A Work Project, presented as part of the requirements for the Award of a Master's degree in Business Analytics from the Nova School of Business and Economics.

ENHANCING PRODUCT RECATEGORIZATION IN THE TOYS INDUSTRY: A COMPREHENSIVE APPROACH INTEGRATING CLASSIFICATION MACHINE LEARNING MODEL AND INDUSTRY TREND ANALYSIS

AMAIA SALAZAR OCHOA DE
OCARIZ

Work project carried out under the supervision of:

Rodrigo Belo

11/01/2024

Abstract

This project encompasses a complete business analysis of the retail toys industry in a renowned company of this sector, which was conducted, first, with the development of a machine learning model which served to correctly categorize toys products into eleven predefined categories and, secondly, by performing a deep analysis at different granularity levels regarding the exanimated toys items, which served to analyse the most important pillars trends regarding the retailing business model, these being: pricing trends, availability trends, selection analysis and search trends. This industry analysis was accomplished by the usage of different automatized data pulling techniques such as complex SQL queries and python programs, in addition to the usage of data visualization methods.

Keywords

- DRI – Directed Research Internship
- KPI – Key Performance Indicator
- EAN – European Article Number
- ML – Machine Learning
- MP – Marketplace
- OPS – Order Product Sales
- TTM – Trailing Twelve Months
- AWS – Amazon Web Services

This work used infrastructure and resources funded by Fundação para a Ciência e a Tecnologia (UID/ECO/00124/2013, UID/ECO/00124/2019 and Social Sciences DataLab, Project 22209), POR Lisboa (LISBOA-01-0145-FEDER-007722 and Social Sciences DataLab, Project 22209) and POR Norte (Social Sciences DataLab, Project 22209).

1. Introduction

This thesis presents the culmination of my Directed Research Internship (DRI) undertaken in pursuit of the master's degree in Business Analytics at Nova School of Business and Economics. This internship was conducted at an e-commerce tech company in Madrid's branch within the Toys Retail Department, where I assumed the role of Product Manager Intern for a duration of 6 months.

The internship consisted of two phases or subprojects namely: (i) toys items classification into product categories with a machine learning model and (ii) toys industry trends analysis through a weekly newsletter.

The challenge that the development of this project aims to tackle consists of addressing the lack of an appropriate categorization model that could recategorize, in an automatic way, the extensive inventory of items in the toys retailing department. This deficiency leads to a consequential issue: the inability to provide precise Key Performance Indicators (KPIs) that reflect the accurate performance of the entire product line within the company.

1.1. Objective of the project

The main goal of this project is (i) to correctly categorize items that were not yet in the toy's main categories, with the aim of tracking more accurately the performance of the distinct categories of the toy's group line. Due to the standardization of the eleven main categories, around one million unique items fell outside them, and the performance of those was not being properly accounted for. Outside the main eleven categories, there are approximately 2,400 other categories that are deprecated and should be in disuse.

Once achieved this reclassification, the second objective is to (ii) provide the toys team with weekly updates on actionable opportunities related to concrete items, providers, and product categories. This was accomplished through the creation of a newsletter and a report that tracks how the trends in the toys industry are evolving week over week. The focus is on four major pillars: pricing, availability, selection, and search trends.

The integration of these two project components will support the team and enhance their negotiating with providers and business decision-making by delivering a complementary precise and up-to-date valuable information to the toys department team, since this goes beyond merely assessing the status of our providers within the company; it extends by offering a broader picture of their trends and their positioning within the analysed market segment of this industry. In addition, by leveraging on correct product categorization, a more accurate analysis of the existing products is achieved.

2. Literature review

In this section I will delve into the existing literature in the product categorization field, making a special reference to the product categorization area in e-commerce in order to understand and explore different methodologies or techniques. Product categorization in every online retail platform takes a key role in terms of categorizing a vast number of products due to its aid on different purposes such as measuring the performance of the sales of the different categories.

2.1. AutoML for structured data

As this project's machine learning categorization model was developed by using AutoGluon framework, examples of the usage of this same framework for a similar task as the goal of this project can be found on AutoGluon-Tabular: Robust and Accurate AutoML for Structured Data (Nick Erickson, 2020). This paper reaffirms the ease that this framework provides in machine learning model developing tasks by clamming that by a single line of Python code, a highly accurate ML model can be trained on unprocessed tabular datasets. Levering on AutoGluon's tabular predictor to ensemble multiple models by stacking them in multiple layers, rather than seeking the best model. In addition, this work praises the efficiency, robustness, and accuracy of this framework against its top competitors.

2.2. E-commerce product categorization via Machine Translation

As for the domain that this work project encompasses there are other existing papers that take other approaches for e-commerce product categorization. This research article titled E-Commerce Product Categorization via Machine Translation (Liling Tan, 2020), presents an approach of product categorization proposes the translation of product's natural language description into a sequence of tokens representing a root-to-leaf path in a product taxonomy. They also demonstrated achieving a better predictive accuracy than a state-of-the-art classification system for product categorization.

2.3. Multi-class and Hierarchical Product Categorization

Another existing research paper that encompasses a similar problematic as this work project is the one titled Large-scale Multi-class and Hierarchical Product Categorization for an E-commerce Giant (Ali Cevahir, 2016). This paper tackles down the problematic of organizing a large number of products listed in e-commerce sites assigned to one of the multi-level categories in the taxonomy tree. They had the objective of easing the time-consuming and complex tasks of carefully choosing a category among the considerable number of options for the products that are sold. To achieve this, they proposed an automatic classification tool build from the combination of two different neuronal models, i.e., deep belief nets and deep autoencoders for both product titles and descriptions. For this, they employed a selective reconstruction approach at the input layer during the training phase of deep neural networks, facilitating scalability for handling large-sized sparse feature vectors. To ensure training could be conducted within a reasonable time period, they utilized Graphics Processing Units (GPUs). The models underwent training on a dataset comprising approximately 150 million products, organized within a taxonomy tree of no more than five levels and encompassing 28,338 leaf categories. Evaluations conducted on several million products revealed that the initial predictions of the model aligned with 81% of the merchants' categorizations, provided that 'others' categories were omitted.

In conclusion, the literature review has provided a foundational understanding of the complexities and advancements in the field of product categorization within the e-commerce domain. By examining the methodologies and findings of pivotal studies such as "Large-scale Multi-class and Hierarchical Product Categorization for an E-commerce Giant" by Ali Cevahir (2016), "E-Commerce Product Categorization via Machine Translation" by Liling Tan (2020), and "AutoGluon-Tabular: Robust and Accurate AutoML for Structured Data" by Nick Erickson (2020), this review has underscored the significance of innovative approaches to improving product categorization. These works collectively highlight the potential of machine learning models and AutoML tools in enhancing the accuracy and efficiency of product categorization processes, particularly in the context of large and complex datasets.

