Johnny Bolívar & Patricia Romero

Problem Statement

A shoe store specialized in the sale of sports shoes decides to reopen its doors after 6 months of being closed for quarantine reasons due to the COVID-19 pandemic. During this 6 month period of inactivity, the owner decides to restructure the business and implement a new more effective sales strategy once the store reopens. The objective is that the total sales revenue of this new period will exceed, in less time, the total sales revenue of the previous period, from the opening of the shoe store until its closure due to the pandemic.

As a first step in realizing the new sales strategy, the owner decides to hire a data analyst to conduct a study of sales performance over the entire period during which the shoe store remained in business; the shoe store opened on March 1, 2018 and closed its doors, as a result of the pandemic, on April 12, 2020. The reopening of the shoe store will coincide with the start of the holiday season, which is expected to be favorable for sales.

After obtaining the sales data, cleaning it and analyzing it, the data analyst notices a peculiar pattern: there are a considerable number of sneaker models that are discounted but have had little or no sales, and, conversely, there are sneaker models with a large number of sales that have no discount. Following this discovery, the analyst arranges a meeting with the shoe store owner to show him and offer him a proposed sales strategy in two phases to be implemented in unison. One phase would consist of taking out of stock the models that report little or no sales, liquidating them at prices that are even below cost in order to get out of them as soon as possible and recover some money. The other phase would consist of leaving the best-selling models in stock and discounting them; although this reduces the profit margin per unit, it is also true that, as they are the best-selling models, the number of sales would increase even more, due to the discounts, and the total profit per sales volume would be even higher.

After being presented with the proposal, the owner of the shoe store decides to accept it and instructs the data analyst to write it down in detail in a report with its respective analysis, as well as a dashboard with the sneaker models that will remain in stock and to which discounts will be applied. With this strategy, it is expected that in less than two years total sales can exceed the sales of the previous period and that the business can obtain more revenue with a smaller stock of models.

Results and Analysis of Results

The data with the information of the models in stock of the sports shoe store was extracted from an SQLite database from the Kaggle website. After cleaning the database, a total of 3120 models of sneakers in stock were obtained. See Figure 1.
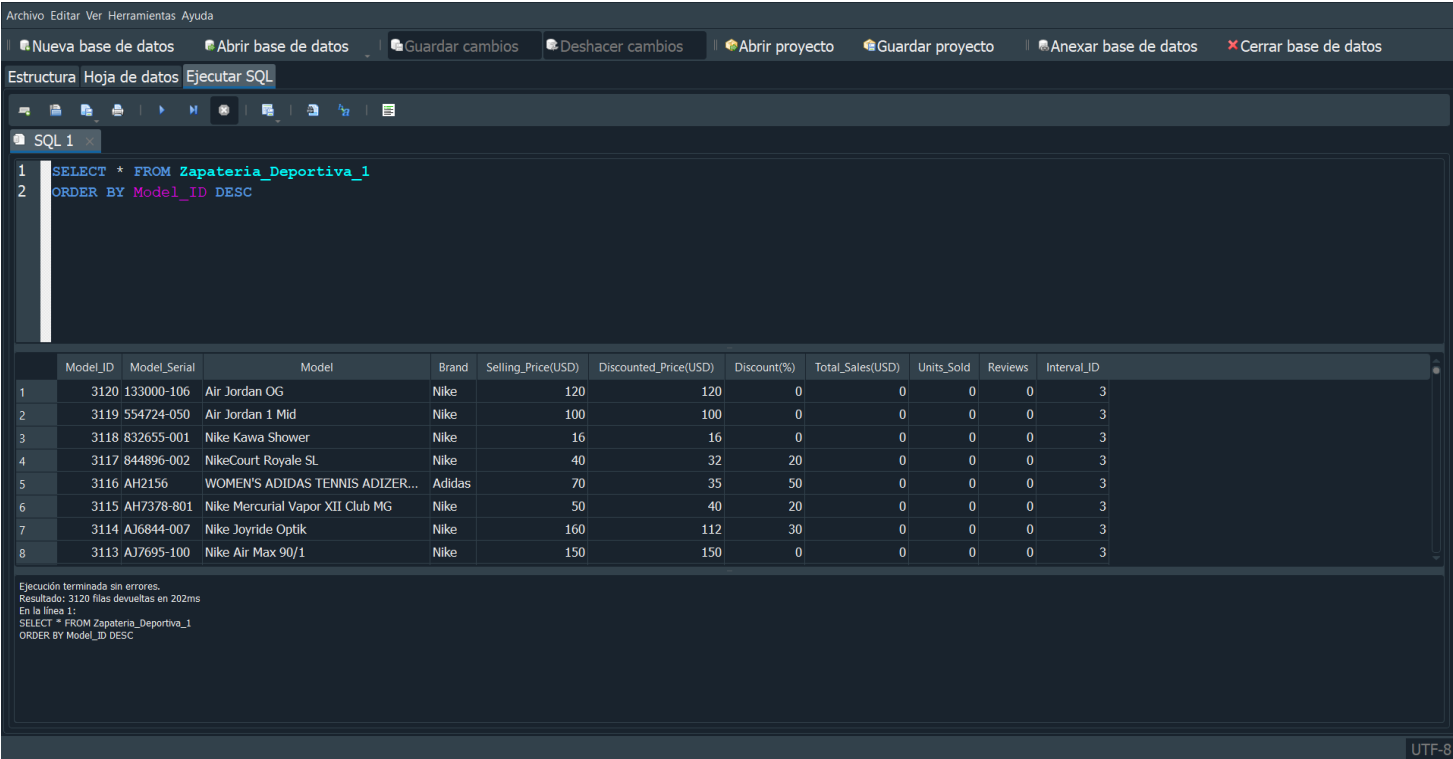

