

## **Introduction**

Turtle Games are a manufacturer and retailer of video games with a worldwide clientele. The company's product offerings include books, board games, video games, and toys. Due to its extensive product line, Turtle Games would like to increase overall sales.

In particular, data and data-driven systems are utilised to explain customer trends or predict future outcomes. For instance, customer reviews or sales data can identify popular products and social media platforms, target specific market segments, and much more. Given this, Turtle Games would better understand its customer base and make more suitable products and marketing to gain more sales.

## **Analytical approach**

Before conducting any analysis, it would be wise to assess the data to ensure its validity and quality. Examples include correcting mistakes, making sure the data is structured appropriately, looking for any missing values, and much more.

I used the commands, `.head()`, `.info()`, `.columns()` and `.dtypes()` to gain a broad overview of the DataFrame. Furthermore, to identify any missing values I used the `isnull().sum()`. I found no missing values for all DataFrames and began to look at some descriptive statistics using the `.describe()` command. I then got used the `summary()` and `is.na()` commands to get an overview of the dataset and to see any missing values. Again, I find there were no missing values and went on to look at the `mean()`, `minimum()` and `maximum()` of certain columns, such as `Global_Sales`, `EU_Sales` and `NA_Sales`.

Furthermore, when analysing any customer trends, it is essential to remember that some columns may need to be more informative. For instance, the `language` column might be removed because all reviews are written in English. Alternatively, giving appropriate names to columns.

I start analysing the data with some regression analysis. Using either linear or multiple linear regression to evaluate any possible linear relationships between certain variables. I run three regressions which are between loyalty points and age, loyalty points and spending, loyalty points and remuneration. To find the line of best fit, data analysts attempt to minimise residuals, which aim to reduce the separation between the data points and regression line. Spending and loyalty points had the closest correlation of the three regressions. Additionally, the regression's  $R^2 = 0.452$  is the greatest of the three regressions. The better the model, the closer the value is to 1 (or 100%). Therefore, to an extent we

could use spending to predict loyalty points and the regression results are displayed below in the figures.

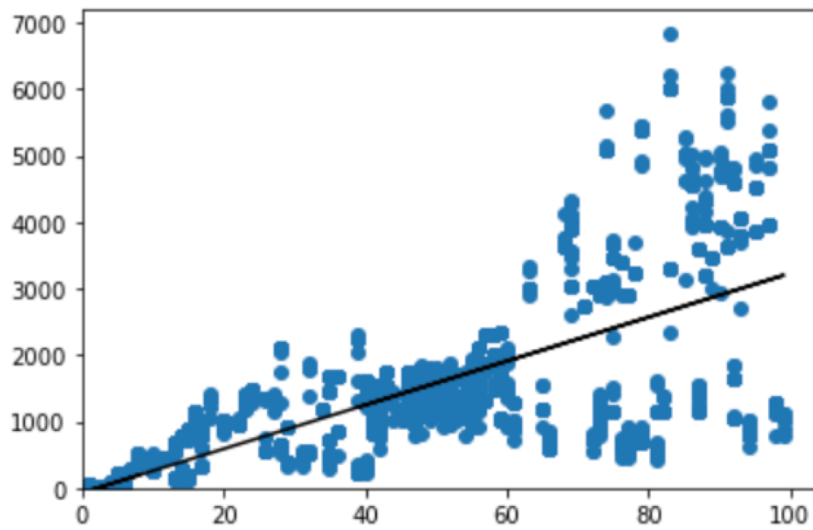


Figure 1: Regression graph for spending and loyalty points

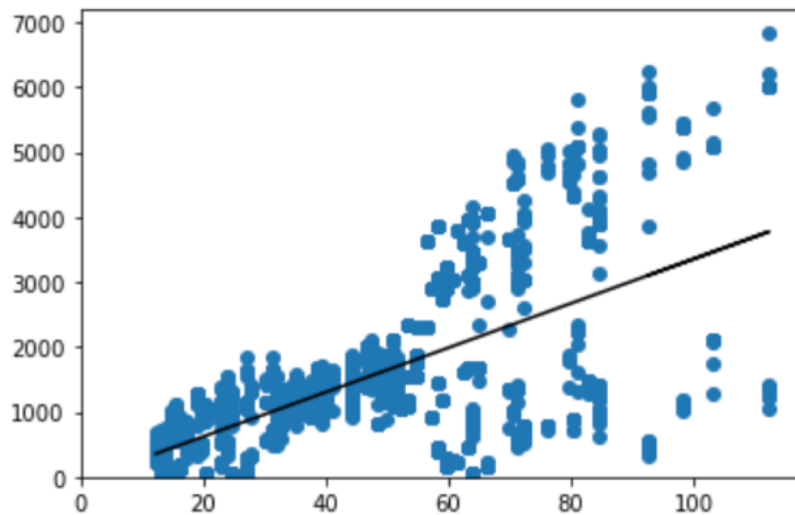


Figure 2: Regression graph for remuneration and loyalty points

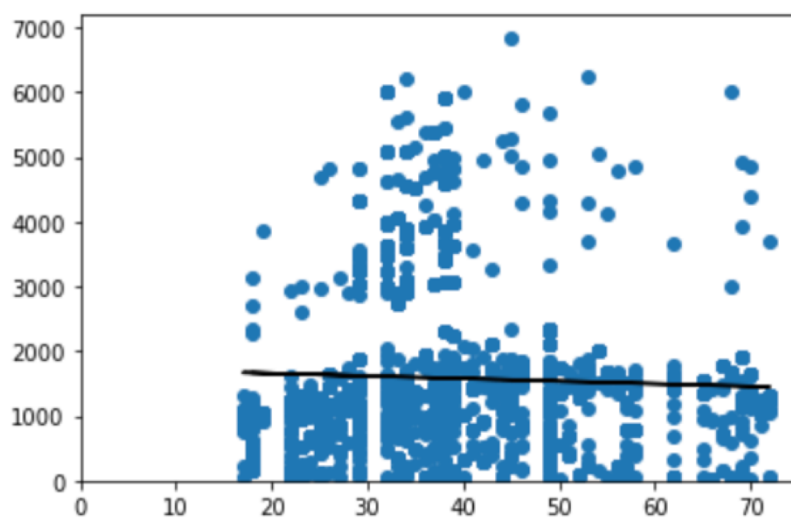


Figure 3: Regression graph for age and loyalty points

Using the regression approach as before I ran simple and multiple linear regressions to identify potential connections between global, North American, and European sales. I ran the same regression on the original dataframe and the aggregated dataframe. Below, figures 1 and 2 display the regression output. Columns 1,3 in both figures are for the original data frame and columns 2 and 4 are for the aggregated data frame.

