

# Turtle Games Technical Report

## Background / Business context

Gaming industry is in a period of high growth (in terms of platforms, markets and products). The industry is expected to exceed £191bn revenue in 2023.<sup>1</sup>

Turtle Games peak sales (2006) represented 0.0086% of current market. Sales are dropping across all markets.

Goal: use data to improve sales and marketing performance by:

- targeting loyal customer groups
- utilising review content
- reviewing product range by market
- forecast sales

Analysis developed for Sales and Marketing Directors, in Python (marketing) and R (sales) and uploaded to Github for access and further team analysis.

Data provided was review and sales data. There is a Product column in both but it's not clear that the data cover the same period, so they have not been joined.

Assumptions:

- Year of release - biggest year for sales
- Loyalty and spending score linked
- Cost to support products and platforms
- Report created in 2023

## Analytical approach

Data was reviewed in excel for an overview with pivot tables and charts.

Reviews imported into Python (marketing preferred).

Data:

- Checked for blanks
- Removed any columns not used for this analysis
- Removed duplicates
- Renamed columns (remove spaces and non-alphabetical characters)

Linear regression to check for variable correlation impacting loyalty points. Education and gender were converted to numerical values to include in multiple linear correlation. No multicollinearity detected.

---

<sup>1</sup> GWI Gaming Report

Independent variables reduced to age and remuneration. (Based on meta description, spending\_score and loyalty\_points are both spend-based - not independent).  
Linear regression used to predict loyalty\_points from remuneration.

Used elbow, silhouette and dendrogram to evaluate optimal k-means cluster numbers.  
Selected 3 clusters based on equality between groups and clear grouping for remuneration.

Summary and review content merged for maximum words, stopwords removed. Wordclouds, frequency distribution and sentiment and subjectivity polarity created. Subset created for target remuneration group, sentiment and subjectivity polarity benchmarked along with creating CSVs to export top positive and negative reviews.

Sales data imported into R (sales preferred platform).

Data cleaned:

- Titles to lowercase.
- Checked for missing data (missing data removed on subsets of data including Year).
- Calculated other\_sales column by product (global\_sales minus NA plus EU sales).
- Product column converted to character
- Added date column from Year

Aggregated sales columns by sum of platform and product sales. Product subset of top 20 by global\_sales used to focus on top products driving sales. Sales insights gathered from scatterplots, histograms, boxplots.

Reviewed normality of aggregated data using Q-Q plots, Shapiro-Wilk, skew and kurtosis to create benchmarks and review for model use.

Simple linear model created for global\_sales from year or NA\_sales correlation. Plot of residuals checked (no pattern). Tested log transformation to get a better fit (not better). Used Year is an independent variable (NA\_sales forms part of global\_sales).

Multiple linear regression to create forecast model from variables. Split data into 80% training (1982-2009), 20% test data is 2010-2015 (so we can review predicted against actual).

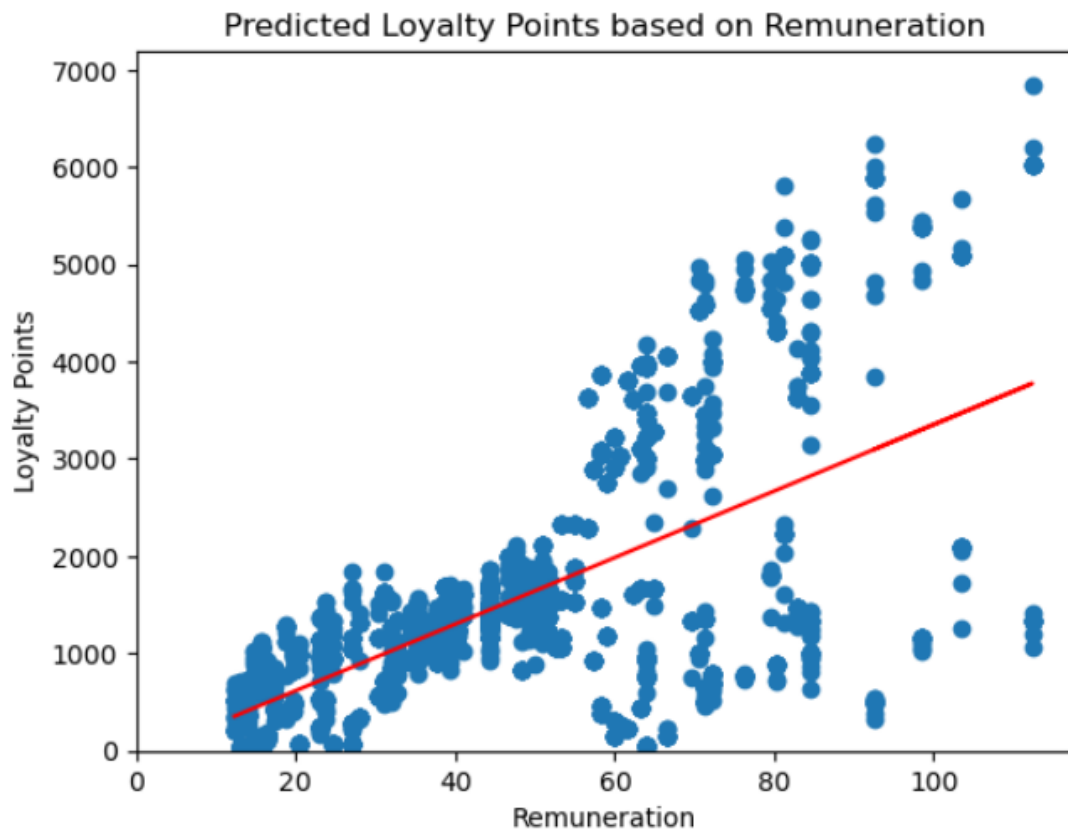
Tested 3 models against actual data to review best model. Modelc which used year and na\_sales delivered most accurate results based on test data.

Created new csv for new NA and EU sales data, added year (assumed this is 2016 data onwards). Output and visualised forecasts from modelc and new data.

## Visualisation & insights

Visualisation colours checked for accessibility.<sup>2</sup> Used Turtle Games brand colour (green) for global\_sales and contrast colours for other regions.

