August 2023

**DATA ANALYSIS PROCESS FOR PREDICTING THE CONVERSION RATE OF PRODUCTS**

Luiz Pedro Marques Filho,

Computer Science Graduating

Pontifícia Universidade Católica de Minas Gerais

Poços de Caldas, MG - Brazil.

August 2023

**Summary**

1. **Introduction** ………………………………………………………………………………………. 3
2. **Studying marketing metrics** …………………………………………………………………… 4
3. **Selecting main metrics** ………………………………………………………………………… 4
4. **Parent-Child Asin (+)** …………………………………………………………………………… 6
5. **Searching for datasets** ………………………………………………………………………… 7
6. **Dataset Generator** ……………………………………………………………………………… 9
7. **Data Analysis Process** ……………………………………………………………………….. 10
8. **Dashboards using PowerBI** …………………………………………………………………. 14
9. **App & Database Development** ………………………………………………………………. 15
10. **Conclusions** ……………………………………………………………………………………. 16
11. **References** ……………………………………………………………………………………… 17
12. **Introduction**

A large seller has his own online business using the Amazon platform to advertise his thousands of products. In order to analyze how his campaigns are doing in relation to how much his potential customers actually make a purchase, he starts studying e-commerce statistics and manages (with the help of a data specialist) to build a set of data that contains your business analytics for each of his products. One of these metrics, the conversion rate, allows the seller to know how effective their campaigns are and how much their buyers buy when clicking on their ads. However, he notices large differences in the same taxa for different items, which indicates that the taxa are related to other numbers that may or may not be being mapped by him.

With a new batch of products to be launched by the end of the month, the seller would like to forecast the conversion rate of these new items in order to make assertive lessons on whether their campaigns should undergo major structural changes or just small adjustments to some details. Therefore, it aims to build a machine learning model that is capable of making this prediction with at least a certain level of significance.

The following assignment brings together a mix of theoretical knowledge, technical study, and practical applications for this scenario. Using the data analysis process and based on the Amazon sales platform, I sought to develop my knowledge and understand how the company's day-to-day operations work and the challenges of seeking to increase performance in digital marketing companies.

Throughout the task, I had to deepen studies in the area of ​​digital marketing, not mainly referring to the metrics used, but also covering topics such as granularity, parent-child asin, among others. In terms of technology, I had to delve deeper into concepts of machine learning, artificial intelligence, Python language, database and cloud service platforms. It is important to highlight that throughout the task, such as obtaining randomly generated fictitious data (scarcity of this type of content on the internet), the main objective at the end of the analysis is not a model evaluation metric that can predict perfectly (or even accurately). satisfactory) the desired data. The main point here is to be able to understand the concepts and how to use them correctly so that it is possible to increase efficiency in the company's day-to-day operations.

1. **Studying marketing metrics**

To acquire outstanding results in your online sales, you have to know where to look so you can understand where/when you can improve in order to increase your purchases. First of all, it's crucial to determine which metrics we have to measure instead of tracking dozens of them. For that, we pick those which have the most impact in your business sales. It’s also important to not make confusion between metrics and KPIs (**Key Performance Indicators**). There is a little but significantly difference that we are going to explain below:

Ecommerce metrics are quantitative measurements that provide insights into various aspects of your online store's performance. These metrics help you understand how your business is performing and where improvements might be needed. Key performance indicators are specific metrics that are selected as the most important indicators of an ecommerce business's performance against its goals and objectives. KPIs are more strategic and are used to track progress toward specific targets. **While many metrics can be considered KPIs, not all metrics are necessarily KPIs.** According to the magazine The Good, you should ask yourself three questions to understand deeper what you are trying to analyze: **If this metric changed, how big would the impact be on my company?** That is, if the values of these metrics changed significantly, how is that gonna affect your business in general? The second one is, **Will Improving this metric contribute to our strategic goals?** That is, if you invest more time, resources and effort into enhancing this particular certain metric, will it help you to move closer to achieving your strategic objectives? And the last one, **Is this a metric that will also improve other metrics?** So, if you invest those resources to improve some metric, would that have influence on other metrics like a domino effect? (If your answer is yes, this is probably a KPI, by the way).

1. **Think Selecting the metrics**

Before we talk about which metrics we should select, it is important to say that we did this based on an Amazon e-commerce context about products of large durability (Electronics), from the perspective of a seller. So, metrics that measure data outside the platform, metrics that the platform itself should track and metrics about non-durable goods will not be mentioned here. Keep that in mind for better understanding.

**Conversion Rate (CR)** - percentage of website visitors or users who take a desired action or "convert" in response to a call-to-action (CTA). Conversion rate is determined by dividing the number of conversions by the total number of visitors who were given the opportunity to take the action.

