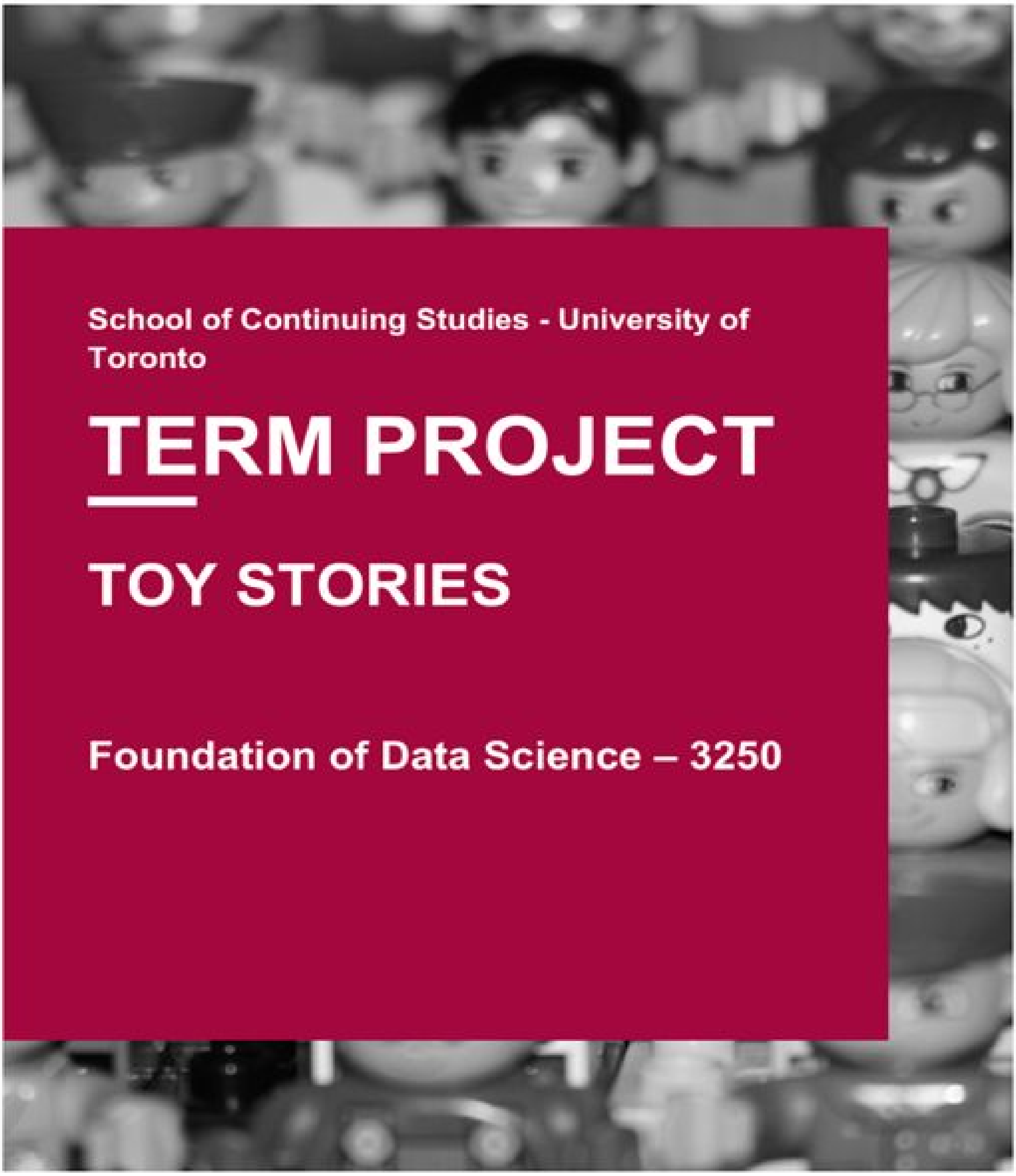


School of Continuing Studies - University of Toronto

TERM PROJECT

TOY STORIES

Foundation of Data Science – 3250

Group 4: Akhat Maneshev, Alexandra Storozheva, Bibizeinab Torabian, Gurpreet Singh, Kushagra Dixit, Rithika Shaw

University of Toronto

August 08, 2019

Contents

Introduction

Amazon has seen amazing growth as a company and as a business model over the last decade. Not only does it present an interesting showcase as a company, it more importantly possesses and also constantly processes enormous amounts of data, while simultaneously opening huge opportunities for companies joining in to easily access a wide range of markets. Selling on Amazon seems enticing, but the most common question arising would be as to What kind of new product can be sold on Amazon for it to be successful?

