

Meli - Data analytics Engineer challenge

by: Juan Camilo Daza

Problema

- Actualmente el equipo comercial no tiene una segmentación clara de sellers.
- Es difícil priorizar a quién dedicar tiempo, qué estrategia aplicar y cómo escalar decisiones

Potencial

- Segmentar correctamente permite enfocar recursos comerciales donde más impacto hay.
- Una clasificación clara + recomendaciones escalables habilitan decisiones masivas con lógica de negocio.

Producto

- Segmentar los sellers basados en tamaño y performance.
- Un Módulo de GenAI que genera una estrategia comercial personalizada por la clasificación del seller

Propuesta

- Llevar la segmentación a la acción, con estrategias generadas a escala.

Agenda

1. Problema de negocio y motivación
2. Datos, hipótesis y EDA
3. Diseño de la solución analítica
4. Resultados y métricas clave
5. Módulo GenAI (Opción B)
6. Conclusiones y próximos pasos



1. Problema Abordado y Motivación

El problema que quise resolver:

- Hoy el equipo comercial no cuenta con una clasificación clara y comparable dentro de su mismo grupo de pares
- hoy es difícil responder preguntas como:
 - **Que tipo de sellers hay? Como los estamos midiendo?**
 - **Cual es la mejor estrategia comercial para cada seller?**

El Enfoque que utilicé:

- Define una clasificación (clustering en dos niveles)
 - **Nivel 1 - Tamaño**
 - **Nivel 2 - Performance**
- De esta forma, cada seller se evalúa contra su propio segmento, no contra todo el universo (Ej: Un seller local no compete directamente con una gran marca)

Motivación de negocio:

- Hacer una clasificación más interpretable y accionable para negocio: Que un KAM pueda entender rápido "Que tipo de seller es" y "Como está performeando dentro de su grupo"
- Permitir crear un KPIs de seguimiento por seller y por segmento, por ejemplo
 - Objetivo que se aumente un x% **de que ciertos sellers pasen de un bajo performance a un medio o alto performance.**
- Usar esta base como punto de partida para
 - Definir **Metas claras** para el equipo comercial
 - **Recomendaciones al Seller** para que tenga un mejor performance y ventas
 - Construir **estrategias comerciales específicas** y campañas dirigidas por segmento

2. ANÁLISIS EXPLORATORIO - HIPOTESIS

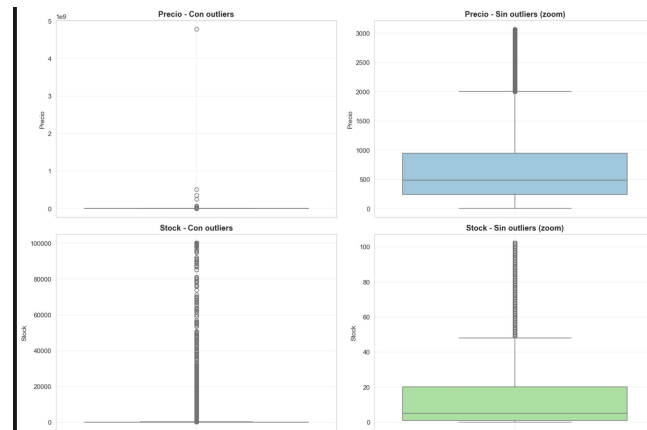
Partí de un análisis exploratorio profundo por variable del dataset original utilizando pandasprofiling (ver notebook o reporte HTML). Donde el foco estuvo en entender bien la distribución y calidad de estos tre ejes claves:

- Precio
- Stock
- Reputación
- $\text{Total_Value} = \text{stock} * \text{precio}$

Se detectaron valores atípicos importantes en:

- Precio: Colas muy largas con presencia de precios en cero
- Stock: Valores muy altos, poco consistentes con la realidad

Estados dos variables son críticas porque condicionan la lectura del tamaño del comportamiento del seller a nivel de ítem



box plot de Total value

2. ANÁLISIS EXPLORATORIO - HIPÓTESIS

Hipótesis Precio: los precios extremadamente altos o en cero no representan un comportamiento real del negocio.