**Click-Through Rate (CTR)** - measure the effectiveness of a marketing campaign, indicating the percentage of people who clicked on a specific link, advertisement, or call-to-action (CTA) out of the total number of people who were exposed to it.

**Cost-Per-Click (CPC)** - the cost an advertiser pays each time a user clicks on their advertisement.

**Return on Investment (ROI)** - evaluate the profitability and effectiveness of an investment relative to its cost.

**Visits to the Website** - number of times people access and interact with the pages of your website. This metric can be divided into:

**Unique Visits:** number of distinct individuals who have accessed the site.

**Time on Page:** average amount of time a visitor spends on a page of your site.

**Devices:** It can be useful to know if visitors are accessing the site from each device.

**Geographical Location:** Knowing which region or country visitors are from

**Return On Advertising Spend (ROAS)** - amount of money spent on an advertising campaign to sell the product. Dúvida: Campanhas dentro da amazon?

**Average Order Value (AOV)** - measure of how much your customers typically spend on a single order from you.

**Cost of goods sold (COGS)** - it represents the direct costs associated with the production or procurement of the goods or services that a company sells during a specific period

**Add to Cart Rate:** It represents the percentage of visitors who place at least one item in their shopping cart while visiting your website.

**Customer Reviews:** collecting and analyzing opinions, feedback, and ratings provided by customers after purchasing products or experiencing services.

Besides this, in this context, it is also important to track other values that are not necessarily related to the product itself, such as the ASIN (product identification code on Amazon), tax and freight values, campaign’s id’s and the date of the purchase.

Formulas of each metric mentioned

1. **CR** = (number of conversions / number of total ad interactions) / 100
2. **CTR** = number of people who clicked on the ad / number of people who saw the ad
3. **CPC** = total cost of your clicks / total number of clicks
4. **Net Income** = Money “Gained” - Cost of investment
5. **ROI** = (Net income / Cost of investment) x 100
6. **ROAS** = revenue attributed to your ad campaign / total cost of that campaign
7. **AOV** = total revenue / total of orders
8. **CartRate** = sessions a user added a product to cart / site sessions during that period

| Conversion | CTR | CPC | ROI |
| --- | --- | --- | --- |
| AOV | ROAS | Reviews | ASIN |
| Taxes | Freight | COGS | Campaign ID |
| Visits | Day | Week | Year |

1. **Parent & Child ASINs**

ASIN stands for "Amazon Standard Identification Number." It is a unique identifier assigned by Amazon to products in its catalog. ASINs are used to identify and manage products on the Amazon platform. Each product listed for sale receives a distinct ASIN.

Parent ASIN and Child ASIN refer to relationships between different variations of the same product. These variations are typically seen in product categories like clothing, electronics, or any other category where products come in different sizes, colors, or configurations. This system is often referred to as Amazon's "Variation Relationships."

**Parent ASIN:** The Parent ASIN is the main product listing in a variation group. It represents a general product that has multiple options or variations. For example, a Parent ASIN might represent a T-shirt in different sizes and colors.

**Child ASIN:** Child ASINs are the individual variations of the Parent ASIN. They represent specific versions or options of the main product. For instance, different sizes (small, medium, large) and colors (red, blue, green) of the T-shirt would each have their own Child ASINs. Each Child ASIN is linked to the Parent ASIN to show that they are variations of the same product.

****

Figure 1 - ASIN Examples taken from each link of the dataset

1. **Searching for datasets**

When selecting the metrics, the next step is to find a public dataset that allows me to develop the analyses and draw conclusions at the end of the process. However, in addition to being a niche area, selecting metrics in such a specific way makes this search very difficult. During the research, several datasets on different platforms were analyzed and none gathered the information that we would like to work on in one. Below is a list of possible datasets to be used in this task.

1. Electronic Store Sales Data

*(*[*https://www.kaggle.com/datasets/saumaydhaundiyal/electronic-store-sales-data*](https://www.kaggle.com/datasets/saumaydhaundiyal/electronic-store-sales-data)*)*

1. Daily website visitors

*(*[*https://www.kaggle.com/datasets/bobnau/daily-website-visitors*](https://www.kaggle.com/datasets/bobnau/daily-website-visitors)*)*

1. eCommerce behavior data from multi category store

*(*[*www.kaggle.com/datasets/mkechinov/ecommerce-behavior-data-from-multi-category-store*](http://www.kaggle.com/datasets/mkechinov/ecommerce-behavior-data-from-multi-category-store)*)*

1. Brazilian E-Commerce Public Dataset by Olist

*(*[*https://www.kaggle.com/datasets/olistbr/brazilian-ecommerce*](https://www.kaggle.com/datasets/olistbr/brazilian-ecommerce)*)*

