

# Data Analysis Report

## FicName Company

**Felipe Bitencourt Martins**

### **1. Introduction**

FicName is a leading platform in the market for distribution and sale of digital products, offering a comprehensive infrastructure for digital entrepreneurs and content producers who wish to market online courses, books, e-books, videos, and other digital materials. With advanced marketing and management solutions, FicName provides a secure and simplified experience, as well as offering data analysis and insights to help producers optimize their sales strategies. The company has a strong international presence and is constantly innovating to meet the needs of digital entrepreneurs around the world.

In this report, we will present the strategies, analyses, and findings achieved while working with a dataset, with the goal of highlighting the candidate's knowledge in data science. The objective is to provide valuable insights to support strategic decisions, reflecting the daily responsibilities of a Data Scientist in the organization.

### **2. Materials and methods**

For the completion of this work, a dataset provided by FicName was used, consisting of a sample of purchases made on the company's platform in 2016, totaling over 1.5 million records. The data was made available through a relational database in MySQL hosted on an AWS server.

To access the data, MySQL Workbench client was used. All analyses were conducted through SQL queries, and graphical visualizations were performed in Microsoft Excel. The main objective of the work was to generate insights, considering especially the category and niche of the products, the device used for the purchase, the top producers, and the top referral websites. Additionally, temporal analyses with linear predictions were conducted. Tables and codes are available in their respective files.

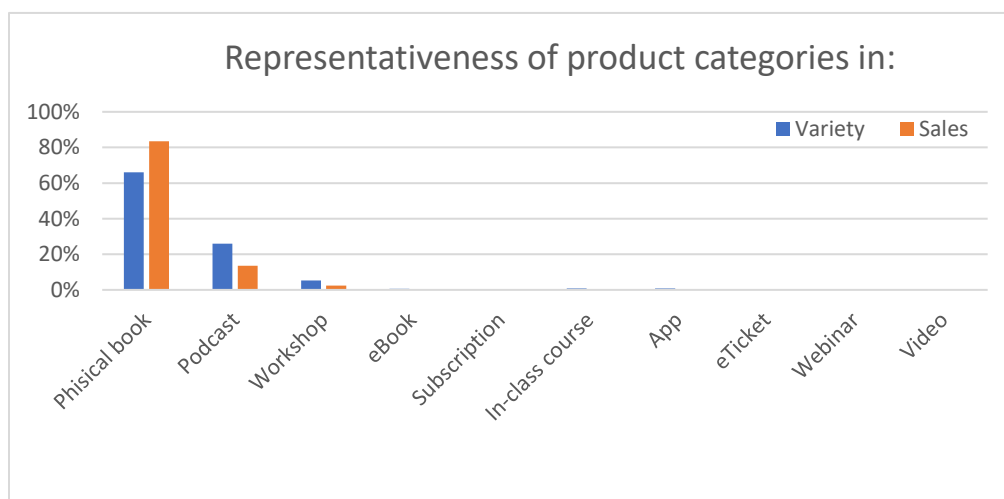
### 3. Results, insights and discussion

#### 3.1. Products and top sellers

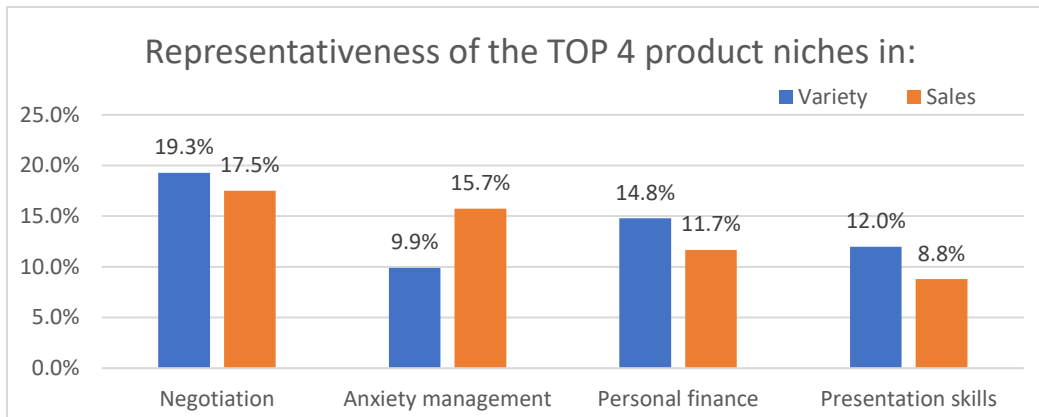
To gain an initial understanding of the data, descriptive analyses of the variables present in the dataset were performed. The count of unique IDs revealed that over a six-month period in 2016, there were 1,599,828 sales to 1,100,649 buyers. These sales involved 17,883 products, produced by 8,020 producers, and involved the participation of 22,947 affiliates. In terms of other variables, the purchases originated from 9,603 different websites and were made through 5 types of devices. It is worth noting that only 36,851 purchases originated from social media networks. The products were categorized into 10 categories and 25 distinct niches (Table 1, Codes 1 and 2).