The insights garnered from these studies are instrumental in informing the methodology and approach adopted in this thesis. The integration of classification machine learning models with a nuanced understanding of industry trends, as proposed in this thesis, aims to address the specific challenges encountered in the toys industry's product categorization. Moving forward, the subsequent sections of this thesis will delve into the methodological design, implementation, and evaluation of the proposed solution, with the ultimate goal of advancing the state-of-the-art in product categorization within the toys industry.

3. Data and Context

The retailing department of this e-commerce company operates in nine distinct marketplaces throughout Europe, five of them being the main countries: Spain, Italy, **Germany**, United Kingdom, and France, and four emerging countries, namely the Netherlands, Belgium, Sweden, and Poland.

Regarding retailing, it is important to understand that each product listed on the platform, referred to as an EAN item throughout this document, owns multiple attributes which describe the product specifications and features. Most of these features are visible in the product's detail page. Each EAN has an internal identification number, item name, brand, category id, bullet points with information, marketplace id, etc. In addition, there is also financial data related to each EAN item, stored in timestamped datasets which records different figures such as the revenue share, units sold, vendor information, etc.

Within the toys department, there is a vast array of toy categories, however, recently, the department has been divided into eleven main ones: 'Dolls', 'Infants/Preschool', 'Arts, Crafts and School Supplies', 'Games and Puzzles', 'Vehicles', 'Plush', 'Action Figures & Accessories', 'Outdoor & Sports Toys', 'Learning & Exploration', 'Building Sets' and 'Other'. The categories were narrowed down to these main ones in order to better analyse the performance and profitability of each type of toy.

For the analysis of the trends in the toys industry, four essential pillars were considered: pricing, availability, search trends and product selection, at three distinct levels: category level, vendor level and EAN item level. These different granularities serve to dive deep into the highest granularity, namely EAN item level, to analyse the root cause of the trend variation of the pillars. For this analysis and reporting, each pillar's data will be explained in the following sections.

3.1. Pricing data

The gathered pricing data resembles the price that each EAN within the toys' retailing department has in a defined timeline. Along with that, the prices that competitors have for matching products at the company's retailing site was used to study the industry's trend of its prices. Competitors' prices are gathered and stored through automatically scrapping other retailers on a daily basis, which is carried out by a program developed prior to this project.

By combining this data, the average weekly price point of the current and the last four weeks was calculated to analyse its trend at the different mentioned granularities, enabling spotting the top offenders that most variated the price, both in an increasing or decreasing trend whether it be a particular EAN item, vendor, or category.

3.2. Availability data

When it comes to availability, the company tracks the daily glance views for each EAN item, which represents the number of views that each EAN receives on its product page in a day. This provides insights into the demand and popularity of a product. To check the availability, a distinction is made between the glance views that occurred while the EAN was available on the site and when the view occurred while it was out-of-stock. With the tracking of the out-of-stock glance view metric, the company can easily discern the items that have a potential high demand which need to be restocked.

In order to enhance this metric, it was combined with another metric, which flags the other retailer sites' availability. This was employed to measure the ratio of our out-of-stock glance views that have at least one in-stock signal from other retailers over the total out-of-stock glance views. This data is also gathered by scrapping other retailers' sites with the previously mentioned program and stored in its correspondent database system.

3.3.Search trends

To tackle search trends analysis, the company owns an internal tool that allows retail employees to observe top level anonymized search trends data (query frequency, clicks from search, sales from search etc.), to deep dive into the keywords customers are searching for and to understand subsequent associated business actions. This tool shows aggregated raw search data to a weekly grain at a marketplace and keyword level and never at customer level, meaning that only weekly or aggregated weekly data can be generated from the tool. For keywords, the tool provides directional figures when aggregating across weeks as each week of data only includes keywords with a search frequency above a defined cute off. This data is highly valuable to understand the trends in the toys industry when well-presented and to know what customers are looking for, in addition to allowing to understand ahead of time the demand that certain toys, toy types or brands are having or going to have.

3.4. Product Selection data

For any e-commerce retail site, it is fundamental to have a wide selection of products, as to specially in e-commerce, where it is particularly quite easy to switch to another website and find what you are looking for.

In order to remain competitive in this industry in online retailing, a particularly important pillar to consider is product selection. This pillar plays a pivotal role in evaluating the breadth of our site's product offerings across various brands, often referred to as *brand coverage*. The brand coverage is measured as the percentage of items of each brand's catalogue is present in our site's catalogue. This data is gathered on a monthly basis and is obtained by an internal team in the company which collaborates with the providers.

In addition, a broader data source used for measuring selection was used by which not only the providers catalogue is compared but also the universal external catalogue is used to measure the gaps in our site.

4. Development

This section provides an overview of the methodologies used and the tasks undertaken to successfully complete this project. Firstly, delving into the machine learning classification model employed for categorizing the items, and followed by a comprehensive breakdown of the reporting designed to analyse industry trends within the toys sector.

4.1. Category classification ML model

As previously mentioned, the objective of this part of the project is to correctly recategorize items into the eleven defined categories. For this, AutoGluon framework was chosen to build a classification model. AutoGluon is an open-source machine learning framework developed by AWS, which eases the process of building, training, and deploying machine learning models (AutoGluon AWS, s.f.). The rationale for selecting this framework against others arises from its user-friendly development and deployment capabilities. Unlike alternative frameworks, this choice enables end users, even those without a technical background in machine learning, to use it. This, in turn, streamlines the handover to the team and enhances the potential for future model improvements.

This part of the project is divided into distinct phases detailed in the following subsections.

4.1.1. Data pulling and feature selection

The first phase of this project consisted of pulling the item data instances and the selection of the features necessary for the training of the model.

The item's attributes selected as features were:

Item name	Item brand	Bullet points	Category code
Name or title shown in the product detail page.	The brand name to which the product belongs to.	Five bullet points of the product, which describe the product and its features.	The target variable, which identifies each of the eleven defined product categories.

In order to pull this data, a set of queries were built which, first, compute each EANs total revenue for the current year across all eleven marketplaces, and second, selects the top 10,000 items with most revenue for each of the 11 final categories, and lastly filters by 1 marketplace to get the items' attributes in a single language for training purposes.

It was assumed that the items with highest revenue share are already correctly classified, as prior to this project, there was an initiative that with the collaboration of the providers, their selections' products had been manually categorized.

4.1.2. Data preparation

All the subsequent development of the project was conducted using Amazon SageMaker, which is a fully managed machine learning service, which is a tool within AWS that serves to quickly and easily build and train machine learning models, providing an integrated Jupyter authoring notebook instance (Amazon, s.f.).

In order to prepare the training data, a series of data cleaning steps were undertaken to enhance the dataset's quality. Firstly, numeric values and substrings were removed to discard potential sources of noise and bias, particularly in cases where numerical information, such as item measurements, or any other numeric information present in the item name or bullet points, could skew the model's understanding.

Secondly, the removal of html tags and especial characters was performed in order to preserve a consistent text representation, in some cases, the bullet points or item name contain these especial characters, namely, emojis, hyphens or other type of undesired characters.