1. The Klarna Product-Page Dataset

*(*<https://registry.opendata.aws/klarna_productpage_dataset/>*)*

1. Amazon Product Dataset 2020

*(*[*https://www.kaggle.com/datasets/promptcloud/amazon-product-dataset-2020*](https://www.kaggle.com/datasets/promptcloud/amazon-product-dataset-2020)*)*

1. E-Commerce Data

*(*[*https://www.kaggle.com/datasets/carrie1/ecommerce-data*](https://www.kaggle.com/datasets/carrie1/ecommerce-data)*)*

1. Sales Conversion Optimization

*(*[*https://www.kaggle.com/datasets/loveall/clicks-conversion-tracking*](https://www.kaggle.com/datasets/loveall/clicks-conversion-tracking)*)*

1. Sales Store Product Analysis

*(*[*https://www.kaggle.com/code/faridrizqis/sales-store-product-analysis*](https://www.kaggle.com/code/faridrizqis/sales-store-product-analysis)*)*

As has been said, despite having found several very good data sets that allow for quality analysis, none of them gathered the information that was defined in the previous stages of the task. Therefore, it was decided to work with a simpler dataset and use the technique of adding dummy columns to simulate the data by generating random data. The chosen dataset actually contains data on various categories of products within the Amazon platform sold during the year 2023. This set originally has the following columns expressed in the table below and can also be accessed through the Kaggle platform via the link ([*https://www.kaggle.com/datasets/lokeshparab/amazon-products-dataset*](https://www.kaggle.com/datasets/lokeshparab/amazon-products-dataset)).

| **Amazon Products Sales Dataset 2023** | |
| --- | --- |
| **name** | **description** |
| *name* | The name of the product |
| *main\_category* | The main category of the product belong |
| *sub\_category* | The main category of the product belong |
| *image* | The image of the product look like |
| *link* | The website reference link of the product |
| *ratings* | The ratings given by amazon customers |
| *no of ratings* | The number of ratings given to this product |
| *discount\_price* | The discount prices of the product |
| *actual\_price* | The actual MRP of the product |

Within the large block of datasets available within the "Amazon Products Sales Dataset 2023", the specific dataset "All Electronics" was chosen, which brings together about nine thousand products available on Amazon, all of an electronic nature such as cell phones, devices, notebooks , etc. The 2 and 3 figures below represent some basic descriptions of the data initially.

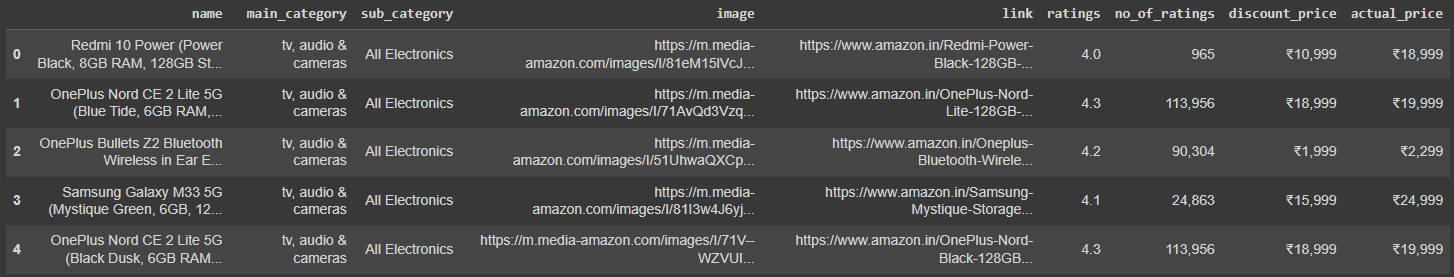
Figure 2 - Dataset header showing five records.

Figure 3 - Statistical representations over the raw dataset.

1. **Datasets Generator**

As stated in one of the previous topics, from the dataset search it was not possible to find any public dataset that was compatible with the metrics found and analyzed during the search. Thus, the only way to carry out an analysis in this context would be to simulate data through their random generation. Therefore, techniques for creating fictitious values ​​were used using some external libraries of the Python language (Pandas and NumPy, mainly).

However, several obstacles were encountered at the time of this generation, such as the difficulty in finding distribution forms compatible with existing data. For example, the only column with some resemblance to a normal distribution (making the process easier) is the 'rating' column, as its values ​​are within a limited range. However, all other numerical columns that already exist do not have any upper limit (because they are prices and absolute values) and are more similar to a uniform distribution. This uniform distribution makes each value equally likely to happen, which can confuse the generation algorithm.

