Marketing team needs to understand how the users accumulate loyalty points. To check if there are any relationships between loyalty points, spendings, remuneration and age of the users and if they can predict, how many loyalties the user accumulates. We define dependant and independent variables answering the question “How many loyalties user can accumulate relative to what’s his remuneration (or spending score or age)?”.

Table1: Spending score and loyalty

OLS Regression Results

Dep. Variable:	y	R-squared:	0.452
Model:	OLS	Adj. R-squared:	0.452
Method:	Least Squares	F-statistic:	1648.
Date:	Fri, 16 Dec 2022	Prob (F-statistic):	2.92e-263
Time:	17:04:58	Log-Likelihood:	-16550.
No. Observations:	2000	AIC:	3.310e+04
Df Residuals:	1998	BIC:	3.312e+04
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	-75.0527	45.931	-1.634	0.102	-165.129	15.024
x	33.0617	0.814	40.595	0.000	31.464	34.659

Omnibus:	126.554	Durbin-Watson:	1.191
Prob(Omnibus):	0.000	Jarque-Bera (JB):	260.528
Skew:	0.422	Prob(JB):	2.67e-57
Kurtosis:	4.554	Cond. No.	122.

Table2: Remuneration and loyalty

OLS Regression Results

Dep. Variable:	y	R-squared:	0.380
Model:	OLS	Adj. R-squared:	0.379
Method:	Least Squares	F-statistic:	1222.
Date:	Fri, 16 Dec 2022	Prob (F-statistic):	2.43e-209
Time:	17:18:18	Log-Likelihood:	-16674.
No. Observations:	2000	AIC:	3.335e+04
Df Residuals:	1998	BIC:	3.336e+04
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	-65.6865	52.171	-1.259	0.208	-168.001	36.628
x	34.1878	0.978	34.960	0.000	32.270	36.106

Omnibus:	21.285	Durbin-Watson:	3.622
Prob(Omnibus):	0.000	Jarque-Bera (JB):	31.715
Skew:	0.089	Prob(JB):	1.30e-07
Kurtosis:	3.590	Cond. No.	123.

Table 3: Age and loyalty

OLS Regression Results

Dep. Variable:	y	R-squared:	0.002
Model:	OLS	Adj. R-squared:	0.001
Method:	Least Squares	F-statistic:	3.606
Date:	Fri, 16 Dec 2022	Prob (F-statistic):	0.0577
Time:	17:22:57	Log-Likelihood:	-17150.
No. Observations:	2000	AIC:	3.430e+04
Df Residuals:	1998	BIC:	3.431e+04
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	1736.5177	88.249	19.678	0.000	1563.449	1909.587
x	-4.0128	2.113	-1.899	0.058	-8.157	0.131

Omnibus:	481.477	Durbin-Watson:	2.277
Prob(Omnibus):	0.000	Jarque-Bera (JB):	937.734
Skew:	1.449	Prob(JB):	2.36e-204
Kurtosis:	4.688	Cond. No.	129.

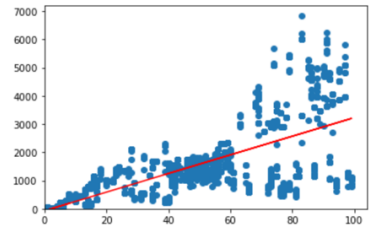
Spendings vs loyalty model has the highest R^2, meaning that 45.2% of the total variability of loyalties, is explained by the variability of spendings. The probability of t-value shows the slope is significant.

Remuneration vs loyalty model has smaller $R^2$of 38%, and statistically significant coefficient of 34.1878.

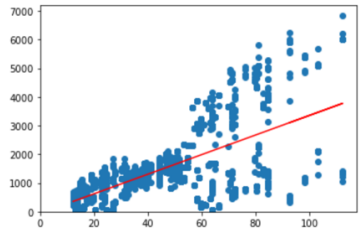
Age vs loyalty model shows the worst fit with R^2 only 2% and a negative coefficient (underlying the reverse relationships with loyalty), but it is not statistically significant, and we can’t consider this model for prediction of loyalties.

The line is the best fit found by minimising residuals. It confirms that age can’t be used as a predictor for loyalties.

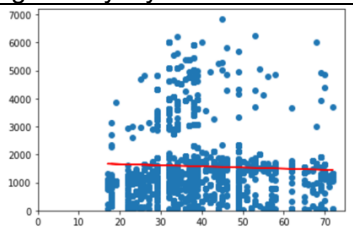
Spending score vs loyalty



Remuneration vs loyalty



Age vs loyalty



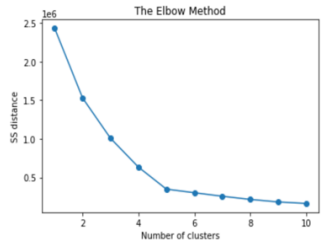
To find a better fit for the model Multiple Linear Regression (MLR) used with two independent variables we looked at before. Apart from already added libraries, we

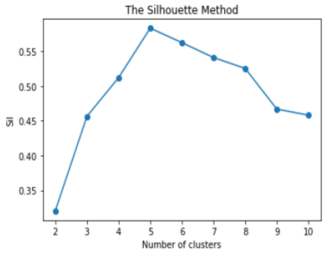
will need sklearn to apply the LR algorithm and statsmodels API for statistical analysis. The R^2 is 82.8% now, gives better fit and all coefficients are statistically significant.

```
                            OLS Regression Results
==============================================================================
Dep. Variable:         loyalty_points   R-squared:                       0.828
Model:                            OLS   Adj. R-squared:                  0.828
Method:                 Least Squares   F-statistic:                     3843.
Date:                Thu, 22 Dec 2022   Prob (F-statistic):               0.00
Time:                        14:39:39   Log-Likelihood:                -12302.
No. Observations:                1600   AIC:                         2.461e+04
Df Residuals:                    1597   BIC:                         2.463e+04
Df Model:                           2
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const      -1686.4121     39.749    -42.426      0.000   -1764.379   -1608.446
remuneration    33.7573      0.576     58.627      0.000      32.628      34.887
spending_score    32.9480      0.512     64.377      0.000      31.944      33.952
==============================================================================
Omnibus:                        2.689   Durbin-Watson:                   2.020
Prob(Omnibus):                  0.261   Jarque-Bera (JB):                2.591
Skew:                           0.079   Prob(JB):                        0.274
Kurtosis:                       3.117   Cond. No.                         222.
==============================================================================
```

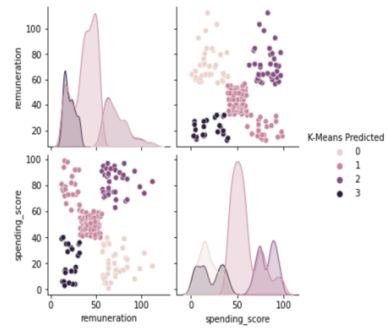
There was no multicollinearity found, based on VIF factor. However, p-value for LM statistic is less than 0.05 and there is an evidence of heteroscedasticity present which can make coefficients less precise and less statistically significant.

