Predicting future outcomes for Turtle Games with Python and R by Aleksandra Koerkemeier

The problem

Turtle Games has a business objective of improving overall sales performance by utilising customer trends. Decision-makers at Turtle Games want to understand:

1. How customers accumulate loyalty points.
2. How groups within the customer base can be used to target specific market segments.
3. How social data (e.g. customer reviews) can be used to inform marketing campaigns.
4. The impact that each product has on sales.
5. How reliable is the data (e.g. normal distribution, skewness, or kurtosis)?
6. What the relationships are between North American, European, and global sales?

To help Turtle Games reach its business objective, I had access to two files containing shoppers and sales data.

Customer persona

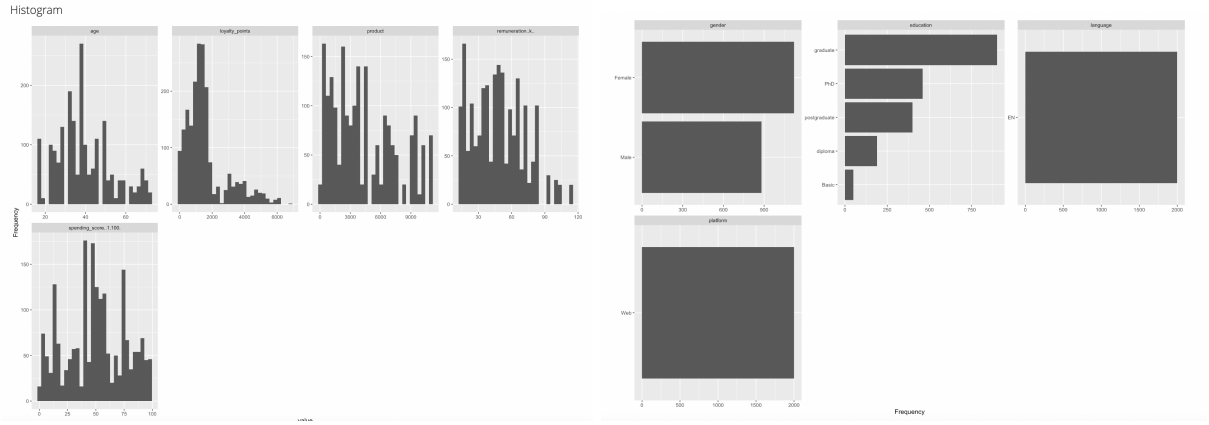
Let's have a closer look at a typical customer of Turtle Games. They are between 17 and 72 years old, with a mean age of 39.5. They earn anywhere from 12.300 to 112.340 pounds a year, with a mean income of 48.000 pounds a year.

In [4]:
```
# Descriptive statistics
reviews.describe()
```

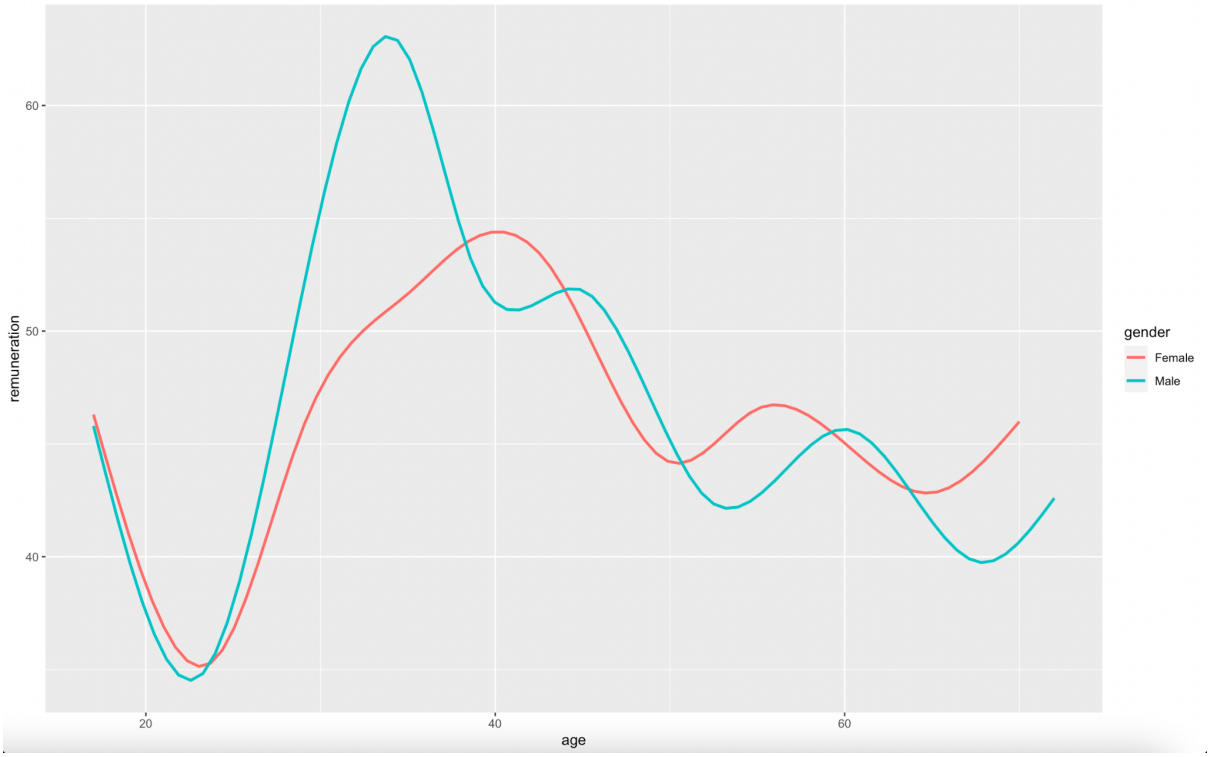
Out[4]:

	age	remuneration (k£)	spending_score (1-100)	loyalty_points	product
count	2000.000000	2000.000000	2000.000000	2000.000000	2000.000000
mean	39.495000	48.079060	50.000000	1578.032000	4320.521500
std	13.573212	23.123984	26.094702	1283.239705	3148.938839
min	17.000000	12.300000	1.000000	25.000000	107.000000
25%	29.000000	30.340000	32.000000	772.000000	1589.250000
50%	38.000000	47.150000	50.000000	1276.000000	3624.000000
75%	49.000000	63.960000	73.000000	1751.250000	6654.000000
max	72.000000	112.340000	99.000000	6847.000000	11086.000000