Dependent variable:				
	Global_Sales (1)	Global_ProductSales (2)	Global_Sales (3)	Global_ProductSales (4)
NA_Sales	1.718*** (0.035)			
NA_ProductSales		1.635*** (0.054)		
EU_Sales			2.714*** (0.079)	
EU_ProductSales				2.237*** (0.106)
Constant	1.012*** (0.148)	2.458*** (0.370)	0.873*** (0.207)	3.334*** (0.479)
Observations	352	175	352	175
R2	0.874	0.839	0.770	0.720
Adjusted R2	0.874	0.839	0.769	0.719
Residual Std. Error	2.226 (df = 350)	3.266 (df = 173)	3.008 (df = 350)	4.313 (df = 173)
F Statistic	2,430.495*** (df = 1; 350)	904.720*** (df = 1; 173)	1,172.449*** (df = 1; 350)	445.182*** (df = 1; 173)
Note: *p<0.1; **p<0.05; ***p<0.01				

Figure 4: Regression results for Global Sales

Dependent variable:				
	NA_Sales (1)	NA_ProductSales (2)	EU_Sales (3)	EU_ProductSales (4)
EU_Sales	1.187*** (0.064)			
EU_ProductSales		0.917*** (0.088)		
NA_Sales			0.419*** (0.023)	
NA_ProductSales				0.420*** (0.040)
Constant	0.564*** (0.166)	2.027*** (0.398)	0.589*** (0.095)	1.179*** (0.274)
Observations	352	175	352	175
R2	0.498	0.386	0.498	0.386
Adjusted R2	0.496	0.382	0.496	0.382
Residual Std. Error	2.420 (df = 350)	3.582 (df = 173)	1.438 (df = 350)	2.424 (df = 173)
F Statistic	346.883*** (df = 1; 350)	108.555*** (df = 1; 173)	346.883*** (df = 1; 350)	108.555*** (df = 1; 173)
Note: *p<0.1; **p<0.05; ***p<0.01				

Figure 5: Regression results for North America (NA) and European (EU) Sales

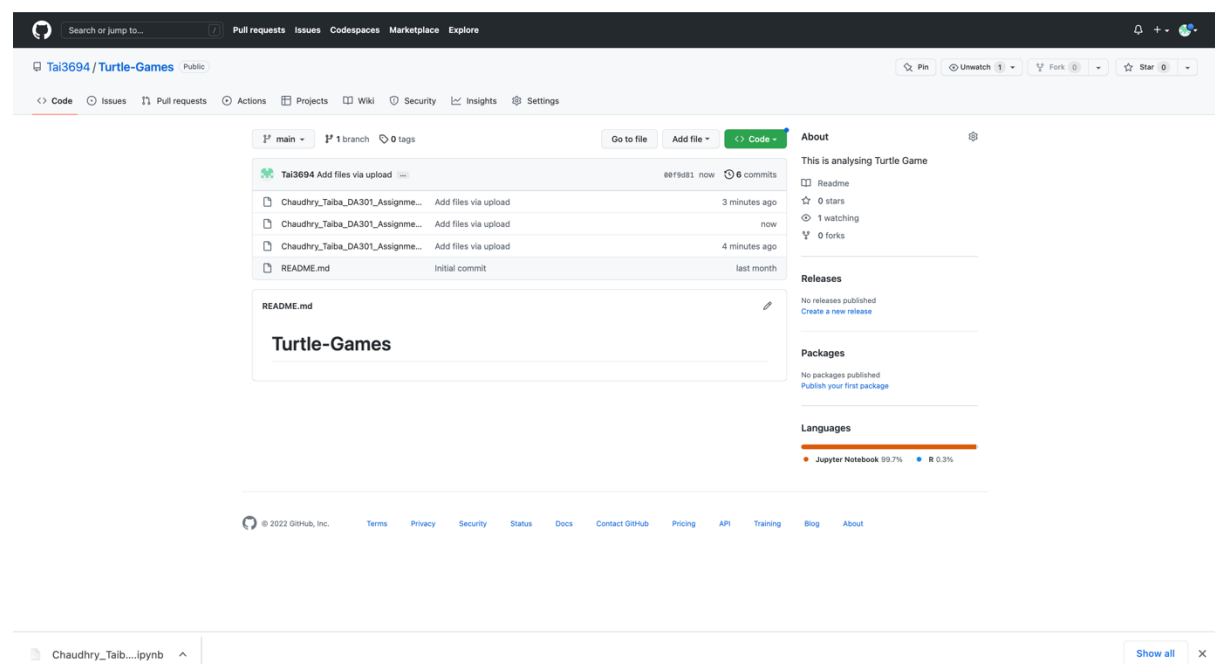
Overall, there tends to be a positive relationship between all 3 sales. The  $R^2$  can evaluating how well our model fits the data. In the first column of figure 1, 87.4% of the variation within Global sales is explained by NA\_Sales. Therefore, this model's goodness of fit and accuracy can help to make predictions and further analysis. Suggest that not all models will be reliable and suitable for further analysis or forecasting. This may require us to improve the model. For example, we could adjust the model by taking logarithms, exponentials, including other variables and much more. From here, small insights can be seen, such as the

GitHub: <https://github.com/Tai3694/Turtle-Games>

level of consumer spending, the popular platforms, and the number of sales, which can be further examined by looking at specific locations.

