Turtle Games Analysis: Predicting Future Outcomes

Executive Summary

Prepared for Fiona Chucks, Head of Marketing Department and Lucas Wood, Head of Sales Department, at Turtle Games

Prepared by Alejandra Anguil Vanier, Data Analyst at dAtaAnalysisAV

April 2023

Contents:

Background and Context

Turtle Games is a game manufacturer and retailer. They manufacture and sell their own products, along with sourcing and selling products manufactured by other companies. Their product range includes books, board games, video games and toys. They have a global customer base and have a <u>business objective of improving overall sales performance</u> by utilizing customer trends.

dAtaAnalysisAV has been tasked to help inform Turtle Games how best to achieve this goal. The stakeholders workflow and the main areas of analysis are shown below.

Stakeholders Workflow

Problem-solving Framework and Potential Causes

Analytical Approach

As part of the brief, dAtaAnalysisAV received three files:

- turtle_reviews.csv – analysed in Python (marketing department)
- turtle_sales.csv – analysed in R (sales department)
- metadata_turtle_games.txt – datasets details

1. Data Ingestion

We first proceeded to import the data into Python using the Pandas library and into R using the tidyverse library, on both cases with the correct format and names (descriptive names will make the next steps easier to follow).

2. Describe the Data

To make sure the data has been loaded correctly and familiarise ourselves with it, we performed a descriptive statistical analysis (info(), describe(), isnull().sum() in Python; class(), glimpse(), summary(), sum(is.na()) in R), which helps us to describe the characteristics of the datasets in a meaningful way, and potentially spotting errors, unwanted columns, missing values, etc.

turtle_reviews	turtle_sales
gender	Ranking
age	Product
remuneration (k£)	Platform
spending_score (1-100)	Year
loyalty_points	Genre
education	Publisher
language	NA_Sales
platform	EU_Sales
product	Global_Sales
review	
summary	

Original Column Distribution for both Datasets

3. Data Wrangling

We proceeded to clean the data in Python by:

- a. Identifying and removing unwanted columns with the drop() function.
- b. Identifying duplicates in Excel on the basis of duplications in all columns: no duplicates found.
- c. Fixing inconsistencies by renaming the columns with the rename() function.

We proceeded to clean the data in R by:

a. Making sure each column is recorded with the correct data type using mutate() and as.factor() functions.

b. Identifying and removing unwanted columns with the select() function. Using the same function we also added Other_Sales column using the information provided in the metadata file:
    Global_Sales – (NA_Sales + EU_Sales).

4. Predictions and Sentiment Analysis in Python (Marketing)

We performed various prediction exercises to respond the questions posed by the marketing team while helping them build strategies to reach Turtle Games goal.

a. Linear Regression: to perform linear regression there are six assumptions that need to be checked to determine the accuracy of a predictive model: exogeneity (independent variables are not dependent on the dependent variable), independent residuals (Durbin-Watson test), outliers or missing values (using boxplots and quartiles), normality of y (using histograms), linearity (using scatterplots) and homoscedasticity (Breusch-Pagan test).

b. Clustering: using scatterplots, pairplots, and the Elbow and Silhouette methods we determined the number of clusters, and checked three different number of clusters to make sure our final choice was the best one.

c. Sentiment Analysis: there are various steps to prepare the data to perform this type of analysis, including changing all words to lower case, removing punctuation and removing stopwords.

5. Exploratory Data Analysis and Predictions in R (Sales)

We performed an exploratory analysis in R to understand the sales data by:

a. Evaluating the impact on sales per product, platforms, genres and publishers.

b. Evaluating the reliability of the data.

c. Looking for relationships between the sales columns.

d. Assumptions: we have identified outliers in the sales data that correspond to products with sales above £20M. We have proceeded with the analysis keeping what we can call the best selling products. However, please can we get confirmation from Turtle Games that this is the case and not an error on the data?

Visualisation and insights

This step aims to create a series of visualisations to present the uncovered insights to the stakeholders. The visualisations have been chosen carefully to be simple to understand while conveying the message intended. The colour palette is consistent and avoids any colours that could be challenging for people with a colour vision deficiency.

These visualisations and insights try to respond the initial queries posed by Turtle Games:

MARKETING

1. How customers accumulate loyalty points? Relationship between loyalty points and age/ remuneration/ spending score.

   We have disregarded the variable age, as it shows no relationship with loyalty points.

   The other two variables (remuneration and spending score) don't work well either in isolation, but working together in a Multiple Linear Regression model they become significant and can predict the loyalty points with an accuracy of 82%. However, we have to acknowledge the large errors this model can have and we would recommend to collect more data to improve these errors and try to include other variables to check the performance of the model.

Customers Loyalty Points: Predicted vs Observed values
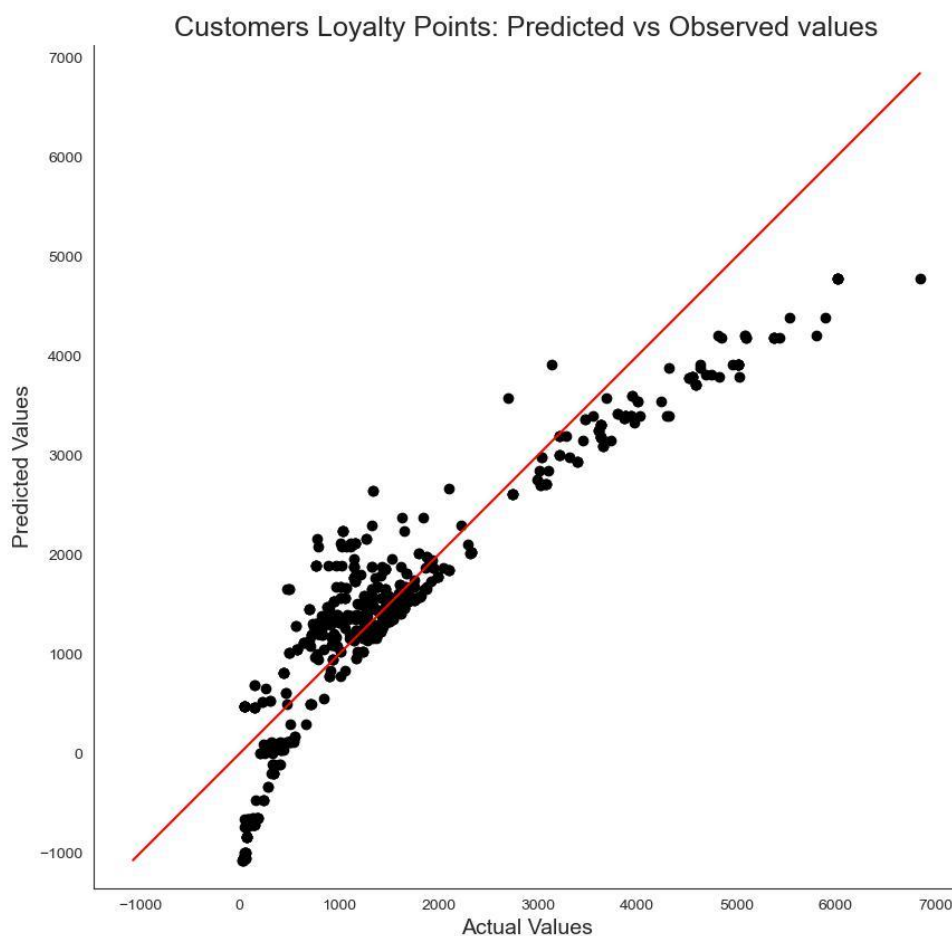
2. How groups within the customer base can be used to target specific market segments? Usefulness of renumeration and spending score.