After initial analysis in Python, I decided to dig deeper into the reviews data set using R and its visualisations. Data Explorer confirmed my first insights into the shopper persona as well as provided some new data. It is very clear on one of the histograms that most of the users are right before their 40th birthday. We have also discovered that the majority of the gamers are female and the prevailing education level is graduate.

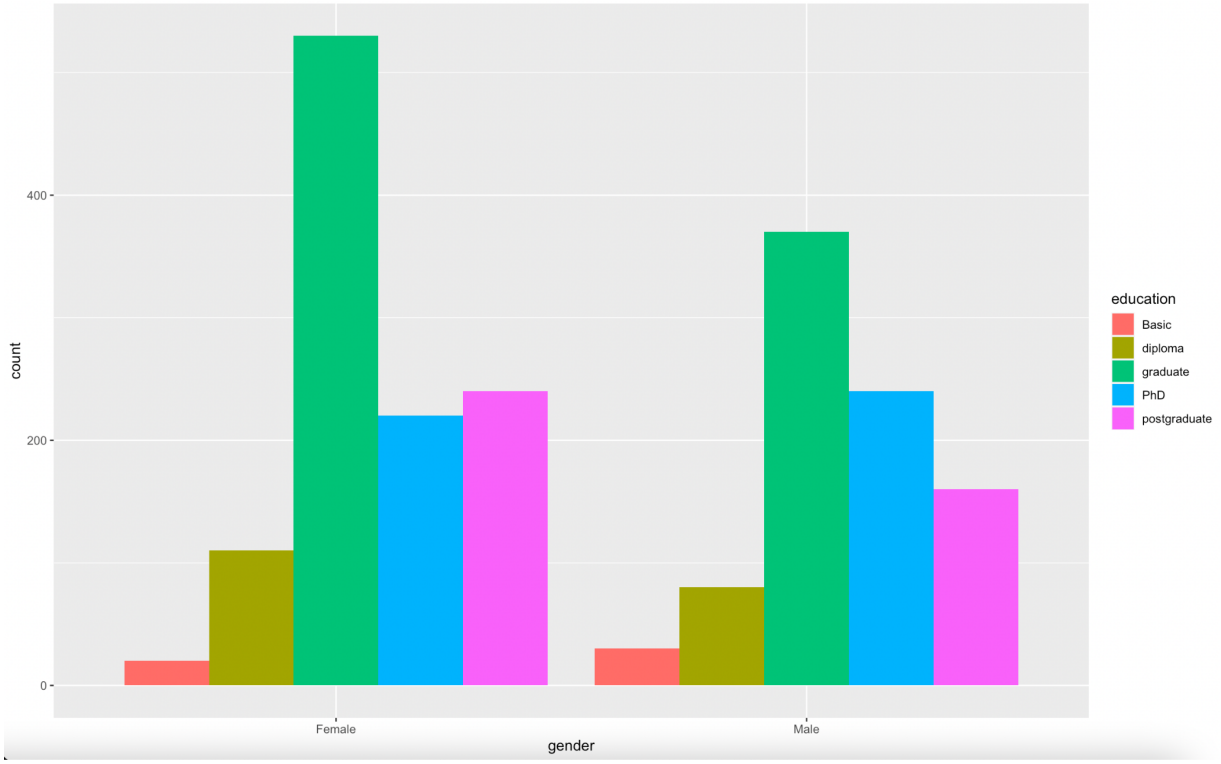


A look into the salary distribution yielded some interesting insights:



The graph, quite surprisingly, starts with a relatively high income at a very young age. It is not so uncommon though for young people with a high interest in gaming to start their online businesses and earn well. What is interesting for our core audience is that males start earning very well around the age of 30, while females seem to have their peek around the age of 40. I would therefore suggest men and women in their 30s and 40s should be a core target audience for Turtle Games.

