Rebuild

30 years with

DATA VISUALIZATION
2022/23

BEATRIZ ANDRADE | R20191252
GONÇALO COUTINHO | 20221011
MARIANA LAVINHA | 20220565
MARIANA RODRIGUES | 20220627

Introduction

LEGO was just a humble building toy. Over the years, it has transformed into a cultural phenomenon that surpassed movies, video games, and even amusement parks. From the classic or the iconic LEGO to the more complex and technic sets of today, LEGO has become a long way in the past 30 years, continuing to conquer adults and children.
Our motivation was to build up a journey through the evolution of LEGO and discover how this toy has changed and grown, while exploring the fascinating history of LEGO and how it has continued to inspire and encourage creativity and imagination.

The idea came up after a few hours of research in a Lego set guide and because it has a lot of LEGO articles, we downloaded the dataset locally, but despite this, the official site is: https://brickset.com. Anyway, our full dataset is available in our GitHub page.

A visualization tool was created to analyze the evolution of the toy industry and demonstrate how visualization concepts and techniques can be used to transform data into an interactive representation that provides insight into the significant changes that have occurred in the world of Lego over the years. In the context of Data Visualization course, this project aims to present in an organised and funny way, a LEGO interaction Dashboard that shows and brings some evolutions and curiosities over the years. The project was implemented in Tableau and the idea in our Dashboard is to allow the users to selecta certain timeline and their favourite Set.

Data and Methods | Dataset Description

To build the visualizations presented in the dashboard, we started to preprocess the dataset in python, exploring the duplicated records and filling the missing values. Due to the fact that our dataset was very large and since we were working with Tableau for the first time, we thought that it would be a good approach to optimize the working time, creating different datasets, according to the analysis we wanted to perform. The following table shows all variables presented in our dataset and a brief description.

Variable	Type	Description
Number	Object	Set number ID
Theme	Object	Theme Name
Subtheme	Object	Subtheme Name
Year	Int64	Year when the set was launched
Set_name	Object	Set Name
Minifigs	Float64	How many minifigures the set has
Pieces	Float64	How many pieces the set has in total
Price	Float64	Price per set without the inflation effect and retail price
Value_new	Float64	Current price of the set (in new conditions)
Value_used	Float64	Price of a set second hand, after devaluation
Launch_date	Datetime64	Represents the date of the first commercial sale
Eol_date	Datetime64	"End-of-life" (eol) represents the final stage of a set's existence
Category	Object	If a record is a set or minifigure
Price_unadjusted	Float64	Target price before inflation adjustment
Value_change_per	Float64	The change of a value between the retail price and the current price as a "new set".

Some decisions considering the dataset were made due to some limitations. Undesired fields for the purpose in-hands were deleted, like records concerning 2022, since the information was incomplete, turning our analysis inconsistent. Regarding the end-of-life dates (eol), we considered the dates between 31-12-1996 and 31-12-2020, since the eol dates before that were not available and after that we have incomplete information. The rest of the features were not influenced by this decision and our analysis focuses on the past 30 years (1991-2021) of Lego's company.

Data and Methods | Technical Aspects

When creating the final dashboard with the visualizations, different data visualization approaches and concepts acquired were taken into consideration.
Firstly, for the project implementation, we used the jupyter notebook to import the datasets and perform some cleaning and treatment of the data. Afterwards, in order to produce the visualizations, we used Tableau as the data visualization tool since it usually had fast performance, has a user-friendly interface and enables the creation of interactive dashboards.
Secondly, to ensure the effectiveness of the graphical representations, the validation of our dashboard was performed. This crucial step allowed us to verify that the dashboard accurately represents the underlying data an effectively communicates the intended message to the audience. We started by testing the dashboard functionality by checking the interactivity of the filters and other features. Furthermore, through some colleagues who do not have knowledge in this area, the dashboard was tested to ensure that the intended message was effectively communicated.
Also, for the project development we used a GitHub repository, were all the datasets used are available https://github.com/BeatrizVargem/Data-Visualization.

Data and Methods | Visualization and interaction choices

Growing up with Lego, we believe that approaching it from a Data Science perspective adds an extra layer of fascination. In order to ensure that all the information was accessible and easy to understand, we decided to create a dashboard that only by looking at it, the user could recognise it was about Lego, so we decided to apply their color scheme throughout the project.
Also, in order not to overcomplicate the dashboard and hinder the audience visualization we included 5 graphical representations, as our overview Tab. Hence, we opted to invest more in the interactivity and give the user freedom to explore more in-depth certain sets. Regarding the first chart, using a scatterplot, we can analyze the evolution of the number of sets, as well as the number of themes by using different circle sizes for this second feature. The second chart, that shows the average price evolution per set over the years, we considered an interesting analysis, to understand why the price suffered sudden and sharp descents and ascents. To achieve that, we found that Lego suffered a major crises, leading it almost to bankruptcy. In 2004, the Danish company was on the brink of financial collapse. Ten years later, Lego surpassed the american giants Mattel and Hasbro to become the largest company in the sector. In the first half of 2015 Lego continued to increase its turnover and Christmas, that is the crucial season for the toy industry, hadn´t arrived yet. Its revenues increased by 23% and profit soared by 27%. These events are evidently reflected in the price chart as for example in 2004, the average price per set was $22.5 by taking one of the lowest recorded values. To contrast, in 2014 it rose to $33.18. Since 2014 the average price per set has been increasing exponentially. In conclusion, both show a positive trend. While the number of themes has more than doubled since 1991, the total number of sets has grown by just under 6 times.