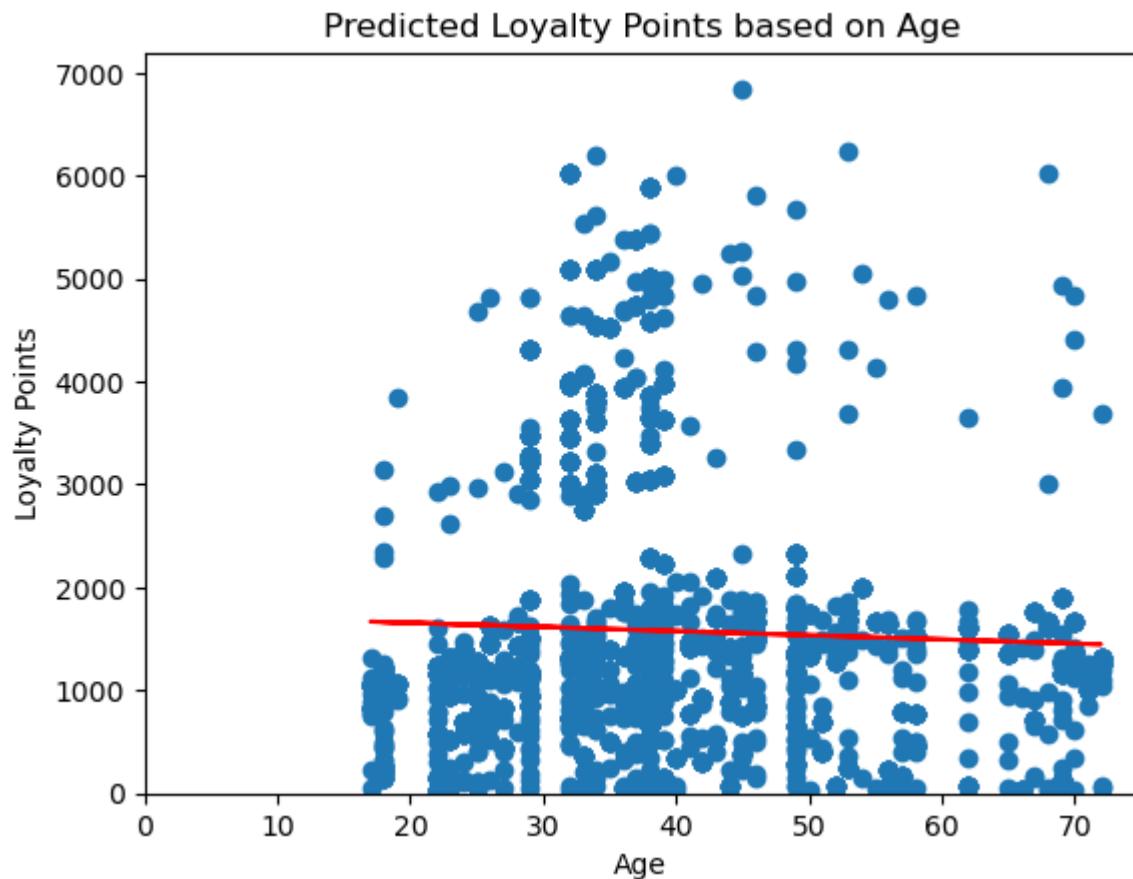
Reviewed which variables can be used to predict loyalty\_points. Spending\_score impacts loyalty\_points, however (from meta description) both loyalty\_points and spending\_score are based on spend, so spending\_score is not independent. Used linear regression with remuneration and age to predict loyalty points:



Plot shows heteroscedasticity - higher remuneration shows wider spread of loyalty points.  
Every £1000 increase in remuneration leads to 34 additional loyalty points.  
Marketing focus on targeting higher remuneration group with higher loyalty points.

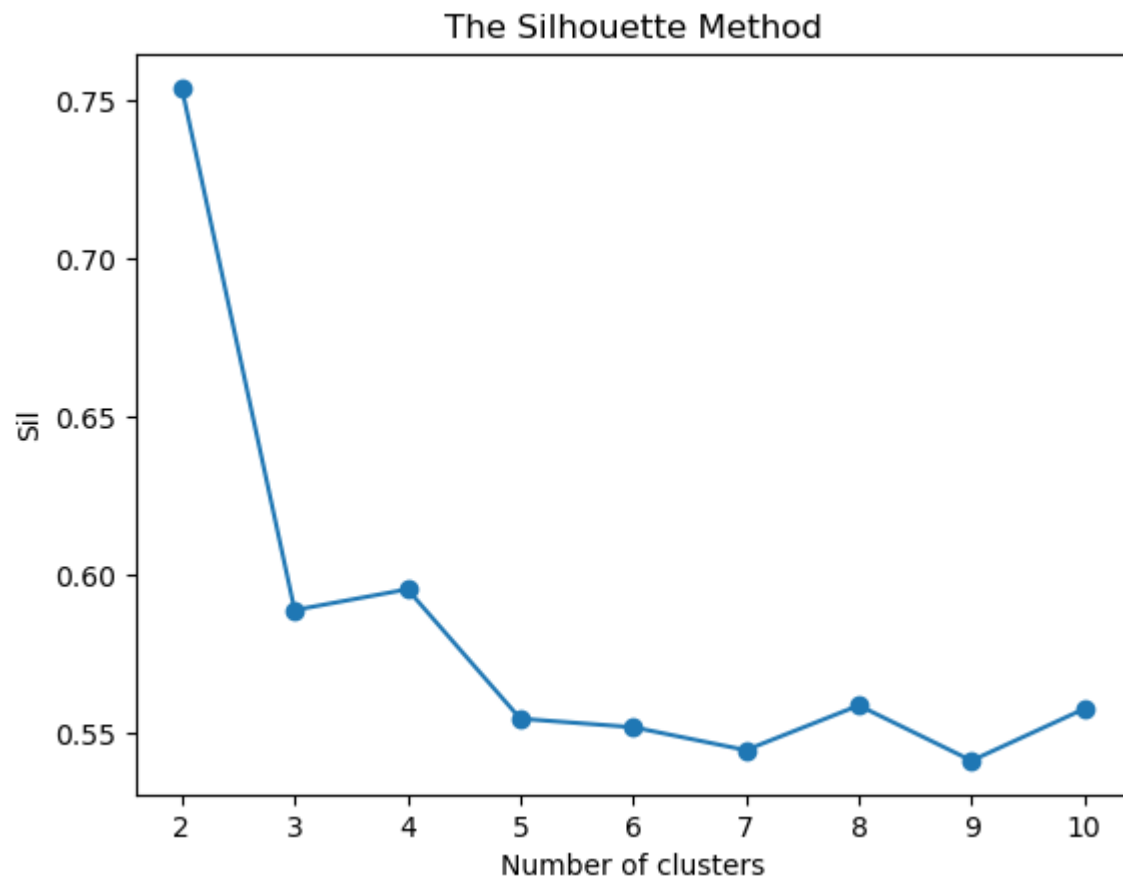
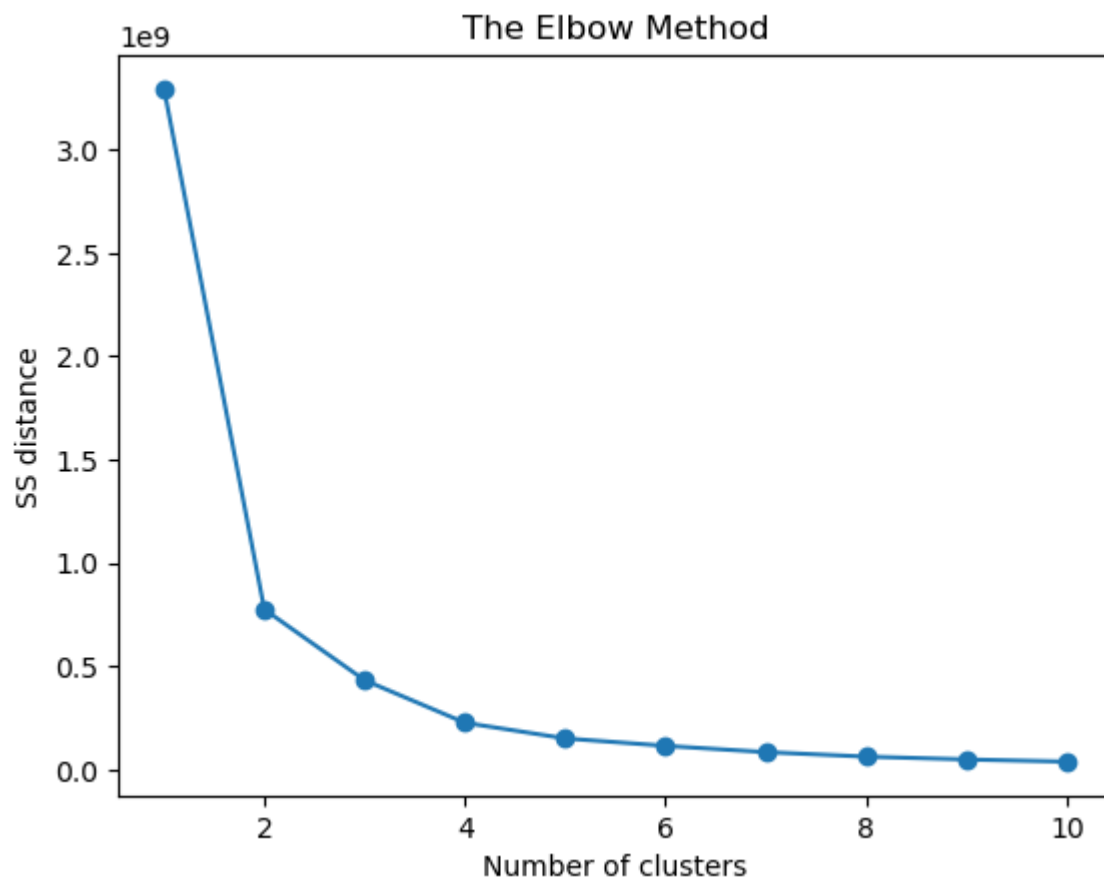
---