For analysis, Amazon is openly offering Big Data datasets to the public. As a research group, the dataset dedicated to Toy Reviews left over a span of years since 1998 to 2015 with over 5 million reviews was decided to be explored in hopes to find clues for answering what kind of toy product a client startup could be selling on Amazon to be successful as a seller.

The Toy Dataset was carefully chosen because toys constitute a category of products appealing to anyone. It is reasonable to assume that the creation of a new toy is primarily restricted by imagination limits. If it is a valid assumption, it would fairly possible to come up with a toy product that can be new to the market, and still can become successful with high potential of building a recognizable brand.

This report provides an analysis of Amazon Toy Reviews Dataset for toy products on Amazon marketplace from 1998 to 2015.

Data Overview

Link to the dataset is listed in Sources section. This dataset consists of reviews of toy products sold from 1998 to 2015 on Amazon.com in the U.S. market. The dataset contains 4859607 rows and 15 columns. Columns and corresponding NaN values are as follows:

customer_id	0
review_id	0
product_id	0
product_parent	0
product_title	0
star_rating	1
helpful_votes	1
total_votes	1
vine	1
verified_purchase	1
review_headline	0
review_body	0
review_date	24
split_title	0
category	0
title_count	0

For the purpose of data integrity, rows that contain NaN values were dropped because their amount is insignificant to the rest of the data. Certain data cleaning was also performed, specifically converting the columns product_title, review_headline and review_body to strings because they had mixed data types that could create issues with running functions consistently. The columns product_category and marketplace was dropped because their values were identical for every row, and corresponded to “Toy” and “US”.

Subsequently, several new columns were engineered. The “category” column was engineered using an algorithm for matching product titles to categories adopted by Amazon, based on keywords found in product titles. Approximately 22% of the dataset could not be categorized as the titles and descriptions were not descriptive enough to indicate the category they belong to, and therefore, the uncategorized rows were dropped to be able to run analysis on categories clearly. Furthermore, a column with estimation of potential expenses associated with many products within each category was engineered for dataset convenience. This estimation was done based on how large and heavy a typical category product is, whether it is a mechanism or whether it uses any electronics.

The algorithms used for engineering of new categories always have a potential of further refinement, but the data analyzed does help with answering important questions for potential startups.

Data Preparation and Validation

Before proceeding any further, the data was validated to make sure that it made sense, to add a degree of confidence to the data findings.

Data Cleansing And Steps Taken to Procure Better Data Quality

The rich “customer reviews data” made available by Amazon was cleaned up for several attributes as follows:

1. Records with malformed dates were removed as the number was really small
2. Non-verified purchases were not considered in the data set for analysis in order to maintain legitimacy of reviews
3. Product_Title, Review_Headline, Review_Body were converted into string types
4. Product_Category and Marketplace columns were dropped as they just contained 1 irrelevant data point as Toy and US respectively

Tools Used
Python was used to perform data analysis using libraries like pandas and numpy. Matplotlib and seaborn were employed for generating graphs on Jupyter.

Data Validation

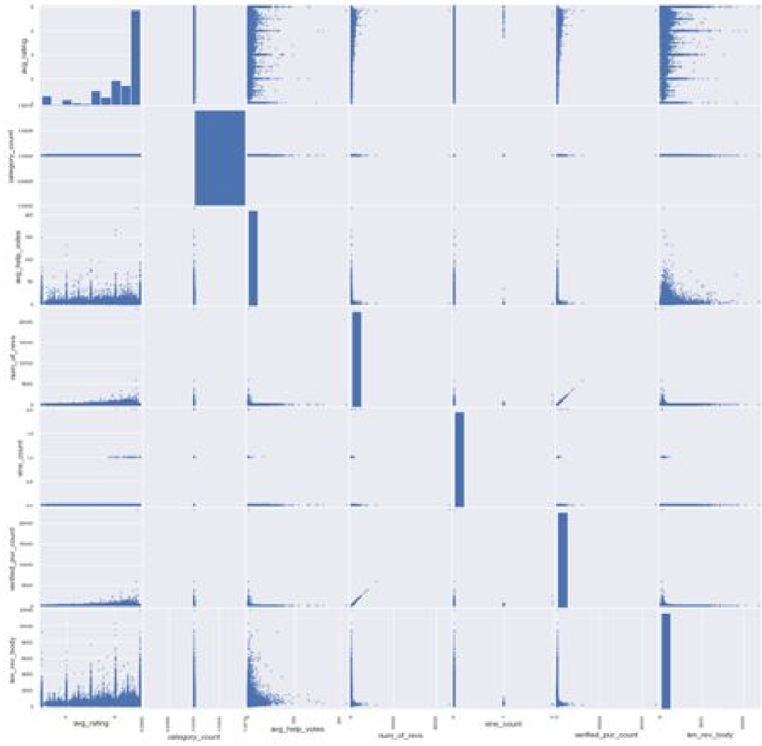
Post data cleansing, data set made available by Amazon was analyzed to check if it made general sense. The following attributes were engineered to build better statistical understanding of underlying data and analyze correlations between these new attributes to identify patterns or trends:

- Average Rating, Toy Category Counts, Average helpfulness of reviews, Number of reviews, Vine counts, Verified purchase count, Length of review body

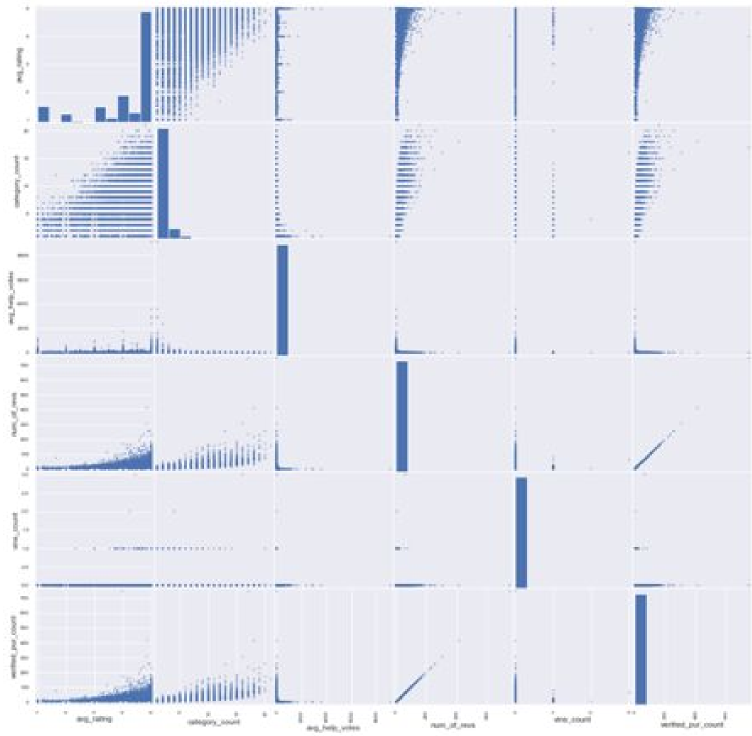
The data shows following interesting observations which also make general sense, and hence validating that data provided is legitimate:

1. More reviews are found when ratings are higher or vice-versa
2. More verified purchases when ratings are higher
3. Length of review body was inversely proportional to the average helpfulness of a review
4. Number of reviews is highly correlated with verified purchases
5. People tend to leave a greater number of smaller review comments
6. Vine count more than zero attracts higher rating
7. Customers shopping in more categories provided higher ratings.
8. Average rating is concentrated towards higher ends suggesting good Amazon performance

The graph below shows correlation between all engineered attributes when viewed/grouped by each product.



The same attributes when viewed/grouped from customer's perspective, show a different set of patterns/trends.



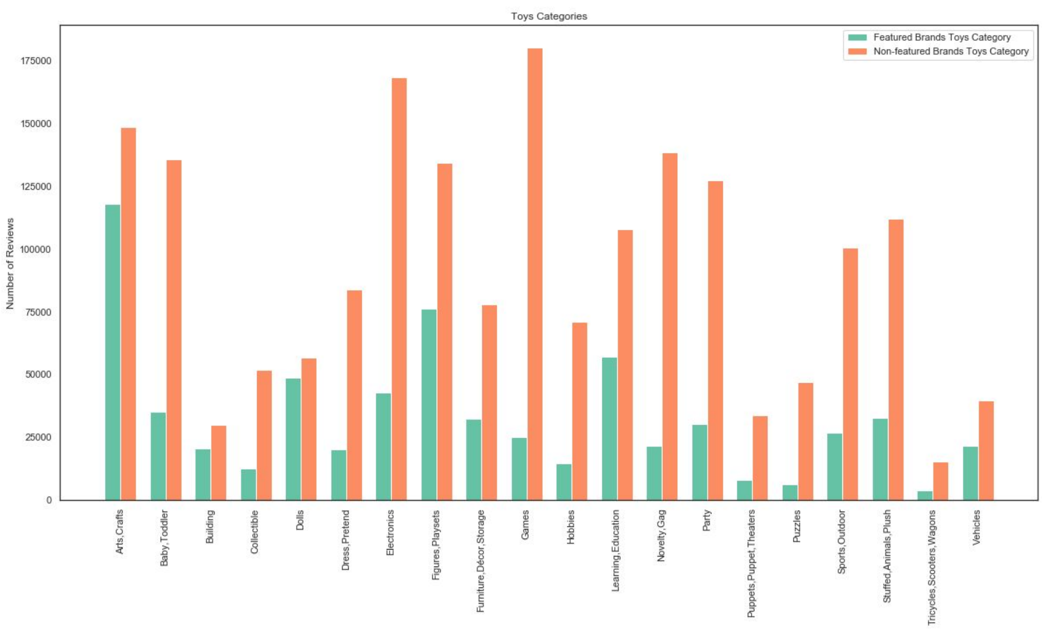
Data validation techniques are proving that the dataset makes sense and meets expectations for a common sense customer behavior. Data quality is high and solid.