Lastly, the elimination of stop words was carried out to refine the dataset by removing commonly occurring words that do not contribute significant meaning towards the training of the model. This data cleaning procedures were fundamental in ensuring that the training data used for the model was reliable and accurate.

Another step in the data preparation process was conducting under sampling to the dataset with the aim of achieving a balanced distribution of instances across all eleven categories. More specifically the dataset was downsized to contain 6,000 instances of each category. This reduction of data instances also ensured that the model training was performed faster and more efficiently.

4.1.3. Model Training

As mentioned before, AutoGluon framework was chosen to develop and train the recategorization model. It provides different options depending on the data types and the target of the prediction to make, for this particular project, AutoGluon's MultiModal Predictor (AutoMM) was chosen not only due to its ability to process different types of data (i.e., text, image, tabular, …) but also because of its cross-lingual transfer learning ability. By employing its built in 'multilingual' preset, AutoMM automatically loads a backbone that is suitable for zero-shot transfer (AutoGluon AWS, s.f.) this enables that the model trained in a certain language can be directly applied to datasets in other languages. This is very suitable for this scenario since the model will be used to make predictions on data from all of the marketplaces with their respective languages. In addition, the hyperparameter of max epochs was set to 10, meaning that the training passes through the entire training dataset ten times enabling the model to update's its parameters based on the information contained in the data. After training, it obtained an accuracy score and F1 score of 0.81 which demonstrates its ability to correctly predict around 81% of the cases and its consistent performance across all the different categories.

4.1.4. Prediction and model validation

When it comes to predicting the correct category of the items that fall outside the main categories, the initial step involved building a query to pull the items that fell outside the main categories. The output threw a total of 958k unique items that are incorrectly categorized in the toy's product line. To proceed with the prediction process, it was decided to perform it in a staggered way, due to the considerable number of items. For this, an analysis was performed by which data related to the items was pulled to prioritize the selection of items. The first bulk of items to be recategorized included those which are present in more than five of the nine marketplaces in which the company operates, as well as those that obtained at least one glance view year to date. This resulted in a preliminary set of 19k items. In order to perform the actual prediction, the *predict_proba()* function was employed as it yields the probability associated with each prediction, offering a more comprehensive insight into the classification output.

To validate the predictions made, two different validation techniques were used. The first one is based on a combination between stacking and cross-lingual validation.

The data pulled to conduct this validation encompassed the features of each EAN item across all the marketplaces where it is available. Consequently, each instance of the same item was predicted as many times as it appeared in the various marketplaces, each in its respective language. For instance, if a particular item is present in seven marketplaces, it has been predicted seven times. This way, it is ensured that the same prediction is achieved regardless of the language of the data.

Figure 2 shows how many EANs have been predicted into how many distinct categories across their presence in the different marketplaces. The table shows that 45% of all EANs have been predicted into a single category, this percentage of EANs show that the cross-lingual capabilities of the model work as expected. In addition, the grand total percentage shows that 18.24% of EANs that are predicted into a single category, are present in 9 marketplaces, and this percentage of grand total depicts a decreasing trend as the marketplace presence number decreases, meaning that the more marketplaces the EAN is present in, the more accurate is the prediction of the category, in other words, the prediction of the same EAN in different marketplaces differs less the more marketplaces' info of the EAN it has.

In case different prediction is made for the same EAN across the nine different marketplaces, the prediction is prioritized by (i) the most repeated predicted category among the marketplaces that the EAN is present and (ii) the highest probability within the predictions.

In Figure 1, the result output has been prepared to show the most predicted category of each EAN across all marketplaces and show the average probability of the prediction, in addition to the old category it belonged to. This file was then shared with the team that is in charge of the database in order to update it.

The second validation method employed was selecting ten subsets of ten random EAN items of different ranges of prediction probability, and manually checking that the category that was predicted is correct. With this, the accuracy percentage of the model is validated to get an estimation of it.

4.2. Toys Industry Trends Analysis

For this part of the project, which aims to analyse the trends of this industry based on data from different pillars: pricing, availability, product selection and search keywords trends. With the final objective of providing valuable insights for the toys team, an automated process that generates a weekly newsletter and a comprehensive report was implemented. This section will provide a clear explanation of the evolution of the project's deliverables, outlining the distinct stages of execution within each of the key pillars.

4.2.1. Pricing

When it comes to pricing trends, the aim was to track the variation of the prices of the items within the toys market and spot the ones that had the highest variation, both positive and negative, in order to understand the increasing and decreasing price trends, respectively.

For this purpose, focus has been placed on the items that are buyable not only on our site but also on other retailers' sites, in order to compare their price variation with ours. The data is pulled with a SQL query which computes the weekly average price point of each EAN for the past four weeks and aggregates it at three different granularities: EAN level, vendor level and category level. These three granularities provide a different level of detail which enables diving deep into the top offenders of the price variations whether it be a certain EAN or a specific vendor or a whole category which can be the cause of the price variation.

For each of the buckets, a visualization graph has been created to enable the team to spot the top offenders at a glance. In addition, in the report it can also be found additional information about the EANs, such as the item name, the brand, and the category it belongs to, to ease the process of diving even deeper into the data.

An example of business analysis to perform with this set of visualizations, involves analysing the different levels that are displayed by the different graphs, starting by focusing on the category level to see which is the trend of the prices of each toy type in order to detect if there is a pattern in the price variations, since depending on the season of the year, the demand of one category or another could vary (i.e.: Outdoor & Sport Toys in Figure 3) which due to the end of the summer season the overall prices within that category are dropping. In addition, it can be used as a tool to compare the categories among each other and get insights about the pricing trend that each toy type is having in the toys industry.

As next step of the analysis, focusing on the next granularity level (see Figure 4), performing an analysis at vendor level serves to callout the team that is accountable for each provider and callout potential promotions that are being carried out in other retailers, this can be leveraged on to action conversations with providers and spot negotiation opportunities. Finally, when analysing the EAN level visualization (see Figure 5) and report, the specific items that are varying their prices the most can be detected. This is particularly useful in deal seasons to pinpoint the items that are having the highest discounts or, on the contrary, a very trendy product is increasing its price point.

The analysis of price trends aids on detecting promotions or deals seasons in the toys industry, or in specific products or providers, which helps the team to rapidly adjust our prices to remain competitive within the market segment that it is being analysed. In addition, it is useful to understand how the toys industry prices vary in time and depending on the season of the year.

4.2.2. Availability

Availability is an especially important input metric in this company, since it is key to pay attention to EANs that are being very visited are in-stock. To analyse this pillar, a new metric has been created, focusing specifically on availability from the perspective of out-of-stock glance views. This approach aims to identify the items that customers frequently search for but find unavailable or out-of-stock. The metric aims to display the number of out-of-stock glance views at our site, along with the percentage of those glance views that do have an in-stock signal in other toys retailers' sites.

With this added information, the team is able to know when there is a high volume of glance views on items that are out-of-stock and most importantly focus on those products that actually are in-stock on other sites. Having this insight could lead to an enhancement in the market share and on a better customer experience, by avoiding them the pain of face non-available products.

