



DATA-DRIVEN SELLER SEGMENTATION: STATISTICAL ANALYSIS FOR BUSINESS INSIGHTS



</> Fernando Henrique Pereira Scuizato



WHAT IS THE GOAL?

The company needs to create a clusterization for its sellers.

The propose is to create a Supervised Machine Learning Model to classify strategically the company's sellers into four clusters.



SOURCES AND MAIN VARIABLES

Type	Variables
 Numeric	Distinct Products, Average Price, Available Stock, Total Stock Value (USD), Total Transactions and Estimated Sales.
 Category	Sellers Status: Platinum, Gold and Silver; Reputation: Green, Light Green, Yellow, Orange and Red.
 API Reputation and Status	https://api.mercadolibre.com/users/{seller_id}
 API Category	"https://api.mercadolibre.com/sites/MLA/categories"

TOOLS USED



1. Python (API Request, pandas, sklearn, Scipy, Matplotlib and Seaborn;
2. Chat GPT, Stack Overflow and YouTube.



EDA ANALYSIS



4.172

SELEERS



81,2 MI

TRANSACTIONS



10.901

PRODUCTS



31

CATEGORIES

These 12 categories represents a share of 80% of the total transactions registered.

Category	Average Price	Available Quantity	Total Value Stock (USD)	Transactions	Estimated Sales	Share
Computación	348,0491837	5.625	1.813.309,40	2347882	722.748.172,79	12,55%
Herramientas	97,21452074	127.191	9.169.422,91	7853963	722.555.810,53	12,55%
Industrias y Oficinas	130,1166381	535.947	16.594.772,20	7554297	656.403.326,25	11,40%
Ropa y Accesorios	42,73653	405.004	12.841.711,43	11813242	477.468.310,10	8,29%
Deportes y Fitness	49,39991989	338.334	16.974.186,92	9826231	458.799.818,76	7,97%
Consolas y Videojuegos	275,2876321	42.490	2.716.047,43	2444557	447.926.867,21	7,78%
Antigüedades y Colecciones	135,6481526	67.883	2.042.968,62	2095060	250.571.166,20	4,35%
Instrumentos Musicales	251,7327045	163.499	7.673.293,05	1284699	225.012.081,80	3,91%
Electrodomésticos y Aires Ac.	166,4606122	24.862	2.574.966,47	1216435	189.484.018,77	3,29%
Electrónica, Audio y Video	161,7641414	25.333	2.432.639,56	1722933	181.195.492,61	3,15%
Hogar, Muebles y Jardín	63,81232323	37.365	1.328.086,14	3114765	152.402.407,28	2,65%
Celulares y Teléfonos	330,8546939	9.791	1.833.897,40	581726	143.423.111,78	2,49%