One could further determine the reliability of the data through other approaches. Organizations can explain their performance using descriptive statistics by synthesizing enormous amounts of data into precise performance indicators. The mean between the three different sales is not similar and indicates the data is not normally distributed. Running this test for the original and aggregated data sets for the 3 different sales. If we assume a 5% significance level, we observe that in all cases that the p-value is lower than this, indicating that the data is not normally distributed. Overall, this would indicate that the data collected by Turtle Games is not entirely trustworthy.

These are just a few simple methods for modifying and using DataFrames. From here, small insights can be seen, such as the level of consumer spending, the popular platforms, and the number of sales, which can be further examined by looking at specific locations.



As a result, I decide to create a GitHub repository where I can keep and update the project files. This is a great environment for software project collaboration, monitoring, and archiving. This promotes teamwork and will help me present the codes and projects to my colleagues.

## Visualisations and insights

Graphs or other visual representations may be more useful when working with complex data. Matplotlib is a Python graphics library for data visualisation that works well with Numpy and Pandas. It is based on Matplotlib and tightly integrates with Pandas data structures. Similarly, the tidyverse package in R is used to build visualisations and further study the data. We can use the ggplot() or qplot() tools.

With this, various graph kinds can be produced using the tidyverse package in R. Specifically utilising the qplot command to collect information from the data set through scatterplots, histograms, and boxplots. Furthermore, looking at the multiple Platforms can help identify popular game consoles. In particular, the DS, PC, PS3, Wii and X360 seem to be the most popular platforms. It would be best to use these popular Platforms to boost sales.

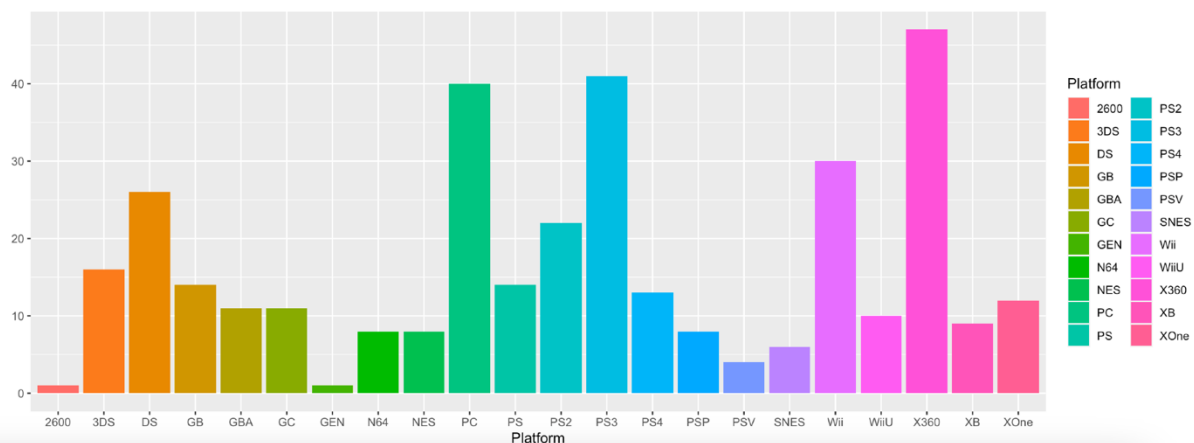


Figure 6: Bar graph for different platform sales

seem to make sense or be easy to understand. Using the original data frame and the aggregated data, I found that in all sales, there were outliers in the dataset. This was more significant for EU Sales and NA sales. It also is a valuable method for visualising variations between regional sales.

