

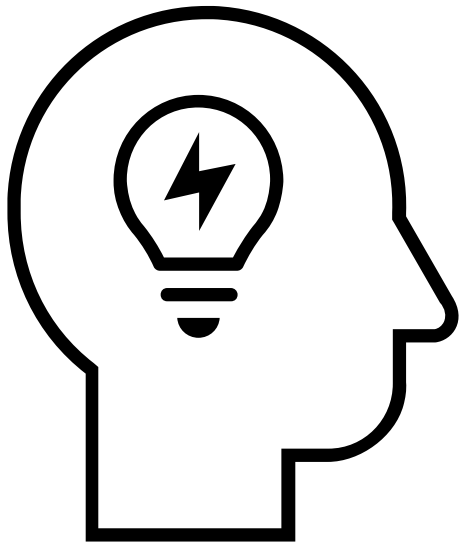
Mercado libre Data Science Challenge



Agenda:

- 1. Caso de Negocio.
- 2. Pregunta Clave de Negocio.
- 3. Levantamiento de la información.
- 4. Contexto.
- 5. Propuesta Analítica.
- 6. MVP de Modelo.
- 7. Caracterización de los grupos.
- 8. Próximos Pasos.

1. Caso de Negocio:



El equipo comercial quiere realizar estrategias focalizadas para los Sellers, pero en este momento no existe una clasificación que permita identificar a aquellos que tienen un buen perfil y son relevantes para el negocio. ¿Cómo podrías ayudar al equipo comercial a identificar estos sellers?

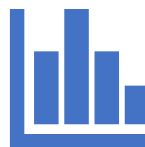
2. Pregunta clave de negocio

¿De los Sellers del país es posible encontrar agrupaciones que permitan identificar a nuestros Sellers más valiosos?



Recopilación de información

Contextualización de caso de uso.
Levantamiento fuentes de información.
Ingesta de Datos.



Identificación de variables significativas

Minería de datos.
Transformación de variables significativas.
Consolidación de Insights.



Calibración Solución Analítica

Propuesta del Modelo.
Calibración del Modelo.
MVP de segmentación Sellers por categoría.

3. Levantamiento de información



Se consulta la información de todos los items en todas las categorías* encontradas en Colombia.



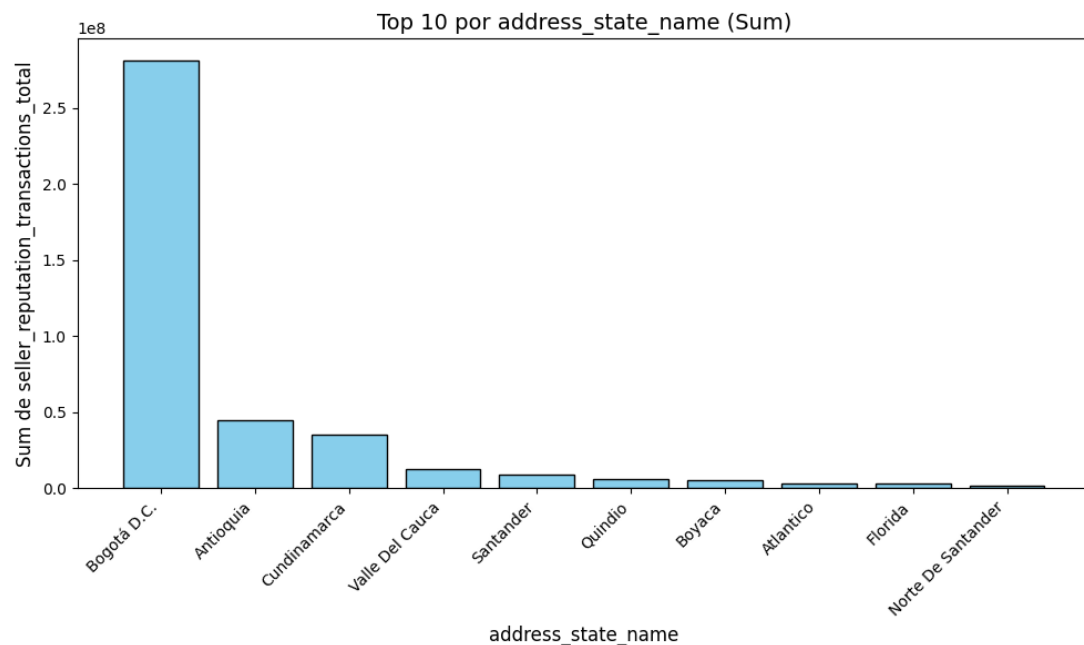
- Precios.
- Descuentos.
- Stock
- Categorías de productos.
- Tipo de envío.
- Pagos a cuotas.
- Ciudad.
- Sellers.
- Campañas.



