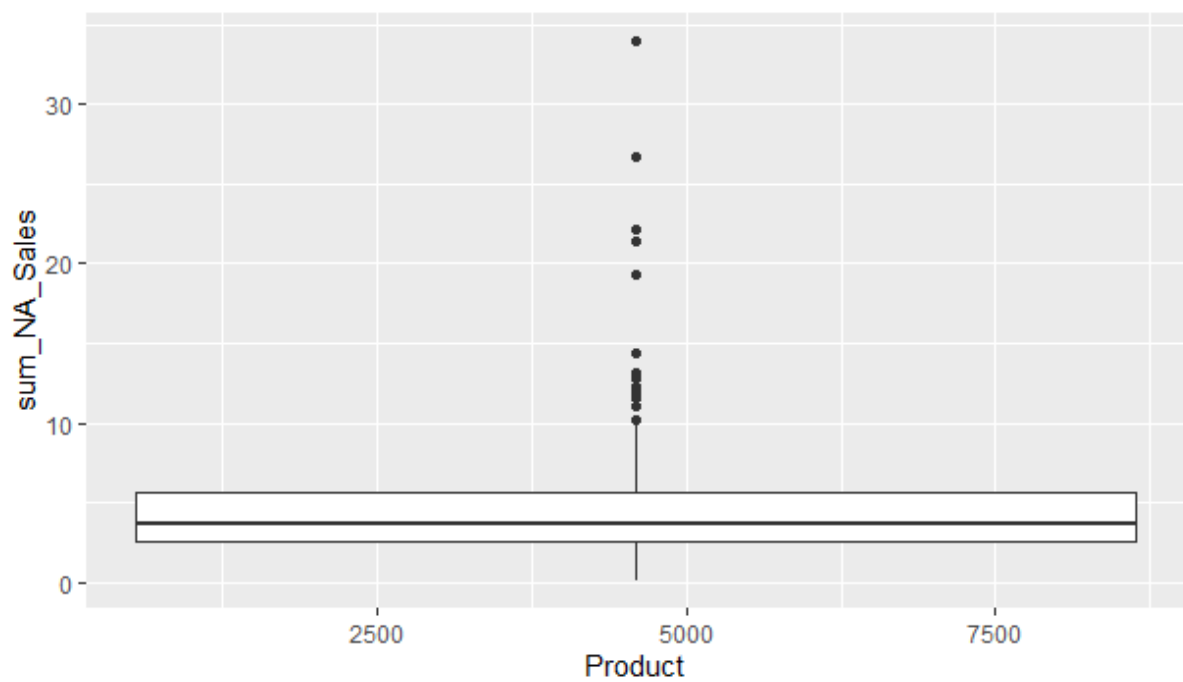
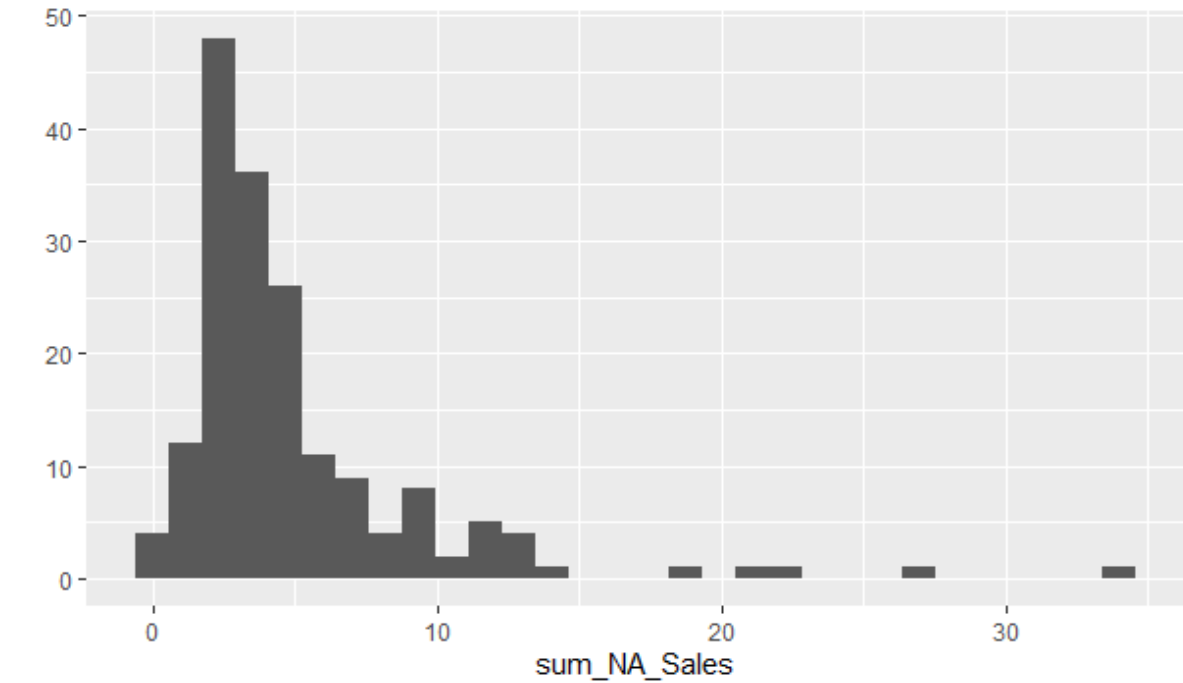
Global sales reflect this with Wii, NES, GB, and DS showing especially good sales. Hence, products using these platforms should perhaps be focused on to boost sales.