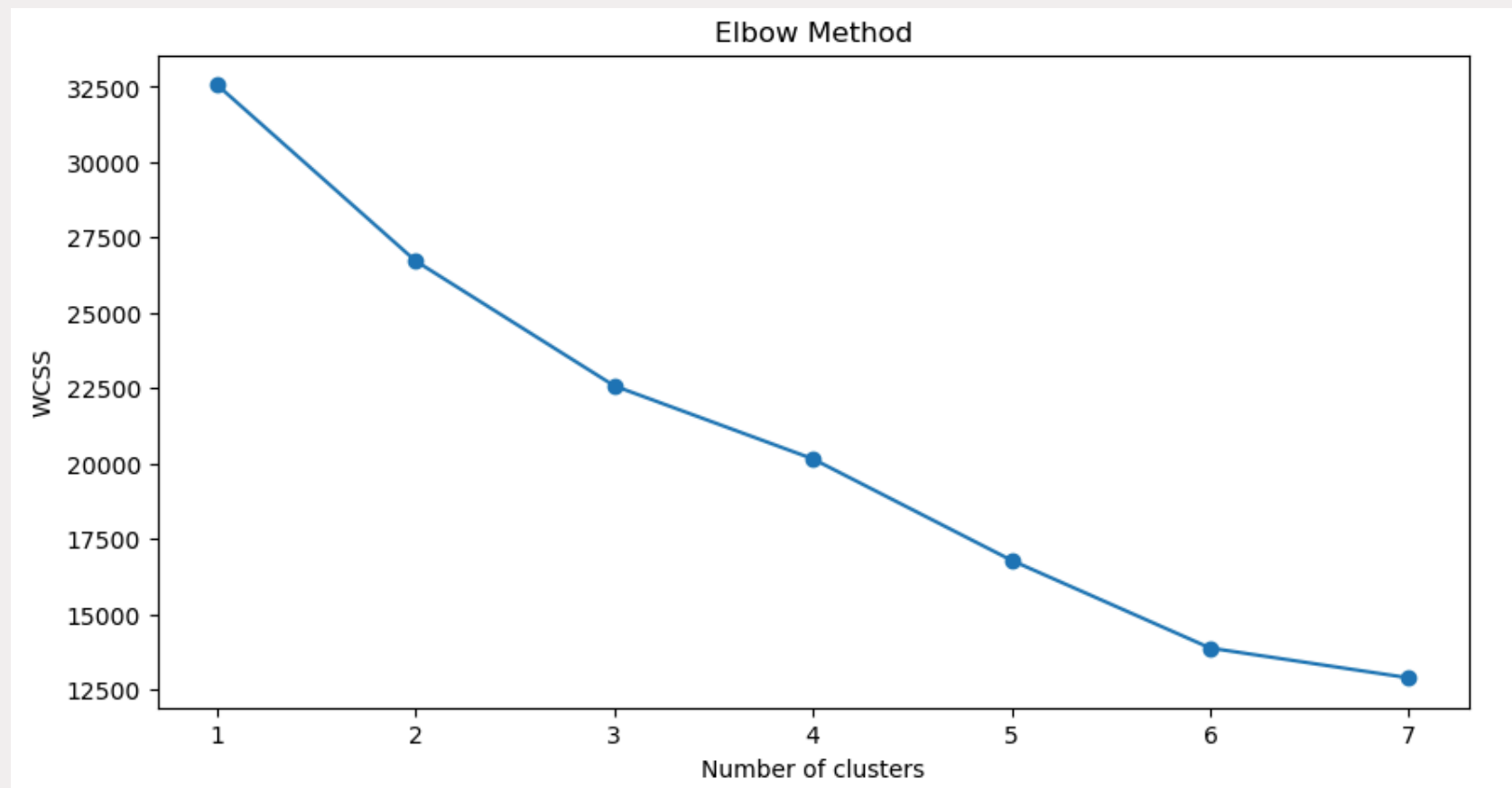


MODEL USED TO SOLVE THE CASE

Unsupervised Machine Learning Model Clustering K-Means

Clustering Quality - Silhouette Score

The K-Means methodology, with 4 clusters, resulted in a Silhouette Score of 0.4163. This value indicates a moderate separation between the clusters, which allows the identification of distinct patterns of salesperson profiles, although with some overlap between groups.