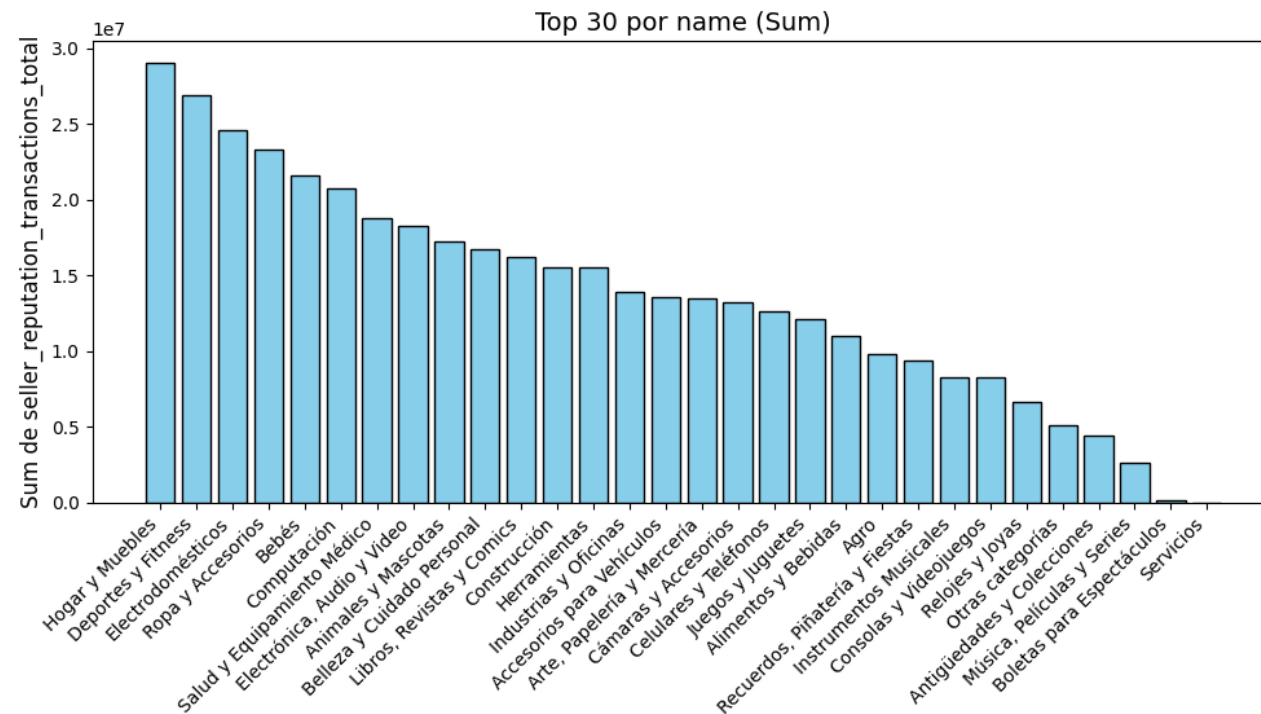
shutterstock.com · 452497843

- **Reputación del seller.**
- **Ventas.**

4. Contexto basado en datos internos

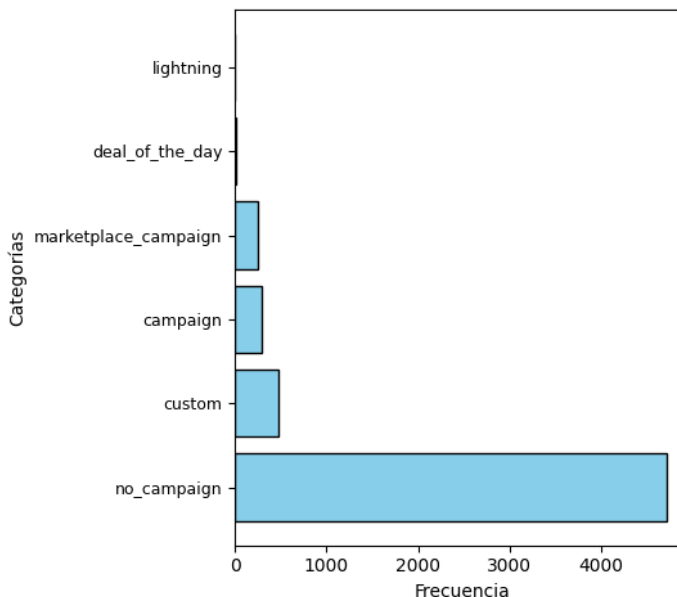
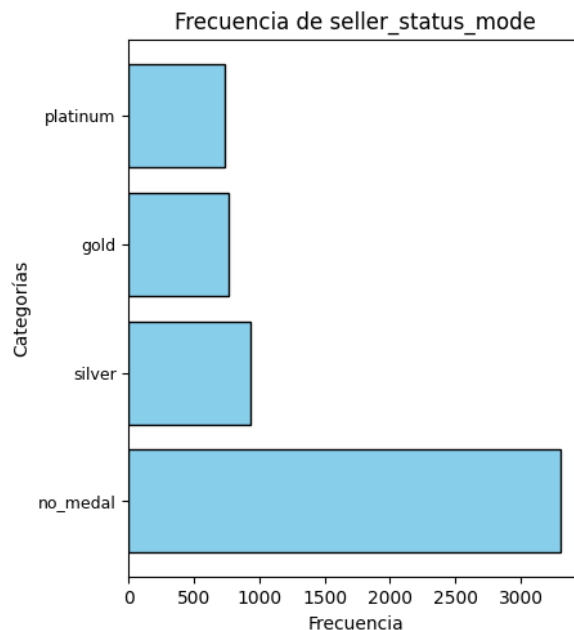


Las transacciones se concentran en Bogotá, Antioquia y Cundinamarca



