Turtle Games Technical Report

Business context

Gaming industry is in a period of high growth and expected to exceed £191bn revenue in 2023.[1]

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

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

Review and sales data provided, with Product column in both but not clear that the data covers the same period, so datasets not 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 as overview with pivot tables and charts.

Reviews imported into Python.

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.

---

[1] GWI Gaming Report

Used age and remuneration as independent variables. (Based on meta description, spending_score and loyalty_points are both spend-based - not independent).
Linear regression used to predict loyalty_points from remuneration.

Elbow, silhouette and dendrogram methods used to evaluate optimal k-means cluster numbers. Three clusters selected based on group equality and clear grouping for remuneration. Group 1 identified as best remuneration group to target for loyalty points.

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.

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). Visualised sales forecast from modelc.

Visualisation & insights

Colours checked for accessibility.[2] Used Turtle Games brand colour (green) for global_sales and contrast colours for other regions.
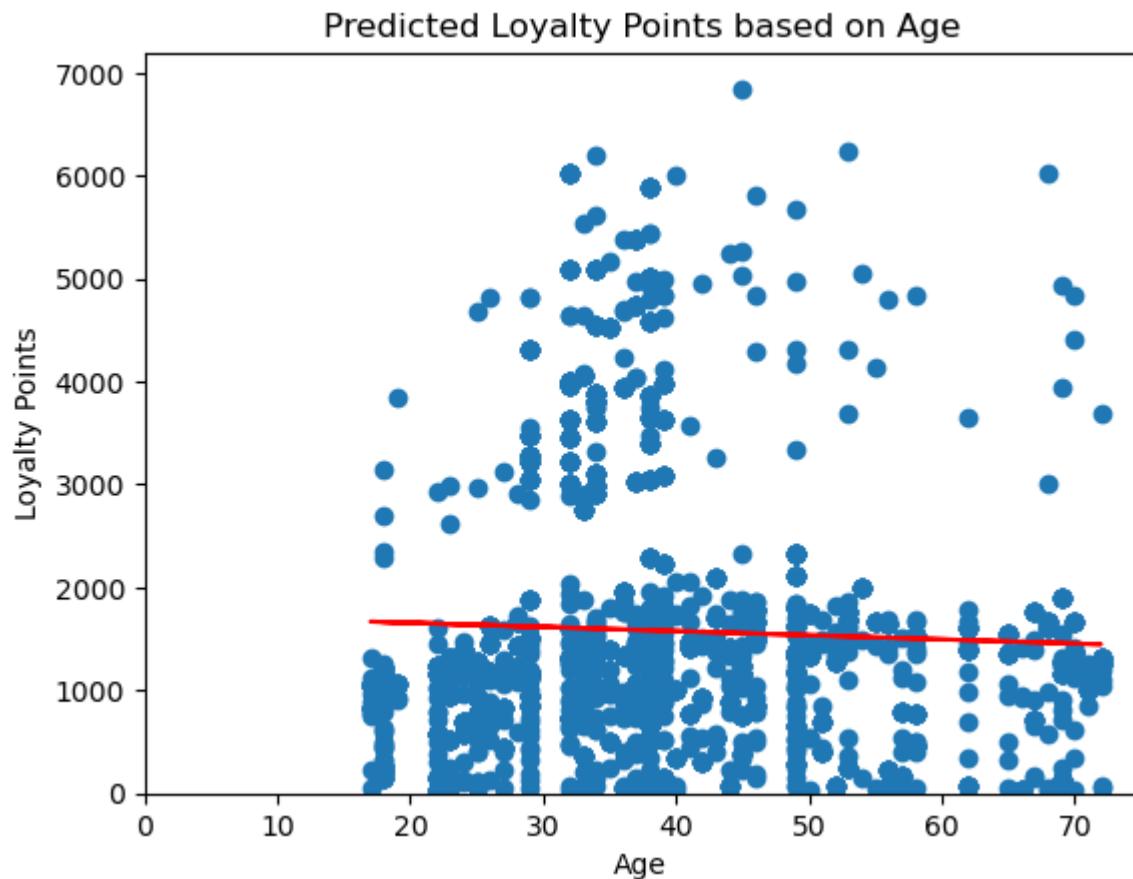
Linear regression used with remuneration and age to predict loyalty points:
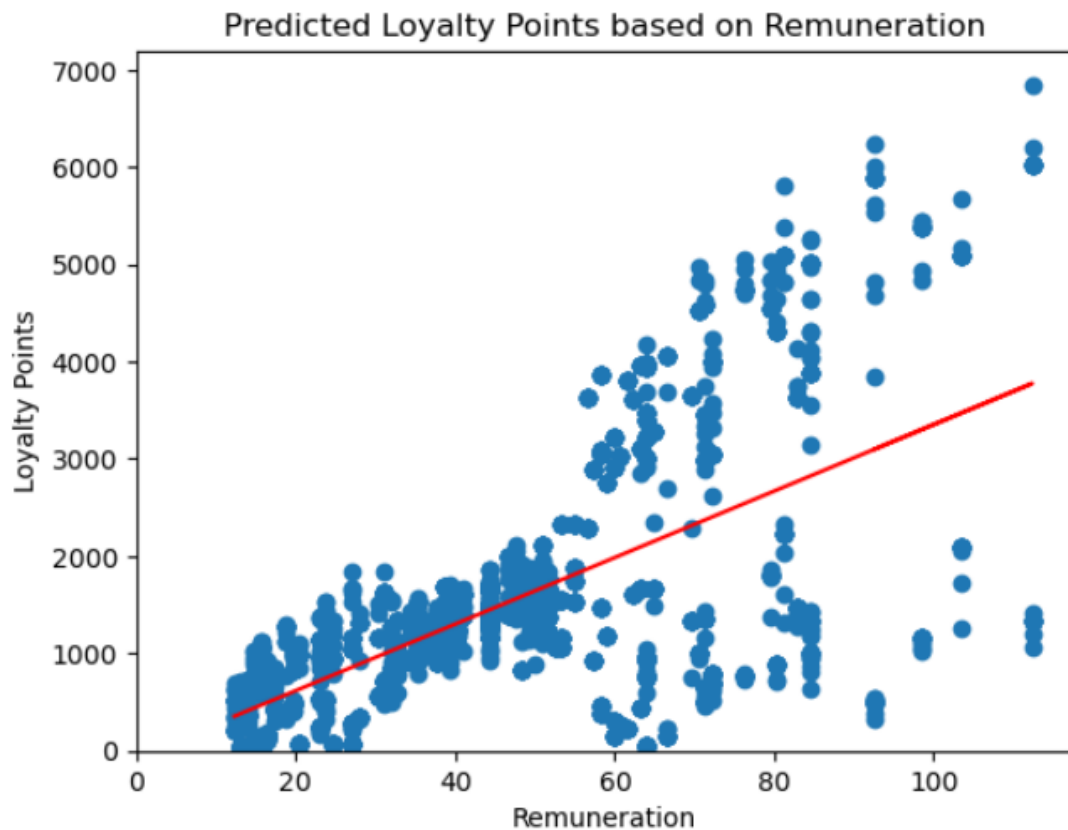
Plot shows heteroscedasticity - higher remuneration shows wider spread of loyalty points.
Every £1000 increase in remuneration leads to 34 additional loyalty points.
Marketing focus on targeting higher remuneration group with higher loyalty points.

