Challenge técnico: Mercado Libre

Santiago Dávila

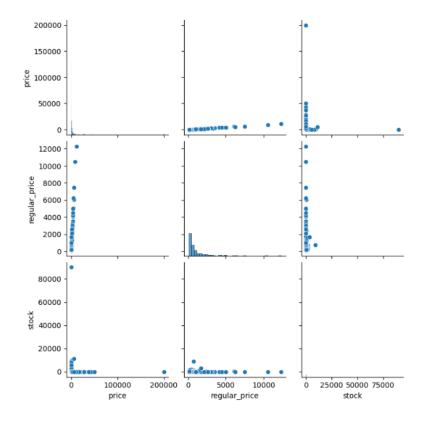
\_\_\_\_\_

**1. Problemática:** Falta de personalización en estrategias comerciales, lo que puede derivar en no desplegar todo el potencial de crecimiento en el Marketplace

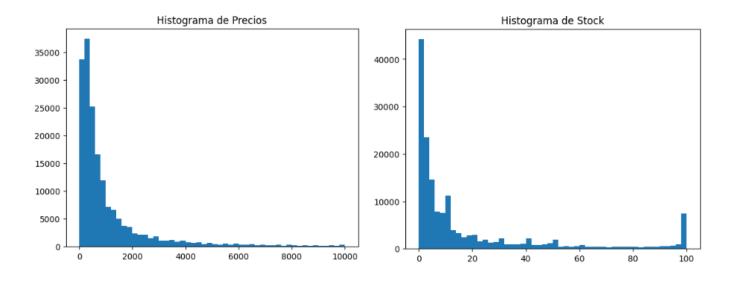
**Motivación:** Con el desarrollo de este reto se busca poder revisar el comportamiento de los diferentes sellers, para perfilarlos y personalizar estrategias que deriven no solo en mejores ventas para ellos, sino que además se dinamice mucho más el Marketplace, en términos de eficiencia e ingresos para Mercado Libre. Con esto se pueden elevar dos motores para la plataforma, por un lado, promover la sana competencia entre los vendedores elevando su reputación, y por otro, atraer nuevas gamas de consumidores que detecten que indirectamente se están viendo beneficiados por las nuevas estrategias creadas por el equipo comercial

- **2. Hipótesis:** Los sellers tienen características heterogéneas y se pueden agrupar en segmentos potenciando su comportamiento comercial
  - Fuente de datos: información de vendedores dentro del Marketplace, con categorías, precios, stock y su respectiva reputación
  - Revisión de variables que aportan valor a la solución

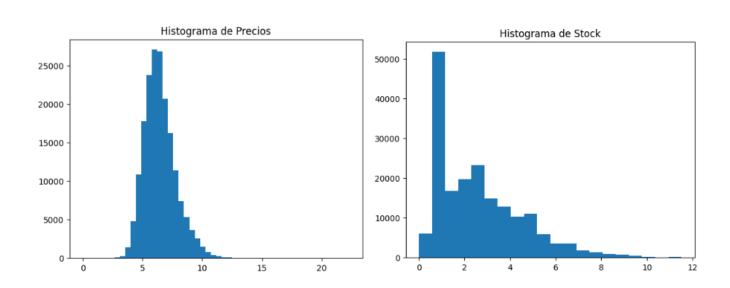
Haciendo el análisis exploratorio nos dimos cuenta con este pairplot que algo no estaba tan claro en la distribución de las variables numéricas



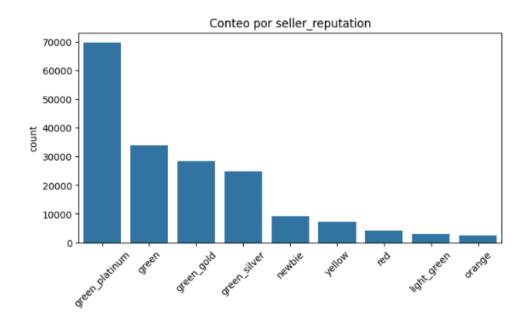
- Los histogramas representados en la diagonal datan la muestra de que era muy probable la existencia de outliers en cada una de las 3 variables
- Escaneamos las variables de precios y stock con una muestra para ver su concentración y ahí se comprobó que hay valores muy altos, lo cual impide ver una distribución típica en los datos tomando todo el conjunto