Donde las categorías más populares son , hogar y muebles, deportes y fitness y electrodomesticos

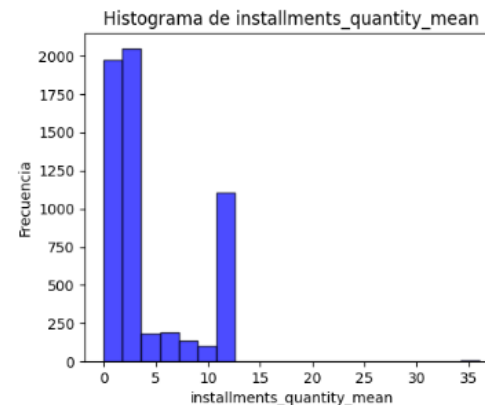
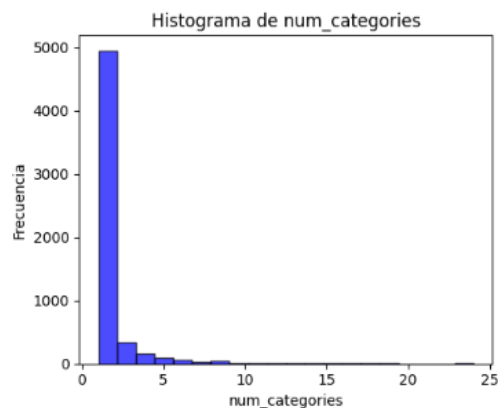
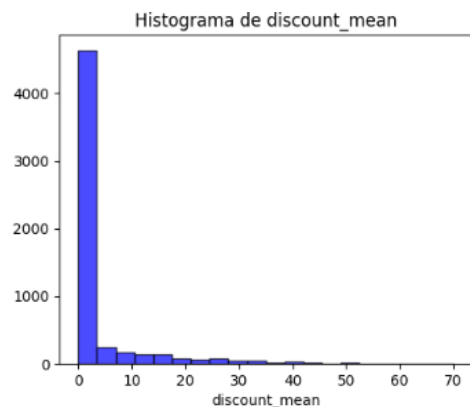
4. Contexto basado en datos



Los vendedores tienden a no pertenecer a alguna campaña y no tener medallas.