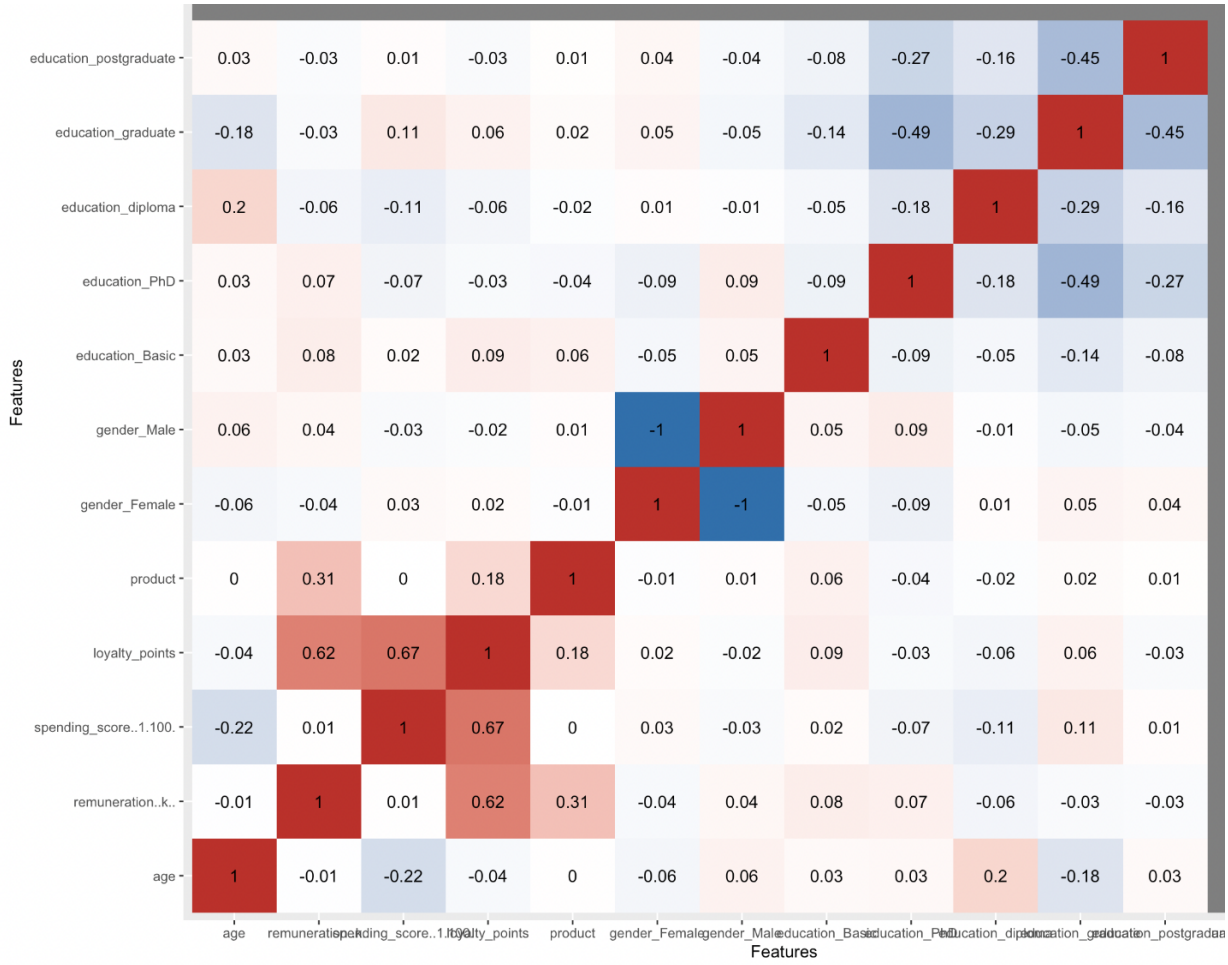
To add to that picture, we can also see that while there are more female buyers with graduate and postgraduate diplomas, more males hold a PhD, suggesting that Turtle Games' target audience is rather well-educated across the board.



Let's next take a look at how education and age both influence remuneration, spending score and loyalty points, separated by gender. Multiple scatterplots confirm our first discovery: both males and females around 30-40 years old, with a graduate degree, hold the most loyalty points and have the highest spending score as well as spending power (high income).

Lucrative correlations

Looking into the Data Explorer file brings attention to one other important aspect - there seems to be a quite strong correlation between loyalty points and spending score (0.67) and loyalty points and remuneration (0.62). Let's revert back to Python to build a linear regression model and see if we can replicate these results.



According to the model built in Python spending score and loyalty points have an r squared of 0.45 which means that 45% of the variation of the data can be explained by the simple regression model. We have also discovered that if the spending score is 1 unit higher there will be an increase in loyalty points by 33 units.

As for loyalty points and remuneration, we get an r squared of 0.38, which means that 38% of the variation of the data is explained by our model and that if the remuneration is 1 unit higher (so 1 thousand pounds higher) then there will be 34.2 increase in loyalty points. Age and loyalty points don't show much correlation.

Out of curiosity, I have taken a look at users with the highest spending score and the biggest amount of collected loyalty points. There indeed doesn't seem to be a close correlation but what is interesting is that holders of the biggest amount of loyalty points are all among the highest earners.

gender	age	remuneration (k£)	spending_score (1-100)	loyalty_points	education	language	platform	product	review
Male	32	112.34	83	6020	PhD	EN	Web	3711	This is a fantastic product! I highly recomm
Male	32	112.34	83	6020	PhD	EN	Web	453	Great therapy gap for kids! Some question
Male	32	112.34	83	6020	PhD	EN	Web	263	Wow! This product is incredible! I was hes
Male	32	112.34	83	6020	PhD	EN	Web	1012	I didn't like how the pieces are made of pa
Male	32	112.34	83	6020	PhD	EN	Web	1497	Played this with my girlfriend. Changed it
Male	32	112.34	83	6020	PhD	EN	Web	1175	just right
Male	32	112.34	83	6020	PhD	EN	Web	618	This is only an expansion, so it wasn't surp
Male	40	92.66	91	6005	PhD	EN	Web	8962	I primarily bought this game to generate H The way this game can generate Hazards My crappy eyes aside, I recommend this g
Female	68	112.34	83	6020	postgraduate	EN	Web	2795	I haven't had much time to play with it - w
Male	34	112.34	83	6208	graduate	EN	Web	979	Our family really enjoyed spending hours c
Female	45	112.34	83	6847	diploma	EN	Web	3478	I love this game. It is so much fun. You will
Female	53	92.66	91	6232	postgraduate	EN	Web	977	Love playing Quiddler and this dictionary h