Perspectives For Startups vs Featured Brands

As the data was found to be making traditional sense, this section would discuss if it is potentially reasonable for a startup to try entering the Amazon marketplace, or it will have low chances of profitability because people could be concentrated on buying from recognized brands, leaving low opportunities for startup products to grow. This was answered by comparing featured brands to regular brands.

Amazon manually selects recognizable brands with high level of sales, and designates such brands as “featured brands” on a subsection of the Amazon website. The rest of the brands do not enjoy special features, and a new startup, which brand is not widely recognized, will be among the non-featured brands.

Amazon and third party sources were used (please refer to the Sources section) to create a list of brands that were selected and featured by Amazon between 1998 and 2015. Furthermore, the popularity of featured brands vs non-featured brands was compared based on the number of reviews that each subset received. Then, the total number of reviews for products of non-featured brands were computed and compared with the number of reviews for featured brands, and the results were then split by toy categories for better comprehension.



The results found through this analysis showed that although featured brands have a high proportion of reviews on the market, which is likely corresponding to overall number of purchases, overall reviews for non-featured brands still constitute a major proportion of reviews. The proportions of reviews for featured and non-featured brands also vary by category, which can be counted in for increasing chances to compete with featured brands successfully.

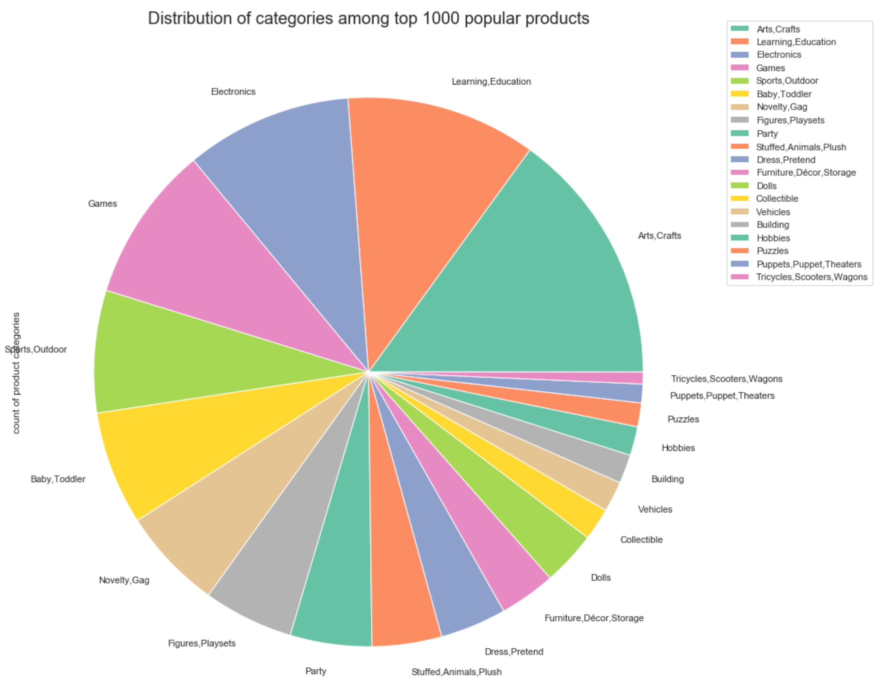
Finding Favourable Categories

This leads to the fact that there is potential for a startup to enter the Amazon market. Next question is “What categories of toys have higher favourability for a startup”.