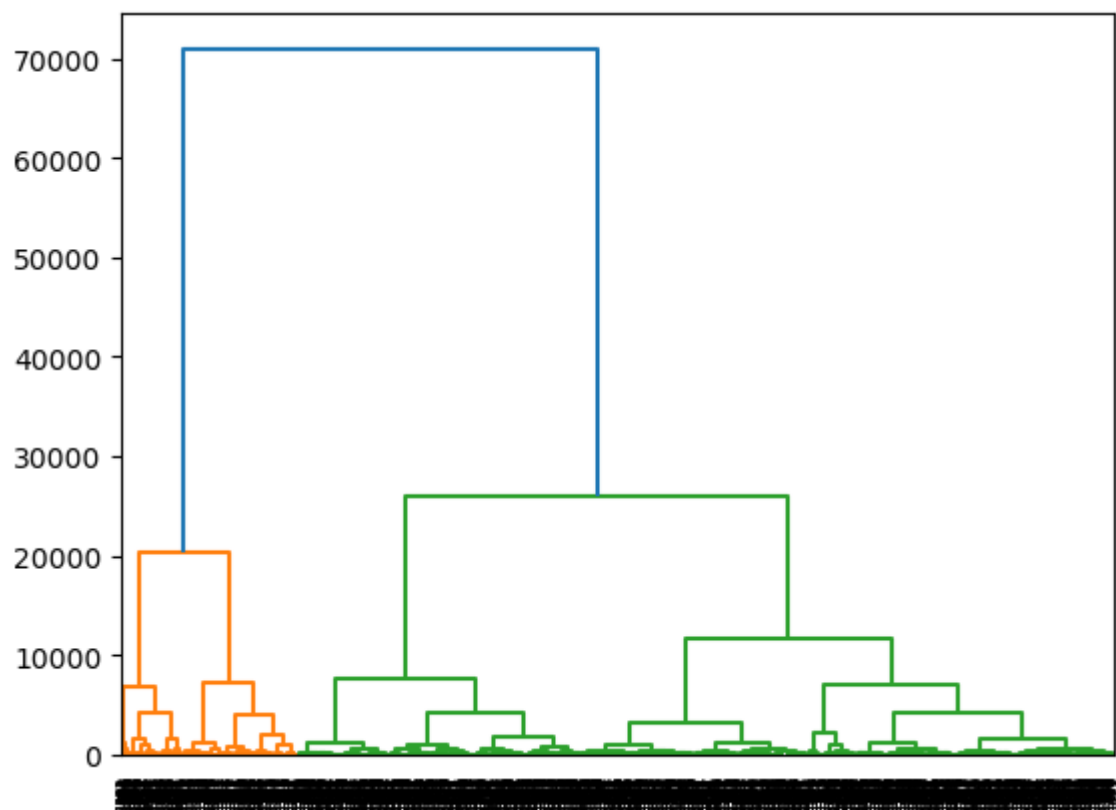
<sup>2</sup> Checked colours in: <https://www.color-blindness.com/coblis-color-blindness-simulator/> for colour blindness



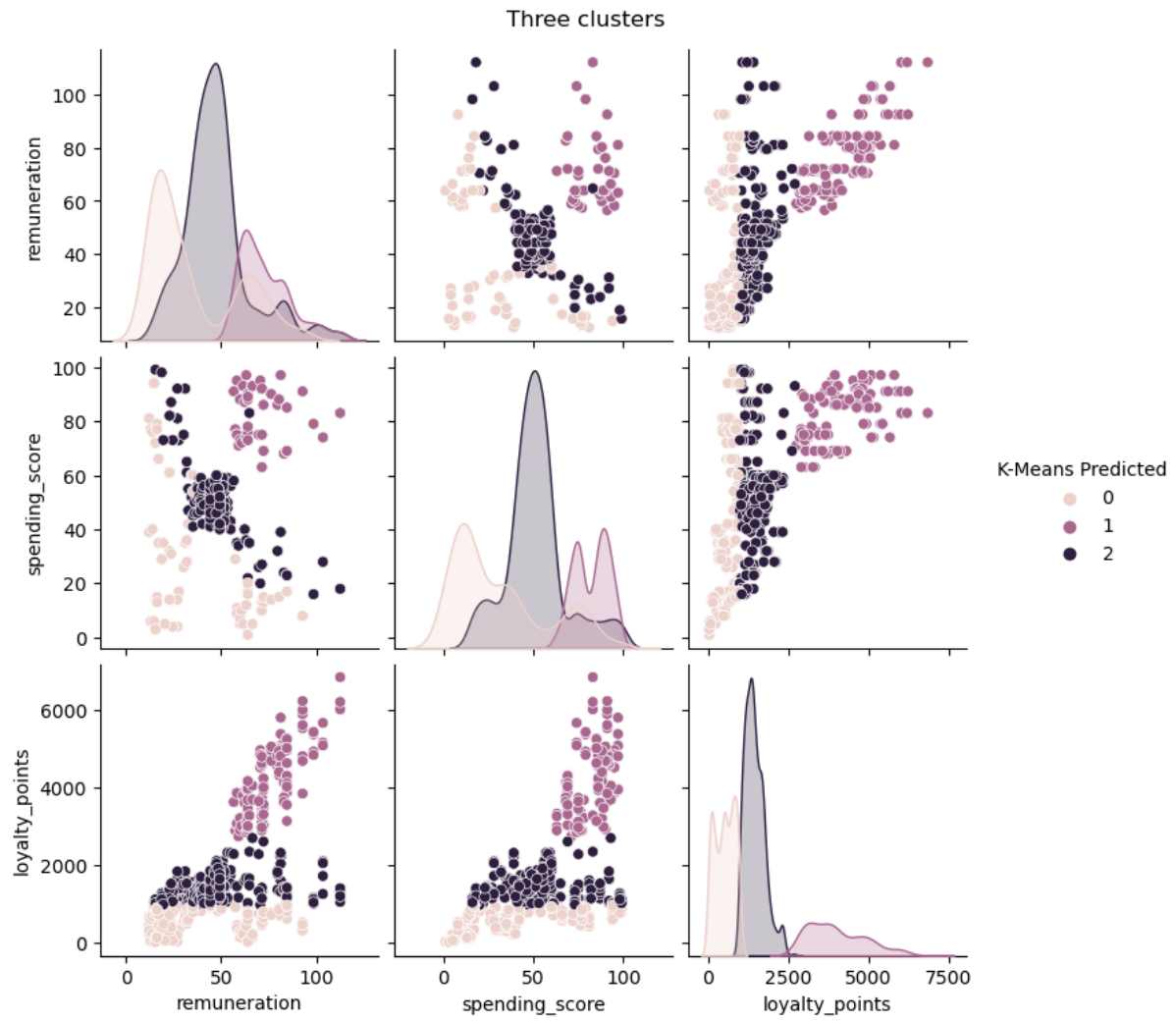
Age shows negative impact on loyalty points - every 1 year older creates -0.04 loyalty points drop. GWI report shows growth in older gamers - opportunity to develop Turtle Games products for older gamers.

