Name: Gabriel Valenzuela  
Date: 2/26/2020  
Title: Amazon Toy Marketplace  
  
**Section 1 – Week 9**

Introduction

When it comes to Amazon, they are currently now the epicenter of purchasing items from anywhere and everywhere. Retail stories are becoming less popular and the ones that are still holding a spot in the competition are now looking towards online transaction and deliveries to compete with Amazon. For me, I was a part of the generation that went to Toys R Us for all the toys they wanted and GameStop for all the video games that I had a huge interest in at the time. Now, these stores are on the decline or all gone, and who has replaced them? Amazon. Hence, I want to see what the current state of the toy market is in how it operates when everything is practically online or on a screen.

Research questions

What are the top toy manufacturers on Amazon?

What is the typical price of the more popular toys?

What is the number of available of toys compared to the price of them?

What is the organization of the toys since it is no longer the primary concern?

How customer reviews played in the number of available items?

Approach

In addressing my problem statement, I plan to approach the data set in determining the major players in the online toy market through Amazon while exploring how reviews through the titan online marketplace influences the prices and availability of toys for anyone to purchase. At the same time, since Amazon is an enormous marketplace when you consider all the products available to anyone at any time, I want to observe how the toys are categorized and referenced throughout the data set.

Approach addresses the problem

With my approach, I believe that it will fully address the problem or situation in determining how the current toy marketplace within Amazon is operating in understanding who are the biggest distributors, the price of items, which toys are more available compared to others, and how the Amazon toy marketplace as a whole has an effect on its other market places or items by seeing the behavior of those that have viewed the toys.

Data

PromptCloud. (2017, September 15). Toy Products on Amazon. Retrieved February 4, 2020, from <https://www.kaggle.com/PromptCloudHQ/toy-products-on-amazon>

Data Set Name: Toy Products on Amazon

The data set is a pre-crawled dataset, taken as a subset of a larger dataset which included more than 115,000 toys that was created by extracting data from Amazon. The data set that I will be using will include about 10,000 toys. The data was created by PromptCloud’s in-house web-crawling service 2 years ago. When it comes to missing values, which the data set includes, they have been set to null within the data.

Variables:

* product\_name
* manufacturer - The item manufacturer, as reported on Amazon. Some common "manufacturers", like Disney, actually outsource their assembly line.
* price
* number\_available\_in\_stock
* number\_of\_reviews
* number\_of\_answered\_questions - Amazon includes a Question and Answer service on all or most of its products. This field is a count of how many questions that were asked actually got answered.
* average\_review\_rating
* amazon\_category\_and\_sub\_category - A tree-based, &gt ;&gt ;-delimited categorization for the item in question.
* customers\_who\_bought\_this\_item\_also\_bought - References to other items that similar users bought. This is a recommendation engine component that played a big role in making Amazon popular initially.
* description
* product\_information
* product\_description
* items\_customers\_buy\_after\_viewing\_this\_item
* customer\_questions\_and\_answers - A string entry with all of the product's JSON question and answer pairs.
* customer\_reviews - A string entry with all of the product's JSON reviews.
* sellers - A string entry with all of the product's JSON seller information (many products on Amazon are sold by third parties).

Required Packages

Stringr, tm, worcloud2, plyr, ggplot2

Plots and Table Needs

Plots: Bar Plot, Scatter Plot, Word Cloud Plot

Tables: Correlation, Summary, Head

Questions for future steps

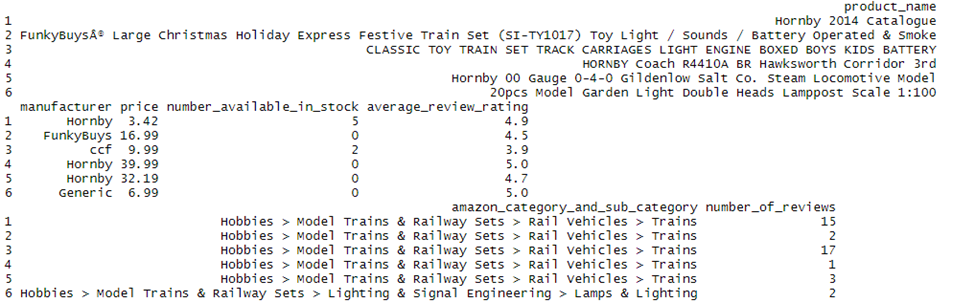
The process of creating a world cloud within R to show which products and descriptions have more of a presence within the data. Now, I possibly will need to learn more to understand how the toy market truly works and how Amazon is organized.