4 factors were taken into account:

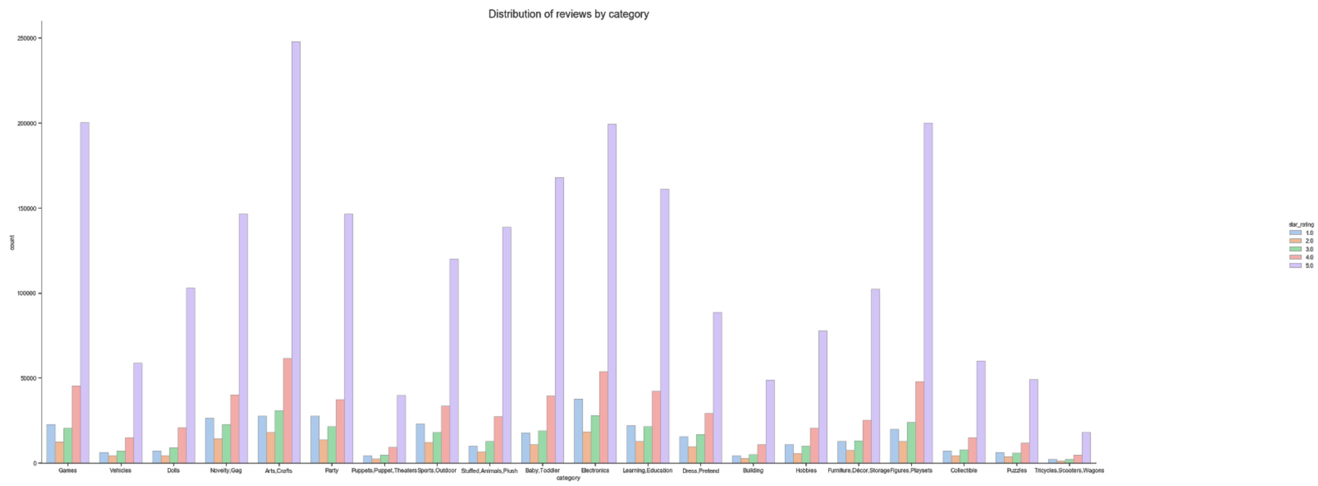
- demand,
- competition,
- likeliness of having negative reviews,
- likeliness of having positive reviews

The results were then split into sub-categories based on previously proposed investment expenses (high, medium and low). The demand evaluation was done through counting all categories of toys that belonged to top 1000 most popular products.

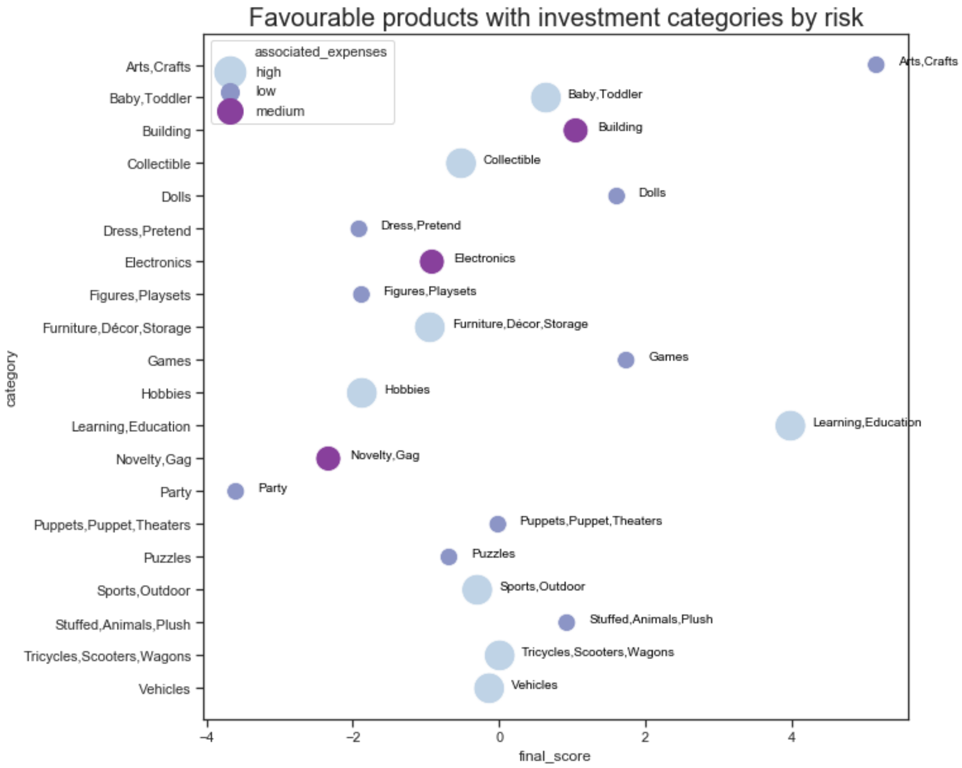


Competition was evaluated by counting proportions of all categories of all products in the dataset.

