

# **Turtle Games Customer Sentiments and Global Sales Analysis**

## ***Executive Summary***

This project is based on analysing data for Turtle Games regarding customer sentiments, product sales, customer loyalty points and their use in predicting future activities. The analysis will help Turtle Games to identify customer sentiments, relationship between various variables when it comes to acquiring loyalty points which, would help in directing marketing campaigns, product sales and how they impact global sales, and correlation between different regions to predict future figures.

## ***Problem and Goal***

Analysing data and exploring questions like: How various customer segments acquire loyalty points, is there any relationship between the variables? What is the general customer sentiment regarding Turtle Game's products, quality, service etc.? Can we group customers based on remuneration and spending score, to help marketing team target them with specific campaigns? Can we identify impact of various products on the global sales? Is there any correlation between different regions in terms of sales?

Goal is to answer all above questions and more, as need arises to conclude recommendations for Turtle Games.

## ***Approach***

Throughout this project various methods, plots, visualisations, and graphs are used to identify trends to predict future activities and establish correlation between various variables.

Linear Regression is used to identify any possible relationship between variables and if they can be used to predict Loyalty Points acquisition by customers. For this purpose, OLS model and statmodels function are used in Python.

K means clustering model is preferred to determine various segments of customers using remuneration and spending score for marketing team. For determining optimal number of clusters Elbow and Silhouette methods were used.

Natural Language Processing is used to establish customer sentiments by analysing their online feedback. Wordcloud is used to visualise most used words, Histogram for sentiment score and polarity, 15 most frequently used words are determined, and 20 top and bottom polarity score comments are also reviewed.

For sales data analysis, 'R' is preferred. Data is analysed and plotted on various graphs to identify underlying useful insights and establish relationship amongst variables. Top and bottom turnover games, by platform, are identified across regions and globally.

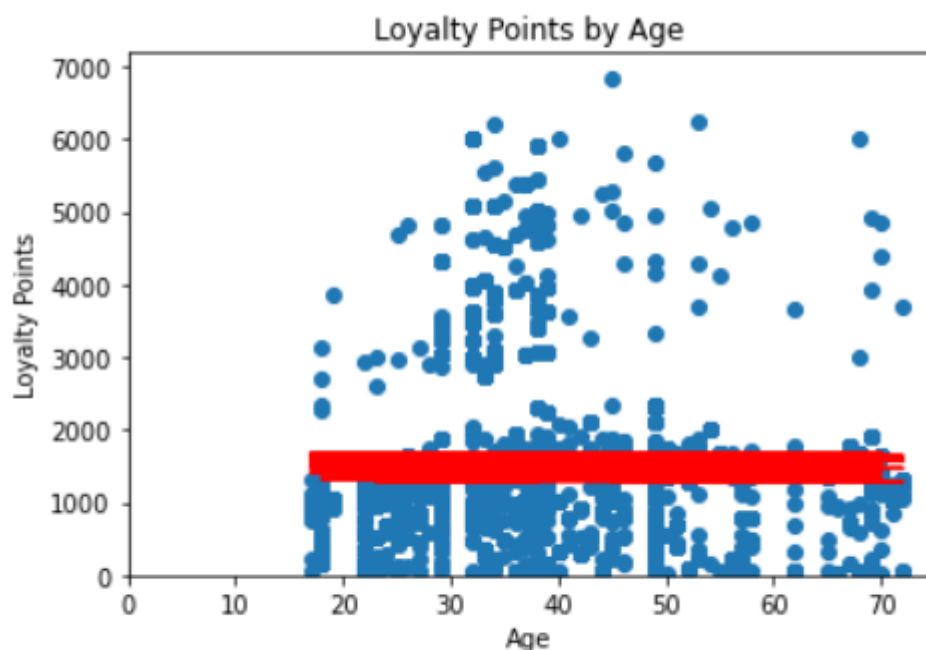
Data is explored, prepared, and analysed to check the normality based on plots, skewness, kurtosis, and Shapiro-Wilko test. Q-Q plots and ggplot2 functions are used to visualise and analyse insights (outliers were removed for better visualisation). Top and bottom revenue generating games were identified.

In the final segment, simple and multiple linear regression is carried out in R to establish correlation between regions, and if this can be used to predict future sales.

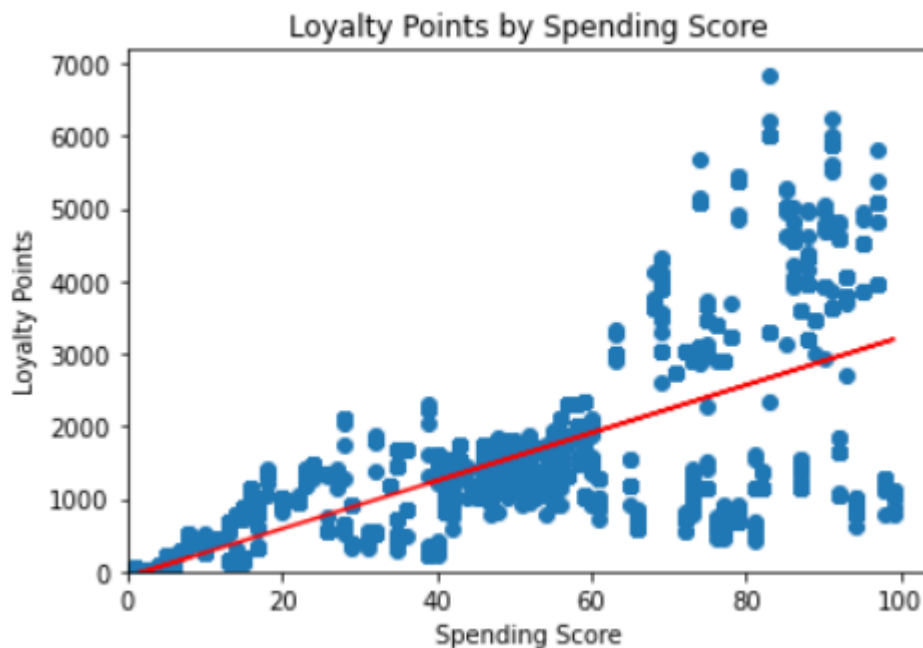
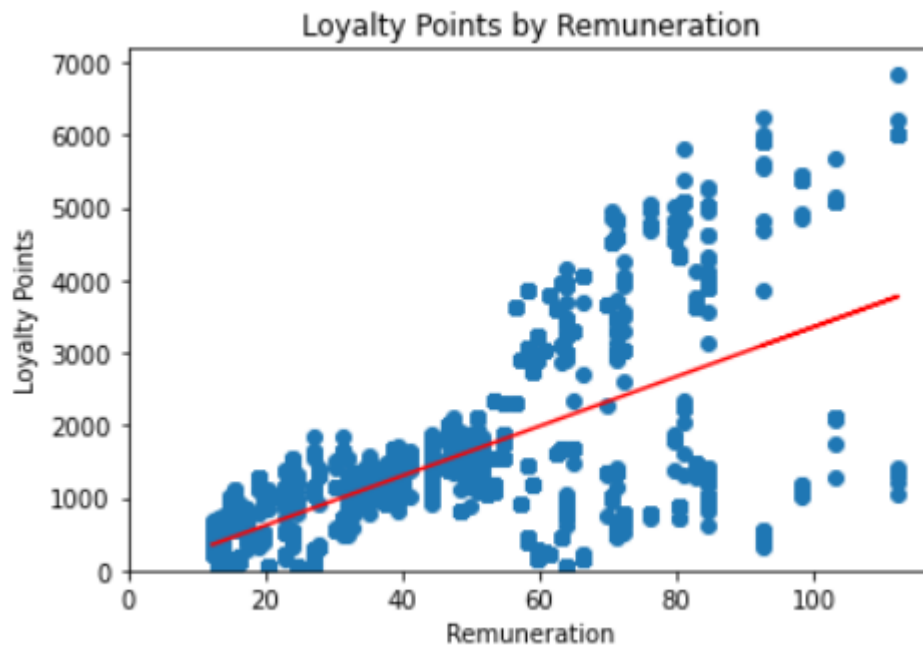
## Analysis

In the first segment relationship between Loyalty Points and age, remuneration and spending score are established, with a view to determine if these variables affect loyalty points acquisition. Observations are:

- Age as a variable is highly unlikely to affect variation in loyalty points as R-square (0.2%) is extremely low.