Fitting a clustering model we have been able to organise Turtle Games customers by remuneration and spending score, creating five groups that the marketing department can use to target specific market segments:
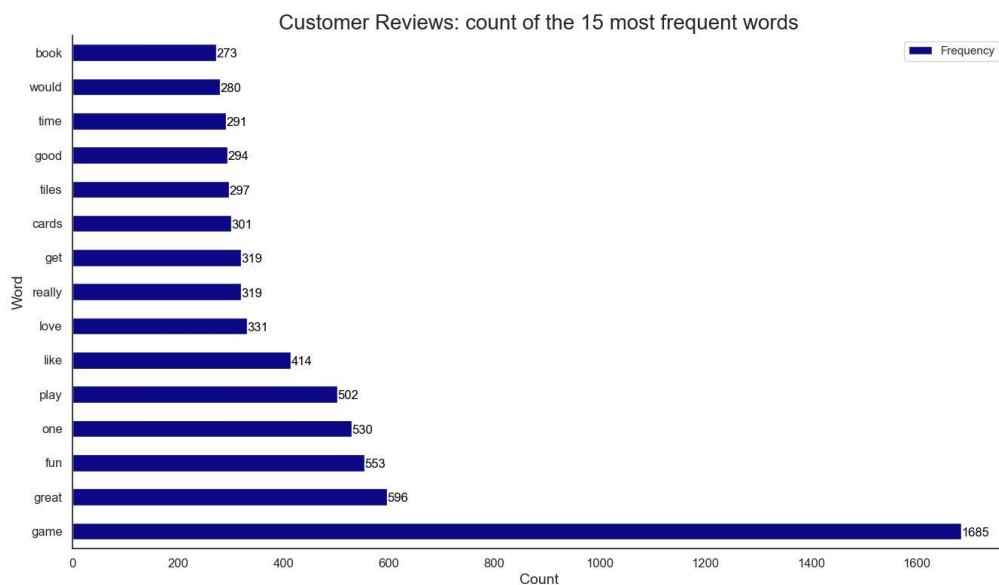
- low remuneration/ low spending score

- low remuneration/ high spending score

- medium remuneration/ medium spending score

- high remuneration/ low spending score

- high remuneration/ high spending score

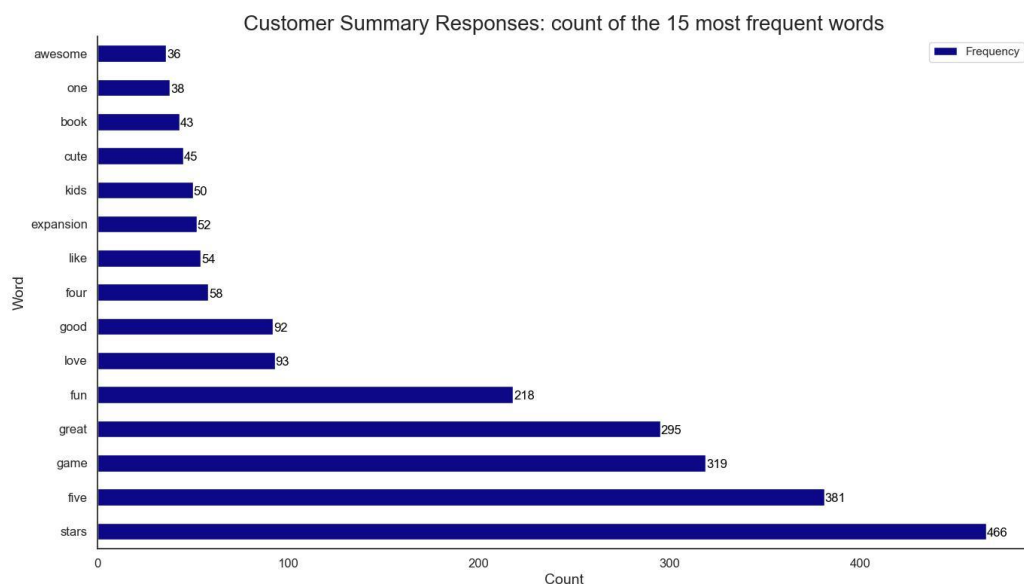Distribution of Customers with 5 Clusters
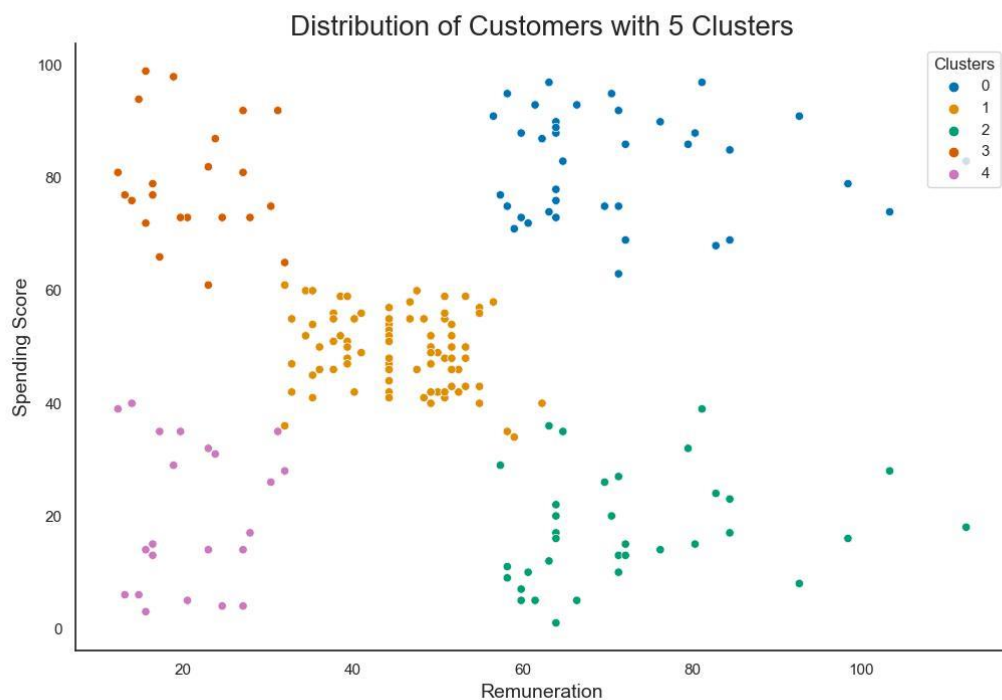
3. How social data can be used to inform marketing campaigns?

    We are providing here the information the marketing team has requested. However, we have also identified that the reviews tend to be more subjective than the summaries of the reviews, so we would recommend to extract the information from the summaries.

    -   15 most common words used in online product reviews:

Customer Reviews: count of the 15 most frequent words

Word	Frequency
book	273
would	280
time	291
good	294
tiles	297
cards	301
get	319
really	319
love	331
like	414
play	502
one	530
fun	553
great	596
game	1685

    -   15 most common words used in online product summaries:

Customer Summary Responses: count of the 15 most frequent words

Word	Frequency
awesome	36
one	38
book	43
cute	45
kids	50
expansion	52
like	54
four	58
good	92
love	93
fun	218
great	295
game	319
five	381
stars	466

- Top 20 positive and negative reviews from the website:

	review	polarity_review	subjectivity_review
7	came in perfect condition	1.000000	1.000000
165	awesome book	1.000000	1.000000
194	awesome gift	1.000000	1.000000
496	excellent activity for teaching selfmanagement skills	1.000000	1.000000
524	perfect just what i ordered	1.000000	1.000000
591	wonderful product	1.000000	1.000000
609	delightful product	1.000000	1.000000
621	wonderful for my grandson to learn the resurrection story	1.000000	1.000000
790	perfect	1.000000	1.000000
933	awesome	1.000000	1.000000
1037	awesome	1.000000	1.000000
1135	awesome set	1.000000	1.000000
1168	best set buy 2 if you have the means	1.000000	0.300000
1177	awesome addition to my rpg gm system	1.000000	1.000000
1301	its awesome	1.000000	1.000000
1401	one of the best board games i played in along time	1.000000	0.300000
1550	my daughter loves her stickers awesome seller thank you	1.000000	1.000000
1609	this was perfect to go with the 7 bean bags i just wish they were not separate orders	1.000000	1.000000
1715	awesome toy	1.000000	1.000000
1720	it is the best thing to play with and also mind blowing in some ways	1.000000	0.300000

|	review	polarity_review	subjectivity_review
208	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	1.000000
182	incomplete kit very disappointing	-0.780000	0.910000
1804	im sorry i just find this product to be boring and to be frank juvenile	-0.583333	0.750000
364	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
117	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
227	this was a gift for my daughter i found it difficult to use	-0.500000	1.000000
230	i found the directions difficult	-0.500000	1.000000
290	instructions are complicated to follow	-0.500000	1.000000
301	difficult	-0.500000	1.000000
1524	expensive for what you get	-0.500000	0.700000
174	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
347	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
538	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
306	very hard complicated to make these	-0.439583	0.852083
427	kids i work with like this game	-0.400000	0.400000
437	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
497	my son loves playing this game it was recommended by a counselor at school that works with him	-0.400000	0.400000
803	this game is a blast	-0.400000	0.400000
806	i bought this for my son he loves this game	-0.400000	0.400000
824	was a gift for my son he loves the game	-0.400000	0.400000