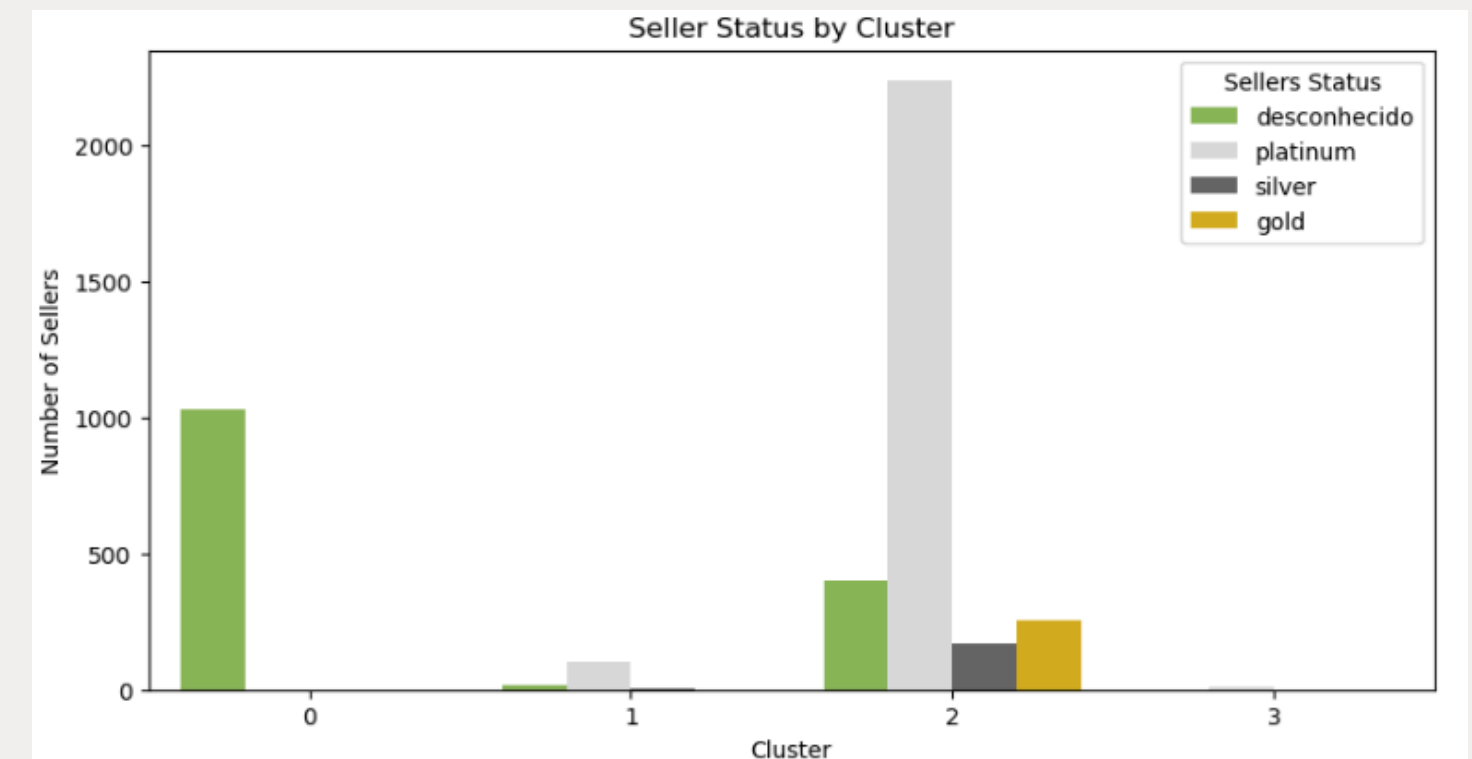
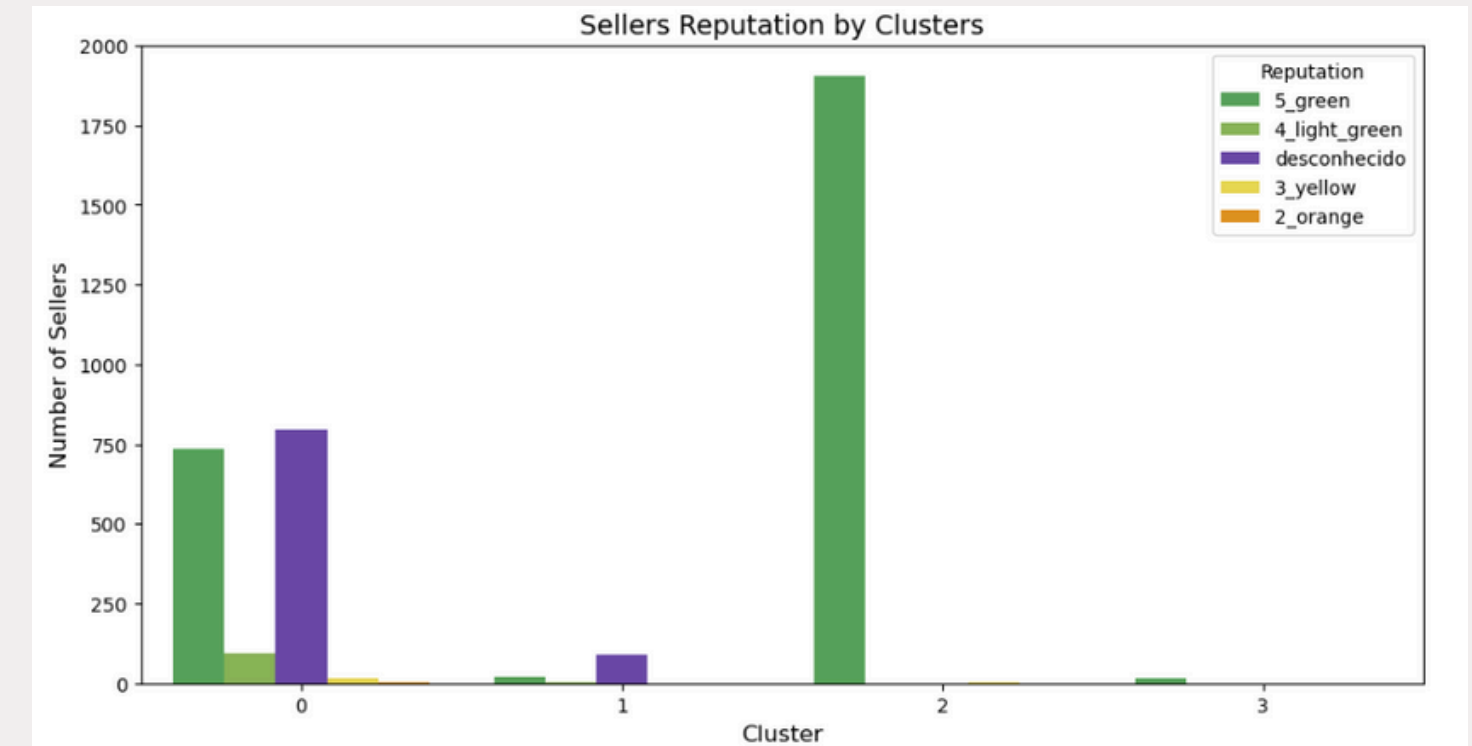