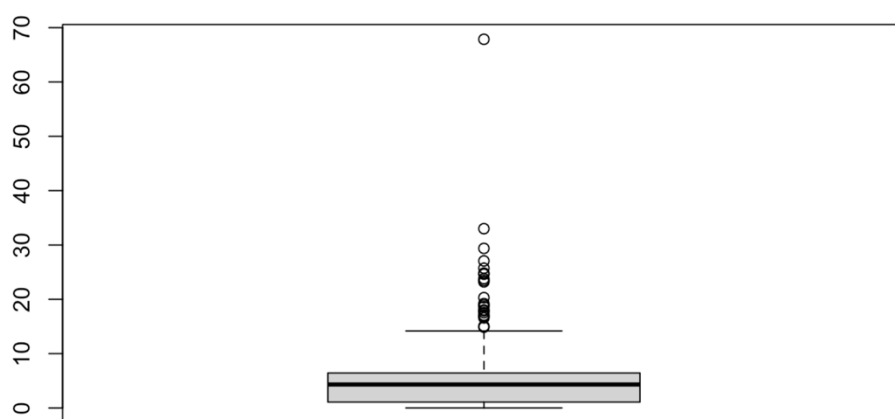


Figure 7: Boxplot for Global sales

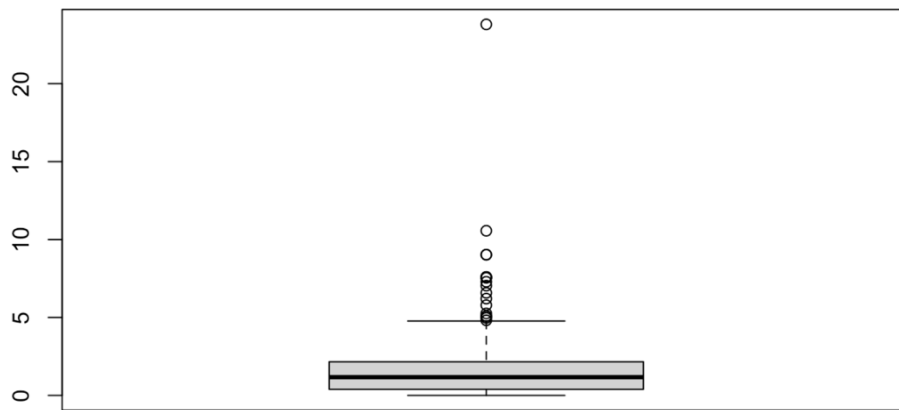


Figure 8: Boxplot for EU sales

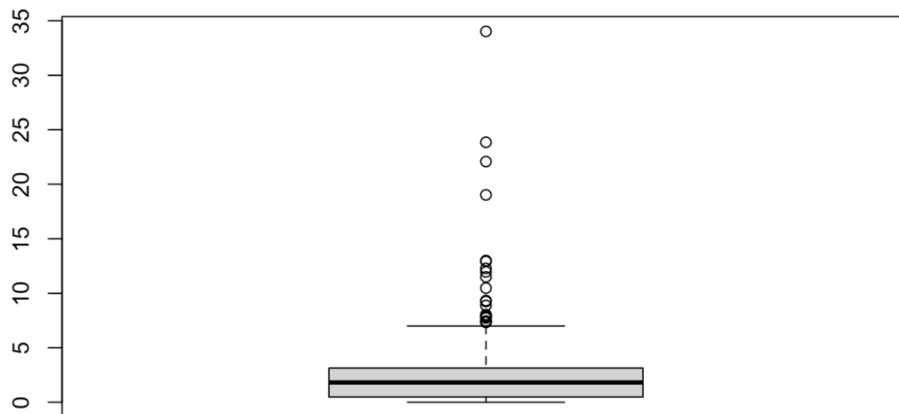


Figure 9: Boxplot for NA sales

In all boxplots, we notice significant outliers in all regional sales. This seems logical since Global\_Sales is made up from NA\_Sales and EU\_Sales. Yet, there also seem to be more outliers, which are more spread out for NA\_Sales. We could further assess the reliability through other plots, such as a histogram or a QQ-Plot.

#### Histogram of Global\_Sales

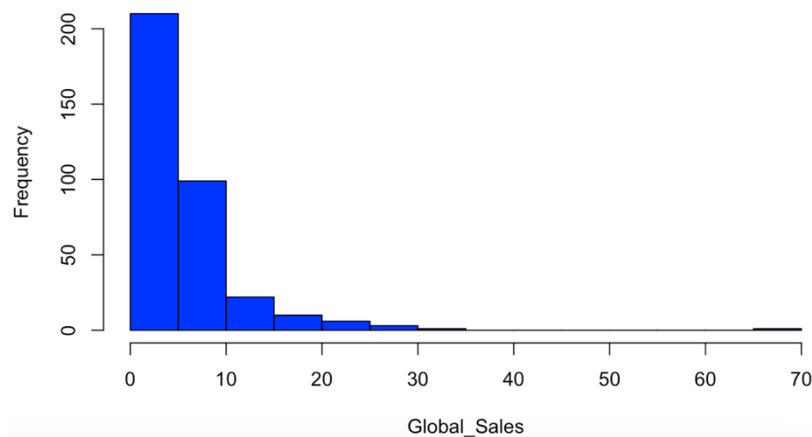


Figure 10: Histogram of Global sales

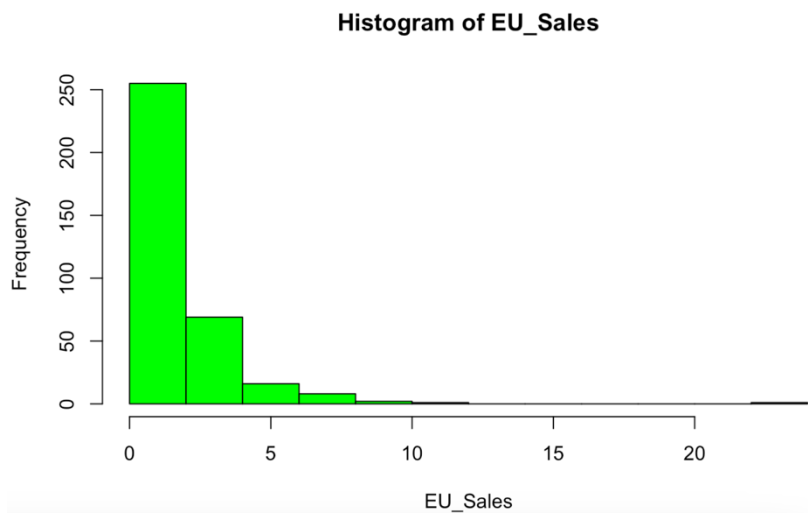


Figure 11: Histogram of European (EU) sales

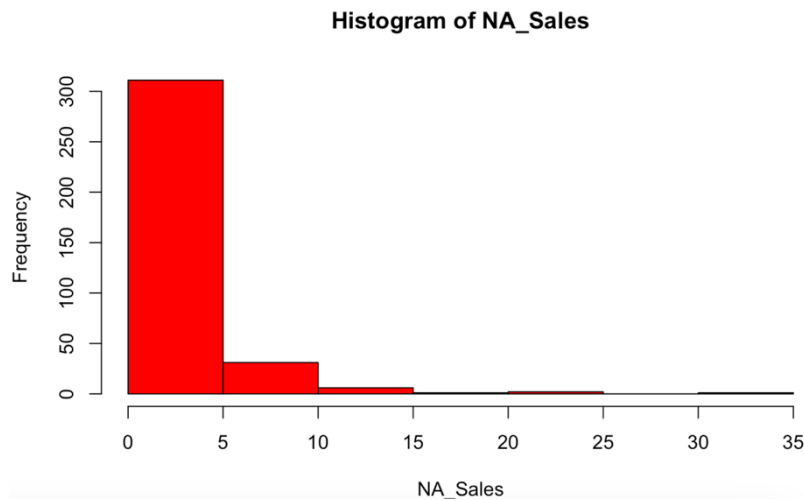


Figure 12: Histogram of North America (NA) sales

I did a similar procedure for the aggregated data and got identical results. This helps in identifying that the data doesn't follow a normal distribution. Additionally, a QQ-Plot can be used to identify whether the data follows a normal distribution, and so far, it seems not to.