- Top 20 positive and negative summaries from the website:

	summary	polarity_summary	subjectivity_summary
6	best gm screen ever	1.000000	0.300000
28	wonderful designs	1.000000	1.000000
32	perfect	1.000000	1.000000
80	theyre the perfect size to keep in the car or a diaper	1.000000	1.000000
134	perfect for preschooler	1.000000	1.000000
140	awesome sticker activity for the price	1.000000	1.000000
161	awesome book	1.000000	1.000000
163	he was very happy with his gift	1.000000	1.000000
187	awesome	1.000000	1.000000
210	awesome and welldesigned for 9 year olds	1.000000	1.000000
418	perfect	1.000000	1.000000
475	excellent	1.000000	1.000000
543	excellent	1.000000	1.000000
548	excellent therapy tool	1.000000	1.000000
580	the pigeon is the perfect addition to a school library	1.000000	1.000000
599	best easter teaching tool	1.000000	0.300000
647	wonderful	1.000000	1.000000
651	all f the mudpuppy toys are wonderful	1.000000	1.000000
657	awesome puzzle	1.000000	1.000000
662	not the best quality	1.000000	0.300000

|	summary	polarity_summary	subjectivity_summary
21	the worst value ive ever seen	-1.000000	1.000000
208	boring unless you are a craft person which i am	-1.000000	1.000000
829	boring	-1.000000	1.000000
1166	before this i hated running any rpg campaign dealing with towns because it	-0.900000	0.700000
1	another worthless dungeon masters screen from galeforce9	-0.800000	0.900000
144	disappointed	-0.750000	0.750000
631	disappointed	-0.750000	0.750000
793	disappointed	-0.750000	0.750000
1620	disappointed	-0.750000	0.750000
363	promotes anger instead of teaching calming methods	-0.700000	0.200000
885	too bad this is not what i was expecting	-0.700000	0.666667
890	bad qualityall made of paper	-0.700000	0.666667
178	at age 31 i found these very difficult to make	-0.650000	1.000000
101	small and boring	-0.625000	0.700000
518	mad dragon	-0.625000	1.000000
805	disappointing	-0.600000	0.700000
1015	disappointing	-0.600000	0.700000
1115	disappointing	-0.600000	0.700000
1804	disappointing	-0.600000	0.700000
1003	then you will find this board game to be dumb and boring	-0.591667	0.633333

SALES

4. What is the impact on sales per product?

The total sales per region is:

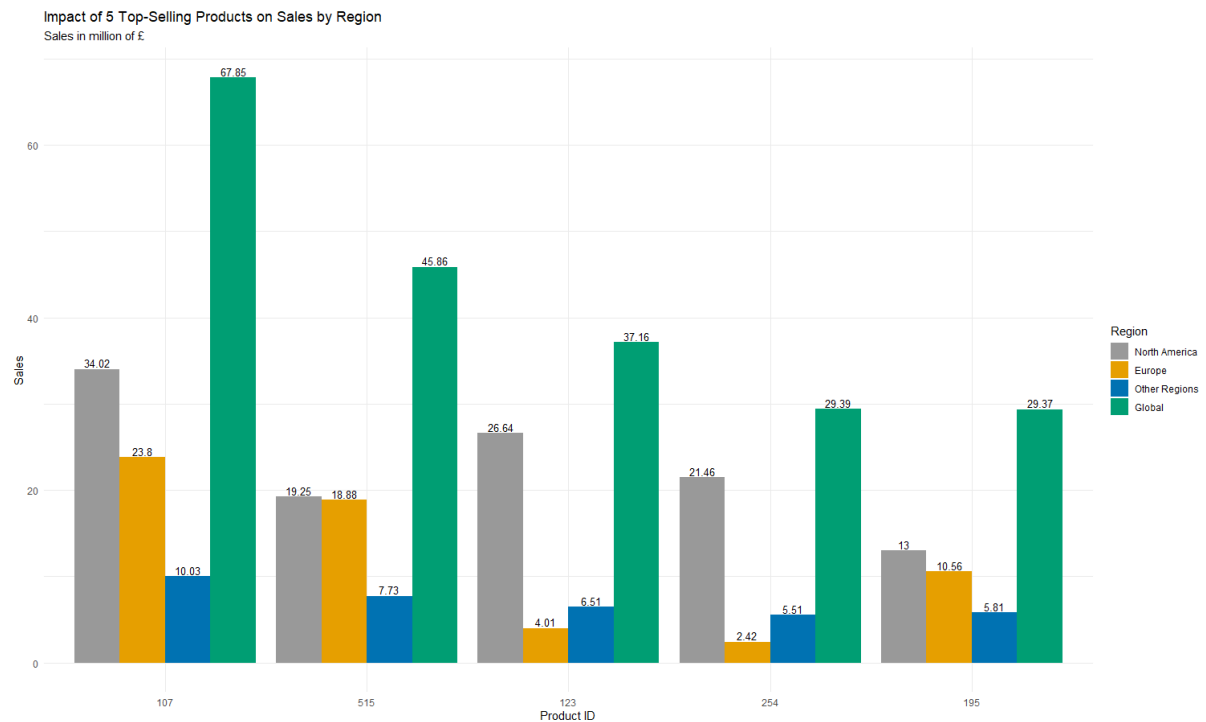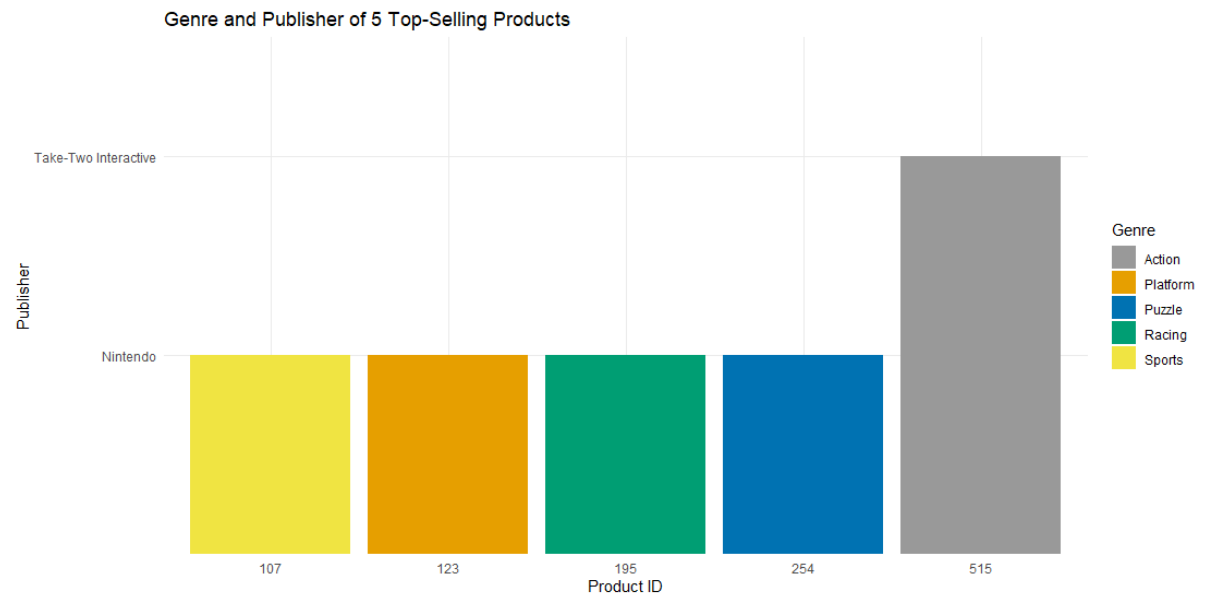Total Sales across Regions and Globally
Sales in million of £
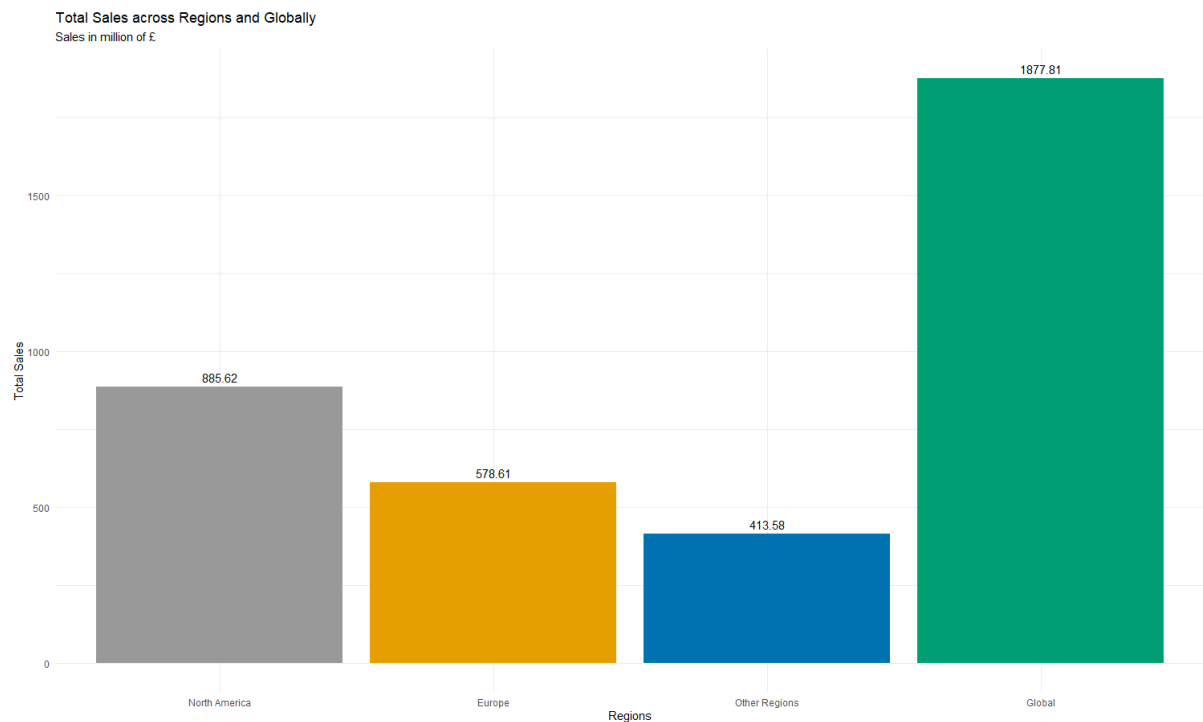
- Popular products across regions?