It is worth mentioning that some care was taken so that the simulated data approximated a real situation. The addition of outliers in this process, defining values ​​based on other columns already established, and defining upper limits for some of the columns were some of the concerns that prioritized that the fictitious data set did not prove to be absurd, completely preventing the analysis. However, as the values ​​are completely random, several obstacles were encountered at the time of this generation, such as the difficulty in finding distribution forms compatible with the existing data. For example, the only column with some resemblance to a normal distribution is the 'rating' column, as its values ​​are within a limited range. However, all other numerical columns that already exist do not have any upper limit (because they are prices and absolute values) and are more similar to a uniform distribution. This uniform distribution makes each value equally likely to happen, which can confuse the generation algorithm.

This entire data generation process is documented in the “Data Analysis Process” section and it is also possible to access the details by opening the code article available next to this task or through the link shared in the Collab tool

([*https://colab.research.google.com/drive/1Y4uF6QXQn8pz80H5eKtoxD9vzhofVEwI?usp=sharing*](https://colab.research.google.com/drive/1Y4uF6QXQn8pz80H5eKtoxD9vzhofVEwI?usp=sharing)).

1. **Data Analysis Process**

In order to make predictions about the conversion rate of products sold, the entire data analysis process was based on treating the data and adding columns of randomly distributed dummy data so that we can build a machine learning model to apply on such a set. In the first step, after importing the libraries and the .csv dataset into a DataFrame in pandas, statistical and data type analyzes were done so that we have an idea of ​​what to do initially. Verifying that the fields that involved values ​​were in Indian rupees, it was necessary (for reasons of better visualization) to convert these values ​​to “reais” (converting them to float first) and making the proportion between the currencies (1 rupee is equivalent to approximately 0.060 reais). After transforming all the columns that needed to be converted from text to numbers, we chose to create a new column (*'percent\_discount'*) to visualize the percentage of discount between the real price of the product (*'actual\_price'*) and the price with the applied discount (*' discount\_price'*).

An important moment of the analysis was when we observed that the **ASIN** (product identification code within the Amazon platform) was inside each product link, information which we had in the *'link'* column. Therefore, in order to create a column with the ASIN of each item, we used Python regular expressions to create a filter that identified the 10-digit code of each product and placed it respectively within a new column (*'asin1*). The curious thing was that we found two identical ASINs, which could mean that maybe it was a product with a Parent/Child ASIN scheme.

Time to remove the null values, it was identified that the dataset in its total had **598 lines with some null value** (484 prices and 119 ratings). Bearing in mind that our dataset, when imported, had 9600 records, this number of nulls represents only about 6% of our entries. So it's okay to remove instead of trying to replace them.

After applying the necessary changes to the original data in our set, we now had to insert dummy columns (columns that were not in the original set) that represent the digital marketing metrics studied at the beginning of this task. In addition, it was also essential to create random data to fill all these attributes uniformly, avoiding data bias. Another important point was the addition of outliers (points “outside the curve”) so that the data had greater similarity with a real business situation. Outlier rates varied by column.

The first generic column created was the number of views, which in this case represents the number of people who clicked on a certain ad for each of the products. The only specification was that the number of views had to be greater than the number of ratings.

The second column was to generate the number of conversions, in this situation representing the number of people who actually bought a certain product (remembering that a conversion does not necessarily need to be a purchase in other situations). The specification made was that the number of conversions should be in the first three data quadrants of the number of views column, that is, the random value generated in 95% of the cases should be between 1% and 75% of the number of views, while outliers could cross this mark being up to 100%.

The third column to be created was the cost of the click, that is, how much each advertising seller pays each time a user clicks on their ad (that is, this column will later be closely related to the number of views/clicks). Knowing that this value usually varies between 0.1 and 1 (disregarding currency devaluation and other economic factors), the randomly generated data also have this range for 99% of the cases, while the outliers with a rate of only 1% vary between one and two.

The fourth fictitious column was the revenue generated, that is, how much value each product generated in total. This number would be needed to later calculate the net income of each item in terms of profitability. Here, random values ​​were created that used the value of each product as a basis for determining revenue. Another important detail was that, for didactic purposes, the maximum revenue value for each product was limited to its own value times five hundred. So revenue varies between *'discount\_price'* times 5 and 500, with outliers ranging from 500 to 1000 in just 1% of cases.

Similarly to revenue, the fifth column created refers to the total expenses for each product (ranging from the campaign to the cost of the click), which in theory should not be greater than the revenue for the seller to be able to obtain a profit margin. These values ​​typically vary (95% of the time) between 30% and 70% of the revenue generated, while the outliers vary between 70% and 90% of the revenue in 5% of the cases.

Having generated these last two metrics (*'total\_revenue'* and *'total\_expenses'*), it is now possible to calculate the net income of the product. To do this, the new column must only subtract the total expenses from the amount of revenue generated.

Similarly, the metric we call cost per click (CPC) can be easily calculated by multiplying the cost per click values ​​by the number of views in each row.