**Section 2 – Week 10**

Importing and Cleaning my Data

In importing this data, it was contained within a simple CSV file that allowed for me to import it similarly to how we have done this entire semester. Once I was able to import the data set, I sliced the data to only contain variables that were necessary for my research questions and tests that would be done throughout the approach. For the data within this collection, many of the values contained unnecessary characters such as currency markets as well as extra words that did not provide any more substantial information. In cleaning the data, I did want to keep rows that did not have complete information such empty fields. However, for some empty fields such as available items, I simply converted those values to zero since they did not currently have any of those items in stock. At the same time, I changed some factors into characters simply because I had no need for levels when it came to those certain variables.

Final Data Set (after cleaning)



Questions for future steps

With Amazon categories, they are split it into different sections, so I want to map about all the different levels within each of the values given within the data. For example, Names and Brands then Star Wars followed by Toys. I think I must learn how to properly split these different main and subcategories to understand if they follow the same process as other subcategories.

**Section 3 – Week 11**

Not self-evident information

Information that I believe is not self-evident would be the distribution between the different categories within the Amazon marketplace. For many of the toys, I could simply look at the manufacturers along with the prices and actual toys. But, to take a deeper look into how categories compare to one another would be useful information to see how the marketplace operates while incorporating the toy side of the entire process. For many of the categories, they are not explicitly divided into toy sections, but under different labels such as “Brands and Names”.

Different ways to look at data

When it comes to this data set for the Amazon marketplace, I could imagine that you investigate this dataset to predict the behaviors that customers will have when they have the mind set of purchasing toys online. A factor that I am not exploring very deeply is the sellers of the toys but more into that are manufacturing said toys. At the same time, I could look further into the actual reviews of these products and see if that plays a more of a factor compared to the rating system within the data set.

Plan to slice and dice the data

I plan on slicing the data into top the tier marketplace upon determining which manufacturers have the largest presence within the marketplace. At the same time, I have sliced my data into only incorporating the minimum price for the toys since a collection of them had a range of prices for the toys.

Summarize your data to answer key questions

Within my data, I could summarize my data into aspects that revolve around price of the toys or the available number of items a manufacturer has of that certain toy. Also, it can be grouped into how all the toys are represented through the five-star rating that Amazon has for all its products.

Types of plots and tables to help illustrate the findings

To help illustrate my findings, I plan to utilize a correlation table to determine how customer reviews influenced available number of items in stock. At the same time, I would use a scatterplot in visualizing the prices and toys as well as a word plot to represent the differences in how much a toy is mentioned or how a manufacturer is throughout the marketplace. I will also be using a word plot to also visualize the spread of the top manufacturers since the aim is to understand how the top manufacturers are competing in a marketplace that is mostly online. 

Incorporating any machine learning techniques to answer research questions

At this time, I do not think I will be employing machine learning techniques at this time because I am not seeking to test the occurrences of certain events and or seeking to determine what will happen towards the future. I am only researching to understand the current market in how it operates within Amazon’s online store.

Questions for future steps

At this time, a question that I have is to better understand how to focus on the measures related to the top prices, products, and manufacturer within my dataset. In my research, I am seeking who are now the major players and influence ratings now have on products rather than other avenues that were used during the retail age.

**Section 4 – Week 12**

Introduction & Addressing Problem Statement

In my project, I attempting to understand the current state of the toy marketplace within Amazon in its online market that includes not only toys but numerous types of products. At this time in our society, retails stores are slowly disappearing or at least retail stores that primarily focus on supplying toys due to Amazon allowing the ease of having these items delivered to our doors and offering a wide selection of products. With the stores closing, manufacturers are no longer looking at these locations to distribute these products. Now, they can almost cut out the middleman and use Amazon to distribute their products and market them to potential customers.