E Regarding the product categories, it was observed that the most predominant type of sale is "Physical book," representing 83.4% of the total sales. Following that, we have "Podcast" products with 13.5% and "Workshop" with 2.4%. When considering the total number of distinct products, "Physical book" represents 66%, while "Podcast" represents 26% and "Workshop" represents 5.2%. The remaining categories have representativeness below 1% in relation to the total number of products and sales (Figure 1, Table 2, Code 3).

Regarding the niches, the best-selling products belong to the "Negotiation," "Anxiety management," "Personal Finance," and "Presentation skills" niches, which together account for 53% of the total sales. Additionally, these same niches account for 56% of the total number of distinct products (Figure 2, Table 3, Code 4). With these initial results, it becomes evident that during the period of 2016, the company's



**Figure 1.** Graph showing the share of each category in total unique products and total sales.



**Figure 2.** Graph showing the representation of the TOP 4 niches in total unique products and total sales.

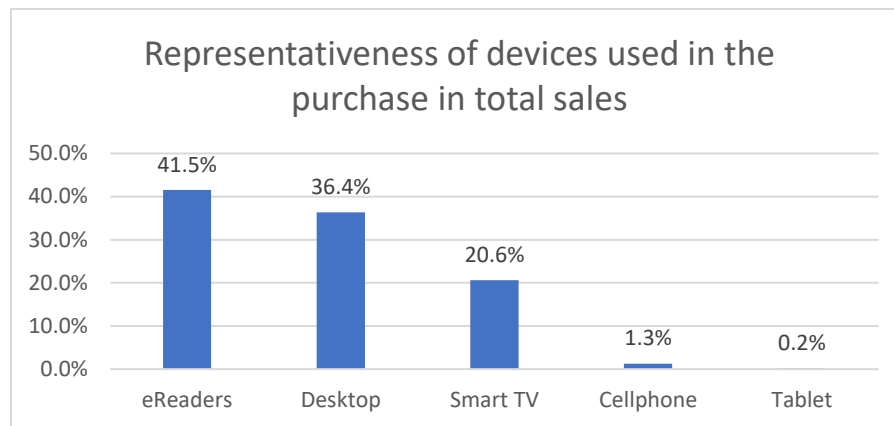
market was concentrated in the sale of physical books, especially those related to business, finance, and interpersonal skills. This indicates that FicName can further explore this category by investing in partnerships with authors, publishers, and book distributors to expand its offering and boost sales in this area. Additionally, the company can direct marketing efforts and product development to meet the demands of the mentioned niches. This strategic segmentation can increase the platform's relevance to specific audiences and drive sales growth.

Furthermore, the difference between the proportions of products and sales in the "Physical book" category (66% and 83.4%) may be evidence that there is still room for the launch and introduction of new products in this category, allowing for further capitalization on this high demand. On the other hand, when we analyze this difference for the "Podcast" category (26% and 13.5%), it may indicate that the demand is already being met by a sufficient supply, and it may not be advantageous to increase the variety of products in this category. However, it is important to note that drawing definitive conclusions would require a more in-depth analysis, considering factors such as product quality, financial performance, market competition, and audience interest/feedback.

### **3.2. Sell to "eReaders"**

Regarding the devices used in purchases, "eReaders" were the most representative, accounting for 42% of the purchases. However, both "Desktops" and "Smart TVs" had considerable values, with 36% and 21% respectively. The other devices had proportions below 1.3% (Figure 3, Table 4, Code 5). It is interesting to note in these results that despite a very low percentage of products sold being "eBooks" (0.68% of the products and 0.25% of the sales), "eReaders" were the primary purchasing device.

Additionally, 76.7% of all purchases made on "eReaders" were physical books, and 20% were "Podcasts," which are two types of products that are not typically played on "eReaders" (Code 6).