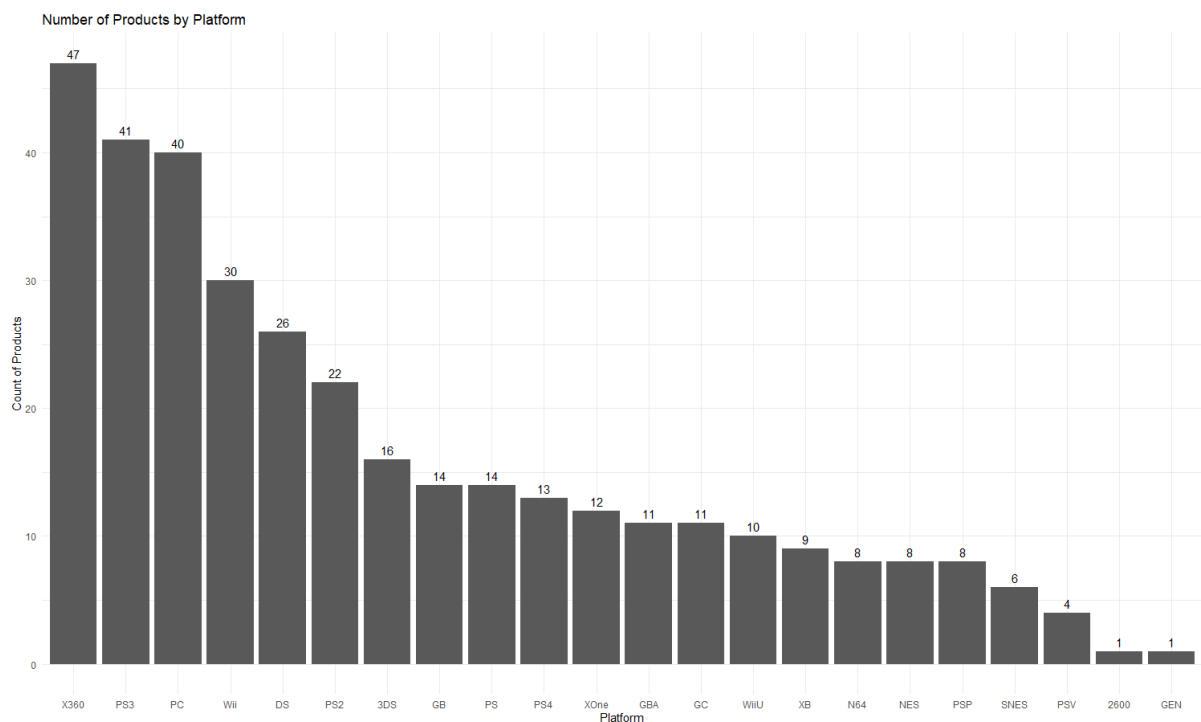There are 175 products. The ones with Global Sales bigger than £20M are:

Impact on Sales by Region of Products with Glabal Sales bigger than £20M
Sales in million of £
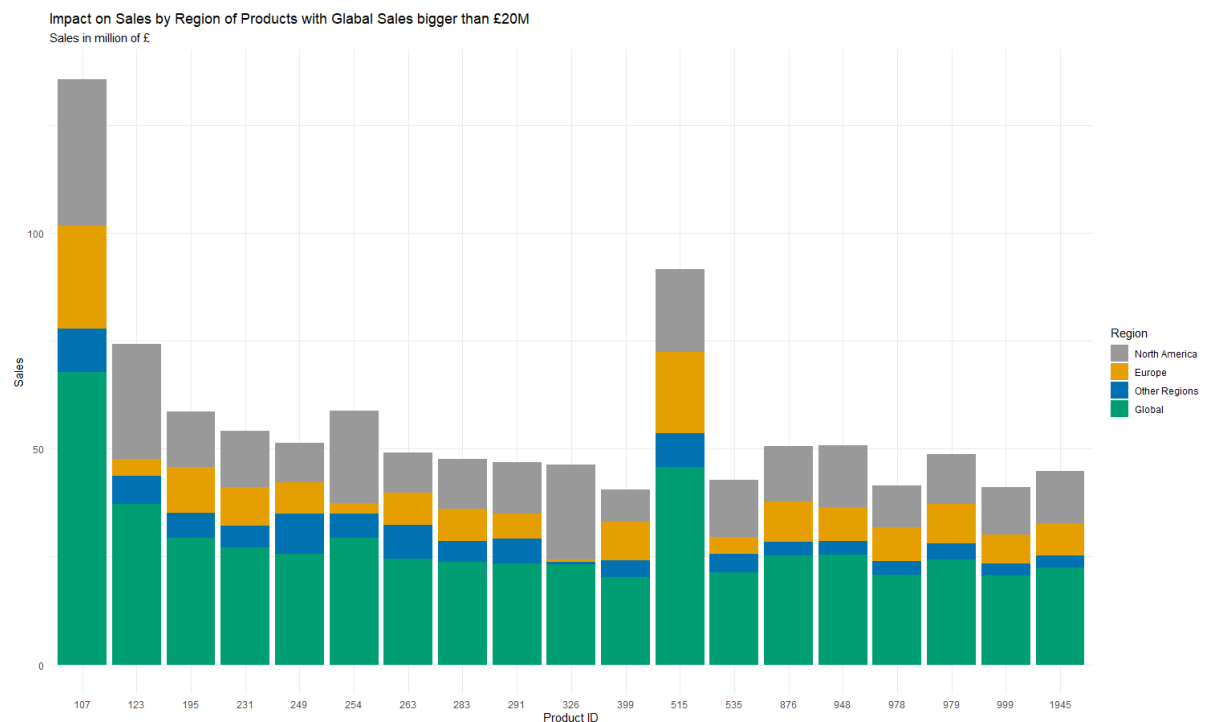
The top 5 are (performance not consistent across regions):

