

TURTLE GAMES REPORT

TABLE OF CONTENTS

- 01** Recommended Lego prices
- 02** Sentiment & reviews analysis
- 03** Customers' segmentation
- 04** Overview of past sales trends
- 05** Linear regression and prediction of future sales

PART 1

PREDICTING LEGO PRICE

Assisting Turtle Games, a global games manufacturer and retailer, to achieve its business objective, improve overall sales performance.



In part 1 of this analysis I will help them determine the optimal price at which they should sell a particular products all based on historical data.

By initially exploring the numeric dataset, a descriptive statistic summary from Excel shows the distribution of prices and pieces count data variables, checked the value of the Kurtosis and skewness.

I have conducted a single and multiple regression analysis to predict the optimal price at what they should sell a Lego sets of 8,000 pieces

As well as the predicted optimal price I will predict the optimal price for a Lego sets with 8,000 pieces likely to be purchased by 30-year-old customers.

Kurtosis identifies whether the tails of a distribution contain extreme values.

Very high kurtosis values could indicate that there are prices making an unusually high or low number of sales as the general guideline is that kurtosis between -3 and 3 is acceptable.

FROM THE DATA

The Excel table shows 31.65 and 19.86 for kurtosis for Lego price and pieces, which are very high figures, meaning that prices and sale price and piece count are high.

list_price		piece_count	
Mean	65.1395524	Mean	493.4059
Standard Error	0.83068962	Standard Error	7.453888
Median	36.5878	Median	216
Mode	12.1878	Mode	1
Standard Deviation	91.9817629	Standard Deviation	825.3646
Sample Variance	8460.64471	Sample Variance	681226.7
Kurtosis	31.6510184	Kurtosis	19.86268
Skewness	4.68052677	Skewness	3.968058
Range	1104.87	Range	7540
Minimum	0	Minimum	1
Maximum	1104.87	Maximum	7541
Sum	798676.052	Sum	6049650
Count	12261	Count	12261

Skewness measures the symmetry of the distribution tails (which relates to assumptions of normality in linear regression).

Skewness would indicate that the data set is biased towards either low or high sale prices.

The general guideline is that between -0.5 and 0.5 indicates a symmetric normal distribution of data set.

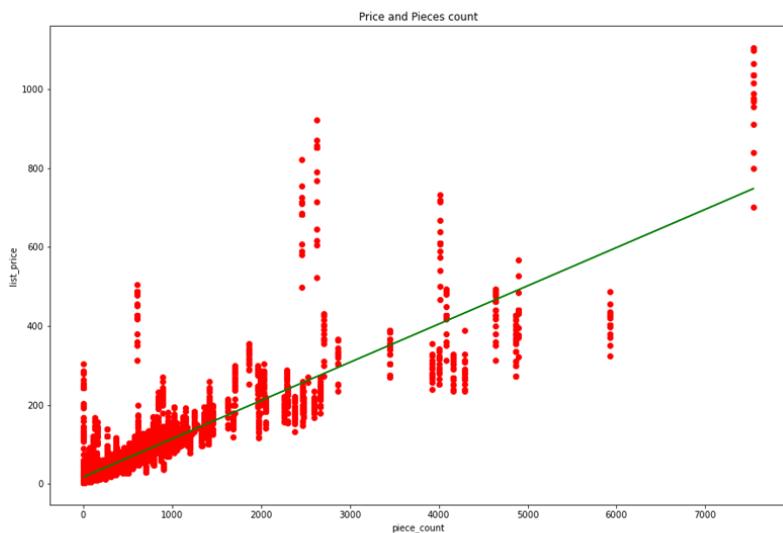
The variable price in our dataset is positively skewed with numbers, above a normal distribution 4.68 and 3.96

MEASURING REGRESSION

Followed by an analysis of the correlation for these two variables, in order to find the optimal price of the Lego pieces.

There's linearity between the two variables 'piece count and price', and the price trend line can be observed from the graph below. The trend line in the graph suggests that as the number of pieces increase so does the price.

Number of Lego pieces & price



0.09

Correlation coefficient

18

Intercept

The correlation coefficient value between the two variables is positive (0.09) meaning that there is a direct linear relationship between the number pieces and the price sold. This coefficient value is also far from 1.00 and can be considered as they are not very strongly correlated.