Figure 1. SQLite table with the 3120 models of sneakers in stock.

When applying a discount to a model, the reduction in the profit of each unit sold must be justified by a greater number of sales, since if, despite the discount, the number of units sold does not increase, there will be no increase in the profit per sales volume and, therefore, the sneaker model will generate losses. Based on this premise, the analysis will be divided into three groups described below:

Group A: Those that have generated 1 or more sales.

Group B: Those that have generated 50 or more sales.

Group C: Those that have generated 100 or more sales.

Group A, B and C will undergo an analysis of total sales, an analysis of the number of discounted models and an analysis of the number of reviews, with the objective of determining which of the three groups will be chosen as the new sales stock of the sports shoe store, based on a higher number of sales that generate a higher profit per sales volume.

Group	Sales Quantity	Models Quantity	Total Sales (USD)	Models Quantity (%)	Total Sales (%)	Analysis of Total Sales
A	>=1	2908	12328946	93	100	93% of the models generate 100% of total sales
B	>=50	1863	11008419	60	89	60% of models generate 89% of total sales
C	>=100	1109	7928584	36	64	36% of models generate 64% of total sales

Table 1. Analysis of total sales.

Group A, in the total sales analysis done from the data in Table 1, makes up the total sales during the Shoe Store's active period from March 1, 2018 through April 12, 2020, and represents the goal to exceed during the store's new active period. While group C generates almost two-thirds of sales, with just over one-third of the models in stock, it has a difference of over $4 million in total sales from group A, so it would be more difficult for it to exceed this goal with a discounting strategy than group B, which has a difference of just over $1 million in sales from group A.

Group	Sales Quantity	Models Quantity	Models Quant. with Discount	Models Quantity (%)	Models Quant. with Discount (%)	Analysis of Quantity of Models with Discount
A	>=1	2908	2050	93	97	93% of models have 97% of models discounted
B	>=50	1863	1388	60	66	60% of models have 66% of models discounted
C	>=100	1109	834	36	39	36% of models have 39% of models discounted

Table 2. Analysis of quantity of models with discount.

The objective of the analysis of the number of discounted models, shown in Table 2, is to eliminate the largest number of discounted models that do not generate sufficient sales. By taking group B, 44% of models that have discounts but do not generate sufficient sales would be eliminated. The question that arises at this point is why group B and not group C? The answer is that within group B there is a larger number of models to which no discount has been applied than in group C; as there is a larger number of models to which discounts can be applied, the higher the sales volume will be and the higher the sales volume gain will be.

Group	Sales Quantity	Models Quantity	Reviews	Models Quantity (%)	Reviews (%)	Analysis of Quantity of Reviews
A	>=1	2908	129622	93	100	93% of models have 100% of number of reviews
B	>=50	1863	117603	60	91	60% of models have 91% of number of reviews
C	>=100	1109	86088	36	66	36% of models have 66% of number of reviews

Table 3. Analysis of quantity of reviews.

Table 3 shows how group B has 91% of the number of reviews, representing 60% of the models; even more remarkable is the fact that group C, with a little more than one third of the models, has two thirds of the number of reviews. As group B contains group C, it can be divided into 2 categories: those models whose units sold range from 50 to 99, and those with 100 or more units sold (group C).

From these three analyses it can be seen that the group chosen to be the new stock is group B, with 1863 models, representing a sales volume of 89% and a reduction of 40% with respect to the previous stock of 3120 models. See Figure 2.