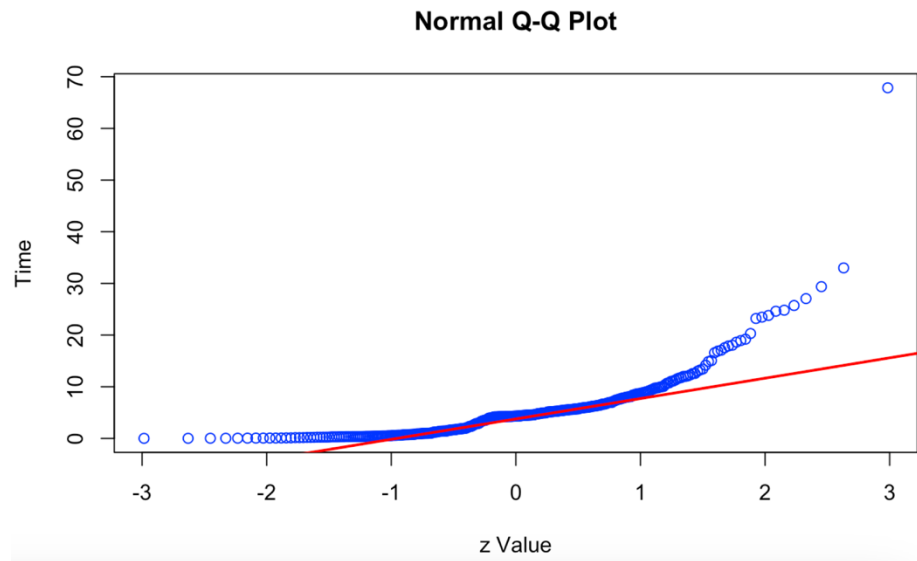


Figure 13: QQ-Plot for Global sales

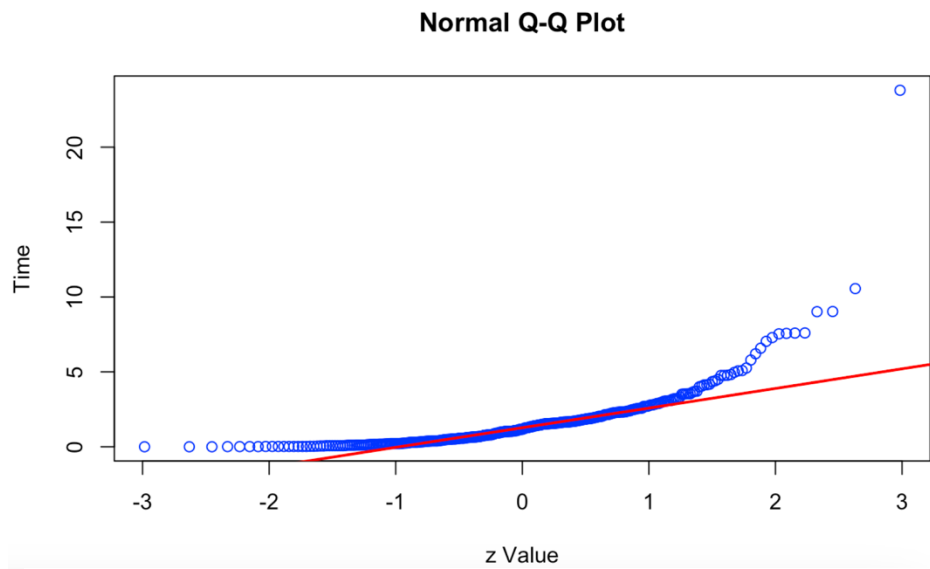


Figure 14: QQ-Plot for European (EU) sales





Besides the sales data, we could also look at the customer reviews. Turtle Games would be able to learn about how their consumers feel about their goods and services. One can visualise this by creating a Word cloud, where a tag is a visual representation of text data; Therefore, we can obtain the most popular or common word using customer reviews and summary variables. The relevant packages, including Numpy, Matplotlib, nltk, OS, and a few others, are needed. I could then make two WordClouds: one for the reviews variable and another for the summary variable.



GitHub: <https://github.com/Tai3694/Turtle-Games>



Figure 17: WordCloud for summary

When looking at reviews the most common words are game, love, great, fun, new and much more. In fact, the common tags in customer summaries is very similar and we see more words such as five stars, excellent and much more. Displayed below are the 15 most frequent words in the customer reviews and customer summaries.

