Elizabeth Earl

DSC680-T301

Project 2 – Toy rankings

Business Problem:

Toys are constantly being made and the market has become a competitive one. Toys based on classic items such as stuffed animals have become rare and heavily sought out items. Taking into consideration how quickly an item goes from average to a ‘collectible’ I have attempted to discover whether the toys produced based on hype are worth the money or just the label.

To analyse this idea of toys being worth their price or just a hype, it is necessary to compare pricing and rating of a toy in a specific category. It is expected a popular item will be sold at a higher price because of the demand for said item.

Background/History:

Social media, such as YouTube and TikTok have created a new competitive market for items to gain traction and ultimately have become what determine what’s “in” and what’s not. Today’s society relies heavily on outsider opinion to determine their next move; if someone with a million followers (whether it be on any social platform) tells you to buy some toy, people are more likely to buy said toy versus the competitors. It is hard to determine with such a constant ‘in your face’ market what choices you should make when it comes to buying toys. The purpose of this project will review top toys, potentially determine why such toys were purchased and if ultimately, they were the right choice.

Data Explanation (Data Prep/Data Dictionary/etc):

Dataset – Top toys as sold on Amazon:

- price
    - Used to determine whether a toy is worth the monetary value
- average_review_rating
    - Rating will help define 'top' toys
    - Average rating will vary based on consumer reviews
- description
    - Helps outline how a company wants a toy to be perceived by the consumers
- customer_reviews
    - Values used to create sentiment values
    - Allows for a better understanding if a toy was worth the purchased based on what is said

Methods:

The main method of analysis was done by sentiment modelling, this involved TextBlob, correlation matrices, and visualization graphs to compare data values mentioned in previous section.

Analysis:

Using TextBlob to determine sentiment and polarity variables allowed for a clear indication of reviews to potentially be used for predictions methods. This was implemented for ease of data reading and creation of visualizations. A confusion matrix of such calculations is seen in supporting illustration sections under Figure 1. Figure 1 as expected show that the top toys were more likely to be toys receiving a

4-5 (out of 5) rating. It is interesting to see a handful of toys with rating from 2-3 which raise questions that will be discussed in the Implementation Plan section below. Additionally, I created another matrix, this time being a correlation matrix, to figure out which variables were crucial to what made a toy fall high on the rating scale. Results are seen in Figure 2.

After having initial results, I created sub datasets which included key words in costumer reviews. The reasoning behind this was because people are typically very honest in their intentions when they write a review versus what a company uses to describe an item. The key words used in the sub datasets are as follow:

- plush
    - Initially I was very interested in how a simple toy like a plush would compete in the market
- soft
    - Goes in hand with the idea of how plush toys would rank in top toys
- squish
    - Introducing a top toy like a squishmallow (which is essentially a glorified plush) emphasized how strong a demand there is for a simple toy
- youtube
    - As mentioned, YouTube can be a strong indication of how popular a toy will become
- video
    - Alongside YouTube are videos included that can be found on other platforms

- collectible
    - This variable comes from the idea that if a consumer believes a toy is a collectible, they are more likely to purchase said item

With this sub datasets it was interesting to then make comparisons on their pricing versus a total price average of all top tops as well as how they do on a ranking scale. The results are shown in Figure 3a and 3b.

In my findings I discovered some ‘top’ items were receiving rating lower than 3.0 (see figure 4). After further investigation I found the lower ranking items were produced by companies that were third-party companies (see figure 4a). The assumption as to why these companies were a third party was based on the idea that third party sellers create names almost identical to real sellers to trick consumers and this can clearly be seen with a company name Wow Wee versus the official company: WowWee (no spacing).

Conclusion:

Based on results shown in figure 3a and figure 3b I noticed that when I created sub datasets using popular keywords (that I figured would be the reason why toys were popular) their ratings were lower than an overall average of rating on the completer dataset. Using this analysis, I assume that people are buying toys based on popular trends, but they are not what they are made out to be. This conclusion comes from this idea that consumers are mentioning these words in their reviews but are not overly content with the item (versus an overall content rating of all top toys). This shows that people are simply buying toys because there is a sort of hype attached to them.