---

[2] Checked colours in: https://www.color-blindness.com/coblis-color-blindness-simulator/ for colour blindness

Predicted Loyalty Points based on Age

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.
Elbow, silhouette and dendrogram methods indicated 3-5 groups as optimal:

The Elbow Method

The Silhouette Method
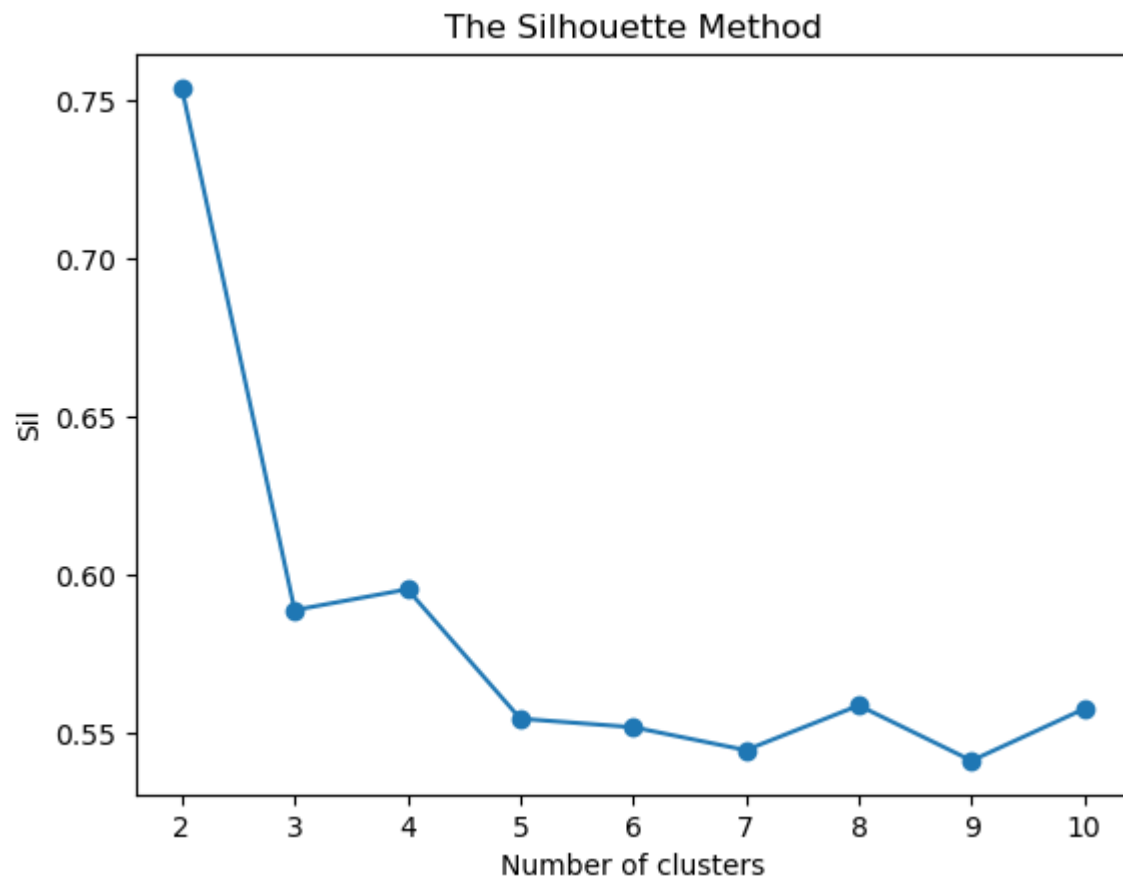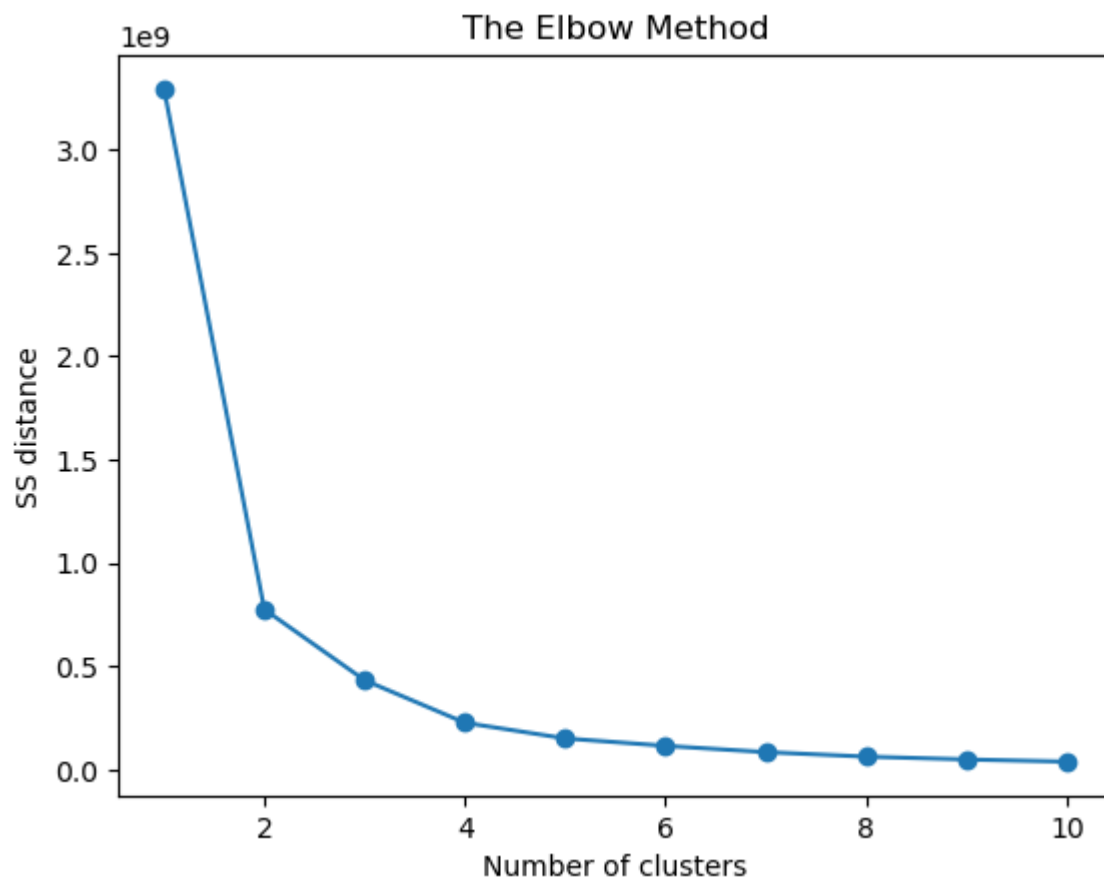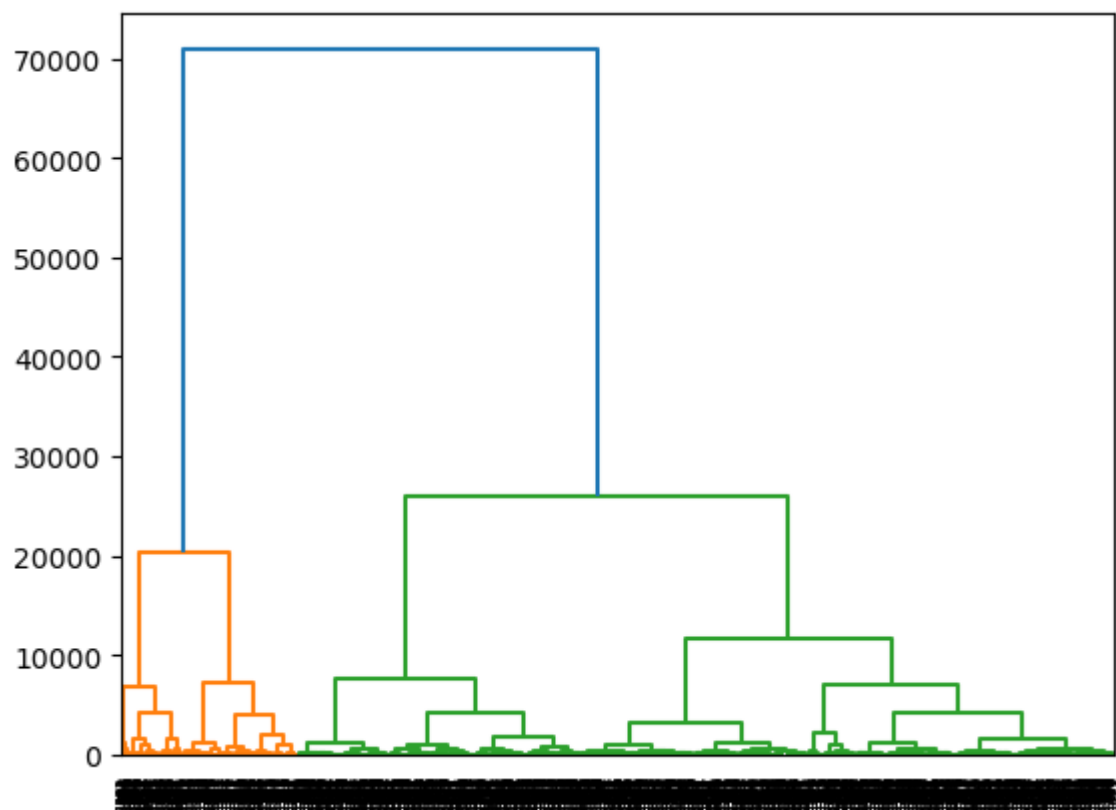
K-means clusters model fitted based on 3 clusters (chosen for equality between clusters and clear cluster) based on remuneration (spending_score as comparison).