- Spending score and remuneration as variables are more likely than age to affect variation in loyalty points considering their R-square values- 45.2% and 38% respectively, and their P values are lower than 5%.

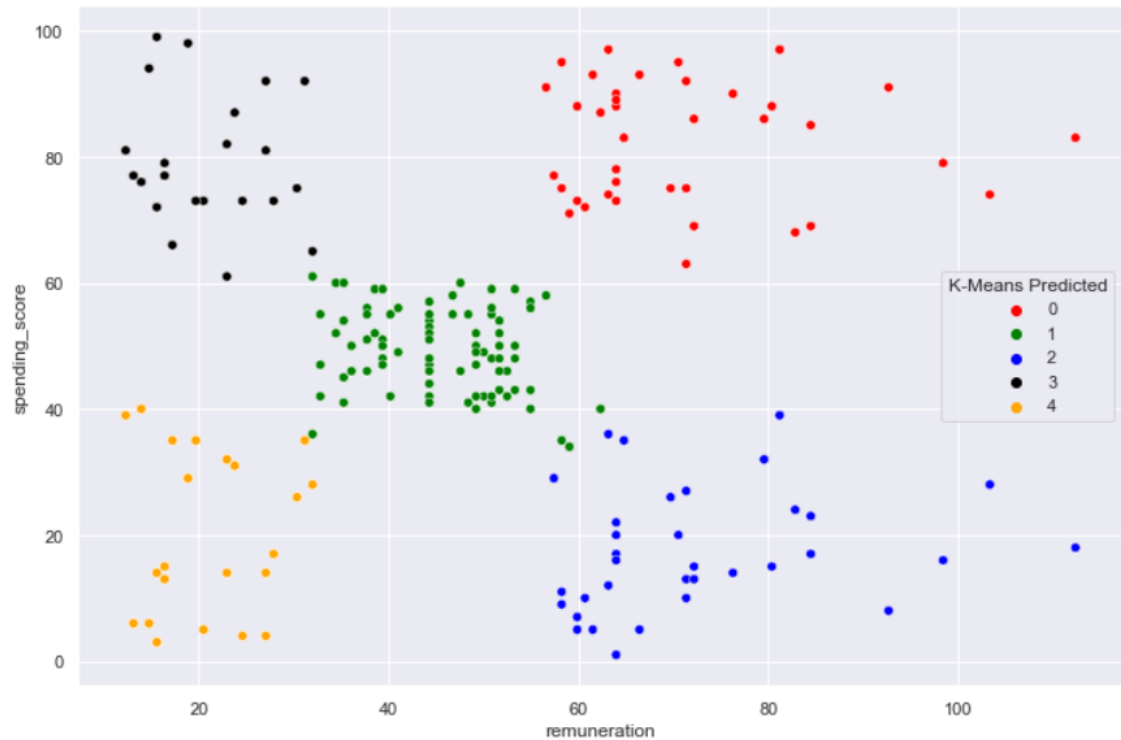


- Would consider exploring these two variables further and testing other variables to predict movement in loyalty points.

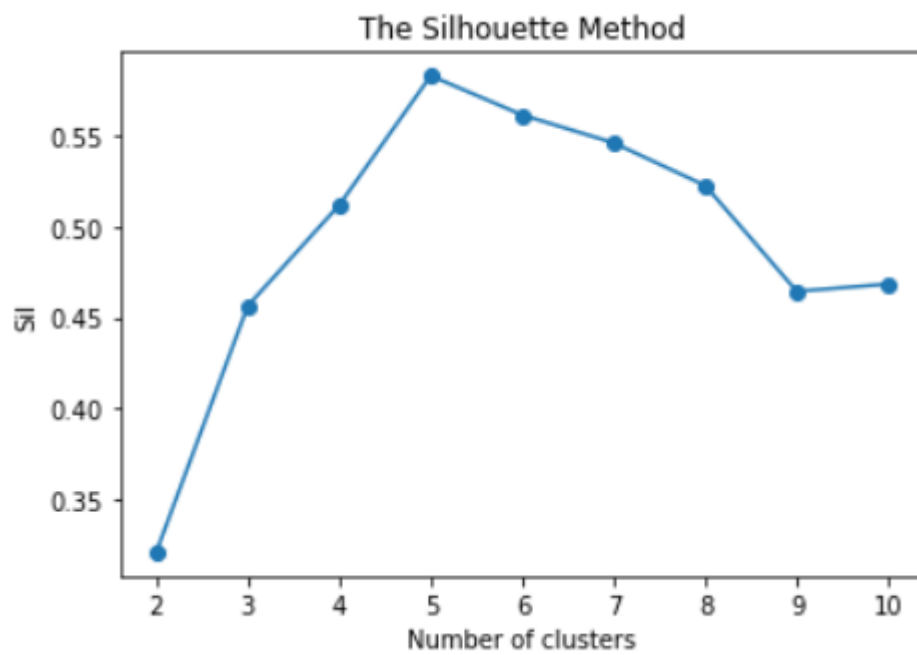
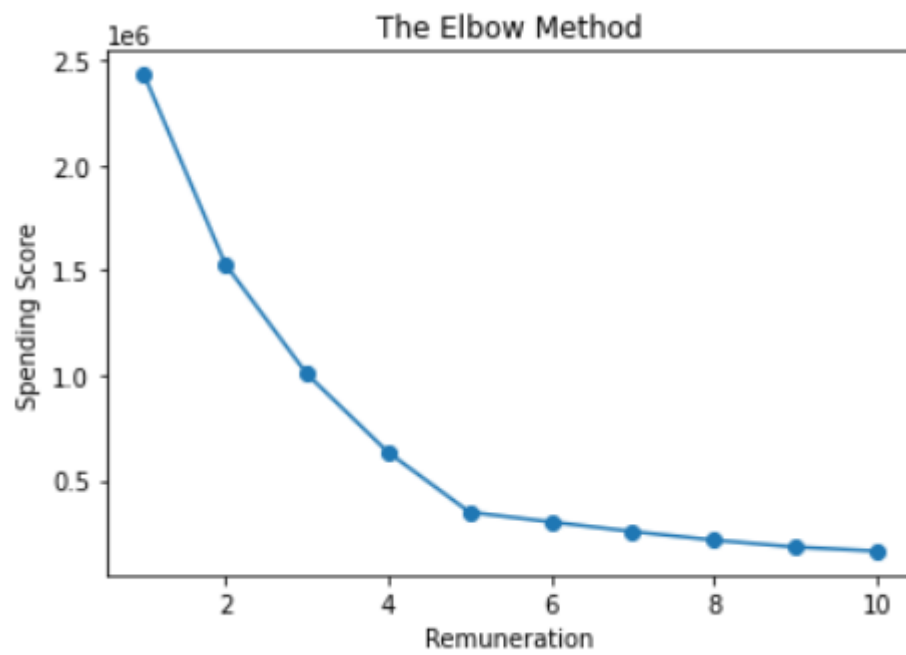
The plotted red line on graphs represent correlation and red line is not affected in the Age graph with the change in customer's age, whereas in the other graphs it does.