Impact of 5 Top-Selling Products on Sales by Region
Sales in million of £

The genre and publisher of these top 5 products are:

Genre and Publisher of 5 Top-Selling Products

Let's check how these products are distributed across the variables:

Number of Products by Genre

Number of Products by Platform

Number of Products by Publisher
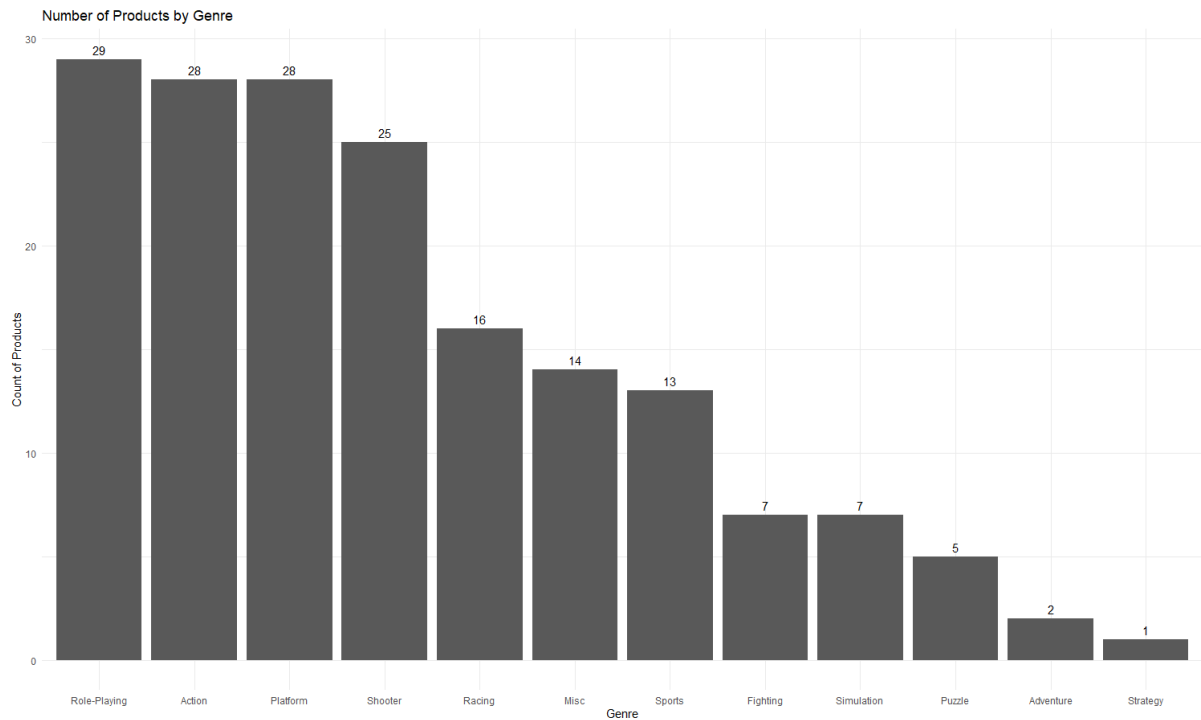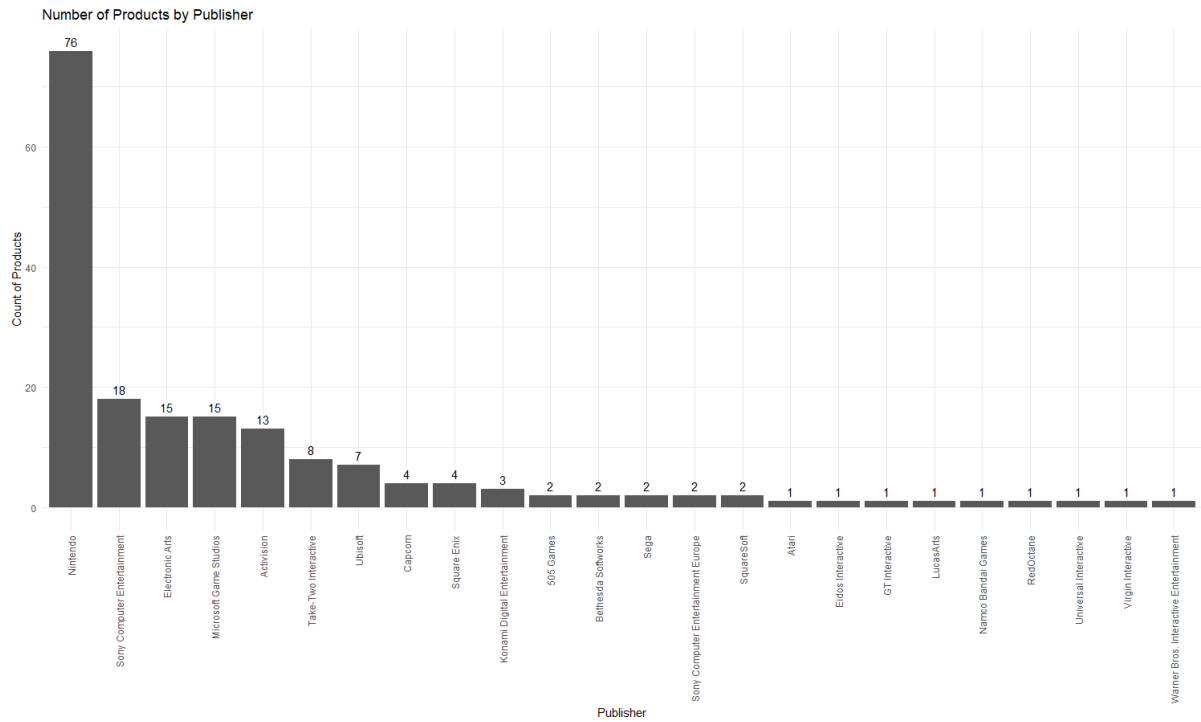
- Popular genres across regions?