- The intercept value of 18 tells us that as the high stat variable increases by 1, the predicted value of Median_s decreases by 18.
- The standard error and the coefficient of determination (R-squared) are two statistics that evaluate the predictive ability of a regression equation. The price standard error is 0.83 for price and 7.45 for piece count. While the R-squared is 0.76 which is telling us that the model is explaining a fraction over 76% itself.
- After performing a linear regression in a sample test data of 70/30 and 100 random. The predicted optimal price for a product of 8,000 Lego pieces is 782 dollars
- And a Lego product with 8,000 or more pieces, likely to be purchased by a 30 year old customer, the recommended price is 784 dollars.

PART 2

SENTIMENT & REVIEWS



In order to maximize the sales performance of products sold by Turtle Games, the second part of the analysis will evaluate:

- Customer's reviews for toys and games from its e-commerce site. Find out what is the general sentiment of customers across all products and
- Check on the polarity shown by the top 20 most positive and negative reviews.

The approach: Uploading the already extracted ready data into Python, considered some duplicates and no null values in the dataset. After careful consideration of the nature of the review, I deleted all the duplicate rows and kept all the verified and unverified reviews.

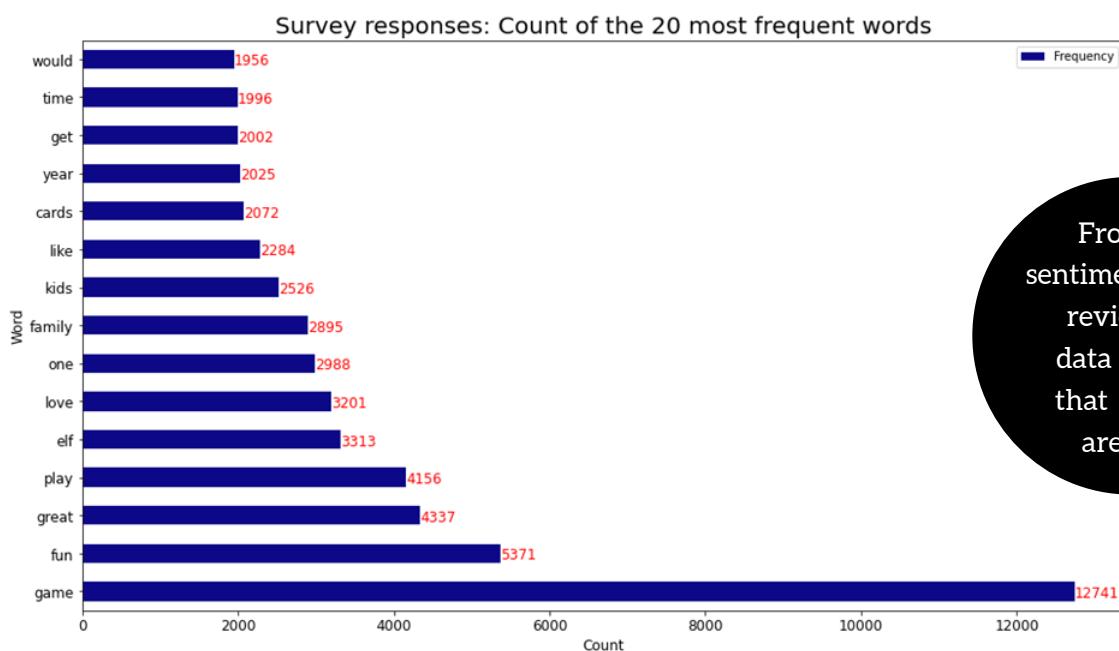
Steps: to standardise the text, first convert the sentences into lowercase, removing punctuation marks, transforms data into tokens and interpret the actual meaning of the texts, ready to analyse the sequence of the words.

WORDS

By plotting the words on a word cloud, it is identified which are the most frequently occurring words. At first, an image showing the most common words showed a lot of stop words, so I removed them and ended up with the second image shown below.