Two variables- remuneration and spending score is further analysed to identify groups within the customer base that can be used to target specific market segments. For this k means clustering model is incorporated and observations are:

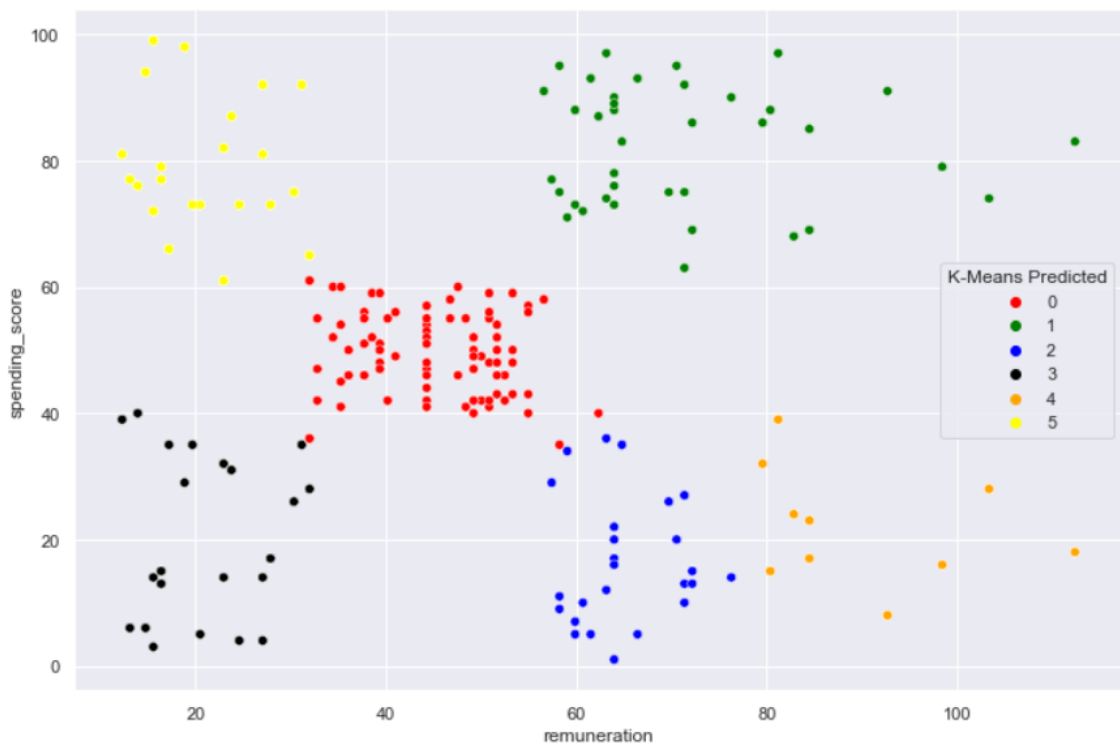
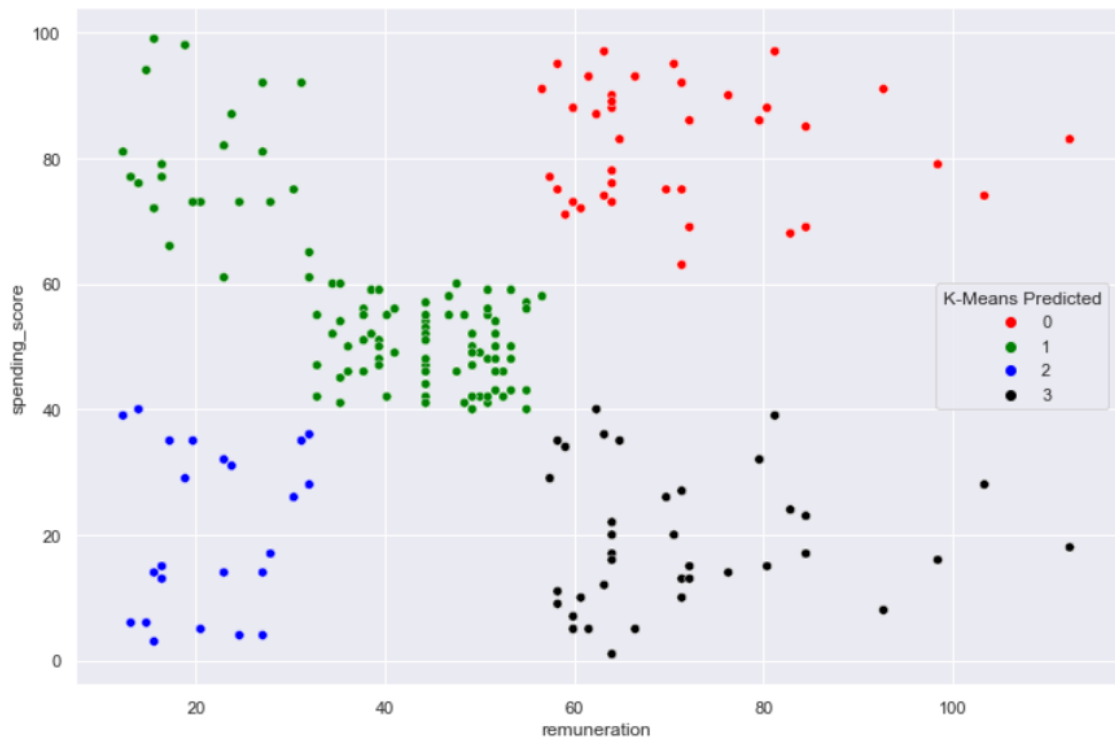
- k=5 is chosen as best fit model.



- The elbow and silhouette method depicting the same.



- k=4 and k=6 was also tested.



- In all models 0 is the largest cluster, but clusters seem to be more evenly distributed in the model where k=5 compared to k=4 and k=6. Or, the number of predicted values per class indicates a better distribution for k=5 than k=4 and k=6.

In process of analysis customer sentiment through NLP, following was observed:

- From wordcloud, nothing negative stands out.
- Summary



## 2. Review

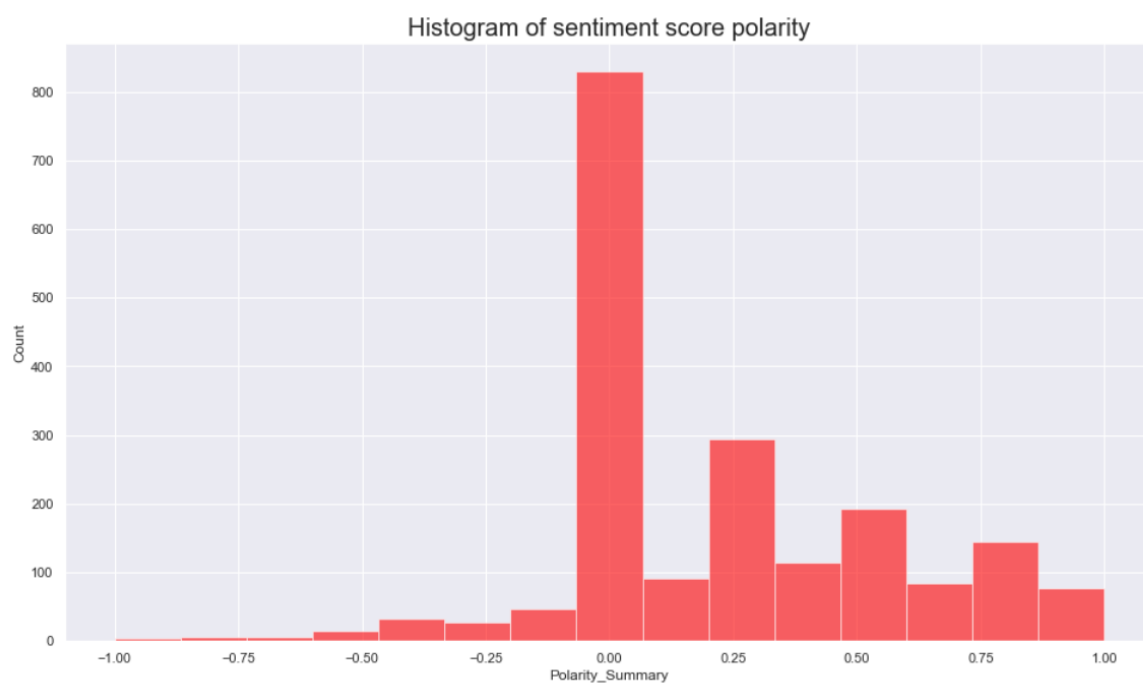
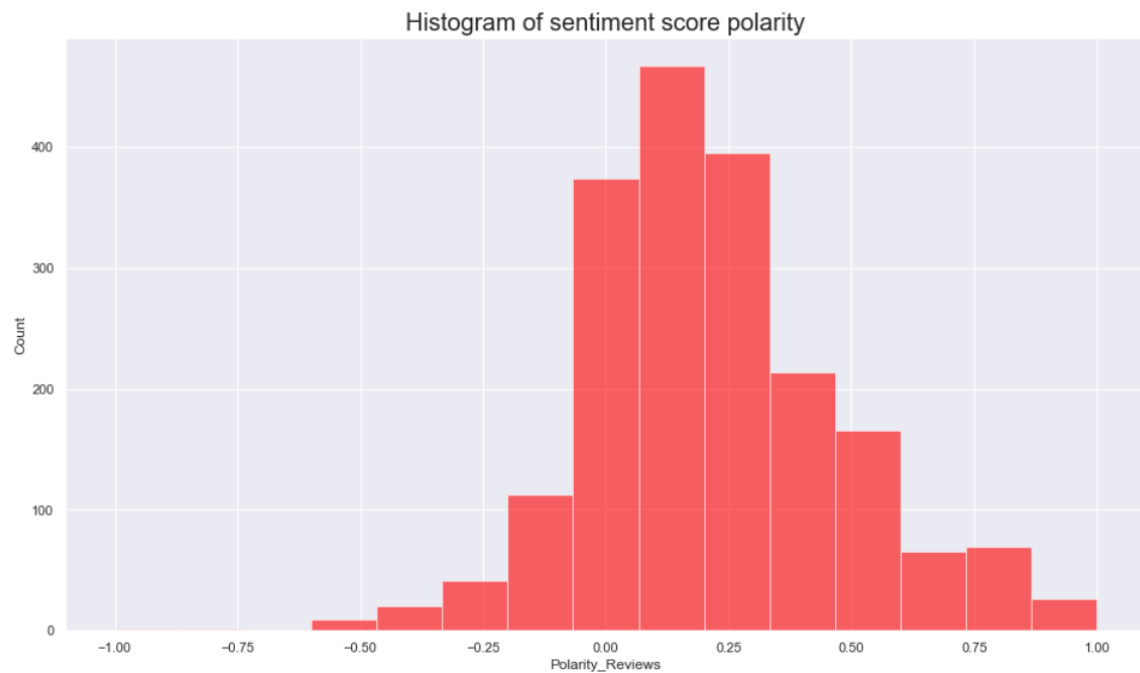


- Top 15 frequent words from both columns does not include any negative words.

| Frequency     |      | Frequency        |     |
|---------------|------|------------------|-----|
| Word          |      | Word             |     |
| <b>game</b>   | 1671 | <b>stars</b>     | 427 |
| <b>great</b>  | 580  | <b>five</b>      | 342 |
| <b>fun</b>    | 552  | <b>game</b>      | 319 |
| <b>one</b>    | 530  | <b>great</b>     | 295 |
| <b>play</b>   | 502  | <b>fun</b>       | 218 |
| <b>like</b>   | 414  | <b>love</b>      | 93  |
| <b>love</b>   | 323  | <b>good</b>      | 92  |
| <b>really</b> | 319  | <b>four</b>      | 58  |
| <b>get</b>    | 319  | <b>like</b>      | 54  |
| <b>cards</b>  | 301  | <b>expansion</b> | 52  |
| <b>tiles</b>  | 297  | <b>kids</b>      | 50  |
| <b>time</b>   | 291  | <b>cute</b>      | 45  |
| <b>good</b>   | 289  | <b>book</b>      | 43  |
| <b>would</b>  | 280  | <b>one</b>       | 38  |
| <b>book</b>   | 273  | <b>awesome</b>   | 36  |



- Histogram of polarity and sentiment score, generally show majority clustered around neutral sentiment but, can clearly see more positivity than negativity.



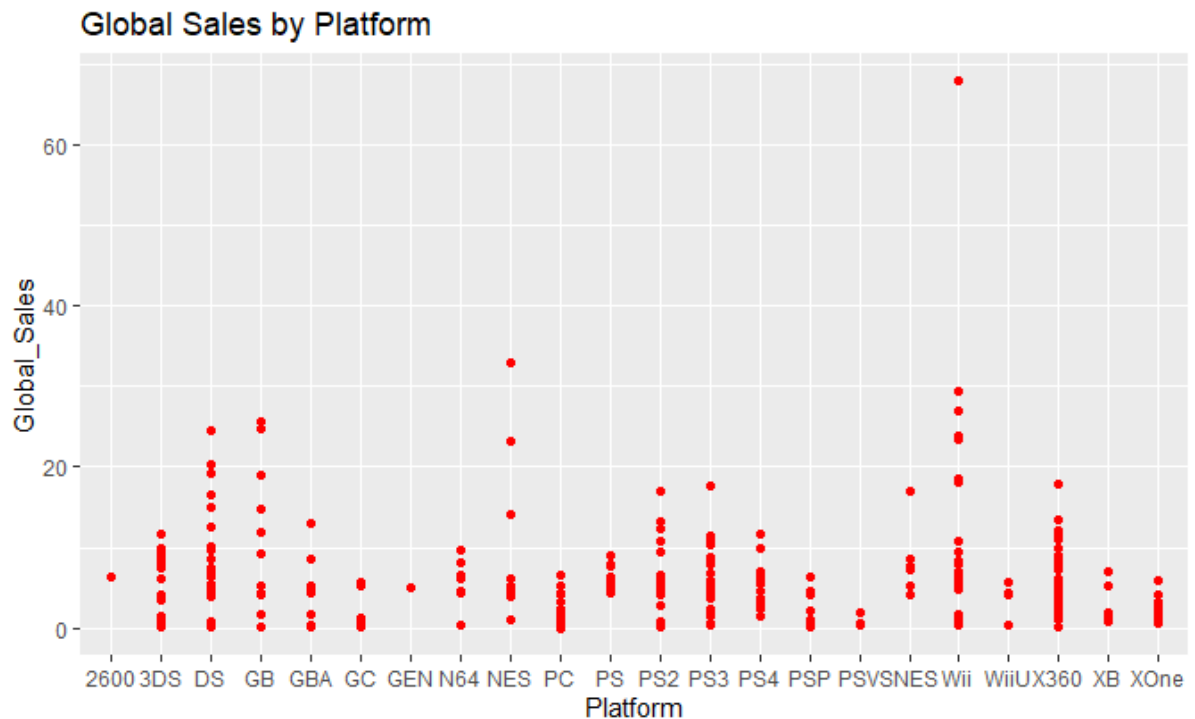
- Can clearly identify certain positive sentiments included in Top 20 negative comments and vice-versa.