In Figure 6, the category view is depicted, in which the trend of the out-of-stock GV count is shown for each category, in addition to the percentage of out-of-stock GVs that are available in other sites.

In the same way as for pricing, this metric is visualized at the three granularities mentioned, letting the team dive deep into the root of the cause of potential availability issues.

An example of business analysis performed based on this new metric consisted on spotting the vendors which have a trend of increasing number of out-of-stock GVs and a high percentage of competitor in-stock signal (i.e.: Vendor #6 in Figure 7Figure 7) and dive deep and locate the vendor's top offender items, which are the ones that are receiving above a set threshold of out-of-stock views. With this information, the brand specialists are able to negotiate with the providers to procure those items that competitors have, and we lack.

The analysis of this pillar’s trend contributes to understanding the toys industry, as per putting side by side two important metrics, presenting a new dimension of the availability. This enables the team to detect more easily the items that are trending but are out of stock at our site and the ratio of them that are in stock in other retailers’ sites. This provides a hint to the team in order to know which EANs to procure and lets us stay ahead of our competitors in terms of having available products that they also have but at a better price.

4.2.3. Selection

Product selection represents the proportion of a vendor's catalogue that our site owns in its retail inventory. A higher score in this metric significantly enhances competitiveness within the market, and it is intricately linked to product availability. The coverage metric is measured at brand and item level, and it focuses on analysing the gap of items that this company does not cover.

There are two data sources to get this metric, (i) one contains all the top brands that this retailing company manages in 5 of the main marketplaces and the other (ii) contains a broader range of brands and items across the 9 marketplaces which our site does not have and that have a potential of being popular items that might be interesting to have. Usually, the team focuses on the first metric, and in order not to miss opportunities on selection that even if it is not part of the top brands could potentially be relevant to have, a data cross-checking has been accomplished between these two data sources.

For this, two different visualizations and reports were created, one showing the coverage status at brand and item level of the running month where two metrics are considered, the head universe, namely the whole catalogue of the brand, and the head gap, which represents the amount of items that retail of this company does not have. This let the team analyse in a glance the percentage of coverage on each brand and where the greatest gap is. The other visualization depicts the potential score of each brand/item which is calculated by the potential glance views that it could have, which is calculated with a machine learning model by another team within the company.

The analysis of the search trends enables understanding what is being and going to be popular in the toys industry, which lets us further understand the customers' preferences. This is particularly useful for the team, as it can serve to provide clues about the customers' tendency to acquire a specific trendy toy or type of toy so that we onboard those trendy items in our retail catalogue.

4.2.4. Search Trends

Analysing search trends in this company is a crucial aspect since it provides valuable insights into customer preferences and demands. This metric synergizes effectively with the rest of the pillars as it can serve as an early indicator for potential price fluctuations driven by popular products that are being actively searched for. Moreover, it aids in proactive inventory management in advance and allows for expanding product selection in response to emerging trends reflected in the search data.

To report this metric, several attributes were considered. For each searched keyword from the last 4 weeks, the search count, product sales amount (OPS), and the purchase rate were observed.

To carry out this analysis, three visualizations have been created to reflect, firstly, the top 15 keywords which accumulated most search counts during the current week in each marketplace. Secondly, showing the purchase rate and OPS of those top searched keywords, and lastly, the search increase rate has been calculated to spot the keywords that had the steepest searched count increasing curve, by computing the percentual increase between the average of search count of the last 3 weeks and the current week as shown in Figure 13.

An example of business analysis carried out around this visualization, consisted of focusing on the keywords that had the highest trend increase rate during the current week, and correlate those keywords to real time events happening at the moment or to a certain micro season. As mentioned before, this analysis and all the other ones, are accountable for the 9 marketplaces present in Europe, this served to dive deep and filter into each marketplace through the report and figure out what is trending around the toy's product line in the different countries, in addition to providing valuable information which was used to detect demand peaks of certain product types within the toys industry. For instance, the visualization in Figure 14, provides a quick view on the top 15 keywords that most increased their search curve WoW across all 9 marketplaces. The team leads can use this visualization to callout brand specialists that are accountable for certain brands or providers that are related to the top keywords and make sure that the stock levels are correct to cover the expected demand.

Analysing search trends allowed us to gain insights into the evolving preferences and inclinations of consumers within this industry. This knowledge can be strategically leveraged to proactively stay ahead of customer trends, ensuring a competitive edge in meeting their changing needs and preferences.

4.3.Automatization

To fully automate the reporting and data visualizations of the toys industry trends, the queries in charge of pulling pricing and availability data were automated to be weekly executed and be stored in a share point folder. In addition, the excel file created for the report was connected through power query with the raw data files of the share point. This way, by updating the data connections, the pivot tables and visualizations generated are also updated with the newly added data each week.

Additionally, to carry out the pulling of selection and search trends data, a python program was developed (see Figure 15). This user-friendly program has a built-in interface by which the user can select which data to download. It also contains two input text boxes to let the user identify itself in order to access certain data that requires authentication.

This program leverages Selenium Web Driver to automatize the process of exporting the data from the designated portals and tools. The data download is fully automated, and it ensures that it retrieves the data from the designated period. In addition, the program is configured so that the download directory is connected to Excel as well, so that the only requirement is refreshing the data sources in the reporting Excel file.

5. Results

This section describes the results obtained by carrying out this project and the impact they had on the business group line that its being analysed. First delving into the results obtained in the classification model and the impact it had in the performance measurement, and secondly analysing the results obtained with the reporting of the industry trends.

5.1. Category classification model

As already mentioned, there were around one million uncategorized Items in the toy's group line across the nine marketplaces. The preliminary prediction and the stacking validation showed a high accuracy which can be trusted to predict the Items that fall outside the main categories. The Items have been gradually recategorized in chunks, considering the number of marketplaces in which they are present. It started with items present in five marketplaces or more, which are around 55,000 unique items, and continued in a descending way until all the items were covered.

Figure 16 shows the impact of the recategorization of each bulk of items represents the marketplaces in which they are available, and the performance they have translated into glance view count and the revenue obtained in the TTM. A total of 958,000 items gathers a total of 161.8 million glance views and a revenue of 20.3 million euros that before performing the recategorization were not being accounted for to track the KPIs of the toy's group line.

In Figure 17 is summarized the impact at category level of the results thrown by the recategorization of the items present in five or more marketplaces. After filtering those which have more than zero glance views, the impact obtained reflects the recategorization of 19k items into the categories shown below, with an added performance of 6.3 million euros of revenue and 20.8 million glance views measured in the TTM. It is remarkable the impact observed in categories such as 'Infants/Preschool' and 'Learning & Exploration.' Although only 1,189 and 521 Items were reassigned to these categories, they contributed substantially, accounting for a combined revenue of 3.7 million euros.

This impact transcends into all performance analysis and KPIs tracking at category level that the team overviews on a daily basis, facilitating a more accurate tracking of the performance of each category.