Sales by Regions across Genres
Sales in million of £
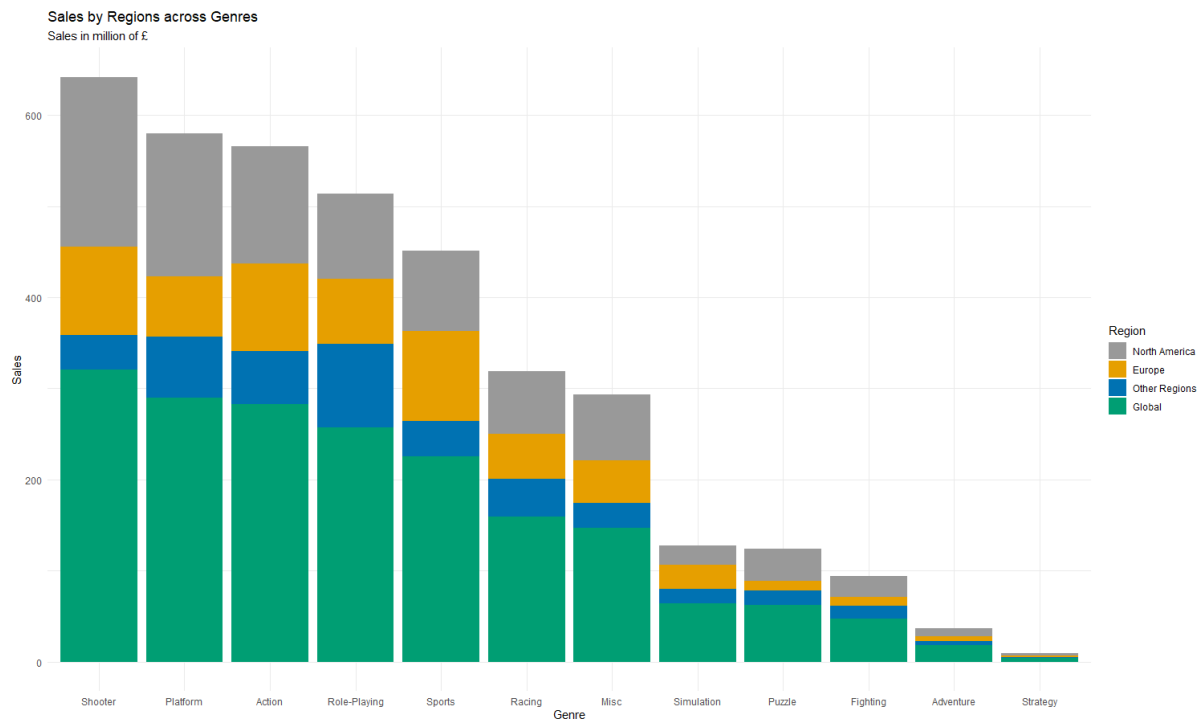
- Popular platforms across regions?

There are 22 platforms:

Sales by Regions across Platforms
Sales in million of £
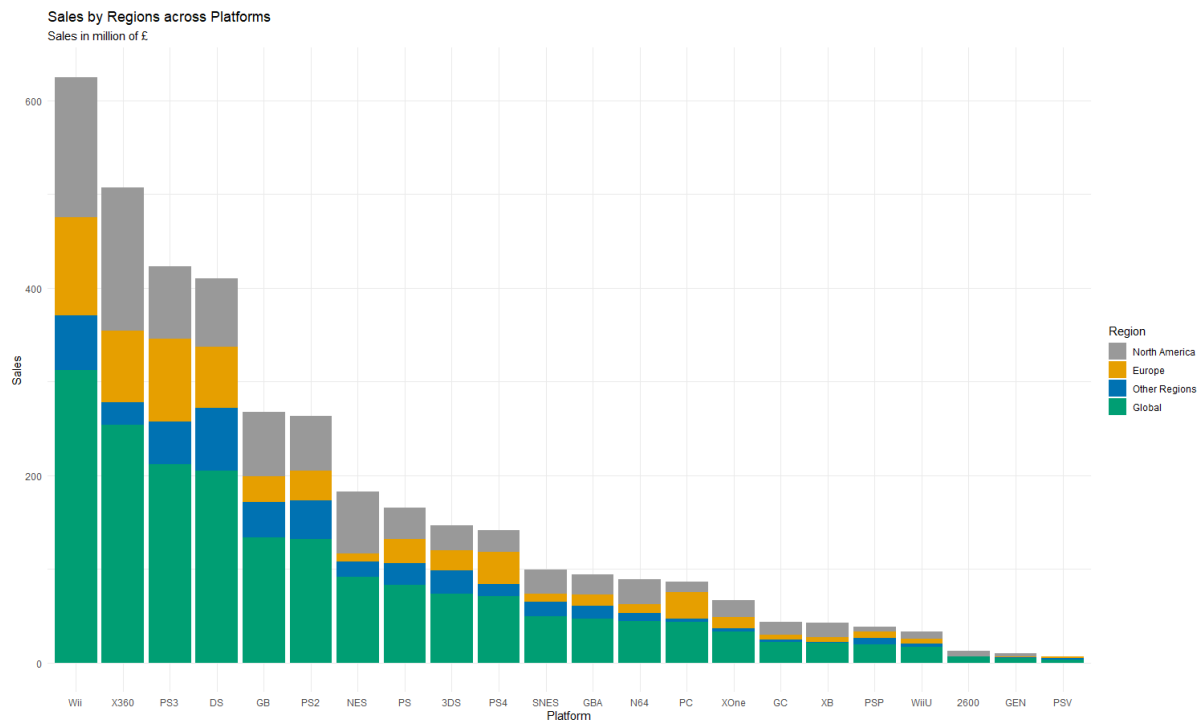
The top 6 platforms by Global Sales (more than £100M):

Impact of 6 Top-Selling Platforms on Sales by Region
Sales in million of £
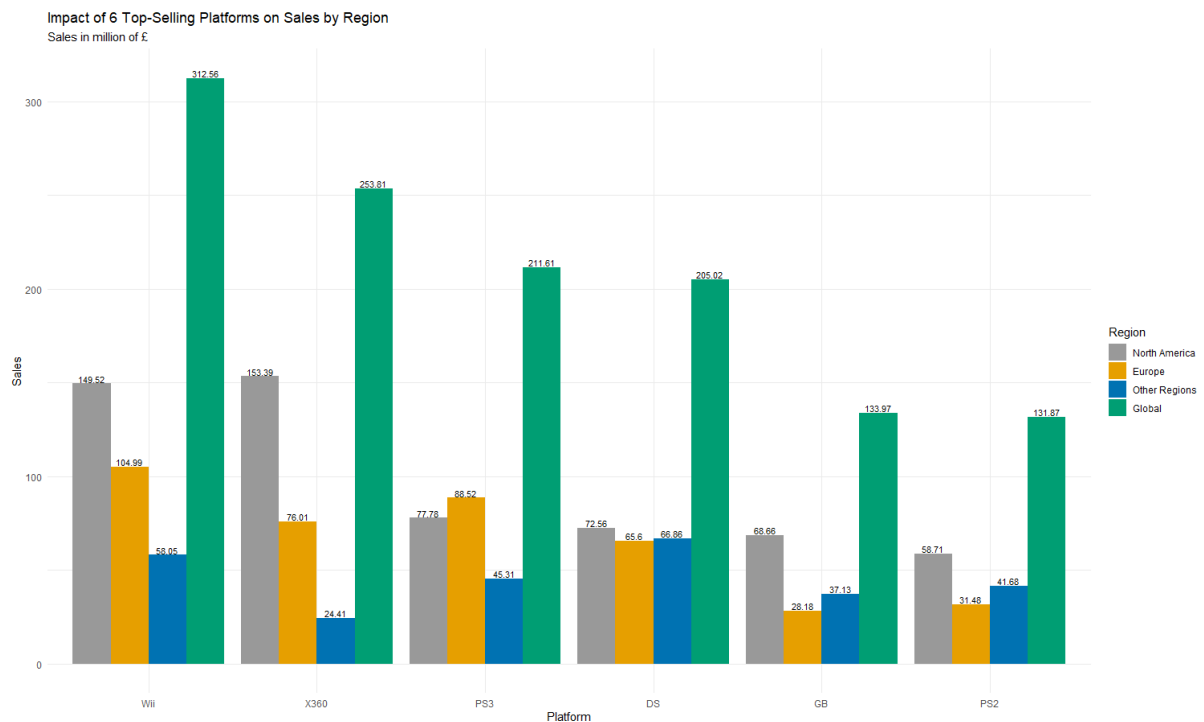
- Popular publisher across regions?