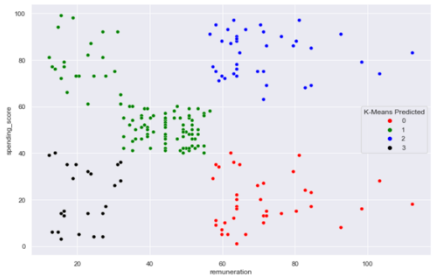
We need to identify groups that can be used to target specific market segments. K-means clustering model is used. Using Elbow method, we can say that there are between 4 and 6 clusters. According to silhouette score 5 or 6 looks more appropriate.





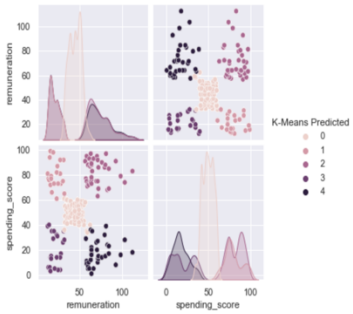
4 clusters:

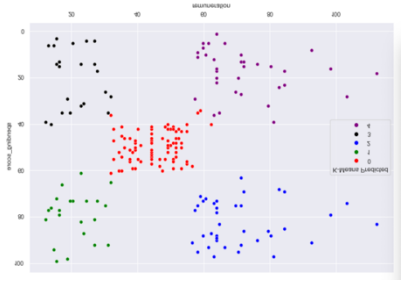




```
0      351
1     1013
2      356
3      280
```

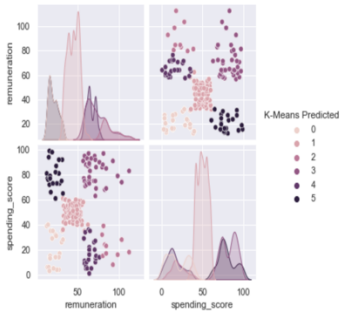
5 clusters:





```
0      774
1      269
2      356
3      271
4      330
```

6 clusters:



```
0      271
1      767
2      123
3      356
4      214
5      269
```

Based on the classification of 2,000 observations of spending scores and remunerations, 5 clusters seem to give the best results to target specific market segments.

Using reviews and summaries we try to use social data for marketing purposes and see general feel about Turtle Games' products among the customers.

We prepare text data for NLP and remove all the "noise": lower case applied, punctuations and duplicates removed. Tokenization of reviews and summaries applied: dividing long sentences into separate words and saving them withing a single row. When we combine all tokens as a single list of words, we can find tokens with maximum frequency.

To visualise the most frequently occurring words WordCloud library was used.

Reviews:

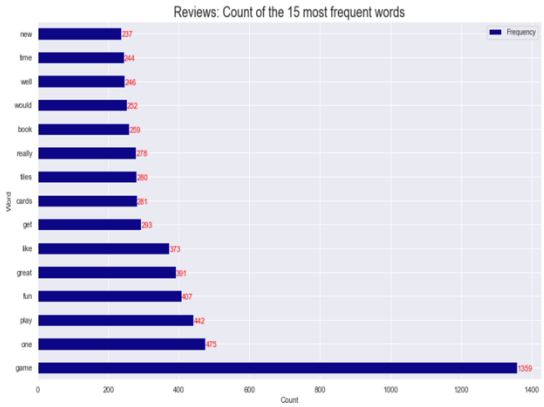
Summaries:

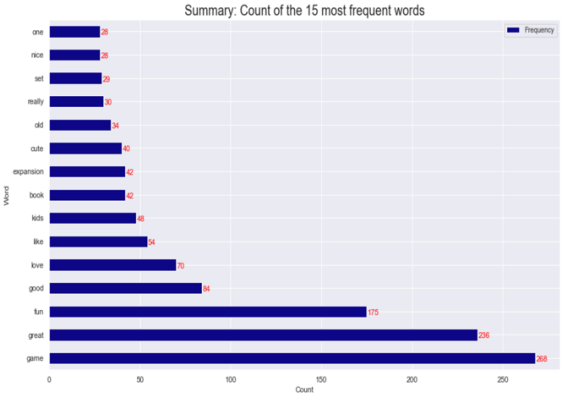
To improve the results, stopwords (English) removed.

Reviews (stopwords removed):

Summaries (stopwords removed):

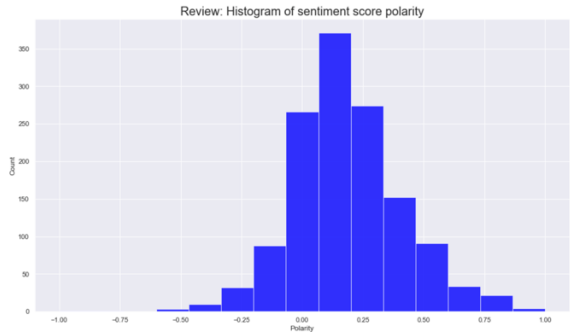
The same results, but with calculated frequency, are presented as bar plots for the top 15.

