

Course 3 Assignment Report

Corina Campian

Background

Turtle Games is a games manufacturer and retailer, looking to do a review of their pricing, determine the product segments, identify the customer sentiments, and predict sales.

Turtle Games manufacture and sell their own products, as well as source and sell products manufactured by other companies. Their product range includes Lego, as well as board games, video games and toys. They have a global customer base and have a business objective of improving overall sales performance.

Analytical Approach

To help Turtle Games answer their business questions, I have carried out data analysis in Python and R, using the datasets provided (lego.csv, game_reviews.csv, game_sales.csv). The methods used include:

- Descriptive statistics
- Single and multiple linear regression in Python
- Natural language processing
- Multi-linear regression in R
- Predictive modelling in R

Visualisation and Insights

Based on exploratory data analysis of the lego dataset we can see that:

- The dataset contains 12,261 observations and 7 variables
- All the data are numeric and there are no missing values
- The average age of players is 17 years old
- The average price is \$65 (minimum \$0 up to a maximum of \$367). The rest of the descriptive statistics are presented below (Fig 1)

	ages	list_price	num_reviews	piece_count	play_star_rating	review_difficulty	country
count	12261.00000	12261.000000	12261.000000	12261.000000	12261.000000	12261.000000	12261.000000
mean	16.68828	65.141998	14.603050	493.405921	3.709689	1.988826	10.015333
std	8.21868	91.980429	34.356847	825.364580	1.641130	1.787565	6.185450
min	0.00000	2.272400	0.000000	1.000000	0.000000	0.000000	0.000000
25%	11.00000	19.990000	1.000000	97.000000	3.600000	0.000000	4.000000
50%	19.00000	36.587800	4.000000	216.000000	4.400000	2.000000	10.000000
75%	23.00000	70.192200	11.000000	544.000000	4.700000	4.000000	15.000000
max	30.00000	1104.870000	367.000000	7541.000000	5.000000	5.000000	20.000000

Figure 1. Descriptive statistics of the lego dataset

In addition, it looks like there is a correlation between the count of lego pieces and the listed prices (Fig. 2 below)

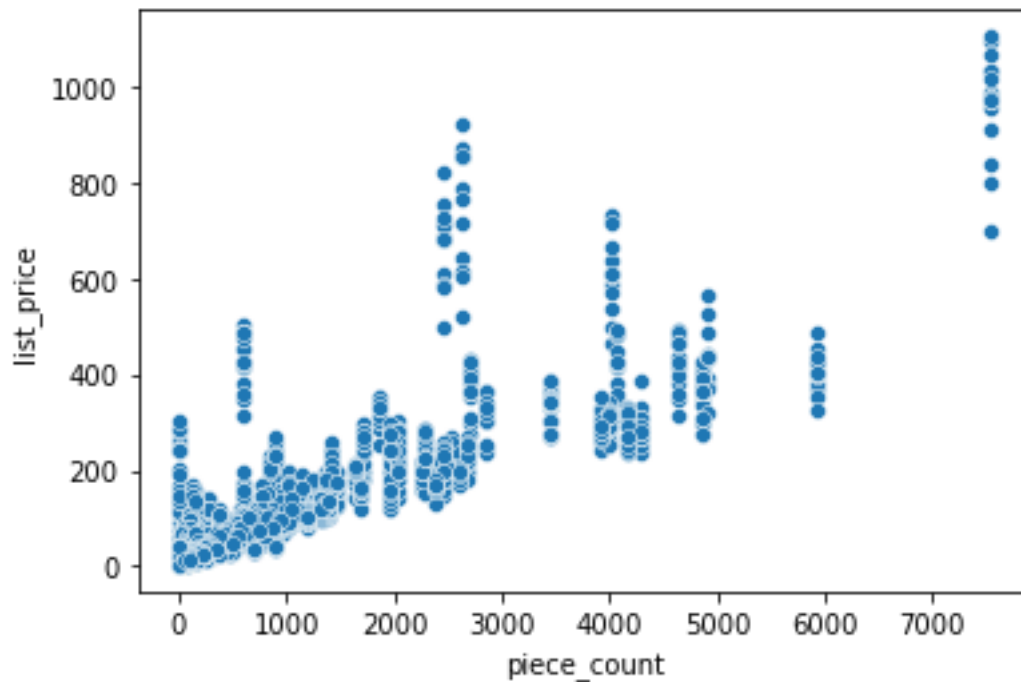


Figure 2. Correlation between count of pieces and price.

Patterns and Predictions

The following sections present the results from the data analysis answering the main business questions.

1. Predicting the prices to inform the sales performance.

The multiple linear regression showed that the predicted price for a Lego set with 8000 pieces for 30 year olds is \$793.

It also showed that when it comes to predictors of price changes, only the number of pieces is a statistically significant determinant of the change in price ($p < 0.05$), not age ($p > 0.05$).

2. Understand customer satisfaction with the products.

Generally the reviews received are positive with an average of 4.5 out of 5 rating. This is confirmed by the sentiment polarity histogram which is slightly skewed to the right (Fig 3 below).