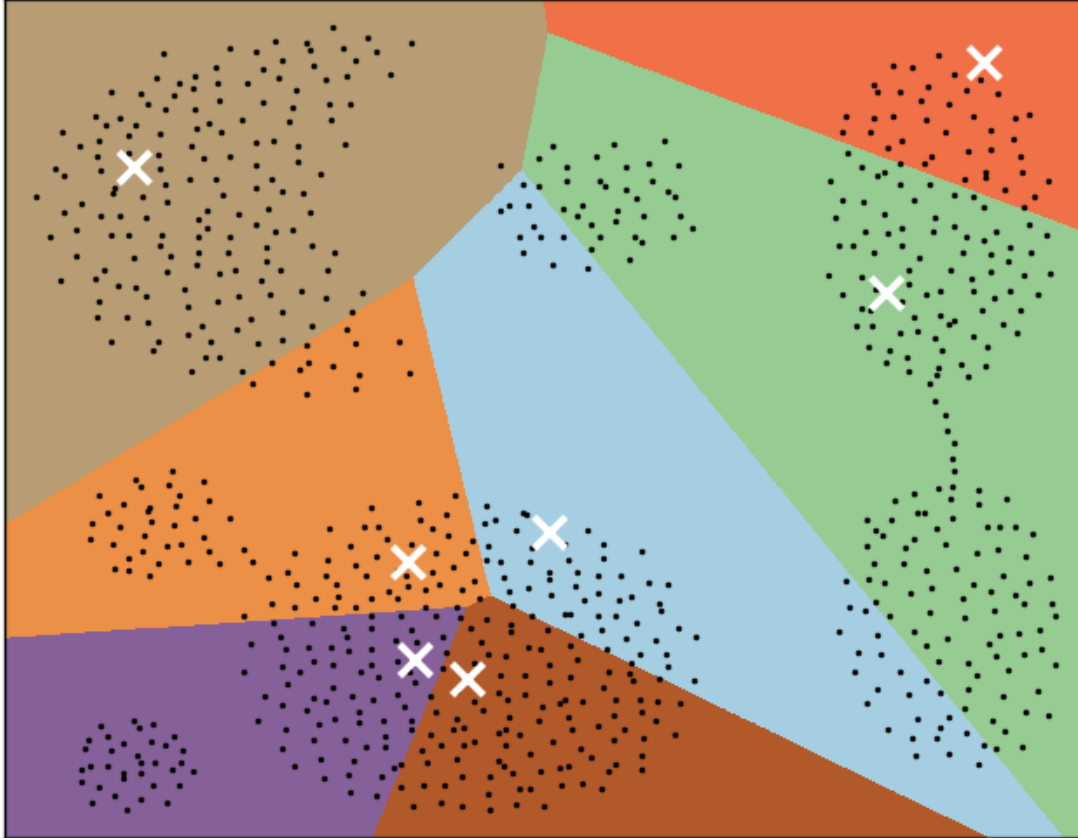
También se puede ver que los Sellers normalmente no tienen incentivos al comprador como descuentos o cuotas de pago.

Finalmente los sellers suelen no tener más de 5 categorías



5. Propuesta analítica : Clustering por Kmeans

KMeans with Iteration:0

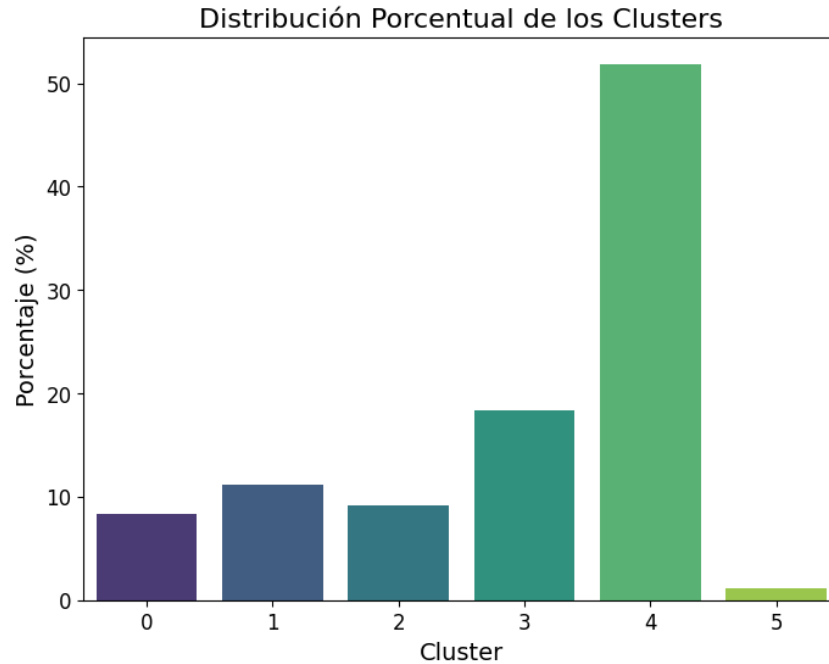


Este algoritmo tiene varios beneficios:

- **Rápida convergencia.**
- **Simple de interpretar.**
- **Bajos costos de entrenamiento.**
- **Fácilmente ajustable a necesidades de negocio.**
- **Escalable.**

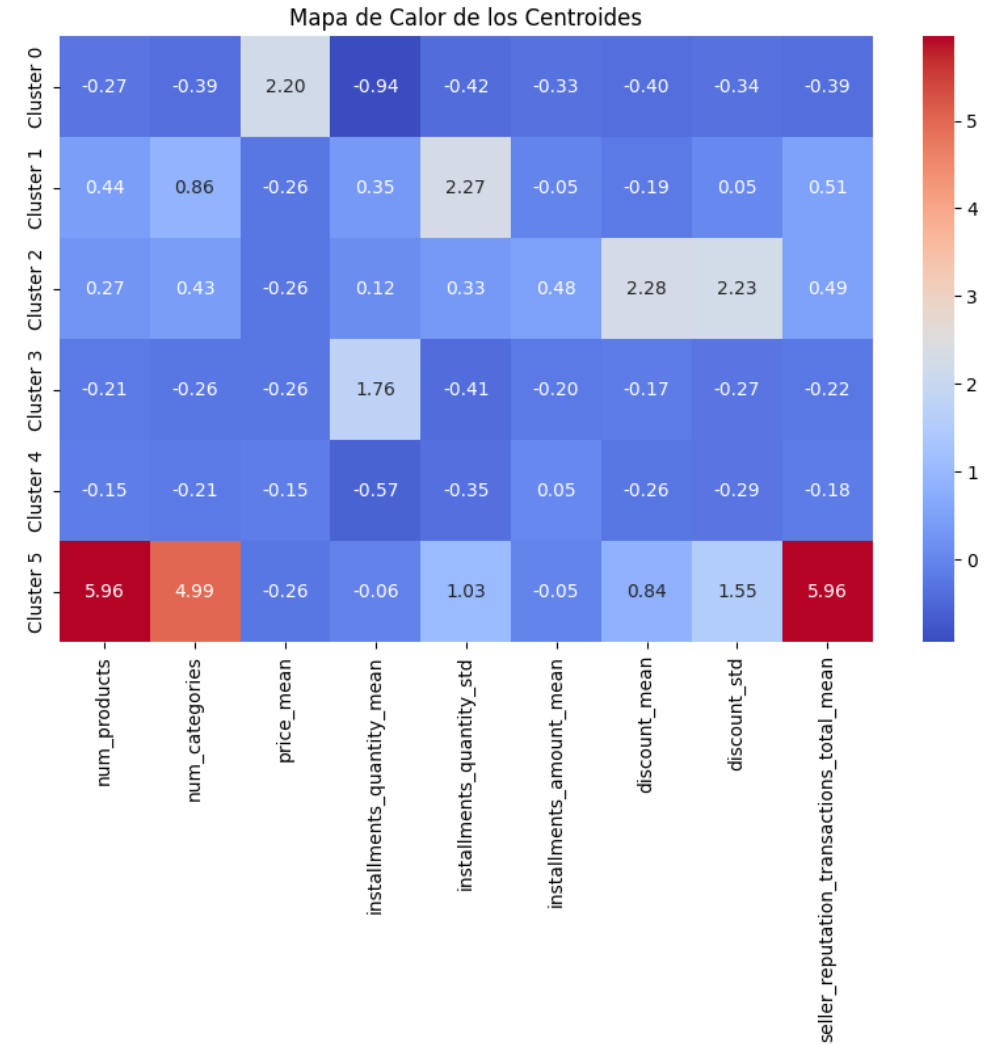
6. MVP del modelo

Participación de los grupos



cluster	0	1	2	3	4	5
Tamaño	463	623	509	1026	2891	62
Participación	9%	12%	10%	20%	57%	1%