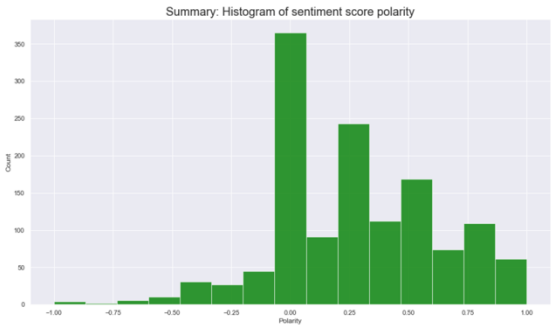




The sentiment analysis with TextBlob library to define polarity and subjectivity.

Reviews:	Summaries:





Most reviews located in positive zone as well as in summaries more positive ones can be seen.

Top 20 negative and positive reviews and summaries are extracted based on polarity scores. Positive polarity reviews ratings match from human perspective as well. However, negative ones are confused sometimes. A good example would be this review marked in top negatives, but in fact it is a positive one.

```
In [123]: 1 negative_sentiment.at[355, 'review']
Out[123]: 'my son loves playing this game it was recommended by a counselor at school that works with him'
```

It is worth doing more research into reviews with complains to find out what is wrong with these products some customers mentioned:

```
In [124]: 1 negative_sentiment.at[141, 'review']
Out[124]: 'i sent this product to my granddaughter the pompom maker comes in two parts and is supposed to snap together to cr
          eate the pompoms however both parts were the same making it unusable if you cant make the pompoms the kit is useles
          s since this was sent as a gift i do not have it to return very disappointed'
```

Top 20 negative reviews:

	review	polarity_review	subjectivity_review
165	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
147	incomplete kit very disappointing	-0.780000	0.910000
267	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
90	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
141	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
251	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
382	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
312	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
355	my son loves playing this game it was recommended by a counselor at school that works with him	-0.400000	0.400000
713	if you like me used to play dd but now you and your friends growed 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
1011	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
723	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
600	i was a bit disappointed in the quality of the cardboard piecesholders and the fact that they changed the names of some hotels otherwise i mean its a terrific game	-0.365625	0.709375
331	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
297	i really like this game it helps kids recognize anger and talk about difficult emotions	-0.350000	0.450000
389	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
338	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
4	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
784	the adventures are tough but you can get throuhg them it all comes down to the die roll just like any dd game	-0.314815	0.507407
631	a crappy cardboard ghost of the original hard to believe they did this but they did shame on hasbro disgusting	-0.305556	0.763889

Top 20 negative summaries:

	summary	polarity_summary	subjectivity_summary
17	the worst value ive ever seen	-1.000000	1.000000
165	boring unless you are a craft person which i am	-1.000000	1.000000
587	boring	-1.000000	1.000000
837	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
116	disappointed	-0.750000	0.750000
266	promotes anger instead of teaching calming methods	-0.700000	0.200000
634	too bad this is not what i was expecting	-0.700000	0.666667
637	bad qualityall made of paper	-0.700000	0.666667
144	at age 31 i found these very difficult to make	-0.650000	1.000000
75	small and boring	-0.625000	0.700000
368	mad dragon	-0.625000	1.000000
575	disappointing	-0.600000	0.700000
723	then you will find this board game to be dumb and boring	-0.591667	0.633333
267	anger control game	-0.550000	0.300000
59	really small disappointed	-0.500000	0.575000
360	its uno for the angry	-0.500000	1.000000
646	50th anniversary is a sad day for acquire	-0.500000	1.000000
808	a disappointing coop game	-0.500000	0.550000
1116	its also really lame that the doll didnt come with the things she	-0.500000	0.750000

Top 20 positive reviews:

	review	polarity_review	subjectivity_review
564	perfect	1.000000	1.000000
1080	my daughter loves her stickers awesome seller thank you	1.000000	1.000000
1334	perfect for tutoring my grandson in spelling	1.000000	1.000000
890	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
498	great quality very cute and perfect for my toddler	0.816667	0.916667
31	the pictures are great ive done one and gave it to a friend of mine who likes dragons	0.800000	0.750000
336	great seller happy with my purchase 5 starrrr	0.800000	0.875000
439	great easter gift for kids	0.800000	0.750000
491	these are great	0.800000	0.750000
692	bought this because i wanted it all these dd games are great	0.800000	0.750000
824	husband seems happy with it	0.800000	1.000000
826	great accessory to use with the playing mat	0.800000	0.750000
828	great price arrived on time with no damage will be a great addition to my collection	0.800000	0.750000
893	this is a great accessory to the starter set i would recommend this to anyone who owns the starter set	0.800000	0.750000
1075	my granddaughter loves these so happy to find peppa pig items for her	0.800000	1.000000
1113	great doll to go with the book animals cant wait to read book with the doll to the grandkids	0.800000	0.750000
1187	a great creation tool it helps me concentrate	0.800000	0.750000
1287	prompt service and a great product	0.800000	0.750000
1333	this is a great tool to have at hand when playing quiddler	0.800000	0.750000
325	this is a great product i use it as a therapeutic tool and it has been very effective	0.790000	0.875000

Top 20 positive summaries:

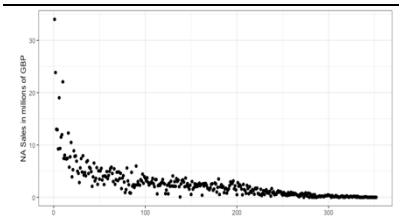
	summary	polarity_summary	subjectivity_summary
5	best gm screen ever	1.000000	0.300000
23	wonderful designs	1.000000	1.000000
27	perfect	1.000000	1.000000
61	theyre the perfect size to keep in the car or a diaper	1.000000	1.000000
107	perfect for preschooler	1.000000	1.000000
112	awesome sticker activity for the price	1.000000	1.000000
132	awesome book	1.000000	1.000000
133	he was very happy with his gift	1.000000	1.000000
150	awesome	1.000000	1.000000
166	awesome and welldesigned for 9 year olds	1.000000	1.000000
337	excellent	1.000000	1.000000
389	excellent therapy tool	1.000000	1.000000
407	the pigeon is the perfect addition to a school library	1.000000	1.000000
423	best easter teaching tool	1.000000	0.300000
462	wonderful	1.000000	1.000000
466	all f the mudpuppy toys are wonderful	1.000000	1.000000
471	awesome puzzle	1.000000	1.000000
476	not the best quality	1.000000	0.300000
514	excellent puzzle	1.000000	1.000000
521	the best feedback i can have	1.000000	0.300000

To help sales department find insights about sales by product in North American (NA), European (EU) markets and globally, R was used.