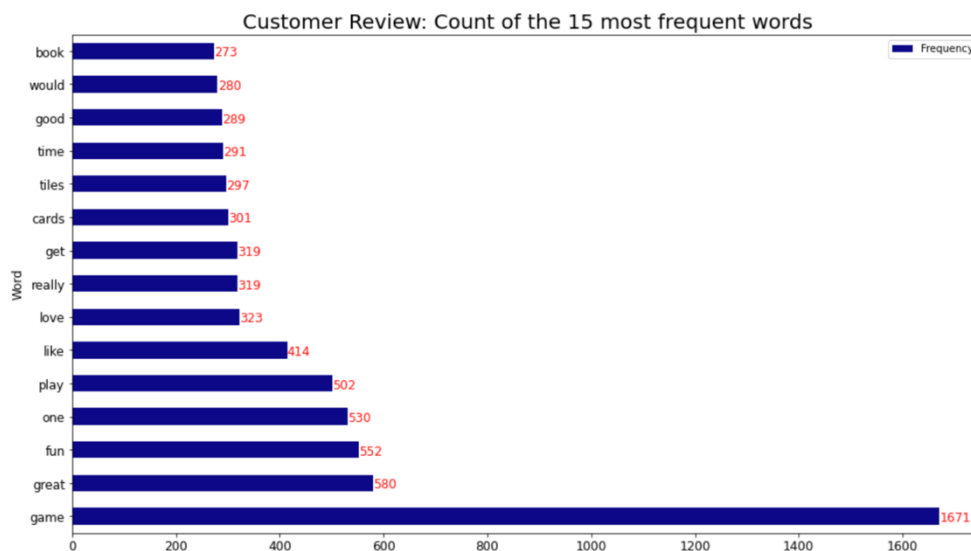


Figure 18: Top 15 common words for reviews

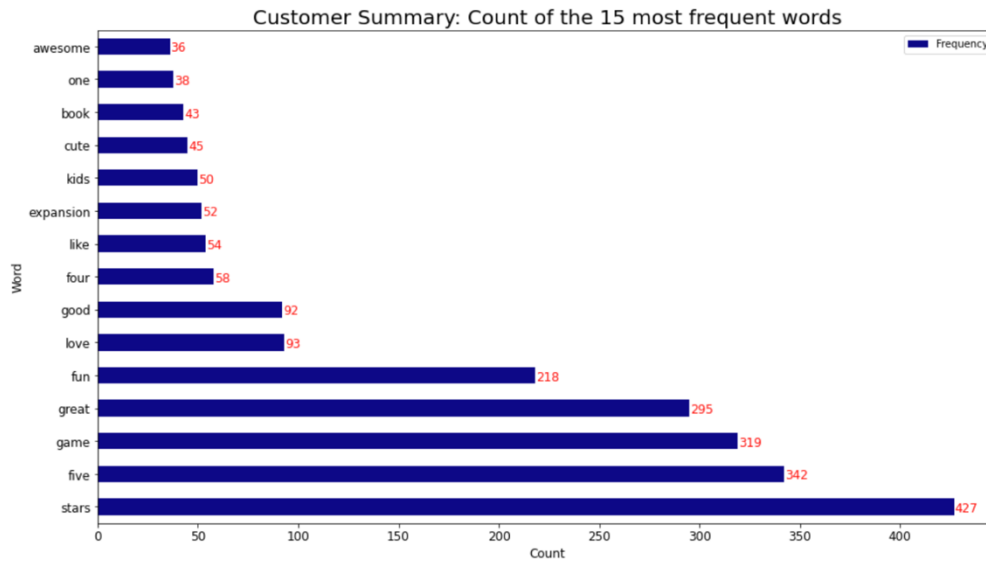


Figure 19: Top 15 common words for summary

We may calculate the percentage of customer evaluations that indicate negative, positive, or neutral sentiments to gain a quick overview of current consumer satisfaction. This is done on the bases of a polarity score. These polarity scores range from -1 to 1, where -1 indicates the speaker's feelings toward the topic are strongly negative, and +1 indicates their feelings toward the subject are strongly positive. Hence, using TextBlob in Python, below shows the top 20 positive and negative comments in reviews and summaries.

index		review	summary	tokens	tokens2	polarity1	polarity2
782	790	perfect	aqire game	[perfect]	[aqire, game]	1.0	-0.4
1695	1726	excellent toy to simulate thought	five stars	[excellent, toy, to, simulate, thought]	[five, stars]	1.0	0.0
193	194	awesome gift	five stars	[awesome, gift]	[five, stars]	1.0	0.0
1929	1967	perfect for tutoring my grandson in spelling	tutoring	[perfect, for, tutoring, my, grandson, in, spe...]	[tutoring]	1.0	0.0
1159	1177	awesome addition to my rpg gm system	five stars	[awesome, addition, to, my, rpg, gm, system]	[five, stars]	1.0	0.0
1150	1168	best set buy 2 if you have the means	five stars	[best, set, buy, 2, if, you, have, the, means]	[five, stars]	1.0	0.0
517	524	perfect just what i ordered	five stars	[perfect, just, what, i, ordered]	[five, stars]	1.0	0.0
1684	1715	awesome toy	five stars	[awesome, toy]	[five, stars]	1.0	0.0
1689	1720	it is the best thing to play with and also min...	three stars	[it, is, the, best, thing, to, play, with, and...]	[three, stars]	1.0	0.0
613	621	wonderful for my grandson to learn the resurre...	five stars	[wonderful, for, my, grandson, to, learn, the,...]	[five, stars]	1.0	0.0
1119	1135	awesome set	five stars	[awesome, set]	[five, stars]	1.0	0.0
1580	1609	this was perfect to go with the 7 bean bags i...	five stars	[this, was, perfect, to, go, with, the, 7, bea...]	[five, stars]	1.0	0.0
1380	1401	one of the best board games i played in along ...	five stars	[one, of, the, best, board, games, i, played, ...]	[five, stars]	1.0	0.0
601	609	delightful product	five stars	[delightful, product]	[five, stars]	1.0	0.0
1282	1301	its awesome	five stars	[its, awesome]	[five, stars]	1.0	0.0
7	7	came in perfect condition	five stars	[came, in, perfect, condition]	[five, stars]	1.0	0.0
164	165	awesome book	five stars	[awesome, book]	[five, stars]	1.0	0.0
583	591	wonderful product	five stars	[wonderful, product]	[five, stars]	1.0	0.0
923	933	awesome	five stars	[awesome]	[five, stars]	1.0	0.0
489	496	excellent activity for teaching selfmanagement...	five stars	[excellent, activity, for, teaching, selfmanag...]	[five, stars]	1.0	0.0