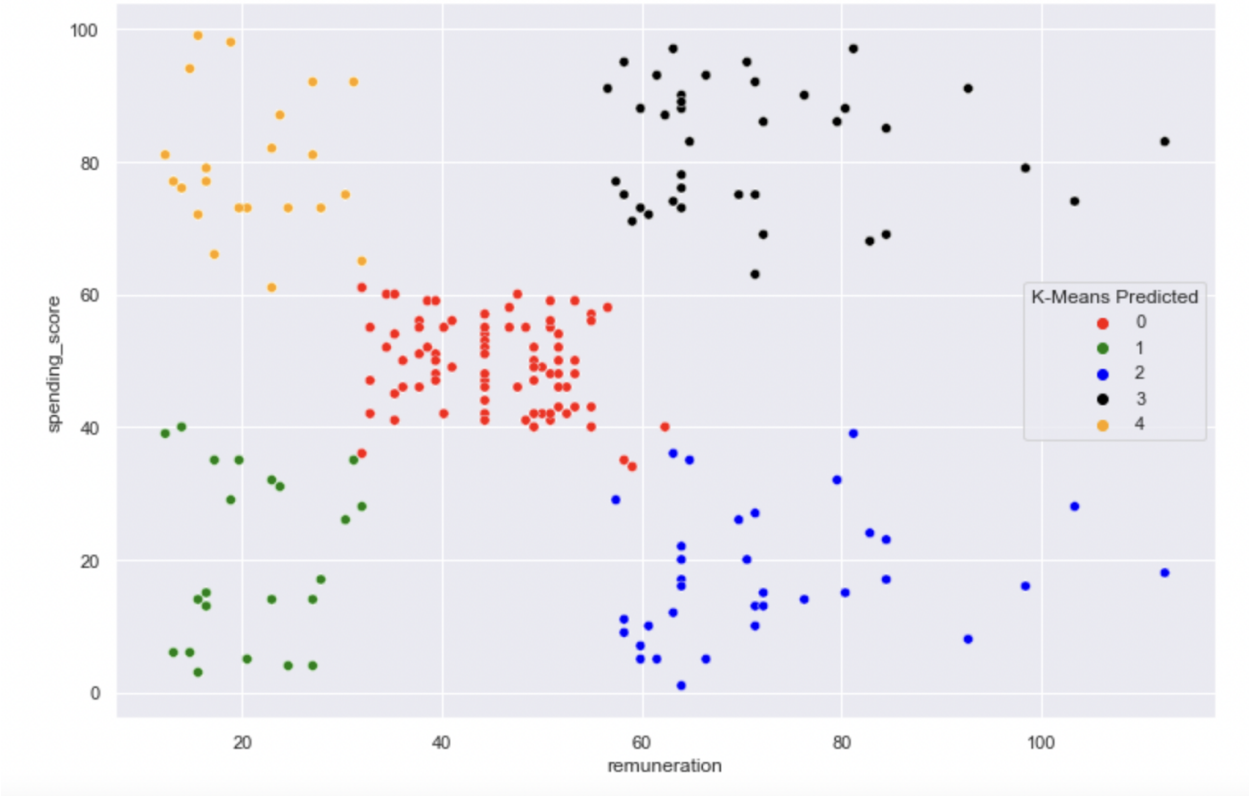
gender	age	remuneration (k£)	spending_score (1-100)	loyalty_points	education	language	platform	product	review
Female	37	15.58	99	1067	postgraduate	EN	Web	1473	Can't wait to use it!
Female	37	15.58	99	1067	postgraduate	EN	Web	1497	Very cute and very well designed.
Female	37	15.58	99	1067	postgraduate	EN	Web	1497	I really like this game, it helps kids
Female	37	15.58	99	1067	postgraduate	EN	Web	123	We are giving a set to each of ou
Female	37	15.58	99	1067	postgraduate	EN	Web	1183	great game
Female	37	15.58	99	1067	postgraduate	EN	Web	577	Great game! Long but that is to b
Female	37	15.58	99	1067	postgraduate	EN	Web	2162	I like the minis and the game is go
Female	33	15.58	99	1012	graduate	EN	Web	2162	Amazing expansion to Lords of W
Female	22	15.58	99	785	graduate	EN	Web	9635	My children have the Alphabet tra
Female	38	15.58	99	1078	graduate	EN	Web	6271	The kids who come to my studio l
Female	57	15.58	99	1122	PhD	EN	Web	811	We've been playing this game for
Male	26	15.58	99	881	PhD	EN	Web	9529	Love this game.
Female	44	15.58	99	1128	PhD	EN	Web	2285	Fun and entertaining.
Male	34	15.58	99	1027	graduate	EN	Web	4047	My wife and I love this game, whi The rules are easy to pick up, and

Market segmentation

Next, I have taken a look at possible market segmentation. Using k-means clustering, following elbow and silhouette methods, it became apparent that Turtle Games could cluster their users into five groups.

In terms of optimising marketing spending cluster number three - of users with high yearly income and high spending score - seems to be the most interesting. They could be targeted with specific email marketing campaigns, for example around premiers, premium releases and limited editions in an effort to upsell a loyal customer base.