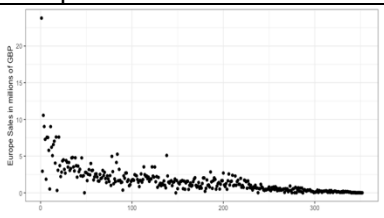
After data was sense-checked, cleaned, unnecessary variables removed and missing data checked, visualisations used to see if there are any patterns.

Scatterplots:

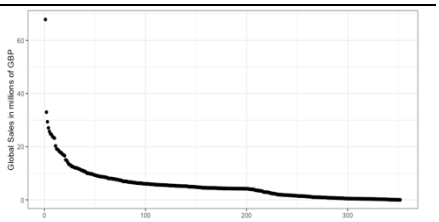
North America Sales:



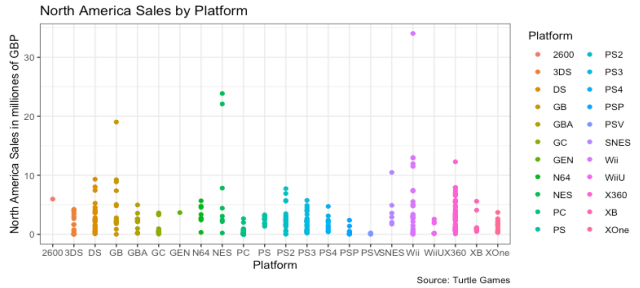
Europe Sales:



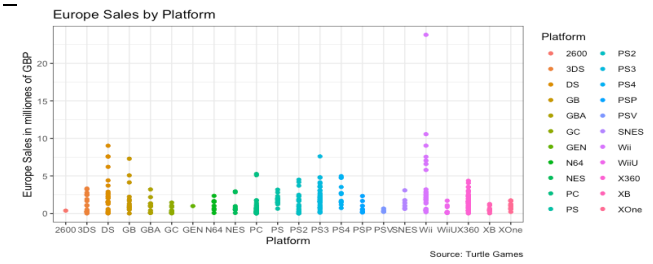
Global Sales:



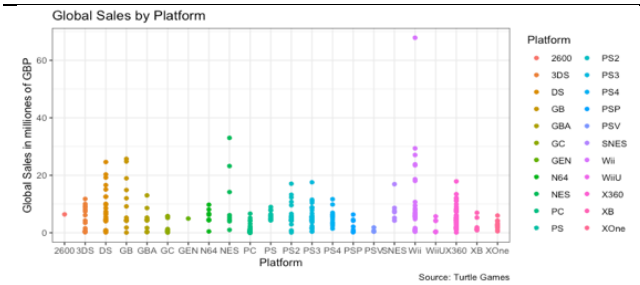
Largest part of NA sales is under £5m, for EU under £2m. There are some noticeable leaders on both markets with sales over £30m. (NA) and over £20m. (EU).



- Wii game over £30m. bestseller
- GB game around £20m.
- NES with 2 games over £20m.
- Xbox360 popular.



- Wii over £20m. game
- PC more popular than NA

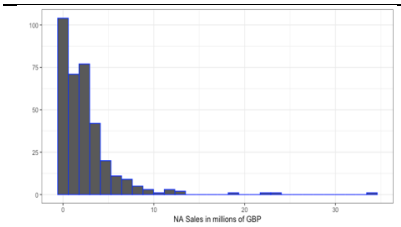


- Wii game has the biggest sale.
- Xbox360 and PS are popular platforms.

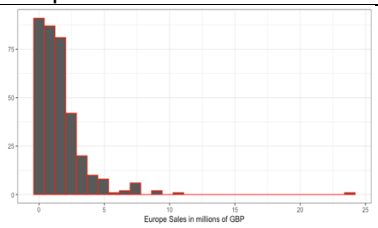
Histograms:

They show more sales in the region below £3m. for N. American and £1.5m. for European market, globally it is £5-6m. NA market has also more sales over £10m and contribute more into Global sales.

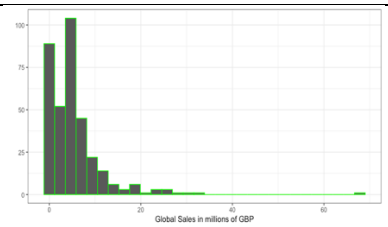
North America Sales:



Europe Sales:



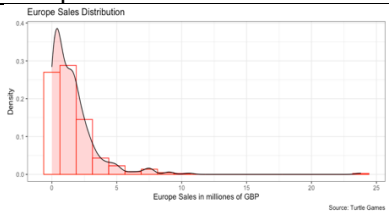
Global Sales:



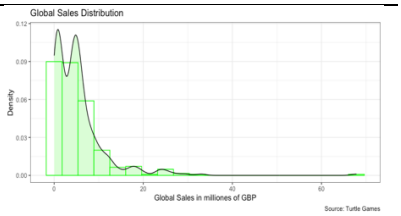
All three shows not normal, right skewed distribution:

North America Sales:

Europe Sales:

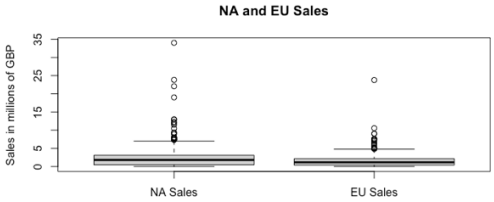


Global Sales:

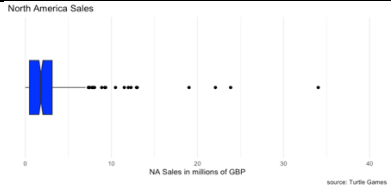


Boxplots:

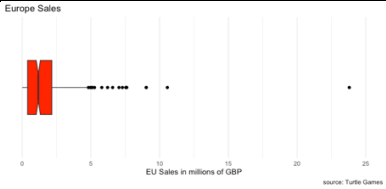
As expected from the previous plots, there are lots of outliers presented with more than 50% of sales below £5m. for NA market, and for EU below £2.5m. NA has larger number of outliers and greater variability. Global sales have the same tendency. Boxplots confirm not normal distribution of right-skewed data.