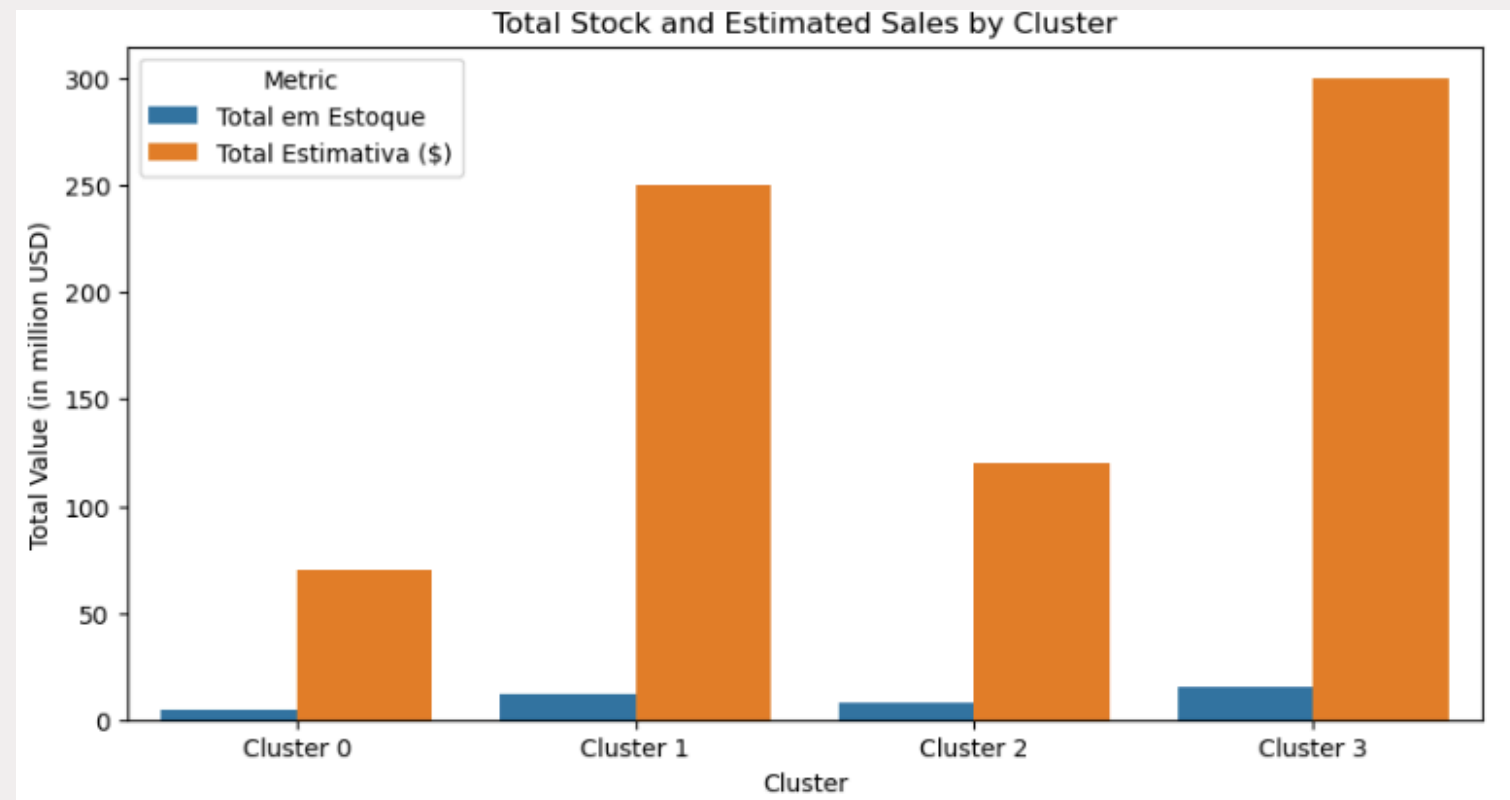
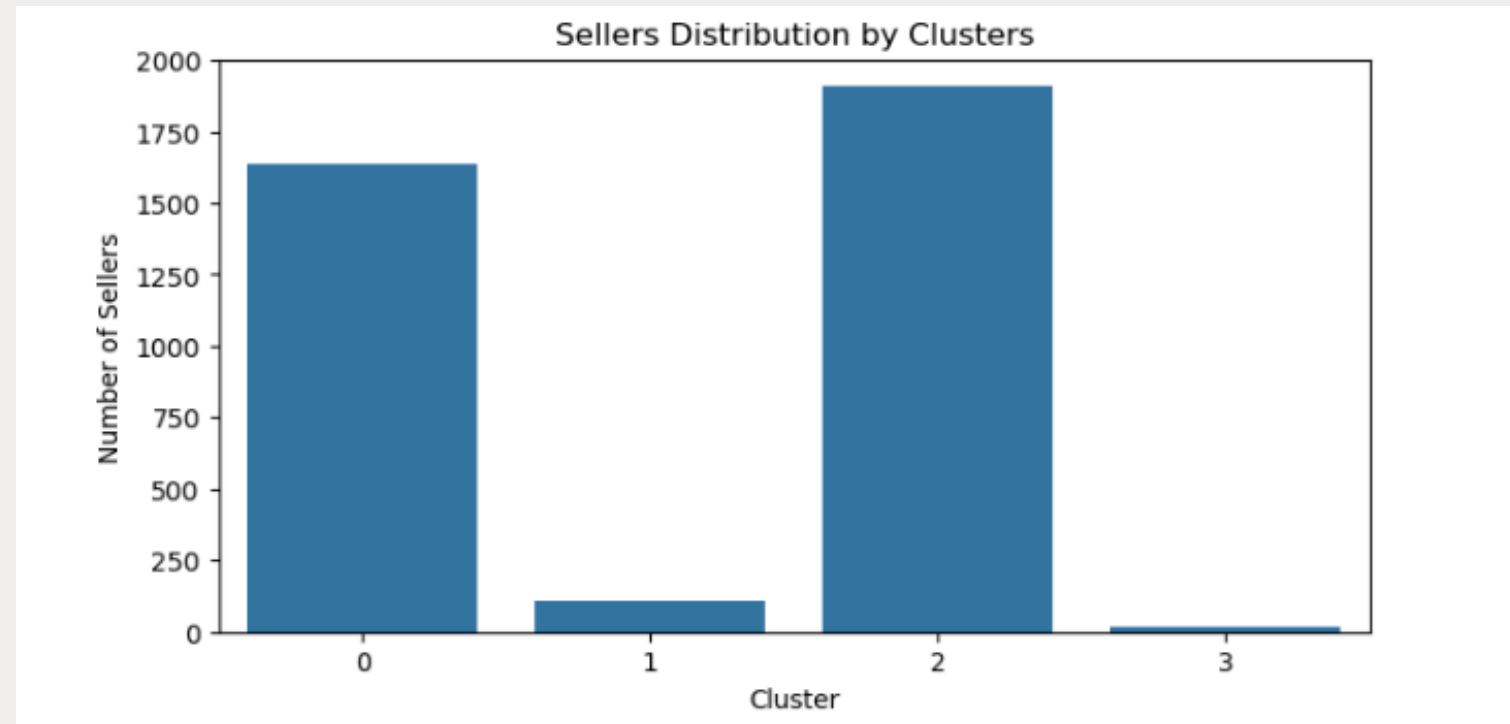
Statistical Validation of Clustering