Catcterización de los grupos*



6. MVP del modelo

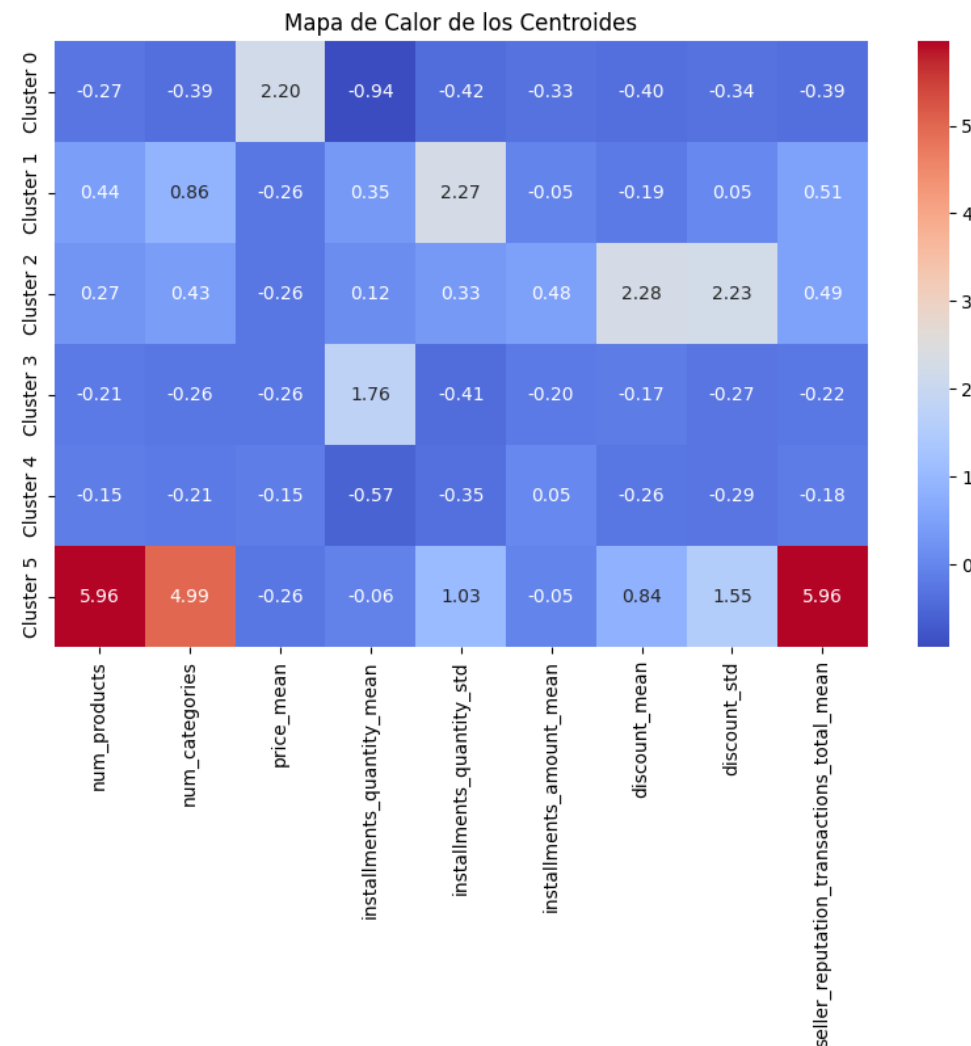
Metodología

Las unidades fueron reescaladas usando un **Standard Scaler*** para hacer comparable la información.

Los valores son más significativos cuanto más altos sean, ya sean positivos o negativos. Esto indica que ese clúster se aleja, por encima o por debajo, de la media en esa variable.

Ejemplo: **num_products** en el clúster 5 tiene 5.96 desviaciones por encima de la media, es decir, este clúster tiene un gran número de productos.

Catcterización de los grupos*



7. Caracterización de los grupos

Grupo 0: Sellers de celulares

Tiene un 8.30% de participación. Principalmente venden celulares, manejan precios altos (2.2 desviaciones por encima de la media) y ofrecen un número de cuotas menor (0.94 desviaciones por debajo de la media). No tienen reputación y no realizan campañas de marketing.

Grupo 2 Gancho por descuento.