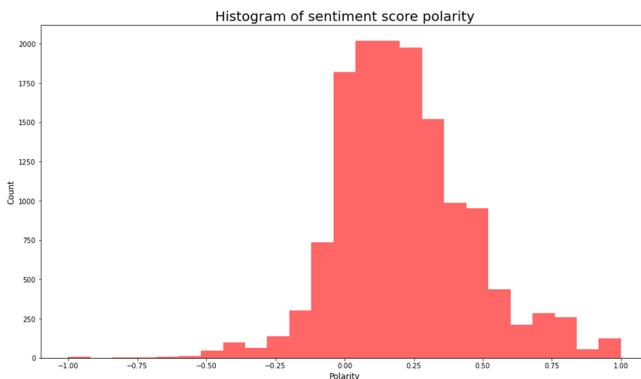


Count of most common words



SENTIMENT

By generating a polarity score, an assessment of the overall sentiment of the users shows below:



As the tail of the histogram is towards the negative values, with a histogram skewed towards the right, this indicates a negative skewness. By looking at this graph I can infer that the general sentiment of the customers is slightly negative.

From the top 20 positive and top 20 negative reviews we understand customers are happy with the delivery, description and it is bought for gifts. They find the items bought as expected.

reviewText	summary	polarity	subjectivity
7 came in perfect condition	Five Stars	1.000000	1.000000
164 awesome book	Five Stars	1.000000	1.000000
193 awesome gift	Five Stars	1.000000	1.000000
489 excellent activity for teaching selfmanagement skills	Five Stars	1.000000	1.000000
517 perfect just what i ordered	Five Stars	1.000000	1.000000
583 wonderful product	Five Stars	1.000000	1.000000
601 delightful product	Five Stars	1.000000	1.000000
613 wonderful for my grandson to learn the resurrection story	Five Stars	1.000000	1.000000
782 perfect	Aquire game	1.000000	1.000000
922 awesome	Five Stars	1.000000	1.000000
1118 awesome set	Five Stars	1.000000	1.000000
1149 best set buy 2 if you have the means	Five Stars	1.000000	0.300000
1158 awesome addition to my rpg gm system	Five Stars	1.000000	1.000000
1279 its awesome	Five Stars	1.000000	1.000000
1376 one of the best board games i played in along time	Five Stars	1.000000	0.300000
1516 my daughter loves her stickers awesome seller thank you	Awesome seller! Thank You	1.000000	1.000000
1573 this was perfect to go with the 7 bean bags i just wish they were not separate orders	Five Stars	1.000000	1.000000
1677 awesome toy	Five Stars	1.000000	1.000000
1682 it is the best thing to play with and also mind blowing in some ways	Three Stars	1.000000	0.300000
1688 excellent toy to simulate thought	Five Stars	1.000000	1.000000

reviewText	summary	polarity	subjectivity
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	BORING UNLESS YOU ARE A CRAFT PERSON WHICH I AM...	-1.000000	1.000000
1987 kids did not like it thought it was boring	Not so much fun	-1.000000	1.000000
3218 some of the suggestions are disgusting	One Star	-1.000000	1.000000
7812 awful we did not receive what was advertised we paid 30 for the boxes set with book we got the elf in a bag without the book	Not What Was Advertised	-1.000000	1.000000
7515 was the elf on the shelf but it didnt have the dvd i was very disappointed	Three Stars	-0.975000	0.975000
8861 i havent even taken it out of the box yet but its already falling apart i contacted customer service and never even got a response i am very disappointed in this product	Poor quality. Falling apart in multiple places	-0.975000	0.975000
8198 i hate the holidays bcuz of the elf he was disgusting i hate him with my life he doesnt leave the shelf alone	I hate the holidays bcuz of the Elf	-0.866667	0.933333
12386 i do not under stand how you keep score or read the scoring i do not like that at all i can never play score with anyone at all i hate that i cant play points	What i do not understand	-0.800000	0.900000
8531 cliche and stupid i should not drink and amazon	Hahaha. Ho Ho Ho	-0.800000	1.000000
8638 just stupid	One Star	-0.800000	1.000000
181 incomplete kit very disappointing	INCOMPLETE KIT	-0.780000	0.910000
13413 i like this product for my daughter she is into the bad kitty book collection so it was an added bonus	Good Kitty	-0.700000	0.666667
4060 ordered for my sons birthday opened it up today to put in the board is damaged before we even take it out of the box. Of the four pieces there is already falling apart and very disappointed	Damaged board out of box	-0.687500	0.687500
4090 id like to upload a photo of the condition of the game bout looks like its been used as a soccer ball 2 corners of the box are smashed in and on a even ripped how i supposed to give this as a gift without it looking like i bought this on clearance very disappointed	X	-0.687500	0.687500
11263 horible and incomplete flash cards not buy not helpful i was too late to return them	One Star	-0.650000	0.800000
2062 this was a bit disappointing my students find it boring and the letters are hard to understand	Not recommended	-0.630556	0.747222
10768 boring did i mention boring well its boring pass on this one there are a lot better games out there	Boring	-0.625000	0.875000
13122 had no idea the extenty you have to go through to put together hundreds and i mean hundreds of pieces that dont snap together it will take my teen age son and i months to put this stupid thing together horrible plan horrible	It will take my teen age son and I months to put this stupid thing together	-0.622500	0.737500