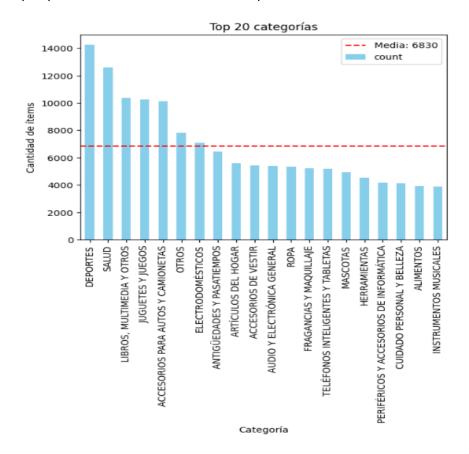
- Ambas variables se consideran muy importantes para la posterior ejecución del modelo
- Creamos dos variables adicionales, que son los logaritmos de precio y stock para mitigar el impacto de los outliers
- Histogramas de todo el conjunto con otra distribución



- Observación de la reputación de los seller, que puede también por el desbalanceo de clases tener un peso ponderado mayor sobre el modelo



- Adicionalmente, de las 54 categorías vimos el top 20 y la media de ese grupo para ver cuales estaban por encima y por debajo de la misma. Esto nos daba una luz de que algunos seller podrían estar en más de una categoría yendo en línea con la reputación, porque esa no cambia en función del producto ofrecido



# 3. Solución y Modelo

# Feature Engineering:

- Eliminamos variables que no aportan al análisis, como la fecha, el producto ofrecido, la url, etc.
- Establecimos un orden predeterminado para la reputación y categorizamos cada una con su respectivo número:

```
reputation_map = {
    'green_platinum': 10,
    'green_gold': 9,
    'green_silver': 8,
    'green': 7,
    'light_green': 6,
    'red': 5,
    'orange': 4,
    'yellow': 3,
    'newbie': 2,
    'otro': 1
}
```

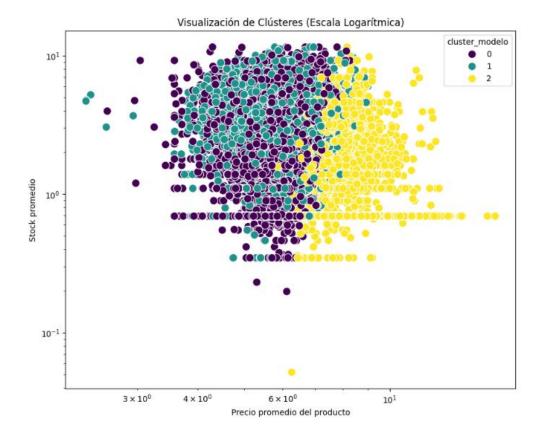
 Volvimos las otras variables categóricas en número, aunque a la hora de armar el dataframe final de esas conversiones solo usamos has\_discount, la cual se creó a partir de comparar price con regular\_price

# • Modelo K-MEANS

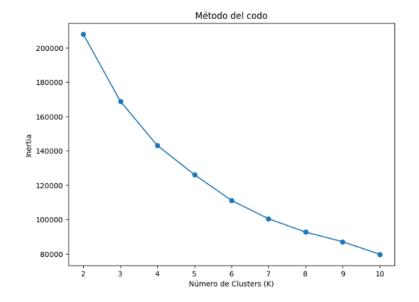
- DataFrame creado agrupando variables por cada vendedor con las respectivas variables numéricas

	seller_nickname	<pre>avg_prod_price</pre>	max_prod_price	avg_stock	unique_cat	reputation_score	has_discount_ratio
0	000631669c	6.684612	6.684612	2.397895	1	2.0	0.0
1	0007153bca	5.991465	5.991465	3.348517	1	7.0	0.0
2	000bee3c3b	5.761745	6.393591	0.000000	1	2.0	0.0
3	000df2bd02	7.346655	7.346655	1.791759	1	7.0	0.0
4	000e27cea2	6.060845	6.431331	1.354025	2	8.0	0.0