Participación del 9.13%, más del 80% poseen campañas de marketing, cerca del 80% tienen medalla, destacan por ofrecer bastantes descuentos donde tienen 2.2 desviaciones sobre la media aunque sus precios no se diferencian mucho de los otros.

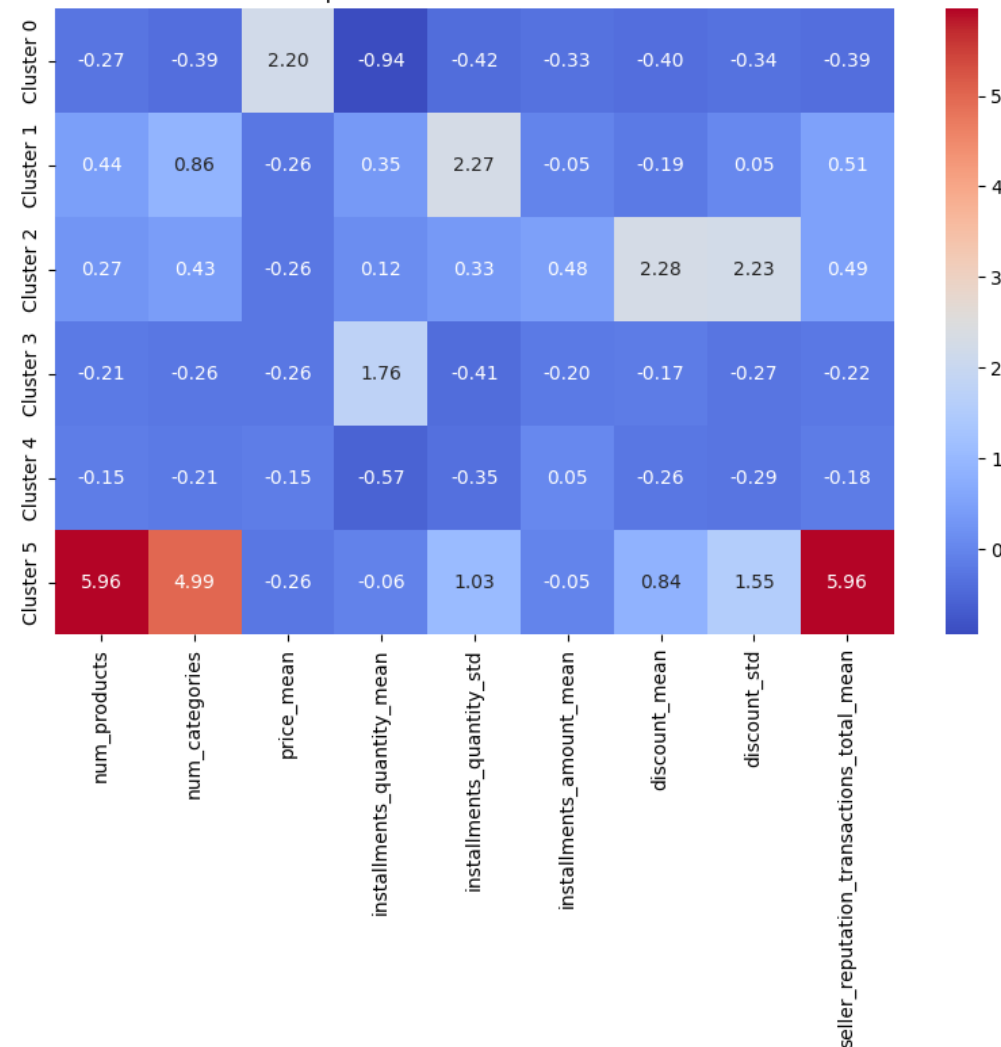
Grupo 1 Emprendedores.

Tiene un 11.17% de participación. Es un vendedor que ofrece una alta variedad de plazos de pago, con 2.25 desviaciones por encima de la media. Su catálogo es ligeramente variado, con 0.86 desviaciones por encima de la media. Menos del 20% de sus productos cuentan con campañas de marketing.

Grupo 3 Pago cómodo.

Participación del 18.40%. Tienen poca reputación, pero destacan por ofrecer facilidades de pago, con un alto número de cuotas disponibles.