The conclusion that popular toys were purchased simply because there is a sort of hype around them, whether it be collector's items or due to social media pressure (influencers) was made based on rating, but this raised a question of whether there was a connection between toys being forced into this market. I had an initial assumption that toys being advertised on social media were simply pushed because of their pricing; this created an idea that toys were expensive but not worth it in quality. After calculations I found that on average the top toys were roughly £3 cheaper than toys that were had key words assumed to be pushed by influencers. All in all the consensus of my findings supported the idea that popular toys are usually expensive and turn out to be a waste of money (when considering rank vs price)

Assumptions:

Because social media is such a strong presence and influencers are creating a new consumer's market it is assumed that sometimes the items influencers promote are not all they claim to be. Companies use influencer to gain revenue and it is done in a way that is not always ethical. Toys that are not worth their price are being pushed to create a positive revenue for companies; from a consumer standpoint it is disappointing but from a company standpoint it is a well thought out tactic. Consumers are being tricked into buying a 'cool' and expensive toy because they believe it is popular or hard to come by and majority of the time, they are disappointed.

Limitations:

My personal limitations for this project came from my location. When searching for data I was hindered by my location and only shown results of sites country locked to England. This created limitations with being able to access bigger

datasets that include USA trends. I did, attempt to access other datasets that included broader locations, but websites were limiting such search. It was very difficult to try and bypass location locked content.

Challenges:

Being limited based on my location was one of the biggest challenges I faced. It was difficult to even begin to make assumptions as it raises a bigger issue of content available per country. If datasets are limited by country, I can assume that influencers and toy commercials are limited to per country audience or perhaps altered. This challenge is something I simply had to move past and assume that is my results were prevalent in a smaller scale (UK based) then similar results would apply to a larger scale (USA and perhaps the rest of the world). Additionally, it was a challenge, perhaps personal, to understand what toys were considered popular in UK versus USA.

Future Uses/Additional Applications:

Going hand in hand with the limitations and challenges, the future for this project can involve other country's toy sales. It would be very interesting to compare country top toys versus the other and see how location can impact who is the top toy company. Alongside this it will also be interesting to see pricing methods in top toys. In this project prices were in British Pounds which with the current exchange rate is less then 1 USD. Comparing the pricing and ranking of toys based on country will be a potential next step in the project.

Implementation Plan:

Q&A:

1. Which manufacturer/company has most top toys?

It took around 930+ results from the dataset to get companies that had less then 5.0 rating (on a 0.0-5.0) scale. There is a good variety of top companies with perfect rating and thus no clear one company.

2. What is the country of origin of said companies?

   Because there is no specific country of origin in the given dataset, this question was too difficult to analyse. A cross reference against company origin of over 900+ companies would have to be done.

3. Are influencers (by name) strictly mentioned in any reviews?

   Using Ranker as a reference for top toy YouTube influencers I cross reference my dataset and found no mentions of any influencers. It was quite interesting to see but I assumed it was due in part with the top influencers are mostly based in the USA whereas the dataset I am using is based in the UK.

4. Who is the top children's influencer?

   Using the same website as reference, I found that Playtime with Ayden was the top kid's influencer, yet he was not mentioned in any of the customer reviews from the dataset.

5. Do the results of influencers in reviews correlate to the top influencer?

   As stated previously there was no mention of kid youtubers in reviews. I could not directly make the correlation with the youtuber but there is still an assumption that the actual products they push are the top toys.

6. What is the cheapest and most expensive top toy?

   According to the dataset the most expensive toy is a Thor collectible figure priced at: £2439.0 while the cheapest was a foil banner priced at: £0.01

7. What is the correlation between price and ranking?

I did notice a correlation between price and rank when it came to my sub dataset. The more expensive an item was the lower it ranked which emphasized my assumption that expensive toys were not worth the hype.

8. Are the previous assumptions/conclusions accurate when price and rank are directly compared?

   Yes, see question 7.

   Considering the lowest ranking toy(s):

9. What is a common trait between the low-ranking top toys as per company descriptions and customer reviews?

   The low-ranking toys were promoted as what they were, simple items (ex. foil banner) and their reviews were as unexpectedly upset over the product. Because the items are described as simple items it is assumed they are of low quality and thus complains are simply unreasonable.

10. What company is more likely to produce low ranking toys?

    There was no repetitive company in the lowest ranking toys thus the conclusion is made that no specific company is more likely than the rest to product a low ranked toy.

11. Is there a type (i.e., car, doll) toy that is more likely to rank low?

    It was not a surprise, but novelty toys are the toy that was mostly seen as low ranking. Novelty toys are often not meant to last more than an afternoon, so their quality is low thus receiving low ratings.