Figure 2. SQLite table showing group B with 1863 models of sneakers in stock.

For the purpose of better understanding the analysis, group B will be divided into two categories of units sold:

Category 1: Those models ranging from 50 to 99 units sold.

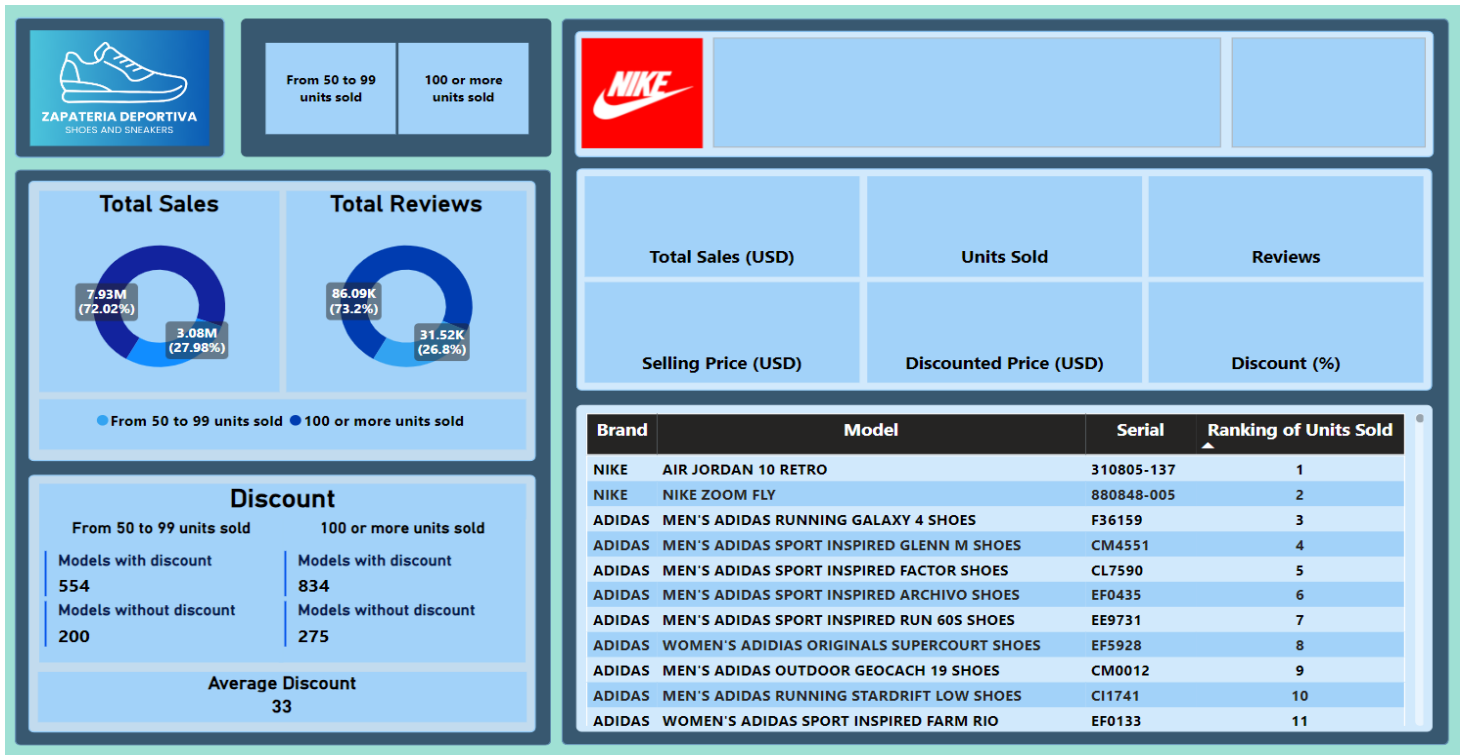**Category 2**: Those models with 100 or more units sold.

Figure 3. Dashboard of the new stock.

Figure 3 shows the dashboard from which the analysis of the new stock will be made to determine what actions should be taken to improve the sales of sneakers. The goal is to achieve that, in the next 2 years, starting from the reopening of the store, the amount of total sales collected will exceed the amount collected in the period of activity of the shoe store prior to the beginning of the quarantine. On the left side there is information on sales, discounts and reviews by category of units sold, and on the right side there is a search engine for all 1863 models with detailed information on each model, such as total sales, units sold, number of reviews, sales price, discounted price and discount.