Figure 20: Top 20 positive reviews

	index	review	summary	tokens	tokens2	polarity1	polarity2
1014	1028	include all items got the package quickly	one of the best	[include, all, items, got, the, package, quickly]	[one, of, the, best]	0.333333	1.0
1899	1935	quick fun easy to learn wide age range fast pl...	excellent	[quick, fun, easy, to, learn, wide, age, range...	[excellent]	0.223810	1.0
805	815	love this game and have been playing it for ov...	one of the best games ever	[love, this, game, and, have, been, playing, i...	[one, of, the, best, games, ever]	0.366667	1.0
1601	1630	received item todaysuper fast shipment i cant ...	awesome learning tool	[received, item, todaysuper, fast, shipment, i...	[awesome, learning, tool]	0.200000	1.0
162	163	this was a christmas present for a nephew who ...	he was very happy with his gift	[this, was, a, christmas, present, for, a, nep...	[he, was, very, happy, with, his, gift]	0.500000	1.0
1152	1170	just what i needed for my collection group of ...	best orcs from wotc	[just, what, i, needed, for, my, collection, g...	[best, orcs, from, wotc]	0.050000	1.0
1367	1388	love this can expand current game with some ne...	awesome expansion	[love, this, can, expand, current, game, with...	[awesome, expansion]	0.005587	1.0
1062	1078	if you have anyone in your life that enjoys bo...	perfect gift	[if, you, have, anyone, in, your, life, that, ...	[perfect, gift]	0.333333	1.0
139	140	i bought 8 of these for my 3 year old daughter...	awesome sticker activity for the price	[i, bought, 8, of, these, for, my, 3, year, ol...	[awesome, sticker, activity, for, the, price]	0.318750	1.0
1153	1171	easy to understand quick start hours of fun wi...	awesome	[easy, to, understand, quick, start, hours, of...	[awesome]	0.291667	1.0
639	647	beautiful we look forward to the easter story ...	wonderful	[beautiful, we, look, forward, to, the, easter...	[wonderful]	0.412500	1.0
643	651	we have several different sets of these magnet...	all f the mudpuppy toys are wonderful	[we, have, several, different, sets, of, these...	[all, f, the, mudpuppy, toys, are, wonderful]	0.090000	1.0
1465	1488	my 5 yearold son absolutely loves this game he...	the perfect gift for preschool construction fans	[my, 5, yearold, son, absolutely, loves, this...	[the, perfect, gift, for, preschool, construct...	-0.041012	1.0
1067	1083	this is in my opinion the best dungeon crawler...	best dungeon crawler	[this, is, in, my, opinion, the, best, dungeon...	[best, dungeon, crawler]	0.361111	1.0
649	657	my son is 2 years and 4 months he is a little ...	awesome puzzle	[my, son, is, 2, years, and, 4, months, he, is...	[awesome, puzzle]	0.351562	1.0
968	980	loads of fun if youve played any of the other ...	the best among the dd boardgames	[loads, of, fun, if, youve, played, any, of, t...	[the, best, among, the, dd, boardgames]	0.026657	1.0
1211	1230	i have recently gotten my sons involved in dd ...	awesome addition to our dd antics	[i, have, recently, gotten, my, sons, involved...	[awesome, addition, to, our, dd, antics]	0.053089	1.0
160	161	my 8 year old son loves this drawing book love...	awesome book	[my, 8, year, old, son, loves, this, drawing, ...	[awesome, book]	0.100000	1.0
1395	1417	yes quick wonderful and accurate	wonderful and	[yes, quick, wonderful, and, accurate]	[wonderful, and]	0.577778	1.0
1431	1454	me and my wife play lords of waterdeep more th...	awesome expansion	[me, and, my, wife, play, lords, of, waterdeep...	[awesome, expansion]	0.134848	1.0

Figure 21: Top 20 positive summaries

	index	review	summary	tokens	tokens2	polarity1	polarity2
207	208	booo unles you are patient know how to measure...	boring unless you are a craft person which i am	[booo, unles, you, are, patient, know, how, to...	[boring, unless, you, are, a, craft, person, w...	-1.000000	-1.00
181	182	incomplete kit very disappointing	incomplete kit	[incomplete, kit, very, disappointing]	[incomplete, kit]	-0.780000	0.00
1773	1804	im sorry i just find this product to be boring...	disappointing	[im, sorry, i, just, find, this, product, to, ...	[disappointing]	-0.583333	-0.60
362	364	one of my staff will be using this game soon s...	anger control game	[one, of, my, staff, will, be, using, this, ga...	[anger, control, game]	-0.550000	-0.55
1501	1524	expensive for what you get	two stars	[expensive, for, what, you, get]	[two, stars]	-0.500000	0.00
116	117	i bought this as a christmas gift for my grand...	stickers	[i, bought, this, as, a, christmas, gift, for...	[stickers]	-0.500000	0.00
229	230	i found the directions difficult	three stars	[i, found, the, directions, difficult]	[three, stars]	-0.500000	0.00
289	290	instructions are complicated to follow	two stars	[instructions, are, complicated, to, follow]	[two, stars]	-0.500000	0.00
300	301	difficult	three stars	[difficult]	[three, stars]	-0.500000	0.00
226	227	this was a gift for my daughter i found it dif...	two stars	[this, was, a, gift, for, my, daughter, i, fou...	[two, stars]	-0.500000	0.00
173	174	i sent this product to my granddaughter the po...	faulty product	[i, sent, this, product, to, my, granddaughter...	[faulty, product]	-0.491667	0.00
345	347	my 8 yearold granddaughter and i were very fru...	frustating	[my, 8, yearold, granddaughter, and, i, were, ...	[frustating]	-0.446250	0.00
531	538	i purchased this on the recommendation of two ...	hmmm	[i, purchased, this, on, the, recommendation, ...	[hmmm]	-0.440741	0.00
305	306	very hard complicated to make these	one star	[very, hard, complicated, to, make, these]	[one, star]	-0.439583	0.00
814	824	was a gift for my son he loves the game	five stars	[was, a, gift, for, my, son, he, loves, the, g...	[five, stars]	-0.400000	0.00
795	803	this game is a blast	five stars	[this, game, is, a, blast]	[five, stars]	-0.400000	0.00
977	989	if you like me used to play dd but now you and...	like me used to play dd	[if, you, like, me, used, to, play, dd, but, n...	[like, me, used, to, play, dd]	-0.400000	0.00
1796	1827	jun game	five stars	[jun, game]	[five, stars]	-0.400000	0.00
1423	1446	you can play the expansions one at a time or a...	two expansions for the price of one	[you, can, play, the, expansions, one, at, a, ...	[two, expansions, for, the, price, of, one]	-0.400000	0.00
798	806	i bought this for my son he loves this game	five stars	[i, bought, this, for, my, son, he, loves, thi...	[five, stars]	-0.400000	0.00