With further exploration, it may prove fruitful to look further into when each of these products were sold to see if there is a relationship between this and what consoles were used.

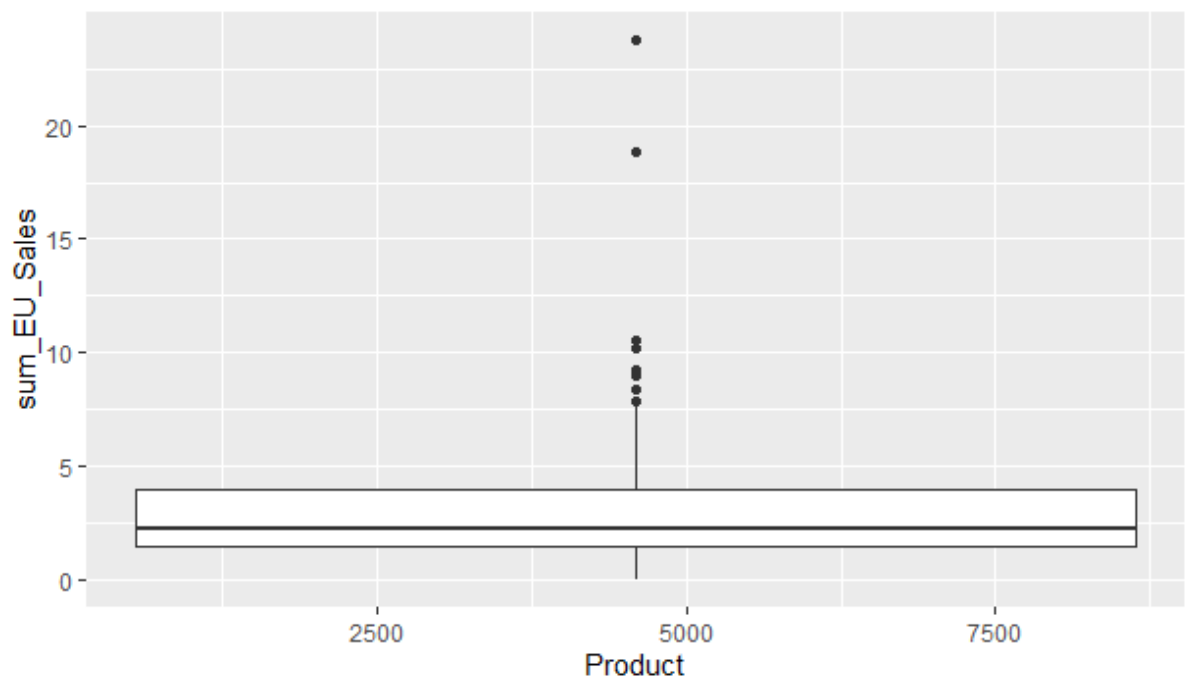
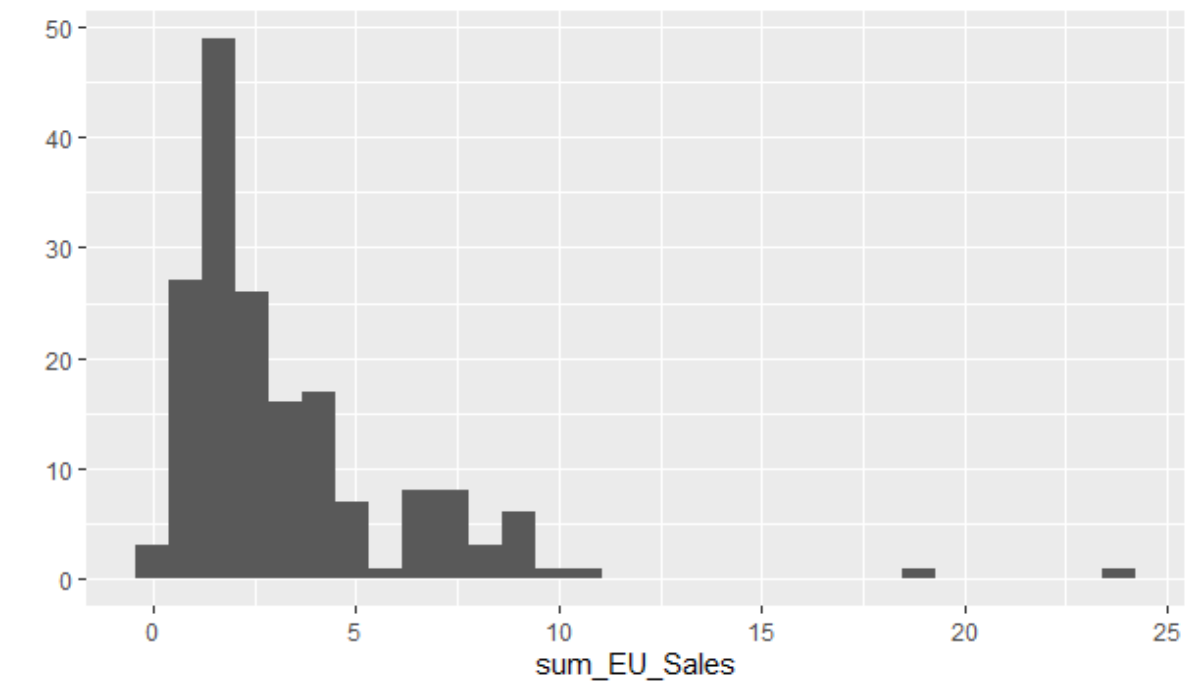
5. The reliability of Turtle Games' data