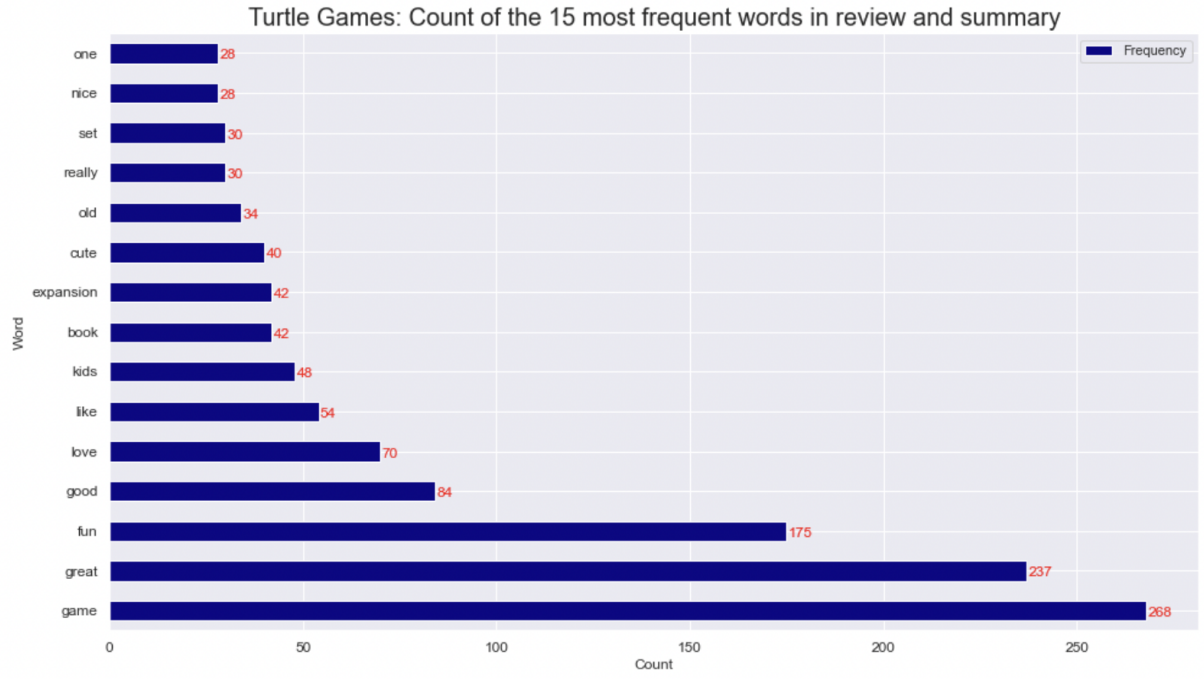
Cluster number two, on the other hand - aka high earners with low spending score - could be targeted with rather different email marketing campaigns highlighting sales and promotions to potentially activate a new segment of spenders.