Sales by Regions across Publishers
Sales in million of £
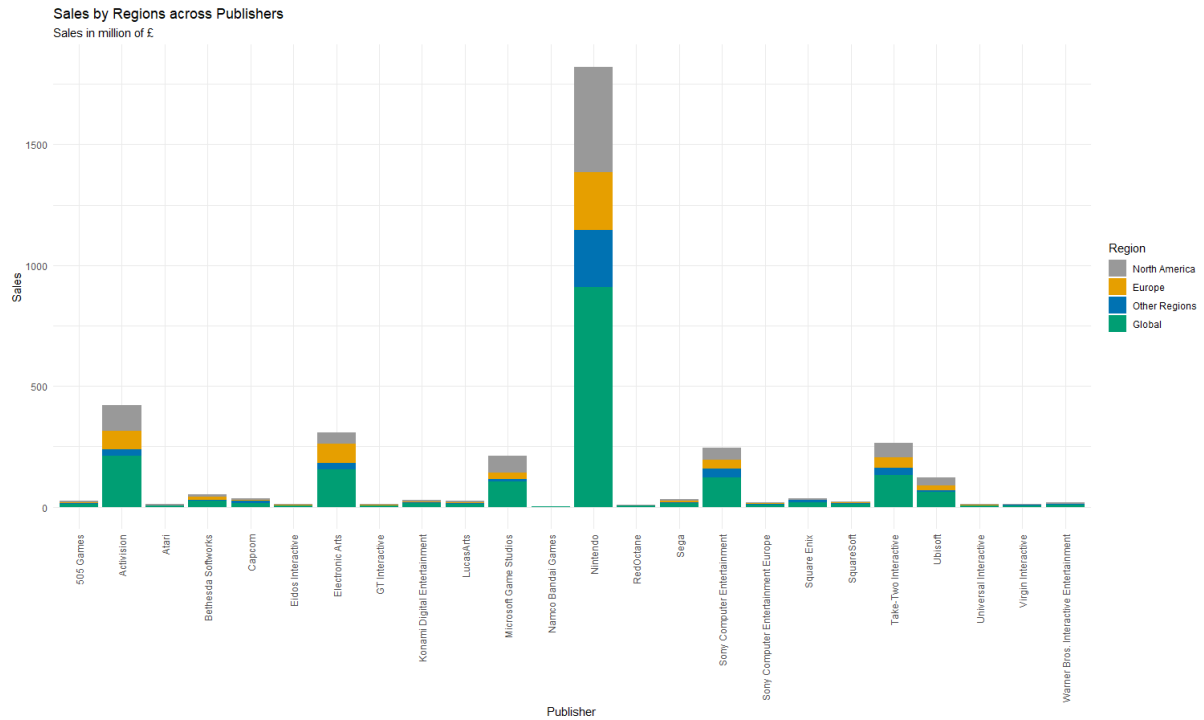
The top 8 publishers by Global Sales are:

**Impact of 8 Top-Selling Publishers on Sales by Region**
Sales in million of £
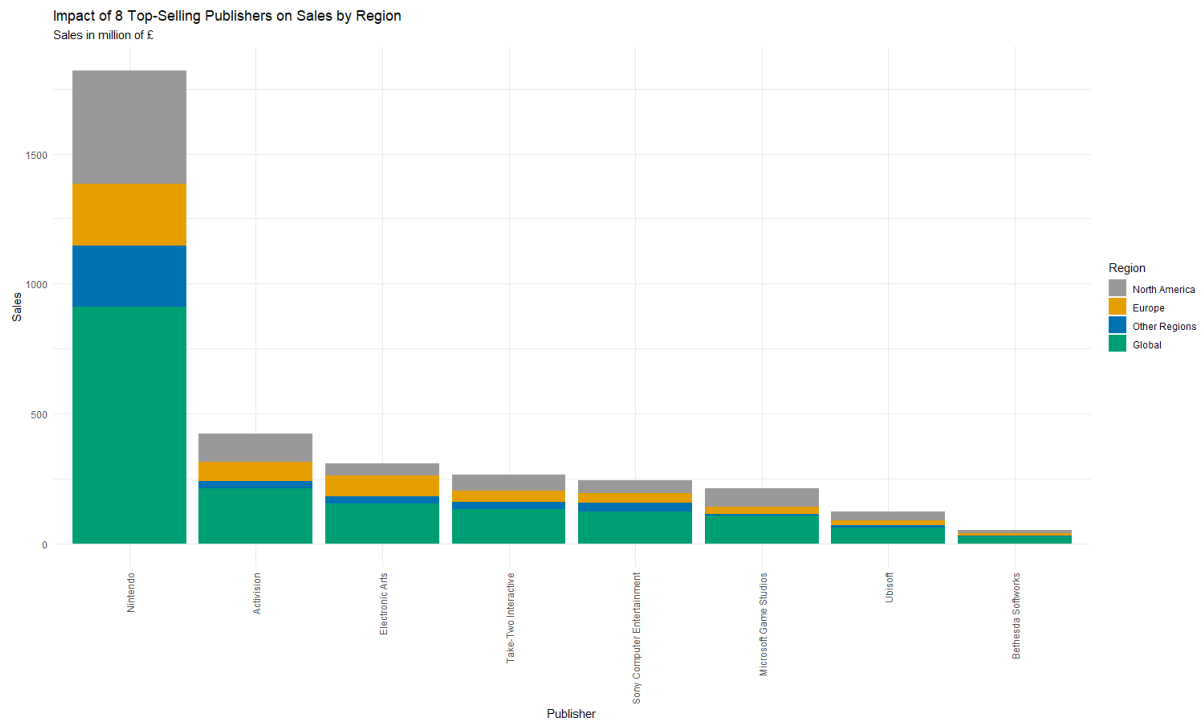
5. The reliability of the data?