Figure 22: Top 20 negative reviews

	index	review	summary	tokens	tokens2	polarity1	polarity2
207	208	booou unless you are patient know how to measure...	boring unless you are a craft person which i am	[booou, unless, you, are, patient, know, how, to...	[boring, unless, you, are, a, craft, person, w...	-1.000000	-1.000000
21	21	these are cute tatoos for those who love pirat...	the worst value ive ever seen	[these, are, cute, tatoos, for, those, who, lo...	[the, worst, value, ive, ever, seen]	0.050000	-1.000000
819	829	ive discovered that im really new school when ...	boring	[ive, discovered, that, im, really, new, schoo...	[boring]	0.156399	-1.000000
1148	1166	before this i hated running any rpg campaign d...	before this i hated running any rpg campaign d...	[before, this, i, hated, running, any, rpg, ca...	[before, this, i, hated, running, any, rpg, ca...	-0.133333	-0.900000
1	1	an open letter to galeforce9 your unpainted mi...	another worthless dungeon masters screen from ...	[an, open, letter, to, galeforce9, your, unpai...	[another, worthless, dungeon, masters, screen,...	0.035952	-0.800000
785	793	my mom already owned an acquire game but she a...	disappointed	[my, mom, already, owned, an, acquire, game, b...	[disappointed]	-0.046364	-0.750000
1591	1620	i was thinking it was a puppet but it is not l...	disappointed	[i, was, thinking, it, was, a, puppet, but, it...	[disappointed]	-0.218750	-0.750000
143	144	my kids grew up with a peg bench and hammer an...	disappointed	[my, kids, grew, up, with, a, peg, bench, and...	[disappointed]	0.108173	-0.750000
623	631	eggs split and were unusable	disappointed	[eggs, split, and, were, unusable]	[disappointed]	0.000000	-0.750000
361	363	i found that this card game does the opposite ...	promotes anger instead of teaching calming met...	[i, found, that, this, card, game, does, the, ...	[promotes, anger, instead, of, teaching, calmi...	-0.126190	-0.700000
875	885	i guess i did not look closely to the informat...	too bad this is not what i was expecting	[i, guess, i, did, not, look, closely, to, the...	[too, bad, this, is, not, what, i, was, expect...	0.200000	-0.700000
880	890	the game tiles board and tile stands are all m...	bad qualityall made of paper	[the, game, tiles, board, and, tile, stands, a...	[bad, qualityall, made, of, paper]	-0.162500	-0.700000
177	178	at age 31 i found these very difficult to make...	at age 31 i found these very difficult to make	[at, age, 31, i, found, these, very, difficult...	[at, age, 31, i, found, these, very, difficult...	-0.275000	-0.650000
511	518	great game kids love to play again and again	mad dragon	[great, game, kids, love, to, play, again, and...	[mad, dragon]	0.300000	-0.625000
100	101	the book is only 4 pages and its about the siz...	small and boring	[the, book, is, only, 4, pages, and, its, abou...	[small, and, boring]	-0.075000	-0.625000
1773	1804	im sorry i just find this product to be boring...	disappointing	[im, sorry, i, just, find, this, product, to, ...	[disappointing]	-0.583333	-0.600000
1001	1015	i wish id watched some of the gameplay videos ...	disappointing	[i, wish, id, watched, some, of, the, gameplay...	[disappointing]	-0.042402	-0.600000
1099	1115	i had high hopes for this game as i am a big f...	disappointing	[i, had, high, hopes, for, this, game, as, i, ...	[disappointing]	-0.038999	-0.600000
797	805	really cheaply produced cardboard playing piec...	disappointing	[really, cheaply, produced, cardboard, playing...	[disappointing]	0.400000	-0.600000
991	1003	if you play dungeons and dragons then you will...	then you will find this board game to be dumb ...	[if, you, play, dungeons, and, dragons, then, ...	[then, you, will, find, this, board, game, to,...	-0.393750	-0.591667

Figure 23: Top 20 negative summaries

## Patterns and Predictions

In conclusion, Turtle Game can gain some insights into how they could improve their sales and develop marketing or advertising strategies, products and services to reach this goal. Therefore, to predict the accumulation of loyalty points, it would be helpful to use spending and spending habits to an extent.

Following this, I further explored spending score and remuneration to recognise these variables' usefulness to understand better what aspects could improve future sales. With the help of current customer data, clustering can be utilised in the marketing industry to discover different client groups. Turtle Games has a global appeal with a large international customer base. Knowing which platforms are most popular can help Turtle Games to grow sales in these specific markets. It would also be helpful to analyse customer reviews and summaries when it comes to creating the most suitable goods or services for consumers. I was specifically able to identify the most frequent words, the top 20 positive and negative comment. Hence, marketing campaigns can focus on solving the problems customers previously faced.

Overall, I conducted some analysis and made some recommendations regarding tools and strategies for Turtle Games to use to achieve their goal of improving their sales. I performed numerous tests to test the reliability of the sales data. Therefore, Turtle Games could build a

GitHub: <https://github.com/Tai3694/Turtle-Games>

more practical approach to collecting data, as well as collecting data in other aspects. For instance, including the time these sales occur, further different regions, etc.