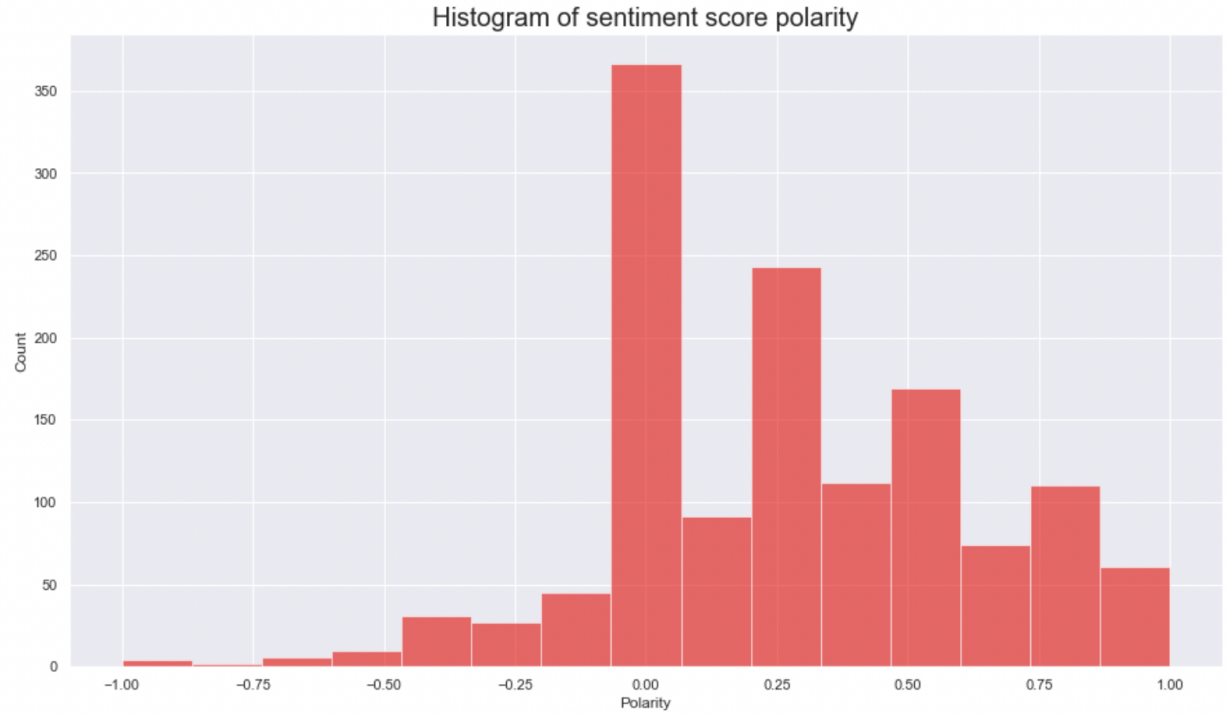
Sentiment toward the brand

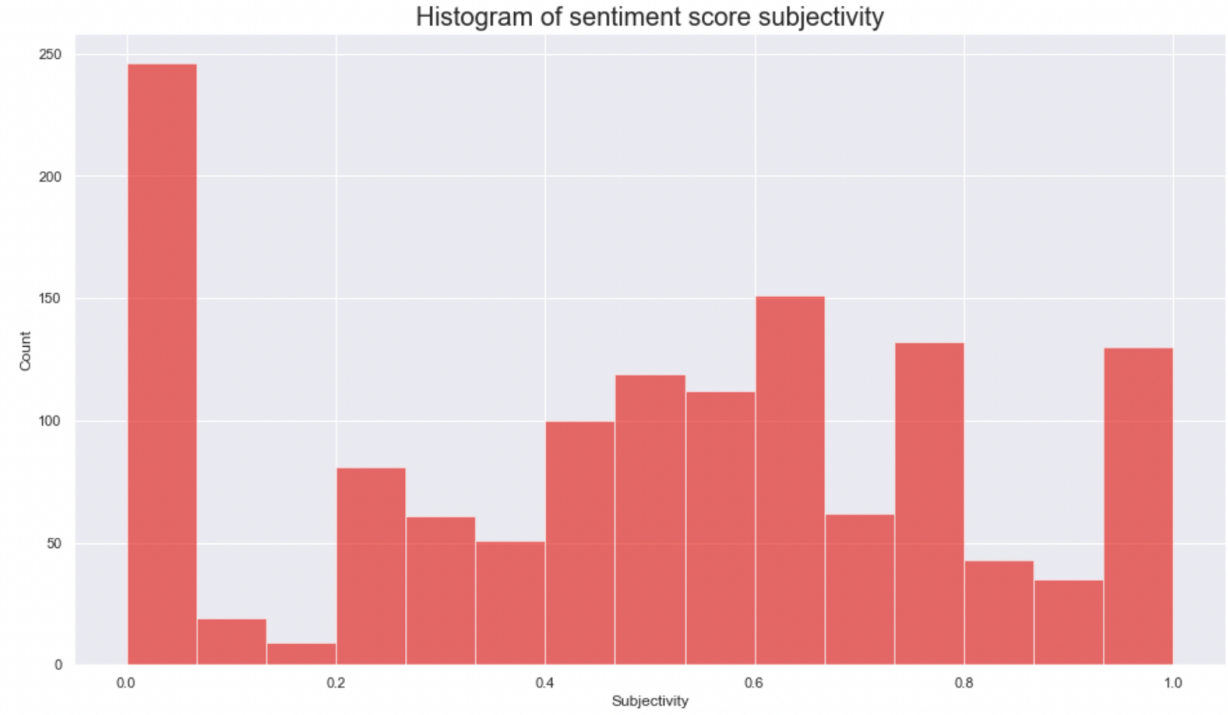
As part of the further analysis, I looked into reviews (and summaries of those reviews) of products bought by users, utilising a very powerful Natural Langauge Processing library in Python to see if Turtle Games have anything to worry about or address. The results were great, though. A word cloud and the count of 15 most used words in the reviews show that

customers are generally really happy with their purchases, with words like great, fun, good, love and like amongst the most commonly used.



Sentiment analysis also showed that there are many neutral reviews, but a significant part of what was written falls on the positive side and that the reviews are rather objective.