The topmost negative reviews are in relation to a board game. The disappointment is with the game design and product features itself. Another product which includes an elf is in these negative sentiments. The percentage of dissatisfaction with the delivery is low as well as there're no recorded dissatisfaction with prices in the top 20 reviews.

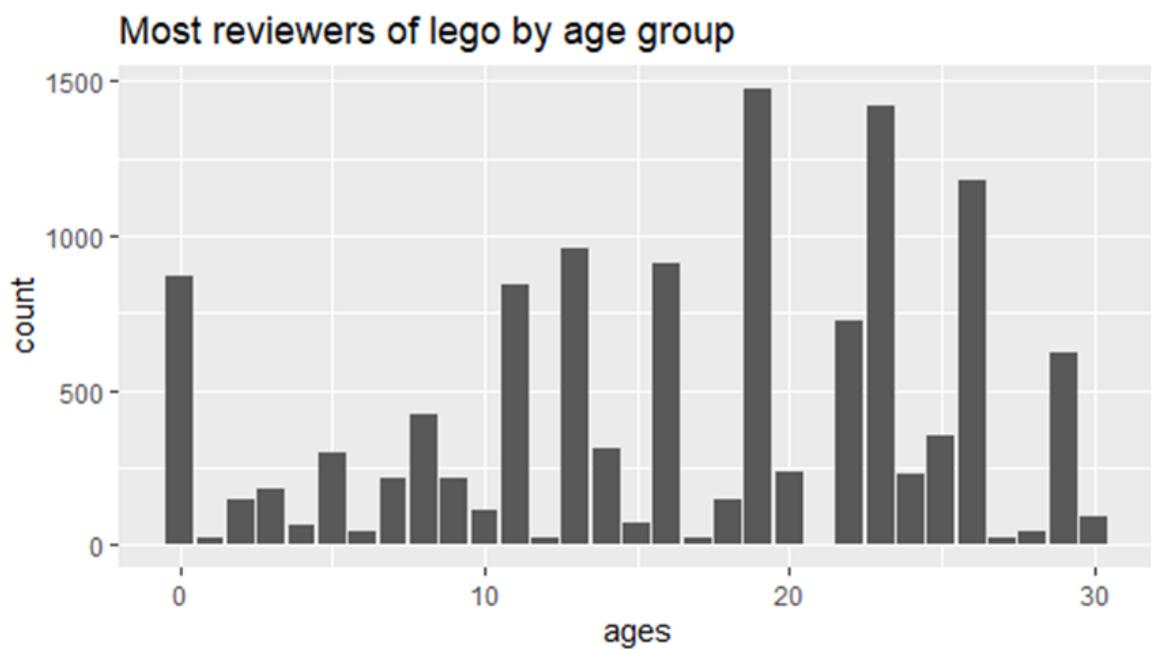
PART 3

CUSTOMERS' SEGMENT

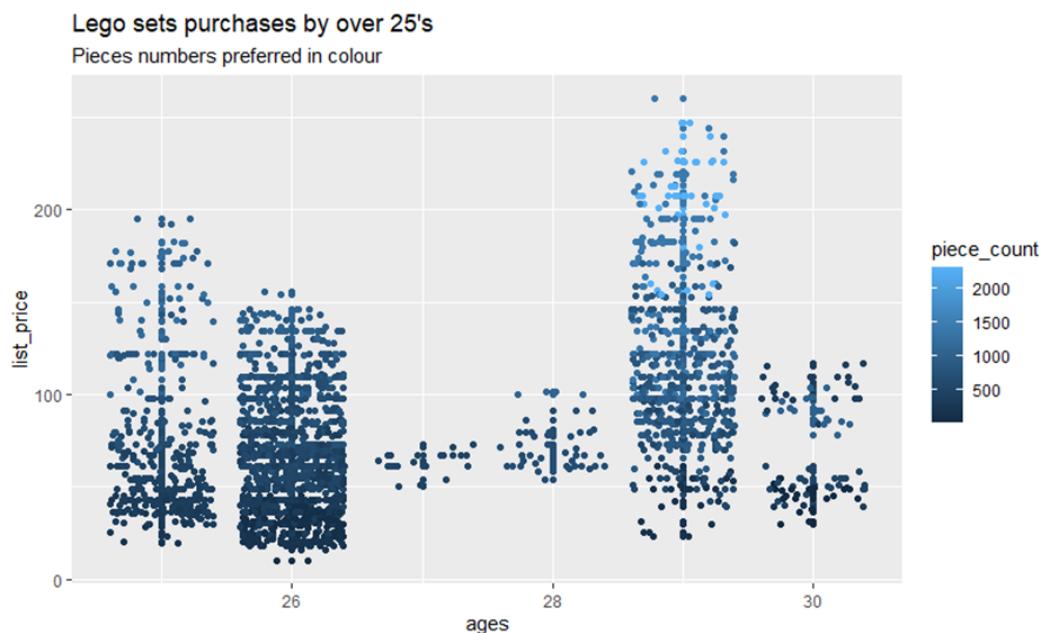
After predicting the optimal price and product segments, I will explore the data to derive information on which customer group will most likely leave a review on the products? Which age group submits the most reviews? As well as to determine the most expensive product purchased by a particular group of customers and what is the most expensive Lego set purchased by customers who are at least 25 years old (>25 years)

The following graphs identify age groups segmentation.