- **Tratamiento aplicado:** **Eliminar** las observaciones con **precio igual a 0** y observaciones con precio por encima del **percentil 99** con el objetivo de quedarme con un rango de precios coherente y robusto.
 - Eliminar esta información sólo representaban el **1%** de los datos

Hipótesis Stock el stock es un dato auto reportado por el seller en la plataforma, por lo que puede estar sesgado (sobreestimado o infraestimado)

- **Tratamiento aplicado:** **Una normalización** de colas en la distribución de stock, reduciendo el peso de los valores extremos con el objetivo mantener la información de stock, pero controlando el impacto de decisiones individuales de carga de inventario.

Hipótesis Reputación: la relación entre seller y reputación es esencialmente 1 a 1.

- **Tratamiento aplicado:** En los casos con valores faltantes, **se imputó** la reputación del seller a partir de los valores disponibles del mismo vendedor (consistencia intra-seller)

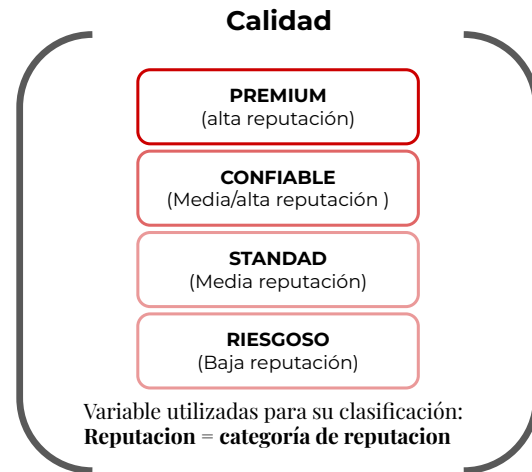
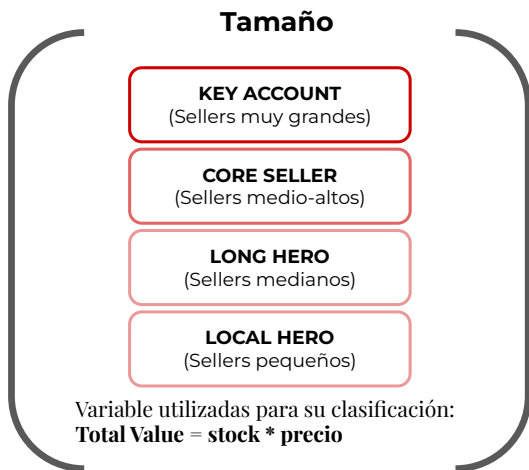
3. SOLUCIÓN TÉCNICA

Dado que desde un inicio siempre quise darle un enfoque más comercial y estratégico opté por hacer una solución **basada en Reglas** que fuera fácil de accionar para el equipo comercial.

Lo primero para hacer la solución fue:

1. Preparar la data (data_prep.py): Aca limpiamos la data previamente analizada en el EDA e hice la agregación por seller con otras variables.
2. Défini 3 ejes de segmentación (segmentation.py);
 - a. Tamaño: Que tan grande es el partner
 - b. Diversificación: Que tan diverso es su portafolio
 - c. Calidad: Que tan buena es la calidad de sus productos

Así quedaron las segmentaciones:



3. SOLUCIÓN TÉCNICA

- Seleccionamos el Eje de Tamano como **primer nivel de clasificación**
- Seleccionamos los ejes de **Diversificación y Calidad** para formar **El cluster de performance como segundo nivel**

La combinación de ambos niveles nos va a permitir generar accionables específicos dependiendo del tamaño y su performance.



Para el **nivel 2** se quedaron los siguientes segmentos:



3. SOLUCIÓN TÉCNICA

Para el **Nivel 2 (Performance)**, la categoría asignada depende del grupo del **Nivel 1 (Tamaño)**. En la práctica, combinamos ambos niveles, obteniendo hasta **16 segmentos posibles** (cada tamaño puede ser Diamante, Top, Expected o Low).