5.2. Toys industry trends newsletter and report

With the accomplishment of this project, it has been possible to dive deep into the most important pillars of the toys group line and provide a clearer and more transparent overview of them to the team leads of the different pillars within the toys group line. As a result, the team has been able to leverage on the generated weekly report and newsletter to visualize data from previously untapped tools and explore new metrics that had not been considered before.

For example, the search trends data was not being consistently tracked by the team, because some were unaware of the tool where this data was allocated. The newsletter and report ease the review of and accessibility to this data, which is crucial for informed decision-making, strategic planning of the stock-levels, by maintaining a comprehensive understanding of evolving trends of the customer searches.

In addition, the use of a new metric was implemented, specifically, the metric used in the availability analysis which illustrates the percentage of out-of-stock GVs that are in-stock in other retailers' sites. The use of this metric enabled to get deeper insights about this pillar offering the additional factor of the competitors' in-stock signal by which, if taken into consideration, it can be prioritize the restocking of items that other retailers have, and our site does not so we remain competitive.

6. Discussion

In the forthcoming section, a comprehensive analysis and interpretation of the results revealed in the preceding pages will be presented. This chapter provides the academic core of this DRI work project, where the data will be discussed.

6.1. Encountered limitations and areas for future research

One of the main limitations encountered during the development of this project has been the inability to gather real-time data. This issue impacts on the freshness of the transmission of the information related to the different pillars analysed in this project, especially the pricing pillar, which is the most variable one day to day even hour to hour due to the matching that is performed to maintain a lower price than other retailers.

Another limitation found involved the category classification model's further validation against correct classified EANs. This limitation was mainly due to time constraints and the massive number of different EANs items that were processed. It would be optimal, for future research, to perform a manual validation of different subsets of the predicted EANs to ensure that they are correctly categorized. Due to the nature of the products information which is very sparse among different toy products even from the same category, a future improvement area would involve categorizing the bullet point information based on similar characteristics of the product features. This approach aims to enhance the model's accuracy by organizing the bullet points information that share common traits. Currently, the top five bullet points that serve as training features contain disparate details of the EAN item, lacking uniformity in the type and sequence of information provided. This suggested refinement seeks to enhance the model's ability to discern patterns and similarities within product features for effective learning and accurate predictions.

6.2. Future steps

This section compiles the different future steps that would have been taken, but, regrettably, could not be taken due to time or resource constraints. Organized into short-term, mid-term, and long-term steps to take.

6.2.1. Short-term

Among the short-term steps, regarding the reclassification model, the first step to proceed with would be executing and performing the impact analysis of the prediction of the remaining Items (around 903,000 Items.), as well as, proceeding with the EAN reassignment procedure, which requires the involvement of another internal team within the company. Unfortunately, time constraints hindered the completion of the finalization of this part of the project.

When it comes to the industry trends reporting part of the project, a short-term measure to proceed with would be the gathering the category leads' feedback in order to tailor the newsletter and the report visualizations with the objective of keep on improving and adapting it to the team's and product line's needs.

6.2.2. Mid-term

A mid-term step to take would be to escalate the recategorization model to other group lines within the retail departments of the company. By adapting it to the categories there are in each group line, this step can be leveraged to add value in the form of measuring better the impact of every EAN and category in the different product lines that face the same issue as in toys.

6.2.3. Long-term

A long-term step would be to improve the ML model's accuracy by adding image data to the features of the training data leveraging AutoGluon's ability to train several types of data with its multimodal predictor. This task was not possible to fulfil due to time constraints which made it difficult to gather all the necessary image resources.

7. Conclusion

In this conclusion chapter, the culmination of extensive data and business analysis tasks is undertaken, encapsulating a brief summary of the key findings, insights and contributions to the industry analysed throughout this academic endeavour.

With the completion of this work project thesis, the task of mass recategorization of incorrectly categorized toy products was accomplished and in consequence, analyse in a more accurate way the KPIs that involved those miss-categorized products. Moreover, it has been possible to automatize reporting processes that are key to analysing the business trends within the toys industry. With this it was possible to understand the industry in more depth and come to different conclusions. Firstly, it was confirmed that this industry is a very seasonal one, due to the different peaks that came across with the trend analysis. Additionally, with the trends reporting, it was possible to tell a story via the data presented, which enabled the team to understand the variations and changes of the analysed pillars.

Even though the preliminary prediction of recategorization of EANs showed a minimal impact in revenue and glance views, this project can be leveraged to efficiently recategorize the remaining EAN items as well as be escalated to englobe other product lines within the company in the near future.

In conclusion, it can be determined that the objectives set at the beginning of this project have been fulfilled and has served as an additional reporting perspective for the toys team which shed light on retailing pillars that were not assessed before leading to a better understanding of the trends of this industry.

8. Annex

ASIN	Predicted Category	Probability
B0CLY3ZTN9	Games & Puzzles	99.98%
B0CLY73TWH	Games & Puzzles	99.98%
B0CLY3Z584	Games & Puzzles	99.98%
B0CLY9NR2C	Games & Puzzles	99.98%
B01HMYTF9S	Games & Puzzles	99.95%
B017T4XAQA	Building Sets	99.95%
B0B3XRMV6R	Action Figures & Accessories	99.94%
B01A7ER77W	Games & Puzzles	99.94%
B0C6FVTMM9	Games & Puzzles	99.94%
B0B4WPBWGZ	Dolls	99.94%
B01HSE6KB8	Games & Puzzles	99.94%
B003F7YI8K	Building Sets	99.94%
B01ACLVH9E	Games & Puzzles	99.94%
B0BPS25P6K	Games & Puzzles	99.94%
B08WM9ZXGD	Games & Puzzles	99.94%
B01BNLKT2I	Dolls	99.94%
B0BSR5RQFZ	Building Sets	99.94%
B09GW1PXNG	Dolls	99.93%