The graphs show a high amount of under 18 reviewers but in this instance, I will assume that the reviewers are guardians purchasing that specific age Lego product. Noted from the graphs the age group that submit the higher number of reviews is the 19 years old. Followed by the 23 years old.



MEASURING SEGMENTATION



PART 4

ANALYSIS OF SALES PAST TREND

Turtle Game provided data with number of video games units in millions sold in North America and Europe to predict the global sales for the next financial year.

After checking the data, I can validate that this data is ordered by ranking of sales, with the most sold product on top.

The data doesn't specify if they if total sales are from 1980 to date, it only says the first game was released in 1980, hence I can't provide a Global Sales statistics due to not having a starting measuring point.

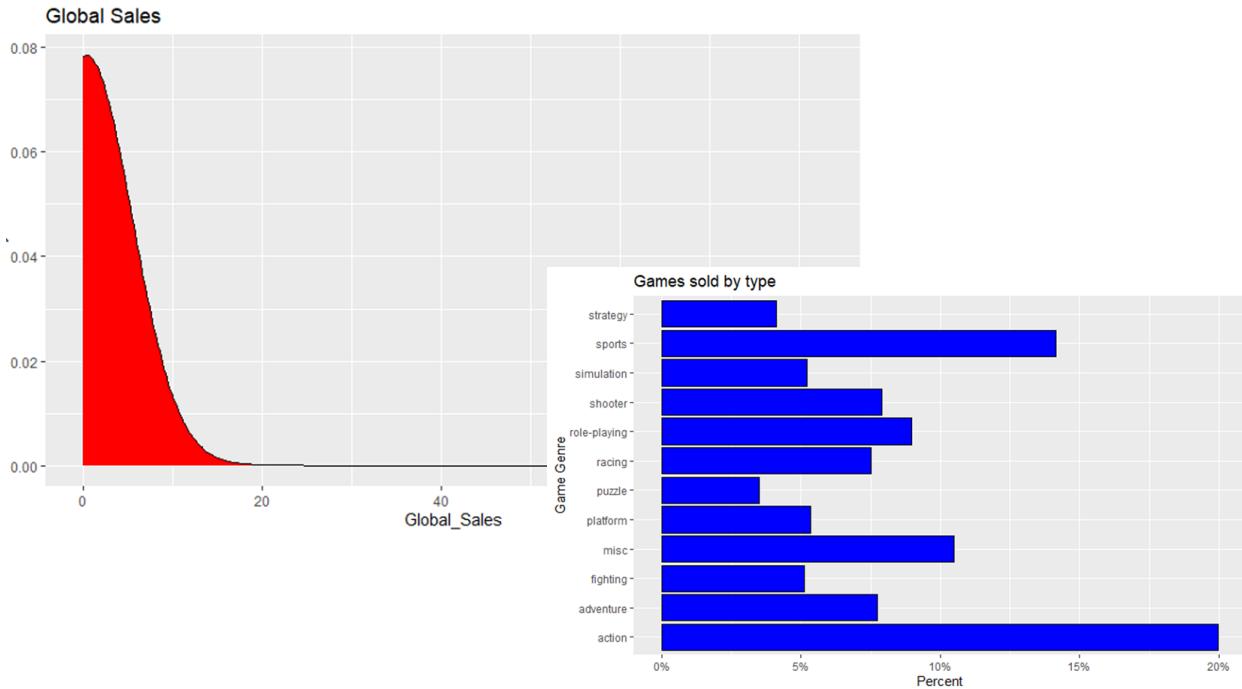
But I will assess the trends and skewness of the total Global Sales given. It's skewed towards the left and the tail of this skew is toward positive values, which suggests a positive skewness.

