

The Toys Industry

The EU has the world's biggest single market for goods and services. The EU market for traditional toys and games, was worth EUR 15.8 billion in 2011 at retail sale prices. In contrast, the US industry is worth EUR 14 billion. The Chinese industry accounted for EUR 4.8 billion in revenue and has a high potential if income levels begin to grow.

Several external factors can have an effect on the market for traditional toys.

The first is the aging of population in developed economies, as shown by a more or less steady number of children in the EU and the US. China's one-child policy and growing wages have also resulted in a sharp reduction in the child population. Another reason is the increased competition from modern ICT devices that can be used as near replacements for traditional toys. Not only video games, but also mobile phones, laptops, and other forms of content contend for children's preferences and spending in mature and developing markets. Electronic devices, such as tablet apps, are direct and inexpensive replacements for preschool toys. At the macroeconomic level, uncertainties about the EU and US recovery from the global crisis, as well as the consequences for market conditions in developing economies, indicate that growth projections remain extremely unpredictable. Nonetheless, global toy demand is expected to grow by around 7.5 percent a year until 2016.

<i>Country</i>	<i>Consumption in million €</i>	<i>Production in million €</i>	<i>Direct employment</i>
<i>EU 28</i>			
<i>EU 28 Total</i>	<i>15,828.40</i>	<i>5,833.61</i>	<i>50,902</i>
<i>Other</i>			
<i>United States</i>	<i>13,971.70</i>	<i>4,382.33</i>	<i>35,037</i>
<i>China</i>	<i>4,802.80</i>	<i>16,011.30</i>	<i>128,012</i>
<i>Japan</i>	<i>5,201.10</i>	<i>2,200.08</i>	<i>17,590</i>

The role of Amazon

Amazon is a platform online created by Jeff Bezos in July 1994 in Seattle.

In 2000 Amazon did a merchant partnership with Toys “R” Us under which it would be the only toys supplier on the Amazon platform. This agreement was a 10-years partnership where Toys “R” Us had to pay \$50 million to Amazon plus a percentage of the sales.

This partnership was stopped because of legal actions taken from the toys company in 2006 but in 2020 Toys “R” Us returned to have their sales through Amazon.

The Data

The data is about the toys put on sale on the platform of Amazon. This is a data collection about Amazon toys that includes almost all that consumers can see on Amazon's website.

We have about 16 types of datas of different kinds, useful for the selling of the product on Amazon.

We have:

- **product_name**
- **manufacturer**
- **price**
- **number_available_in_stock**

Amazon gives the possibility of reviewing the product and rating it from one to five stars, thus our datas shows us:

- **number_of_reviews**
- **number_of_answered_questions:**
- **average_review_rating**

Our data include the categories and sub categories in which each product has been categorized on the Amazon platform. In our data there are also the references of other items that similar customers bought, and the items that other customers bought after viewing the product. In the data we have as well the description of the item on Amazon and the information related to it.

- **amazon_category_and_sub_category**
- **customers_who_bought_this_item_also_bought**
- **product_information**
- **product_description**
- **items_customers_buy_after_viewing_this_item**

On Amazon the users can ask questions about the product to other users that already bought the product, and our data includes the number of questions answered on the page of the product, as well as all the questions and answers related to the item.

Then, in our data we have the customer reviews related to the seller and the seller information about the products they sell on Amazon.

- **customer_questions_and_answers**
- **customer_reviews**
- **sellers**

The variables in the data collection can be broken down into three categories:

- **product data,**
- **seller data,**
- **buyer data.**

These three constitute Amazon's data ocean. We can see the interactions between buyers and sellers on Amazon's website by examining the relationships between the three types of data. For the analysis we are not going to use all of the 16 data, but just some of them. In order to do our analysis we are going to use the item manufacturer reported on Amazon, the number of answered questions on the platform, the categorization of the item on the platform as well as the sub-category, the references to other items that similar users bought, the customer's questions and answers, the customer reviews and the seller page with all the product this person sells on Amazon.

Questions

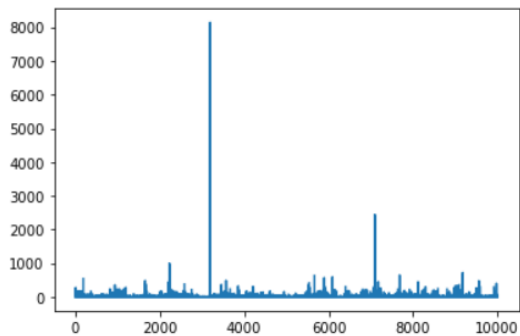
The data collection is aimed at obtaining valuable information that can be used in order to better understand amazon's business and draw meaningful conclusions on the toys segment for amazon's business and how to improve it in order to maximize revenues.