The next metric (conversion rate) can also be obtained according to the dummy columns we created. To do this, we must take the number of conversions obtained for each product and divide it by the respective number of views, resulting in a result usually between 0 and 1 (closer to 1, the better).

Finally, using the net result values ​​(*'net\_income'*) and total expenses, it is possible to calculate one of the most important metrics already analyzed at the beginning of this task: the ROI (Return on Investment). The metric makes it possible to analyze the campaign's profit in relation to its investment cost. To calculate it, just divide the net income by the total expense (the higher, the better).

Having created all these columns, now let's draw some statistical conclusions about the data using some functions. Let's start with a correlation matrix that aims to analyze the relationship between each of the variables with each other. By having created the columns and knowing the production process of each one, the vast majority (we can say all, in fact) of the correlations here become quite obvious, as the relationship between the Net Income and the Product Value, for example. For documentation purposes, as our objective is to predict the conversion rate, it is important to highlight that (obviously) the variable that indicates the number of conversions has significant impacts on this metric. And regarding the conversions metric itself, we have the number of views with a high correlation index associated with this number.

After this analysis, it is important to make some simple statistical measures such as the mean, median, standard deviation, minimum-maximum value of each variable (mainly those that were randomly generated in order to identify grotesque discrepancies in the data set. In addition, it is necessary to analyze the distribution of each column in order to predict whether the number of outliers is really within the forecast and whether the data in general are consistent with what was constructed.

Once this is done, it is important to remove all columns that are not used when training the model, being mandatory to remove at least the text columns. However, it is important to save them somewhere so that, after analysis, these columns can be incremented to the data frame again and exported to a visualization tool or to the database.

Once removed, we only have the columns that will actually be trained in order to predict the dependent variable, which must also be separated from the rest of the columns at this point. Using functions from external libraries of the language, we can separate the data set into four variables: independent training and testing, in addition to the dependent variable training and testing.

Another important step in data pre-processing for better results of the regression model is to scale or normalize them so that all records are in the same range, preventing the model from getting confused by prioritizing large values ​​and giving less significance to small values. For example, without normalization, product values ​​(high) would have greater “importance” in the algorithm than other values ​​as important as metrics (CPC, Conversion, etc). Therefore, in this task we used the scaling technique (**Standard Scaler**) from the SkLearn library where, following the formula, all values ​​must be in a range between minus one and one, with zero being the mean.

Finally, with the data all clean, pre-processed and subject to the application of the learning model, we can use the same library mentioned above to apply a Multiple Linear Regression. This technique consists of modeling the relationship between a dependent variable (conversion rate) and several independent variables (called predictors). The goal here is to find values ​​that best fit the coefficients that multiply each variable, which are later added to find the value of the dependent and generate a general equation (as in Simple Linear Regression).

Almost at the end of the data analysis process, the next step is to evaluate the results generated by applying the model, that is, to compare the “real” data we had (“real” because they were randomly generated, impairing the fidelity of the case with reality) with those same values, but this time generated by the machine learning algorithm. It is possible to generate a table with these two columns (actual and predicted) and compare them manually, which should take absurdly more time than simply using evaluation metrics already available in external libraries. Here, we use four of them that are recommended for regression cases. Mean Squared Error (**MSE**) is basically the average between these two sets, being better whenever it's closer to zero (ps: this metric is more sensitive to outliers). The Root Mean Squared Error (**RMSE**), which is simply the root of the previous metric, allows for more interpretable values ​​because they are in the same metric as the dependent variable. The Mean Absolute Error (**MAE**) is the average of the absolute values ​​of the errors between the prediction and the reality, being less sensitive to outliers. The Coefficient of Determination (**R²**), which varies from 0 to 1 and indicates the proportion of variability in the target variable that is explained by the model, that is, the closer to one, the better.

The last step in the process is sharing. In this specific case, we are now going to export this data to a .csv spreadsheet (same format as they came in the beginning) to later load them into a database and access this database with the function of retrieving the data and building the dashboards that will present the results. In this way, we can have a new dataset that merges all the columns initially obtained and now with the addition of all the randomly generated dummy columns (remembering that we must bring back the columns that had been removed at a time before).

As mentioned in the introduction to this task, the objective here is not to obtain the best evaluation metrics (which is visible when analyzing the metrics derived from the model), as the authority of the data makes it very unfeasible for this to happen. However, knowing how to apply techniques, metrics and knowing the concept behind all these steps and the concept behind the digital marketing area itself is what matters right now.

1. **Dashboards using PowerBI**

To build the dashboards integrated into the application, we used **Power BI**, a powerful data visualization and analysis tool developed by Microsoft (available at https://powerbi.microsoft.com). The software allows individuals and organizations to collect, transform and present data in a visually impactful and interactive way, making it easier for both the data presentation and the listeners to understand.