Figure 1 Prediction output

Predicted Category Count	MP Presence	Count of EANs	% of MP Presence	% of Grand Total
1	In 9 MPs	3,487	40.45%	18.24%
	In 8 MPs	1,724	20.00%	9.02%
	In 7 MPs	689	7.99%	3.60%
	In 6 MPs	507	5.88%	2.65%
	In 5 MPs	1,445	16.76%	7.56%
	In 4 MPs	249	2.89%	1.30%
	In 3 MPs	102	1.18%	0.53%
	In 2 MPs	147	1.71%	0.77%
	In 1 MP	270	3.13%	1.41%
1 Total		**8,620**	**45.09%**	**45.09%**
2	In 9 MPs	2,382	34.15%	12.46%
	In 8 MPs	1,876	26.89%	9.81%
	In 7 MPs	614	8.80%	3.21%
	In 6 MPs	510	7.31%	2.67%
	In 5 MPs	1,279	18.33%	6.69%
	In 4 MPs	178	2.55%	0.93%
	In 3 MPs	100	1.43%	0.52%
	In 2 MPs	37	0.53%	0.19%
2 Total		**6,976**	**36.49%**	**36.49%**
3	In 9 MPs	1,108	43.49%	5.80%
	In 8 MPs	584	22.92%	3.05%
	In 7 MPs	274	10.75%	1.43%
	In 6 MPs	202	7.93%	1.06%
	In 5 MPs	329	12.91%	1.72%
	In 4 MPs	38	1.49%	0.20%
	In 3 MPs	13	0.51%	0.07%
3 Total		**2,548**	**13.33%**	**13.33%**
4	In 9 MPs	406	51.01%	2.12%
	In 8 MPs	184	23.12%	0.96%
	In 7 MPs	86	10.80%	0.45%
	In 6 MPs	59	7.41%	0.31%
	In 5 MPs	56	7.04%	0.29%
	In 4 MPs	5	0.63%	0.03%
4 Total		**796**	**4.16%**	**4.16%**
5	In 9 MPs	93	62.00%	0.49%
	In 8 MPs	31	20.67%	0.16%
	In 7 MPs	15	10.00%	0.08%
	In 6 MPs	6	4.00%	0.03%
	In 5 MPs	5	3.33%	0.03%
5 Total		**150**	**0.78%**	**0.78%**
6	In 9 MPs	18	66.67%	0.09%
	In 8 MPs	4	14.81%	0.02%
	In 7 MPs	3	11.11%	0.02%
	In 6 MPs	2	7.41%	0.01%
6 Total		**27**	**0.14%**	**0.14%**
7	In 9 MPs	2	100.00%	0.01%
7 Total		**2**	**0.01%**	**0.01%**
Grand Total		**19,119**	**100.00%**	**100%**