Higher values in ANOVA (F), Kruskal-Wallis (H), and Chi-Square (X^2) indicate stronger differences or associations. This confirms that the clusters are meaningfully distinct in terms of seller characteristics, product inventory, and sales metrics.

Test	Variable	Statistic	p-value	Interpretation
ANOVA	Number of distinct products	F = 687.47	< 0.001	Significant differences in product variety across clusters.
Kruskal-Wallis	Estimated sales	H = 2246.97	< 0.001	Sales distribution differs significantly between clusters.
Chi-Square	Seller status	X ² = 3565.87	< 0.001	Strong association between cluster and seller status.




SELLER SEGMENTATION AND KEY METRICS







CLUSTER SUMMARY

Cluster	Qty Sellers	Sellers (%)	Stock Share (%)	Sales Share (%)	Avg. Stock per Seller	Avg. Sales per Seller (\$M)	Key Insight
0 - Small Sellers	1.635	44.6%	1.7%	2.1%	28	0.07M	Many sellers with low stock and low sales.
1 - Large Inventory, Low Sales	110	3.0%	58.1%	0.4%	14,115	0.20M	Huge stock, but poor sales conversion.
2 - High-Performing Sellers	1.907	52.0%	36.5%	78.3%	512	2.21M	Best sales efficiency with high transaction volume.
3 - Premium Sellers	16	0.4%	3.6%	19.2%	6,070	64.56M	Few sellers, but highest revenue per seller.