After performing different tests we can conclude the data is not fully reliable as it doesn't perform well comparing it against a normal distribution of data (which will be the best option to get reliable results when performing predictions, see images below for comparison). However, if using the grouped data by products slightly improves the performance. We would recommend to gather more data and confirm the reliability of what we have called the best-selling products, as their sales numbers differ vastly from the rest of products.
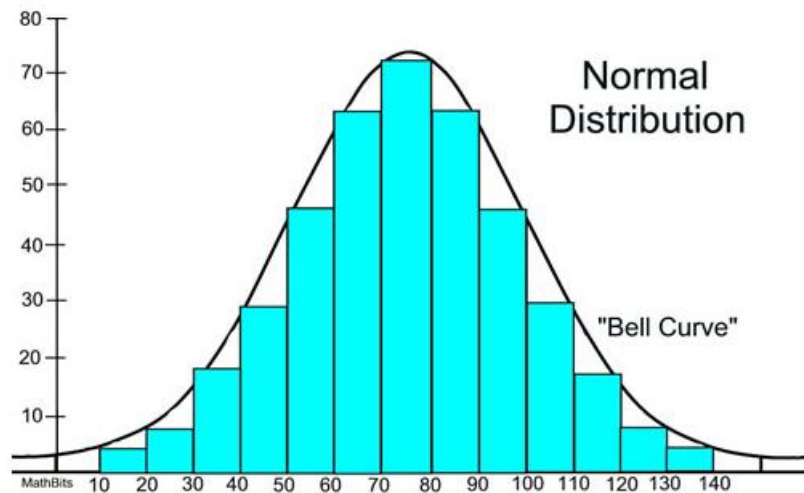
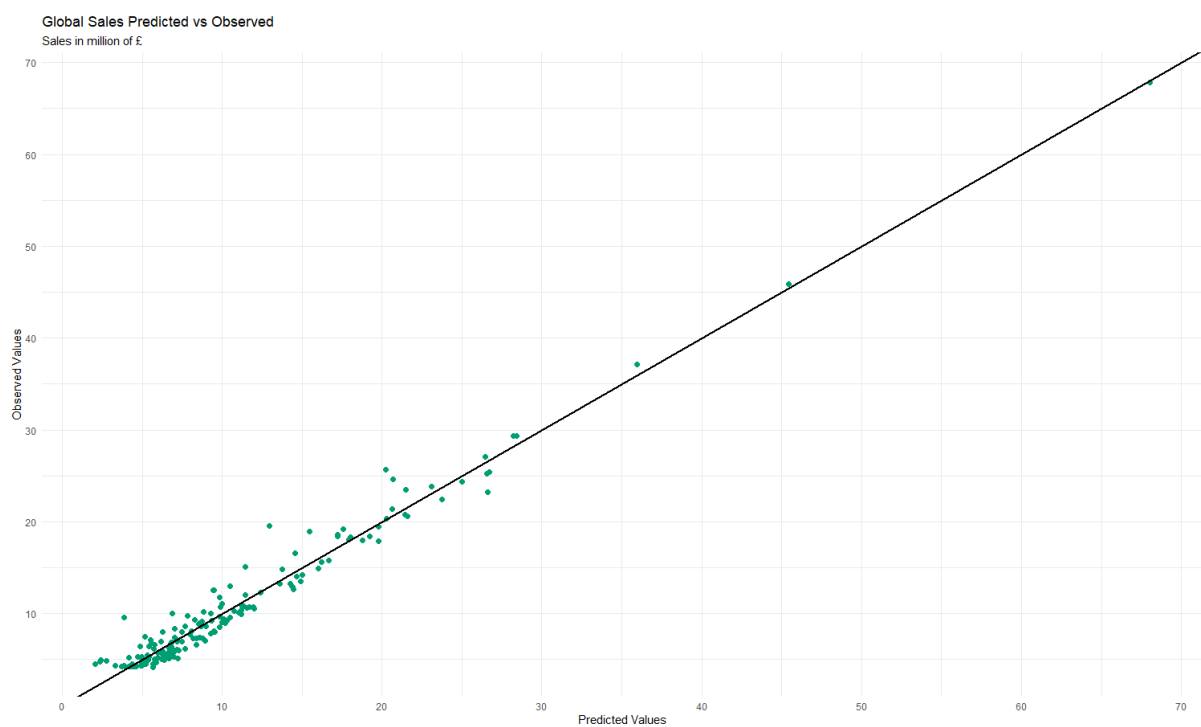*Normal Distribution (Bell Curve)*
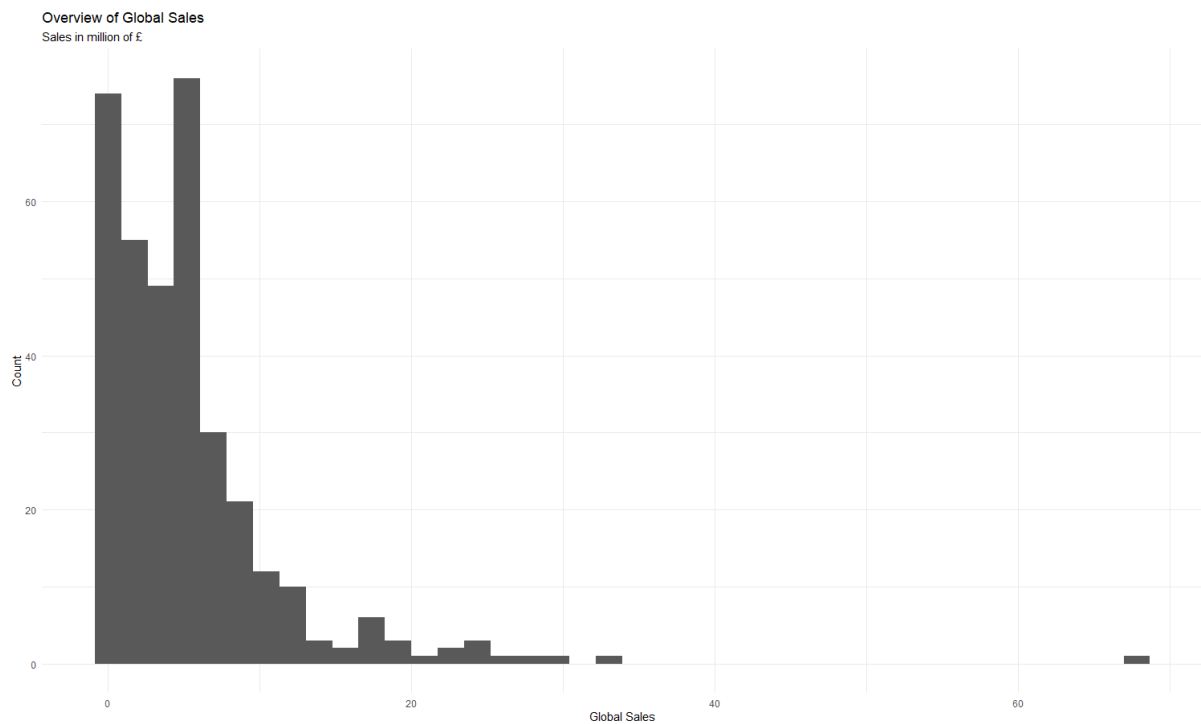
Distribution of Global Sales data

6. Possible relationship in sales between North America, Europe, and global sales?

We have discovered both North America and Europe sales have a positive relationship with Global sales, and when fitted into a regression model they can predict 97% of the global sales. However, the non-normality of the data can be an issue and we would recommend to gather more data and maybe look for other variables to be fitted into the equation.

Correlation plot from data

The model can be visualise as below:

Patterns and predictions

Based on the analysis described above, we found the following limitations:

- The data is not reliable to perform predictions due to non-normal distribution and the presence of outliers, therefore these recommendations are not conclusive and further investigations are recommended.
- The presence of outliers (extreme values) make the predictions unreliable.
- The number of variables at hand is limited to make predictions.

However, taking into account these quality considerations, we can conclude that:

- Loyalty points can be predicted with some accuracy from remuneration and spending score.
- Remuneration and spending score are useful to group customers for marketing purposes.
- Social data can be used for marketing purposes with caution as it tends to be subjective. The most common words show a positive feeling about Turtle Games products.
- North America is the best-selling region. Products from Nintendo and Take-Two Interactive are the best selling ones. However the sales are not consistent across regions.
- The data is not totally reliable and further investigations are needed.
- Global sales are directly related to North America and Europe sales, but these two are also related between them.

Recommendations and next steps:

- Predict loyalty points of first-time buyers to check if they are likely to churn (low loyalty points) or be loyal (high loyalty points), in which case benefits can be offered to secure a fruitful business relationship with them.
- The five groups described earlier can be used to implement different marketing strategies:
    o low remuneration/ low spending score: send offers to attract more sells.
    o low remuneration/ high spending score: send a thank-you voucher to recognise fidelity.
    o medium remuneration/ medium spending score: send a thank-you voucher to attract sells.
    o high remuneration/ low spending score: send product information to attract sells.
    o high remuneration/ high spending score: send a thank-you letter and a voucher if they bring someone else (prospect customers with the same level of remuneration) to attract sells.
- Trust better the summary of the review as supposed the review itself as it tends to be more objective. On the negative side "boring", "difficult" and "expensive" appear quite often, so maybe there is a need to review the game offer.
- Products don't sell consistently across regions; are there other variables we should be aware of? Keep the publishers that are selling the best and maybe drop the lower ones. The range of products is high

but not all of them perform the same way across regions and overall. Turtle Games could focus on the best ones and improve the offer of these.

- Collect more data to perform further investigations and improve performance. The data can be transformed (i.e. log) to improve distribution.
- Focus on North America sales to improve sales globally as it is the region with the best-selling numbers overall. The model fitted to predict Global Sales is fairly accurate, but it could be better if the normality of the data was to be resolved.
- Schedule a session to discuss data assumptions and potential further requirements

References:

- Business Models of Video Games: Past, Present, and Future
  https://medium.com/@mjperrotta46/business-models-of-video-games-past-present-and-future-2b2aafe8ade1
- Understanding The Gaming Industry And Its Business Models
  https://fourweekmba.com/gaming-industry/#:~:text=Gaming%20Consoles%20business%20models,making%20money%20by%20selling%20games.
- New report says there will be 3.5 billion gamers globally by 2024
  https://www.rcrwireless.com/20200717/5g/csps-business-models-booming-gaming-industry-study