Figure 2 Category Prediction Results Assessment

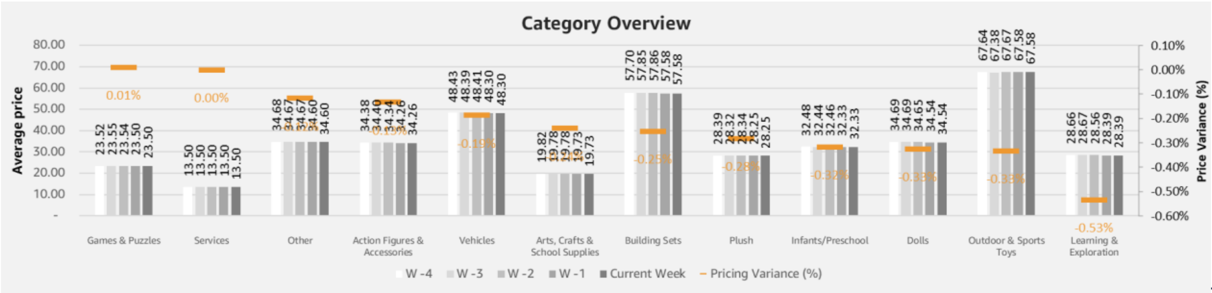

Figure 3 Category overview graph of pricing pillar

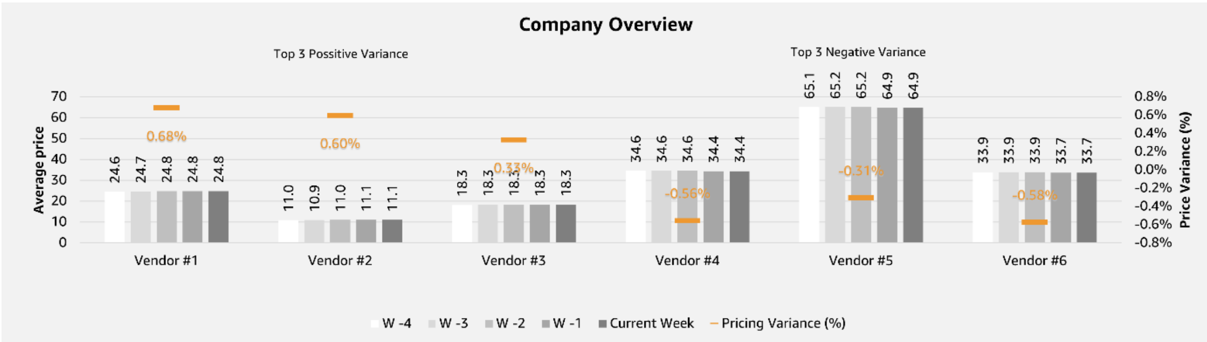

Figure 4 Vendor overview graph of pricing pillar

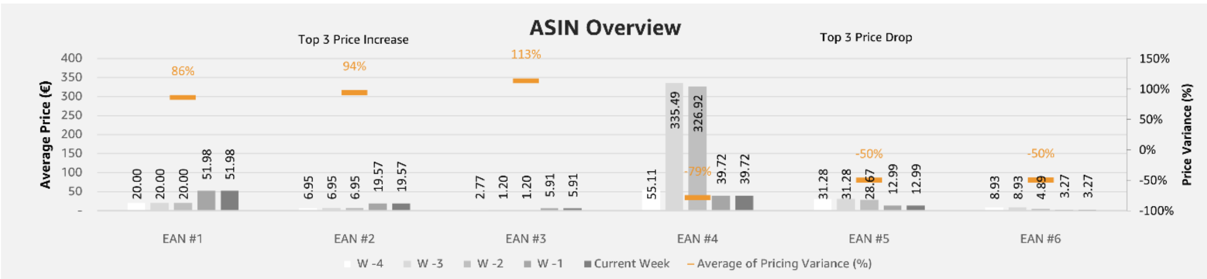

Figure 5 EAN overview graph of pricing pillar

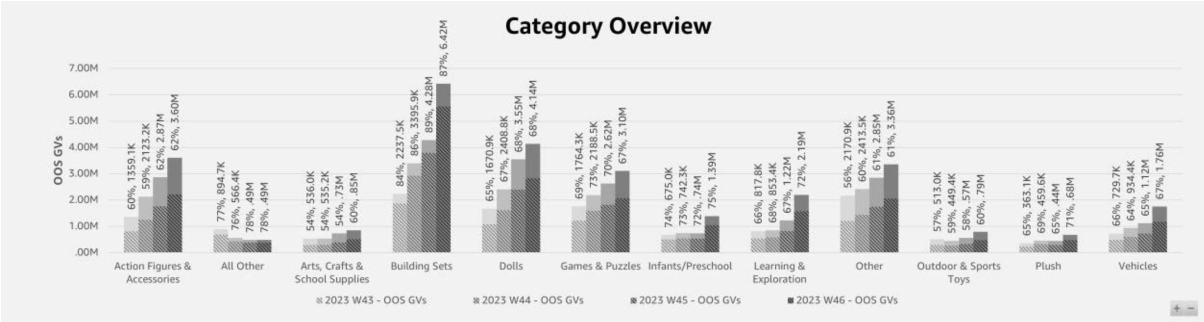

Figure 6 Category overview graph of availability pillar

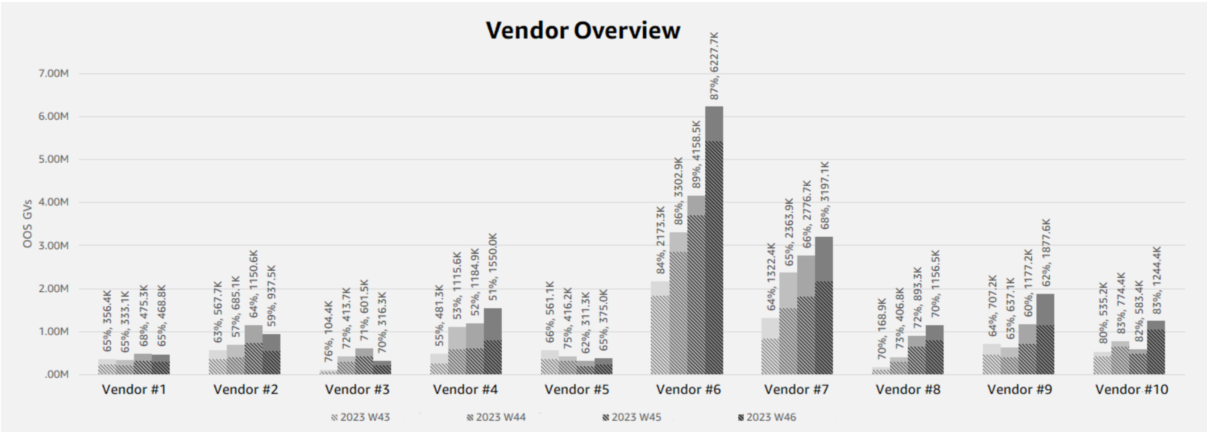

Figure 7 Vendor overview graph of availability pillar

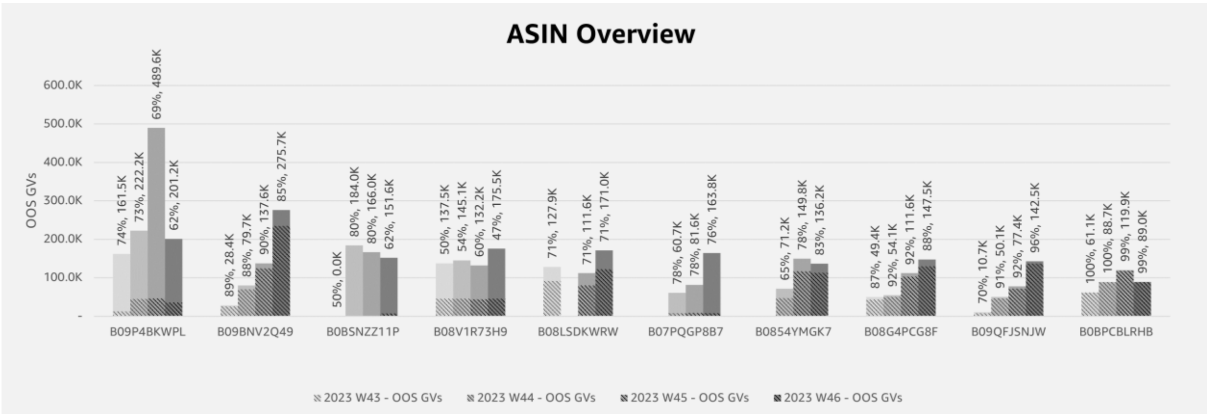

Figure 8 EAN overview graph of availability pillar

Brand Coverage (Brand level)

s	Brand Coverage %	Head Gap	Head Overlap	Head Universe	of which UTA	Total Goal
Brand #1	95%	610	12198	12808	429	91%
Brand #2	96%	356	6191	6547	816	95%
Brand #3	7%	225	3	228	2	10%
Brand #4	0%	188	0	188	0	2%
Brand #5	90%	179	975	1154	81	91%
Brand #6	67%	169	241	410	1	22%
Brand #7	97%	139	4018	4157	351	97%
Brand #8	92%	136	744	880	106	94%
Brand #9	88%	124	210	334	103	98%
Brand #10	91%	120	903	1023	21	54%
Brand #11	97%	116	2777	2893	797	90%
Brand #12	99%	109	7585	7694	303	97%
Brand #13	1%	108	3	111	0	51%
Brand #14	88%	99	272	371	12	96%
Brand #15	89%	96	474	570	131	92%
Brand #16	96%	91	2078	2169	131	84%
Brand #17	39%	85	18	103	1	
Brand #18	96%	69	1025	1094	167	93%
Brand #19	93%	68	462	530	119	97%
Brand #20	6%	67	2	69	0	15%
Total	**83%**	**3,154**	**40,179**	**43,333**	**3,571**	**80%**

Figure 9 Brand Coverage table at brand level

Brand Coverage (Item level)

Item	Predicted GVs	ASP (€)
Item #1	52.7K	32
Item #2	52.7K	199
Item #3	52.7K	249
Item #4	52.7K	75
Item #5	52.7K	75
Item #6	52.7K	77
Item #7	52.7K	184
Item #8	52.7K	183
Item #9	52.7K	184
Item #10	52.7K	24
Item #11	52.7K	24
Item #12	52.7K	24
Item #13	52.7K	17
Item #14	52.7K	17
Item #15	52.7K	24
Item #16	52.7K	17
Item #17	115.5K	0
Item #18	59.8K	
Item #19	53.7K	12
Item #20	106.7K	17
Grand Total	**58.9K**	**75**

Figure 10 Brand Coverage table at item level

SGM Brand Gap Not Covered by BC

Brand	Potential Score
Brand #1	480
Brand #2	432
Brand #3	431
Brand #4	429
Brand #5	426
Brand #6	426
Brand #7	423
Brand #8	422
Brand #9	420
Brand #10	420
Brand #11	420
Brand #12	419
Brand #13	417
Brand #14	409
Brand #15	404
Brand #16	404
Brand #17	403
Brand #18	401
Brand #19	398
Brand #20	398
Grand Total	416

Figure 11 Selection Gap Management at brand level

SGM Item Gap Not Covered by BC

Item	Brand	Potential Score
EAN Item #1	Brand #1	480
EAN Item #2	Brand #2	471
EAN Item #3	Brand #3	468
EAN Item #4	Brand #4	461
EAN Item #5	Brand #5	457
EAN Item #6	Brand #6	456
EAN Item #7	Brand #7	456
EAN Item #8	Brand #8	456
EAN Item #9	Brand #9	456
EAN Item #10	Brand #10	452
EAN Item #11	Brand #11	450
EAN Item #12	Brand #12	450
EAN Item #13	Brand #13	450
EAN Item #14	Brand #14	446
EAN Item #15	Brand #15	445
EAN Item #16	Brand #16	445
EAN Item #17	Brand #17	443
EAN Item #18	Brand #18	443
EAN Item #19	Brand #19	443
EAN Item #20	Brand #20	443