Histograms and boxplots were used to visualise the skewness and kurtosis of NA, EU and Global sales.

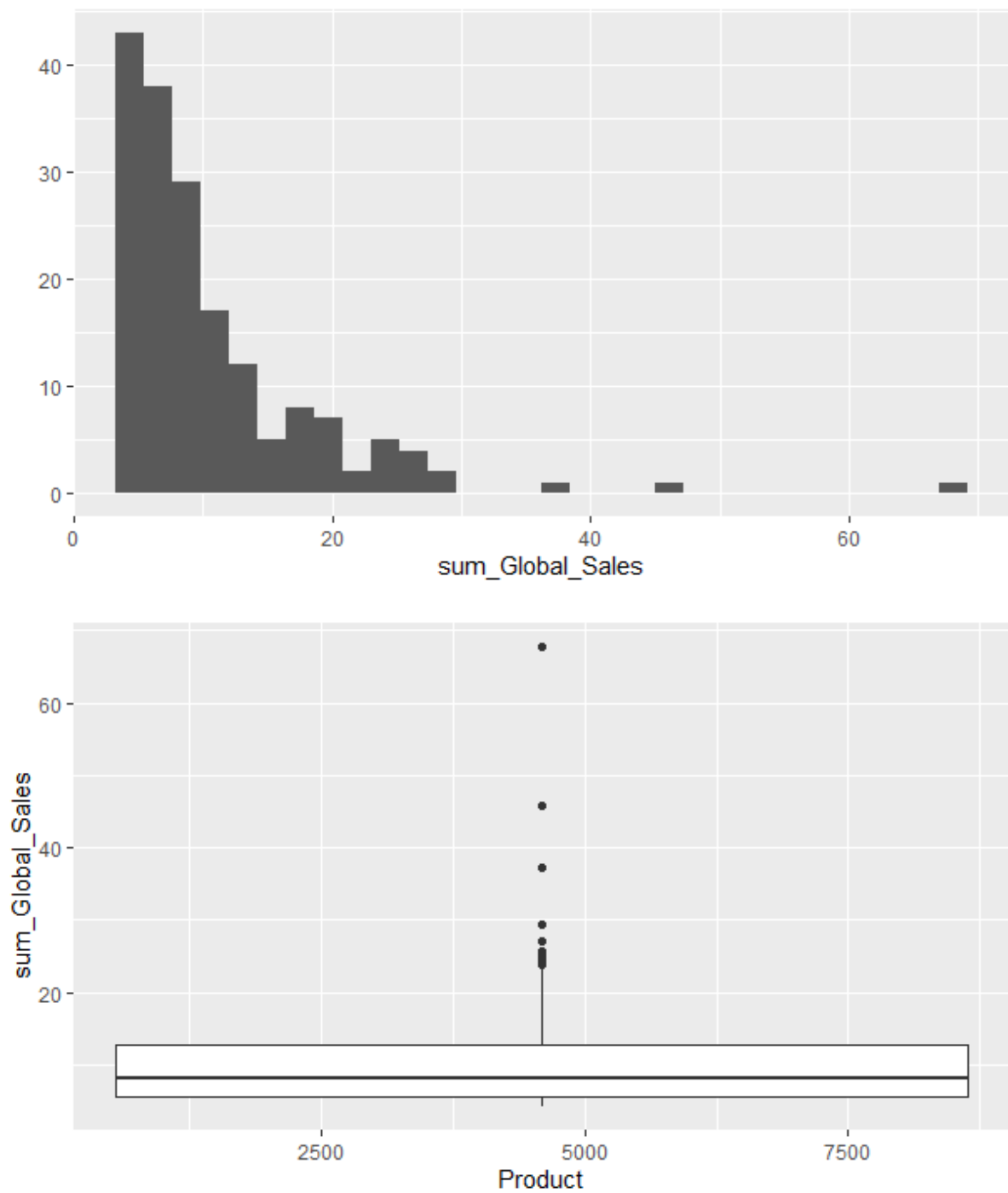
Sum of NA Sales



Sum of EU Sales



Sum of Global Sales



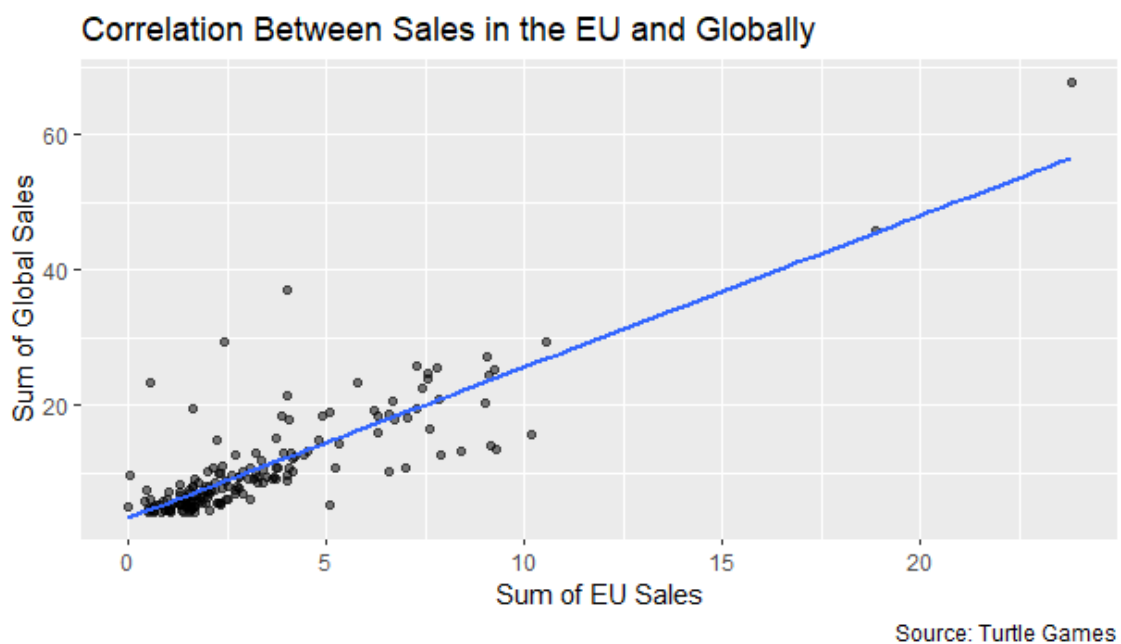
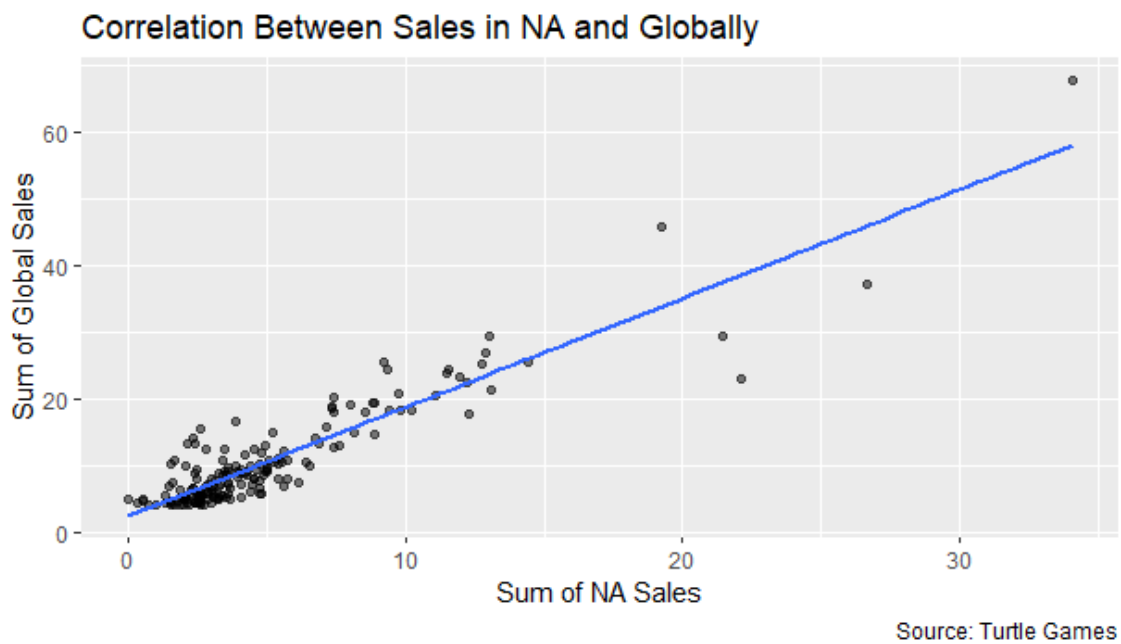
These show that all three variables had a very positively skewed and leptokurtic distribution as is reflected within their values for skewness and kurtosis.

Variable	Skewness	Kurtosis
sum_NA_Sales	3.048198	15.6026
sum_EU_Sales	2.886029	16.22554
sum_Global_Sales	3.066769	17.79072

The Shapiro-Wilk test is <0.05 for all three sales variables suggesting that they are not normally distributed. After testing the impact of removing outliers, they did not seem to impact the result of the analysis nor the outcomes of the tests. Therefore, the outliers were not omitted.

Results show a strong to very strong positive correlation between the sales variables especially between NA and Global sales. This suggests that NA sales potentially have a greater impact on the outcome of Global sales compared to EU sales.