Mapa de Calor de los Centroides



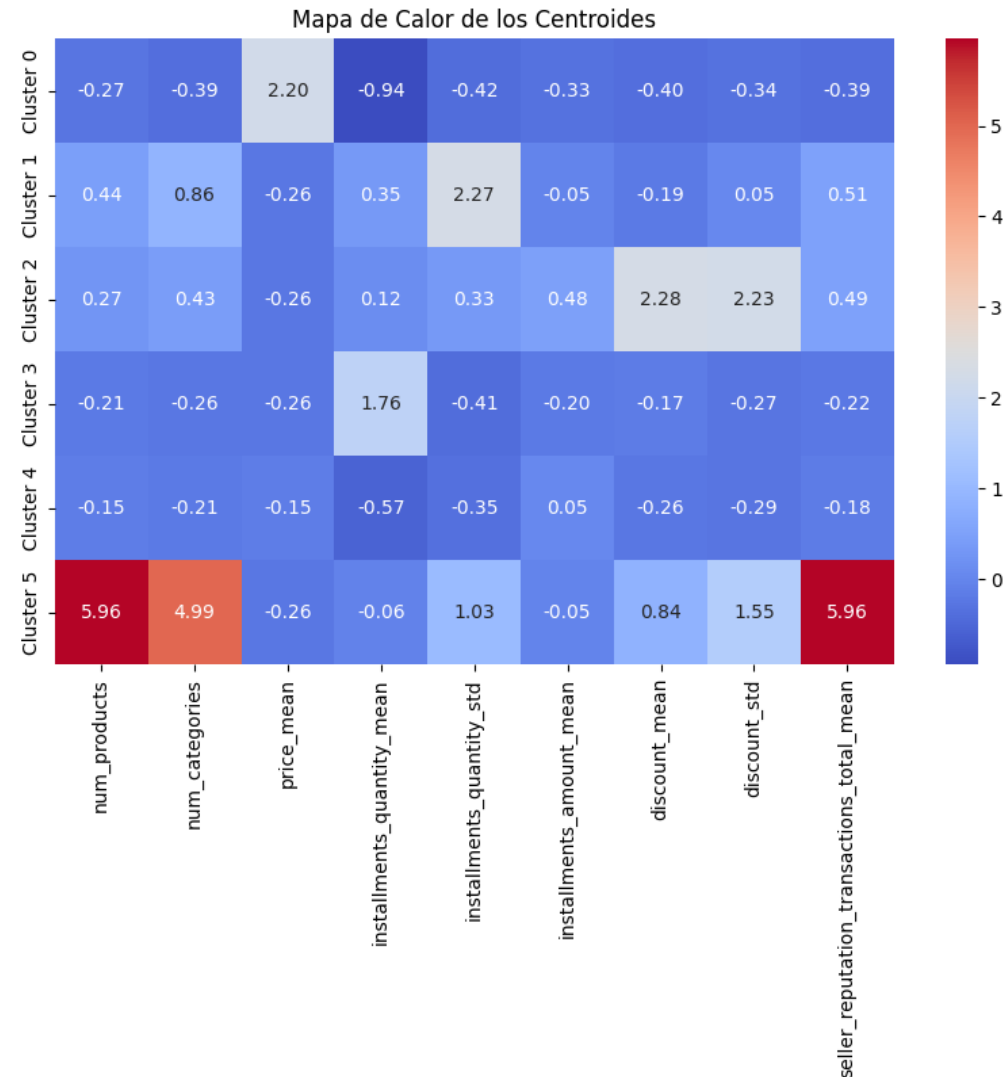
7. Caracterización de los grupos

Grupo 4 Usuario común

Tiene la mayor participación de los grupos con un 51.86%. Tienen poca reputación, y cerca del 60% no poseen medalla.

Grupo 5 Tiendas establecidas.

Representan el 1.1% de los sellers, destacándose con un número de productos, categorías y transacciones realizadas que están más de 5 desviaciones por encima de la media. Además, tienen una excelente reputación, ya que más del 90% poseen la calificación platinum.



7. Caracterización de los grupos: Propuestas



8. Proximos pasos

- **Definición y visualización de métricas de control por cada campaña.**
 - **Aumento en ventas para casuales y emprendedores.**
 - **Número de relaciones establecidas con bancos para el grupo de inmuebles.**
 - **Aumento de ventas producto del aumento del número de cuotas.**
- **Monitoreo del desempeño del modelo en el tiempo para generar alerta de cambio de centroides.**
- **Posible boost del modelo reentrenando con data enriquecida por nuevas variables product de tener retroalimentación con negocio.**