Figure 12 Selection Gap Management at item level

Top Trending Keywords This Week

Keywords	W43	W44	W45	W46	Trend
Keyword #1		1.1K	4.2K	21.0K	682%
Keyword #2		1.9K	12.7K	43.9K	501%
Keyword #3		2.0K	12.6K	39.1K	436%
Keyword #4	1.2K	3.6K	9.7K	22.7K	367%
Keyword #5		1.1K	4.6K	13.1K	358%
Keyword #6	2.5K	9.0K	17.3K	40.9K	327%
Keyword #7	2.2K	4.2K	5.9K	14.9K	266%
Keyword #8	0.9K	4.2K	7.0K	14.1K	249%
Keyword #9	1.7K	4.7K	13.0K	21.9K	240%
Keyword #10		2.8K	27.5K	50.9K	235%
Keyword #11		3.2K	10.1K	22.2K	235%
Keyword #12	0.9K	2.3K	7.0K	11.0K	222%
Keyword #13	0.7K	2.9K	8.5K	12.7K	215%
Keyword #14	1.1K	6.2K	17.2K	25.6K	214%
Keyword #15	1.1K	2.6K	5.3K	9.5K	213%
Keyword #16	2.5K	5.7K	18.8K	27.5K	205%
Keyword #17	1.9K	2.7K	7.9K	12.6K	204%
Keyword #18	4.3K	5.5K	9.4K	19.3K	203%
Keyword #19	3.2K	6.0K	10.0K	19.0K	197%
Keyword #20	1.5K	1.7K	2.4K	16.4K	186%
Grand Total	**25.7K**	**73.6K**	**211.0K**	**458.5K**	**279%**

Figure 13 Weekly top searched keywords table

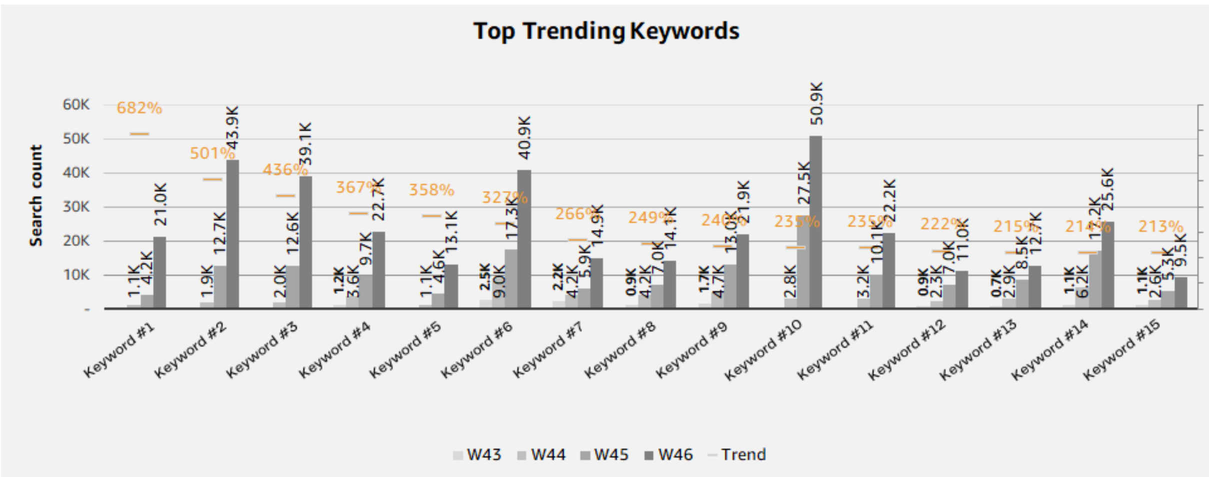

Figure 14 Top trending keywords of the current week

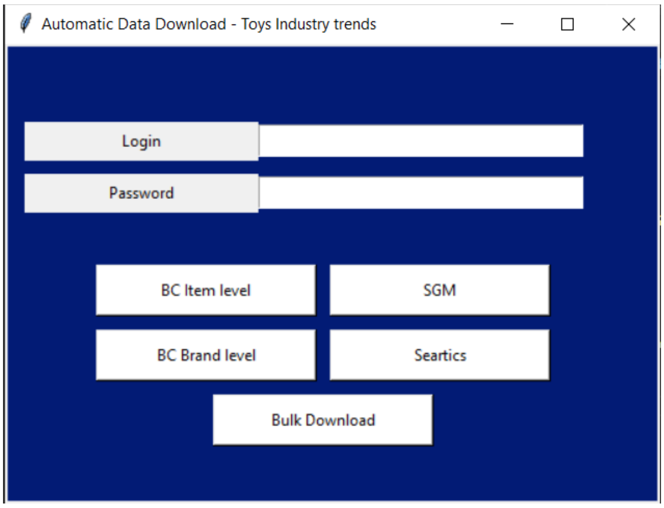

Figure 15 Python program to automatize data download

MP presence	EAN Count	GVs	Revenue
9	1.2K	1.6MM	€543.9K
8	5.7K	2.2MM	€867.1K
7	2.6K	3.3MM	€1.2MM
6	4.3K	5.1MM	€1.6MM
5	41.1K	8.7MM	€2.2MM
4	58.9K	19.4MM	€2.5MM
3	74.1K	22.8MM	€2.5MM
2	165.9K	36.4MM	€3.5MM
1	604.3K	62.4MM	€5.5MM
Grand Total	**958.0K**	**161.8MM**	**€20.3MM**

Figure 16 Impact of recategorization based on MP presence

Predicted Categories	EAN Count	Revenue	GV
Infants/Preschool	1,189	€2.3MM	3.6MM
Learning & Exploration	521	€1.4MM	1.3MM
Arts, Crafts & School Supplies	5,986	€582.2K	3.5MM
Outdoor & Sports Toys	1,340	€501.0K	3.4MM
Building Sets	1,061	€411.1K	1.6MM
Other	6,333	€411.0K	4.4MM
Dolls	215	€210.9K	429.1K
Vehicles	775	€195.5K	452.9K
Action Figures & Accessories	588	€178.6K	828.0K
Games & Puzzles	870	€149.5K	799.2K
Plush	241	€24.3K	362.5K
Grand Total	**19,119**	**€6.3MM**	**20.8MM**

Figure 17 Recategorization impact based on predicted categories

9. References

Amazon. (n.d.). *AWS docs, What is Amazon SageMaker?* Retrieved from https://docs.aws.amazon.com/sagemaker/latest/dg/whatis.html

AutoGluon AWS. (n.d.). *auto.gluon.ai*. Retrieved from https://auto.gluon.ai/stable/tutorials/multimodal/text_prediction/multilingual_text.html

AutoGluon AWS. (n.d.). *Auto.gluon.ai*. Retrieved from https://auto.gluon.ai/stable/index.html

Nick Erickson, J. M. (2020). *AutoGluon-Tabular: Robust and Accurate AutoML for Structured Data.*