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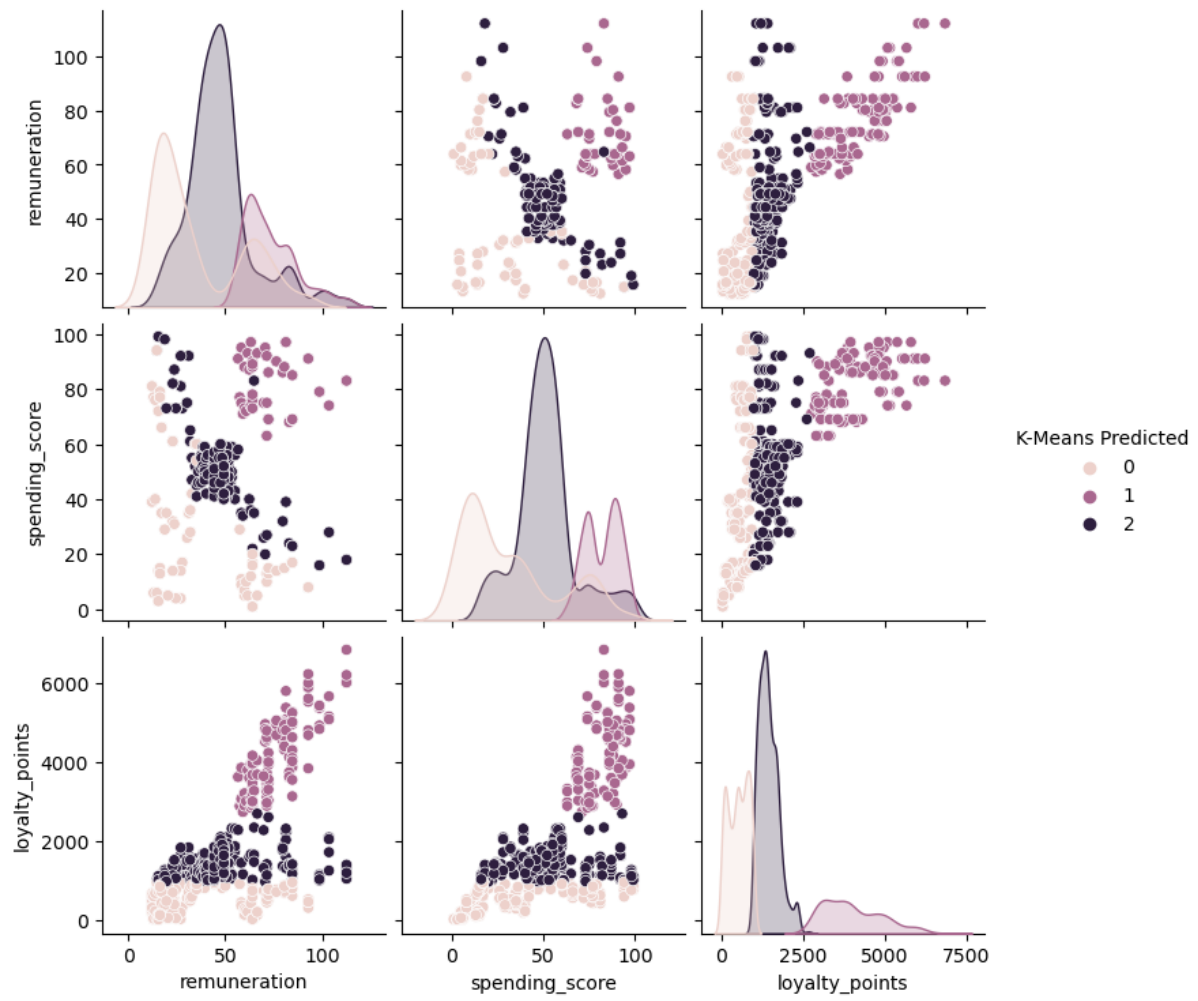
Group 1 - optimal remuneration group for maximum loyalty points (£56.58k-£112.34k)