12. Where do toys not meant to be played with (i.e., collectibles) fall in the ranking scale?

    Out of 10,000 total values in the dataset, the highest-ranking collector item was 1,590/10,000

Supporting illustrations:

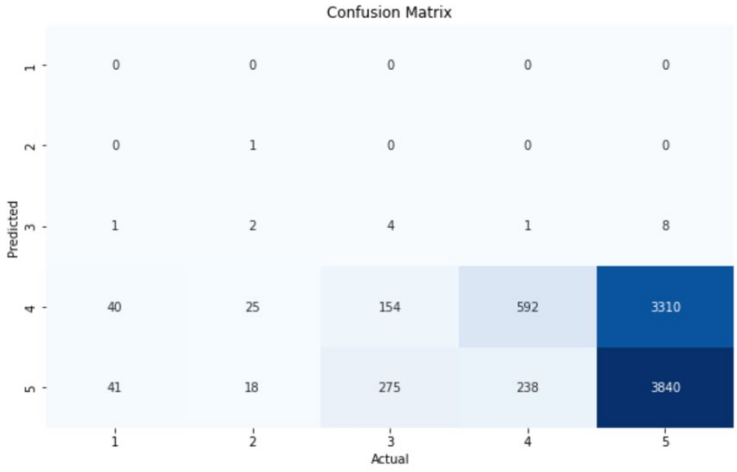

Figure 1:

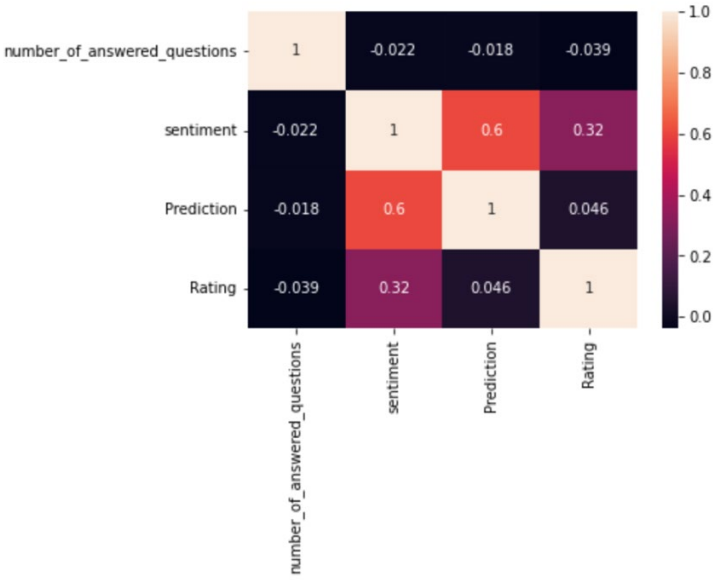

Figure 2:

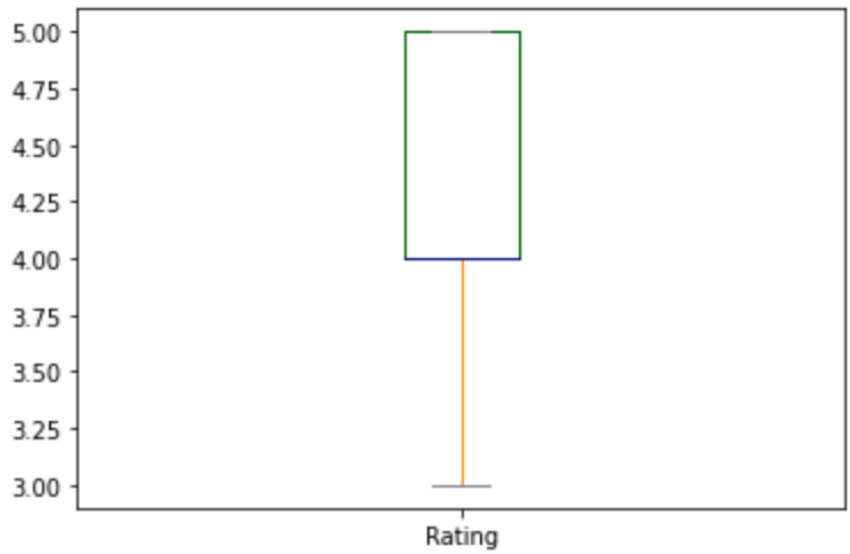

Figure 3a:

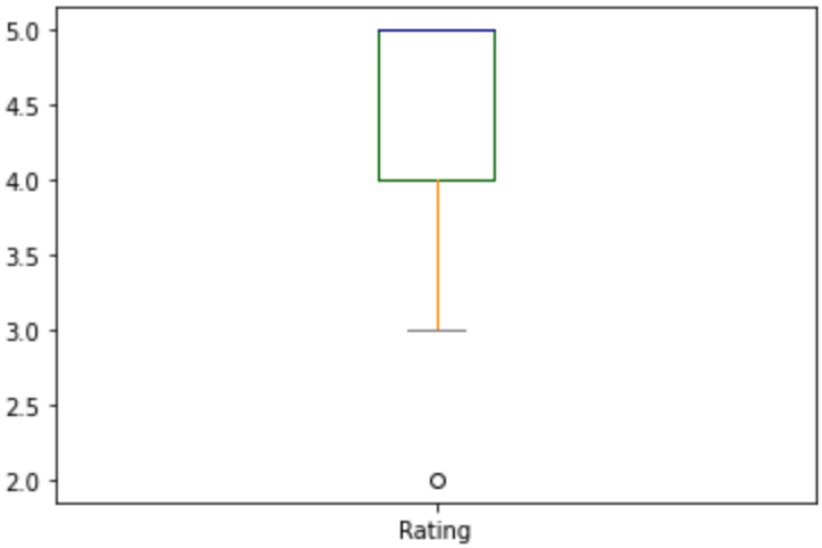

Figure 3b:

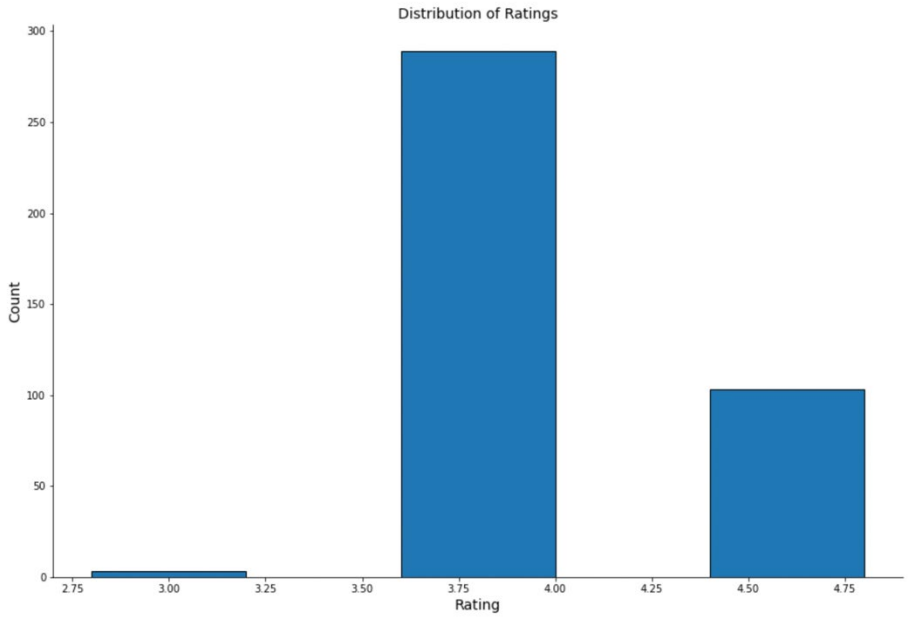

Figure 4:

	manufacturer	Rating
1847	Super Mario	3.000000
1966	The Original Glowstars	3.000000
2218	Wow Wee	3.000000
159	BINGWAN	3.000000
1865	TAYO	3.000000
2268	ccf	3.000000
973	Inflatable Santa	3.000000
368	Cavendish Trading	3.000000
1025	John Adams	3.750000
1158	Lite Brix Moonlight Monsters	4.000000

Figure 4a:

Appendix:

Exploring Toy Products on Amazon. (2022). Retrieved 25 April 2022, from https://www.kaggle.com/code/residentmario/exploring-toy-products-on-amazon/data

Kluyver, T., Ragan-Kelley, B., Fernando P'erez, Granger, B., Bussonnier, M., Frederic, J., Willing, C. (2016). Jupyter Notebooks – a publishing format for reproducible computational workflows. In F. Loizides & B. Schmidt (Eds.), *Positioning and Power in Academic Publishing: Players, Agents and Agendas* (pp. 87–90).

The Best Toy Reviewers On YouTube. (2022). Retrieved 2 May 2022, from https://www.ranker.com/list/best-toys-youtubers/youtuber