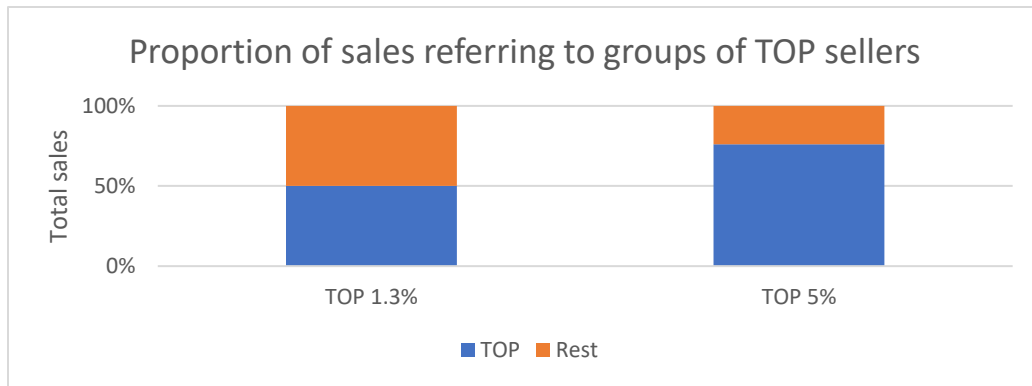


**Figure 3.** Graph showing the representativeness of the devices used for purchases in relation to the total number of products sold.

In this situation, it is evident that there is a significant market of "eReader" users that can still be explored. To seize this opportunity, FicName can pursue strategic partnerships or specific resources targeting users of this device, enhancing the purchasing experience and meeting their needs. Furthermore, diversifying the offering of "eBooks" can be an excellent strategy to increase sales and revenue, including encouraging producers to make their books available in both the "Physical book" and "eBook" categories.

### 3.3. Great producers and the FicName

Investigating the producers who sold the most during the period, no producers with exceptional sales were identified. The top 4 producers, despite selling large quantities (over 30,000 products), do not represent a significant portion of the total sales (9.4%), and they did not sell much more than the next producers in the ranking (Table 5). However, if the analysis is extended to a larger number of top producers, we obtain different results. For example, if we consider the top 108 producers, which represent only 1.3% of the total number of producers, their sales account for 50% of the company's total sales. And when we consider the top 5%, this portion rises to 76% (Figure 4, Codes 7 and 8).

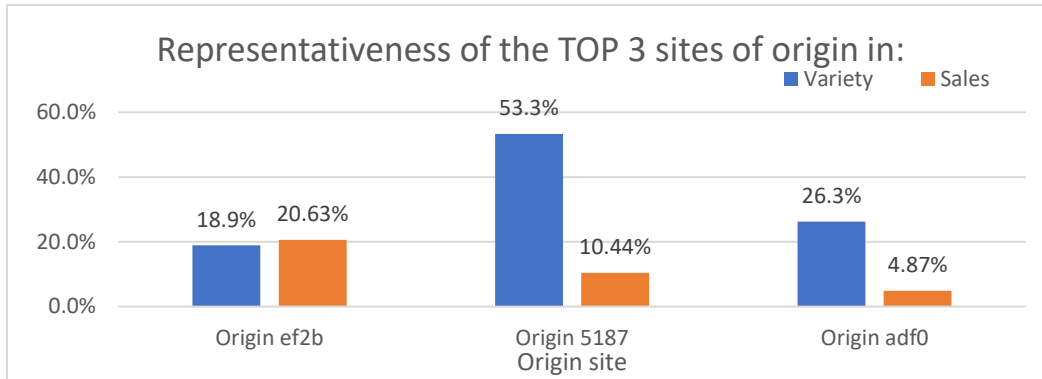


**Figure 4.** Graph showing the proportion of total sales that were made by the TOP 1.3% and TOP 5% largest producer groups.

Due to the lack of data related to revenue, such as sales value, discounts, taxes, and profits, it is not possible to determine if the top-selling producers are responsible for the majority of FicName's revenue. However, based on the presented data, there is a strong indication that a small portion of producers is responsible for a significant portion of the revenue. In this regard, it is strategic to support and promote successful producers through strong relationships, offering support, marketing resources, and collaboration opportunities. Additionally, it is important to initiate projects that facilitate the growth of small producers so that in the future, standout producers do not account for such a significant portion of the revenue. This would make the company more stable and less dependent or susceptible to the decisions of a small group of producers.