Esto significa que **no es lo mismo** lo que necesita un *Key Account* para ser Diamante que lo que requiere un *Local Hero* para alcanzar ese mismo nivel. Cada combinación *tamaño × performance* se trata como un segmento **independiente**, lo que permite:

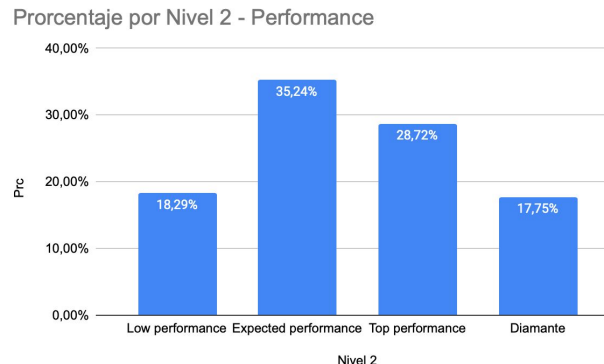
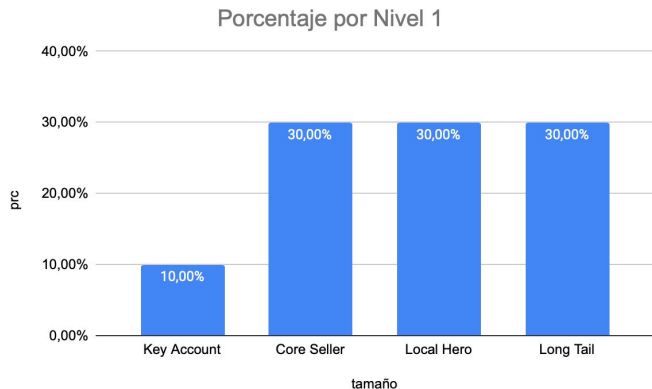
- Definir **estrategias mucho más específicas** por tipo de seller.
- Diseñar **KPIs de crecimiento y movilidad comercial**, por ejemplo, mover sellers de *Low* a *Expected* o de *Expected* a *Top* dentro de su propio grupo comparable.

Además, esta estructura hace que sea **muy sencillo incorporar nuevos ejes** (por ejemplo, foco en experiencia de cliente, servicio post-venta, etc.) si la estrategia comercial evoluciona, permitiendo **iterar y refinar** el modelo sin romper la lógica actual.



4. Resultados Obtenidos

Estos son los resultados obtenidos por tamaño y performance



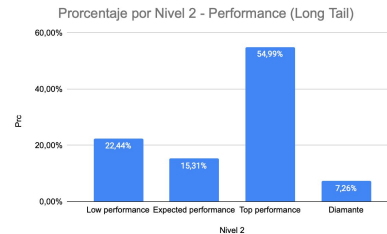
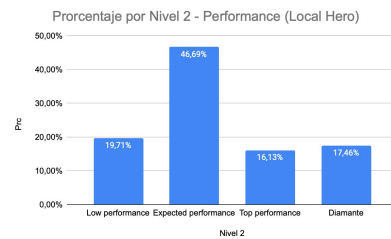
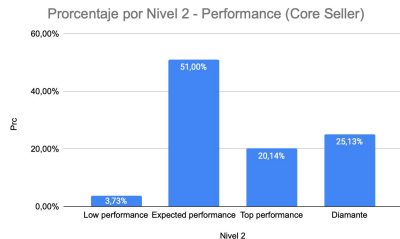
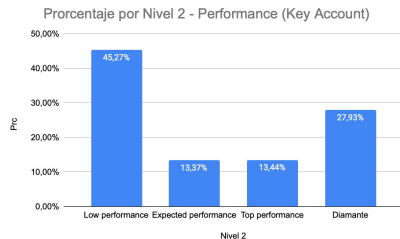
Los Key Accounts representan el grupo **más concentrado y estratégico**, mientras que seller core, Local Hero y Long tail forman tres bloques de volumen similar, lo que facilita comparaciones dentro de cada cohorte; por otro lado en el nivel 2, en conjunto, Diamante + top suman casi el **46% de los sellers**

Con esta foto global, ya sabemos **cómo luce la cartera por tamaño y cómo se distribuye el desempeño en total**.

El siguiente paso es combinar ambos niveles – *tamaño × performance* – para entender **qué tipos de sellers están aportando el mejor performance y dónde deberíamos enfocar las palancas comerciales por segmento**

4. Resultados Obtenidos

Resultados segmentado.