The correlation coefficient between North America and the Global Sales is 0.94, suggesting a strong positive correlation. Moreover, the correlation between the European sales and the online sales is 0.66, also positive. These values are both closer to 1 than to 0, suggesting a strong relationship between the variables.

Kurtosis defines how heavily the tails of a distribution differ from the tails of a normal distribution. In other words, identifies whether the tails of a distribution contain extreme values. In this dataset we have extremely high Kurtosis in value which could indicate that sales are unusual high.

Further visualising the trends of the data I plotted the trends between the different type of games sold globally by genre in the bar chart,

left bottom corner.



PART 5

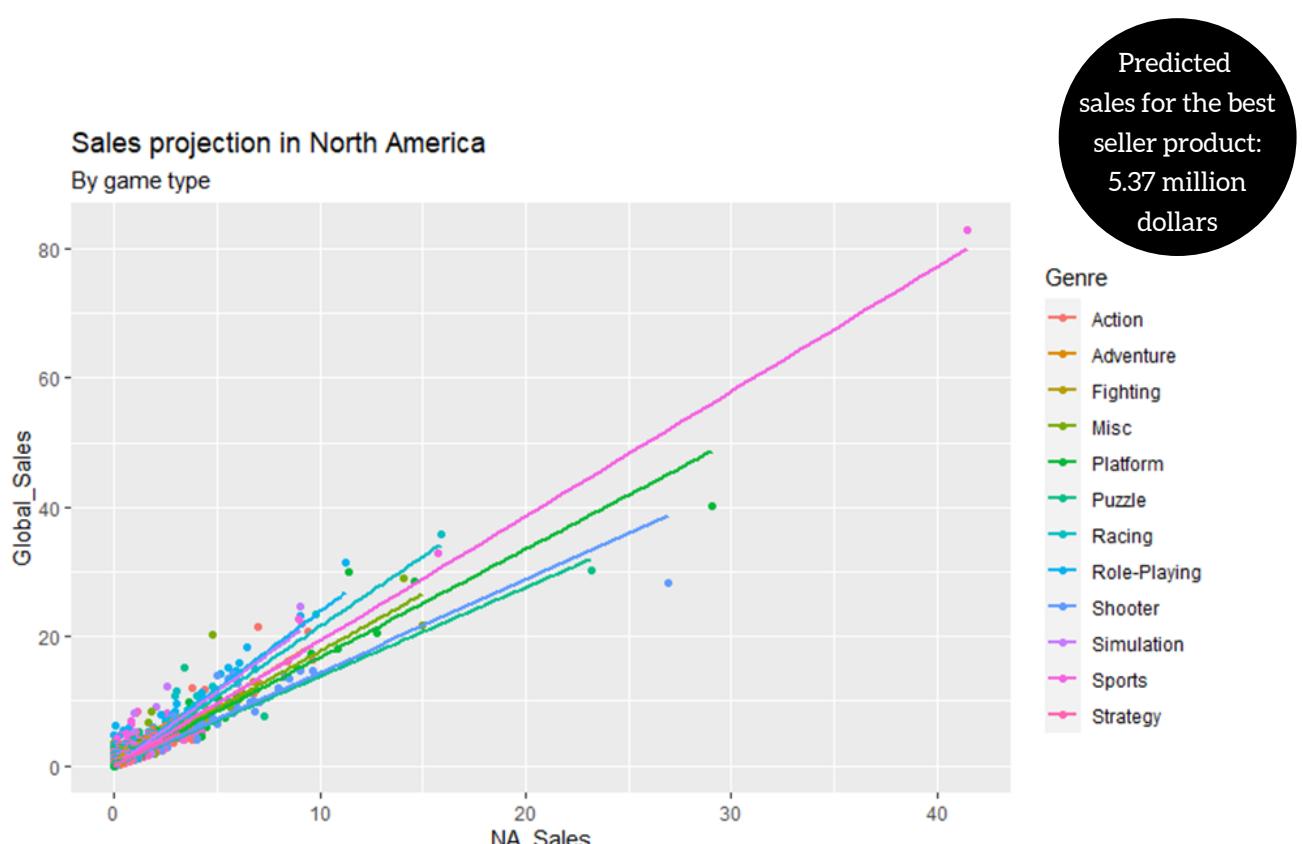
REGRESSION AND PREDICTION

Last but not least, is the Global sales prediction of games for next year based on the North America and Europe top ranked product.

I have analysed the data based on these top ranked product and built a single and multiple linear regression model comparing Global sales data, correlating data between the Global sales with North America and Europe sales. This top ranked product total sales was 82.74 million dollars in total since it year of release 2006 (15 years ago).

Assuming that the yearly sales was 5.51 in 2021 (82.74 divided by 15 years) and will be used the same figure to predict 2022, the predicted yearly sales for 2022 for North America will be 2.72, Europe 1.6 and Globally 5.37 million dollars for this particular top one ranked product, Nintendo Wii Sports 2006 edition.

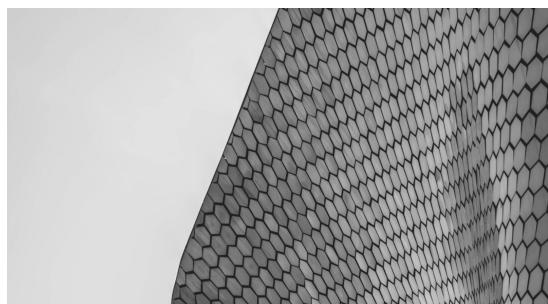
Finding the trend for each video game type, the Global sales in relation with North America Sales, can be seen from the regression lines in the graph below with trends correlation.



CONCLUSION & RECOMMENDATIONS



- A recommended price for a Lego set of 8000 pieces is in the range of \$ 782 and \$784 dollars, regardless the age group.