In particular, our analysis is structured to answer the following questions

- What kind of toy (category/subcategory) has the highest rating? (How many ratings have there been, and what is the average customer rating)?
- What is the average price for a toy?
- How much of each product is currently in stock, and are there enough of the highest-rated products in stock?
- Why are outliers important for our analysis?

Statistical Analysis

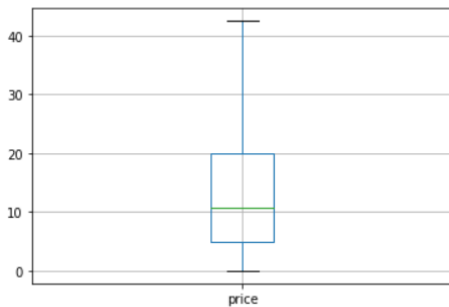
Price



Outliers

Clearly, we observed three highest outliers. In order to make the boxplot of the mean of price more readable, we hide these 3 outliers and analyze them individually.

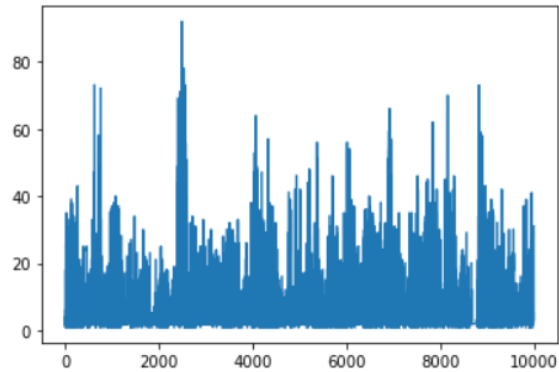
```
count    8565.000000
mean      21.277768
std       99.317922
min        0.010000
25%        4.990000
50%       10.640000
75%       19.990000
max      8140.300000
Name: price, dtype: float64
```



Mean

As we can see from the histogram, the overall price is mostly ranging around 5 to 20. In this case we can conclude that the toys are not selling at a high price.

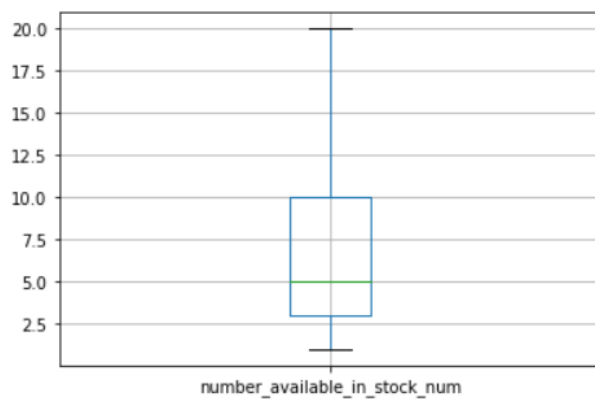
Number available in stock



Outliers

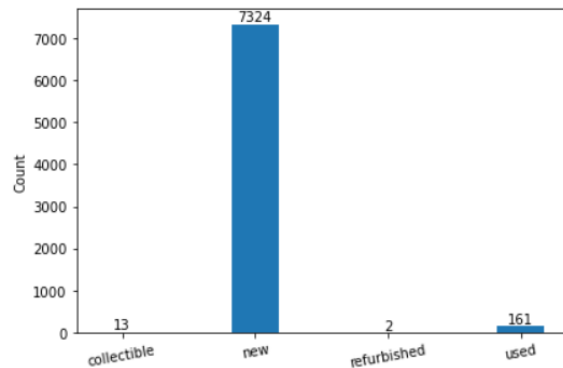
We hide the top 10 highest outliers ranging from 70 to 92.

```
count    7500.000000
mean       7.916933
std        8.805777
min         1.000000
25%         3.000000
50%         5.000000
75%        10.000000
max         92.000000
Name: number_available_in_stock_num, dtype: float64
```



Mean

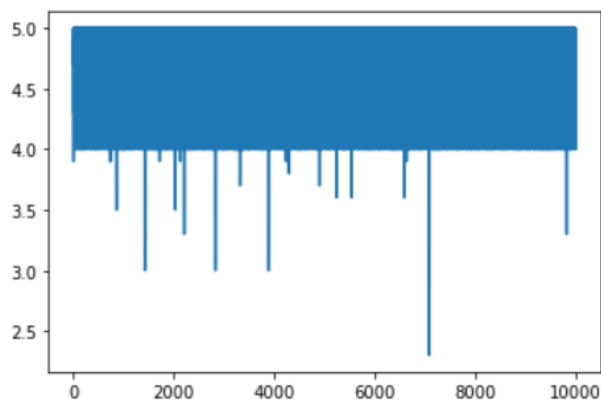
As we can see from the histogram, the number available in stock is concentrated around 4 to 10. So we can conclude that most products have stock near 5.



Stock type

From the graph we can see that most of the stock are new, there is about 1% of stock is used And collectible and refurbished stock have only 13 and 2 out of 7500.