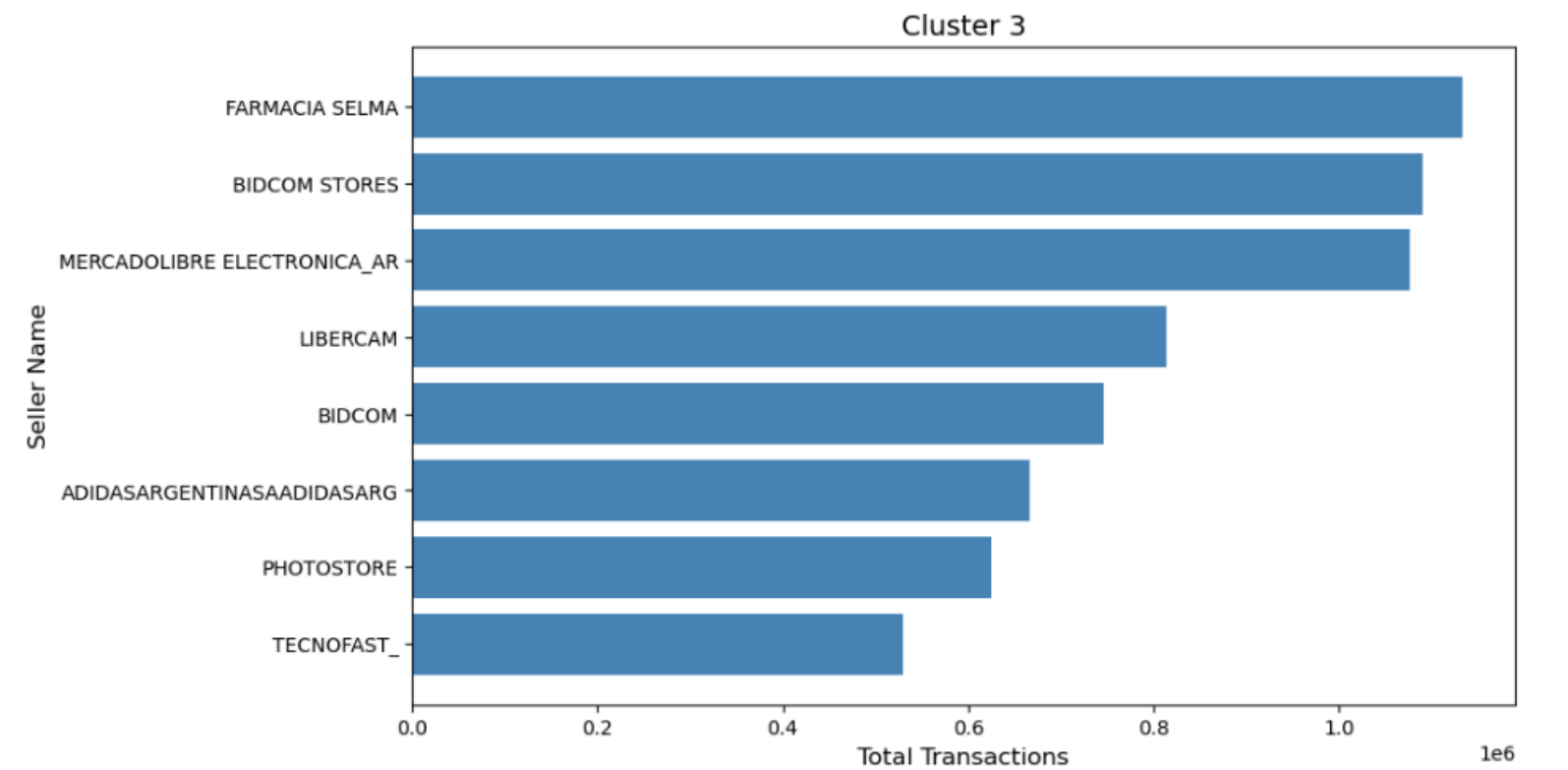
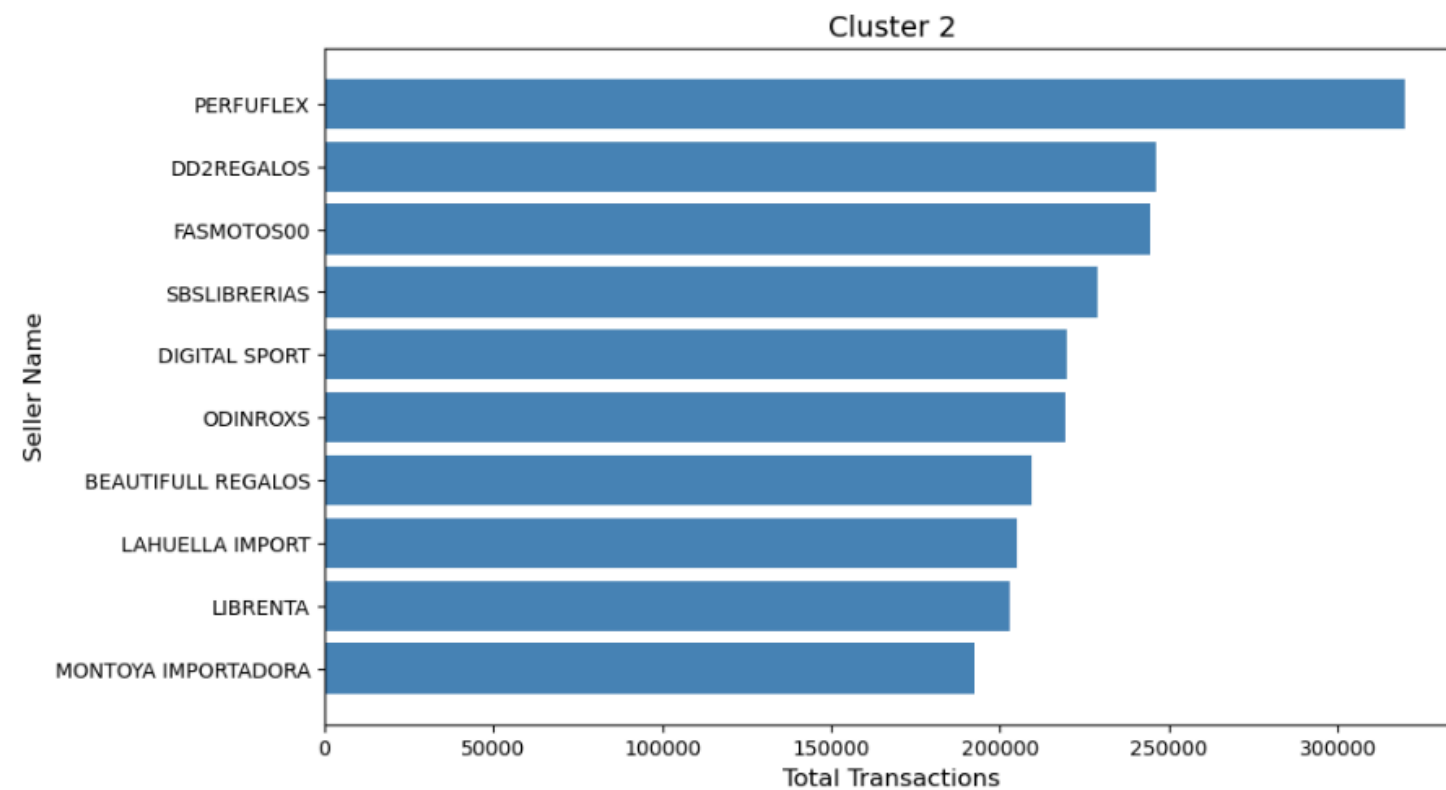
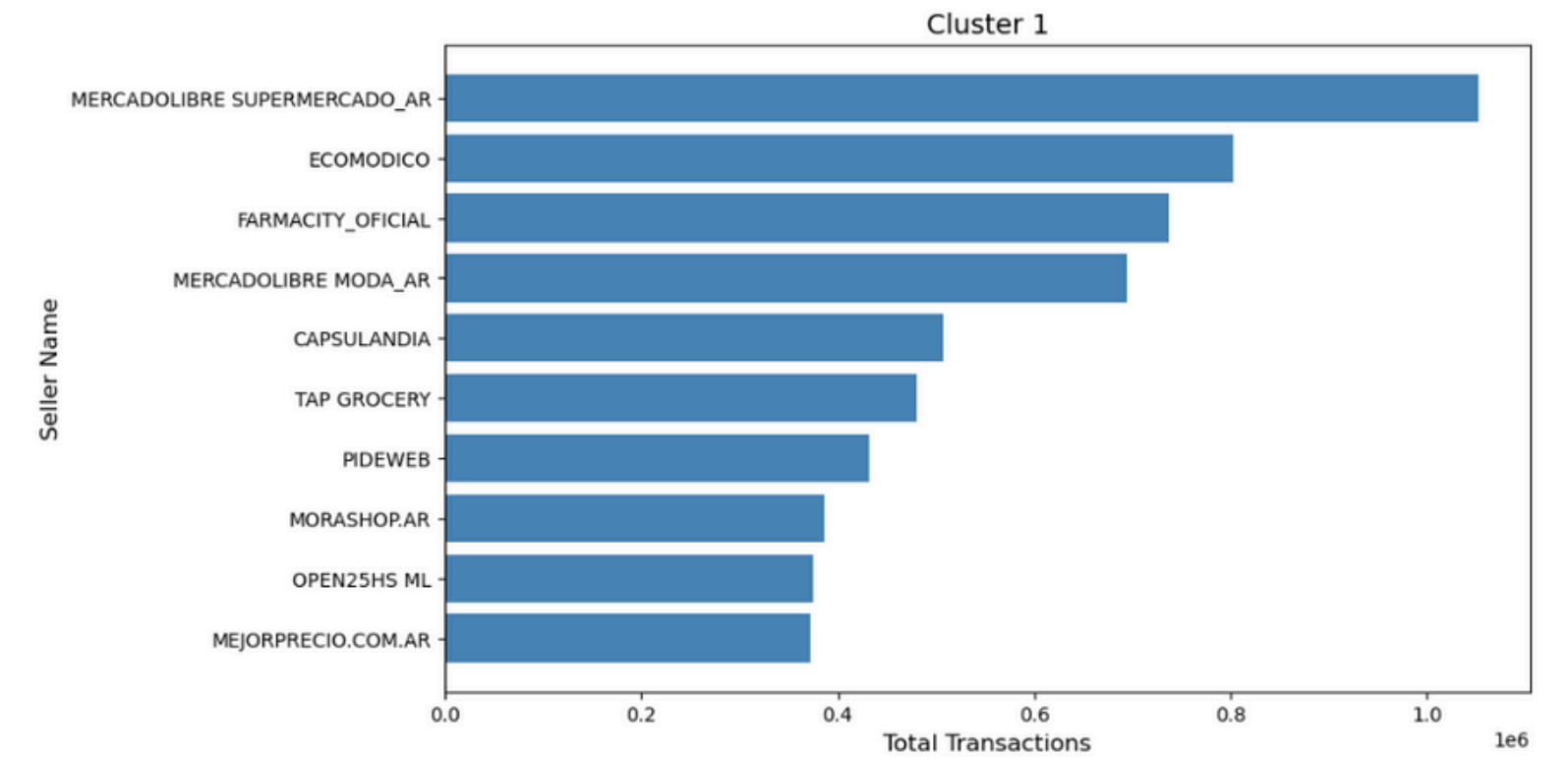