Relationships	Correlations
NA - Global	0.92
EU - Global	0.85

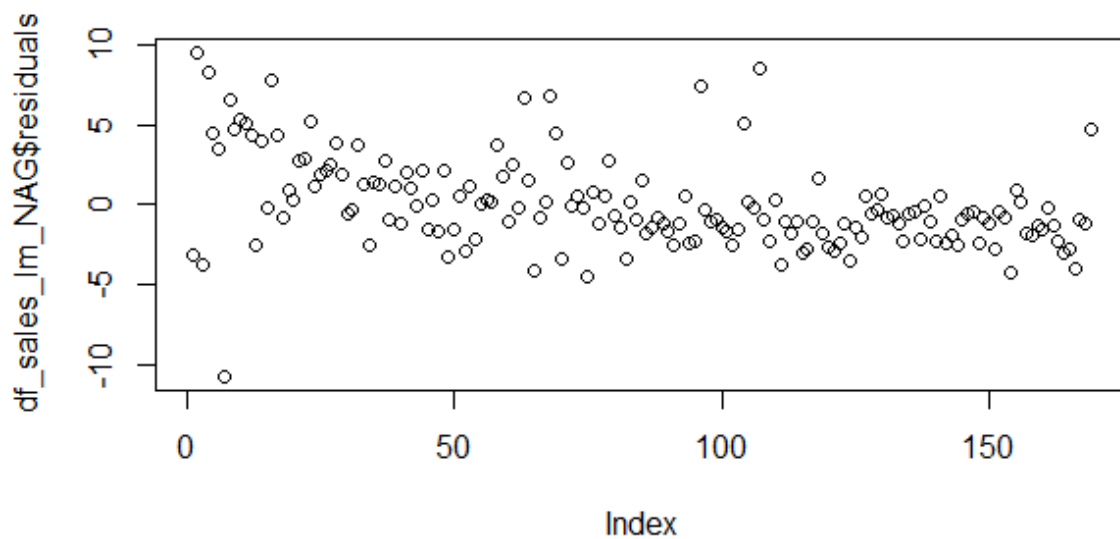
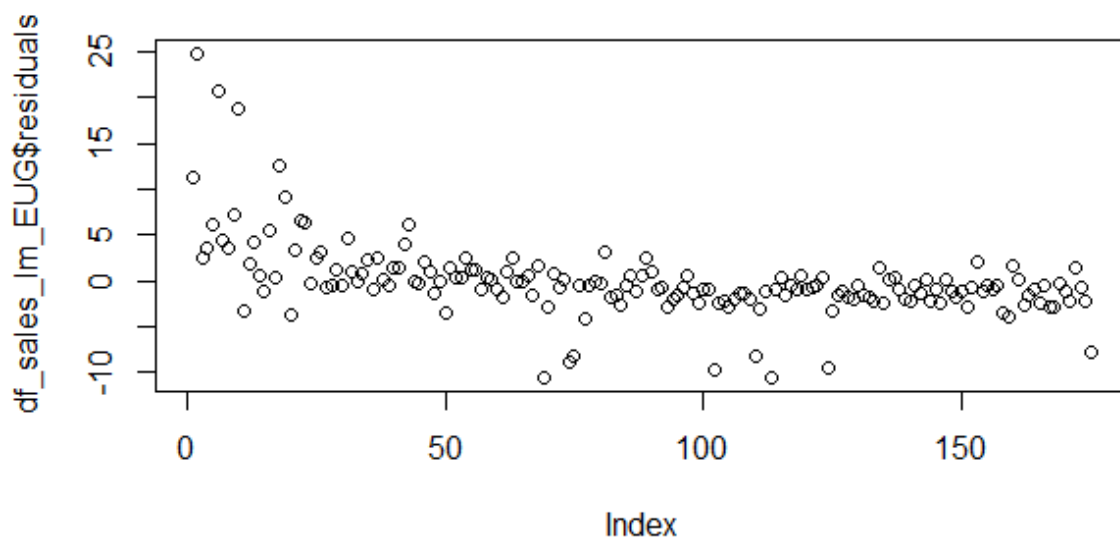