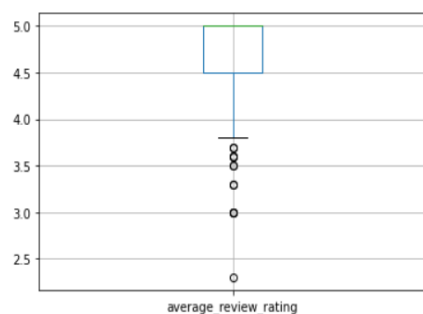
Average review rating



Outliers

This graph is very interesting, because it clearly shows the outliers (the products with low review rating). The blue box represents the large majority of products that have a 4 - 5 stars review rate

```
count    9982.000000
mean      4.707283
std       0.372279
min       2.300000
25%       4.500000
50%       5.000000
75%       5.000000
max       5.000000
Name: average_review_rating, dtype: float64
```



Mean

We clearly see that the average review rating is around 4_5 with few outliers which represent the low ratings.

Sentiment Analysis

We are doing sentiment analysis by using the TextBlob AI library that is trained on analyzing sentences.

We can see that the sentiment analysis is not perfectly accurate (as seen in cell ≤ -0.2)

These errors are due to the limit of the model which cannot understand perfectly the wording of every human being.

A few incoherences were spotted in the dataset.

The reviews extracted from the dataset contain sometimes expressions or words that can be interpreted negatively by the AI

-> #2376 "Nasty catch here" 4/5 rating ("Nasty" understood as a negative word)

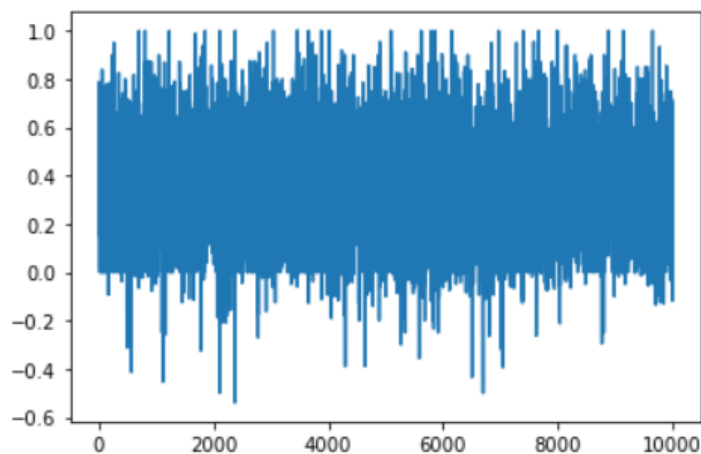
-> #1121 "Stop Grievous!" 5/5 rating ("Stop")

Also most of the reviews written as "Five Stars" are not understood positively by the AI model which again is not trained enough for such exercise : the wording seems very neutral and interprets it to a sentiment value close to 0 and not the real sentiment of the customer (5/5 stars)

The use of TextBlob AI is very experimental

In the graph below the correlation between the sentiment analysis and the average review rating is very low at 20%

	sentiment	average_review_rating
sentiment	1.00000	0.19875
average_review_rating	0.19875	1.00000



Conclusion

Some important variables are missing from this dataset.

We don't have a "number of sales." As a result, determining popularity is difficult.

We don't have a "list of questions." We only have the "number of answered questions." As a result, we are unable to determine the ratio of the question answering. This ratio may indicate the level of interest in a specific product.

We don't have a "cost of product." We just have the price. As a result, we are unable to calculate the profit.

Perhaps the next move should be to collect the missing data listed above in order to complete the study.

At this stage, we haven't analyzed some of the character variables in the dataset. Variables such as "customer questions" and "customer answers" have a lot of hidden value that is only waiting to be discovered by more powerful analytical techniques.

Answer to questions

1. The product with the highest price is Morhsuits Childrens Mr Block Head Costume, for £ 8,140.30 followed by Marvel Statue Thor Classic for £2,439.92 . Under the Top 10 filter, the product with the lowest price is Revell Monogram 1:72 - RVM5089 - Space Shuttle w/Fuel Tank & Boosters.
Without the filter, the average price for these toys is £ 21.28
2. Games has an average customer review of 4.7 (out of 5), with the most reviews at 14,214. Followed by Characters & Brands, then Arts & Crafts (sub-categories are also found within the Amazon category group). The Total Number of Reviews and the Average Customer Review is 4.7.
3. The category with the highest combined price, Hobbies with £2,439.92 , also has the most items in stock with 1,464.
4. According to analytics, certain "outliers" of data that contradict the overwhelming majority can also require analysis, if only to rule them out completely from the inference. Outlier Analysis adds value to the total Analytic Solution. This means that if a buyer discovers even one product review for a toy that may endanger his child, it totally changes the buying paradigm; he will not purchase that product.