|      |   | review  | polarity_review  |
|------|---|---|------------------|
| 207  | booo unless 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        |
| 181  |   | incomplete kit very disappointing   | -0.780000        |
| 1773 |   | im sorry i just find this product to be boring and to be frank juvenile                                   | -0.583333        |
| 362  | 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        |
| 116  |   | 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        |
| 226  |   | this was a gift for my daughter i found it difficult to use   | -0.500000        |
| 229  |   | i found the directions difficult  | -0.500000        |
| 289  |   | instructions are complicated to follow  | -0.500000        |
| 300  |   | difficult   | -0.500000        |
| 1501 |   | expensive for what you get  | -0.500000        |
| 173  | 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        |
| 345  | 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        |
| 531  | 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        |
| 305  |   | very hard complicated to make these   | -0.439583        |
| 421  |   | kids i work with like this game   | -0.400000        |
| 430  | 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        |
| 490  |   | my son loves playing this game it was recommended by a counselor at school that works with him            | -0.400000        |
| 795  |   | this game is a blast  | -0.400000        |
| 798  |   | i bought this for my son he loves this game   | -0.400000        |
| 814  |   | was a gift for my son he loves the game   | -0.400000        |
|      |   | summary   | polarity_summary |
| 6    |   | best gm screen ever   | 1.000000         |
| 28   |   | wonderful designs   | 1.000000         |
| 32   |   | perfect   | 1.000000         |
| 80   |   | theyre the perfect size to keep in the car or a diaper  | 1.000000         |
| 133  |   | perfect for preschooler   | 1.000000         |
| 139  |   | awesome sticker activity for the price  | 1.000000         |
| 160  |   | awesome book  | 1.000000         |
| 162  |   | he was very happy with his gift   | 1.000000         |
| 186  |   | awesome   | 1.000000         |
| 209  |   | awesome and welldesigned for 9 year olds  | 1.000000         |
| 412  |   | perfect   | 1.000000         |
| 468  |   | excellent   | 1.000000         |
| 536  |   | excellent   | 1.000000         |
| 541  |   | excellent therapy tool  | 1.000000         |
| 572  |   | the pigeon is the perfect addition to a school library  | 1.000000         |
| 591  |   | best easter teaching tool   | 1.000000         |
| 639  |   | wonderful   | 1.000000         |
| 643  |   | all f the mudpuppy toys are wonderful   | 1.000000         |
| 649  |   | awesome puzzle  | 1.000000         |
| 654  |   | not the best quality  | 1.000000         |

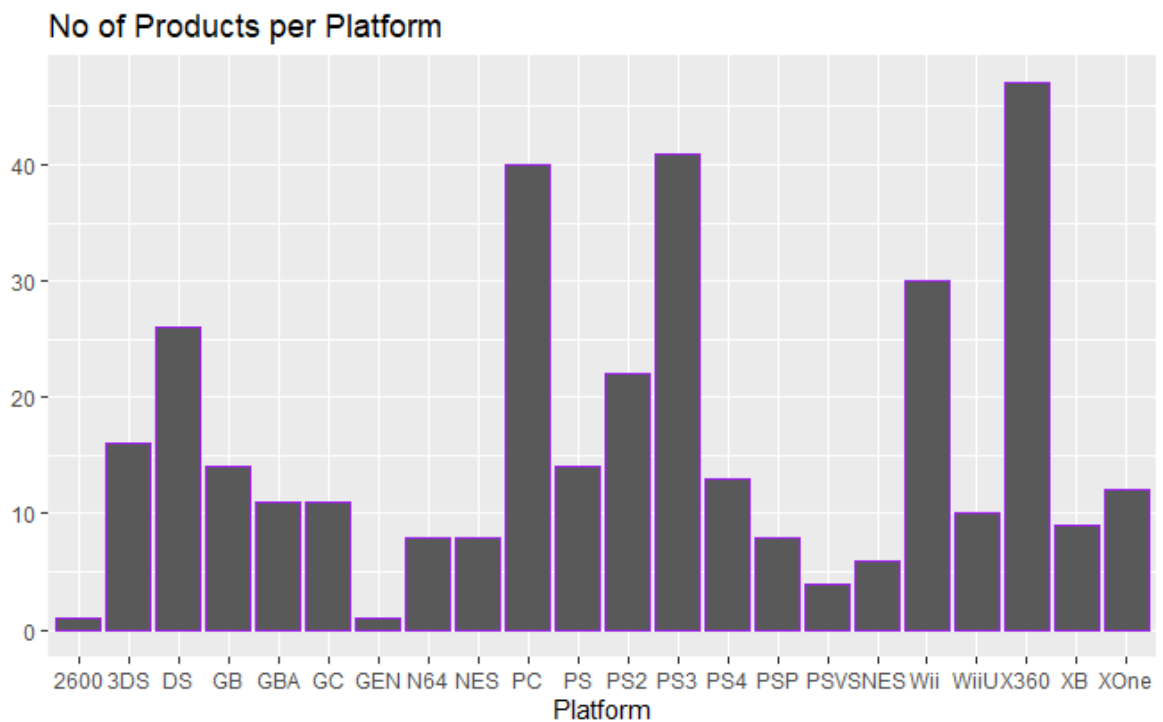
- The topmost common words used for online feedback (both columns combined) are game, great, fun, one, play, star, love, like and five.

Initial exploration of sales data highlighted the following:

- Globally, out of top 10 turnover products by platform, 5 belong to wii platform. Highest sales product belongs to wii platform. 7 out of lowest turnover 10 products, platform wise, belong to PC platform.



- Maximum number of games are sold on x360 platform.



- 17 products in NA registered no sale on at least one platform compared to only 3 in EU. Out of 17 in NA 10 belong to PC.

|    | Product | Platform | NA_Sales | EU_Sales |
|----|---------|----------|----------|----------|
| 1  | 5510    | PS2      | 1.54     | 0.00     |
| 2  | 7143    | 3DS      | 0.00     | 0.00     |
| 3  | 5453    | GB       | 0.00     | 0.00     |
| 4  | 1459    | PS2      | 2.47     | 0.01     |
| 5  | 5758    | GC       | 0.03     | 0.01     |
| 6  | 4702    | PC       | 0.00     | 0.01     |
| 7  | 1012    | PC       | 0.00     | 0.01     |
| 8  | 8933    | X360     | 1.65     | 0.02     |
| 9  | 7141    | XB       | 0.84     | 0.02     |
| 10 | 2253    | DS       | 0.79     | 0.02     |
| 11 | 9064    | PS2      | 0.64     | 0.02     |
| 12 | 3657    | DS       | 0.09     | 0.02     |
| 13 | 518     | X360     | 0.07     | 0.02     |
| 14 | 3887    | PS2      | 0.00     | 0.02     |
| 15 | 629     | PC       | 0.00     | 0.02     |
| 16 | 4399    | PC       | 0.01     | 0.02     |
| 17 | 8962    | PC       | 0.01     | 0.03     |
| 18 | 3436    | PC       | 0.00     | 0.03     |
| 19 | 948     | DS       | 0.44     | 0.04     |
| 20 | 2404    | PC       | 0.01     | 0.05     |
| -- | ----    | ---      | ---      | ---      |

|    | Product | Platform | NA_Sales |
|----|---------|----------|----------|
| 1  | 2253    | PC       | 0.00     |
| 2  | 518     | PC       | 0.00     |
| 3  | 7143    | 3DS      | 0.00     |
| 4  | 2518    | PSV      | 0.00     |
| 5  | 3645    | Wii      | 0.00     |
| 6  | 3967    | PC       | 0.00     |
| 7  | 3645    | 3DS      | 0.00     |
| 8  | 3645    | PSP      | 0.00     |
| 9  | 978     | PC       | 0.00     |
| 10 | 9080    | PC       | 0.00     |
| 11 | 3887    | PS2      | 0.00     |
| 12 | 3657    | PC       | 0.00     |
| 13 | 5453    | GB       | 0.00     |
| 14 | 3436    | PC       | 0.00     |
| 15 | 629     | PC       | 0.00     |
| 16 | 4702    | PC       | 0.00     |
| 17 | 1012    | PC       | 0.00     |
| 18 | 6424    | PC       | 0.01     |
| 19 | 999     | PC       | 0.01     |
| 20 | 1501    | PC       | 0.01     |
| 21 | 6287    | PC       | 0.01     |
| 22 | 3645    | PC       | 0.01     |
| 23 | 2404    | PC       | 0.01     |
| 24 | 8962    | PC       | 0.01     |
| 25 | 4399    | PC       | 0.01     |

- Scatterplot is most appropriate to compare Game sales as its easier to interpret and visualise in this case.