K-means clustering used to find the optimal remuneration group for marketing targets. Elbow, silhouette and dendrogram methods indicated 3-5 groups as optimal:

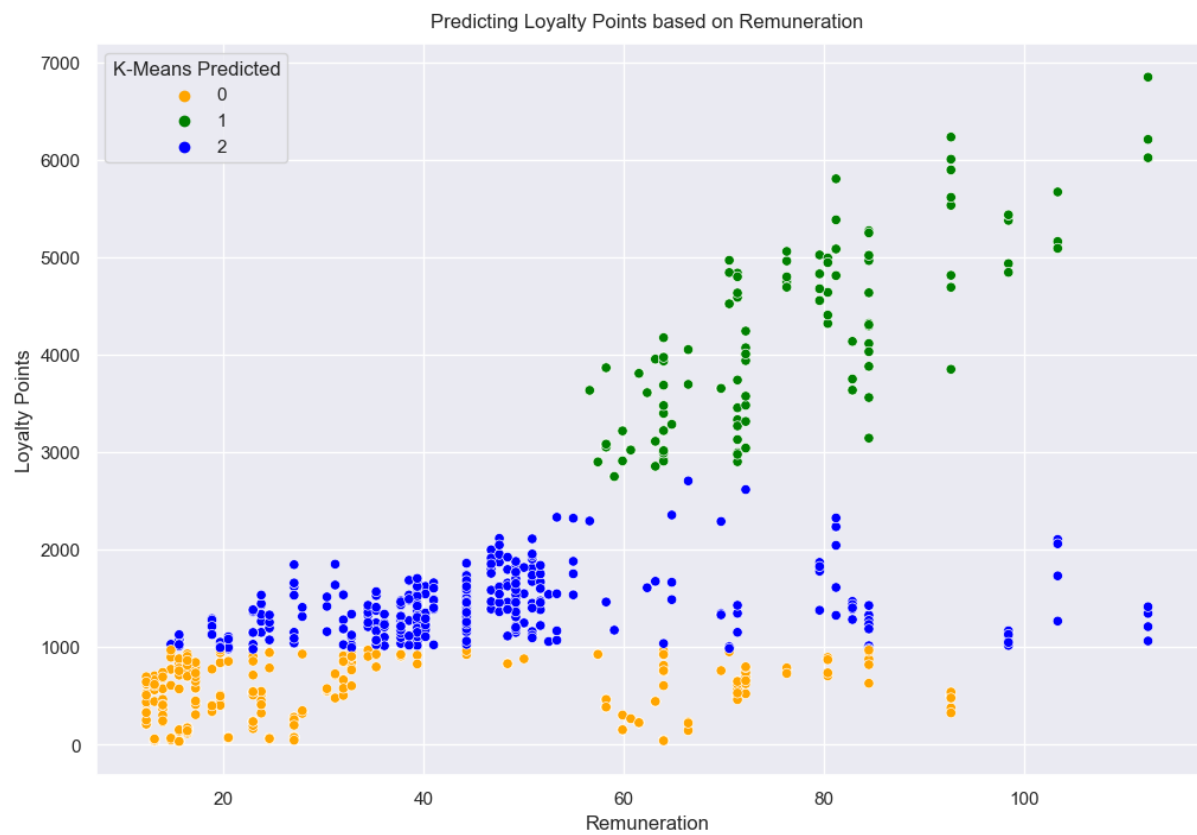




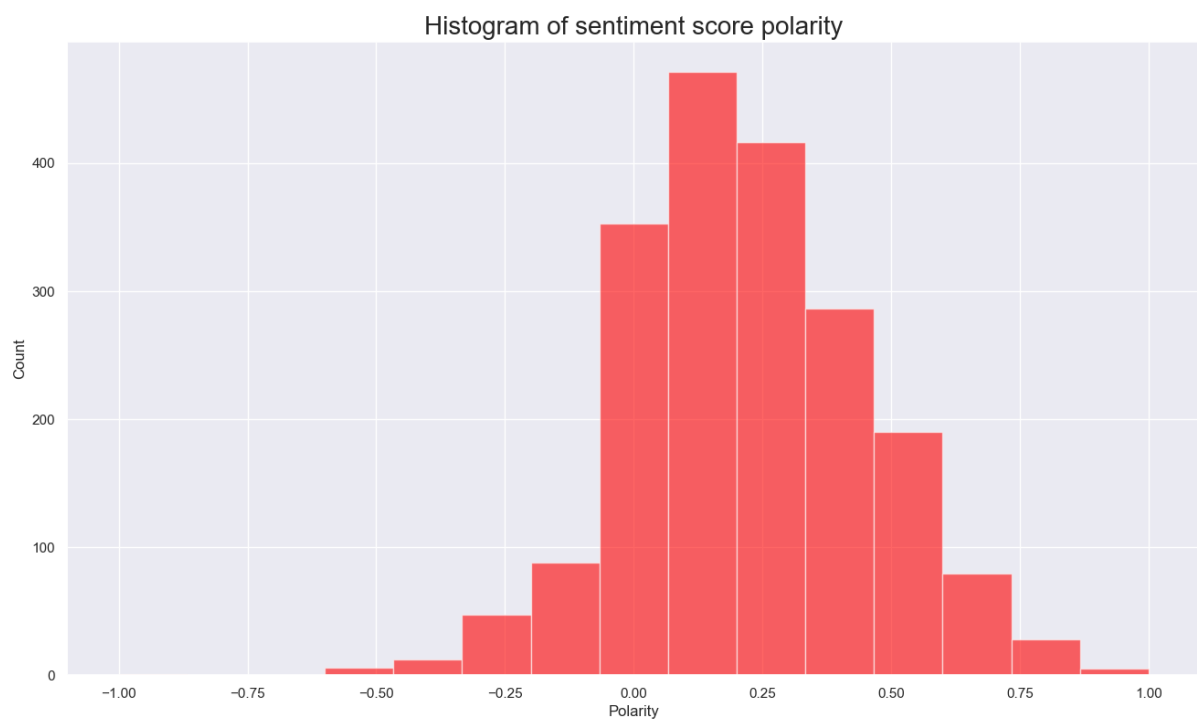
K-means clusters model fitted based on 3 clusters - based on remuneration (spending\_score as comparison). Three clusters chosen for equality between clusters and clear cluster.



Group 1 is optimal remuneration group for maximum loyalty points (£56.58k-£112.34k)



Review and summary content merged and benchmarked at 0.22 for sentiment and 0.45 on subjectivity. Then subset created for target remuneration group and benchmarked at 0.24 (indicates target group happier with product than whole market).