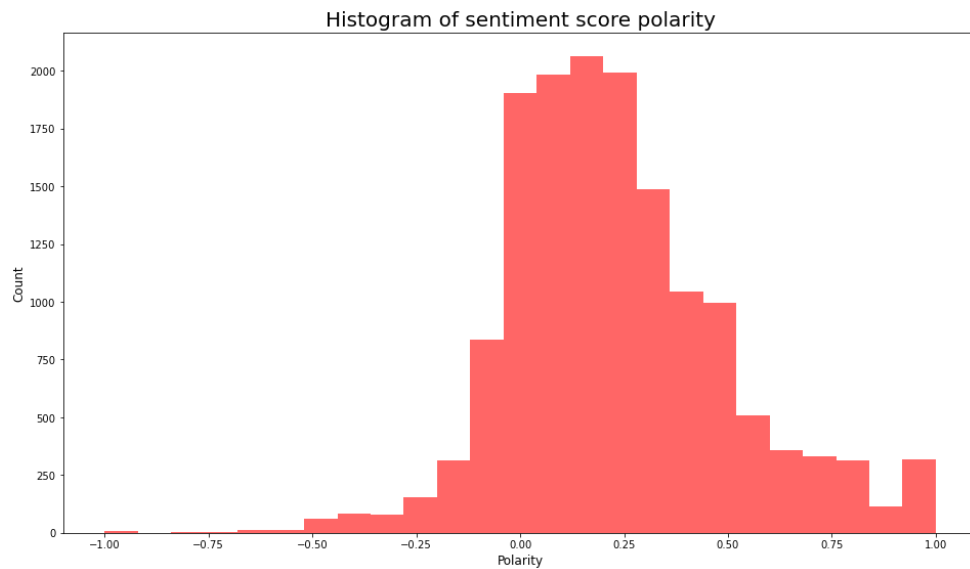


Figure 3 Histogram of sentiment score polarity of the review texts

Top 20 most positive reviews:

reviewText	summary	polarity
Came in perfect condition.	Five Stars	1
Absolutely great pictures even before coloring!	Five Stars	1
Great!	Five Stars	1
Awesome book	Five Stars	1
Awesome gift	Five Stars	1
Great product! Arrived on time.	Five Stars	1
Great buy!! My granddaughter loves it!!	Five Stars	1
Great!	Five Stars	1
Great resource for BHIS care coordinators!! Works well with kids and teens on what it says it does!!	Perfect!	1
Great Seller!!! Happy with my purchase!!! 5 starrrr	Great for kids	1
Excellent activity for teaching self-management skills!	Five Stars	1
Great game...I use it a lot!	Five Stars	1
Great therapy tool!	Five Stars	1
Perfect, just what I ordered!!	Five Stars	1
Wonderful product	Five Stars	1
Delightful product!	Five Stars	1
Great Easter gift for kids!	Great Easter gift for kids!	1
Wonderful for my grandson to learn the resurrection story.	Five Stars	1
These are great!	Cute letters	1
Perfect!	Aquire game	1

Top 20 most negative reviews:

reviewText	summary	polarity
BOOO UNLES YOU ARE PATIENT KNOW HOW TO MEASURE I DIDN'T 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

Kids did not like it. Thought it was boring.	Not so much fun	-1
some of the suggestions are disgusting	One Star	-1
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
Going to be a crazy Christmas at our house this year!!!!	Five Stars	-1
I do not under stand how you keep score or read the scoring. I i do not like that at all. I can never play score with anyone at all. I HATE THAT I CAN'T PLAY POINTS!!!!...	What i do not under stand	-1
Incomplete kit! Very disappointing!	INCOMPLETE KIT!	-0.975
Was the elf on the shelf but it didn't have the DVD I was very disappointed	Three Stars	-0.975
I haven't even taken it out of the box yet but it's 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.	-0.975
It is not waterproof and is dangerous!!	One Star	-0.9375
I hate the holidays bcuz of the Elf, he was disgusting, i hate him with my life, he doesnot leave the shelf alone.	I hate the holidays bcuz of the Elf	-0.8666667
Cliche and stupid. I should not drink and Amazon....	Hahaha. Ho Ho Ho.	-0.8
Just stupid.	One Star	-0.8
Family loves this game!!!	Well worth the purchase!	-0.78125
My son obsessed!!	Five Stars	-0.78125
The dog ate him. :(Five Stars	-0.75
Does not come with wire in the legs or arms...:(Needs wire in arms and legs	-0.75
I'd like to upload a photo of the condition of the game box...it looks like it's been used as a soccer ball! 2 corners of the box are smashed in and on is even ripped. How am I supposed to give this as a gift without it looking like I bought this on clearance? Very disappointed.	:(-0.7375
I like this product for my daughter. She is into the Bad Kitty book collection, so it was an added bonus.	Good Kitty	-0.7
Ordered for my sons birthday, opened it up today to play and the board is damaged. Before we even take it out of the box the game is already falling apart. Very disappointed...	Damaged board out of box	-0.6875

3. Predicting video games sales (in millions) for the next financial year

After running a correlation test between all the numerical variables, we can see that global sales are correlated firstly with both North American sales and then with Europe sales.

	Rank	NA_Sales	EU_Sales	Global_Sales
Rank	1.0000000	-0.4013622	-0.3791235	-0.4274066
NA_Sales	-0.4013622	1.0000000	0.7677267	0.9410474
EU_Sales	-0.3791235	0.7677267	1.0000000	0.9028358
Global_Sales	-0.4274066	0.9410474	0.9028358	1.0000000

Looking at North American sales specifically through a regression analysis, we can see that NA_Sales is a highly significant value ($p < 0$). It explains 88% of the variation in Global Sales.