Predicting Loyalty Points based on Remuneration
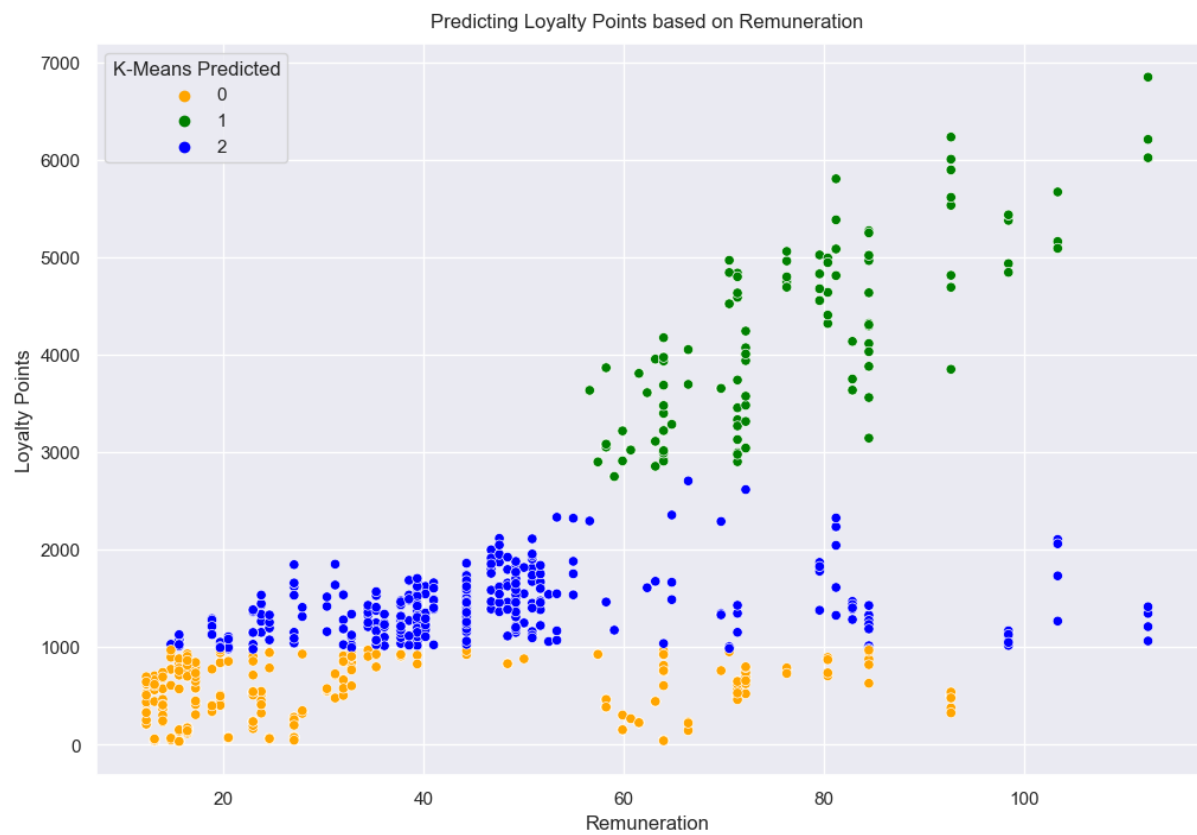
Review and summary content merged and benchmarked at 0.22 for sentiment and 0.45 on subjectivity. Subset created for target remuneration group and benchmarked at 0.24 (target group more positive towards product than whole market).
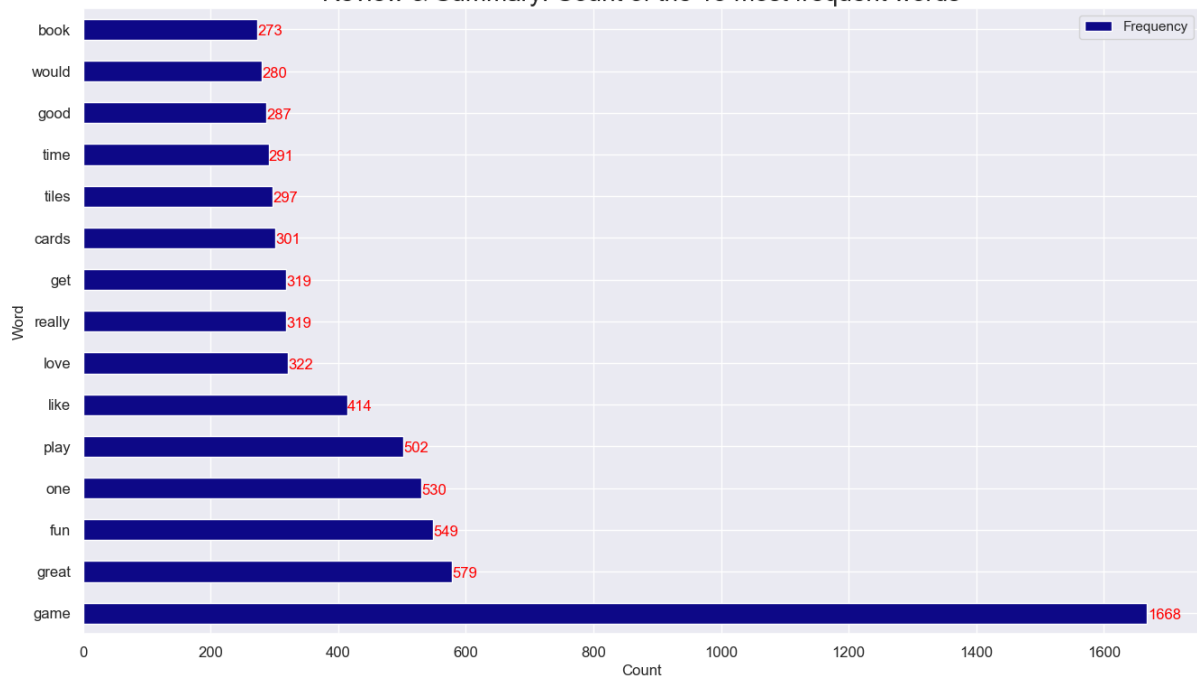
Histogram of sentiment score polarity
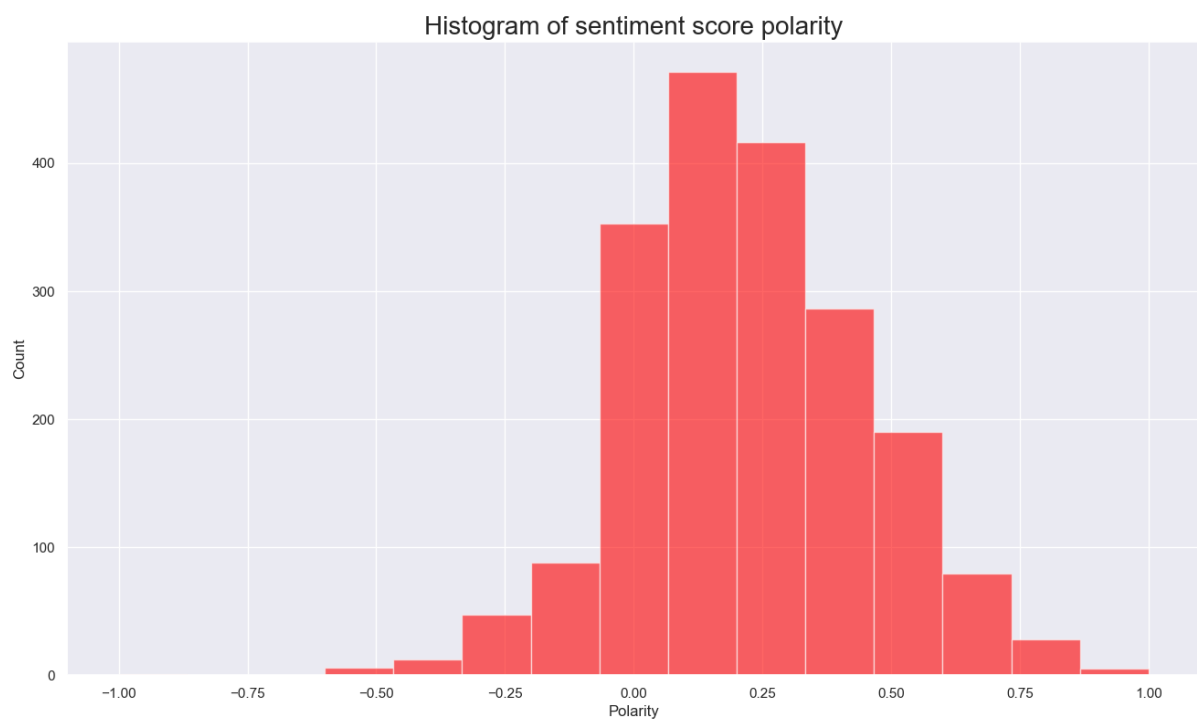
Wordcloud created for use for search engine marketing.
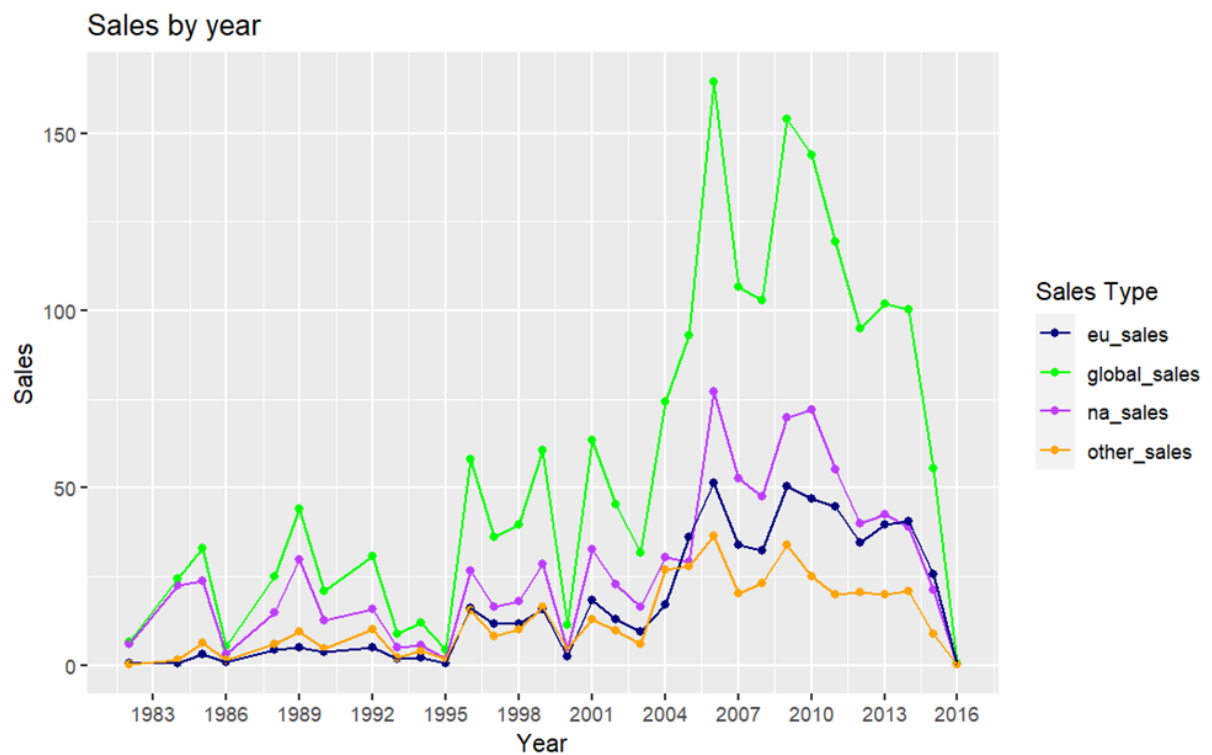
Reviews & Summary words: stopwords excluded