The total sales of Category 2 are 72.02%, while those of Category 1 are 27.98%. In terms of the number of reviews, Category 2 has 73.2% of the reviews and Category 1 has 26.8%. In approximate terms, Category 2 has 3 times more total sales and total reviews than Category 1.
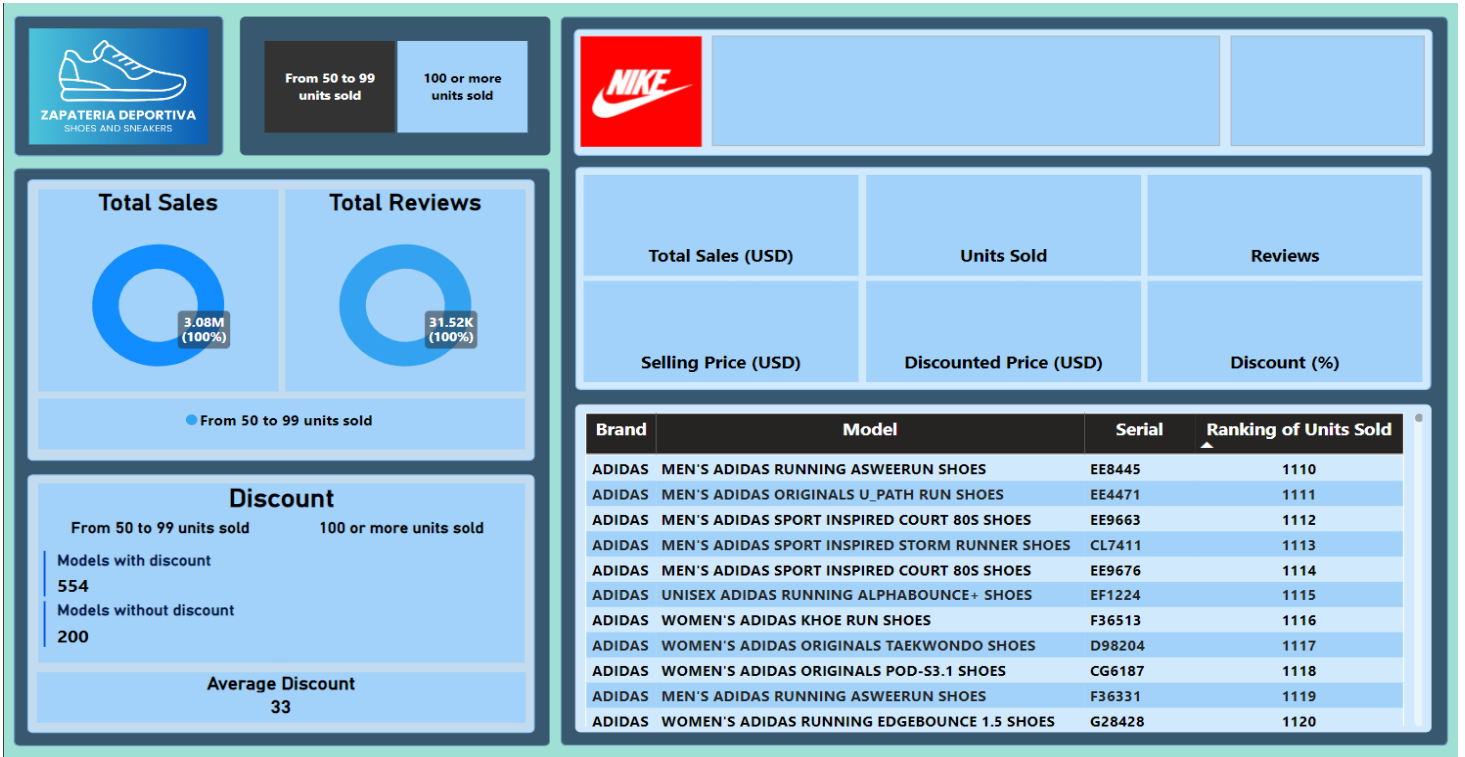
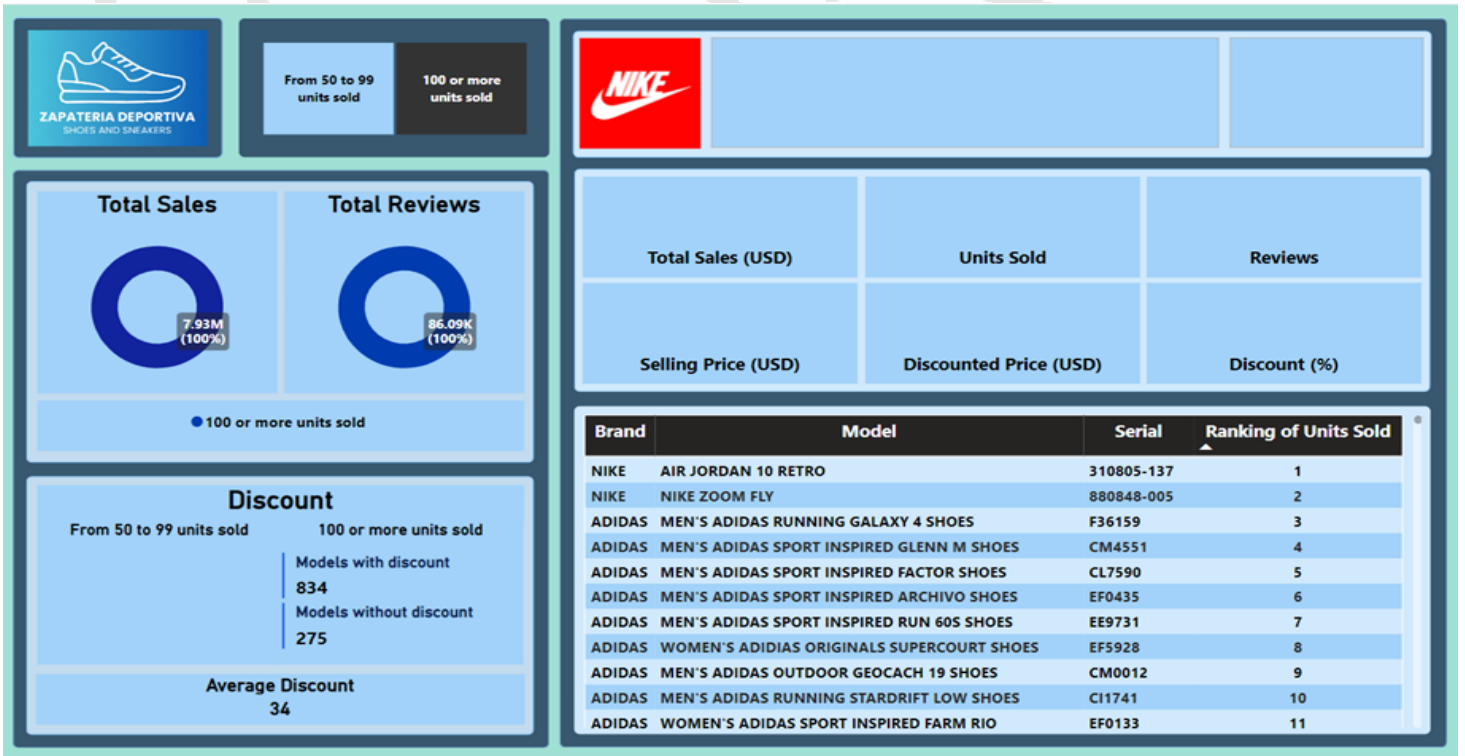**Figure 4.** Category 1 (50 to 99 units sold).