During the task, only two dashboards were built for the dataset in question. The first, related to the rating prediction analysis, consists of a bar graph that shows the number of ratings for each product rating. That is, how many products had an evaluation equal to one, two, three, etc. A slider was also inserted that allows filtering this amount of evaluations according to the price of the product. The complete dashboard can be accessed through the web application, but below is an image for initial visualization (Figure 4).

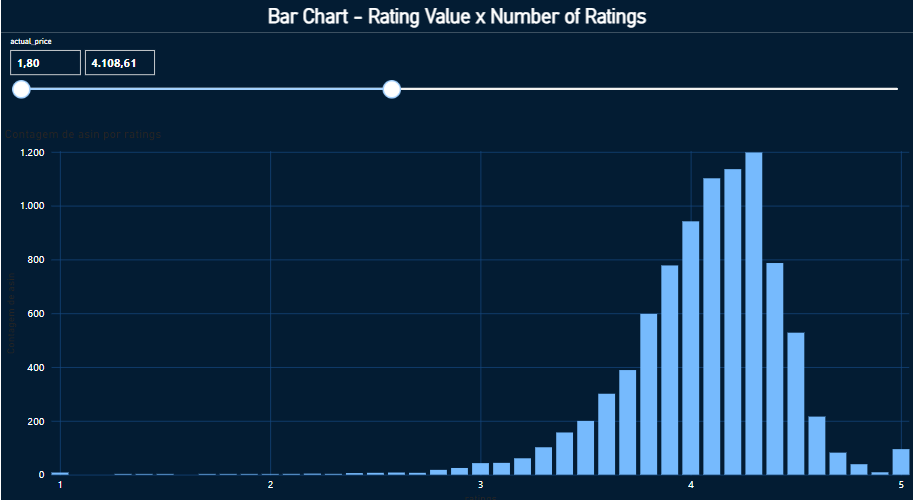


Figura 4 - Dashboard - Rating Prediction Model (Made in PowerBI)

1. **App & Database Development**

A simple system was built using HTML, CSS and Javascript on a local server in order to integrate the dashboards built in PowerBI and display them interactively in the application in question. The data has been placed in a local database using XAMPP (tool to make a local server in your machine). That data was taken by PowerBI into its system, and it was necessary to export the dashboard to the cloud, and integrate it using the “Developer Playground” tool, available in the service itself. This tool formats the dashboard in question in IFrame format so that it can be inserted in HTML format within the desired page (Figure 6). Furthermore, the dashboard must be published with a public configuration, so that everyone who accesses the system can view it without necessarily having private access to the item.

| Product Analysis | |
| --- | --- |
| *images* | folder with all the images inserted on the site |
| *dashboard1.html* | ratings x number of ratings |
| *dashboard2.html* | another dashboard |
| *script.js* | the authentication Supabase API |
| *style.css* | stylization of the pages |