Frequency of top 15 words reflects product demand (games above books) and can be used for search engine marketing.
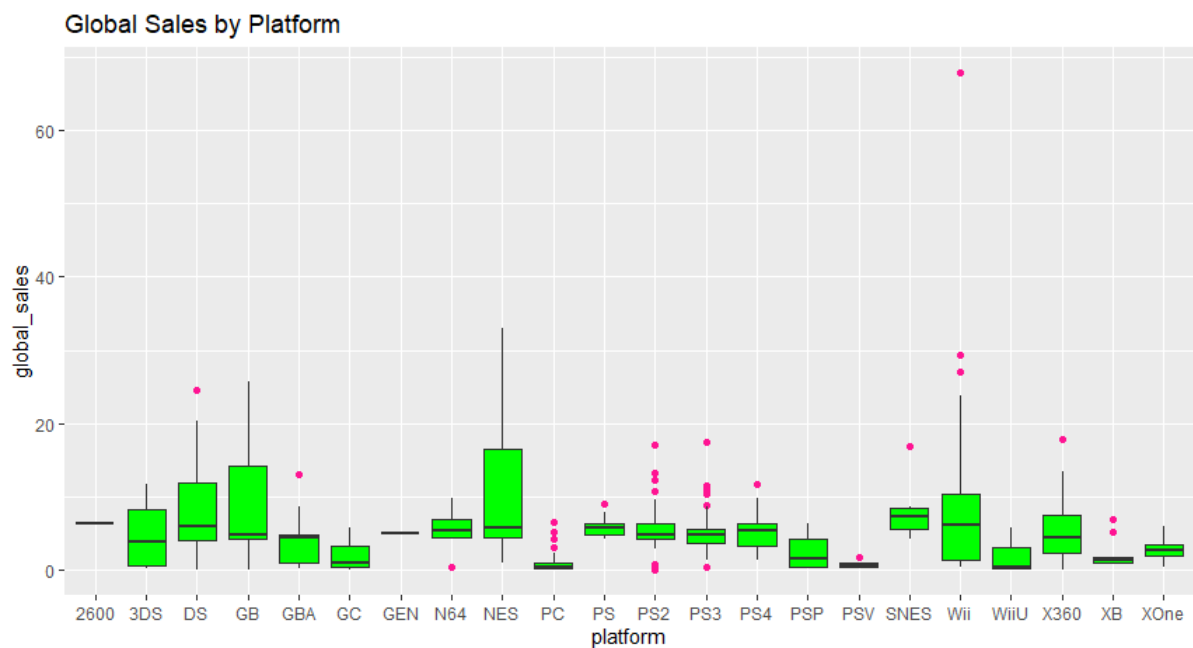
Review & Summary: Count of the 15 most frequent words