While checking normality of the data and statistical figures, following observations came to light:

- NES platform has highest mean sales per product, followed by wii. PSV has the lowest mean. Highest turnover platform is wii.

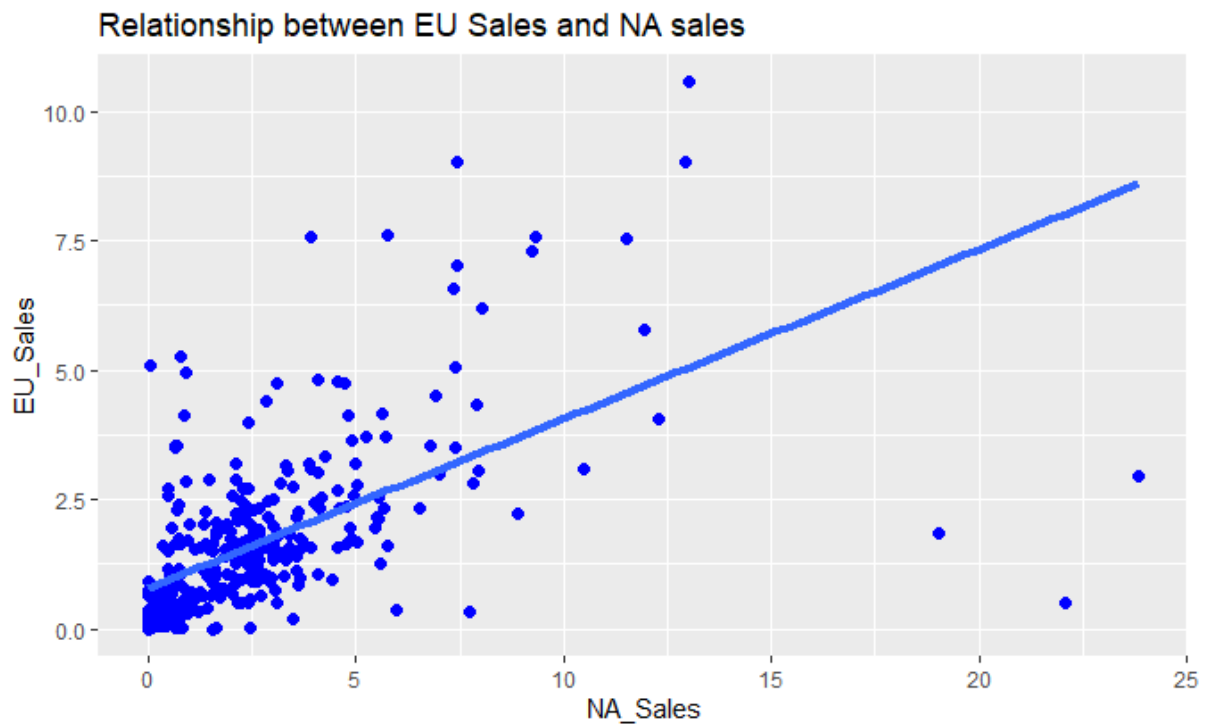
#### SUM

|    | Platform | Total_sales |
|----|----------|-------------|
| 1  | 2600     | 6.40        |
| 2  | 3DS      | 73.20       |
| 3  | DS       | 205.02      |
| 4  | GB       | 133.97      |
| 5  | GBA      | 47.10       |
| 6  | GC       | 21.66       |
| 7  | GEN      | 4.94        |
| 8  | N64      | 44.50       |
| 9  | NES      | 91.40       |
| 10 | PC       | 43.08       |
| 11 | PS       | 82.92       |
| 12 | PS2      | 131.87      |
| 13 | PS3      | 211.61      |
| 14 | PS4      | 70.54       |
| 15 | PSP      | 19.04       |
| 16 | PSV      | 3.34        |
| 17 | SNES     | 49.75       |
| 18 | wii      | 312.56      |
| 19 | wiiU     | 16.41       |
| 20 | X360     | 253.81      |
| 21 | XB       | 21.18       |
| 22 | XOne     | 33.51       |