En el corte por tamaño se observa un comportamiento muy distinto en cada grupo. En **Key Account** la cartera está fuertemente polarizada: alrededor de **45,3%** de los sellers quedan en *Low performance*, solo **13,4%** en *Top performance* y **13,4%** en *Expected performance*, mientras que un **27,9%** alcanza nivel *Diamante*; es decir, conviven cuentas muy débiles con un bloque relevante de cuentas excelentes. En **Core Seller** la distribución es mucho más sana: apenas **3,7%** quedan en *Low performance*, la mayoría (**51,0%**) se ubica en *Expected performance*, un **20,1%** en *Top performance* y **25,1%** en *Diamante*, lo que convierte a este grupo en el “core gestionable” más equilibrado. En **Local Hero** el peso principal también está en *Expected performance* con **46,7%**, seguido de **19,7%** en *Low performance*, **16,1%** en *Top performance* y **17,5%** en *Diamante*, lo que sugiere un segmento de proximidad con mucho espacio para desarrollo hacia arriba. Finalmente, en **Long Tail** domina claramente *Top performance* con **55,0%** de los sellers, mientras que **22,4%** se ubica en *Low performance*, **15,3%** en *Expected performance* y solo **7,3%** llega a *Diamante*; aquí aparecen muchos pequeños “campeones ocultos” que ya performan bien y son candidatos naturales a estrategias escalables de crecimiento.

5. Módulo GenAI

Para la extensión de GenAI seleccioné la **Opción B – Recomendador Generativo de Estrategias Comerciales**. El principal dolor del negocio no es solo clasificar sellers, sino **traducir esa segmentación en acciones concretas y escalables** para el equipo comercial. Una vez resuelto el mapa de segmentos (tamaño + performance), tiene sentido usar un LLM como **capa de orquestación** que, apoyado en un playbook por segmento, proponga qué hacer con cada tipo de seller.

En esta versión, el modelo genera para cada seller una estrategia estructurada (objetivo, palancas sugeridas, ideas de campaña y KPIs). El comportamiento es consistente con la información disponible: entrega recomendaciones generales pero accionables, que el equipo comercial puede adaptar y ejecutar. Con más señal de negocio (p. ej. ciudad, categoría principal, datos históricos de ventas, performance en el portal) el recomendador podría ser mucho más preciso y específico. Aun así, para el challenge demuestra cómo, a partir de la segmentación, GenAI puede **cerrar el gap entre análisis y acción comercial**, generando estrategias personalizadas a escala.

Muy brevemente, el flujo técnico es: el pipeline genera seller_profile.csv, luego el módulo GenAI (carpeta src/meli_challenge/genai/) toma cada fila, playbook.py determina las reglas base por segmento, prompt_builder.py construye el prompt y strategy_generator.py llama al modelo de OpenAI para devolver la estrategia, que se consolida en data/outputs/strategies_sample.csv para ser consumida por negocio.

6.1. Conclusiones

La solución entrega un **mapa claro de la cartera de sellers**, combinando tamaño y desempeño (performance) para que cada vendedor sea evaluado dentro de su grupo comparable. Sobre esta base, el módulo GenAI añade una **capa de acción**, generando estrategias comerciales por segmento que permiten pasar de un ejercicio puramente analítico a un **paquete completo: clasificación + recomendación**. En términos de negocio, esto facilita que perfiles no técnicos interpretan rápidamente “qué tipo de seller tengo delante” y “qué palancas puedo accionar” para capturar más valor.

6.2. Próximos pasos

El siguiente paso natural es **socializar este modelo con el equipo comercial**, validando que los segmentos y ejemplos de estrategias reflejen su realidad del día a día. A partir de allí, definir un **baseline** claro (distribución actual por performance dentro de cada tamaño) y fijar **KPIs de movilidad**: por ejemplo, qué porcentaje de sellers queremos mover de *Low* a *Expected* o de *Expected* a *Top* en un periodo determinado. Finalmente, acompañar este marco con un **catálogo de acciones sugeridas por segmento** (salida del GenAI) para que cada KAM o ejecutivo comercial tenga un punto de partida concreto al trabajar con sus cuentas. De esta forma, la segmentación deja de ser solo un diagnóstico y se convierte en una herramienta operativa para alinear a analytics, negocio y fuerza comercial alrededor de un mismo objetivo: **eleva el nivel de los sellers y, con ello, el desempeño del marketplace en su conjunto**.



Repositorio:

<https://github.com/camilo-daza-ac/Mercado-Libre-Data-Analytics-Engineer-Challenge>

MUCHAS GRACIAS.

Juan Camilo Daza