The pieces chart, shows the evolution of pieces' number and price. The conclusion is that the number of pieces per set is increasing, while the average price per piece is decreasing. Which means that in the beginning, the sets had less pieces but more expensive pieces and now, the sets have much more pieces, but cheaper per piece. To kill some curiosities that arose during the research and development of the work, we decided to know what were the top 5 record-breaking set by piece count and we conclude that The World Map is the largest set of all time, with 11.695 pieces. We also can see in the pie chart, that "pack" is the most frequent word in set names.
Furthermore, some details that conveyed information about each set of Lego, were created for the user to be able to select the Set Name and be able to see what theme it belongs to, the launch year, its number of pieces and minifigures, the price and the percentage change since the set was discontinued.

Results and Discussion

During the development of the visualization product, we incorporated the data visualization techniques and concepts learned throughout the course. At first, the data was organized into items and attributes, forming a structured dataset in the form of a table. As for the attribute types, we dealt with both quantitative, nominal and ordinal types, in all ordering directions. Regarding the marks throughout the product, there are points - Scatterplots, lines - Line Plot and Bar Plot, and also Areas - Pie Chart.
In what concerns the channels used, we employed horizontal position, color – picked with different luminance and saturation, different shapes and size – through length and area. In making decisions related to this matter, the principles of both expressiveness and effectiveness were always taken into consideration, as the most highly ranked channels were utilized. Through the interaction techniques, the user can, by placing the cursor over the graphic, access more detailed information about the plot in question. Preventing doubts or misinterpretations in this way.

The user has the ability to interactively select and filter the information they wish to access and evaluate, such as details according to a chosen set and overview Tab graphs filtered by year range. Example of that, is on the Set Name where the user can choose his favorite one, get the launched year and analyze lego's profile overview at that time.

Conclusion

We can conclude that this project resulted in an interactive dashboard that provides a wide range of interaction and some details regarding Lego items. During the process, we were able to create a variety of visualizations that offers the viewers relevant information. This way we achieved our initial goals, analyzing data to support reasonning and communicating information to users about a topic that would reach people with nostalgia.

Regarding the limitations, there were some that hindered and slowed down the progress of our visualizations. Foremost, the dataset used with the information about Lego had a lot of missing data, which dificults the creation of some visualizations and requires more time dedicated to the preprocessing step. Also, some errors and inconsistencies were found throughout the analysis, such as the set name being repeated even though the launch year was different or the same sometimes. However, this issue can complicate the analysis, since the set name should be different for each set for a better differentiation. Furthermore, the dataset found lacked information on interesting features for the analysis, in particular the amount sold of each unit, the countries with the most sales and the fluctuations of the number of sales and price throughout each year.

Due to the limitations that appeared during the process, some future work would be essential to improve the current visualizations. Firstly, it would be interesting to find more datasets about other information that could be relevant to compare with the Lego data. Moreover, adding new information on the recent years, would complete our dashboard and guarantee up-to-date information about this topic.

References

Soldado, C. (2015, September 14). *Lego: De Quase Falida a Líder da Indústria de Brinquedos.* PÚBLICO. Retrieved April 9, 2023, from https://www.publico.pt/2015/09/14/economia/noticia/lego-de-quase-falida-a-lider-da-industria-de-brinquedos-1707694

Lego History. LEGO® History - LEGO.com IN. (n.d.). Retrieved April 9, 2023, from https://www.lego.com/en-in/history/

Dashboard integration: Theory. campus.datacamp.com. (n.d.). Retrieved April 9, 2023, from https://campus.datacamp.com/courses/dashboard-design-concepts/dashboard-success-and-integration?ex=10

Cotton, R. (2022, April 22). *Data Visualization Cheat Sheet.* DataCamp. Retrieved April 9, 2023, from https://www.datacamp.com/cheat-sheet/data-viz-cheat-sheet

Pritika Bahad et al 2021 IOP Conf. Ser.: Mater. Sci. Eng. 1099 012019

Oliveres, C. (2022, June 26). *Lego analysis dataset with Power Bi.* Medium. Retrieved April 9, 2023, from https://medium.com/@Christophe_Ol/lego-analysis-dataset-with-power-bi-b5dad228cb41

Finding lego treasure using data analytics: Kids in data. KiD. (n.d.). Retrieved April 9, 2023, from https://kidsindata.com/blog/lego-treasure-with-data

Dashboards. Tableau. (n.d.). Retrieved April 9, 2023, from https://www.tableau.com/learn/get-started/dashboards