For the analysis of reviews, all reviews were analysed by category for the distribution of ratings. The following graph represents a general overview of the findings:



Negative reviews of 1 and 2 and positive reviews of 4 and 5 were calculated separately and split by categories. Weights were assigned to every factor and used in an algorithm for getting the final score for each product category. The results obtained can be visualized on the scatter plot below:



As seen above, the categories are listed in alphabetical order. Favorability of each category is measured by being closer to the right side of the graph, e.g. *Arts,Crafts* is the most favourable category, while *Party* is the least favourable. Associated expenses related to each category are represented by the point size and point color.

The estimation is enabling businesses with a tool for choosing a category of product to invest in, based on the business' expertise and resources.

Finding Features of Products

Once a business decision with regards to choosing the right category for a particular startup is made, attention can be taken to the features of the products. Specifically, a startup may want to answer such questions as: “Do we need to invest more in product quality or it is not entirely relevant”, or “Is logistics an important success factor”? The analysis for these questions can be performed best through carefully looking at the review body of poorly rated products.

The analysis of identifying important product features was approached by reading the star rating of the review dataset. The star rating is a simplified metric to determine the success or failure of the product with the respective target buyer. This was followed by an analysis of the review body itself to see what specific problems were encountered by the products causing the poor rating and the product’s subsequent failure.

Below is a summary table of categories of toys split by proportions of ratings.

<table>
<tr><th>Star rating</th><th>associated_expenses</th><th>Category of Toys</th></tr>
<tr><td rowspan="7">High 4 stars or more rating (82-87%)</td><td rowspan="3">Low</td><td>Dolls</td></tr>
<tr><td>Stuffed, Animals, Plush</td></tr>
<tr><td>Figures, Playsets</td></tr>
<tr><td rowspan="2">Medium</td><td>Building toys</td></tr>
<tr><td>Electronics</td></tr>
<tr><td rowspan="2">High</td><td>Baby Toddler</td></tr>
<tr><td>Play vehicles</td></tr>
<tr><td rowspan="8">High 2 stars or less rating (14-16%))</td><td rowspan="3">Low</td><td>Sports, Outdoor</td></tr>
<tr><td>Furniture, Décor, Storage</td></tr>
<tr><td>Learning, Education</td></tr>
<tr><td rowspan="2">Medium</td><td>Novelty, Gag</td></tr>
<tr><td>Electronics</td></tr>
<tr><td rowspan="3">High</td><td>Party Supplies</td></tr>
<tr><td>Dress, Pretend</td></tr>
<tr><td>Puzzles</td></tr>
</table>

High and medium level investment categories received a higher percentage of less than 2 star reviews probably because in a competitive marketplace like Amazon where cost determines sale, it can be hard to estimate the quality of a product when a buyer only has pictures and written specifications as a reference point. Categories like Electronics, Sports and Outdoor items need to perform as required by the user and also be very durable at a competitive cost. Many of these “toy” products also cater to adults as opposed to exclusively catering to small children.

This is in contrast to the highly rated categories with a very high percentage of more than 4 star rated reviews. These categories have a tendency to have mostly low associated expenses. This is probably because it can be relatively easy and cheaper to produce dolls, stuffed animals, action figures and playsets. The appearance that these products have on the website is potentially what you receive in person, and small children tend to be exclusively using these products as opposed to more poorly rated reviews received by categories that also cater to adults.

Electronics seems to polarize the buyers as it peaks in both the lists. UDI U818A 2.4GHz 4 CH 6 Axis Gyro RC Quadcopter with Camera RTF Mode 2 was a very interesting popular product as it was either loved by reviewers or was returned due to issues with its batteries, charger and poor flying after the 1'st flight. This is congruent to our previous findings as Electronics can either perform up to the buyers expectations with immediate effects causing a prompt positive review or fail due to technical/other issues and have the buyer return the product resulting in a negative review.

As an extension, the most common keywords in the review body of the most frequently 2 or less star rated products of each poorly rated category can be seen below:

a. **HIGH**
    i. Joissu Flingshot Slingshot Flying Screaming Monkey
    monkey, broke, first time, scream didn't work, quality
    ii. 1 Pack of Shocking Gum, Funny Shock Gag Random Color
    didn't work, didn't even get one shock, broke, gum bad, won't work
    iii. Rainbow Loom Crafting Kit Includes Loom, Metal Hook, Mini Rainbow Loom, 600 Rubber Bands + 24 Clips
    Not Real Product, Counterfeit, Fake, Not Original Broke
b. **MEDIUM**
    i. Money Maze Bank by Dragonpad
    Broke, Plastic, Waste, Gift, Christmas, Arrive Late
    ii. UDI U818A 2.4ghz 4 CH 6 Axis Gyro RC Quadcopter with Camera RTF Mode 2
    Battery, Charger, Drone, Fly First Only Good
c. **LOW**
    i. Gazillion Bubble Hurricane Machine
    use worked, only one time, topped, batteries, motor, broke
    ii. Disney Frozen Enchanting Elsa Dress
    dress material, itchy, scratchy, sleeves, cheap quality
    iii. 4M Glow 3D Solar System
    stars, don't glow, adhesive, light off, stick, fell, waste

As seen above, the most common issues encountered with these worst rated products in each of the expense category that was found to be statistically relevant were mostly concerned with the poor quality of product. The buyer was unable to judge the quality based on a picture and the vendors description and therefore bought a product that was far inferior compared to the buyer's expectations.