Patterns & Predictions
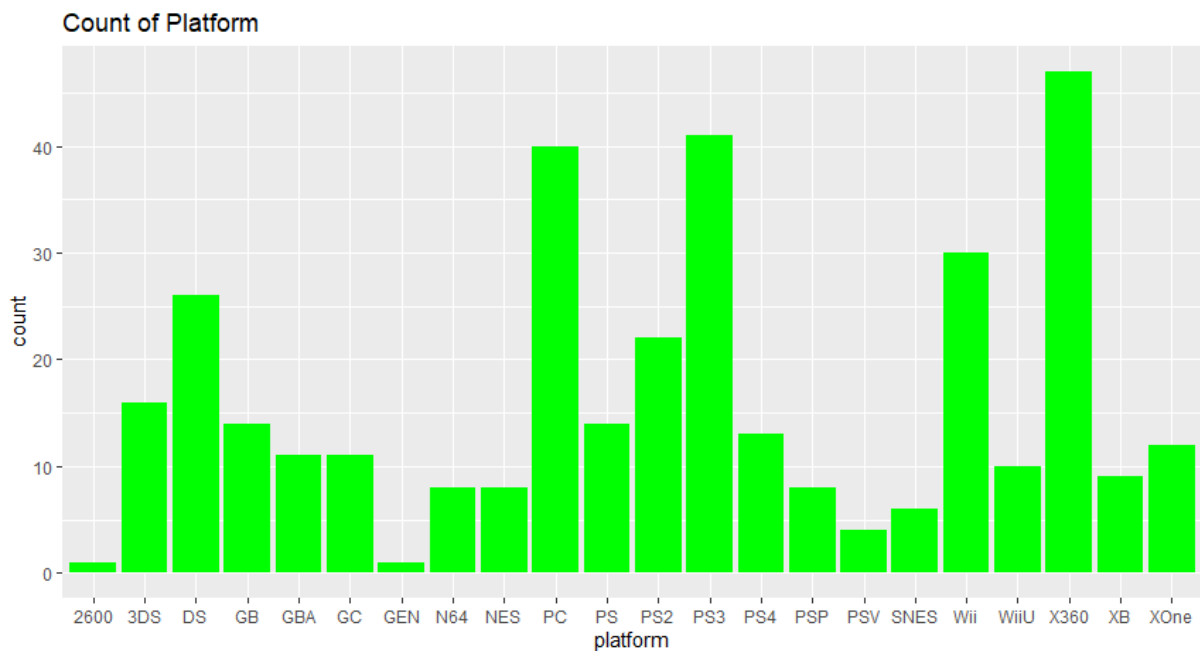
Sales by year

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

Platform analysis shows outlier for sales - wii (although not the highest median sales), with 3 large outlier products.
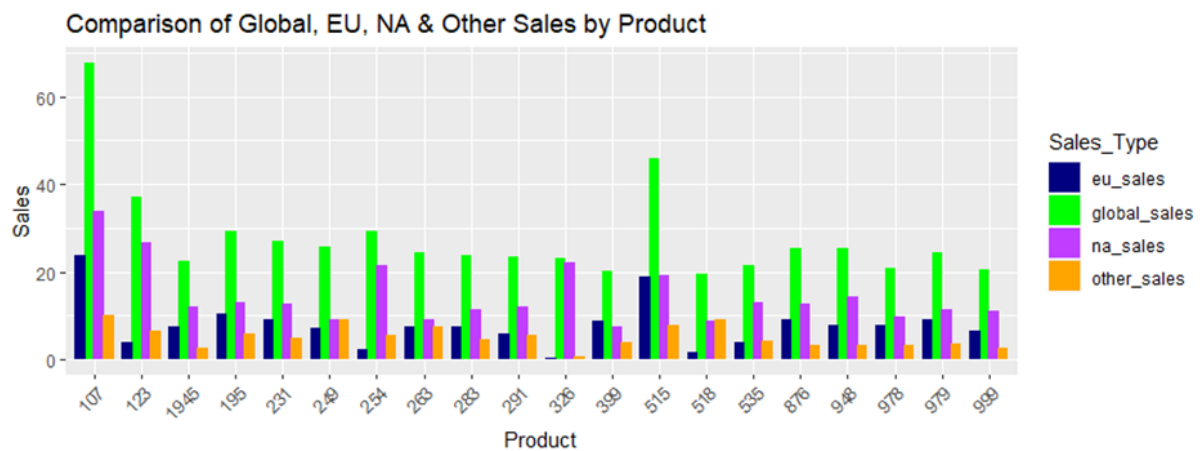
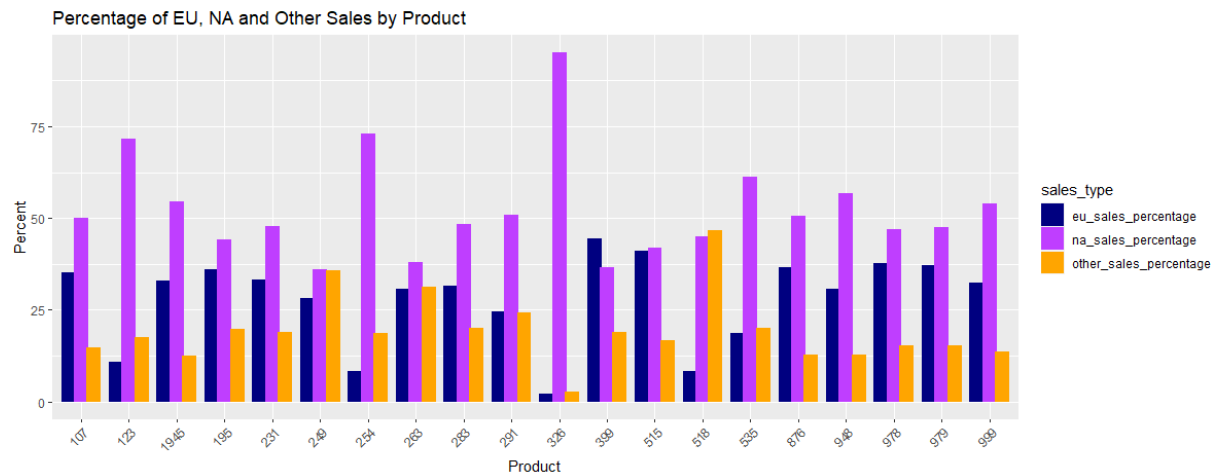Count by platform: wii is lower than X360, PS3 and PC.

Count of Platform

Recommend focus on wii and reduce spend on other platforms, but needs further data on profits, costs and recent data.

Product list subsetted top 20 by global_sales. Highest selling products 107 and 515:

Comparison of Global, EU, NA & Other Sales by Product

Each market as a percentage of global_sales:

Percentage of EU, NA and Other Sales by Product

Dfferent products sell better in different markets. NA 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. Diversify product portfolio by market to spread risk from one high selling product.

Reviewed correlation of Year and NA, EU and other_sales:

Correlation plot from data
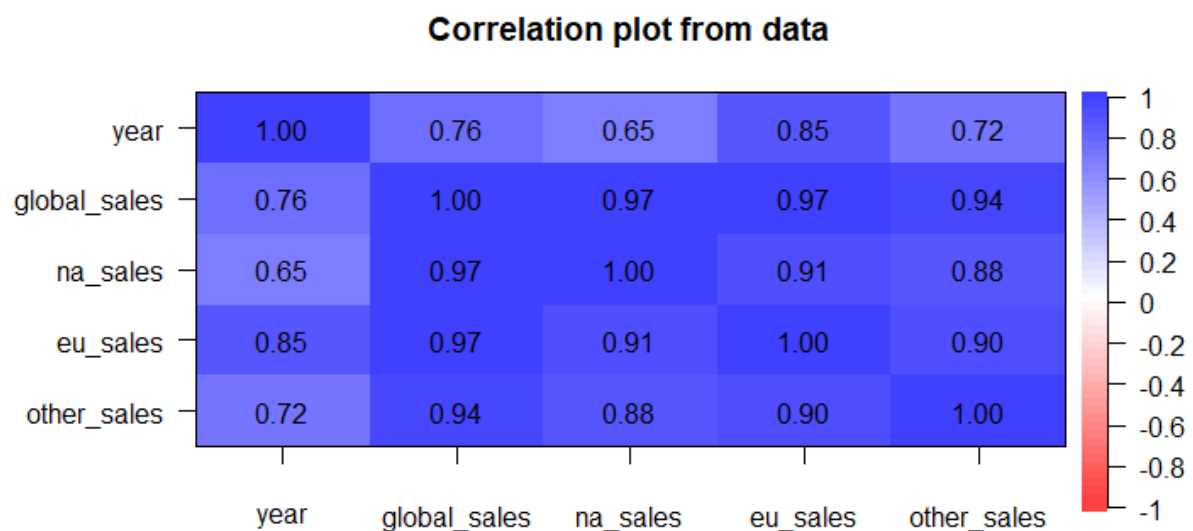
EU, NA and other should be closely correlated as they make up global_sales, but year highly correlated too.