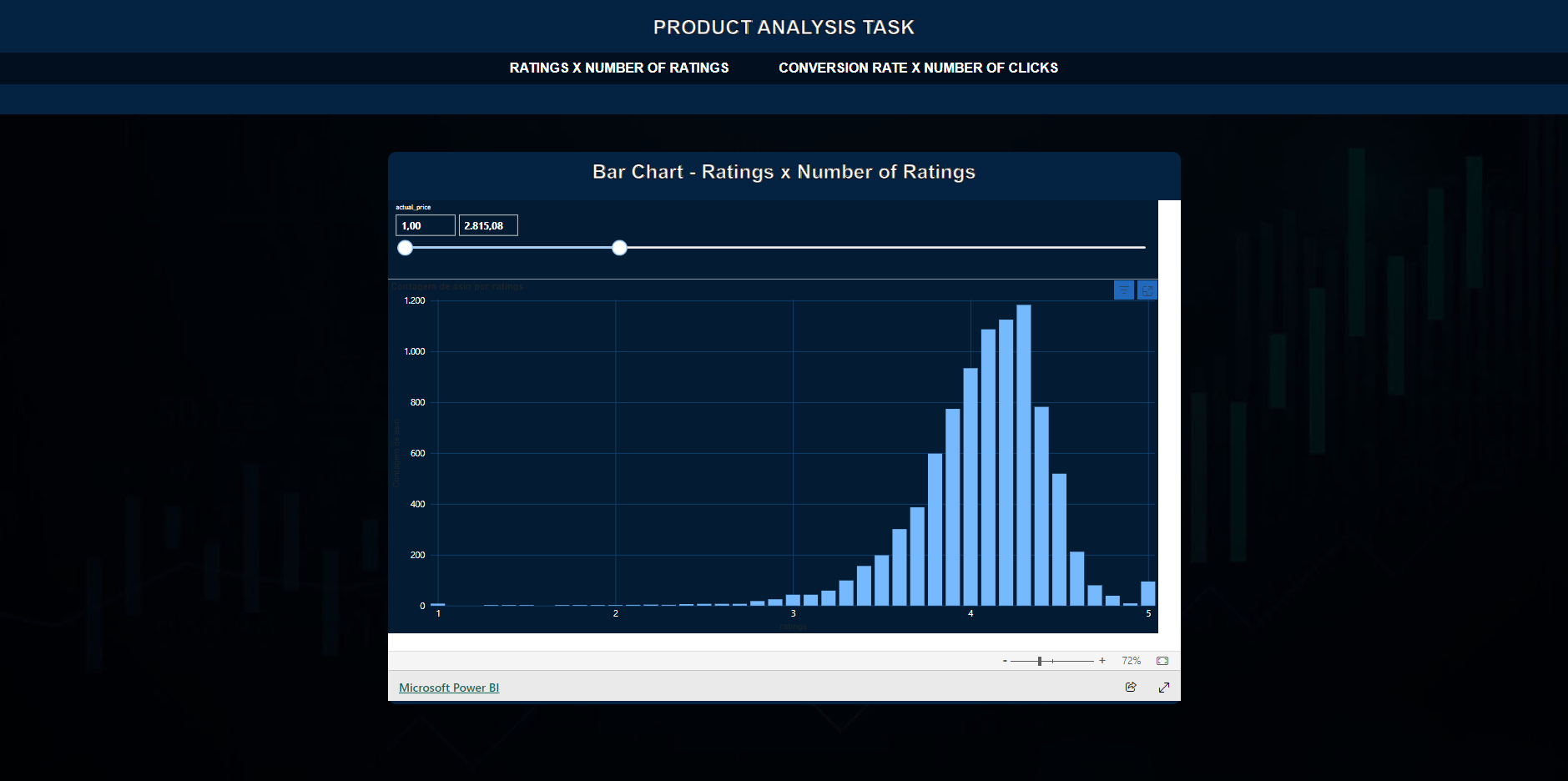


Figure 6 - System using html, css, javascript and PowerBI integrated.