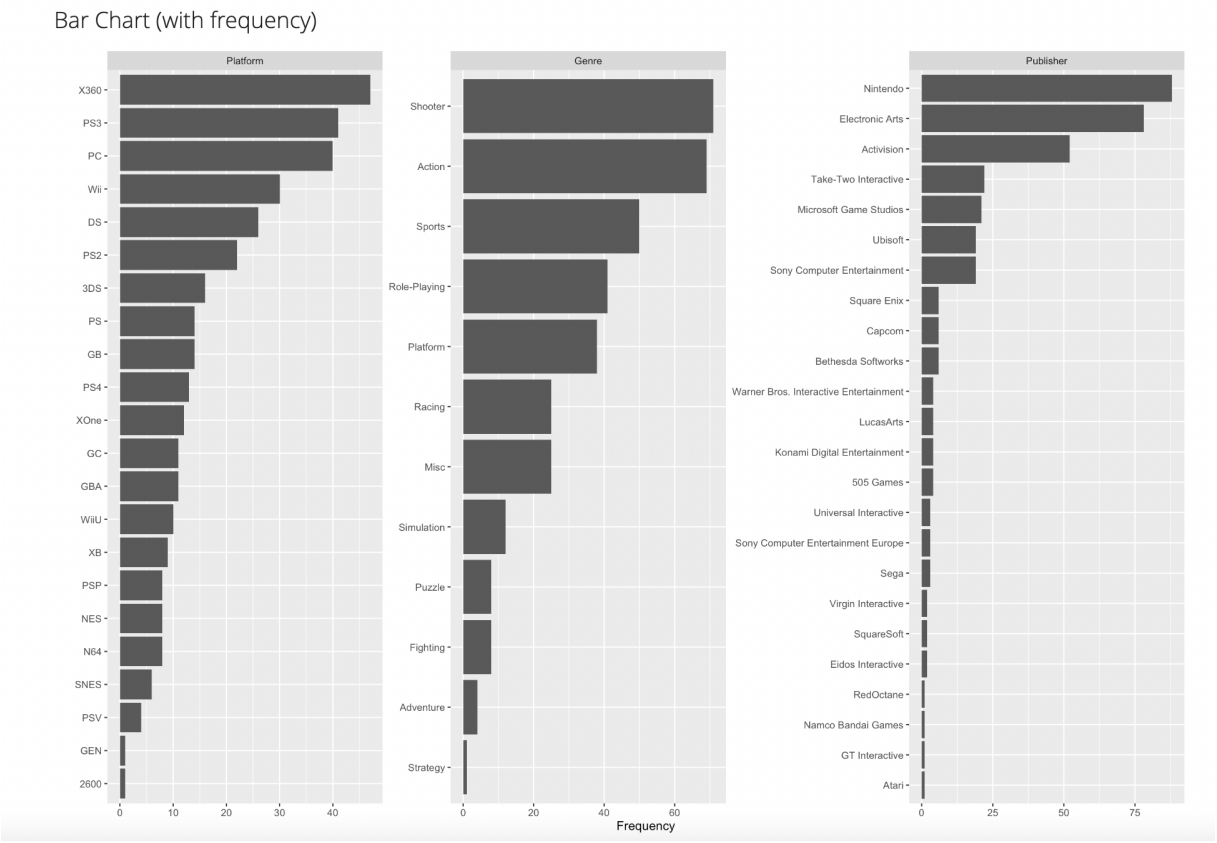


Sentiment analysis is a powerful tool for marketing departments: authors of most positive reviews could be targetted with campaigns suggesting further purchases, while buyers who were not satisfied with a purchase could be addressed by a customer service department in a customised outreach aiming to keep them engaged with the store and not churn.

Furthermore, Turtle Games could utilise social listening (for example analysing specific hashtags on Twitter) to track trends in the gaming industry to promote the right products at the time they are being widely discussed on social media.

Product sales

A first look into the sales data shows some interesting insights into the platform, game publisher and genre of games that are being chosen the most. I have decided to follow that trail, before looking into which exact product generates the most sales.



Looking into Global Sales confirmed that Nintendo is by far the best-selling publisher (with 911 million pounds in global sales) and shooter games (with 321 million pounds in global sales) are the most sold genre. In terms of a platform. In terms of a platform, while Xbox, PS3 and PC sell the most items, in terms of global sales Wii is leading (with 311 million pounds earned).

	Platform	NA_Sales_sum	EU_Sales_sum	Global_Sales_sum
	<chr>	<dbl>	<dbl>	<dbl>
1	Wii	150.	105.	313.
2	X360	153.	76.0	254.
3	PS3	77.8	88.5	212.
4	DS	72.6	65.6	205.
5	GB	68.7	28.2	134.
6	PS2	58.7	31.5	132.
7	NES	66.0	9.14	91.4
8	PS	34.0	25.6	82.9
9	3DS	26.4	21.6	73.2
10	PS4	23.1	34.1	70.5

Next, let's have a look at sales per product. Three most sold products by Turtle Games are products number 107 (which is a sports game released by Nintendo for Wii), 515 (which is an action game for PS3 released by Take-Two Interactive) and finally game number 123 (which is a Nintendo game sold for DS marked as miscellaneous).

	Product	NA_Sales_sum	EU_Sales_sum	Global_Sales_sum
	<int>	<dbl>	<dbl>	<dbl>
1	107	34.0	23.8	67.8
2	515	19.2	18.9	45.9
3	123	26.6	4.01	37.2
4	254	21.5	2.42	29.4
5	195	13	10.6	29.4
6	231	12.9	9.03	27.1
7	249	9.24	7.29	25.7
8	948	14.4	7.79	25.4
9	876	12.8	9.25	25.3
10	263	9.33	7.57	24.6

Data reliability

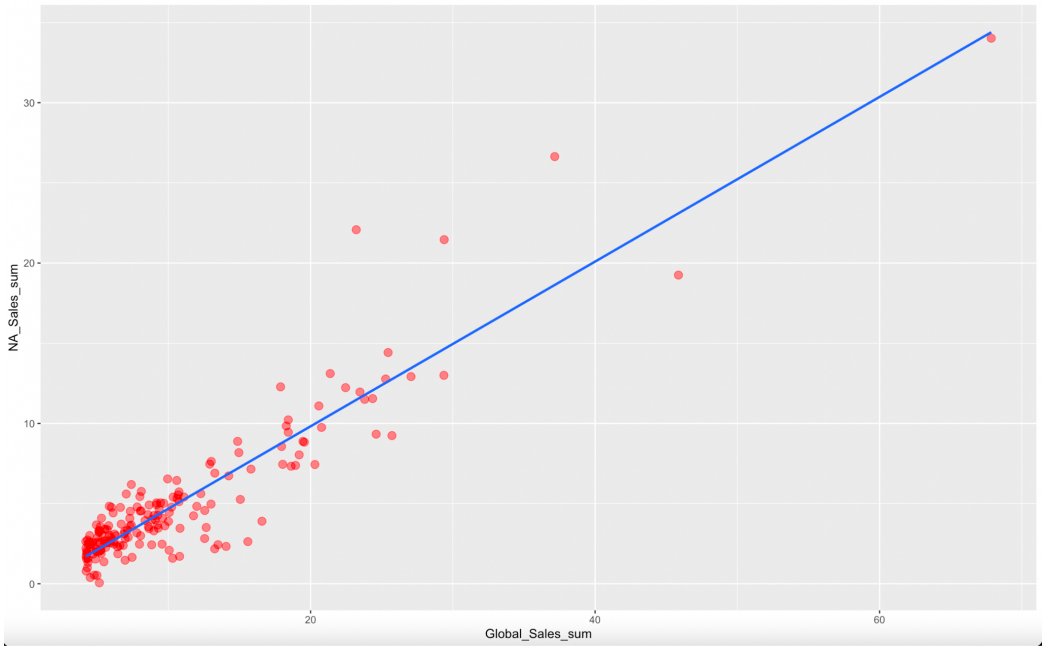
Looking into data distribution, I have performed the Shapiro-Wilk normality test and in all cases - North American sales, European sales and global sales - it resulted in a p-value above 2. The null hypothesis of Shapiro's test is that the population is distributed normally and with a p-value above 0.05 we can conclude that the data, in this case, is normally distributed.