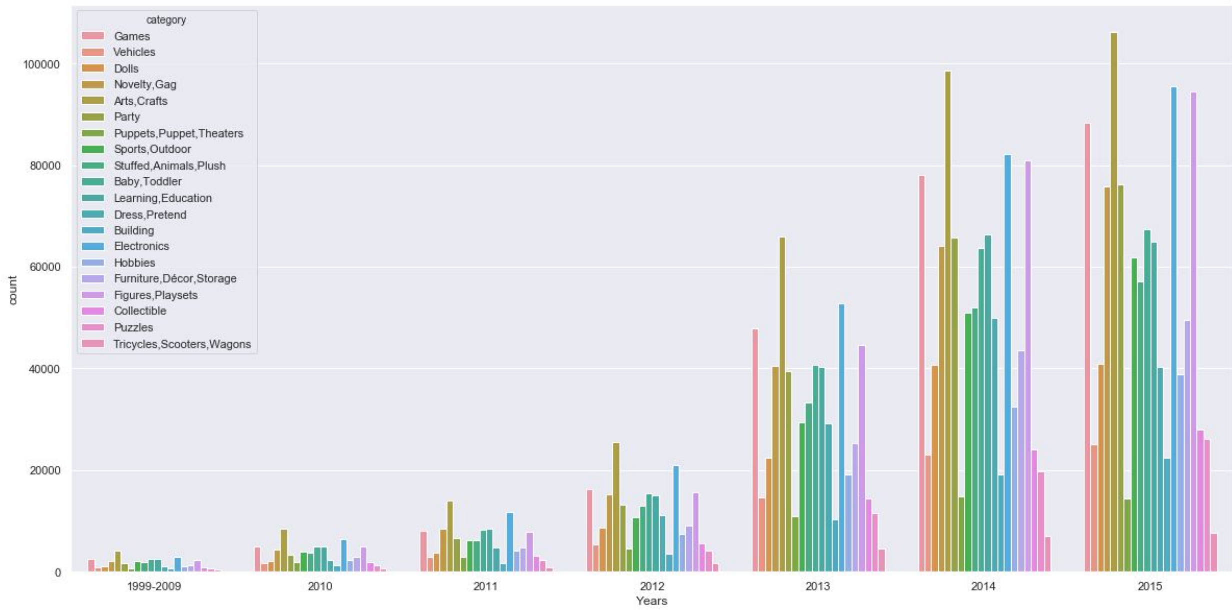
Exploring Product Launch Time

Once the startup understands what features the product is going to have and gets done with the manufacturing matters, question shifts to the marketing. General market overview seemed very promising as every category has been experiencing growth over time.

There are two important trends that appear from year-to-year (YOY) and month-to-month (MOM) data analysis.