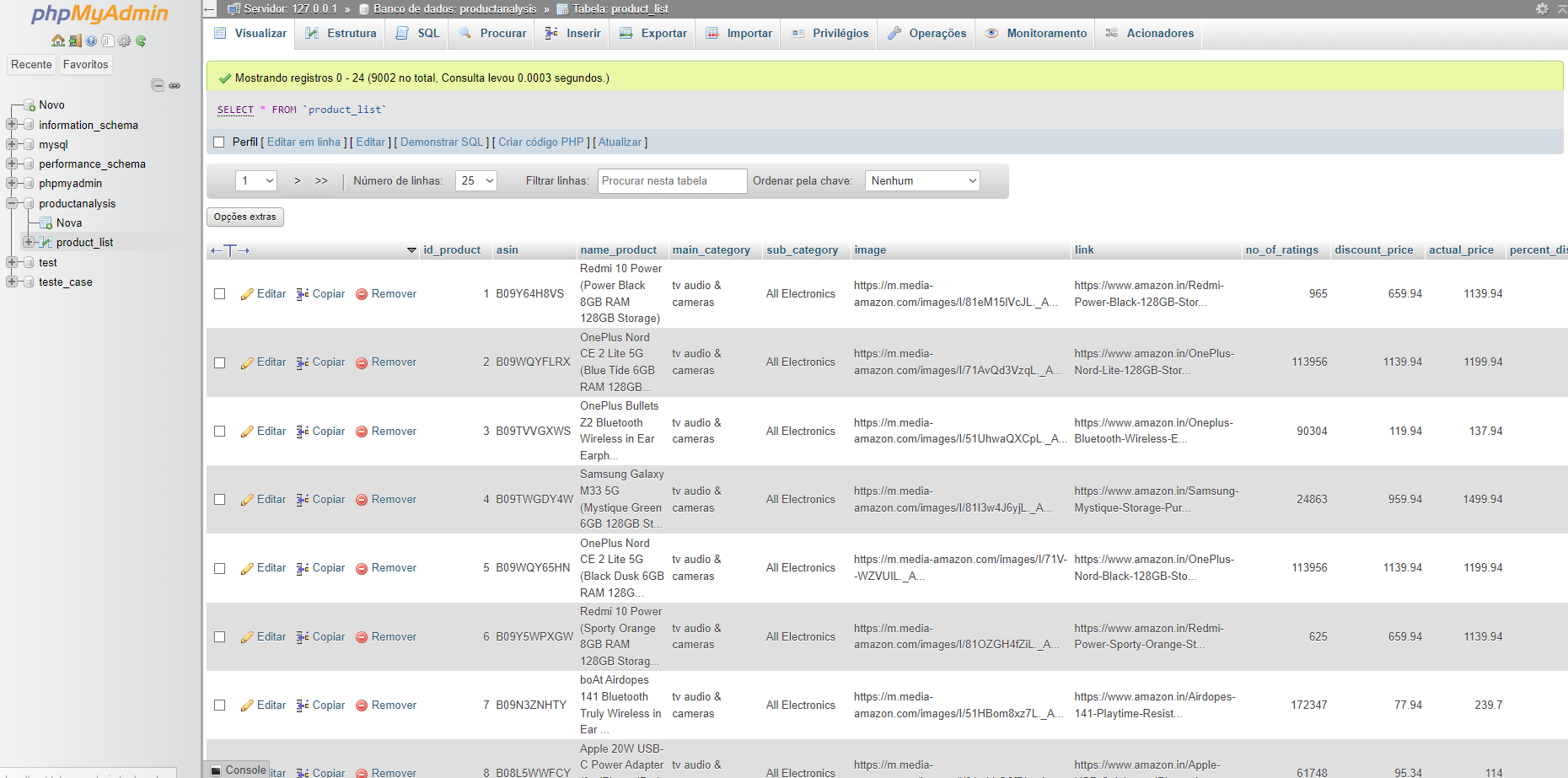


Figure 7 - Local database using XAMPP

1. **Conclusion**

Throughout this experience, I have learned a great deal about marketing digital and its metrics. Besides this, I’ve also learned about Amazon Advertising Methods and how to increase sales in e-commerce platforms. This journey has provided me with valuable insights into the data analysis area and improved my skills in Python language and its external libraries, database management and queries, etc. It has expanded my knowledge in ways I had not anticipated, and I am excited about the opportunity to continue learning and growing in this area.

I recognize that there is always room for improvement, and this experience has highlighted a few areas where I can focus my efforts. Learning how to use API’s properly is something that I have to explore deeper at this point, such as focus on my visualization tool skills. I am committed to honing my skills and knowledge in these areas to become more proficient and effective in the future, looking forward to learning so much more in the next weeks and months.

On the positive side, I believe I have done some good work in my data analysis process considering the resources that I had in terms of datasets, which I am particularly proud of. These successes have given me confidence in my abilities and will serve as experience for next projects in the future. However, I also encountered difficulties, including some issues during my system integration, when I faced the Supabase API not working when I was trying to connect it to my visualization tool. I really believe that, with more time available to understand the error, I can perform much better at this aspect. These obstacles were at times frustrating, but they also provided valuable learning experiences.

For the future, I believe that I would like to try to deepen my knowledge in the researched development tools and apply them in the next projects. Despite not having much knowledge, the use of APIs was also an area that added new knowledge to me during this task, which I am very grateful for. On the other side, I think that the area I least enjoyed working in was actually the system construction using HTML, CSS and Javascript, despite this being an important part of the process, because that made it possible to show the dashboards online.

In summary, this experience has been a huge and significant step in my professional development and career. It has taught me important aspects about this “Data World”, highlighted areas for improvement, showcased my strengths, exposed my weaknesses, and challenged me in various ways. I am grateful for this journey and look forward to applying these lessons and insights in my future pursuits.

1. **References**

MACDONALD, Jon. **The 9 Key Ecommerce Metrics You Should Be Tracking No Matter What**. 2023. Disponível em: <https://thegood.com/insights/ecommerce-metrics/>

LEBLANC, Dan. **The 2023 Seller’s Guide to Amazon Metrics and Amazon Reports**. 2023. Disponível em: <https://www.daasity.com/post/amazon-metrics-and-amazon-reports>

PIOVESANA, Luiz. **O que são métricas de marketing digital e conheça as 20 principais.** 2023. Disponível: <https://www.nuvemshop.com.br/blog/metricas-de-marketing-digital/>

KEENAN, Michael. **15 Critical Ecommerce Metrics You Must Track. 2022**. Disponível em: <https://www.shopify.com/blog/basic-ecommerce-metrics>

D’ALAMA, Laryssa. **Glossário de Marketing Digital – com vídeos!**. 2020. Disponível em: <https://resultadosdigitais.com.br/marketing/glossario-de-marketing-digital/>

Feedvisor. **Defining Amazon ASINs, Variation Relationships, and Parent-Child ASINs**. Data Indisponível. Disponível em: <https://feedvisor.com/university/asin/>

PARAB, Lokesh. **Amazon Products Sales Dataset 2023**. 2023. Disponível em: <https://www.kaggle.com/datasets/lokeshparab/amazon-products-dataset>