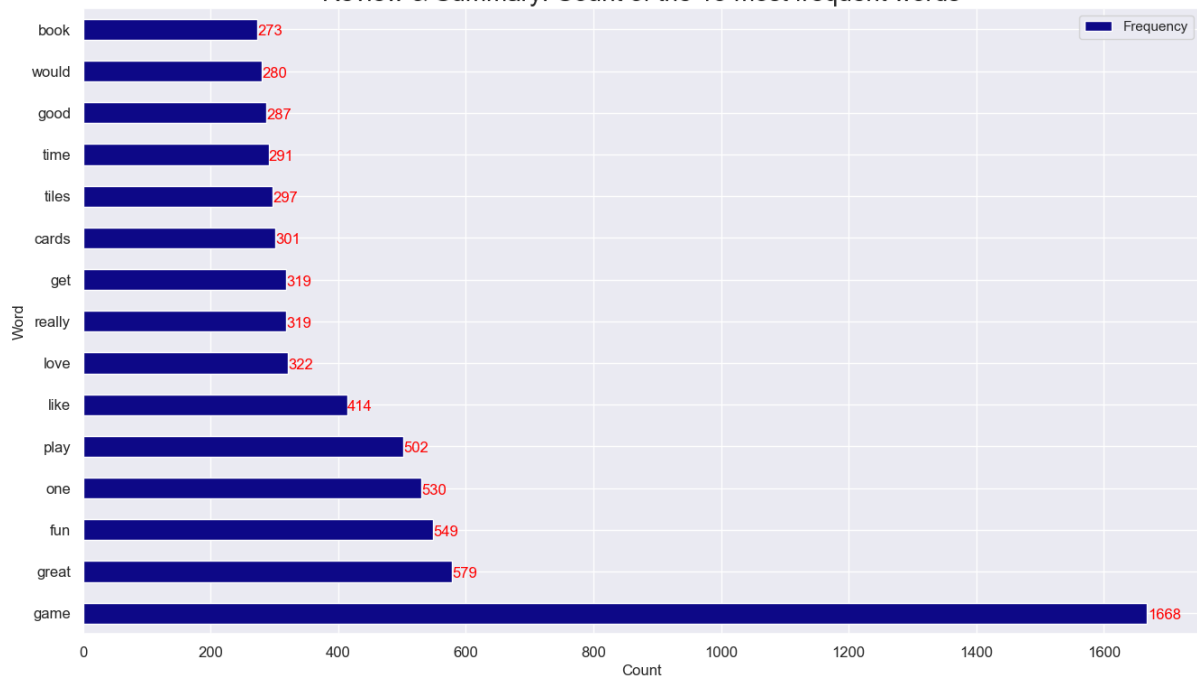
Wordcloud created for use for search engine marketing.

### Reviews & Summary words: stopwords excluded

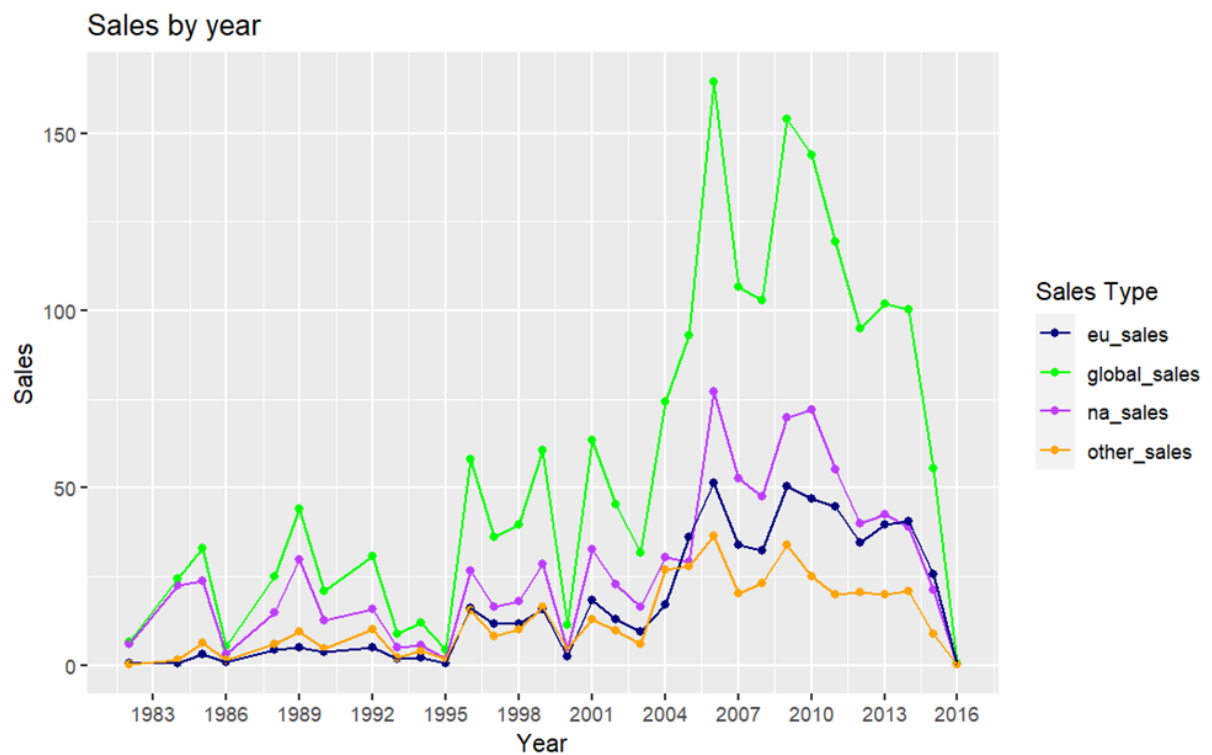


Frequency of top 15 words reflects product demand (games above books) and can be used for search engine marketing.