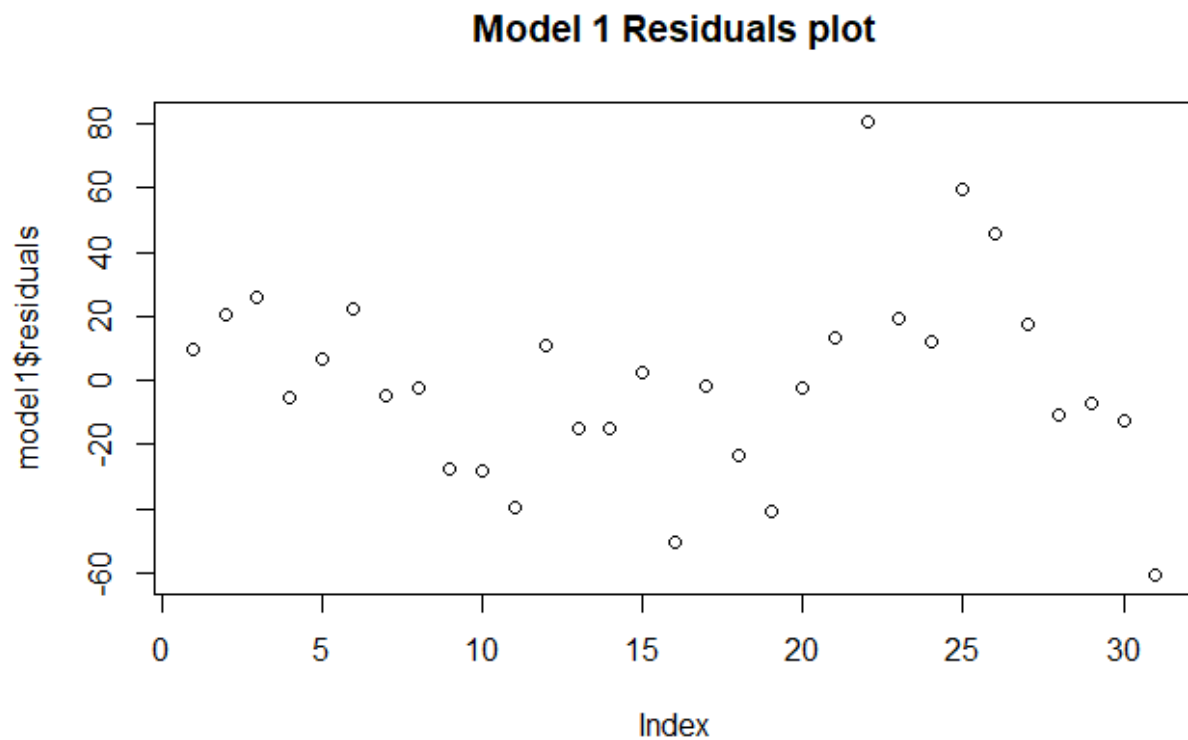Global sales is not normal distribution:

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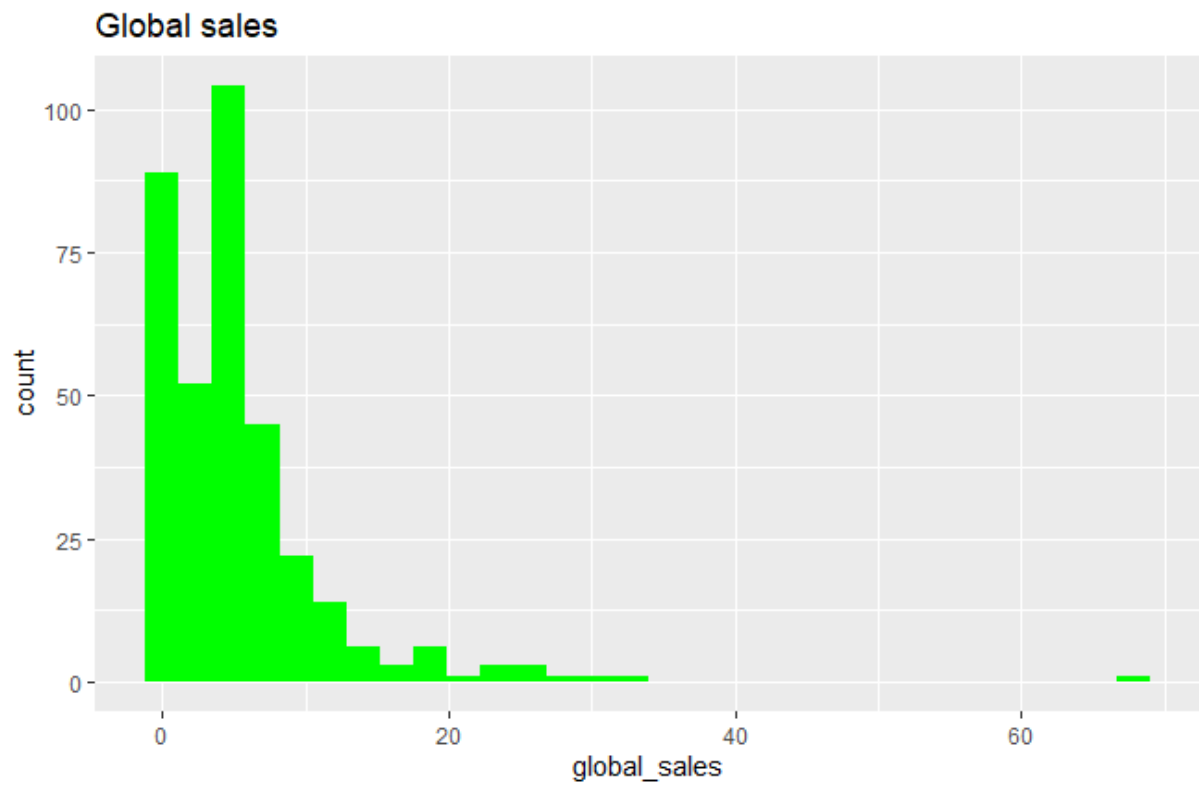
Heavy positive skew and kurtosis (benchmarked for annual review).
Shows sales reliant on single product.

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

Normal Q-Q Plot
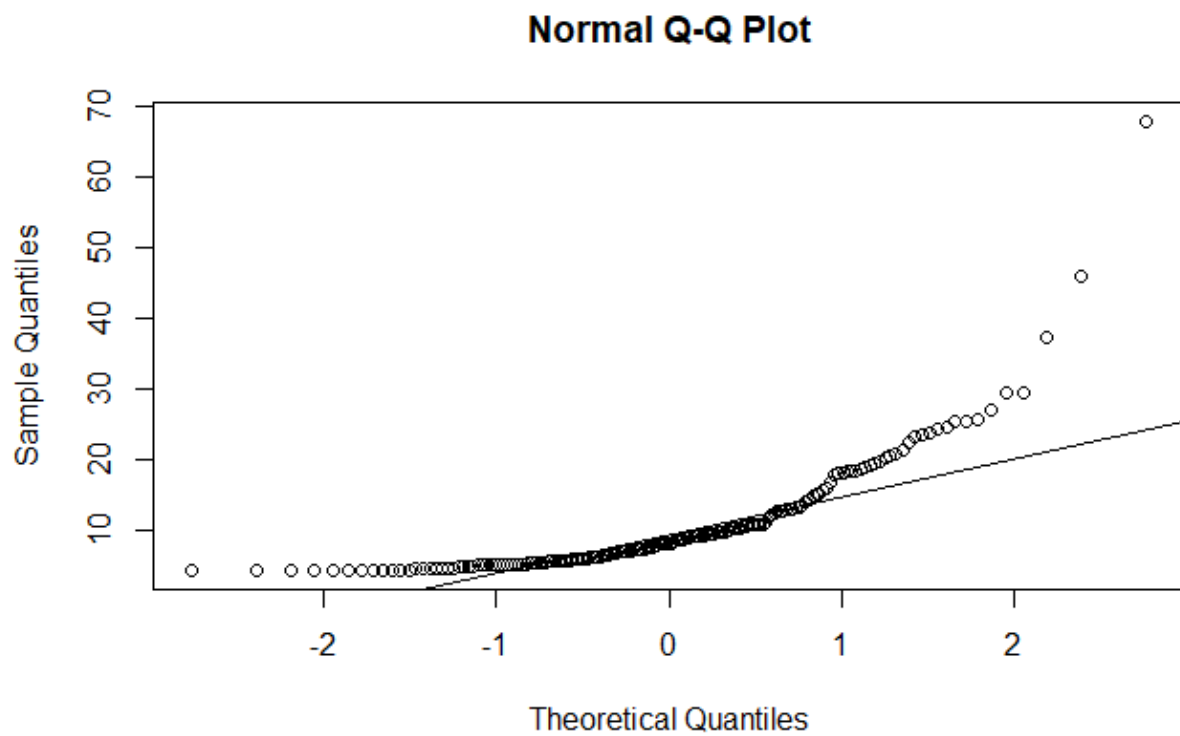
Residuals checked for pattern (no pattern so used linear regression):