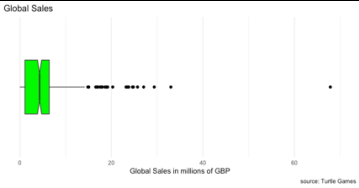
North America Sales:



Europe Sales:



Global Sales:

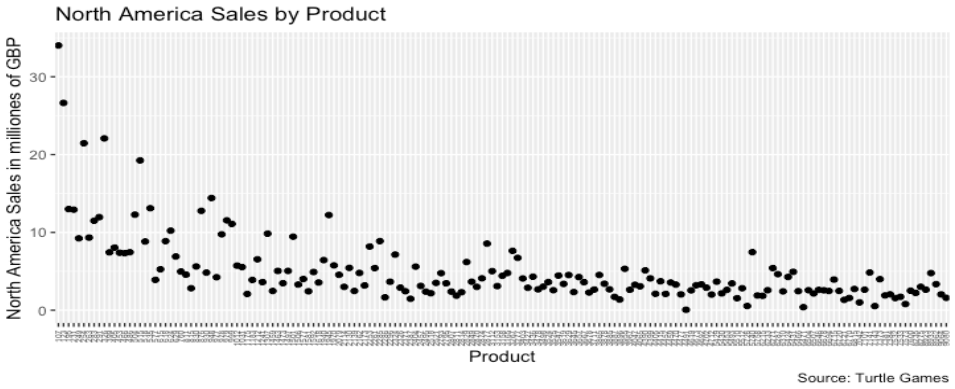


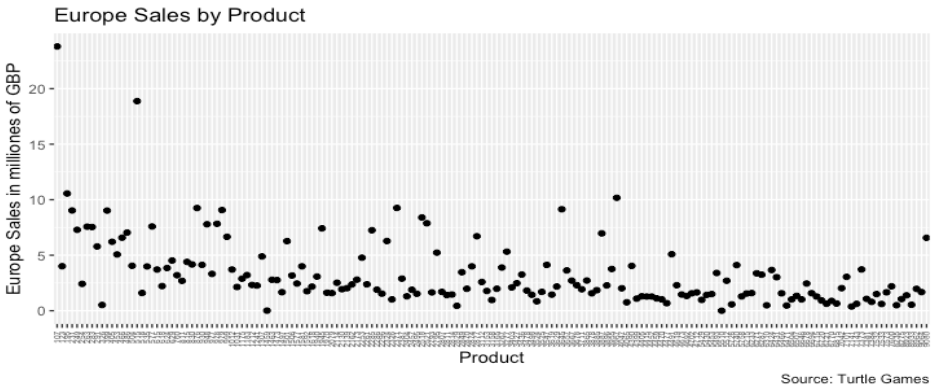
Let's check impact on sales per product to get more insights about the market sales.

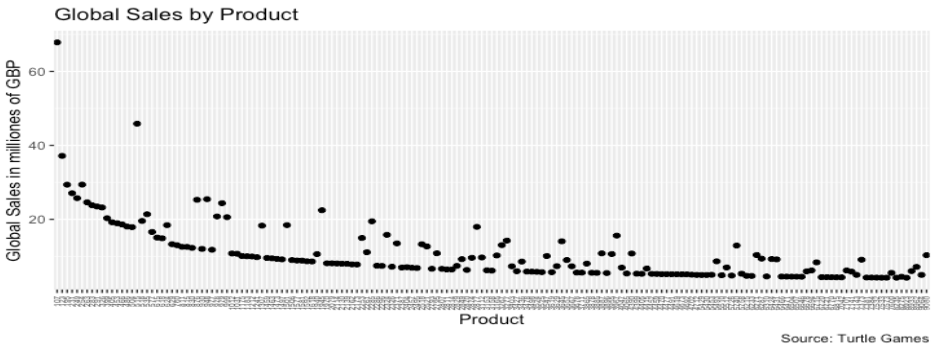
	Min	1st Quartile	Median	Mean	3rd Quartile	Max
NA Sales	0.0000	0.4775	1.8200	2.5160	3.1250	34.020
EU Sales	0.0000	0.390	1.170	1.644	2.160	23.800
Global Sales	0.010	1.115	4.320	5.335	6.435	67.850

- Average Global sales are £5.34m. with maximum of £ 67.85m. in one game sales.
- North America has higher average sales compared to Europe £2.52m.
- European average sales are £1.64m.

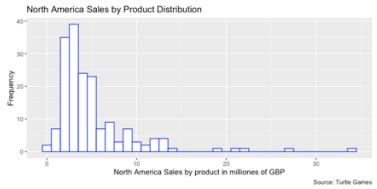
Most of games revenues are below £5m. Clear outliers with the highest game sale over 30m. EU sales by product are smaller: most of them under £2.5m. Globally, most of the product sales are under £10m.

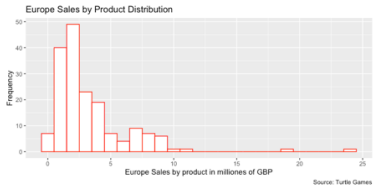


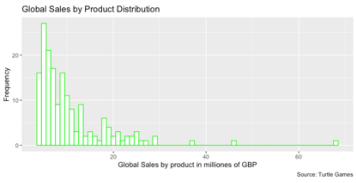




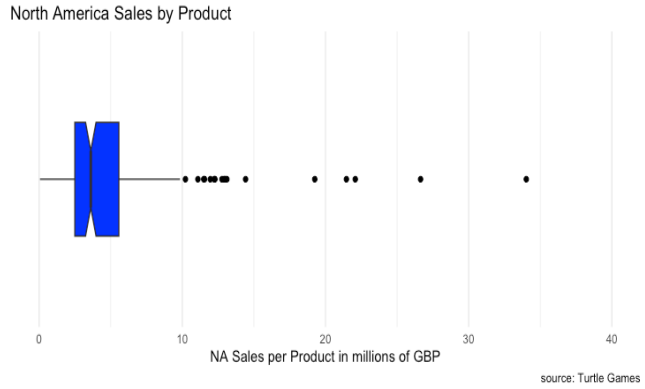
All three sales by product frequency distributions are not normal and look right skewed.