### Review & Summary: Count of the 15 most frequent words

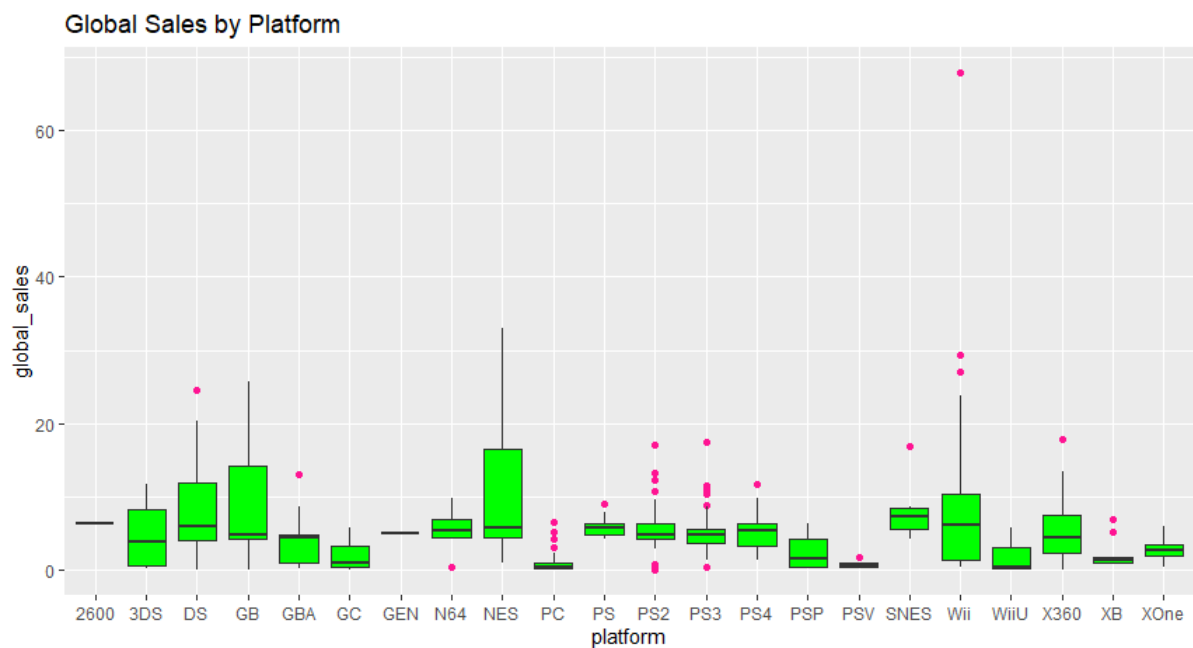


## Patterns & Predictions

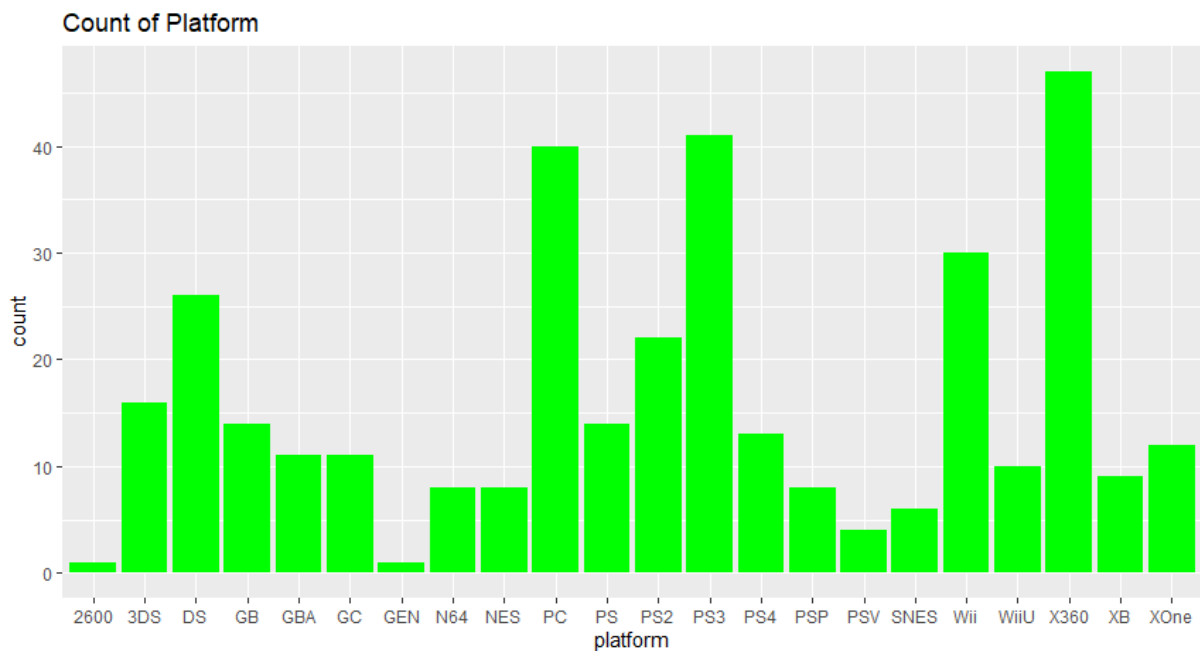


Sales dropping across all regions from peak in 2006. 2016 data is very low, indicating data not complete for the year.

Platform analysis shows clear outlier for sales - wii (although not the highest median sales), with 3 large outlier sales in total.

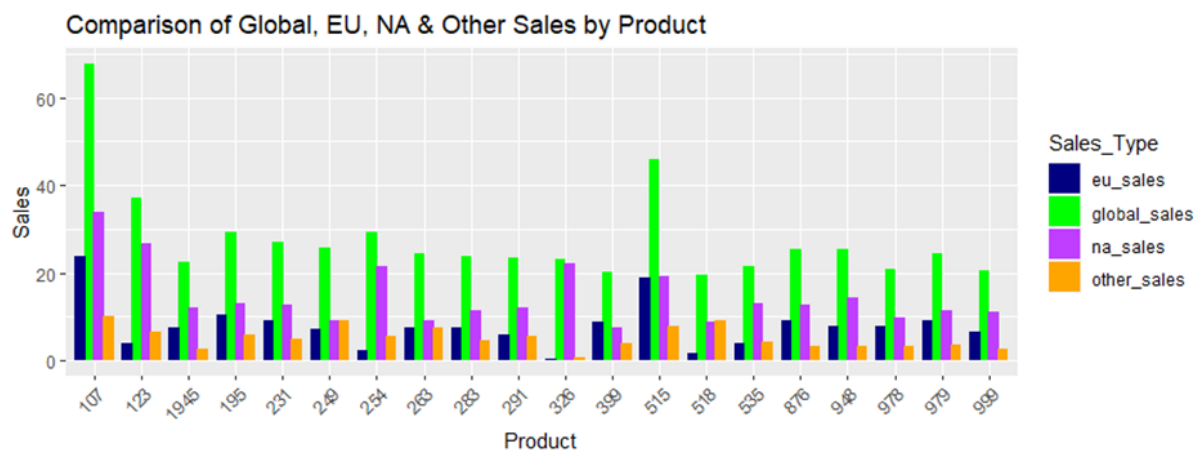


Count by platform: wii is lower than X360, PS3 and PC.

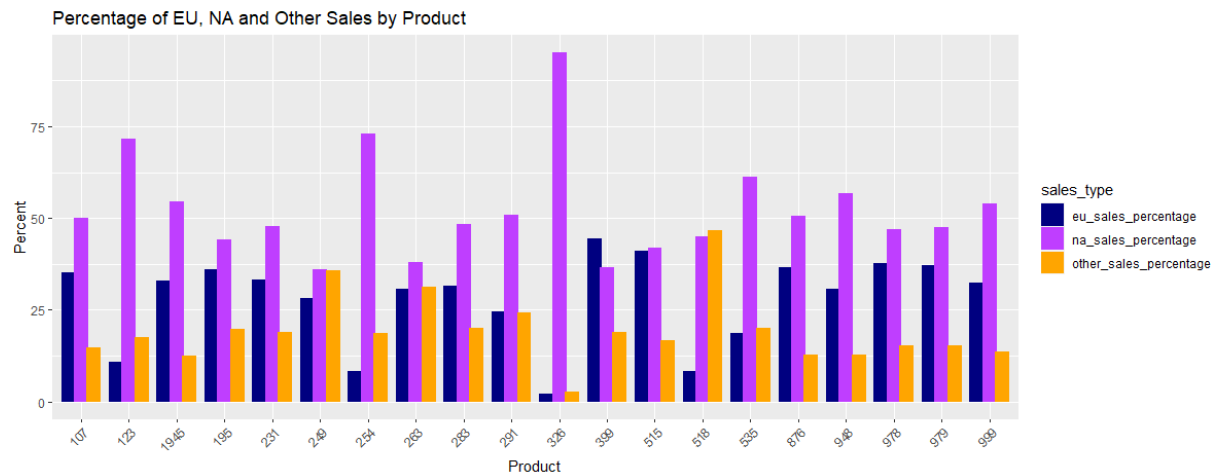


Recommend focus on wii and reduce focus (and costs) on other platforms. Needs further information on profits, costs and recent platform data.

Product list reduced to top 20 products by global\_sales to find highest selling products (107 & 515):



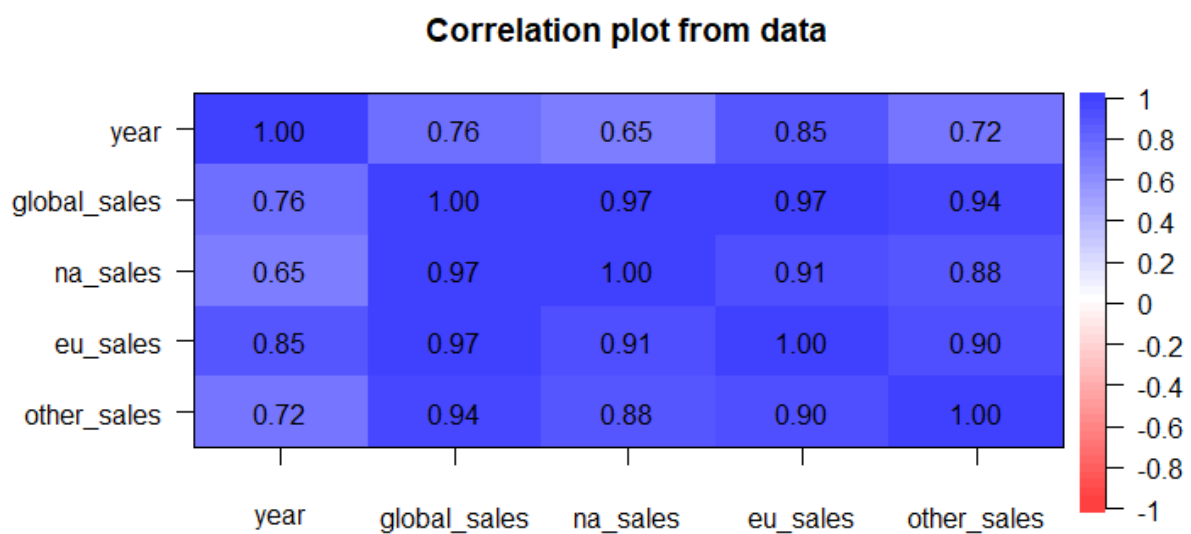
Each market as a percentage of global\_sales:



Different products sell better in different markets. NA generally makes up most of global\_sales, but 399 sells more in EU, and other\_sales higher sales than EU on products 518 and 249.