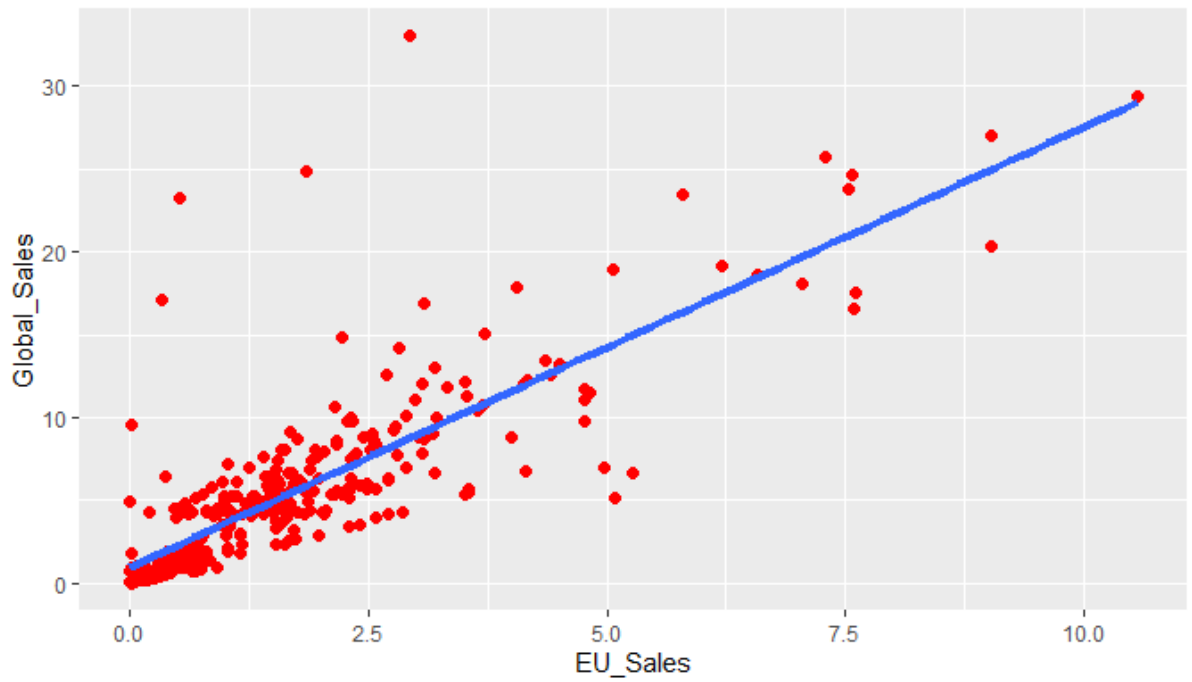
#### MEAN

|    | Platform | Total_sales |
|----|----------|-------------|
| 1  | 2600     | 6.400000    |
| 2  | 3DS      | 4.575000    |
| 3  | DS       | 7.885385    |
| 4  | GB       | 9.569286    |
| 5  | GBA      | 4.281818    |
| 6  | GC       | 1.969091    |
| 7  | GEN      | 4.940000    |
| 8  | N64      | 5.562500    |
| 9  | NES      | 11.425000   |
| 10 | PC       | 1.077000    |
| 11 | PS       | 5.922857    |
| 12 | PS2      | 5.994091    |
| 13 | PS3      | 5.161220    |
| 14 | PS4      | 5.426154    |
| 15 | PSP      | 2.380000    |
| 16 | PSV      | 0.835000    |
| 17 | SNES     | 8.291667    |
| 18 | wii      | 10.418667   |
| 19 | wiiU     | 1.641000    |
| 20 | X360     | 5.400213    |
| 21 | XB       | 2.353333    |
| 22 | XOne     | 2.792500    |

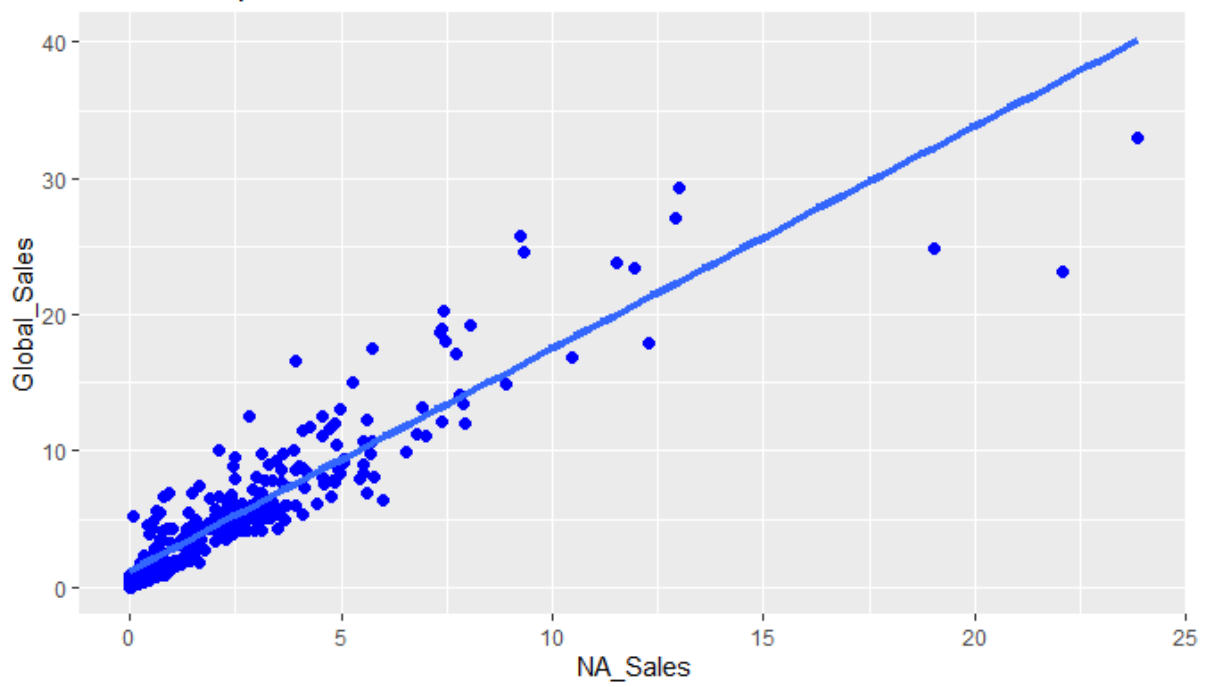
- Lowest turnover game registered 4.2 million sales while highest registered 67.8 million. Average sales were 10.7 million per product.
- NA: highest sales - 107: 34.02, lowest - 4491: 0.06 and for EU: Highest sales - 107: 23.80, Lowest - 5510: 0.0
- p value for all 3 data sets were lower than 0.05, so can conclude that data is not normally distributed.
- All 3 sales data sets are positively skewed.
- Kurtosis is higher than 3, indicating Leptokurtic distributions.
- Strong correlation exists between 3 sales figures, suggesting similar choice of games across the globe.