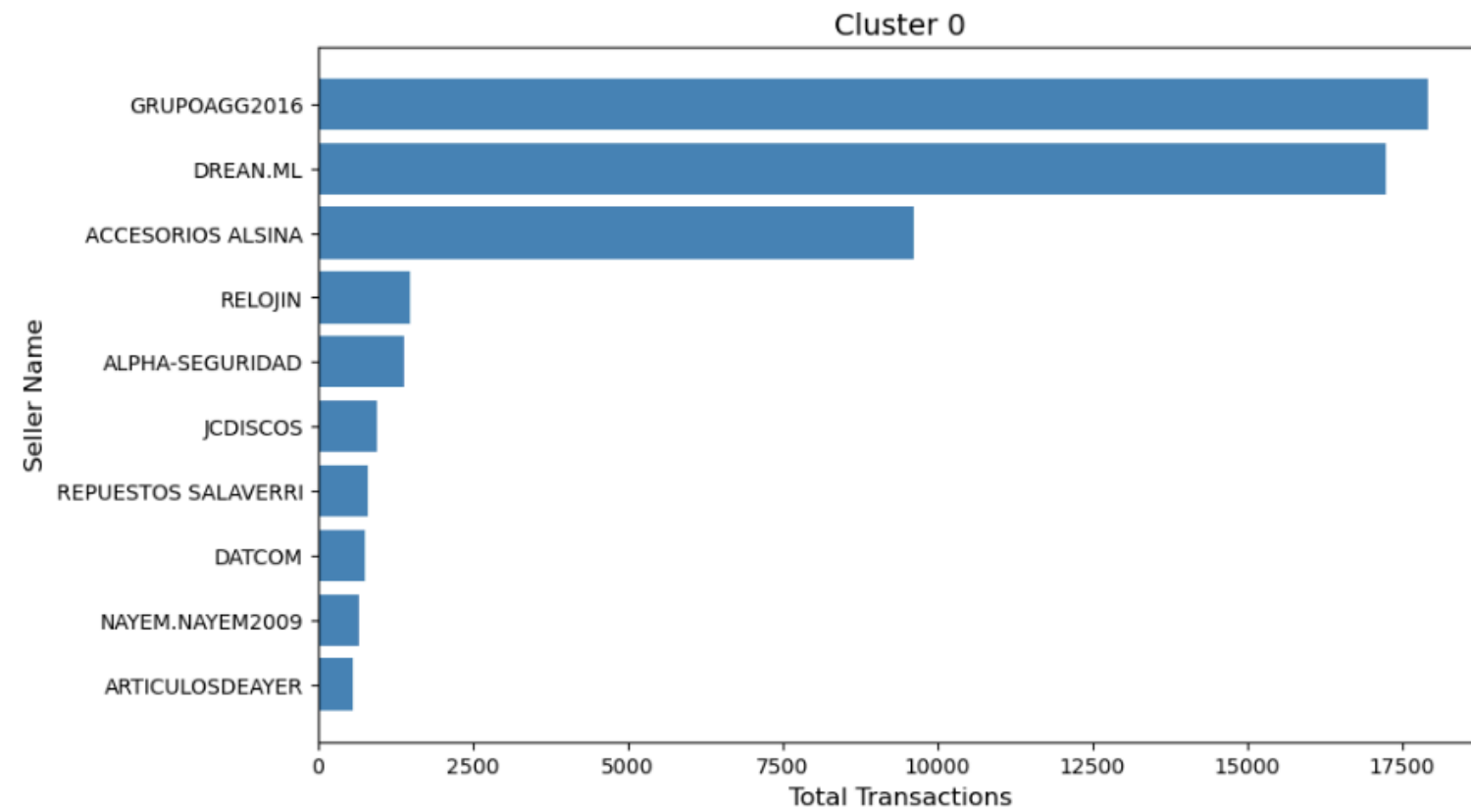
 **Cluster 0 has many small sellers with low impact on total sales.**