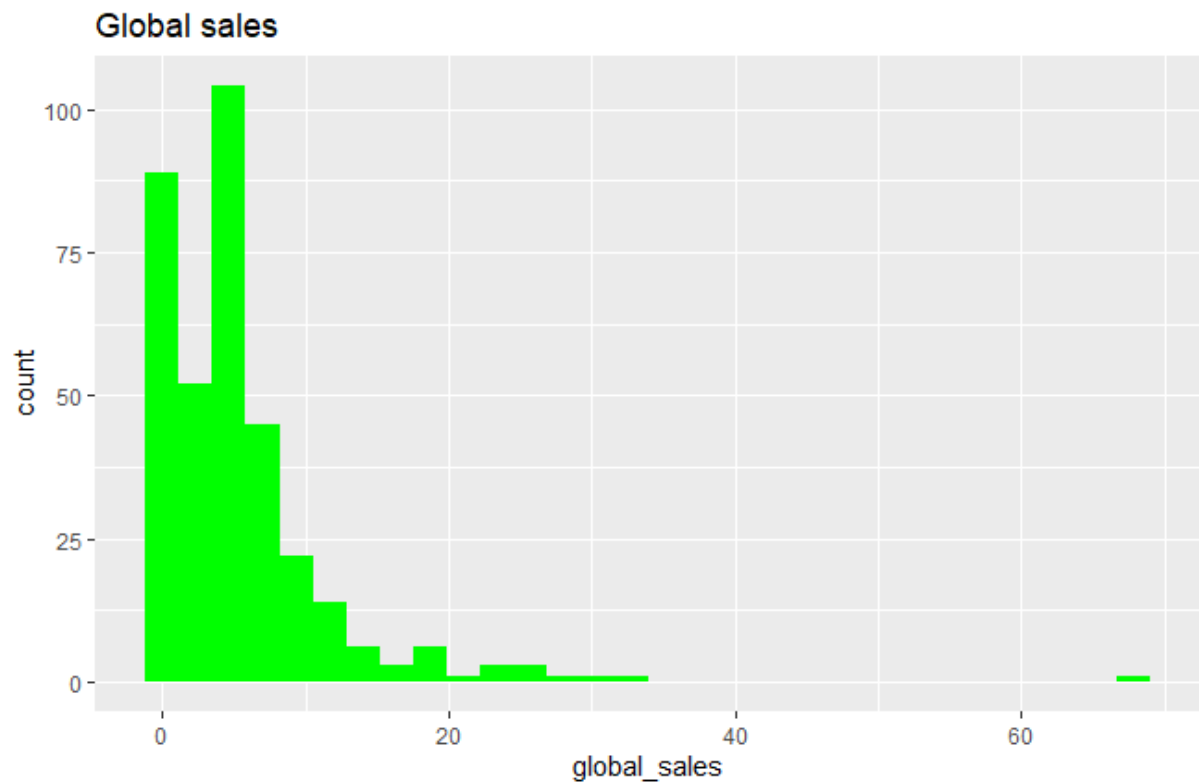
NA is a key market for most products. Focus on bestsellers per market. Try to diversify product portfolio to spread risk from one high selling product.

Reviewed correlation of Year and NA, EU and other\_sales:



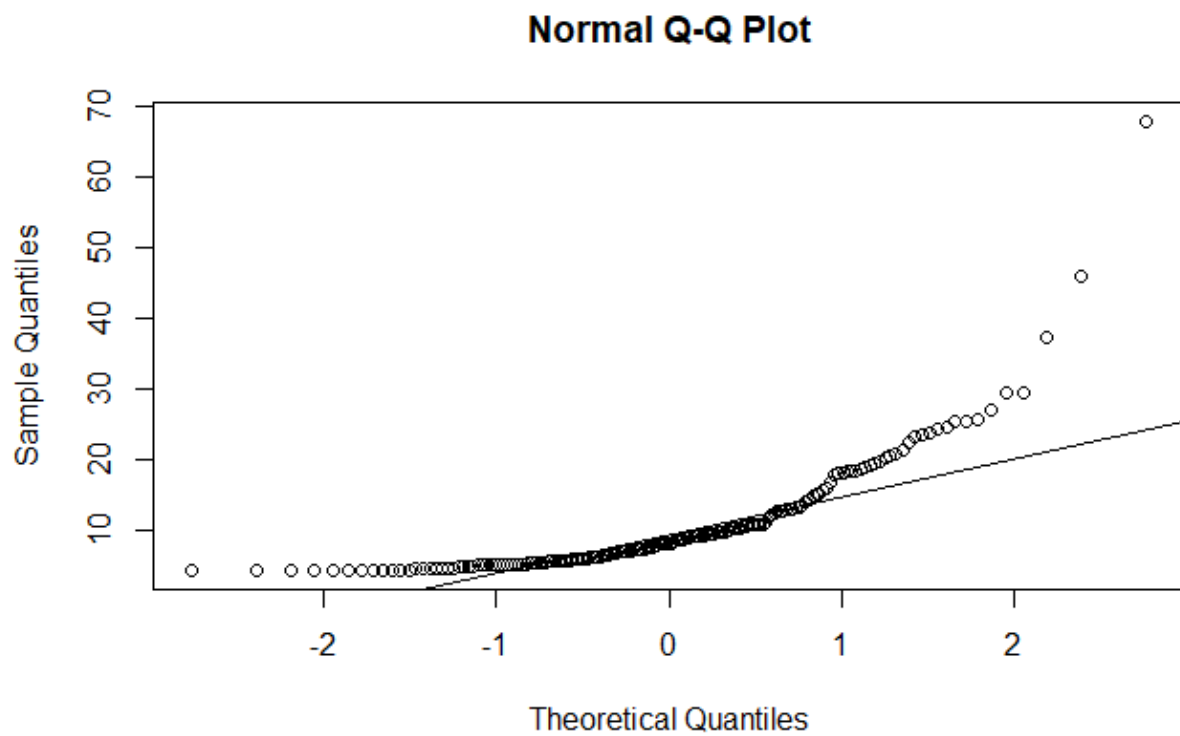
EU, NA and other should be closely correlated as they make up global\_sales, but year high correlated too.

Global sales is not normal distribution:

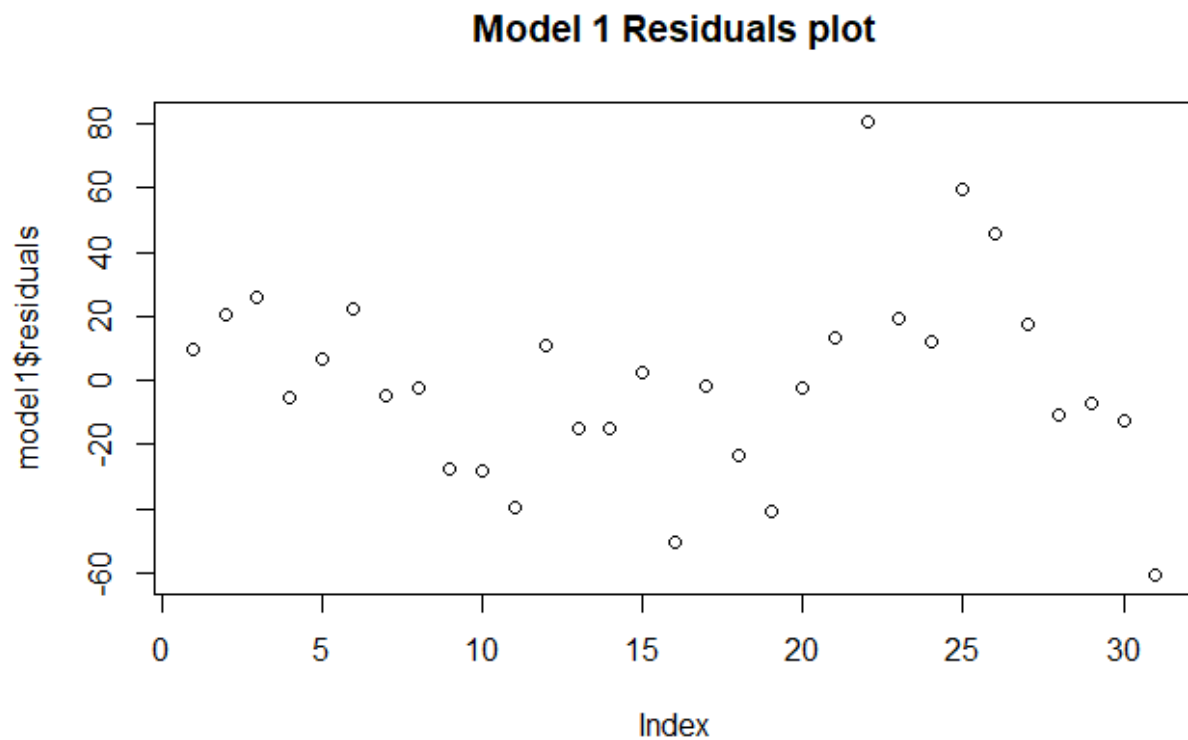


Heavy positive skew and kurtosis (benchmarked for annual review).  
Shows an outlier (sales reliant on single product).

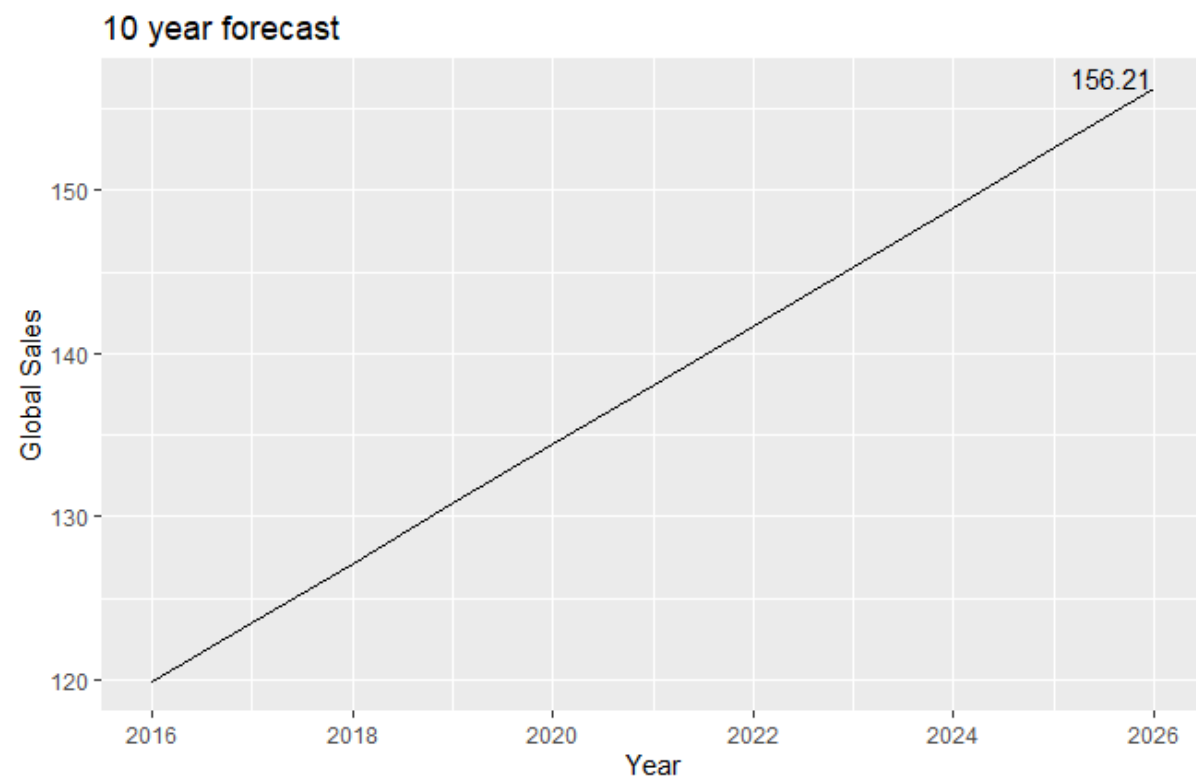
QQnorm shows outliers at the top and bottom, but all above the line:



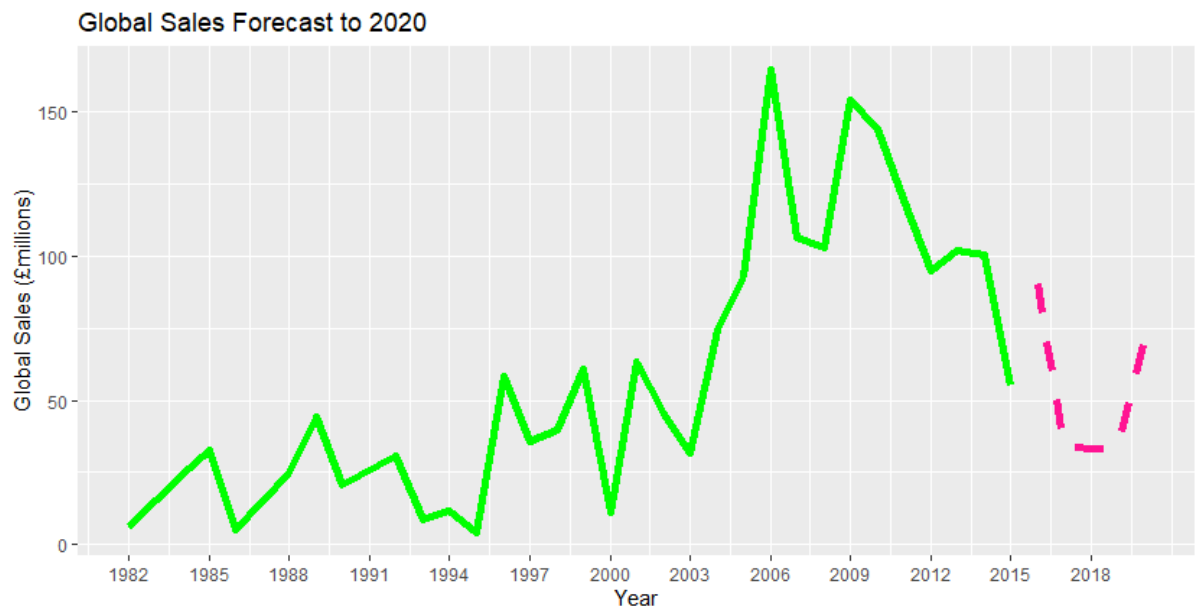
Residuals checked for pattern (no pattern so can use linear regression on data):



Simple linear regression based on year shows 2026 sales at £156.21 million. Each year increases sales by £3.63 million.



Multiple linear models trained and tested on supplied data showed NA and Year to be the best predictors for global\_sales. New data added, and run using NA and year to predict sales for 2016-2020 inclusive:



Overall trend down from 2010, despite 2016 and 2020 peaks. (Forecasting doesn't take into account pandemic impact on gaming market).

Your feedback: <https://forms.gle/zMDe1oZfvyPKguK87>

Wordcount: 1093