### 3.4. Origin sites and the FicName

When analyzing the data related to the source website addresses of the purchases, a similar but more extreme situation was observed compared to the top producers, with two websites showing exceptional numbers. The top 1 and top 2 were responsible for 20.6% and 10.4% of the sales, respectively, while the top 3 and 4 represented 4.9% and 1.8%. The remaining websites on the list accounted for less than 1% of the sales. It is interesting to note that the three websites with the highest sales are also among the top three websites that sold the most distinct products, with one of them being responsible for 53.3% of the unique products in the dataset (Figure 5, Table 6, Code 9). This result suggests a possible correlation between the diversity of products promoted by the websites and the volume of sales, which can be an opportunity to increase sales on similar websites.

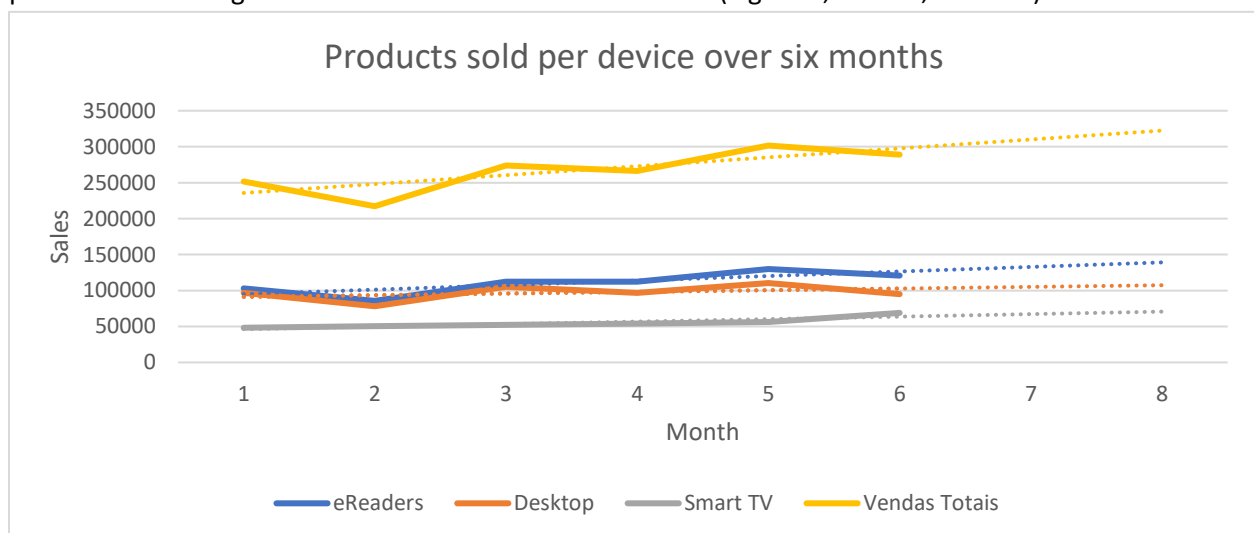


**Figure 5.** Graph showing the representation of the TOP 3 origin sites in total unique products and total sales.

In this context, the strategic positioning of FicName should be similar to what was described in the previous topic. It is important to strengthen the relationship with the websites that showed exceptional numbers, seeking partnership opportunities to maximize the reach and impact of marketing campaigns. At the same time, it is necessary to investigate ways to increase sales from the other websites, aiming to reduce the company's dependency on the exceptional websites. This way, FicName can diversify its sources of sales and mitigate potential risks associated with an excessive reliance on a few websites.