Figure 5. Category 2 (100 or more units sold).

Category 1 has a total of 754 of the 1863 models. Of these 754 models, 200 have no discount. The average discount for Category 1 is 33%. See Figure 4.

Category 2 has a total of 1109 of the 1863 models. Of these 1109 models, 275 have no discount. The average discount for Category 2 is 34%. See Figure 5.

Since there are more non-discounted models in Category 2, and since it is the category that has approximately three-quarters of the total sales and reviews, a higher discount can be applied to them than to the non-discounted models in Category 1.

With a maximum discount ceiling of 33% for Category 1 models and 34% for Category 2 models, low discounts can be applied to increase the number of sales per model sufficiently without reducing the profit per unit sold too much. At the same time, the discount has to be attractive enough to make the customer want to buy his favorite sneaker model at a more affordable price.

CONCLUSIONS

1.   Category 1, with models selling 50 to 99 units, has 27.98% of total sales.

2.   Category 2, with models selling 100 or more units, has 72.02% of total sales.

3.   Category 1, with models selling 50 to 99 units, has 26.8% of total reviews.

4.   Category 2, with models selling 100 or more units, has 73.2% of the total reviews.

5.   Category 1, with models selling 50 to 99 units, has 200 undiscounted units.

6.   Category 2, with models selling 100 or more units, has 275 units without discount.

RECOMMENDATIONS

1.   Maintain the same discounts within the models of both categories that already had it.

2.   Apply a 10% discount to the 200 models in Category 1 that do not have it.

3.   Apply a 15% discount to the 275 Category 2 models that do not have it.

4.   Make an inexpensive media and social media advertisement of the shoe store with the new model discounts.

References

Zapatería Deportiva. Open Data. Available in:

https://www.kaggle.com/datasets/angelobejaranociotti/retail-db, 2024