Model 1 Residuals plot

Simple linear regression based on year shows 2026 sales at £156.21 million. Each year increases sales by £3.63 million. Note this includes the outlier (highest sales) product, so any instability could affect forecast.

10 year forecast
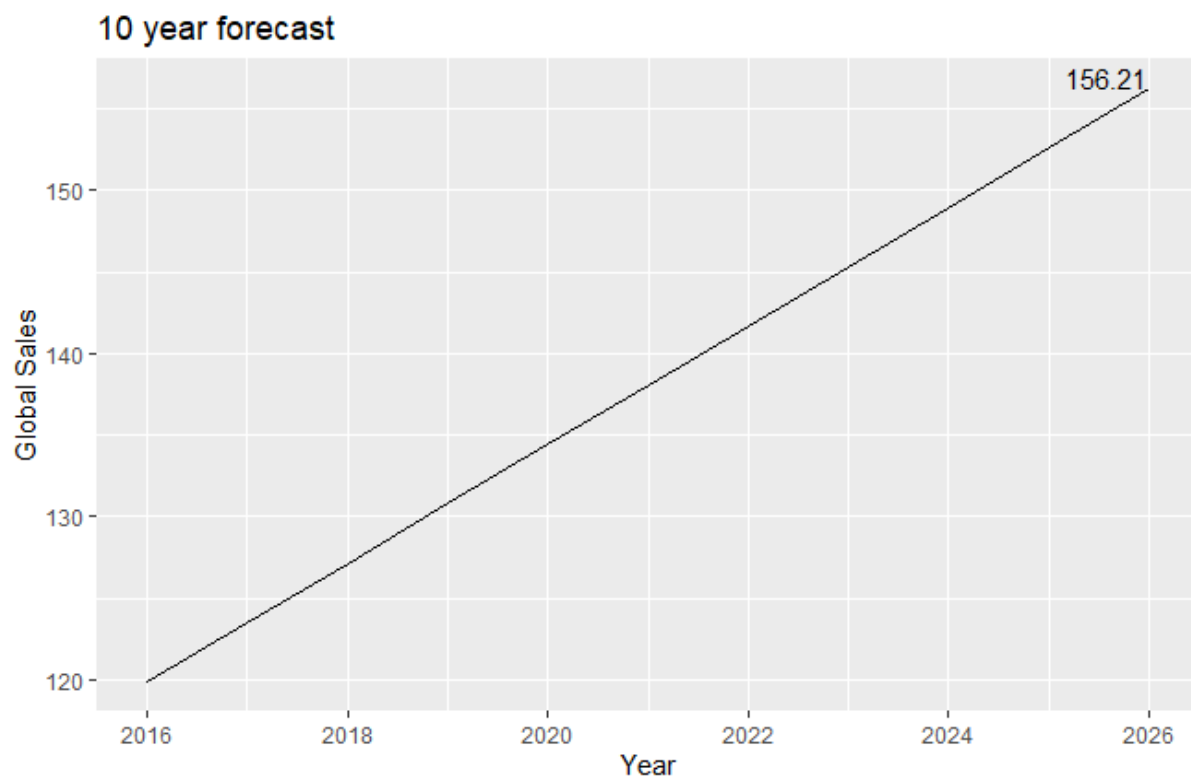
Multiple linear models trained and tested. NA and Year are most accurate predictor variables for global_sales. New data added to model to predict sales for 2016-2020 inclusive:

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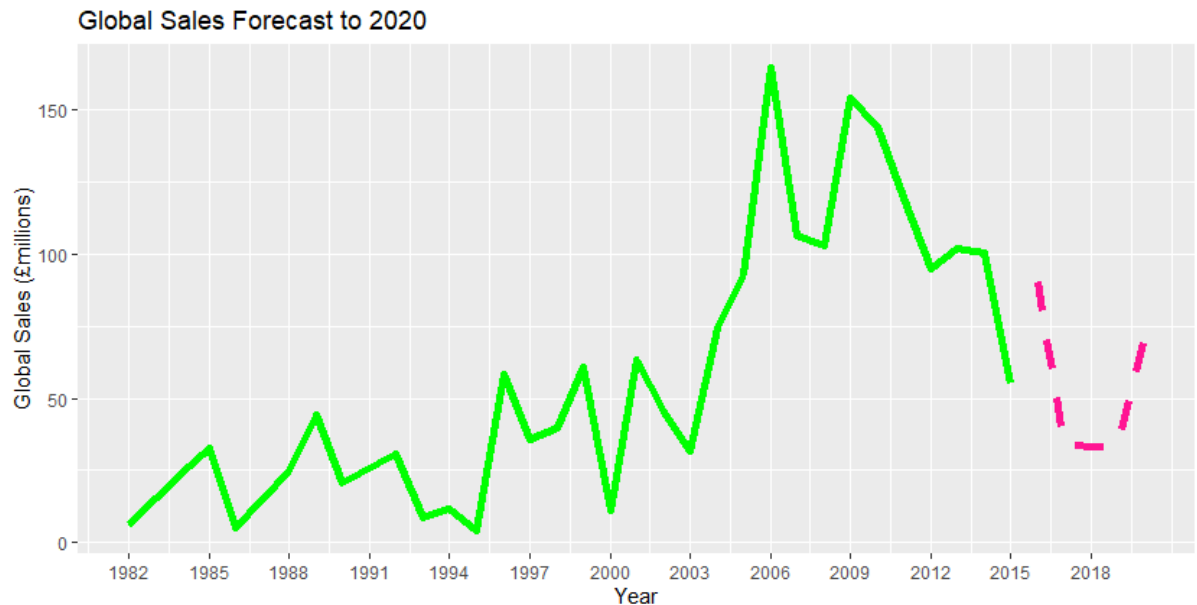
Overall trend continues down from 2010, despite 2016 and 2020 peaks. (Forecasting doesn't take into account pandemic impact on gaming market).

Summary

- Focus on remuneration customers £57-112k driving loyalty
- Use positive reviews for marketing and negative reviews to improve product.
- Use sentiment benchmarks to measure success
- Improve products for older market
- Focus on best performing products by market
- Diversify product portfolio to reduce reliance on highest selling product - benchmark success using skew and kurtosis
- Explore in-game sales and new geographical markets

Feedback: https://forms.gle/zMDe1oZfvyPKguK87

Wordcount: 1088