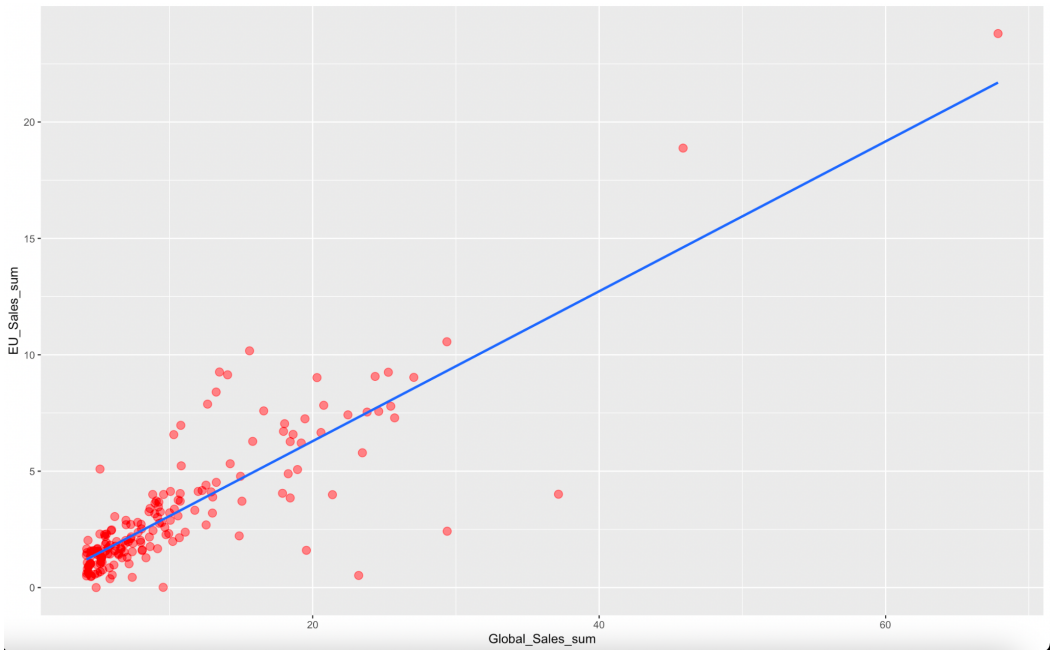
```
        Shapiro-Wilk normality test

data:  sales_df1$NA_Sales_sum
W = 0.69813, p-value < 2.2e-16

> shapiro.test(sales_df1$EU_Sales_sum)

        Shapiro-Wilk normality test

data:  sales_df1$EU_Sales_sum
W = 0.74058, p-value = 2.987e-16

> shapiro.test(sales_df1$Global_Sales_sum)

        Shapiro-Wilk normality test

data:  sales_df1$Global_Sales_sum
W = 0.70955, p-value < 2.2e-16
```

Let's look into skewness and kurtosis next. Positive excess values of kurtosis (>3) indicate that the distribution is peaked and possesses thick tails, which would indicate leptokurtic (or heavy-tailed) distribution.

```
> skewness(sales_df1$NA_Sales_sum)
[1] 3.048198
> kurtosis(sales_df1$NA_Sales_sum)
[1] 15.6026
> skewness(sales_df1$EU_Sales_sum)
[1] 2.886029
> kurtosis(sales_df1$EU_Sales_sum)
[1] 16.22554
> skewness(sales_df1$Global_Sales_sum)
[1] 3.066769
> kurtosis(sales_df1$Global_Sales_sum)
[1] 17.79072
```

In terms of correlation - there is a very strong correlation of 91% between North American and global sales and a strong correlation of 84% between European sales and global sales. That could inform Turtle Games to invest more into marketing in the US for a better ROI.





And to finalise the data analysis process I have created a multiple linear regression model with North American, European and global sales per product which has an excellent R squared of 0.97. Using that model Turtle Games can predict sales of products - the model predicted a global sales price for product 107 (used here as an example) at 66.35 which is a very close fit with an observed value of 67.85.

Final recommendations

- According to Statista, there will be 3.3 billion gamers in the world by 2024.
- According to this report, the average age of a current gamer is 35. Research into Turtle Games' buying persona confirms that 30-40+ is a perfect target audience in terms of buying power and accumulated spending score and loyalty points.
- For Turtle Games, female audience is bigger than the male audience and that mirrors a global trend where the gender gap is a myth and the split is closer to 50-50. My recommendation would be to conduct further research into channels where female gamers can be best reached with advertisements to monetise on this trend.
- Loyalty points are a good metric for Turtle Games to target their users as holders of the majority of points appear to also be the highest earners.
- Market segmentation: two very interesting segments for Turtle Games are high earners with a high spending score and high earners with a low spending score. Both could be reached by customised email marketing efforts.
- Turtle Games should utilise sentiment analysis to identify highly satisfied users and offer them loyalty programmes as well as address negative feedback to avoid churn. On top of that Turtle Games could regularly analyse a set of keywords to understand online trends and feed their marketing campaigns with relevant content.
- It is not surprising that Nintendo is by far the most-selling publisher for Turtle Games as they are the third best-selling company in the world. It would make sense to look into two of the biggest gaming companies worldwide - Sony (which products are on offer currently in Turtle Games store) and Tencent Games (currently not offered)
- Shooter games are best sold worldwide (according to Statista)
- We have also identified the best-selling products, Turtle Games can use that knowledge for example in the upcoming Cyber Month and Christmas campaigns.
- And finally, we have seen that there is a very strong correlation between North American and global sales. Indeed, for Turtle Games, North America is the best-earning market, which could inform focusing marketing spending there for a better ROI. One market bigger than the US is China (possible expansion?) and in terms of Europe where Turtle Games operate Germany is in the lead in terms of revenue (a German-centric campaign could be an option to raise European sales).