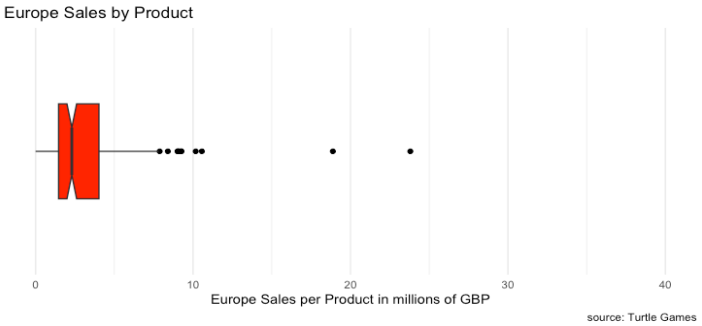


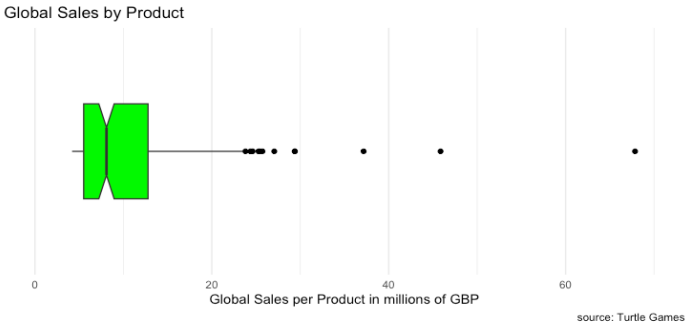




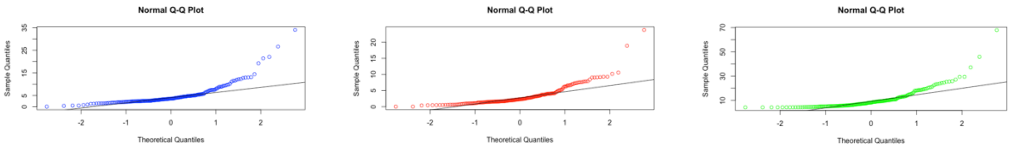
50% of NA sales by product is under £6m. while EU is under £4m. Global sale has IQR of about £4m. Everywhere median value is closer to lower quartile (right skewed) and the notches don't overlap (with 95% confidence, the true medians differ). There are more outliers in NA market, and they are more extreme, than in Europe.







Normality:

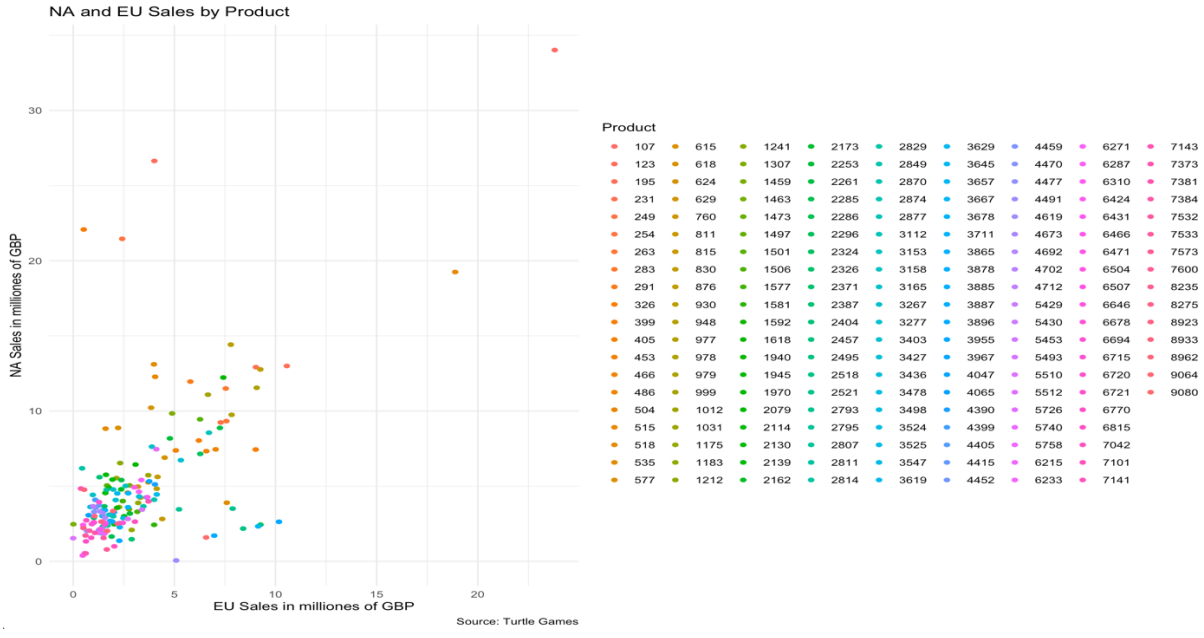
	NA Product Sales	EU Product Sales	Global Product Sales
W =	0.69813	0.74058	0.70955
p-value<	2.2e-16 <0.05	2.987e-16<0.05	2.2e-16<0.05
QQ Plots			
Skewness	3.048198	2.886029	3.066769
Kurtosis	15.6026	16.22554	17.79072

The p-value < 0.05 for all sales implying that the distributions of the data are significantly different from normal. QQ plots shows right skewness, with real sales go much over theoretical (normal) starting 1SD especially after 2SD. Skewness is confirmed. Excess kurtosis confirms heavy tails with outliers.

Correlation is positive for all data, with the highest between NA and Global sales, showing the bigger influence of NA market.

	NA Product Sales	EU Product Sales	Global Product Sales
NA Product Sales	1		
EU Product Sales	0.6209317	1	
Global Product Sales	0.9162292	0.8486148	1

Scatterplots are the most informative for the purpose of sales research. It is worth looking into which games are popular on different markets. How markets compare? Which market would be better for launching new products? Which platforms are more popular in EU and NA? Games by which publishers have biggest sales?