Relationship between Global Sales and EU sales

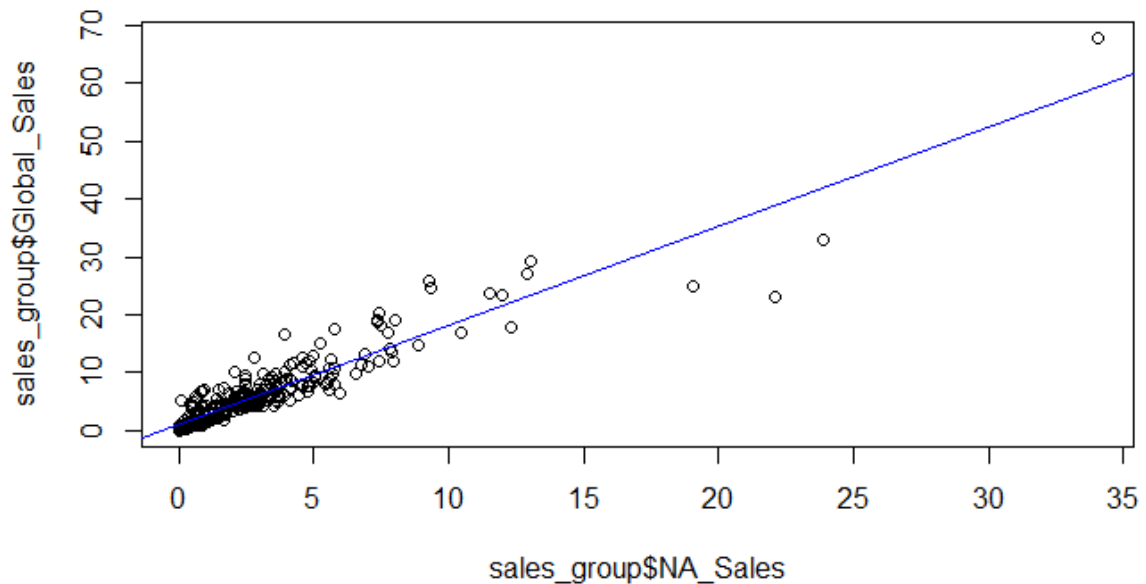


Relationship between Global Sales and NA sales

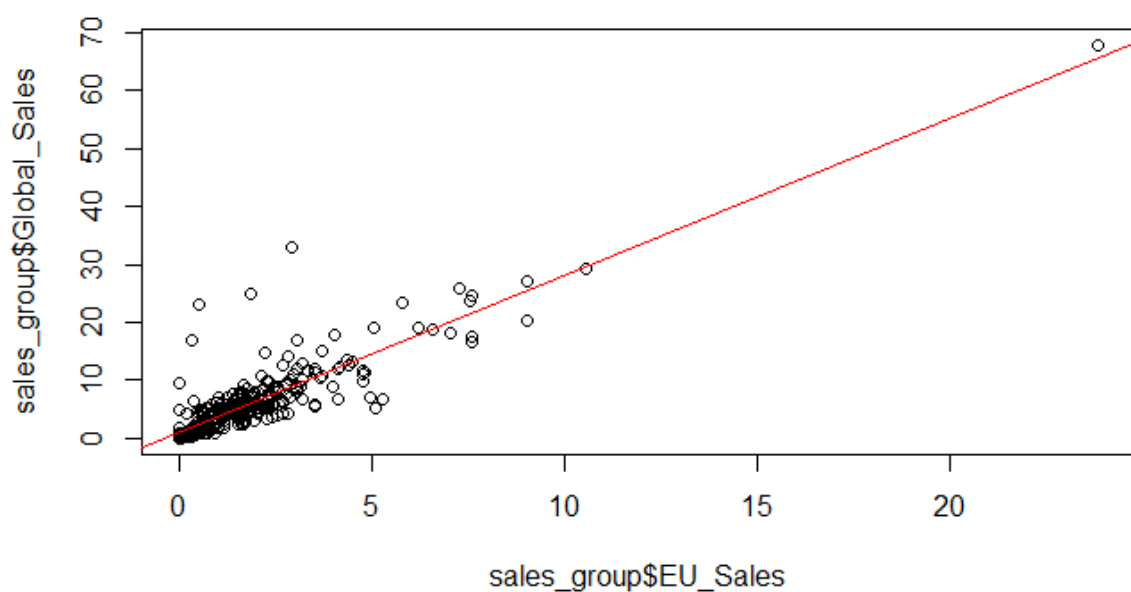


Moving towards the end of analysis, correlation between regions is determined with following results:

- For model1: NA Sales is a highly significant value, explaining over 87.41% of variability in Global Sales.

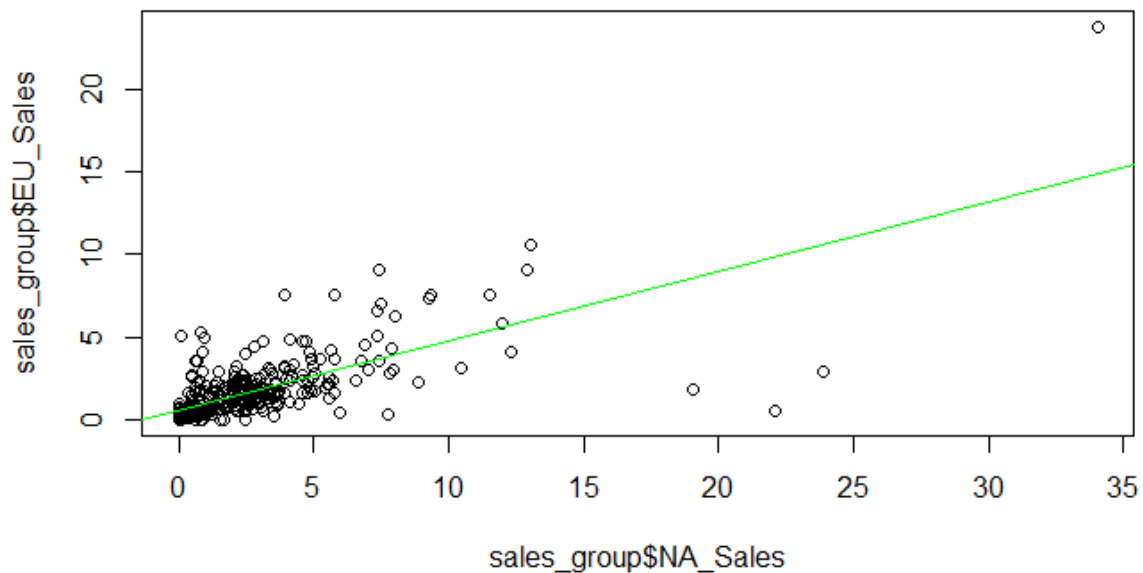


- For model2: EU Sales is a significant value, but not as much as NA Sales. EU Sales explains 77.01% of variability in Global Sales.

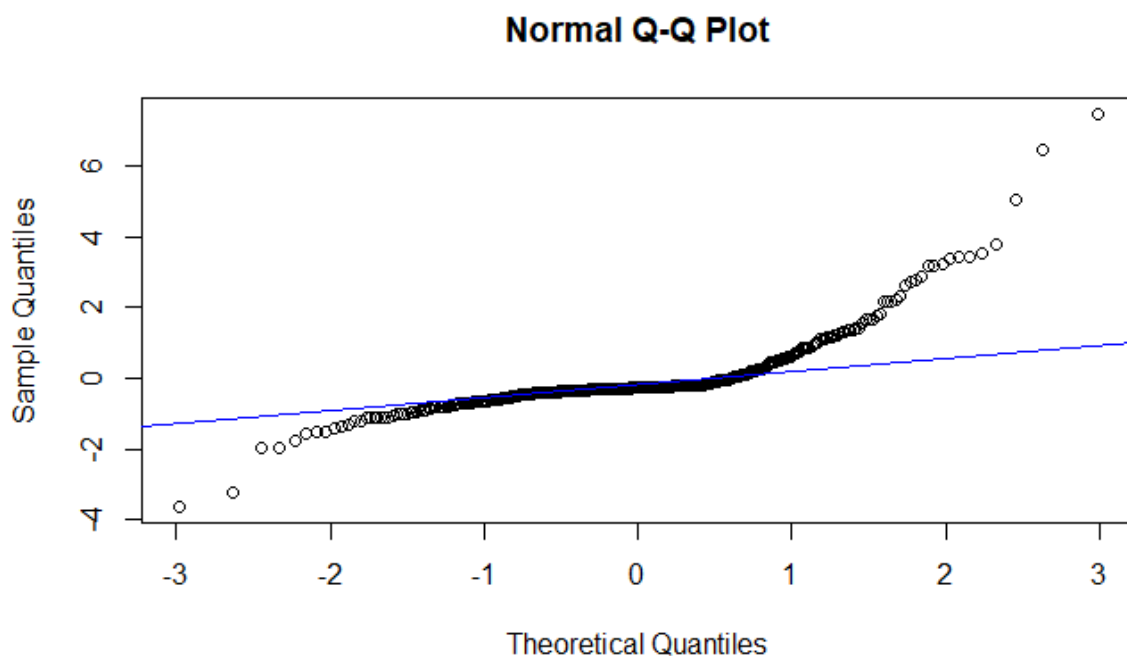




- For model3: NA Sales explains 49.78% of variability in EU Sales. NA Sales are not good predictors for EU Sales and vice-versa.



- Model4: Both NA Sales and EU Sales are significant and explain 96.85% variability in Global Sales.



- Predictions are quite close to the observed values using model4. Model can be used to predict global sales.

## ***Conclusion***

Based on analysis, it can be interpreted that 'age' doesn't affect loyalty points, whereas 'remuneration' and 'spending score' does to an extent but further variables and factors can be investigated for better loyalty points prediction.  $K=5$  is the most suitable model for segmenting customers for specific targeted marketing based on remuneration and spending score, but are they the most suitable variables?

Sentiment analysis shows more positive feedback than negative but need to check what time do they belong to. Are these recorded recently? And what span of time? Does the frequency change after a new product launch?

Sales data is dominated by NA and EU sales, so no surprises that these two regions would eventually affect the total global sales. NA even more than EU as it shares a bigger chunk of global sales.

What other regions does company sells its products in? Does it focus and sell in all continents? Is the strategy going to change in future? To what time range does the sales data belong to?

Products are ranked by total sales value, but it's hard to say their profitability unless the cost and selling price is known and need data for number of units sold as well. Some of the games were released long back, do they still sell? Most recently released games would probably be higher in price than older games, hence generating more turnover but hard to say about the volume. Sales data doesn't include other products like books and board games, which are mentioned in online customer feedback. Unless, the required data is made available, its hard to comment on future sales trend, pricing, and other variables.