- The customer segment with highest number of reviews for all Turtle Games products is the 19 and 23 years old as well but less the 11, 13 and 16 years old.
- Future sales and marketing strategies for sets with higher numbers of pieces is advisable to target 29 years old customers, as they are the one that spent the most in this products.
- Based on the sentiment of the customers across all the products, some board games are not engaging and should be considered, paying special attention to the elf game which attract many negative reviews. Other than that the service and engagement of the games is positive.
- From the historic data on games sold in stores across Europe and North America, the predicted Global sales for the next financial year for the best seller product (Nintendo Sport Wii) is estimated to be \$5.37 Million.



Thank you for reading.

The end

TABLE OF CONTENTS

Section 5. Excel Statistical analysis

NA_Sales		EU_Sales		Global_Sales		Online_Sales	
Mean	0.26466743	Mean	0.146652006	Mean	0.537440656	Mean	0.126121219
Standard Error	0.006339072	Standard Error	0.003922523	Standard Error	0.012070085	Standard Error	0.003153004
Median	0.08	Median	0.02	Median	0.17	Median	0.03
Mode	0	Mode	0	Mode	0.02	Mode	0
Standard Deviation	0.816683029	Standard Deviation	0.505351231	Standard Deviation	1.555027936	Standard Deviation	0.406211657
Sample Variance	0.66697117	Sample Variance	0.255379867	Sample Variance	2.41811188	Sample Variance	0.16500791
Kurtosis	649.1302677	Kurtosis	756.0277958	Kurtosis	603.9323462	Kurtosis	219.9342395
Skewness	18.7996274	Skewness	18.8755351	Skewness	17.4006451	Skewness	11.94352427
Range	41.49	Range	29.02	Range	82.73	Range	12.24
Minimum	0	Minimum	0	Minimum	0.01	Minimum	-0.01
Maximum	41.49	Maximum	29.02	Maximum	82.74	Maximum	12.23
Sum	4392.95	Sum	2434.13	Sum	8920.44	Sum	2093.36
Count	16598	Count	16598	Count	16598	Count	16598



Attached files:

- 2 Jupyter Notebooks
- 2 R Notebooks
- 6 Excel files
- 11 PNG Files