We investigated best-selling platforms, publishers, and games:

Platforms:

- NA has X360 (can be explained as it is Microsoft), followed by Wii (Nintendo) and PS3 (Sony product).
- For EU: Wii, followed by PS3 and X360.
- For X360, where NA revenues are twice more than EU.
- NES games, where NA sales 7 times more, is not popular in Europe.
- PC games are more popular in Europe as well as PS3 and PS4.
- Globally, Wii is number one, X360 and PS3.
- Overall, NA and EU have almost 80% of the market.

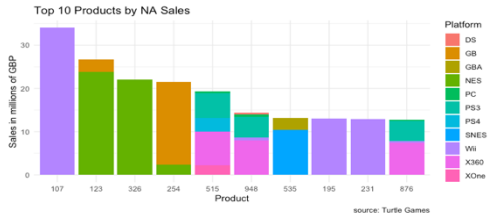
Publishers:

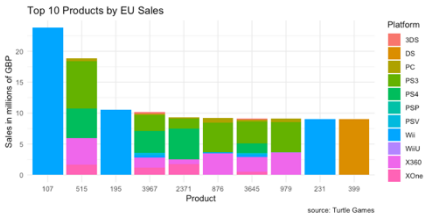
- Nintendo is an absolute leader everywhere with almost 50% of sales.
- In EU Electronic Arts is the second, explained by FIFA most probably, which is much more popular in EU.
- Activision is in top 3.

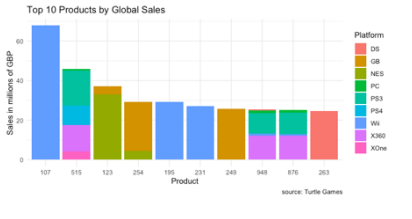
Individual games:

- Product 107 for Wii is the most popular everywhere.
- 123 for NES and GB, being number 2 in NA, is not in top10 in EU (as any other NES or GB), as well as 326 (NES).
- At the same time EU has 3967 and 2371 which have smaller revenues on NA market or 399 (PC) game.
- Globally we can see that games 107, 195 and 231 are sold the best, all of them Wii platform.

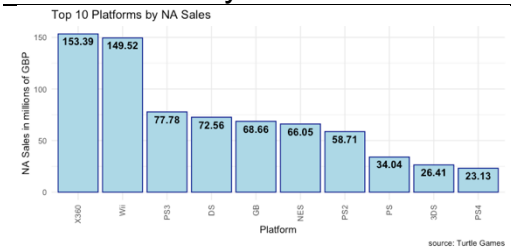
Top 10 Games by Sales:

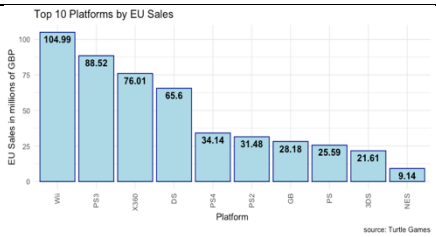




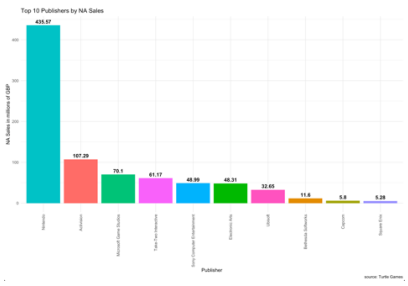


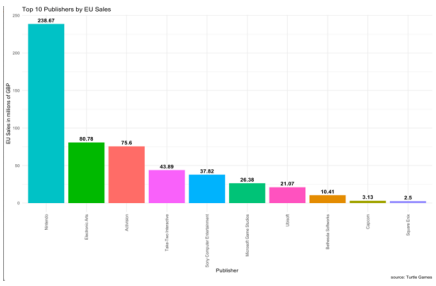
10 Platforms by Sales:

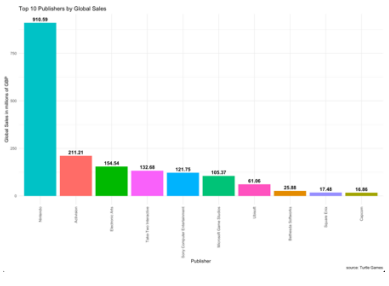




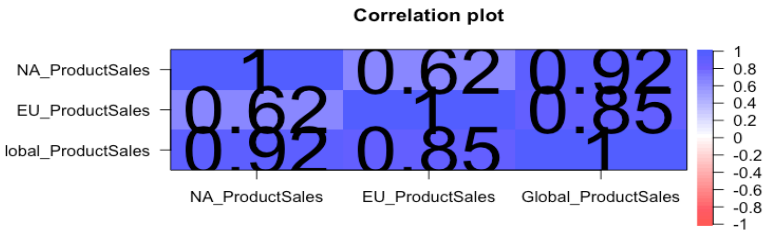
Top 10 Publishers by Sales:







As NA, EU and Global sales are positively correlated (if one goes up another goes up too), we can use regression model to find if there are any relationships between them.



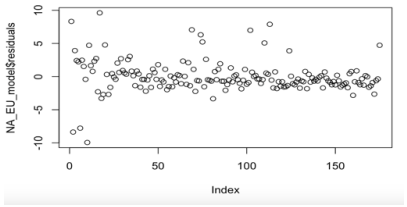
SLR: NA product sales predict EU?

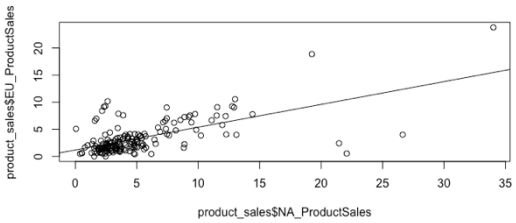
```
Residuals:
    Min      1Q  Median      3Q     Max
-9.9391 -1.1930 -0.4267  0.7023  9.6102

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   1.17946    0.27433   4.299 2.85e-05 ***
NA_ProductSales  0.42028    0.04034  10.419  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.424 on 173 degrees of freedom
Multiple R-squared:  0.3856,    Adjusted R-squared:  0.382
F-statistic: 108.6 on 1 and 173 DF,  p-value: < 2.2e-16
```

- R squared is 38.56%
- Slope is statistically significant (EU product sales go up 0.4243 when NA changes 1 unit).