- Usamos StandardScaler para preservar las relaciones de distancia en el modelo que naturalmente es un algoritmo basado en la distancia euclidiana. Con esto acotamos el efecto de los outliers, transformando los datos con una distribución con media 0 y desviación estándar 1, en lugar de usar MinMaxScaler.
- Hicimos una prueba inicial de la segmentación con 3 clústers para ver el comportamiento y la media de cada variable usada como insumo



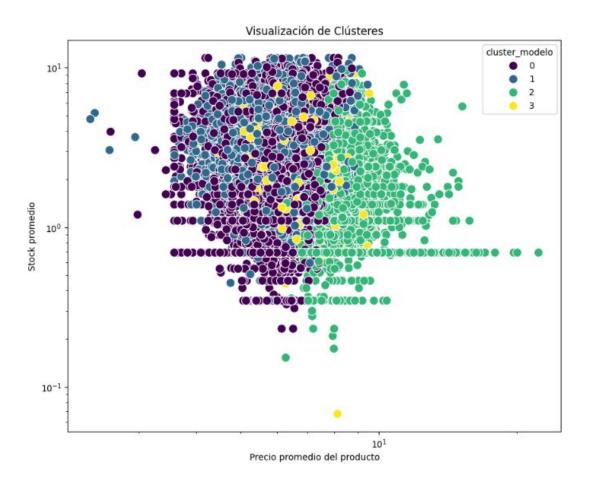
- Se observan los 3 grupos con una dinámica interdependiente entre el grupo 0 y 1, probablemente hay variables muy cercanas que distan de las características de otras como puede ser el precio y el stock
- Si bien, parece tener una buena distribución en el plano, decidimos usar el método del codo (Elba Method) en conjunto con Silhouette Score para determinar el número óptimo de clústers



k=2: Silhouette Score=0.247 k=3: Silhouette Score=0.267 k=4: Silhouette Score=0.286 k=5: Silhouette Score=0.275 k=6: Silhouette Score=0.275 k=7: Silhouette Score=0.273 k=8: Silhouette Score=0.262 k=9: Silhouette Score=0.261 k=10: Silhouette Score=0.266

- Como se aprecia en la gráfica, la inercia disminuye con cierta velocidad hasta k = 4, de ahí en adelante la pendiente se empieza a aplanar un poco, con lo cual usar un número mayor a 4 puede derivar en grupos menos compactos y poco diferenciados del resto
- Ese número óptimo que nos propone la gráfica también es consistente con el Silhouette Score con un coeficiente de 0.286 siendo el más grande en las pruebas realizadas. De cara a la propuesta de estrategias comerciales también es un hecho que es un buen punto de partida para el ejercicio, dado que en un número mayor podría generar cuellos de botella en la diferenciación y aplicación de las mismas.

### Modelo con clústers k = 4



- Observamos que los 4 grupos tienden hacia un espectro del plano, aunque el grupo 3 tuvo muy pocos sellers y tratan de converger hacia el medio de los otros 3, además tiene una alta varianza comparando estas dos variables.
- Con base en estos grupos, hicimos el siguiente perfilamiento comercial y propusimos diversas estrategias

Cluster	Precio Promedio	Stock Promedio	Variedad de Categorías	Reputación	% Descuentos
0	Medio-Bajo	Medio-Bajo	Baja	Media	Muy Bajo
1	Medio-Bajo	Alto	Media	Alta	Muy Alto
2	Muy Alto	Bajo	Baja	Baja	Muy Bajo
3	Medio-Alto	Alto	Muy Alta	Alta	Medio

## Clúster 0

- Se caracterizan por un precio promedio bajo, tienen stock moderado y una reputación media (entre light green y red)
- Son fuertes con su catálogo y casi no ofrecen descuentos
- Estrategias:
  - Incentivos y asesoría para manejar más stock, condiciones logísticas y mejora de la reputación
  - Ayudar a mejorar el mensaje para llegar a su nicho y maximizar ventas
  - Herramientas de comparación de precio con su competencia, para mejorar estructura financiera de sus ventas