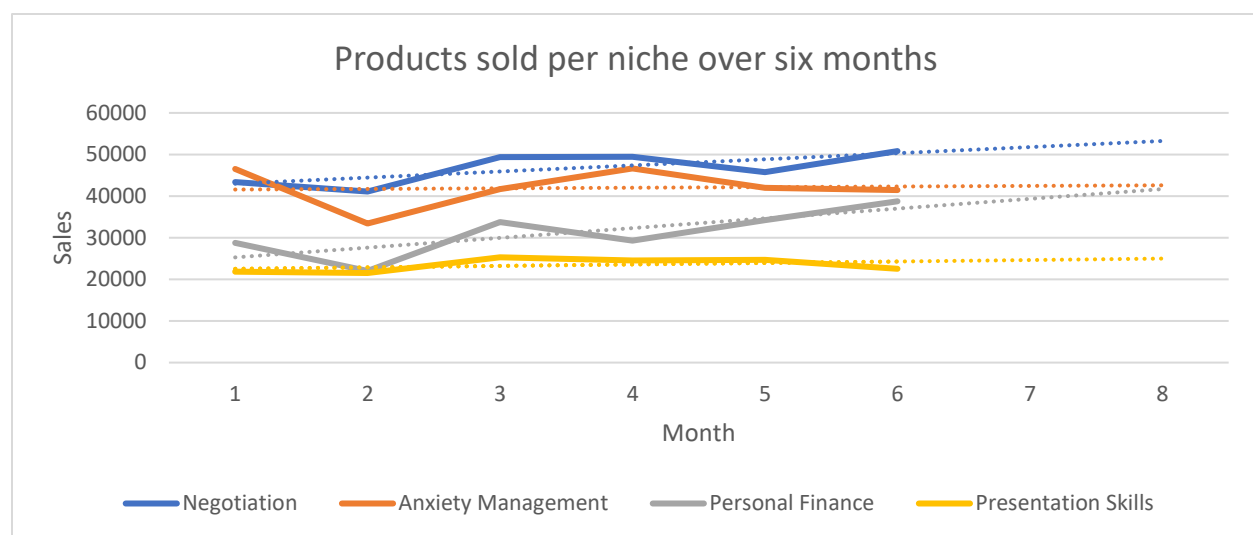
### 3.5. Analysis over the months

When analyzing the quantity of sales over the six available months in the dataset, it was found that the monthly average was 266,638 products, with a standard deviation of 27,238 (Code 10). Additionally, a positive linear trend in total sales was observed, indicating an increase in sales over the period and allowing for sales estimation for the near future (Figure 6, Table 7, Code 11).



**Figure 6.** Graph showing total sales and total sales by purchasing device over six months, with linear forecast for the next two months.

When comparing sales separated by the device used in the purchase, a high correlation was found between sales from "eReaders" and "Desktops". The observed series in total sales is influenced by the variations in sales of these two devices. On the other hand, sales from "Smart TVs" showed a steady growth in the first five months and a significant increase in the sixth month, not following the same pattern as the other two devices (Figure 6, Table 7). This difference in sales patterns indicates that the behavior of users of these devices is distinct. Therefore, conducting a more in-depth study to identify the reasons behind these results can provide valuable insights to guide marketing strategies. By considering this market segmentation, it is possible to boost sales by adopting specific and personalized approaches for each user group.



**Figure 7.** Graph showing the representation of the TOP 4 niches in total unique products and total sales.

The same analysis was conducted considering the TOP 4 niches of products with the highest representation in sales. Among these niches, "Negotiation," "Anxiety Management," and "Personal Finance" showed similar patterns of variation. "Negotiation" and "Personal Finance" ended the period with an increase in sales, indicating a growth trend and confirming the existing relationship between these topics. On the other hand, the "Presentation Skills" niche had little variation over time, and the observed linear trend indicates a maintenance in the quantity of sales (Figure 7, Table 8, Code 12). These results suggest possible strategies to increase sales, such as cross-promotion that integrates the marketing of the "Negotiation" niche with "Presentation Skills" through referrals and promotions that include products from both niches. By leveraging the increasing sales trend of "Negotiation," it is possible to boost the sale of products related to the "Presentation Skills" niche. This approach allows for the exploration of synergies between the niches and maximizes sales results.

Despite linear trends allowing predictions of product sales quantities in the coming months, which is important for both company and producers' planning, it is important to highlight that the method used in these forecasts is not very reliable, with linear regression being the simplest method of forecasting. For a solid positioning regarding future sales, it is recommended to use longer time series data. This would allow the use of techniques such as time series decomposition, which separates the series into trend, seasonality, and residual variation components, as well as the use of more robust models such as autoregressive moving average (ARMA) models, autoregressive integrated moving average (ARIMA) models, and exponential smoothing models.

#### **4. Conclusion and perspectives**

With the analysis of the provided dataset, we were able to obtain a comprehensive overview of FicName's market activities during the analyzed period, identifying its strengths, weaknesses, and potential opportunities. This analysis provided valuable insights into sales trends, top producers, product categories, and devices used in purchases. We identified the importance of standout producers and websites for the company's total sales and emphasized the need to strengthen relationships with these partners and diversify this base to reduce dependence on a select group. Additionally, we identified an opportunity to explore the market of eReader users and enhance the purchasing experience for this specific audience. These insights provide important strategic guidance for FicName, aiming to maximize sales, increase revenue, and drive company growth.

Future perspectives for this work include the integration of more data into the analysis, especially related to revenue, to enable more comprehensive studies involving the financial aspect of the company and support decisions for greater profitability and growth. The use of more data would also enable the application of advanced data science methods such as dimensionality reduction analysis, predictive modeling, and clustering analysis, which have great potential in investigating consumer profiles and increasing company efficiency. This would allow FicName to adopt more solid and reliable strategic positions.