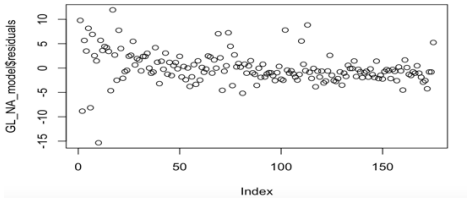
SLR: How Global product sales change if NA change?

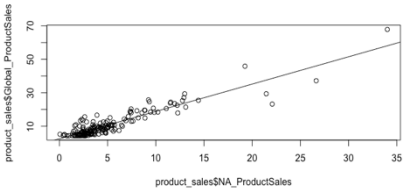
```
Residuals:
     Min       1Q   Median       3Q      Max
-15.3417  -1.8198  -0.5933   1.4322  11.9345

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   2.45768    0.36961   6.649 3.71e-10 ***
NA_ProductSales  1.63469    0.05435  30.079  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.266 on 173 degrees of freedom
Multiple R-squared:  0.8395,    Adjusted R-squared:  0.8385
F-statistic: 904.7 on 1 and 173 DF,  p-value: < 2.2e-16
```

- R squared is 83.95%.
- Slope is statistically significant (Global product sales go up 1.635 when NA changes 1 unit).





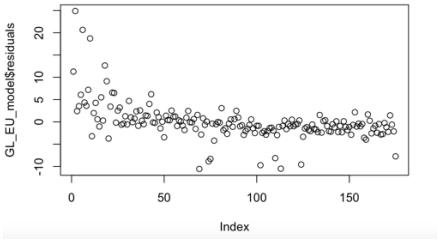
SLR: How Global sales change if EU change?

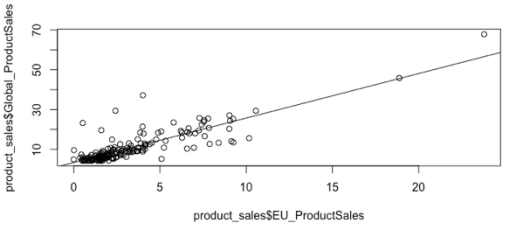
```
Residuals:
     Min       1Q   Median       3Q      Max
-10.5583  -1.7530  -0.5371   0.9586  24.8556

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)       3.3343     0.4787   6.965 6.57e-11 ***
EU_ProductSales   2.2369     0.1060  21.099  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4.313 on 173 degrees of freedom
Multiple R-squared:  0.7201,    Adjusted R-squared:  0.7185
F-statistic: 445.2 on 1 and 173 DF,  p-value: < 2.2e-16
```

- R squared is 72.01%.
- Slope is statistically significant (Global product sales go up 2.237 when EU changes 1 unit).





MLR: Can NA and EU Sales predict Global sales?

```
Residuals:
    Min      1Q  Median      3Q     Max
-3.4156 -1.0112 -0.3344  0.6516  6.6163

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)      1.04242    0.17736   5.877 2.11e-08 ***
NA_ProductSales  1.13040    0.03162  35.745  < 2e-16 ***
EU_ProductSales  1.19992    0.04672  25.682  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.49 on 172 degrees of freedom
Multiple R-squared:  0.9668,    Adjusted R-squared:  0.9664
F-statistic:  2504 on 2 and 172 DF,  p-value: < 2.2e-16
```

- Adj. R2 is 96.64%.
- Slope coefficients are statistically significant.
- Residual std. error is less than in SLR, residuals have smaller variance.
- The MLR gives a good fit.

Predicted global sales:

	NA_ProductSales	EU_ProductSales	fit	lwr	upr
1	34.02	23.80	1 68.056548	66.429787	69.683310
2	3.93	1.56	2 7.356754	7.099418	7.614090
3	2.73	0.65	3 4.908353	4.614521	5.202185
4	2.26	0.97	4 4.761039	4.478855	5.043223
5	22.08	0.52	5 26.625558	25.367353	27.883763

Actual global sales:

	Product	NA_ProductSales	EU_ProductSales	Global_ProductSales
	<fct>	<dbl>	<dbl>	<dbl>
1	107	34.0	23.8	67.8
2	326	22.1	0.52	23.2
3	6815	2.73	0.65	4.32

Overall, MLR model fits much better, than SLR we created before.

Conclusions and recommendations:

- The best explanation to loyalties (45%) can be given by spending score. Based on remuneration on its own model explains 38%. No evidence found that age can be used. Overall, MLR gives better result, but we found heteroscedasticity present.
- Based on remuneration and spending score defined that 5 groups would fit best. Going forward, marketing department can create 5 profiles to approach users based on different spending scores and remuneration and tailor make to best way of targeting these groups.
- Summaries and reviews give us a general insight about how the users feel about the products they bought and can highlight the areas of concern as well. In top 15, we can see mostly positive words and it's really encouraging. In reviews there are "cards" and "tiles" mentioned, most probably related to board and card games produced by Turtle Games, indicating a high response from the customers.
- Customers are expressing positive sentiment in their reviews and summaries mostly.
- Positive or neutral reviews were quite often misinterpreted by NLP as negative, underlying that it requires human check. NLP doesn't pick up sarcasm well or words about the game context (for aggressive or violent content) and interpret it as how the customers feel about the product.
- Reviews with complaints about the missing parts and games quality should be addressed and followed.
- North America has higher average sales compared to Europe £2.52m. European average sales are £1.64m
- There are number of bestsellers: maximum of £ 67.85m. in one game sales. It is game 107 for Wii platform.
- NA region games on Wii, X360, PS3 are more popular and in EU it is Wii, PS3 and X360.
- NES and GB are mostly popular in NA, but not in EU.
- PC more popular in Europe.
- Nintendo is the leader among publishers. Electronic Arts is number 2 in EU, as football is more popular in Europe than US. It shows that local market knowledge is very important.
- The data for sales is not normally distributed, right skewed and has thick tails
- Positive correlations are between all sales data. Global market influenced more by NA, than EU.
- EU and NA product sales data can be used to predict global sales with MLR model. Adding sales data from other markets might improve the predicted values.