## Clúster 1

- Este grupo tiene competencia directa en el volumen
- Stock alto, precios bajos y una reputación alta
- Estrategias:
  - Ofrecer servicios logísticos que maximicen el manejo del stock y puedan reducir costos
  - Buen acople en la comunicación de campañas masivas como Hot Sale, Black Friday, Ciber Monday, etc. priorizando su participación y disminuyendo comisiones
  - Herramientas financieras que ayuden en la optimización de su estrategia de precios, asegurando su rentabilidad sin comprometer stock ni reputación

## Clúster 2

- Precios promedio más altos, stock bajo, reputación baja y de vez en cuando ofrecen descuentos
- Estrategias:
  - Analizar el flujo de publicación, venta y envío para identificar los puntos débiles que pueden estar afectando su reputación y no tener tanto movimiento en sus productos
  - Implementación de seguimiento a la calidad de los productos, esto debido a que pueden ser productos de lujo y no sean objeto de reclamaciones
  - Prestar asesoría personalizada para volver el clúster mucho más competitivo cumpliendo las expectativas de los clientes

#### Clúster 3



- Pueden ser catalogados como vendedores generales, pero con una reputación que los remarca
- Manejan un stock amplio y se mueven por varias categorías y suelen implementar descuentos
- Estrategias:
  - Ofrecer herramientas de tracking en tiempo real para optimizar los descuentos y así maximizar sus ingresos
  - Brindar beneficios por su buen desempeño dentro del Marketplace, dando acceso prioritario a nuevas funcionalidades y percibir una disminución en las comisiones
  - Extraer buenas prácticas de fidelización y de crear una marca solida que puede ser implementado en otros grupos con la experiencia que este clúster ha construido

#### 3. Gen Al

- Opción A como catalizadora de aprovechar los insights derivados de la segmentación para nuevos vendedores
- Se manejó la solución con SentenceTransformer para vectorizar el comportamiento de los sellers, entrenando un modelo de regresión logística para la clasificación de los nuevos vendedores que lleguen al Marketplace y el equipo comercial pase directamente a aprovechar el potencial con las estrategias derivadas
- Con la entrada de nuevos vendedores la solución busca entender el comportamiento comercial, pasar esto a un vector numérico y posteriormente ser clasificado en alguno de los 4 clústers propuestos

text_for_embedding	cluster_modelo	has_discount_ratio	reputation_score	unique_cat	avg_stock	max_prod_price	avg_prod_price	seller_nickname	
El seller tiene un precio promedio de 6.684611	0	0.0	2.0	1	2.397895	6.684612	6.684612	000631669c	0
El seller tiene un precio promedio de 5.991464	2	0.0	7.0	1	3.348517	5.991465	5.991465	0007153bca	1
El seller tiene un precio promedio de 5.761744	0	0.0	2.0	1	0.000000	6.393591	5.761745	000bee3c3b	2
El seller tiene un precio promedio de 7.346655	0	0.0	7.0	1	1.791759	7.346655	7.346655	000df2bd02	3
El seller tiene un precio promedio de 6.060845	0	0.0	8.0	2	1.354025	6.431331	6.060845	000e27cea2	4

- En este caso la nueva columna text\_for\_embedding se encarga de agrupar el comportamiento comercial de cada seller y esto es lo que codifica como un vector numérico para entrenar el modelo de regresión

#### 4. Conclusiones

- La solución no se limitó únicamente a la creación del modelo k-means, requirió limpieza de datos, investigación y creación de nuevas features para armar un buen insumo
- La segmentación nos permitió identificar perfiles comerciales, explotando las características innatas de cada grupo
- Se pueden explorar más formas de segmentar la información con base en los resultados del método del codo, probar en caliente si otro número de clústers pueden aportar mejor al modelo
- El valor de la solución no se quedó en la parte predictiva, sino que entramos al terreno de la analítica prescriptiva con la aplicación de la inteligencia artificial generativa

### **Próximos pasos:**

- Afinar el funcionamiento de la solución de Gen AI y complementar con otras herramientas como el manejo del lenguaje natural para enfocar esfuerzos no solo en una buena clasificación, sino que amplie el espectro a recomendación de otras estrategias comerciales
- Una revisión continua del modelo, porque las estrategias pueden cambiar en función de las directrices del mercado y del comportamiento de los vendedores
- Integrar la solución en un módulo que el equipo comercial pueda consultar, que ofrezca retroalimentación permanente, buscando su adopción y hacerla escalable