6. The relationships between sales in NA, in EU, and Globally

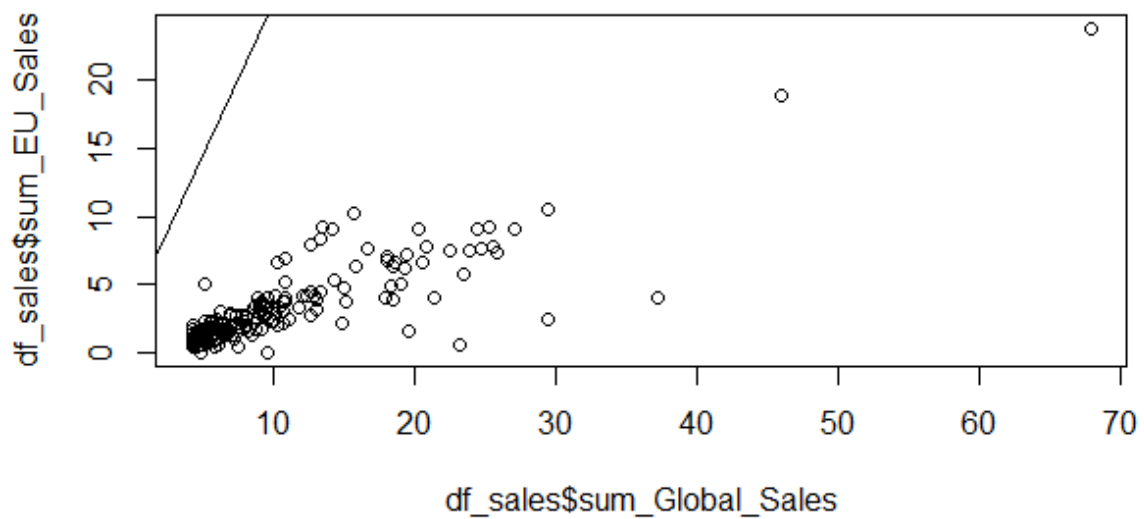
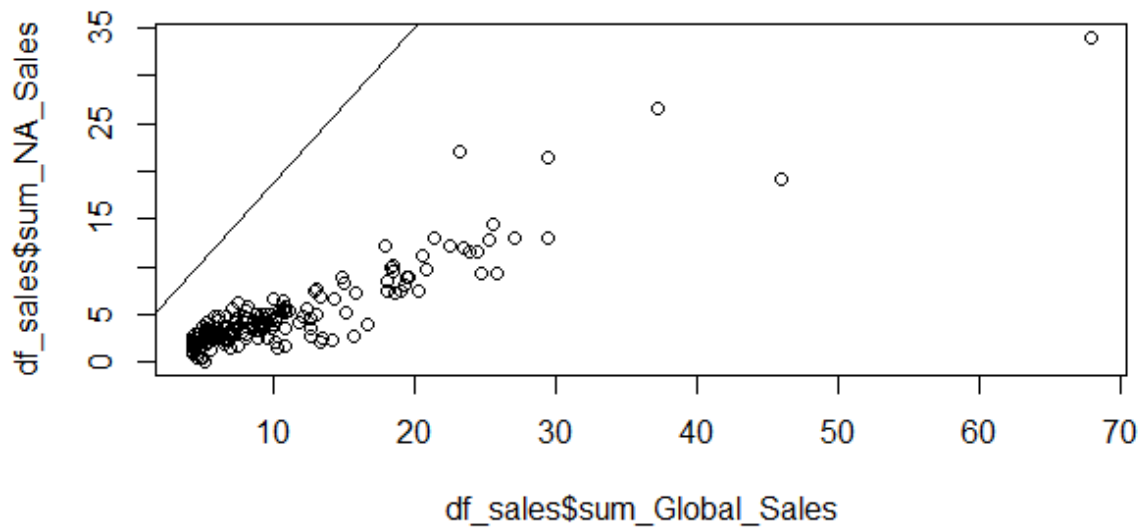
Simple Linear Regression Models yielded the following R-Squared Values.

Predictor for Sum of Global Sales	R-Squared Value
sum_NA_Sales	0.8395
sum_EU_Sales	0.7201

Slight pattern on residuals for sum_EU_Sales simple linear regression model, but no clear pattern on residuals for sum_NA_Sales model.



Neither of the simple linear regression models seem to be adequate as shown by the line-of-best-fits which are not in line with data points at all.



MLR model using both sum of NA and EU sales had adjusted R-Squared value of 0.9664 which is an extremely good fit.

Predictions were then made using the Multiple Linear Regression Model to test accuracy.

Predicted value	Observed value
68.06	67.85
7.36	6.04
4.91	4.32
4.76	3.53
26.63	23.2

The predicted values are relatively close to the observed values. Hence, the MLM model using both the sum of NA and EU sales is a very good predictor of global sales.