 **Cluster 1 holds most of the inventory but struggles to sell.**

 **Cluster 2 dominates sales & transactions despite lower stock.**

 **Cluster 3 has the highest-value sellers, even though they are few in number.**

TOP 10 SELLERS BY CLUSTER



STRATEGIES FOR EACH CLUSTER

Summary

Result

Recomendation

CLUSTER 0



Small Sellers

(Many sellers, low stock, low sales)



44.6% of sellers, but only 1.7% of stock and 2.1% of sales.

Low inventory and low impact on total sales.



Upsell & Support Growth:

Offer incentives to increase stock levels. Provide education on marketplace best practices, pricing strategies, and ad campaigns to boost visibility.

CLUSTER 1



Large Inventory, Low Sales

(Stock-heavy sellers with poor sales conversion)



Holds 58.1% of total stock but contributes to only 0.4% of estimated sales.



Sales Optimization

Investigate pricing and product-market fit. Offer promotional support, better search ranking incentives, and advertising solutions to improve conversion.

CLUSTER 2



High-Performing Sellers

(Efficient sellers with high sales volume)



Largest share of sales (78.3%) and transactions (82%).

Well-balanced stock and strong seller performance.



Retention & Expansion

Encourage exclusivity deals and loyalty programs. Provide advanced analytics to optimize pricing and sales. Keep them engaged with premium seller benefits.

CLUSTER 3



Premium Sellers

(Few sellers, highest revenue per seller)



Only 0.4% of sellers, but 19.2% of total sales.

The highest revenue per seller (\$64.56M per seller on avg).



VIP Treatment

Provide priority customer service, exclusive promotions, and early access to platform innovations. Tailor premium ad solutions for further growth.



NEXT STEPS

1. **Implement clustering in a dynamic system to track changes in salesperson profiles;**
2. **Cross-reference clusters with consumer behavior data to understand the relationship between salespeople and customers;**
3. **Explore predictive models to identify salespeople with the greatest growth potential and salespeople at risk of declining performance;**
4. **Refine segmentation by including more variables such as sales seasonality and pricing policies;**
5. **Create a Dashboard to follow the changes of our Sellers.**

THANKS



**mercado
livre**