Addressing the problem statement

Like with any data set you acquire in search of answers, you must ensure that the data set is clean and aggregated to the best of your ability to prevent from further issues down the line during your analysis. With the Amazon market data set, I created a subset of data that was only necessary to determine the top contributors to the market and what was the current status involving the toys. Once I had the clean data, I first gathered a summary of the marketplace as well as the correlation between the variables of number of available now, the average review of the product, and the price using the Pearson method. In working with the correlations, I wanted to gain a standard visualization using a scatterplot to see any possible relationships exist between the variables. Once I saw the correlation results, I sought to understand the most common word frequencies of manufacturers and toy products by matrix using the “tm” package. Since I was unable to use categories to the best of my ability, I wanted to see if there was any association with the actual word, “toy”, for manufacturers and the products. Like how I previously mentioned before in this report, the categories of the marketplace do not precisely say toys as a category. Hence, I sought any insight if the word has any association with the products. For the manufacturers to make a more distinguished view of the frequencies, I created a word cloud of the manufacturers using the “wordcloud2” package. With being able to view the top frequencies, I then wanted to look further into the top manufacturers within the market who seem to have a strong position compared to other suppliers. For these top players, I wanted to see what their average price was for their products to see what is being offered to the public. With Amazon allowing for customers to openly leave ratings for the products and influence these manufacturers if they continue to receive poor ratings, I also gauged to determine the ratings these manufacturers obtained and how many. To obtain this, I simply used the “ggplot2” package to create several bar charts in one visual to compare amongst each other.

Analysis

Throughout my analysis, I was interested to see that overall customers tend to rate the manufacturers with a good rating. For most of the ratings that were given, the lower end of the ratings were around 4.2 out of 5.0 scale. At the same time, the number of items available for customers was more dominant when the cost of the product was less expensive. With understanding this correlation, I remember the idea that with the cost of a product being lower can lead to consumers purchasing more of said product. Looking at the list of the top ten manufacturers:

1. Oxford Diecast
2. Lego
3. Disney
4. Playmobil
5. Puppet Company
6. MyTinyWorld
7. Star Wars
8. Mattel
9. Hasbro
10. Corgi

I am overall not surprised that these are dominant players in the market at this time. With monopoly companies such as Disney acquiring more under their umbrella, it makes sense that their products would be in the top and other child company products be not that far behind as well. Towards the end, I noticed how much of an association the word “toy” had with all of the products and manufacturers. The word had lower association scores that I had initially expected, but I also remind myself that some of the toys are not listed under toy sections technically.

Implications

For the consumer, the applications are that there are dominant players in the Amazon marketplace when it comes to their offering of toy products. Yes, there are other manufacturers of toys that do exist, but overall, you will see the same manufacturers throughout the market. When the price is lower, the more the item will be available as well as when the ratings are higher. When searching through the marketplace through the categories, you will not always be able to locate a toy section; it may be under a different name especially if it is a popular brand. With seeing the number of products and the higher ratings that products are achieving in the Amazon marketplace, it safe to say that this will be the service to use towards the future in acquiring future toys.

Limitations

In my case, I believe that the limitations for the analysis was being able to determine the customer behavior when searching for products and the actions that they take when acquiring a product within Amazon. I was able to determine possible relationships with prices, products, manufacturers, and ratings, but I think someone else can build off that information in determining if consumers will stick to these top players or do, they search for different products across the market. It would be interesting to see consumer behavior within Amazon and then predict for most customers, like how it currently suggests products to you now.

Concluding Remarks

With the information that has been presented throughout this report and within the Markdown report, I believe that I give an insight on how the Amazon marketplace is currently operating when it comes to selling toy products to its consumers by describing relationships between areas of the market as well as determining the top manufacturers and suppliers of toys. In this information, I hope that you can examine how the direct connection to the consumer using the rating system can distinguish the manufacturers and the products that achieve success in the marketplace.