YOY Analysis

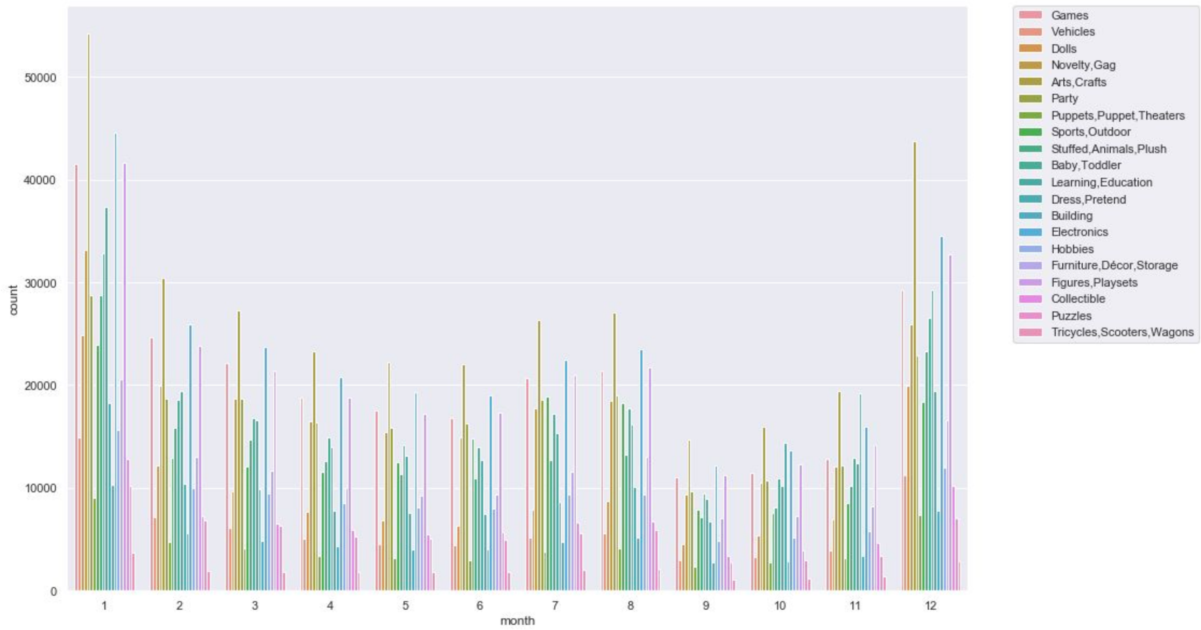
From the year over year analysis, the results demonstrated that the popularity of toy categories Games, Arts & Crafts, Electronics and Figures & Playsets is increasing YOY. With the growing popularity of Amazon and the increase of YOY sales, the popularity of toy categories remains the same.



MOM Analysis

The outcome of MOM analysis is that the most popular months are January (1), December (12). During November's (11) Black Friday just a slight increase is observed.

Seasonality is outlined on the graph below:



Months from September to November were seen to be getting the least number of reviews, which can be corresponding to the number of sold products over these months. This case is of special interest due to the Black Friday sales events, this may indicate that toys are not a popular category for Black Friday sales. This would show that toys are not particularly influenced by the Black Friday sales event as it was primarily introduced as a boost for low season months.

Some outliers were also recognized, specifically toy products related to clothing that enjoy a spike of sales during November. This can be attributed to the celebration of Halloween.

From these findings, a startup can make appropriate decisions for the best time for a product launch. A basic recommendation could be over launching in the spring season as it is preferable over launching in the fall season due to higher sales and review numbers, but the product type also needs to be carefully taken into account.

Conclusion

Amazon is a revolutionary marketplace offering huge opportunities for new products to compete and make a fortune. Amazon has made actual customer reviews data publicly available which opens up great opportunities to analyze data and trends. Focusing just on Toys review data, interesting trends like Amazon's overall review ratings inclines towards higher ratings, which is an excellent indicator of Amazon's performance in the market and solid quality of service were revealed. The data also indicated that the marketplace offers plenty of opportunities to launch non-featured brands that are not yet popular in the market, suggesting that new products can compete with existing popular brands and have high chances of being successful and make money.

Considering diverse set of market parameters like high-investment to low investment Toys, demand, and competition. It can be said that broad categories like Arts & Crafts, Learning & Education, Board Games and Dolls seem to have the highest likelihood of success.

Another interesting observation is that popularity of Toys across all categories seems to be really high during the months of Dec and Jan every year; apparently it is even bigger than Black Friday sale events. Analyzing people's reviews, both negative and positive, it becomes evident that customers truly care for the quality of the product and do not shy away from providing a business with their opinion about the product.

Overall the performed analysis of the data suggests that Amazon offers an open, amazingly competitive marketplace with great opportunities for new products and companies, including startups, to compete with existing brands and make a profit.

Sources

- Amazon dataset:
https://s3.amazonaws.com/amazon-reviews-pds/tsv/amazon_reviews_us_Toys_v1_00.tsv.gz

- Amazon data set metadata

https://s3.amazonaws.com/amazon-reviews-pds/tsv/index.txt

- Header Image Source:
https://pixabay.com/photos/grandstand-toys-males-child-330930/

- Featured toy brands on Amazon (web pages starting from here):
https://www.amazon.com/gp/search/other/ref=sr_in_-2_Z?rh=i%3Atoys-and-games%2Cn%3A165793011&bbn=165793011&pickerToList=lbr_brands_browse-bin&ie=UTF8&qid=1564854325

- Websites used for getting extra information about featured brands:
https://brandirectory.com/rankings/toys-25-2015
https://www.insider.com/popular-christmas-toys-2016-12#2010-ipad-26
https://www.pocket-lint.com/parenting/news/142866-the-most-popular-christmas-toys-and-tech-from-over-the-last-40-years

- Toy categories on Amazon:
https://www.amazon.com/s/ref=lp_166316011_ex_n_1?rh=n%3A165793011&bbn=165793011&ie=UTF8&qid=1564861988